New York State Testing Program 2013: English Language Arts Mathematics Grades 3–8

Technical Report

Pearson
2013

Developed and published under contract with the New York State Education Department by Pearson, 2510 North Dodge Street, Iowa City, Iowa 52245.

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 2013 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, standard-setting, scoring, scaling, and student performance.

Test Purpose

The federal No Child Left Behind Act requires that states annually administer tests in English Language Arts and Mathematics in grades 3 – 8. The 2013 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 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, "Test Design and Development," subsection "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. Nonpublic schools may participate in the testing program, but their participation is not mandatory. In 2013, some nonpublic 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 students with severe disabilities. 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-samcc13rev.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 Common Core Learning Standards in ELA and Mathematics, respectively. 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 Scaling." 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.

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.* Tables 1 and 2 below present the reporting subscore categories and the point values that correspond to each on the 2013 tests. In 2013, 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	Reporting Subscores	
	Reading	Writing to Sources
3	31	24
4	31	24
5	42	24
6	41	24
7	42	24
8	42	24

Table 2. Mathematics Subscore Categories and Total Possible Score Points

Grade	Reporting Subscores		
	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 17	Number and Operations—Fractions 26	Measurement and Data 9
6	Ratios and Proportional Relationships 19	The Number System 11	Expressions and Equations 28
7	Ratios and Proportional Relationships 20	The Number System 14	Expressions and Equations 22
8	Expressions and Equations 30	Functions 11	Geometry 14

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 Education Research Association, American Psychological Association, and National Council on Measurement in Education, 1999), 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 2013 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 2013 Grades 3–8 Common Core ELA and Mathematics Tests are criterion-referenced tests comprised 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 2013 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 online at: http://www.p12.nysed.gov/assessment/sam/ei-samcc13rev.pdf

ELA Tests

The 2013 Grade 3–8Common Core English Language Arts Tests were designed to measure student literacy as defined by the CCLS. The tests assessed Reading, Writing, and Language standards using multiple-choice, short-response, and extended-response questions. All questions were based on close reading 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 answer to an inferential question. 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 2013 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 both Standards and Standards for Mathematical Practices, 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 mainly used 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 2013 Grades 3–8 Common Core ELA Tests were comprised of four books per grade and administered in three sessions over three days. Day 1 consisted of Book 1. Day 2 consisted of Books 2 and 3, completed by the student in one continuous session. Day 3 consisted of Book 4. Book 1 and Book 2 contained literary and informational reading passages and MC items based on the passages. Book 3 and Book 4 contained reading passages with short-response items and an extended-response item based on those passages.

The 2013 Grades 3–8 Common Core Mathematics Tests were comprised 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 2013 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 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 2013 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;
- 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, 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 comment 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 (i.e., 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 2013 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 CCLS 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 it could be used to measure the CCLS for ELA. The criteria documents were used to determine whether each passage suggested for testing use was grade appropriate, fair, and possessed the necessary characteristics to assess each standard. Specifically, passages were evaluated for the presence and quality of key ideas and details, craft and structure, and integration of knowledge and ideas. The full passage review criteria can 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 impact a student's ability to perform the required task. For example, the criteria include checking to make sure the vocabulary of test items are on grade level and that questions avoid technical terms unrelated to the content. Likewise, the Fairness criteria are used to ensure questions are un-biased, non-offensive, and not disadvantageous to any given subgroup. The criteria also address how each item measures a given standard or standards, and articulates the aspects of each standard that the items need to address. Finally, the criteria establish key requirements for each item type, 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 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 CCLS, 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 questions containing art are reasonable and that interpreting a graphic does not confuse the underlying construct. Likewise, the Fairness criteria are used to evaluate whether questions are un-biased, non-offensive, and not disadvantageous to any given subgroup(s). 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 items involve a multi-step processes 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 the tests measured the deep conceptual understanding that CCLS demands rather than focusing on predictable math 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 math standards could be measured within the constraints of 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 using a particular representation and identifies types of math skills (e.g. application of process; explanation of a principle, etc.) that are appropriate to assess 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, principals 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 CCLS-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 copyeditors, fact checked by research librarians, reviewed for subject-specific content issues by science and social studies content specialists, and reviewed for universal-design issues by specifically trained reviewers. After the passages went through these review steps, ELA content specialists posted the passages and completed criteria documents for NYSED's review and approval for moving forward in the process.

NYSED staff retrieved 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, CCLS-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 they were writing items, experience in writing for large-scale assessments, and 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 where Pearson content staff and NYSED staff reviewed and edited all of the items to meet 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 also attended by NYSED staff. For ELA, reviewers first read and discussed the passages. For Mathematics and ELA the educators then 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) 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 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 2012 Field Test Items

Grade	Unique ELA Items by Type		Unique Mathematics Items by Type	
	Multiple-Choice	Constructed-Response	Multiple-Choice	Constructed-Response
3	86	23	74	22
4	82	25	79	24
5	80	15	86	27
6	81	19	83	27
7	94	22	81	23
8	85	22	81	22

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 2012 as embedded field test items within the 2012 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. Two additional rounds of field testing were administered separately from the 2012 operational forms (i.e., as stand-alone tests) – one later in the spring of 2012 and another in the fall of 2012.

In order to better understand how 2012 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 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 2012 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 2013. 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 2012.

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 to 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 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 Pearson content and psychometric teams were satisfied that the content and statistics of the selected items, and 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 January 2013 to finalize item selection and test creation with the 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, 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 2013 administration.

Proficiency and Performance Standards

In summer 2013, after the operational administration of the 2013 tests, a standard-setting meeting occurred in Albany where 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.

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, "IRT Calibration and Scaling."

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* (1999) 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 for 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 purpose 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 1999 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 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* (1999). HumRRO also evaluated the degree to which the 2013 Grades 3–8 Common Core ELA and Mathematics Tests measured the CCLS. HumRRO concluded that the 2013 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, "Reliability and Standard Error of Measurement." For the total population, the ELA reliability coefficients (Cronbach's alpha) ranged from .90–.92, and for all subgroups, the reliability coefficients were greater than or equal to .83. For the total population, the mathematics reliability coefficients (Cronbach's alpha) ranged from .93–.94, and for all subgroups, the reliability coefficients were greater than or equal to .82. 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 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 Scaling 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 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 MC and CR data. The study was conducted on the total population of New York State public and charter school students in each grade. A large first principal component was evident in each analysis, demonstrating essential unidimensionality of the trait measured by each test. In other words, statistical evidence indicates the ELA items are measuring one underlying construct, ELA proficiency, and 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 the grades.

It was found that all the New York State Grades 3–8 Common Core ELA and Mathematics Tests exhibited first principal component accounting for more than 18% and 21% 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. Factor Analysis Results for ELA Tests (Total Population)

	Initial Eigenvalues			
Grade	Component	Total	% of Variance	Cumulative %
3	**1**	**8.65**	**21.09**	**21.09**
	2	1.51	3.67	24.76
	3	1.20	2.93	27.69
	4	1.08	2.64	30.33
	5	1.04	2.54	32.87
4	**1**	**8.89**	**21.69**	**21.69**
	2	1.52	3.70	25.39
	3	1.13	2.75	28.14
	4	1.03	2.50	30.64
5	**1**	**9.75**	**18.74**	**18.74**
	2	1.39	2.68	21.42
	3	1.17	2.25	23.67
	4	1.11	2.13	25.80
	5	1.07	2.05	27.85
	6	1.03	1.98	29.83
6	**1**	**10.27**	**20.14**	**20.14**
	2	1.59	3.12	23.26
	3	1.27	2.49	25.75
	4	1.03	2.02	27.77
	5	1.01	1.99	29.76
7	**1**	**9.98**	**19.19**	**19.19**
	2	1.52	2.93	22.12
	3	1.22	2.36	24.48
	4	1.13	2.17	26.65
	5	1.06	2.04	28.69
	6	1.02	1.95	30.64
	7	1.00	1.93	32.57
8	**1**	**9.49**	**18.25**	**18.25**
	2	1.60	3.08	21.33
	3	1.14	2.19	23.52
	4	1.10	2.12	25.64
	5	1.04	2.00	27.64
	6	1.00	1.93	29.57

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. Factor Analysis Results for Mathematics Tests (Total Population)

	Initial Eigenvalues			
Grade	Component	Total	% of Variance	Cumulative %
3	**1**	**12.24**	**24.98**	**24.98**
	2	1.74	3.54	28.52
	3	1.27	2.58	31.10
	4	1.13	2.31	33.41
	5	1.02	2.08	35.49
4	**1**	**12.03**	**23.14**	**23.14**
	2	1.97	3.78	26.92
	3	1.25	2.40	29.32
	4	1.10	2.11	31.43
	5	1.06	2.03	33.46
	6	1.04	2.00	35.46
	7	1.02	1.96	37.42
5	**1**	**12.07**	**23.21**	**23.21**
	2	2.11	4.07	27.28
	3	1.24	2.38	29.66
	4	1.03	1.97	31.63
	5	1.01	1.94	33.57
6	**1**	**13.28**	**22.89**	**22.89**
	2	2.14	3.69	26.58
	3	1.29	2.22	28.80
	4	1.03	1.78	30.58
	5	1.00	1.73	32.31
7	**1**	**13.13**	**22.64**	**22.64**
	2	1.78	3.07	25.71
	3	1.33	2.30	28.01
	4	1.11	1.92	29.93
	5	1.00	1.73	31.66
8	**1**	**12.56**	**21.66**	**21.66**
	2	1.92	3.32	24.98
	3	1.20	2.07	27.05
	4	1.06	1.83	28.88
	5	1.00	1.73	30.61

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 that are classified with disabilities that 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 large-scale 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 by experts of those flagged items, 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 impact 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 (2013)* located here: http://www.p12.nysed.gov/assessment/sam/scoringleaderhandbook13.pdf.

Test Administration

NYSTP Grades 3–8 Common Core ELA and Mathematics Tests were administered to students during April 2013. The testing window was April 16–18 for the Grades 3–8 Common Core ELA Tests and April 24–26 for the Grades 3–8 Common Core Mathematics Tests. The makeup test administration window was April 19, 22, and 23 for the Grades 3–8 Common Core ELA Tests and April 29–May 1 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 test.

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 detail). 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 Agreement.

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 2012–2013 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 nonpublic schools in an affiliation group (nonpublic 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, nonpublic 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 nonpublic school.

Schools and districts were instructed to carefully analyze their individual needs and capacities to determine their appropriate scoring model. Boards of Cooperative Educational Services (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 were 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 Agreement 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 "Assign Scorer Numbers and Questions to Scoring Committee Members" on page 31, found online at:
http://www.p12.nysed.gov/assessment/sam/ei-samcc13rev.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 98% of the student test records were received from the data warehouse and delivered to Pearson at the end of May 2013. Phase 2 involved submitting "straggler files" to Pearson in June 2013.

The straggler files contained less than 2% of the total population cases and were excluded from the classical, IRT, and reliability analyses (as described in the Sections 5, 6, and 7, respectively) due to late submission. The analyses described in the Section 9 "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 a valid test score were included in the analyses described in the Section 9 "Summary of Operational Test Results." For the analyses described in other sections, however, more stringent data confirmation procedures were applied (see details below).

Data processing here refers to the confirmation 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 confirmation 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 student records so that their test results were scored properly. As mentioned above, there student records were included in the Section 9 analyses.

The major groups of cases excluded from the data set (used for analyses in Sections 5, 6, and 7) were students with incorrect or incomplete grade information. Other deleted cases included schools with a school type other than public, charter, and non-public; duplicate record cases; and no-response record cases. The mathematical data confirmation procedure also excluded records with mismatched form language indicators for translated versions across the three test books for a given student. The data confirmation procedures and accompanying case counts are represented for ELA and mathematics in Tables 6A–6F and 7A–7F, respectively.

Table 6A. ELA Data Confirmation, Grade 3

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		201,065
Wrong Subject	0	201,065
No Grade	3,043	198,022
Wrong Grade	78	197,944
Language Mismatched Form	3	197,941
School Type	524	197,417
No Response	3	197,414
Invalid Score	0	197,414
Out-of-Range CR Scores	1	197,413
Duplicated Record	8	197,405

Table 6B. ELA Data Confirmation, Grade 4

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		207,878
Wrong Subject	0	207,878
No Grade	2,673	205,205
Wrong Grade	89	205,116
Language Mismatched Form	4	205,112
School Type	551	204,561
No Response	1	204,560
Invalid Score	0	204,560
Out-of-Range CR Scores	0	204,560
Duplicated Record	4	204,556

Table 6C. ELA Data Confirmation, Grade 5

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		198,220
Wrong Subject	0	198,220
No Grade	2,675	195,545
Wrong Grade	25	195,520
Language Mismatched Form	4	195,516
School Type	634	194,882
No Response	0	194,882
Invalid Score	0	194,882
Out-of-Range CR Scores	1	194,881
Duplicated Record	4	194,877

Table 6D. ELA Data Confirmation, Grade 6

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		205,954
Wrong Subject	0	205,954
No Grade	3,071	202,883
Wrong Grade	116	202,767
Language Mismatched Form	3	202,764
School Type	732	202,032
No Response	6	202,026
Invalid Score	0	202,026
Out-of-Range CR Scores	0	202,026
Duplicated Record	8	202,018

Table 6E. ELA Data Confirmation, Grade 7

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		204,105
Wrong Subject	0	204,105
No Grade	2,033	202,072
Wrong Grade	115	201,957
Language Mismatched Form	4	201,953
School Type	970	200,983
No Response	1	200,982
Invalid Score	0	200,982
Out-of-Range CR Scores	0	200,982
Duplicated Record	0	200,982

Table 6F. ELA Data Confirmation, Grade 8

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		210,449
Wrong Subject	0	210,449
No Grade	1,631	208,818
Wrong Grade	102	208,716
Language Mismatched Form	3	208,713
School Type	1,227	207,486
No Response	2	207,484
Invalid Score	0	207,484
Out-of-Range CR Scores	0	207,484
Duplicated Record	8	207,476

Table 7A. Mathematics Data Confirmation, Grade 3

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		202,869
Wrong Subject	0	202,869
No Grade	126	202,743
Wrong Grade	84	202,659
Language Mismatched Form	32	202,627
School Type	522	202,105
No Response	2	202,103
Invalid Score	0	202,103
Out-of-Range CR Scores	0	202,103
Duplicated Record	8	202,095

Table 7B. Mathematics Data Confirmation, Grade 4

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		211,053
Wrong Subject	0	211,053
No Grade	98	210,955
Wrong Grade	111	210,844
Language Mismatched Form	17	210,827
School Type	564	210,263
No Response	0	210,263
Invalid Score	0	210,263
Out-of-Range CR Scores	0	210,263
Duplicated Record	4	210,259

Table 7C. Mathematics Data Confirmation, Grade 5

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		200,966
Wrong Subject	0	200,966
No Grade	192	200,774
Wrong Grade	35	200,739
Language Mismatched Form	42	200,697
School Type	656	200,041
No Response	0	200,041
Invalid Score	0	200,041
Out-of-Range CR Scores	0	200,041
Duplicated Record	4	200,037

Table 7D. Mathematics Data Confirmation, Grade 6

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		208,355
Wrong Subject	0	208,355
No Grade	111	208,244
Wrong Grade	138	208,106
Language Mismatched Form	48	208,058
School Type	776	207,282
No Response	2	207,280
Invalid Score	0	207,280
Out-of-Range CR Scores	0	207,280
Duplicated Record	10	207,270

Table 7E. Mathematics Data Confirmation, Grade 7

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		205,726
Wrong Subject	0	205,726
No Grade	104	205,622
Wrong Grade	133	205,489
Language Mismatched Form	41	205,448
School Type	997	204,451
No Response	0	204,451
Invalid Score	0	204,451
Out-of-Range CR Scores	0	204,451
Duplicated Record	4	204,447

Table 7F. Mathematics Data Confirmation, Grade 8

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases		209,923
Wrong Subject	0	209,923
No Grade	111	209,812
Wrong Grade	105	209,707
Language Mismatched Form	56	209,651
School Type	1,272	208,379
No Response	0	208,379
Invalid Score	0	208,379
Out-of-Range CR Scores	0	208,379
Duplicated Record	8	208,371

Classical Analysis and Calibration Sample Characteristics

The demographic characteristics of students in the cleaned calibration and equating 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 clean data sets included over 95% of New York State students and were used for classical analyses presented in the calibrations in this section. 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	96,650	48.96
	Male	100,755	51.04
Ethnicity	Asian	16,988	8.61
	Black	34,259	17.35
	Hispanic	46,873	23.74
	American Indian/Native	1,070	0.54
	Multiracial	2,723	1.38
	Pacific Islander	432	0.22
	White	95,060	48.15
NRC	New York	68,598	34.75
	Big 4 Cities	7,896	4.00
	Urban/Suburban	15,098	7.65
	Rural	10,851	5.50
	Average Needs	54,931	27.83
	Low Needs	26,776	13.56
	Charter	7,180	3.64
	Non-Public	6,075	3.08
SWD	No	170,486	86.36
	Yes	26,919	13.64
SUA	No	149,896	75.93
	Yes	47,509	24.07
ELL	No	179,982	91.17
	Yes	17,423	8.83

Note. The total n-count was 197,405.

Table 8B. ELA Grade 4 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	100,767	49.26
	Male	103,789	50.74
Ethnicity	Asian	16,880	8.25
	Black	35,059	17.14
	Hispanic	47,461	23.2
	American Indian/Native	1,109	0.54
	Multiracial	2,409	1.18
	Pacific Islander	455	0.22
	White	101,183	49.46
NRC	New York	67,244	32.87
	Big 4 Cities	7,854	3.84
	Urban/Suburban	15,131	7.4
	Rural	10,120	4.95
	Average Needs	55,648	27.2
	Low Needs	27,330	13.36
	Charter	5,440	2.66
	Non-Public	15,789	7.72
SWD	No	175,782	85.93
	Yes	28,774	14.07
SUA	No	157,190	76.84
	Yes	47,366	23.16
ELL	No	190,065	92.92
	Yes	14,491	7.08

Note. The total n-count was 204,556.

Table 8C. ELA Grade 5 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	96,074	49.30
	Male	98,803	50.70
Ethnicity	Asian	16,206	8.32
	Black	34,547	17.73
	Hispanic	44,455	22.81
	American Indian/Native	1,087	0.56
	Multiracial	2,035	1.04
	Pacific Islander	323	0.17
	White	96,224	49.38

Table 8C. ELA Grade 5 Sample Characteristics (cont.)

Demographic Category		N-count	% of Total N-count
NRC	New York	65,485	33.60
	Big 4 Cities	7,662	3.93
	Urban/Suburban	14,204	7.29
	Rural	10,800	5.54
	Average Needs	56,017	28.74
	Low Needs	27,871	14.30
	Charter	6,701	3.44
	Non-Public	6,137	3.15
SWD	No	165,208	84.78
	Yes	29,669	15.22
SUA	No	147,978	75.93
	Yes	46,899	24.07
ELL	No	180,798	92.78
	Yes	14,079	7.22

Note. The total n-count was 194,877.

Table 8D. ELA Grade 6 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	99,494	49.25
	Male	102,524	50.75
Ethnicity	Asian	16,230	8.03
	Black	36,561	18.10
	Hispanic	46,324	22.93
	American Indian/Native	1,055	0.52
	Multiracial	1,880	0.93
	Pacific Islander	400	0.20
	White	99,568	49.29
NRC	New York	64,497	31.93
	Big 4 Cities	7,753	3.84
	Urban/Suburban	14,353	7.10
	Rural	10,911	5.40
	Average Needs	56,415	27.93
	Low Needs	28,131	13.92
	Charter	6,033	2.99
	Non-Public	13,925	6.89
SWD	No	172,551	85.41
	Yes	29,467	14.59
SUA	No	157,936	78.18
	Yes	44,082	21.82
ELL	No	191,381	94.73
	Yes	10,637	5.27

Note. The total n-count was 202,018.

Table 8E. ELA Grade 7 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	98,469	48.99
	Male	102,513	51.01
Ethnicity	Asian	16,863	8.39
	Black	36,115	17.97
	Hispanic	44,051	21.92
	American Indian/Native	1,000	0.5
	Multiracial	1,803	0.9
	Pacific Islander	375	0.19
	White	100,775	50.14
NRC	New York	66,631	33.15
	Big 4 Cities	7,549	3.76
	Urban/Suburban	14,368	7.15
	Rural	11,364	5.65
	Average Needs	58,872	29.29
	Low Needs	31,472	15.66
	Charter	5,251	2.61
	Non-Public	5,475	2.72
SWD	No	171,361	85.26
	Yes	29,621	14.74
SUA	No	158,461	78.84
	Yes	42,521	21.16
ELL	No	190,902	94.98
	Yes	10,080	5.02

Note. The total n-count was 200,982.

Table 8F. ELA Grade 8 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	101,958	49.14
	Male	105,518	50.86
Ethnicity	Asian	16,500	7.95
	Black	37,812	18.22
	Hispanic	45,543	21.95
	American Indian/Native	1,015	0.49
	Multiracial	1,709	0.82
	Pacific Islander	392	0.19
	White	104,505	50.37

Table 8F. ELA Grade 8 Sample Characteristics (cont.)

Demographic Category		N-count	% of Total N-count
NRC	New York	66,271	31.94
	Big 4 Cities	7,318	3.53
	Urban/Suburban	14,159	6.82
	Rural	11,122	5.36
	Average Needs	58,325	28.11
	Low Needs	30,955	14.92
	Charter	4,322	2.08
	Non-Public	15,004	7.23
SWD	No	178,426	86.00
	Yes	29,050	14.00
SUA	No	166,041	80.03
	Yes	41,435	19.97
ELL	No	196,626	94.77
	Yes	10,850	5.23

Note. The total n-count was 207,476.

Table 9A. Mathematics Grade 3 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	98,748	48.86
	Male	103,347	51.14
Ethnicity	Asian	17,741	8.78
	Black	35,406	17.52
	Hispanic	48,822	24.16
	American Indian/Native	1,097	0.54
	Multiracial	2,771	1.37
	Pacific Islander	448	0.22
	White	95,810	47.41
NRC	New York	72,553	35.90
	Big 4 Cities	7,959	3.94
	Urban/Suburban	15,210	7.53
	Rural	10,918	5.40
	Average Needs	55,799	27.61
	Low Needs	26,931	13.33
	Charter	7,354	3.64
	Non-Public	5,371	2.66
SWD	No	174,190	86.19
	Yes	27,905	13.81
SUA	No	150,758	74.60
	Yes	51,337	25.40
ELL	No	182,692	90.40
	Yes	19,403	9.60

Note. The total n-count was 202,095.

Table 9B. Mathematics Grade 4 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	103,418	49.19
	Male	106,841	50.81
Ethnicity	Asian	17,509	8.33
	Black	35,920	17.08
	Hispanic	49,171	23.39
	American Indian/Native	1,126	0.54
	Multiracial	2,479	1.18
	Pacific Islander	450	0.21
	White	103,604	49.27
NRC	New York	70,822	33.68
	Big 4 Cities	7,998	3.8
	Urban/Suburban	15,464	7.35
	Rural	10,901	5.18
	Average Needs	57,632	27.41
	Low Needs	27,980	13.31
	Charter	5,795	2.76
	Non-Public	13,667	6.5
SWD	No	180,147	85.68
	Yes	30,112	14.32
SUA	No	158,533	75.4
	Yes	51,726	24.6
ELL	No	193,691	92.12
	Yes	16,568	7.88

Note. The total n-count was 210,259.

Table 9C. Mathematics Grade 5 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	98,748	48.86
	Male	103,347	51.14
Ethnicity	Asian	17,741	8.78
	Black	35,406	17.52
	Hispanic	48,822	24.16
	American Indian/Native	1,097	0.54
	Multiracial	2,771	1.37
	Pacific Islander	448	0.22
	White	95,810	47.41

Table 9C. Mathematics Grade 5 Sample Characteristics (cont.)

Demographic Category		N-count	% of Total N-count
NRC	New York	72,553	35.90
	Big 4 Cities	7,959	3.94
	Urban/Suburban	15,210	7.53
	Rural	10,918	5.40
	Average Needs	55,799	27.61
	Low Needs	26,931	13.33
	Charter	7,354	3.64
	Non-Public	5,371	2.66
SWD	No	174,190	86.19
	Yes	27,905	13.81
SUA	No	150,758	74.60
	Yes	51,337	25.40
ELL	No	182,692	90.40
	Yes	19,403	9.60

Note. The total n-count was 200,037.

Table 9D. Mathematics Grade 6 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	102,039	49.23
	Male	105,231	50.77
Ethnicity	Asian	16,764	8.09
	Black	37,521	18.10
	Hispanic	47,896	23.11
	American Indian/Native	1,085	0.52
	Multiracial	1,957	0.94
	Pacific Islander	411	0.20
	White	101,636	49.04
NRC	New York	68,231	32.92
	Big 4 Cities	7,902	3.81
	Urban/Suburban	14,332	6.91
	Rural	11,293	5.45
	Average Needs	58,608	28.28
	Low Needs	28,650	13.82
	Charter	6,614	3.19
	Non-Public	11,640	5.62
SWD	No	176,524	85.17
	Yes	30,746	14.83
SUA	No	159,228	76.82
	Yes	48,042	23.18
ELL	No	194,471	93.82
	Yes	12,799	6.18

Note. The total n-count was 207,270.

Table 9E. Mathematics Grade 7 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	100,087	48.95
	Male	104,360	51.05
Ethnicity	Asian	17,375	8.5
	Black	37,018	18.11
	Hispanic	45,861	22.43
	American Indian/Native	1,018	0.5
	Multiracial	1,794	0.88
	Pacific Islander	392	0.19
	White	100,989	49.4
NRC	New York	69,709	34.1
	Big 4 Cities	7,667	3.75
	Urban/Suburban	14,692	7.19
	Rural	11,483	5.62
	Average Needs	59,202	28.96
	Low Needs	31,314	15.32
	Charter	5,326	2.61
	Non-Public	5,054	2.47
SWD	No	174,127	85.17
	Yes	30,320	14.83
SUA	No	158,194	77.38
	Yes	46,253	22.62
ELL	No	192,275	94.05
	Yes	12,172	5.95

Note. The total n-count was 204,447.

Table 9F. Mathematics Grade 8 Sample Characteristics

Demographic Category		N-count	% of Total N-count
Gender	Female	102,193	49.04
	Male	106,178	50.96
Ethnicity	Asian	16,815	8.07
	Black	37,926	18.20
	Hispanic	46,283	22.21
	American Indian/Native	1,021	0.49
	Multiracial	1,724	0.83
	Pacific Islander	393	0.19
	White	104,209	50.01

Table 9F. Mathematics Grade 8 Sample Characteristics (cont.)

Demographic Category		N-count	% of Total N-count
NRC	New York	68,852	33.04
	Big 4 Cities	7,436	3.57
	Urban/Suburban	14,323	6.87
	Rural	11,489	5.51
	Average Needs	58,710	28.18
	Low Needs	31,086	14.92
	Charter	4,400	2.11
	Non-Public	12,075	5.79
SWD	No	178,692	85.76
	Yes	29,679	14.24
SUA	No	164,112	78.76
	Yes	44,259	21.24
ELL	No	195,498	93.82
	Yes	12,873	6.18

Note. The total n-count was 208,371.

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 (i.e., 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 .27 to .90. For Grade 3, the item p-values were between .30 and .81, with a mean of .54. For Grade 4, the item p-values were between .39 and .81, with a mean of .59. For Grade 5, the item p-values were between .27 and .77, with a mean of .57. For Grade 6, the item p-values were between .28 and .90, with a mean of .61. For Grade 7, the item p-values were between .30 and .89, with a mean of .62. For Grade 8, the item p-values were between .32 and .90, with a mean of .62. These p-value statistics are Appendix M Tables M1-M6, along with pbis statistics of the key.

Item difficulties (p-values) on the mathematics tests ranged from .13 to .96. For Grade 3, the item p-values were between .31 and .89, with a mean of .63. For Grade 4, the item p-values were between .25 and .96, with a mean of .59. For Grade 5, the item p-values were between .20 and .95, with a mean of .58. For Grade 6, the item p-values were between .24 and .89, with a mean of .58. For Grade 7, the item p-values were between .13 and .91, with a mean of .50. For Grade 8, the item p-values were between .18 and .91, with a mean of .54. 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 which 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 .57 as shown in Appendix M by Tables M1-M6. For Grade 3, the pbis were between .17 and .53, with a mean of .39. For Grade 4, the pbis were between .18 and .53, with a mean of .40. For Grade 5, the pbis

were between .12 and .52, with a mean of .36. For Grade 6, the pbis were between .24 and .54, with a mean of .39. For Grade 7, the pbis were between .14 and .54, with a mean of .36. For Grade 8, the pbis were between .22 and .57, with a mean of .36.

Point biserial correlations for the correct answer options on the mathematics tests ranged .09 to .62 as shown in Appendix M by Tables M7-M12. For Grade 3, the pbis were between .20 and .62, with a mean of .46. For Grade 4, the pbis were between .15 and .57, with a mean of .42. For Grade 5, the pbis were between .09 and .62, with a mean of .42. For Grade 6, the pbis were between .27 and .59, with a mean of .43. For Grade 7, the pbis were between .12 and .62, with a mean of .40. For Grade 8, the pbis were between .10 and .56, with a mean of .40.

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., Agreement) 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 item 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 were typically between 0% and 3%. As may be expected, omit rates tended to increase for items at the end of the test booklets. However, omit rates rarely exceeded 3%, even for the last items within a booklet. 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 among 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 analysis results are presented in Tables 10A–10B. It can be seen that the correlation coefficients were sufficiently high and consistent across books for both ELA and Mathematics tests. These patterns in ELA and mathematics reflect what would be expected for tests that do not exhibit speededness.

Table 10A. ELA Correlations (across Books)

Grade	Book	Other Books
3	1	0.74
	2	0.69
	3	0.71
	4	0.71
4	1	0.75
	2	0.60
	3	0.72
	4	0.73
5	1	0.80
	2	0.76
	3	0.72
	4	0.77
6	1	0.81
	2	0.79
	3	0.72
	4	0.73
7	1	0.80
	2	0.74
	3	0.77
	4	0.74
8	1	0.81
	2	0.76
	3	0.69
	4	0.75

Table 10B. Mathematics Correlations (across Books)

Grade	Book	Other Books
3	1	0.89
	2	0.88
	3	0.84
4	1	0.88
	2	0.89
	3	0.86
5	1	0.88
	2	0.88
	3	0.87
6	1	0.89
	2	0.89
	3	0.88
7	1	0.86
	2	0.88
	3	0.86
8	1	0.86
	2	0.87
	3	0.86

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	Format	Books	Correlation
3	MC	1, 2	0.65
	CR	3, 4	0.68
4	MC	1, 2	0.59
	CR	3, 4	0.71
5	MC	1, 2	0.73
	CR	3, 4	0.71
6	MC	1, 2	0.77
	CR	3, 4	0.70
7	MC	1, 2	0.71
	CR	3, 4	0.72
8	MC	1, 2	0.75
	CR	3, 4	0.70

Table 11B. Mathematics Correlations (Book by Book)

Grade	Book	Book 1	Book 2	Book 3
3	1	1.00	0.87	0.82
	2	0.87	1.00	0.81
	3	0.82	0.81	1.00
4	1	1.00	0.86	0.83
	2	0.86	1.00	0.84
	3	0.83	0.84	1.00
5	1	1.00	0.85	0.84
	2	0.85	1.00	0.84
	3	0.84	0.84	1.00
6	1	1.00	0.86	0.85
	2	0.86	1.00	0.85
	3	0.85	0.85	1.00
7	1	1.00	0.84	0.81
	2	0.84	1.00	0.84
	3	0.81	0.84	1.00
8	1	1.00	0.83	0.81
	2	0.83	1.00	0.82
	3	0.81	0.82	1.00

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 experts to exhibit DIF, 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 2013. Turning to the analysis of the 2013 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 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. Tables 12 and 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	34,259	46,873	95,060	96,650	100,755	102,443	81,707
4	35,059	47,461	101,183	100,767	103,789	100,349	82,978
5	34,547	44,455	96,224	96,074	98,803	98,151	83,888
6	36,561	46,324	99,568	99,494	102,524	97,514	84,546
7	36,115	44,051	100,775	98,469	102,513	99,912	90,344
8	37,812	45,543	104,505	101,958	105,518	98,870	89,280

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	35,406	48,822	95,810	98,748	103,347	106,640	82,730
4	35,920	49,171	103,604	103,418	106,841	105,185	85,612
5	35,756	46,682	97,221	98,550	101,487	102,104	85,399
6	37,521	47,896	101,636	102,039	105,231	101,758	87,258
7	37,018	45,861	100,989	100,087	104,360	103,551	90,516
8	37,926	46,283	104,209	102,193	106,178	102,100	89,796

Tables 14 (ELA) and 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	Number of Flagged Items
3	0
4	3
5	3
6	4
7	4
8	7

Table 15. Mathematics Items Flagged for DIF

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

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 Scaling

IRT Models and Rationale for Use

IRT allows for comparisons among 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 *k*-th level of the *j*-th 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 j-th 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 sample data were used for calibration and scaling of NYSTP 2013 Grades 3–8 Common Core ELA and Mathematics Tests. It should be noted that the sample sizes were adequate as the calibration and scaling 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–17C, the 2013 operational test samples were comparable to 2012 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

Demographics	2012 Grade 3 Population	2013 Grade 3 Sample	2012 Grade 4 Population	2013 Grade 4 Sample
	%	%	%	%
GENDER				
Female	48.94	48.96	49.09	49.26
Male	51.06	51.04	50.91	50.74
ETHNICITY				
Asian	8.31	8.61	8.31	8.25
Black	18.02	17.35	18.24	17.14
Hispanics	23.79	23.74	23.42	23.20
American Indian	0.55	0.54	0.53	0.54
Multiracial	1.02	1.38	0.87	1.18
Other	0.20	0.22	0.16	0.22
White	48.11	48.15	48.47	49.46
NRC SUBGROUPS				
NYC	35.06	34.75	32.74	32.87
Big 4 Cities	3.93	4.00	3.76	3.84
Urban/Suburban	7.67	7.65	7.18	7.40
Rural	5.43	5.50	5.23	4.95
Average Needs	28.41	27.83	27.15	27.20
Low Needs	13.68	13.56	13.27	13.36
Charter	2.92	3.64	2.25	2.66
Non-Public	2.90	3.08	8.42	7.72
DISABILITY				
No	85.76	86.36	84.76	85.93
Yes	14.24	13.64	15.24	14.07
ACCOMMODATIONS				
No	86.21	75.93	88.10	76.84
Yes	13.79	24.07	11.90	23.16
ELL STATUS				
No	91.36	91.17	91.80	92.92
Yes	8.64	8.83	8.20	7.08

Table 16B. ELA Grades 5 and 6 Demographic Statistics

Demographics	2012 Grade 5 Population	2013 Grade 5 Sample	2012 Grade 6 Population	2013 Grade 6 Sample
	%	%	%	%
GENDER				
Female	49.07	49.30	48.89	49.25
Male	50.93	50.70	51.11	50.75
ETHNICITY				
Asian	8.02	8.32	8.36	8.03
Black	18.61	17.73	18.60	18.10
Hispanics	23.01	22.81	22.33	22.93
American Indian	0.50	0.56	0.50	0.52
Multiracial	0.78	1.04	0.73	0.93
Other	0.16	0.17	0.17	0.20
White	48.91	49.38	49.31	49.29
NRC SUBGROUPS				
NYC	33.83	33.60	31.99	31.93
Big 4 Cities	3.93	3.93	3.61	3.84
Urban/Suburban	7.43	7.29	6.90	7.10
Rural	5.67	5.54	5.35	5.40
Average Needs	29.23	28.74	28.35	27.93
Low Needs	14.14	14.30	14.05	13.92
Charter	2.97	3.44	2.57	2.99
Non-Public	2.80	3.15	7.18	6.89
DISABILITY				
No	84.38	84.78	84.75	85.41
Yes	15.62	15.22	15.25	14.59
ACCOMMODATIONS				
No	87.16	75.93	89.68	78.18
Yes	12.84	24.07	11.32	21.82
ELL STATUS				
No	93.12	92.78	94.33	94.73
Yes	6.88	7.22	5.67	5.27

Table 16C. ELA Grades 7 and 8 Demographic Statistics

Demographics	2012 Grade 7 Population	2013 Grade 7 Sample	2012 Grade 8 Population	2013 Grade 8 Sample
	%	%	%	%
GENDER				
Female	48.76	48.99	49.18	49.14
Male	51.24	51.01	50.82	50.86
ETHNICITY				
Asian	7.86	8.39	7.99	7.95
Black	18.77	17.97	18.64	18.22
Hispanics	22.01	21.92	21.76	21.95
American Indian	0.51	0.50	0.52	0.49
Multiracial	0.70	0.90	0.58	0.82
Other	0.16	0.19	0.17	0.19
White	49.99	50.14	50.35	50.37
NRC SUBGROUPS				
NYC	33.12	33.15	31.83	31.94
Big 4 Cities	3.74	3.76	3.38	3.53
Urban/Suburban	7.24	7.15	6.70	6.82
Rural	5.82	5.65	5.44	5.36
Average Needs	29.65	29.29	28.23	28.11
Low Needs	15.50	15.66	14.86	14.92
Charter	2.29	2.61	1.59	2.08
Non-Public	2.63	2.72	7.96	7.23
DISABILITY				
No	84.90	85.26	85.05	86.00
Yes	15.10	14.74	14.95	14.00
ACCOMMODATIONS				
No	88.57	78.84	91.27	80.03
Yes	11.43	21.16	9.73	19.97
ELL STATUS				
No	94.53	94.98	94.60	94.77
Yes	5.47	5.02	5.40	5.23

Table 17A. Mathematics Grades 3 and 4 Demographic Statistics

Demographics	2012 Grade 3 Population	2013 Grade 3 Sample	2012 Grade 4 Population	2013 Grade 4 Sample
	%	%	%	%
GENDER				
Female	48.93	48.86	49.07	49.19
Male	51.07	51.14	50.93	50.81
ETHNICITY				
Asian	8.48	8.78	8.47	8.33
Black	17.95	17.52	18.17	17.08
Hispanics	24.02	24.16	23.63	23.39
American Indian	0.54	0.54	0.53	0.54
Multiracial	1.01	1.37	0.87	1.18
Other	0.21	0.22	0.17	0.21
White	47.79	47.41	48.16	49.27
NRC SUBGROUPS				
NYC	35.36	35.90	33.02	33.68
Big 4 Cities	3.95	3.94	3.78	3.80
Urban/Suburban	7.67	7.53	7.18	7.35
Rural	5.39	5.40	5.20	5.18
Average Needs	28.17	27.61	26.98	27.41
Low Needs	13.61	13.33	13.20	13.31
Charter	2.89	3.64	2.23	2.76
Non-Public	2.95	2.66	8.41	6.50
DISABILITY				
No	85.85	86.19	84.89	85.68
Yes	14.15	13.81	15.11	14.32
ACCOMMODATIONS				
No	88.44	74.60	89.42	75.40
Yes	11.56	25.40	10.58	24.60
ELL STATUS				
No	98.12	90.40	98.06	92.12
Yes	1.88	9.60	1.94	7.88

Table 17B. Mathematics Grades 5 and 6 Demographic Statistics

Demographics	2012 Grade 5 Population	2013 Grade 5 Sample	2012 Grade 6 Population	2013 Grade 6 Sample
	%	%	%	%
GENDER				
Female	49.04	48.86	48.87	49.23
Male	50.96	51.14	51.13	50.77
ETHNICITY				
Asian	8.17	8.78	8.48	8.09
Black	18.54	17.52	18.55	18.10
Hispanics	23.23	24.16	22.55	23.11
American Indian	0.50	0.54	0.51	0.52
Multiracial	0.78	1.37	0.72	0.94
Other	0.17	0.22	0.18	0.20
White	48.61	47.41	49.01	49.04
NRC SUBGROUPS				
NYC	34.10	35.90	32.27	32.92
Big 4 Cities	3.95	3.94	3.63	3.81
Urban/Suburban	7.43	7.53	6.90	6.91
Rural	5.62	5.40	5.32	5.45
Average Needs	29.00	27.61	28.17	28.28
Low Needs	14.05	13.33	13.84	13.82
Charter	2.94	3.64	2.55	3.19
Non-Public	2.90	2.66	7.32	5.62
DISABILITY				
No	84.50	86.19	90.07	85.17
Yes	15.50	13.81	9.93	14.83
ACCOMMODATIONS				
No	86.66	74.60	93.43	76.82
Yes	13.34	25.40	6.57	23.18
ELL STATUS				
No	98.21	90.40	98.00	93.82
Yes	1.79	9.60	2.00	6.18

Table 17C. Mathematics Grades 7 and 8 Demographic Statistics

Demographics	2012 Grade 7 Population	2013 Grade 7 Sample	2012 Grade 8 Population	2013 Grade 8 Sample
	%	%	%	%
GENDER				
Female	48.73	48.95	49.17	49.04
Male	51.27	51.05	50.83	50.96
ETHNICITY				
Asian	8.01	8.50	8.12	8.07
Black	18.71	18.11	18.55	18.20
Hispanics	22.22	22.43	21.98	22.21
American Indian	0.51	0.50	0.52	0.49
Multiracial	0.70	0.88	0.57	0.83
Other	0.17	0.19	0.18	0.19
White	49.69	49.40	50.08	50.01
NRC SUBGROUPS				
NYC	33.43	34.10	32.13	33.04
Big 4 Cities	3.76	3.75	3.42	3.57
Urban/Suburban	7.25	7.19	6.72	6.87
Rural	5.77	5.62	5.42	5.51
Average Needs	29.44	28.96	28.08	28.18
Low Needs	15.40	15.32	14.78	14.92
Charter	2.28	2.61	1.57	2.11
Non-Public	2.68	2.47	7.89	5.79
DISABILITY				
No	85.02	85.17	85.17	85.76
Yes	14.98	14.83	14.83	14.24
ACCOMMODATIONS				
No	89.81	77.38	91.97	78.76
Yes	10.19	22.62	9.03	21.24
ELL STATUS				
No	97.93	94.05	98.07	93.82
Yes	2.07	5.95	1.92	6.18

Calibration Process
The item parameters were estimated using MULTILOG software (Thissen, 1991). MC and CR items were calibrated simultaneously using marginal maximum likelihood procedures.

The calibration of NYSTP 2013 Grades 3–8 Common Core ELA and Mathematics Tests did not exhibit any test-level issues. The estimated parameters were in the original theta metric, and all 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.

Table 18. ELA Calibration Results

Grade	Largest *a*-parameter	*b*-parameter/ Step value Range		Theta Mean	Theta SD	# Students
3	1.419	–2.407	2.859	0.01	0.930	197,405
4	1.706	–2.933	2.603	0.01	0.933	204,556
5	1.392	–2.871	2.773	0.01	0.935	194,877
6	1.437	–3.435	1.901	0.01	0.941	202,018
7	1.627	–2.819	1.597	0.01	0.942	200,982
8	1.527	–3.234	2.021	0.00	0.936	207,476

Table 19. Mathematics Calibration Results

Grade	Largest *a*-parameter	*b*-parameter/ Step value Range		Theta Mean	Theta SD	# Students
3	1.673	–2.381	1.802	0.01	0.933	202,095
4	1.614	–3.379	2.382	0.01	0.947	210,259
5	2.754	–2.954	3.707	0.01	0.950	200,037
6	2.065	–2.033	1.889	0.01	0.955	207,270
7	2.915	–1.630	3.975	0.03	0.951	204,447
8	1.989	–2.125	2.408	0.01	0.950	208,371

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

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

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 700,000 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_{Q_1} critical. If the obtained Z_{Q_1} was less than Z_{Q_1} critical, the items were rated as having acceptable fit. The fact that the majority of the items in the NYSTP 2013 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 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 on all ELA and mathematics tests and no items were found to be significant in terms of local dependency in mathematics Grade 8. Items that were found to be significant in local dependency vary depending on the subject and grade: two pairs of items were found in ELA Grades 3 and 4; five pairs of items were found in ELA Grades 5, 6, and 7; and four pairs of items were found in ELA Grade 8. One pair of items was found in mathematics Grades 3, 5, and 6; two pairs of items were found in mathematics Grade 4; and six pairs of items were found in mathematics Grade 7. The magnitudes of these statistics were not sufficient to warrant any concern (with the Q_3 values ranging from .20–.26 for ELA tests and .20–.33 for mathematics tests).

Scaling

With the new assessments being implemented in 2013, a new scale was established after the data were collected. The following major steps were involved in the scaling process:

- Operational items were calibrated in MULTILOG.

- Raw score to theta conversion chart was produced using the test characteristic curve (TCC) method (Stocking and Lord, 1983) and implemented in POLYEQUATE (Kolen & Cui, 2004).
- 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 end 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 (for example, ELA Grade 3, -5.3; mathematics Grade 3, -4.9). The following adjustment/interpolation was conducted:

  For any theta estimates 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. See the table below for an example in 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	Theta	Adjusted theta
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

- Once theta values were either estimated or interpolated for all raw scores, the raw score to theta relationship was applied to the students, yielding a theta estimate corresponding to his or her raw score.
- The mean and standard deviation (SD) of the thetas were then computed.

- Using the formula below, together with the desired mean and SD values for the scale scores, which were 350 and 35, respectively, M1 and M2 (scaling linear transformation intercept and slope) were obtained.

$$M1 = \frac{\sigma_s}{\sigma_\theta}$$

$$M2 = \bar{S} - \left[\left(\frac{\sigma_S}{\sigma_\theta}\right) \times \bar{\theta}\right]$$

where

σ_S = Standard deviation of scale score (desired)

σ_θ = Standard deviation of theta (observed)

$\bar{S}$ = Average of scale score (desired)

$\bar{\theta}$ = Average of theta (observed)

M1 is defined as the multiplicative adjustment and *M2* is the additive adjustment. *M1* and *M2* are applied to a true score in order to obtain a scale score:

$$\text{Scale score} = \hat{\theta} \times M1 + M2$$

- *M1* and *M2* were applied to the theta values to derive the scale scores.
- Scale scores range approximately from 100–400 across grades. The lowest and highest observed scale score (LOSS and HOSS, respectively) may vary by grade.
- For conditional standard error of measurement (CSEM), the scale scores (both estimated and interpolated) were used to compute the information function and CSEM

In this process, the NYSED psychometricians and a senior scientist from HumRRO have independently verified the results generated by the Pearson psychometricians. For future years, the same process will be applied:

- Item parameter estimates will be equated to the base scale established in spring 2013
- TCC estimates will be calculated using equated item parameter estimates
- Adjustments/interpolation will be made to thetas outside of the -2.5 to 3 range or when there are no estimates provided by POLYEQUATE
- *M1* and *M2* established from this year will be applied to the estimated and interpolated thetas from TCC approach
- LOSS and HOSS will be computed accordingly

Item Parameters

The test item parameters were estimated by the software MULTILOG (Thissen, 1991) and are presented in Appendix O, in Tables O13–O24. The parameter estimates are expressed in scale score 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;
- for CR items:
    - *alpha* is a discrimination parameter; and
    - *step* is a difficulty parameter for category m_j.

As described in the Section 6 "IRT Calibration and Scaling," 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*- and *b*-parameter per item in the 3PL model.

Test Characteristic Curves

Test Characteristic Curves (TCCs) provide an overview of the tests in the IRT scale score metric. The 2013 TCCs were generated using final item parameters for all reporting test items administered in spring 2013. 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.

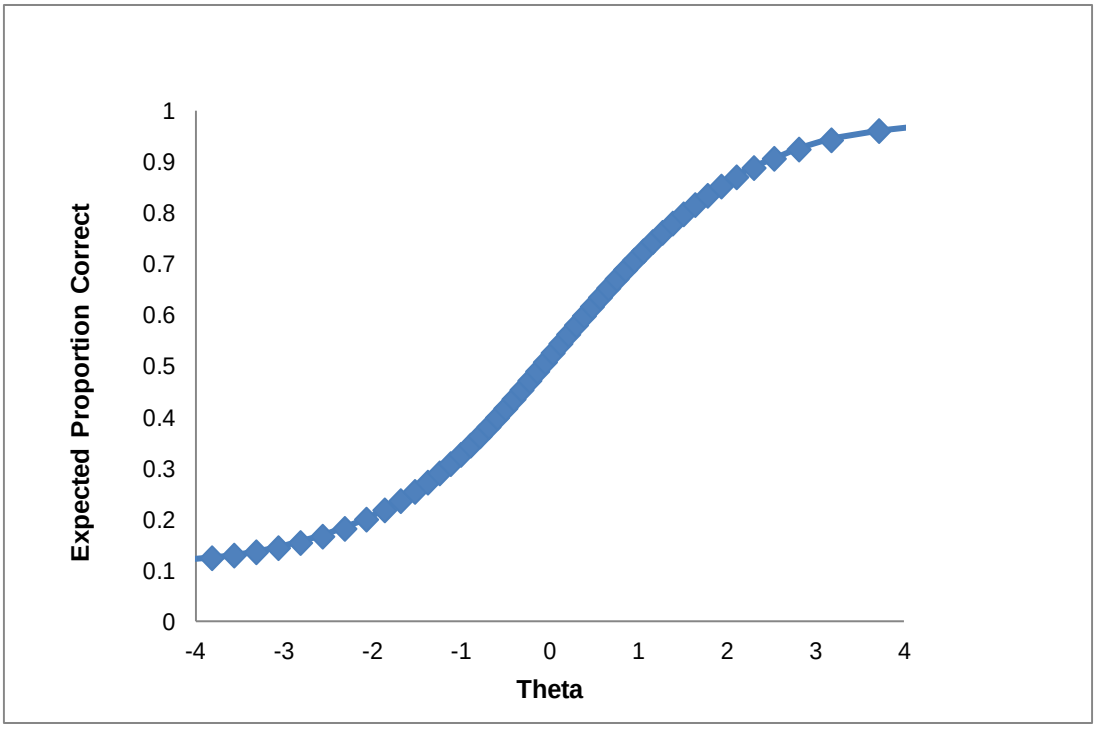

Figure 1. ELA Grade 3 TCC

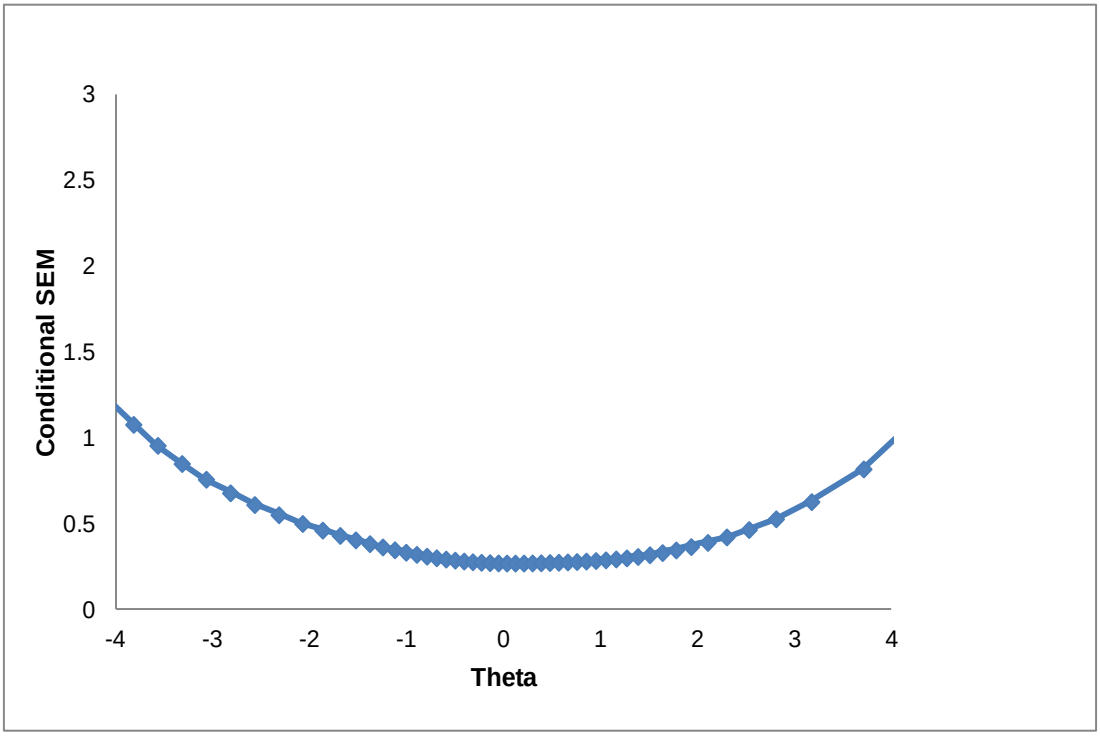

Figure 2. ELA Grade 3 CSEM Curve

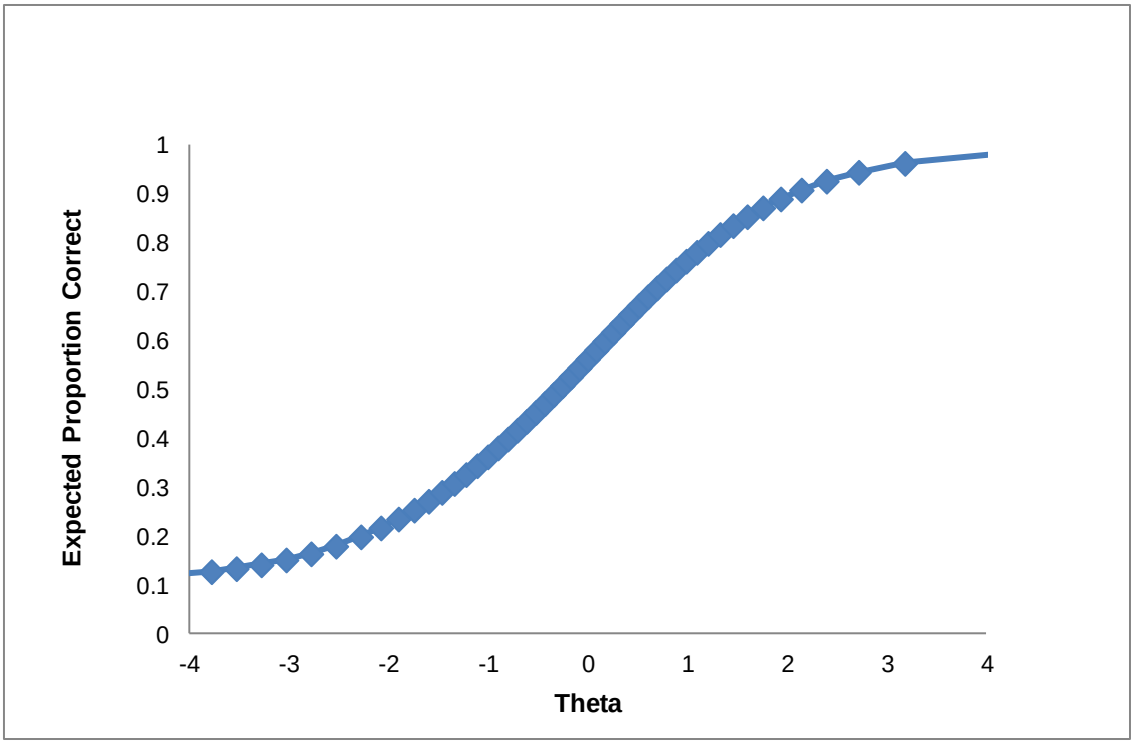

Figure 3. ELA Grade 4 TCC

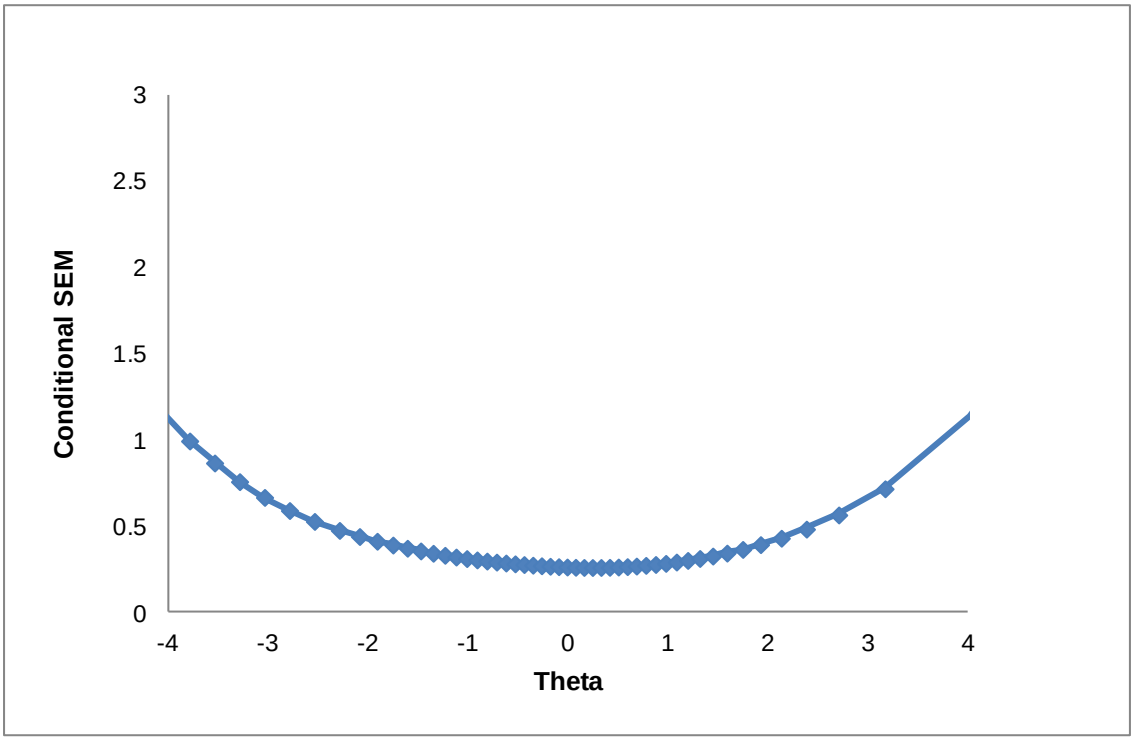

Figure 4. ELA Grade 4 CSEM Curve

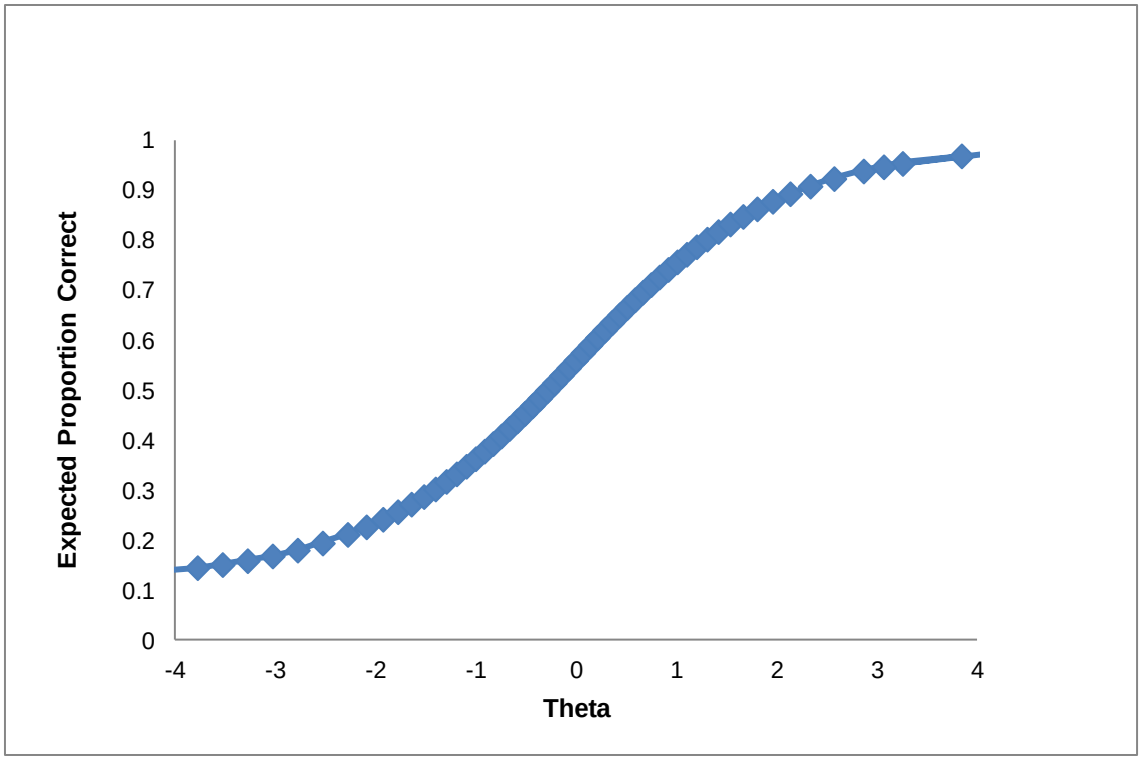

Figure 5. ELA Grade 5 TCC

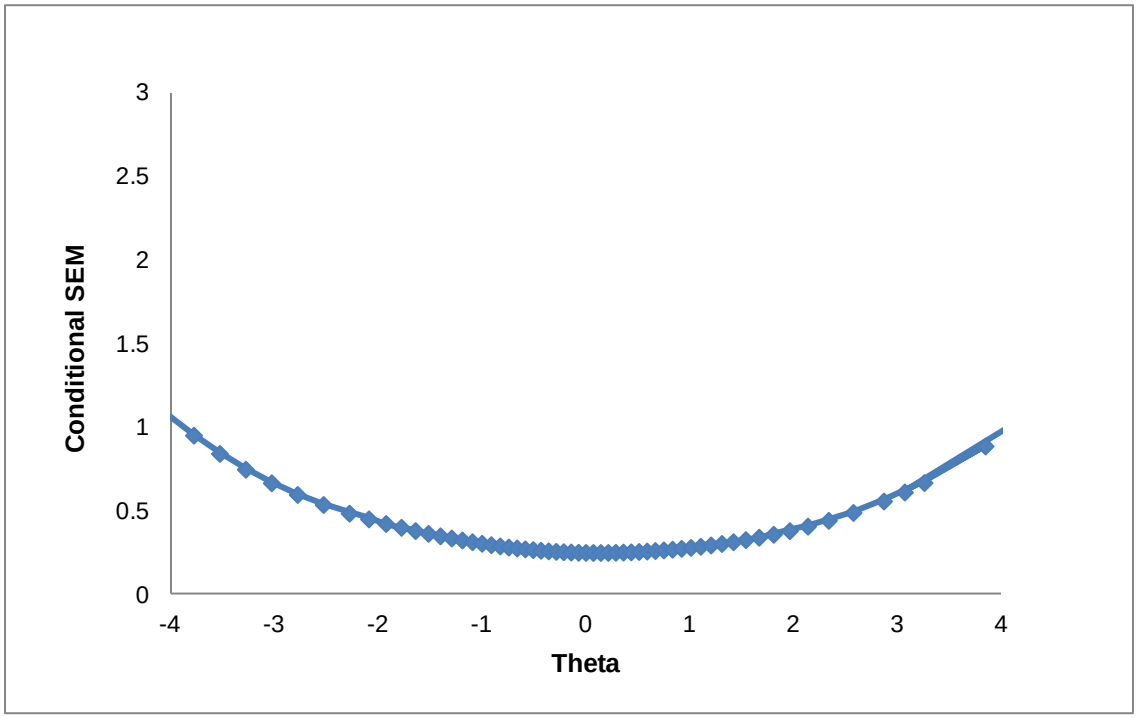

Figure 6. ELA Grade 5 CSEM Curve

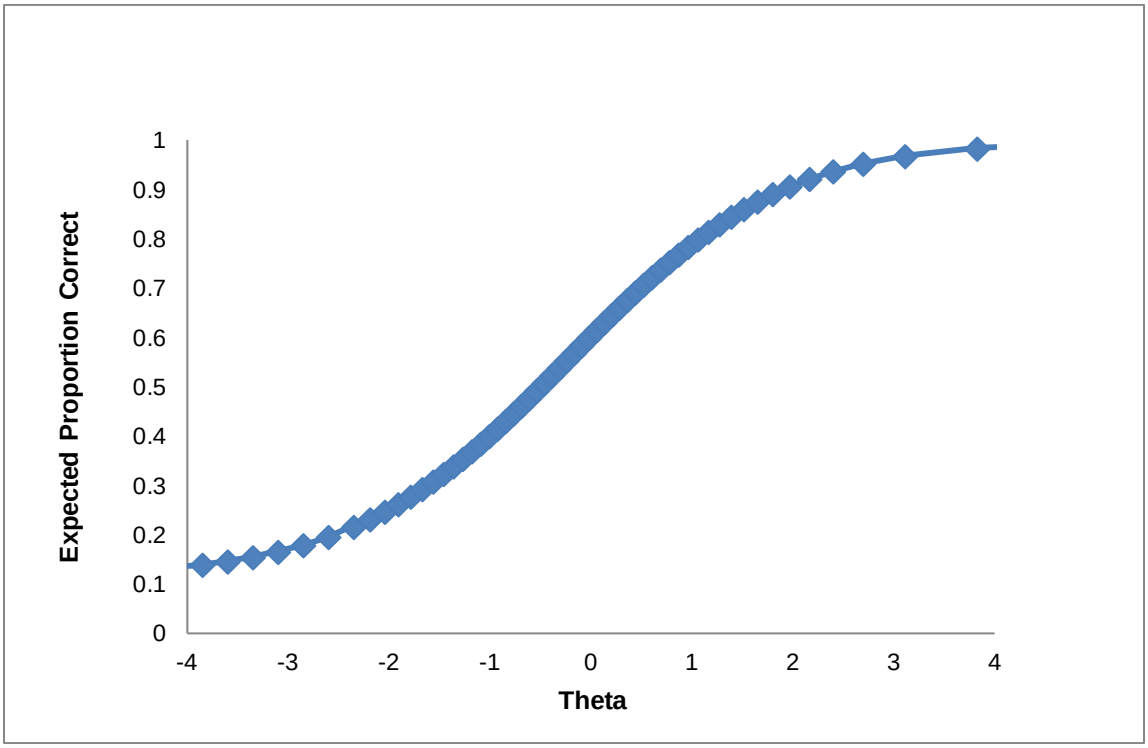

Figure 7. ELA Grade 6 TCC

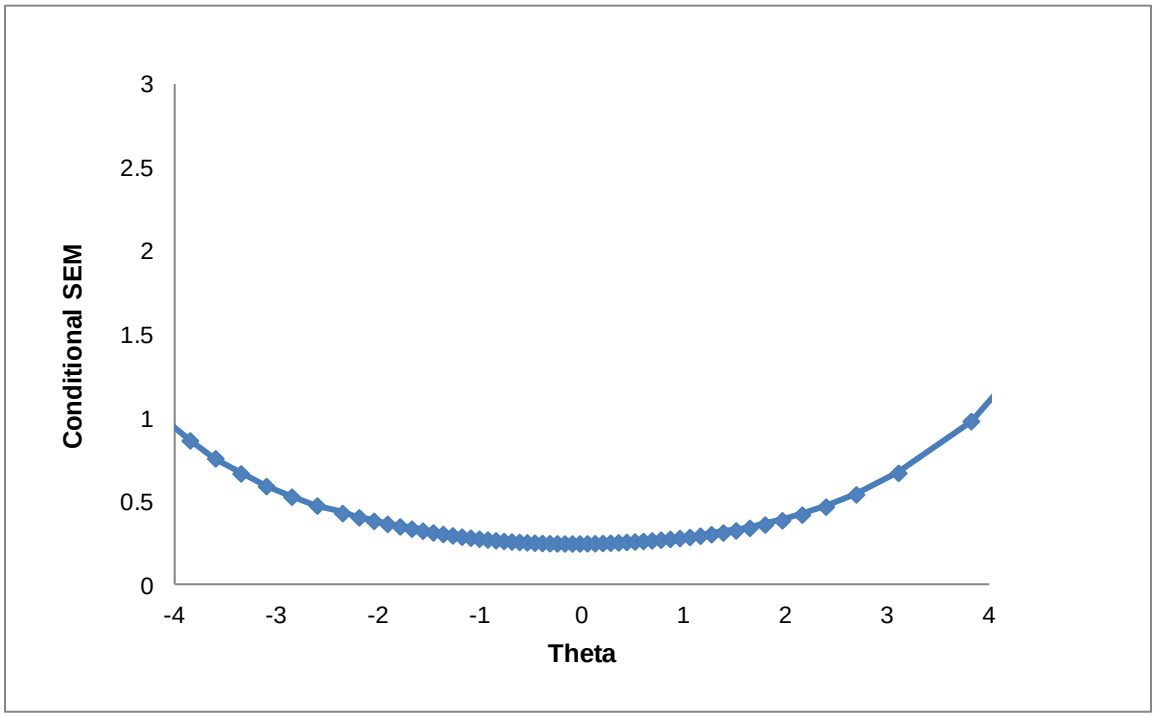

Figure 8. ELA Grade 6 CSEM Curve

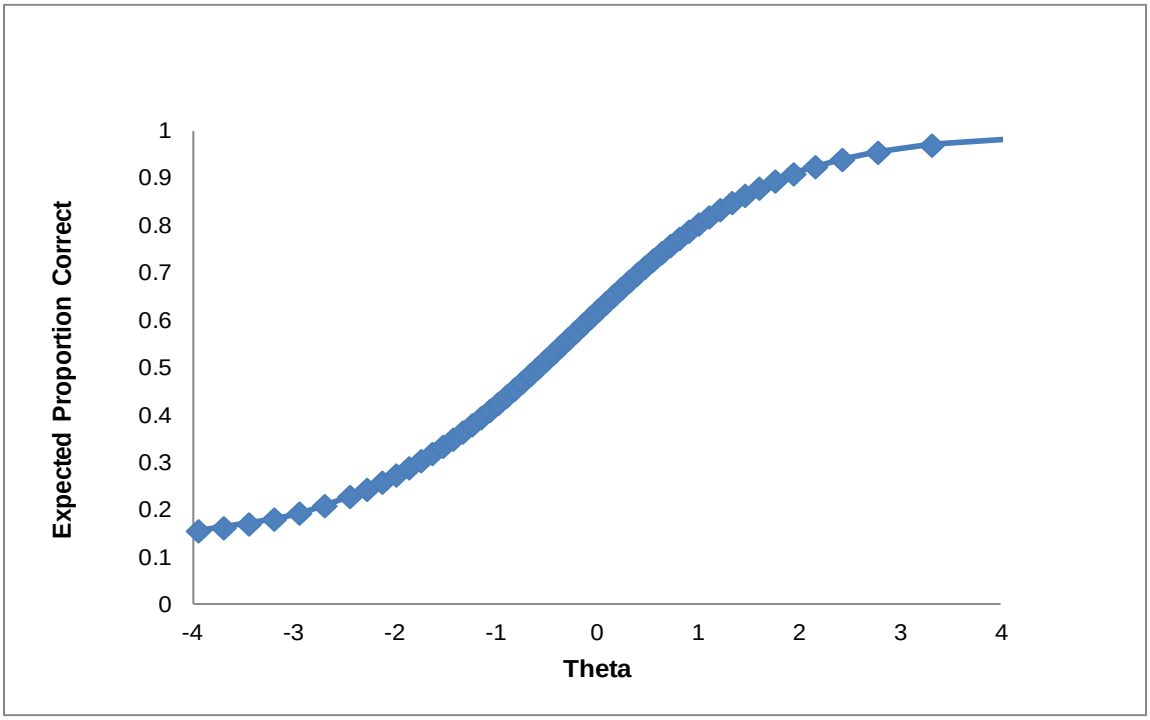

Figure 9. ELA Grade 7 TCC

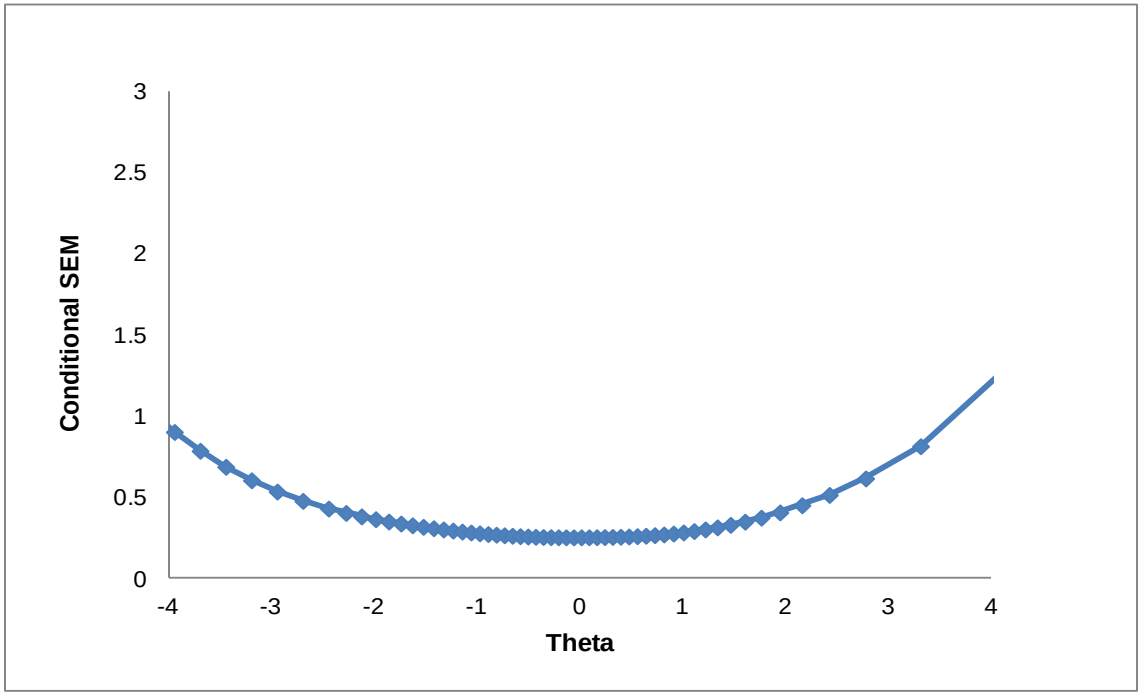

Figure 10. ELA Grade 7 CSEM Curve

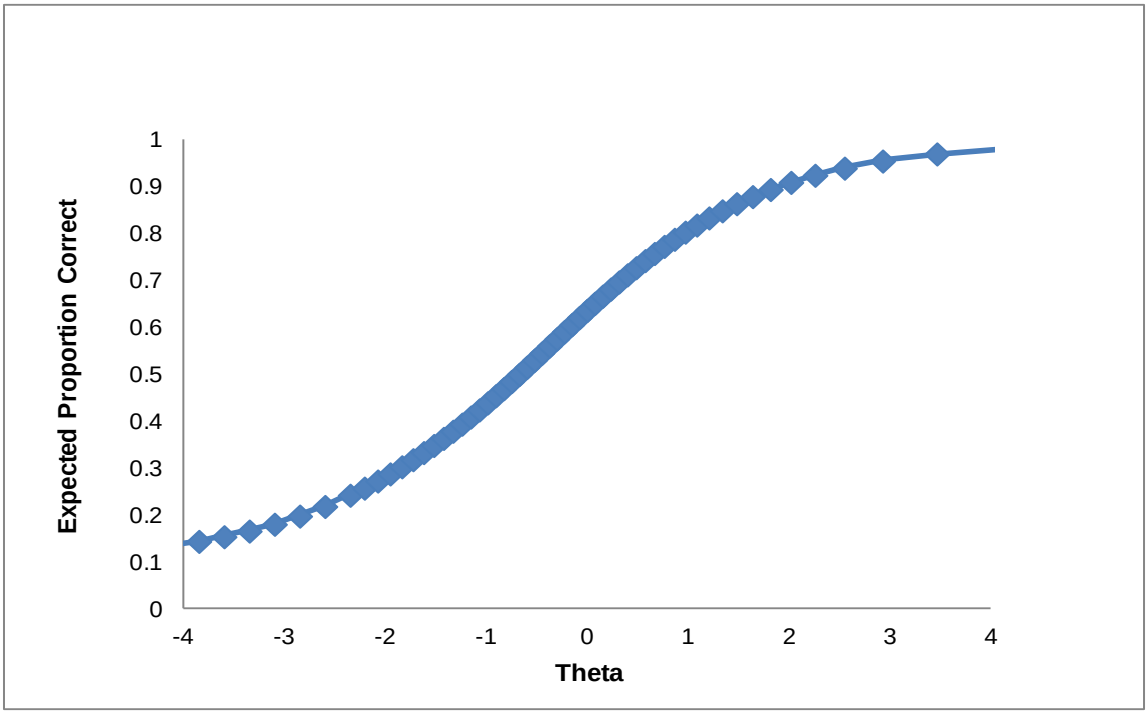

Figure 11. ELA Grade 8 TCC

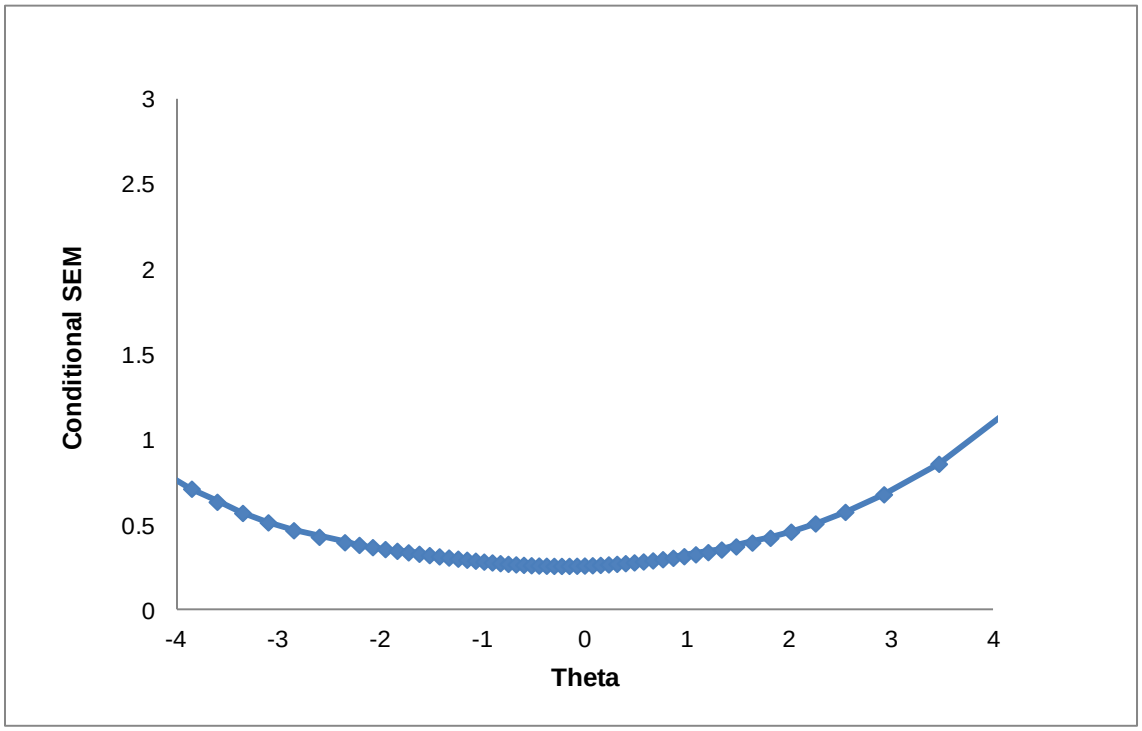

Figure 12. ELA Grade 8 CSEM Curve

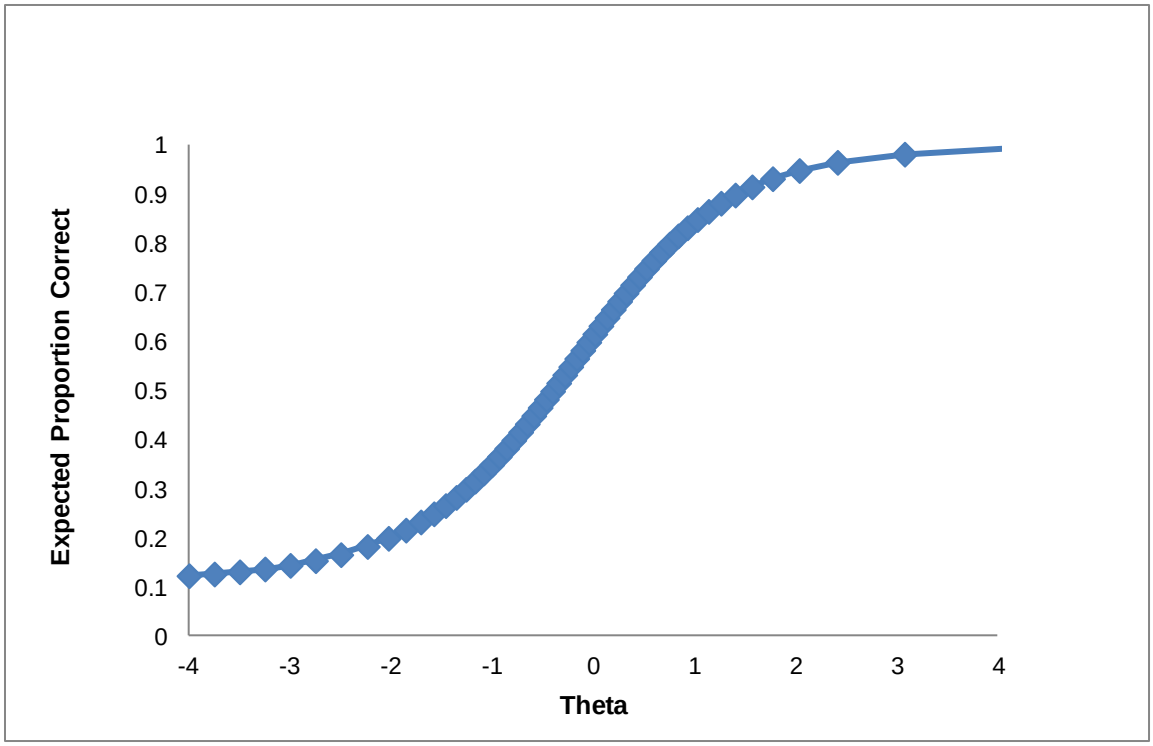

Figure 13. Mathematics Grade 3 TCC

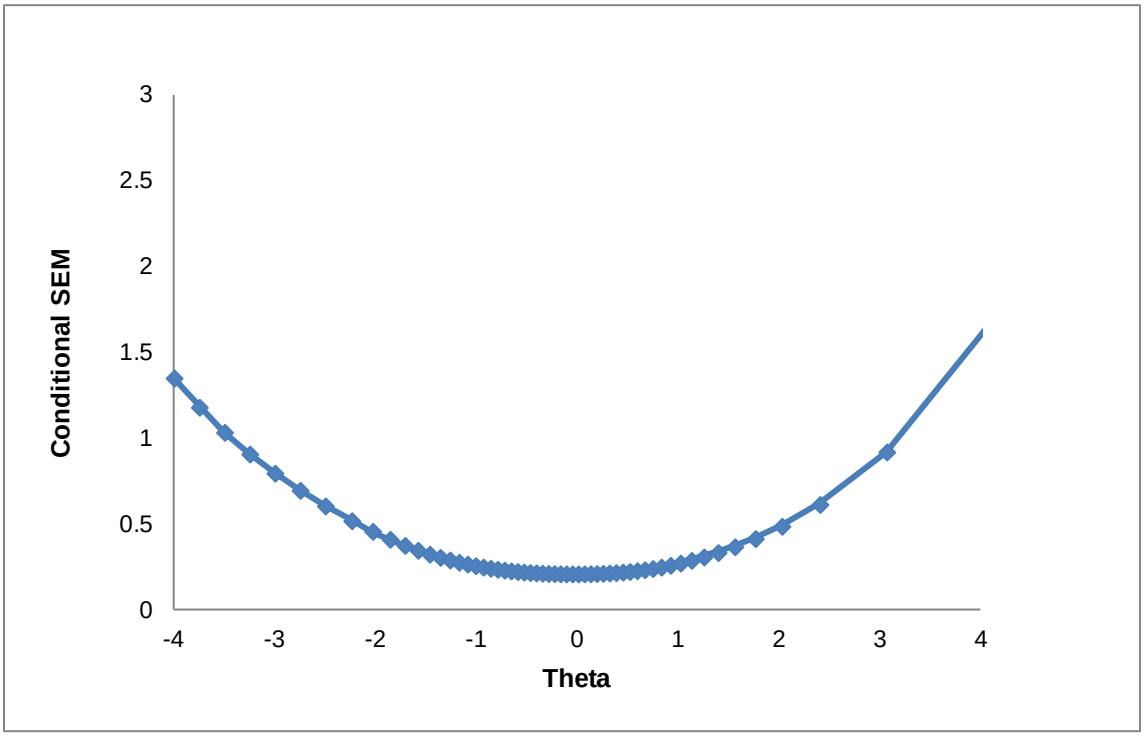

Figure 14. Mathematics Grade 3 CSEM Curve

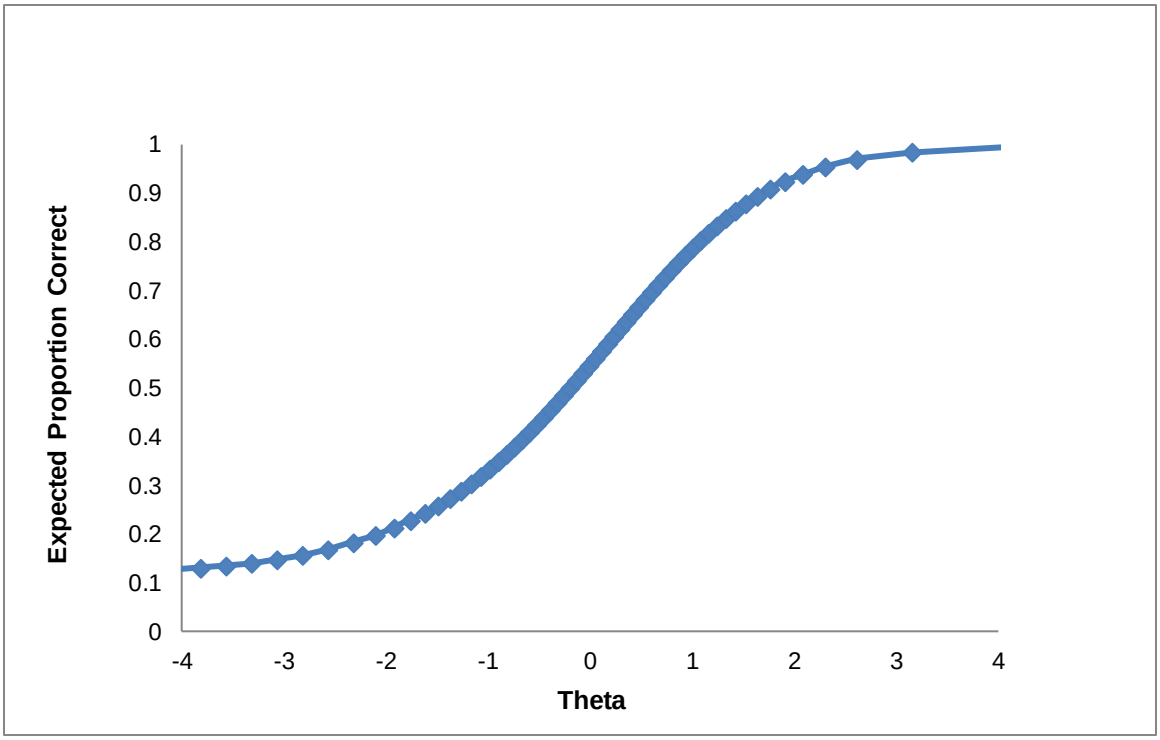

Figure 15. Mathematics Grade 4 TCC

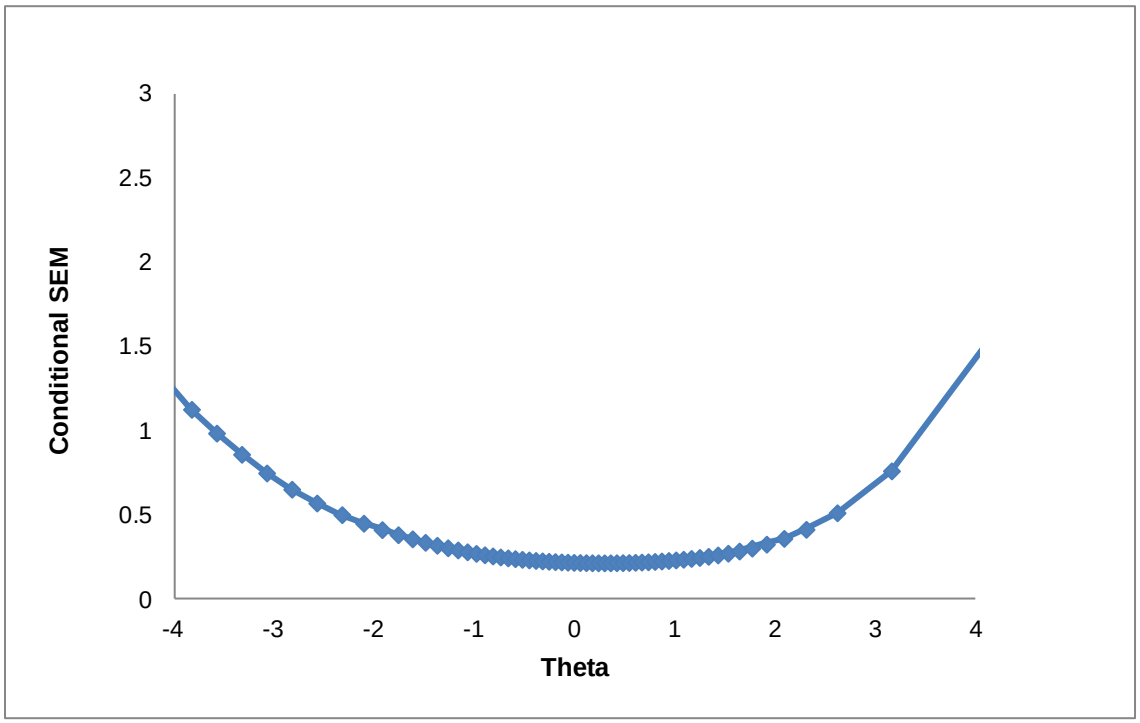

Figure 16. Mathematics Grade 4 CSEM Curve

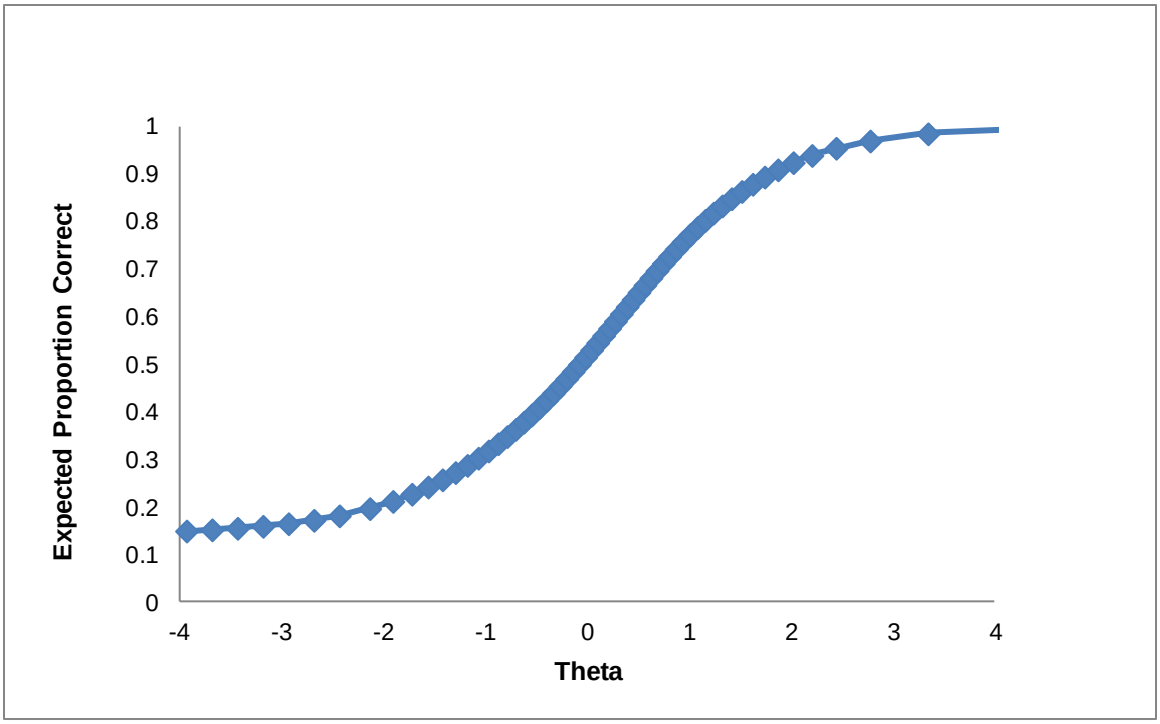

Figure 17. Mathematics Grade 5 TCC

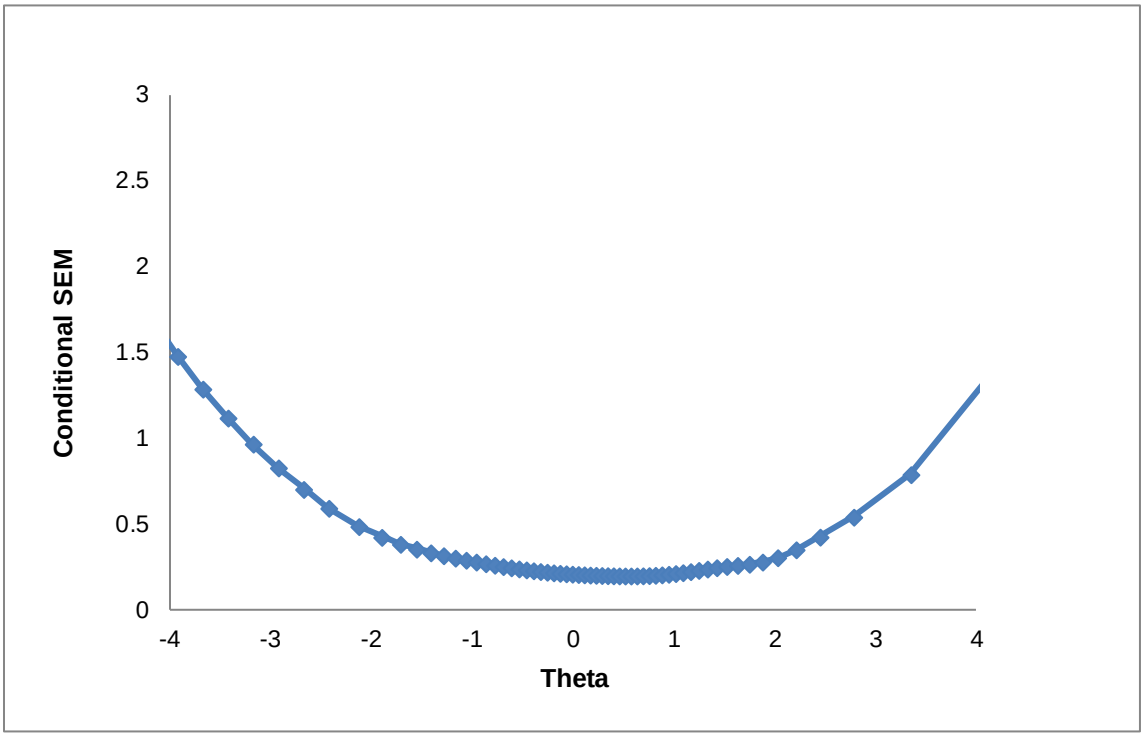

Figure 18. Mathematics Grade 5 CSEM Curve

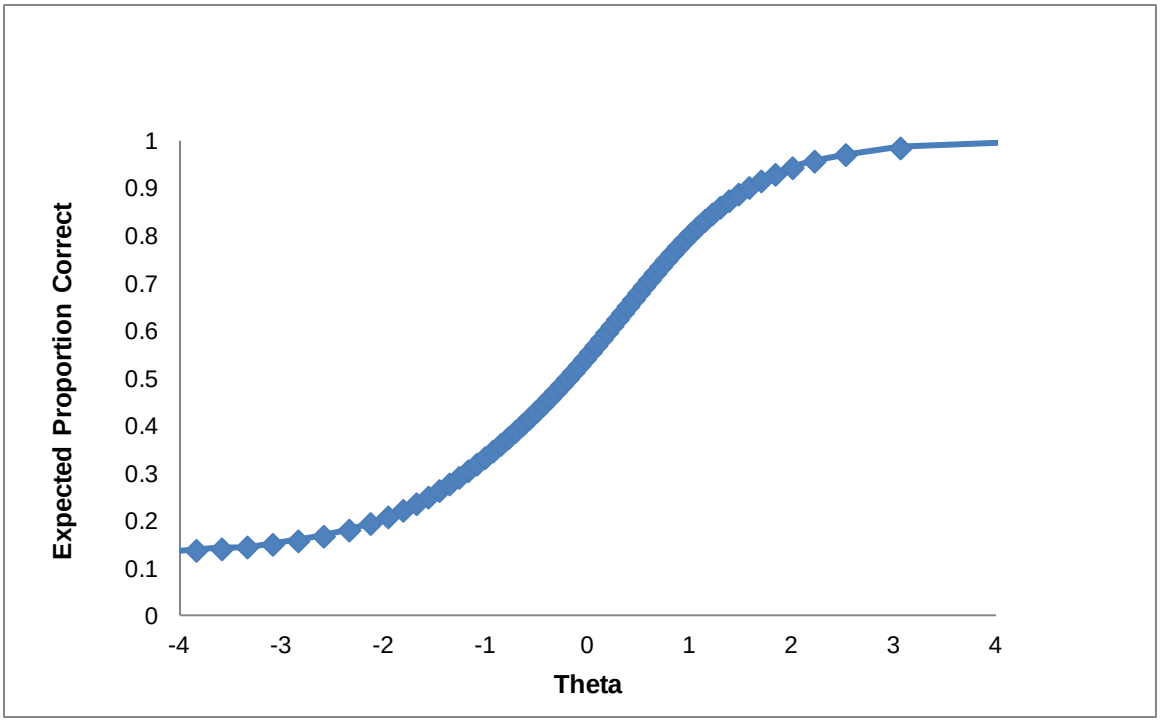

Figure 19. Mathematics Grade 6 TCC

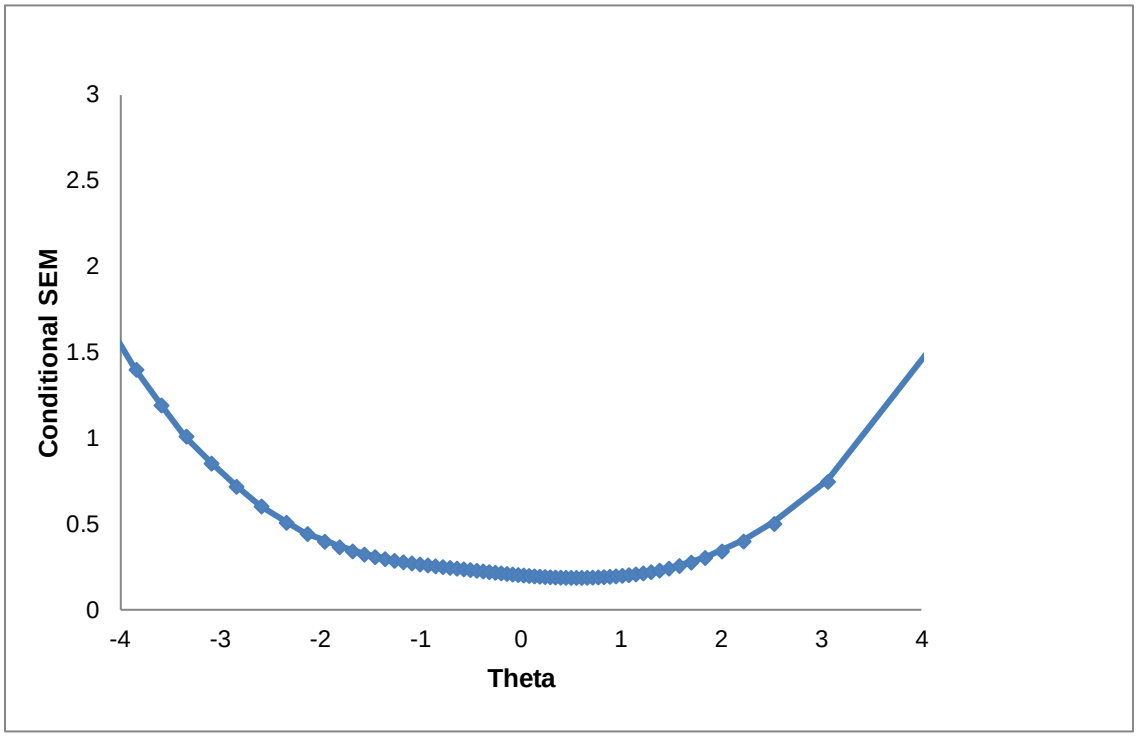

Figure 20. Mathematics Grade 6 CSEM Curve

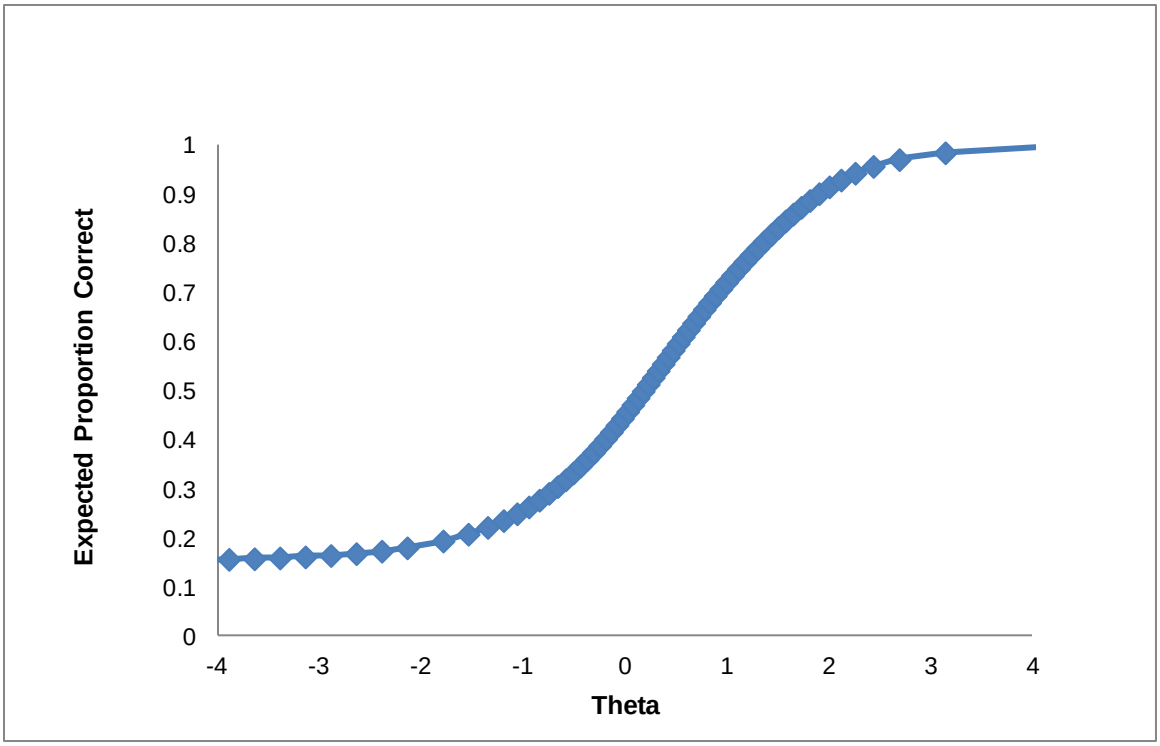

Figure 21. Mathematics Grade 7 TCC

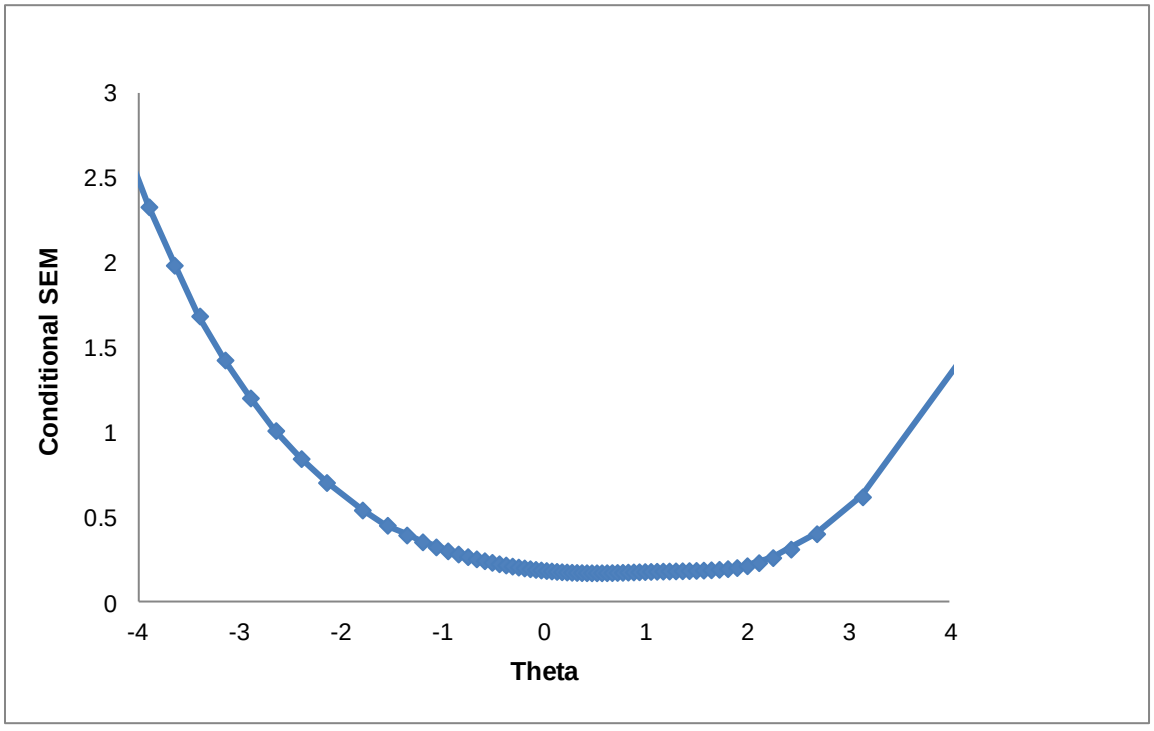

Figure 22. Mathematics Grade 7 CSEM Curve

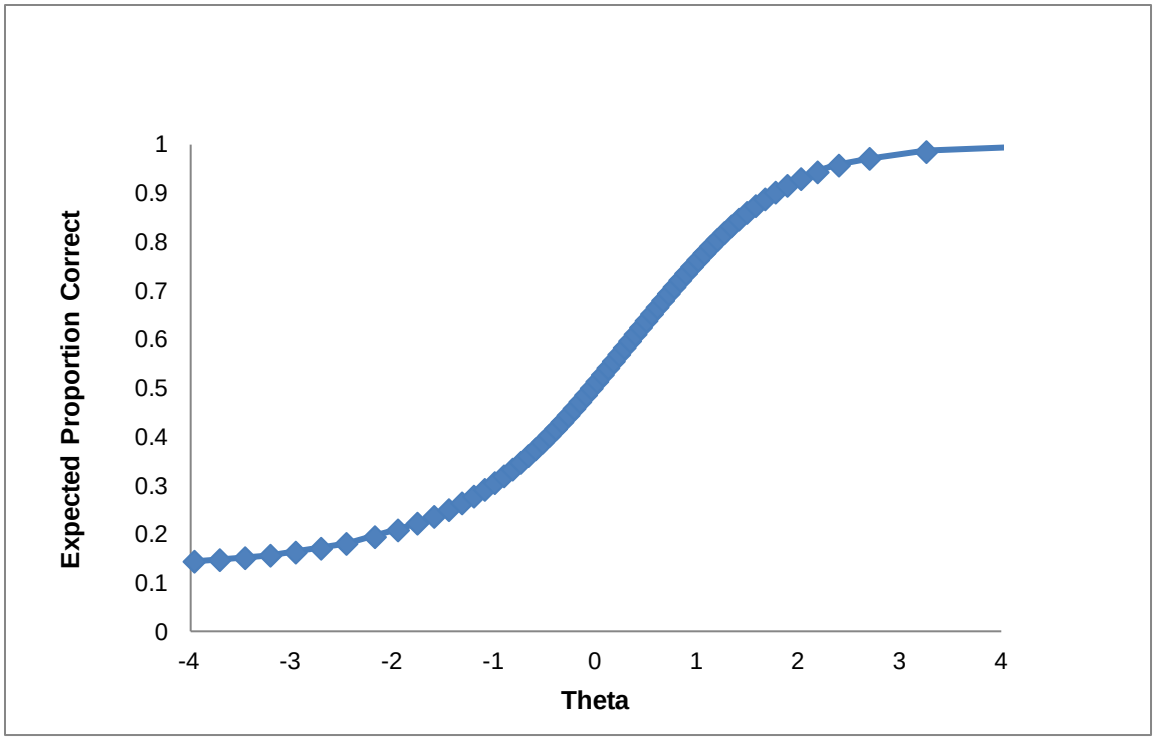

Figure 23. Mathematics Grade 8 TCC

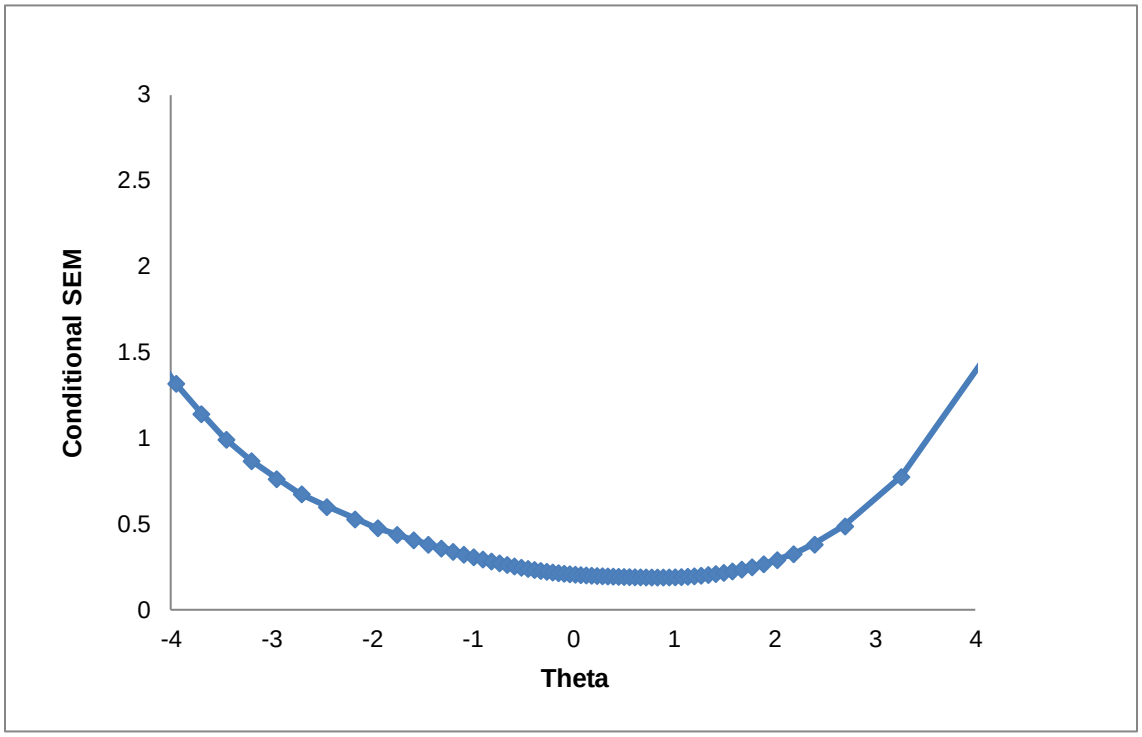

Figure 24. Mathematics Grade 8 CSEM Curve

Scoring Procedure

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

As described in the "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 (M1 and M2) 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 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:

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

where

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

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

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

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 conversion tables based on the total number correct are presented in Appendix R, as Tables R1-R12.

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.

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 agreement 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 Tables 20 and 22 for ELA and mathematics, respectively. These statistics give the necessary context for Tables 21 and 23, 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 .90–.92. Grades 3–8 mathematics reliability estimates ranged from .93–.94 (Cronbach's alpha and Feldt-Raju). The lowest reliability was observed for ELA Grade 3 and Grade 4; however, as that test had the lowest number of 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 6, 7, and 8. All reliabilities exceeded .90 across grades, which is a good indication that the NYSTP Grades 3–8 Common Core ELA and Mathematics Tests are acceptably reliable.

Table 20. ELA Test Form Statistics

Grade	Max RS	RS Mean	RS SD	P-value Mean	Minimum P-value	Maximum P-value
3	55	28.50	10.28	0.54	0.30	0.81
4	55	31.06	10.49	0.59	0.39	0.81
5	66	36.82	12.26	0.57	0.27	0.77
6	65	38.73	12.17	0.61	0.28	0.90
7	66	40.13	12.20	0.62	0.30	0.89
8	66	40.98	11.98	0.62	0.32	0.90

Table 21. ELA Test Reliability and Standard Error of Measurement

Grade	N-count	# Items	Cronbach's Alpha	SEM of Alpha	Feldt-Raju Coefficient	SEM of Feldt-Raju
3	203,041	41	0.90	3.27	0.91	3.14
4	213,673	41	0.90	3.23	0.91	3.11
5	201,117	52	0.91	3.68	0.92	3.53
6	211,594	51	0.92	3.54	0.92	3.40
7	205,800	52	0.91	3.66	0.92	3.48
8	214,090	52	0.91	3.66	0.91	3.50

Table 22. Mathematics Test Form Statistics

Grade	Max RS	RS Mean	RS SD	P-value Mean	Minimum P-value	Maximum P-value
3	60	36.13	13.36	0.63	0.31	0.89
4	66	36.66	13.87	0.59	0.25	0.96
5	66	35.27	13.72	0.58	0.20	0.95
6	72	40.36	15.36	0.58	0.24	0.89
7	72	34.88	15.58	0.50	0.13	0.91
8	72	37.82	15.13	0.54	0.18	0.91

Table 23. Mathematics Test Reliability and Standard Error of Measurement

Grade	N-count	# Items	Cronbach's Alpha	SEM of Alpha	Feldt-Raju Coefficient	SEM of Feldt-Raju
3	204,504	49	0.93	3.56	0.94	3.36
4	215,394	52	0.93	3.73	0.94	3.50
5	202,579	52	0.93	3.66	0.94	3.44
6	213,147	58	0.94	3.87	0.94	3.67
7	207,251	58	0.93	3.97	0.94	3.68
8	214,932	58	0.93	3.94	0.94	3.71

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. Tables 24 and 25 present reliabilities for the subsets of MC items.

Table 24. ELA Reliability and Standard Error of Measurement—MC Items Only

Grade	N-count	# Items	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju	SEM of Feldt-Raju
3	203,041	31	0.84	2.45	0.84	2.44
4	213,673	31	0.85	2.44	0.85	2.43
5	201,117	42	0.86	2.92	0.86	2.90
6	211,594	41	0.87	2.77	0.87	2.76
7	205,800	42	0.85	2.84	0.85	2.83
8	214,090	42	0.86	2.85	0.86	2.85

Table 25. Mathematics Reliability and Standard Error of Measurement—MC Items Only

Grade	N-count	# Items	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju	SEM of Feldt-Raju
3	204,504	41	0.91	2.60	0.91	2.57
4	215,394	42	0.90	2.64	0.90	2.61
5	202,579	42	0.89	2.68	0.89	2.66
6	213,147	48	0.91	2.93	0.91	2.92
7	207,251	48	0.89	3.04	0.90	3.02
8	214,932	48	0.90	2.98	0.90	2.96

Reliability of CR Items

Reliability coefficients were also computed for the subsets of CR items. The results are presented in Tables 26 and 27.

Table 26. ELA Reliability and Standard Error of Measurement—CR Items Only

Grade	N-count	# Items	# RS Points	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju	SEM of Feldt-Raju
3	203,041	10	24	0.84	1.98	0.85	1.90
4	213,673	10	24	0.86	1.91	0.87	1.84
5	201,117	10	24	0.86	2.00	0.87	1.92
6	211,594	10	24	0.86	1.96	0.87	1.88
7	205,800	10	24	0.87	2.02	0.88	1.93
8	214,090	10	24	0.86	2.02	0.87	1.92

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

Table 27. Mathematics Reliability and Standard Error of Measurement—CR Items Only

Grade	N-count	# Items	# RS Points	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju	SEM of Feldt-Raju
3	204,504	8	19	0.82	2.21	0.83	2.14
4	215,394	10	24	0.85	2.38	0.86	2.31
5	202,579	10	24	0.87	2.23	0.87	2.16
6	213,147	10	24	0.86	2.26	0.87	2.20
7	207,251	10	24	0.90	2.16	0.91	2.04
8	214,932	10	24	0.88	2.24	0.89	2.20

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 Tables 28A through 29F, the estimated reliabilities for subgroups were close in magnitude to the test reliability estimates of the population. Cronbach's alpha reliability coefficients were all greater than .80. Feldt-Raju reliability coefficients, which tend to be larger than the Cronbach's alpha estimates for the same group, were also all larger than .80. These indicate a very good test internal consistency (reliability) for analyzed subgroups of examinees.

Table 28A. ELA Test Reliability by Subgroup, Grade 3

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	203,041	0.90	3.27	0.91	3.14
Gender	Female	99,173	0.89	3.27	0.90	3.14
	Male	103,868	0.90	3.26	0.91	3.13
Ethnicity	Asian	17,411	0.90	3.20	0.90	3.07
	Black	35,947	0.89	3.29	0.89	3.17
	Hispanic	48,391	0.88	3.28	0.89	3.17
	American Indian	1,099	0.89	3.27	0.90	3.14
	Multiracial	2,793	0.91	3.25	0.92	3.10
	Native Hawaiian or Other Pacific Islander	439	0.90	3.26	0.91	3.13
	White	96,961	0.90	3.24	0.90	3.11
NRC	NYC	71,676	0.90	3.25	0.91	3.13
	Big 4 Cities	7,921	0.89	3.25	0.89	3.13
	High Needs Urban/Suburban	15,208	0.88	3.28	0.89	3.17
	High Needs Rural	11,007	0.89	3.25	0.90	3.13
	Average Needs	56,084	0.89	3.26	0.90	3.13
	Low Needs	27,192	0.88	3.19	0.89	3.08
	Charter	7,368	0.87	3.31	0.88	3.22
	Private School	6,585	0.90	3.39	0.91	3.21
SWD	All Codes	28,051	0.88	3.21	0.88	3.10
SUA	All Codes	27,870	0.88	3.26	0.89	3.14
ELL	ELL=Y	17,494	0.85	3.27	0.86	3.17
SWD/SUA	SUA=504 plan codes	15,750	0.86	3.18	0.87	3.08
ELL/SUA	SUA & ELL codes	7,028	0.84	3.27	0.85	3.18

Table 28B. ELA Test Reliability by Subgroup, Grade 4

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	213,673	0.90	3.23	0.91	3.11
Gender	Female	105,061	0.90	3.23	0.91	3.11
	Male	108,612	0.91	3.22	0.91	3.10
Ethnicity	Asian	17,408	0.90	3.13	0.91	3.01
	Black	37,085	0.89	3.29	0.90	3.17
	Hispanic	49,439	0.89	3.29	0.90	3.18
	American Indian	1,135	0.90	3.25	0.91	3.12
	Multiracial	2,544	0.91	3.21	0.92	3.08
	Native Hawaiian or Other Pacific Islander	477	0.88	3.25	0.89	3.15
	White	105,585	0.90	3.19	0.91	3.07
NRC	NYC	69,892	0.90	3.24	0.91	3.12
	Big 4 Cities	7,878	0.90	3.27	0.91	3.15
	High Needs Urban/Suburban	15,432	0.90	3.28	0.90	3.18
	High Needs Rural	11,026	0.90	3.22	0.91	3.12
	Average Needs	57,920	0.90	3.19	0.91	3.09
	Low Needs	28,128	0.88	3.09	0.89	3.00
	Charter	5,732	0.87	3.25	0.88	3.18
	Private School	17,660	0.90	3.32	0.91	3.18
SWD	All Codes	30,278	0.89	3.23	0.89	3.12
SUA	All Codes	26,444	0.89	3.26	0.90	3.14
ELL	ELL=Y	14,634	0.85	3.31	0.86	3.19
SWD/SUA	SUA=504 plan codes	15,755	0.88	3.19	0.88	3.09
ELL/SUA	SUA & ELL codes	4,687	0.84	3.28	0.85	3.17

Table 28C. ELA Test Reliability by Subgroup, Grade 5

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	201,117	0.91	3.68	0.92	3.53
Gender	Female	98,798	0.91	3.67	0.91	3.53
	Male	102,319	0.91	3.67	0.92	3.52
Ethnicity	Asian	16,571	0.91	3.53	0.92	3.38
	Black	36,258	0.89	3.74	0.90	3.60
	Hispanic	46,141	0.89	3.73	0.90	3.60
	American Indian	1,111	0.90	3.72	0.91	3.58
	Multiracial	2,114	0.92	3.66	0.93	3.49
	Native Hawaiian or Other Pacific Islander	336	0.90	3.63	0.91	3.50
	White	98,586	0.91	3.64	0.92	3.48
NRC	NYC	68,029	0.91	3.69	0.92	3.53
	Big 4 Cities	7,697	0.90	3.70	0.90	3.57
	High Needs Urban/Suburban	14,849	0.90	3.73	0.90	3.60
	High Needs Rural	11,022	0.90	3.71	0.91	3.56
	Average Needs	57,619	0.90	3.67	0.91	3.52
	Low Needs	28,319	0.89	3.52	0.90	3.41
	Charter	6,942	0.88	3.73	0.88	3.63
	Private School	6,634	0.92	3.75	0.93	3.54
SWD	All Codes	30,841	0.88	3.65	0.88	3.54
SUA	All Codes	28,230	0.89	3.69	0.89	3.57
ELL	ELL=Y	14,247	0.85	3.72	0.86	3.61
SWD/SUA	SUA=504 plan codes	17,495	0.87	3.62	0.87	3.52
ELL/SUA	SUA & ELL codes	5,299	0.85	3.71	0.86	3.61

Table 28D. ELA Test Reliability by Subgroup, Grade 6

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	211,594	0.92	3.54	0.92	3.40
Gender	Female	103,943	0.91	3.51	0.92	3.38
	Male	107,651	0.92	3.54	0.92	3.41
Ethnicity	Asian	16,720	0.92	3.39	0.92	3.25
	Black	39,005	0.90	3.63	0.90	3.52
	Hispanic	48,240	0.90	3.61	0.90	3.50
	American Indian	1,086	0.91	3.58	0.91	3.46
	Multiracial	2,023	0.92	3.51	0.93	3.35
	Native Hawaiian or Other Pacific Islander	414	0.91	3.48	0.92	3.34
	White	104,106	0.91	3.46	0.92	3.33
NRC	NYC	67,443	0.91	3.56	0.92	3.43
	Big 4 Cities	7,784	0.91	3.65	0.91	3.53
	High Needs Urban/Suburban	14,797	0.90	3.60	0.91	3.50
	High Needs Rural	11,320	0.91	3.55	0.92	3.43
	Average Needs	59,256	0.91	3.48	0.92	3.36
	Low Needs	28,795	0.90	3.34	0.91	3.23
	Charter	6,498	0.87	3.61	0.88	3.54
	Private School	15,688	0.91	3.60	0.92	3.45
SWD	All Codes	31,025	0.88	3.59	0.89	3.49
SUA	All Codes	28,080	0.89	3.62	0.90	3.51
ELL	ELL=Y	10,694	0.84	3.62	0.85	3.51
SWD/SUA	SUA=504 plan codes	18,308	0.88	3.58	0.89	3.49
ELL/SUA	SUA & ELL codes	3,588	0.84	3.62	0.85	3.52

Table 28E. ELA Test Reliability by Subgroup, Grade 7

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	205,800	0.91	3.66	0.92	3.48
Gender	Female	100,426	0.90	3.61	0.91	3.45
	Male	105,374	0.91	3.68	0.92	3.50
Ethnicity	Asian	17,099	0.91	3.46	0.92	3.29
	Black	37,655	0.89	3.76	0.90	3.61
	Hispanic	45,309	0.89	3.74	0.90	3.59
	American Indian	1,019	0.90	3.73	0.91	3.57
	Multiracial	1,845	0.92	3.63	0.92	3.45
	Native Hawaiian or Other Pacific Islander	383	0.92	3.56	0.93	3.36
	White	102,490	0.91	3.56	0.92	3.41
NRC	NYC	68,732	0.90	3.69	0.91	3.51
	Big 4 Cities	7,590	0.90	3.80	0.91	3.62
	High Needs Urban/Suburban	14,810	0.90	3.75	0.91	3.59
	High Needs Rural	11,599	0.91	3.70	0.92	3.53
	Average Needs	59,859	0.91	3.60	0.91	3.45
	Low Needs	31,615	0.89	3.43	0.90	3.32
	Charter	5,394	0.87	3.72	0.87	3.62
	Private School	6,188	0.93	3.71	0.94	3.44
SWD	All Codes	30,831	0.88	3.74	0.89	3.60
SUA	All Codes	26,042	0.89	3.77	0.90	3.61
ELL	ELL=Y	10,132	0.84	3.74	0.85	3.61
SWD/SUA	SUA=504 plan codes	17,467	0.87	3.74	0.88	3.60
ELL/SUA	SUA & ELL codes	3,410	0.84	3.76	0.85	3.61

Table 28F. ELA Test Reliability by Subgroup, Grade 8

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	214,090	0.91	3.66	0.91	3.50
Gender	Female	104,876	0.90	3.60	0.91	3.46
	Male	109,214	0.91	3.69	0.92	3.53
Ethnicity	Asian	16,850	0.91	3.49	0.92	3.33
	Black	39,575	0.89	3.77	0.90	3.64
	Hispanic	46,975	0.89	3.77	0.90	3.63
	American Indian	1,044	0.90	3.74	0.90	3.60
	Multiracial	1,839	0.91	3.62	0.92	3.47
	Native Hawaiian or Other Pacific Islander	392	0.91	3.66	0.91	3.50
	White	107,415	0.90	3.54	0.91	3.41
NRC	NYC	68,108	0.90	3.73	0.91	3.57
	Big 4 Cities	7,373	0.90	3.78	0.91	3.62
	High Needs Urban/Suburban	14,551	0.90	3.75	0.90	3.62
	High Needs Rural	11,632	0.90	3.65	0.91	3.51
	Average Needs	59,860	0.90	3.58	0.91	3.45
	Low Needs	31,455	0.88	3.40	0.89	3.31
	Charter	4,424	0.86	3.70	0.86	3.63
	Private School	16,649	0.90	3.62	0.91	3.47
SWD	All Codes	30,367	0.88	3.76	0.89	3.62
SUA	All Codes	24,449	0.89	3.77	0.90	3.62
ELL	ELL=Y	10,964	0.83	3.80	0.85	3.65
SWD/SUA	SUA=504 plan codes	16,591	0.88	3.76	0.89	3.61
ELL/SUA	SUA & ELL codes	2,840	0.83	3.78	0.84	3.63

Table 29A. Mathematics Test Reliability by Subgroup, Grade 3

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	204,504	0.93	3.56	0.94	3.36
Gender	Female	99,775	0.92	3.57	0.93	3.37
	Male	104,729	0.93	3.55	0.94	3.34
Ethnicity	Asian	17,939	0.93	3.24	0.94	3.01
	Black	36,064	0.92	3.59	0.93	3.42
	Hispanic	49,162	0.92	3.61	0.93	3.44
	American Indian	1,102	0.93	3.60	0.93	3.41
	Multiracial	2,794	0.93	3.56	0.94	3.34
	Native Hawaiian or Other Pacific Islander	445	0.93	3.49	0.94	3.28
	White	96,998	0.92	3.52	0.93	3.33
NRC	NYC	72,758	0.93	3.54	0.94	3.32
	Big 4 Cities	7,980	0.92	3.59	0.93	3.43
	High Needs Urban/Suburban	15,366	0.92	3.63	0.93	3.46
	High Needs Rural	11,031	0.92	3.63	0.92	3.47
	Average Needs	56,089	0.92	3.56	0.93	3.38
	Low Needs	27,281	0.92	3.39	0.92	3.21
	Charter	7,452	0.91	3.48	0.92	3.28
SWD	All Codes	28,832	0.92	3.58	0.93	3.43
SUA	All Codes	33,593	0.92	3.60	0.93	3.44
SWD/SUA	SUA=504 plan codes	17,938	0.91	3.57	0.92	3.43
ELL/SUA	SUA & ELL codes	10,145	0.91	3.58	0.92	3.43
ELL	English	15,884	0.92	3.58	0.92	3.43
	Chinese	575	0.92	3.41	0.93	3.22
	Haitian Creole	44	0.91	3.40	0.92	3.25
	Korean	45	0.92	3.03	0.93	2.88
	Russian	73	0.94	3.53	0.95	3.32
	Spanish	2,895	0.89	3.58	0.90	3.45
	All Translations	3,632	0.92	3.58	0.93	3.42

Table 29B. Mathematics Test Reliability by Subgroup, Grade 4

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	215,394	0.93	3.73	0.94	3.50
Gender	Female	105,936	0.92	3.74	0.93	3.51
	Male	109,458	0.93	3.72	0.94	3.48
Ethnicity	Asian	17,924	0.93	3.53	0.94	3.26
	Black	37,219	0.91	3.71	0.92	3.52
	Hispanic	50,239	0.91	3.72	0.92	3.53
	American Indian	1,140	0.92	3.73	0.93	3.52
	Multiracial	2,537	0.93	3.71	0.94	3.47
	Native Hawaiian or Other Pacific Islander	476	0.92	3.74	0.93	3.49
	White	105,859	0.92	3.72	0.93	3.50
NRC	NYC	70,981	0.93	3.74	0.94	3.47
	Big 4 Cities	8,030	0.91	3.63	0.92	3.46
	High Needs Urban/Suburban	15,566	0.91	3.68	0.92	3.51
	High Needs Rural	11,027	0.91	3.72	0.92	3.53
	Average Needs	57,933	0.92	3.72	0.93	3.51
	Low Needs	28,173	0.92	3.61	0.93	3.39
	Charter	5,806	0.91	3.71	0.92	3.49
SWD	All Codes	31,161	0.91	3.60	0.92	3.45
SUA	All Codes	30,895	0.91	3.62	0.92	3.47
SWD/SUA	SUA=504 plan codes	18,082	0.90	3.55	0.90	3.42
ELL/SUA	SUA & ELL codes	6,608	0.90	3.57	0.90	3.43
ELL	English	13,476	0.90	3.66	0.91	3.50
	Chinese	538	0.91	3.72	0.92	3.50
	Haitian Creole	52	0.86	3.49	0.88	3.32
	Korean	31	0.95	3.43	0.96	2.87
	Russian	76	0.93	3.79	0.93	3.55
	Spanish	2,661	0.87	3.54	0.88	3.43
	All Translations	3,358	0.92	3.65	0.93	3.46

Table 29C. Mathematics Test Reliability by Subgroup, Grade 5

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	202,579	0.93	3.66	0.94	3.44
Gender	Female	99,557	0.92	3.66	0.93	3.44
	Male	103,022	0.93	3.67	0.94	3.43
Ethnicity	Asian	17,003	0.93	3.47	0.94	3.18
	Black	36,425	0.91	3.61	0.92	3.46
	Hispanic	47,043	0.91	3.64	0.92	3.47
	American Indian	1,122	0.92	3.67	0.93	3.45
	Multiracial	2,130	0.93	3.68	0.94	3.45
	Native Hawaiian or Other Pacific Islander	345	0.92	3.66	0.93	3.44
	White	98,511	0.92	3.66	0.93	3.43
NRC	NYC	69,104	0.93	3.65	0.94	3.40
	Big 4 Cities	7,816	0.91	3.57	0.92	3.42
	High Needs Urban/Suburban	14,992	0.91	3.63	0.92	3.48
	High Needs Rural	11,042	0.91	3.65	0.92	3.48
	Average Needs	57,641	0.92	3.66	0.93	3.46
	Low Needs	28,355	0.92	3.58	0.93	3.35
	Charter	7,079	0.92	3.61	0.92	3.43
SWD	All Codes	31,675	0.90	3.53	0.91	3.41
SUA	All Codes	32,552	0.91	3.58	0.91	3.44
SWD/SUA	SUA=504 plan codes	19,571	0.89	3.50	0.90	3.39
ELL/SUA	SUA & ELL codes	7,523	0.90	3.58	0.91	3.43
ELL	English	12,890	0.90	3.58	0.90	3.45
	Chinese	494	0.91	3.65	0.93	3.38
	Haitian Creole	75	0.74	3.45	0.74	3.40
	Korean	31	0.92	3.44	0.93	3.12
	Russian	84	0.93	3.65	0.94	3.42
	Spanish	2,836	0.86	3.46	0.87	3.37
	All Translations	3,520	0.92	3.58	0.93	3.40

Table 29D. Mathematics Test Reliability by Subgroup, Grade 6

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	213,147	0.94	3.87	0.94	3.67
Gender	Female	104,829	0.93	3.87	0.94	3.68
	Male	108,318	0.94	3.86	0.95	3.65
Ethnicity	Asian	17,218	0.95	3.56	0.95	3.33
	Black	39,107	0.92	3.93	0.93	3.75
	Hispanic	49,176	0.92	3.91	0.93	3.75
	American Indian	1,092	0.92	3.90	0.93	3.73
	Multiracial	2,027	0.94	3.86	0.94	3.66
	Native Hawaiian or Other Pacific Islander	424	0.94	3.79	0.95	3.59
	White	104,103	0.93	3.81	0.94	3.63
NRC	NYC	68,465	0.94	3.88	0.95	3.65
	Big 4 Cities	7,933	0.92	3.90	0.93	3.71
	High Needs Urban/Suburban	14,762	0.92	3.90	0.92	3.74
	High Needs Rural	11,435	0.92	3.88	0.92	3.72
	Average Needs	59,186	0.93	3.84	0.94	3.66
	Low Needs	28,777	0.93	3.69	0.94	3.50
	Charter	6,701	0.92	3.89	0.93	3.72
SWD	All Codes	31,841	0.90	3.86	0.91	3.70
SUA	All Codes	32,323	0.91	3.88	0.92	3.72
SWD/SUA	SUA=504 plan codes	20,628	0.89	3.84	0.90	3.69
ELL/SUA	SUA & ELL codes	5,607	0.91	3.82	0.91	3.68
ELL	English	9,301	0.91	3.87	0.92	3.72
	Chinese	642	0.94	3.71	0.94	3.52
	Haitian Creole	143	0.85	3.75	0.86	3.61
	Korean	47	0.95	3.49	0.96	3.19
	Russian	95	0.89	3.93	0.89	3.83
	Spanish	2,835	0.87	3.78	0.88	3.65
	All Translations	3,762	0.93	3.85	0.94	3.67

Table 29E. Mathematics Test Reliability by Subgroup, Grade 7

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	207,251	0.93	3.97	0.94	3.68
Gender	Female	101,198	0.93	3.97	0.94	3.68
	Male	106,053	0.94	3.97	0.95	3.68
Ethnicity	Asian	17,553	0.95	3.69	0.95	3.44
	Black	37,786	0.91	3.92	0.92	3.69
	Hispanic	46,251	0.91	3.97	0.92	3.72
	American Indian	1,027	0.92	3.98	0.93	3.70
	Multiracial	1,836	0.94	3.96	0.95	3.67
	Native Hawaiian or Other Pacific Islander	398	0.95	3.88	0.96	3.57
	White	102,400	0.93	3.94	0.94	3.69
NRC	NYC	70,020	0.94	3.96	0.95	3.66
	Big 4 Cities	7,716	0.90	3.82	0.91	3.60
	High Needs Urban/Suburban	14,955	0.91	3.95	0.92	3.69
	High Needs Rural	11,648	0.91	4.00	0.92	3.74
	Average Needs	59,652	0.92	3.97	0.93	3.71
	Low Needs	31,525	0.93	3.81	0.94	3.60
	Charter	5,355	0.92	3.95	0.93	3.72
SWD	All Codes	31,379	0.88	3.75	0.89	3.58
SUA	All Codes	26,091	0.89	3.82	0.91	3.62
SWD/SUA	SUA=504 plan codes	17,390	0.86	3.74	0.88	3.57
ELL/SUA	SUA & ELL codes	3,805	0.90	3.69	0.91	3.51
ELL	English	8,372	0.88	3.73	0.89	3.58
	Chinese	703	0.93	3.81	0.94	3.61
	Haitian Creole	157	0.73	3.54	0.74	3.46
	Korean	36	0.94	3.82	0.94	3.56
	Russian	129	0.93	3.89	0.94	3.64
	Spanish	2,910	0.82	3.61	0.83	3.53
	All Translations	3,935	0.92	3.79	0.93	3.57

Table 29F. Mathematics Test Reliability by Subgroup, Grade 8

Group	Subgroup	N-count	Cronbach's Alpha	SEM of Cronbach	Feldt-Raju Coefficient	SEM of Feldt-Raju
State	All Students	214,932	0.93	3.94	0.94	3.71
Gender	Female	105,389	0.93	3.94	0.94	3.72
	Male	109,543	0.94	3.93	0.94	3.70
Ethnicity	Asian	17,282	0.94	3.64	0.95	3.42
	Black	39,633	0.91	3.96	0.92	3.76
	Hispanic	47,763	0.92	3.97	0.92	3.77
	American Indian	1,038	0.92	3.97	0.92	3.77
	Multiracial	1,817	0.93	3.93	0.94	3.71
	Native Hawaiian or Other Pacific Islander	410	0.94	3.91	0.95	3.69
	White	106,989	0.93	3.88	0.93	3.69
NRC	NYC	69,217	0.94	3.94	0.95	3.69
	Big 4 Cities	7,492	0.90	3.85	0.91	3.67
	High Needs Urban/Suburban	14,651	0.91	3.97	0.92	3.76
	High Needs Rural	11,707	0.91	3.97	0.91	3.78
	Average Needs	59,339	0.92	3.90	0.93	3.71
	Low Needs	31,249	0.92	3.73	0.93	3.57
	Charter	4,405	0.92	3.91	0.93	3.72
SWD	All Codes	30,749	0.89	3.82	0.90	3.66
SUA	All Codes	25,996	0.90	3.88	0.91	3.70
SWD/SUA	SUA=504 plan codes	17,166	0.88	3.82	0.89	3.66
ELL/SUA	SUA & ELL codes	3,882	0.92	3.80	0.93	3.62
ELL	English	8,879	0.90	3.84	0.91	3.69
	Chinese	776	0.93	3.79	0.94	3.57
	Haitian Creole	165	0.89	3.74	0.90	3.59
	Korean	33	0.94	3.70	0.95	3.41
	Russian	109	0.93	3.90	0.94	3.64
	Spanish	3,186	0.86	3.76	0.87	3.65
	All Translations	4,269	0.93	3.86	0.94	3.66

Standard Error of Measurement

The SEMs, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, are presented in Tables 21 and 23 for ELA and mathematics, respectively. The SEMs ranged 3.11–3.97 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 Tables 28A-29F. The SEMs associated with all reliability estimates for all subjects, grades, methods of estimation, and subpopulations are in the range 2.87–4.00, 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 Agreement

This subsection describes the analyses conducted to estimate performance level classification consistency and agreement 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 Agreement 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 tests that classify students into performance levels. 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 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 R.

Classification consistency and agreement were estimated using the IRT procedure suggested by Lee, Hanson, and Brennan (2002) and Wang, Kolen, and Harris (2000). Appendix Q 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. Tables 30 and 31 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. The inconsistency index is equal to the "1 – agreement index." Kappa is a measure of agreement corrected for chance.

Table 30 depicts the ELA and Mathematics consistency study results based on the range of performance levels for all grades. For ELA, between 71% and 73% 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 .59 to .61. These are between "moderate" and "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For Mathematics, between 75 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 .65 to .70. 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 30. Decision Consistency (All Cuts)

	Grade	N-count	Agreement	Kappa
ELA				
	3	203,041	.73	.61
	4	213,673	.72	.60
	5	201,117	.73	.61
	6	211,594	.73	.61
	7	205,800	.73	.61
	8	214,090	.71	.59
Math				
	3	202,095	.75	.65
	4	210,259	.76	.66
	5	200,037	.77	.67
	6	207,270	.78	.68
	7	204,447	.80	.70
	8	208,371	.79	.69

Table 31 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, between 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 .73 to .75. These are considered "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. In addition, between 91% and 93% 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 .79 to .82. As with ELA, these statistics indicate "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa.

Table 31. Decision Consistency (Level III Cut)

Grade	N-count	Agreement	Kappa
ELA			
3	203,041	.89	.74
4	213,673	.90	.75
5	201,117	.90	.75
6	211,594	.90	.75
7	205,800	.89	.75
8	214,090	.88	.73
Mathematics			
3	202,095	.91	.79
4	210,259	.91	.80
5	200,037	.92	.80
6	207,270	.92	.82
7	204,447	.93	.82
8	208,371	.93	.81

Agreement

The results of classification agreement for ELA and mathematics across all grades are presented in Table 32. Included in the table are case counts (N-count) and classification agreement for all performance levels (All Cuts) and for the Level III cut score. Note that, by definition, agreement 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 agreement rates indicate that the categorization of a student's observed performance is in agreement with the location of his or her underlying proficiency between 79% and 81% of the time across all performance levels and approximately 92%–93% of the time in regards to the Level III cut score. For Mathematics, the estimated agreement 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 between 94%–95% of the time in regard to the Level III cut score.

Table 32. Decision Agreement Estimates

Grade	N-count	Agreement	
		All Cuts	Level III Cut
ELA			
3	203,041	**.80**	**.92**
4	213,673	**.80**	**.92**
5	201,117	**.81**	**.93**
6	211,594	**.80**	**.93**
7	205,800	**.80**	**.92**
8	214,090	**.79**	**.92**
Mathematics			
3	202,095	**.82**	**.94**
4	210,259	**.83**	**.94**
5	200,037	**.84**	**.94**
6	207,270	**.83**	**.94**
7	204,447	**.86**	**.95**
8	208,371	**.85**	**.95**

Section 8: Standard-Setting

To set performance standards for the new assessments, standard-setting was conducted in summer 2013. The goals of standard-setting were as follows:

- Provide vertically articulated performance standards for the assessments in ELA and mathematics and indicate the degree to which students have met the standards for their grade.
- Incorporate college and career readiness into the performance standards for the new assessments.
- Recommend rigorous and attainable performance standards.
- Incorporate existing and future policy considerations relevant to New York State's educational system into the established performance standards.

With these goals in mind, the process for setting performance standards was expanded beyond the process that was used for the past Grades 3–8 ELA and mathematics assessments that were aligned to the New York State 2005 standards. Specific extensions to the standard-setting process included taking into account the indicators of college readiness to the degree practicable and using external benchmark data points to help inform performance standards.

The following steps were used as the standard-setting process:

1. Identify external benchmarks for the assessments using empirical studies of established, national testing programs.
2. Develop policy- and grade-level-specific performance level descriptors (PLDs).
3. Convene standard-setting committees.
4. Conduct vertical articulation.
5. Approve performance standards.

The evidence-based standard-setting process combines content, data, educator expertise, and policy expertise with a standard-setting procedure. Sources of evidence by category include:

- **Content**: performance labels, policy definitions, curriculum, assessment blueprints, and specific PLDs;
- **Data**: empirically-defined, external benchmarks that use established college-readiness performance levels on existing national tests;
- **Educator expertise**: standard-setting committee reviews of content and data;
- **Policy framework**: review study data and policy implications to help make recommendations for performance standards; and
- **Standard-setting procedure**: use of the Bookmark standard-setting methodology to recommend standards.

The cut score recommendations that came out of the standard-setting committee's work (in Step 3 above) were affirmed during vertical articulation (in Step 4 above) and were then approved by the Commissioner of Education without any further changes (in Step 5 above). The raw score cuts associated with committee recommendation are shown in Tables 33 and 34 for ELA and mathematics, respectively, along with the corresponding scale score cuts.

Table 33. 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)	26 (287)	32 (289)	32 (283)	35 (287)	35 (284)
NYS Level III	35 (320)	38 (320)	45 (320)	47 (320)	48 (318)	48 (316)
NYS Level IV	46 (358)	45 (343)	54 (346)	53 (338)	57 (347)	56 (343)

Table 34. 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	29 (285)	28 (283)	31 (294)	31 (284)	28 (293)	29 (287)
NYS Level III	44 (314)	43 (314)	44 (319)	50 (318)	46 (322)	49 (322)
NYS Level IV	53 (340)	55 (341)	56 (346)	61 (340)	60 (348)	62 (349)

Appendix P presents the full standard-setting report that describes the general process, composition of the committees, ratings from the various rounds, evaluation forms, and other materials.

Section 9: Summary of Operational Test Results

This section summarizes the distribution of operational scale score results on the NYSTP 2013 Grades 3–8 Common Core ELA and Mathematics Tests. These include the scale score means, standard deviations, percentiles, 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 accommodation. 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 R.

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 groups with small number counts that are included in the scale score summaries.

ELA Scale Score and Subscore Distributions

Table 35 shows some key statistics characterizing the distribution of ELA scale scores, while Table 36 summarizes the ELA subscores derived from the test in each grade. Tables 37A-37F 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 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. This pattern of achievement was consistent across all grades.

Table 35. ELA Scale Score Distribution Summary

Grade	N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
3	203,041	299.68	35.24	253	279	303	322	342
4	213,673	299.71	35.23	254	281	303	323	339
5	201,117	299.58	35.25	256	279	301	322	343
6	211,594	299.53	35.31	256	278	302	323	341
7	205,800	299.41	35.40	255	280	301	324	343
8	214,090	299.52	35.40	256	279	301	322	343

Table 36. ELA Subscore Summary

Grade	N-count	Subscore	Max	Mean	SD
3	203,041	Reading	31	16.93	6.13
		Writing	24	11.57	4.97
4	213,673	Reading	31	18.58	6.26
		Writing	24	12.48	5.07
5	201,117	Reading	42	24.04	7.69
		Writing	24	12.78	5.38
6	211,594	Reading	41	24.73	7.80
		Writing	24	14.00	5.26
7	205,800	Reading	42	25.40	7.39
		Writing	24	14.73	5.62
8	214,090	Reading	42	25.17	7.51
		Writing	24	15.81	5.39

ELA Grade 3

Scale score statistics and N-counts of demographic groups for Grade 3 are presented in Table 37A. The population scale score mean was 299.68 with a standard deviation of 35.24. By gender, Females outperformed Males by around 9 scale score points. Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the average scale score, as did students from Low Needs, Average Needs, and Charter districts. Among all ethnic groups, the Asian ethnic group had the highest average scale score mean (315.03). Students from the Big 4 Cities achieved a lower scale score mean than their peers from schools with other NRC designations, and it was about four-fifths of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, slightly less than one standard deviation below the mean scale score for the population. The ELL subgroup, which had a scale score mean about 26 scale score points below the population mean, was the lowest performing group analyzed. At the 50th percentile, the following groups exceeded that of the population (303): Female (305), Asian (317), Multiracial (305), Native Hawaiian or Other Pacific Islander (308), White (308), Average Needs districts (305), and Low Needs districts (320).

Table 37A. ELA Scale Score Distribution by Subgroup, Grade 3

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	203,041	299.68	35.24	253	279	303	322	342
Gender	Female	99,173	304.32	33.89	262	283	305	329	346
	Male	103,868	295.26	35.93	248	273	300	320	338
Ethnicity	Asian	17,411	315.03	33.35	273	294	317	338	354
	Black	35,947	287.80	34.32	242	266	291	311	329
	Hispanic	48,391	289.81	33.15	248	269	291	314	329
	American Indian	1,099	292.51	34.26	248	273	294	314	335
	Multiracial	2,793	301.60	37.68	253	279	305	329	346
	Native Hawaiian or Other Pacific Islander	439	305.24	35.23	258	286	308	329	346
	White	96,961	306.26	34.19	262	288	308	329	346
NRC	New York City	71,676	297.76	34.76	253	276	300	322	342
	Big 4 Cities	7,921	274.30	36.63	220	253	276	300	320
	High Needs Urban/Suburban	15,208	287.44	34.16	242	266	291	311	329
	High Needs Rural	11,007	290.61	34.97	242	269	294	314	332
	Average Needs	56,084	302.83	34.05	258	283	305	326	342
	Low Needs	27,192	315.94	31.18	276	300	320	335	354
	Charter	7,368	301.42	29.68	262	283	303	322	338
	Private School	6,585	298.77	37.98	248	276	303	326	342
SWD	All Codes	28,051	266.47	36.42	212	242	269	291	311
SUA	All Codes	27,870	271.89	36.15	220	248	276	297	317
ELL	ELL=Y	17,494	274.07	31.90	228	258	276	297	311
SWD/SUA	SUA=504 plan codes	15,750	261.12	35.70	212	235	266	286	305
ELL/SUA	SUA & ELL codes	7,028	271.57	31.79	228	253	276	294	308

ELA Grade 4

Scale score statistics and N-counts of demographic groups for Grade 4 are presented in Table 37B. The Grade 4 population (All Students) mean was 299.71, with a standard deviation of 35.23. By gender, Females outperformed Males by around 9 scale score points. Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the average scale score, as did students from Low Needs and Average Needs districts, charter, and private schools. Among all ethnic groups, the Asian ethnic group had the highest average scale score mean (314.85). Students from the Big 4 Cities achieved a lower scale score mean than their peers from schools with other NRC designations, and it was about three-fourths of a standard deviation below the population mean. The SWD/SUA subgroup had a scale score mean nearly 42 scale score points below the population mean, and was at or below the scale score of all listed percentiles for any other subgroup, with the exception of the 90th percentile. At the 50th percentile, the following groups exceeded that of the population (303): Female (306), Asian (317), Native Hawaiian or Other Pacific Islander (306), White (309), Average Needs districts (306), Low Needs districts (320), and Private Schools (309).

Table 37B. ELA Scale Score Distribution by Subgroup, Grade 4

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	213,673	299.71	35.23	254	281	303	323	339
Gender	Female	105,061	304.54	33.83	261	284	306	326	343
	Male	108,612	295.05	35.91	249	275	298	320	336
Ethnicity	Asian	17,408	314.85	33.60	275	298	317	336	352
	Black	37,085	288.29	34.55	245	268	292	311	329
	Hispanic	49,439	289.29	33.90	245	272	292	311	329
	American Indian	1,135	290.67	36.11	245	272	295	314	332
	Multiracial	2,544	300.64	35.97	254	281	303	326	343
	Native Hawaiian or Other Pacific Islander	477	302.41	30.55	261	284	306	323	339
	White	105,585	306.17	33.78	265	287	309	329	343
NRC	New York City	69,892	297.14	35.35	254	278	301	320	339
	Big 4 Cities	7,878	273.77	37.50	220	249	278	301	320
	High Needs Urban/Suburban	15,432	285.84	34.93	240	265	290	309	326
	High Needs Rural	11,026	289.85	34.96	245	272	292	314	329
	Average Needs	57,920	302.64	33.45	261	284	306	326	339
	Low Needs	28,128	315.82	30.14	278	301	320	336	352
	Charter	5,732	301.15	29.05	265	284	303	320	336
	Private School	17,660	304.06	34.73	261	284	309	326	343
SWD	All Codes	30,278	263.22	37.28	211	240	265	290	309
SUA	All Codes	26,444	267.88	37.16	211	245	272	295	311
ELL	ELL=Y	14,634	265.44	33.51	220	245	268	290	303
SWD/SUA	SUA=504 plan codes	15,755	257.73	36.76	203	234	261	284	303
ELL/SUA	SUA & ELL codes	4,687	261.63	33.35	211	240	265	287	301

ELA Grade 5

Scale score summary statistics for Grade 5 students are in Table 37C. Overall, the scale score mean was 299.58, with a standard deviation of 35.25. The difference between mean scale scores by gender groups was about 8 scale score points, with Females outperforming Males. Female, Asian, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the population mean scale score, as did students from Low Needs and Average Needs districts. The students from the Big 4 Cities scored below their peers from schools with other NRC designations, and it was about three-fourths of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored approximately one standard deviation below the mean scale score for the population. The SWD/SUA subgroup, which had a scale score mean nearly 44 scale score points below the population mean, was the lowest performing group analyzed. At the 50th percentile, the following groups exceeded that of the population (301): Female (305), Native Hawaiian or Other Pacific Islander (310), White (308), Average Needs districts (305), and Low Needs districts (317).

Table 37C. ELA Scale Score Distribution by Subgroup, Grade 5

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	201,117	299.58	35.25	256	279	301	322	343
Gender	Female	98,798	303.83	33.87	263	282	305	327	346
	Male	102,319	295.48	36.06	248	274	298	320	339
Ethnicity	Asian	16,571	317.60	34.54	274	298	320	339	359
	Black	36,258	288.09	33.02	248	269	289	310	327
	Hispanic	46,141	290.14	32.83	248	271	292	312	330
	American Indian	1,111	289.88	35.33	244	271	294	315	330
	Multiracial	2,114	299.13	37.48	252	277	301	325	343
	Native Hawaiian or Other Pacific Islander	336	306.58	32.93	266	287	310	327	346
	White	98,586	305.30	34.67	263	287	308	327	346
NRC	New York City	68,029	299.00	34.83	256	277	301	322	343
	Big 4 Cities	7,697	274.01	35.99	228	252	277	298	317
	High Needs Urban/Suburban	14,849	285.92	34.13	244	266	289	310	327
	High Needs Rural	11,022	289.86	35.03	244	269	292	315	330
	Average Needs	57,619	301.96	33.73	259	282	305	325	343
	Low Needs	28,319	315.96	30.82	277	298	317	336	354
	Charter	6,942	295.42	28.79	259	277	296	315	330
	Private School	6,634	295.85	41.33	239	271	301	325	343
SWD	All Codes	30,841	265.05	34.86	220	244	266	289	308
SUA	All Codes	28,230	269.32	35.52	220	248	271	294	312
ELL	ELL=Y	14,247	267.77	31.74	228	248	271	289	305
SWD/SUA	SUA=504 plan codes	17,495	260.24	34.65	212	239	263	284	303
ELL/SUA	SUA & ELL codes	5,299	266.12	31.67	228	248	269	287	303

ELA Grade 6

Scale score summary statistics for Grade 6 students are in Table 37D. The scale score mean was 299.53, with a standard deviation of 35.31. Females tended to outperform males by around 10 scale score points. Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the population mean scale score, as did students from Low Needs, Average Needs, and Private School districts. The students from the Big 4 Cities scored below their peers from schools with other NRC designations, and it was slightly less than three-fourths of a standard deviation below the population mean. The SWD and SUA subgroups scored about one standard deviation below the mean scale score for the population. The ELL subgroup, which had a scale score mean of more than 42 scale score points below the population mean, was the lowest performing group analyzed. At the 50th percentile, the following groups exceeded that of the population (302): Female (305), Asian (317), Native Hawaiian or Other Pacific Islander (309), White (309), Average Needs districts (307), and Low Needs districts (320).

Table 37D. ELA Scale Score Distribution by Subgroup, Grade 6

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	211,594	299.53	35.31	256	278	302	323	341
Gender	Female	103,943	304.36	33.82	262	283	305	328	345
	Male	107,651	294.86	36.09	250	273	297	320	338
Ethnicity	Asian	16,720	314.11	35.40	270	295	317	338	354
	Black	39,005	285.49	32.61	246	265	288	307	325
	Hispanic	48,240	287.77	32.14	246	268	290	309	325
	American Indian	1,086	289.34	34.06	246	268	293	312	331
	Multiracial	2,023	301.44	38.34	256	278	302	325	349
	Native Hawaiian or Other Pacific Islander	414	307.92	32.68	262	285	309	331	349
	White	104,106	307.93	34.17	265	288	309	331	349
NRC	New York City	67,443	293.90	34.83	250	273	295	317	338
	Big 4 Cities	7,784	274.59	36.43	230	253	276	300	320
	High Needs Urban/Suburban	14,797	287.15	33.97	243	268	290	309	328
	High Needs Rural	11,320	294.08	34.68	250	276	297	317	335
	Average Needs	59,256	304.50	33.65	262	285	307	325	345
	Low Needs	28,795	317.64	31.49	278	300	320	338	354
	Charter	6,498	291.72	27.45	256	276	293	309	325
	Private School	15,688	302.90	35.18	259	283	305	325	345
SWD	All Codes	31,025	262.99	33.89	216	243	265	285	305
SUA	All Codes	28,080	268.07	34.81	224	246	270	293	309
ELL	ELL=Y	10,694	257.15	30.65	216	239	259	278	293
SWD/SUA	SUA=504 plan codes	18,308	260.77	34.31	216	239	262	285	302
ELL/SUA	SUA & ELL codes	3,588	255.41	31.01	216	239	259	276	293

ELA Grade 7

Scale score statistics and N-counts of demographic groups for Grade 7 are presented in Table 37E. The population scale score mean was 299.41, and the population standard deviation was 35.40. By gender, Females outperformed Males by about one-third of a standard deviation. Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the population mean scale score, as did students from Low Needs and Average Needs districts. Among all ethnic groups, the Asian ethnic group had the highest average scale score mean (315.08). The students from the Big 4 Cities scored below their peers from schools with other NRC designations, and it was about three-fourths of a standard deviation below the population mean. The SWD and SUA subgroups scored approximately one standard deviation below the mean scale score for the population. The ELL subgroup, which had a scale score mean of about 44 scale score points below the population mean, was the lowest performing group analyzed. At the 50th percentile, the following groups exceeded that of the population (301): Female (306), Asian (318), Multiracial (304), Native Hawaiian or Other Pacific Islander (311), White (311), Average Needs districts (306), and Low Needs districts (318).

Table 37E. ELA Scale Score Distribution by Subgroup, Grade 7

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	205,800	299.41	35.40	255	280	301	324	343
Gender	Female	100,426	304.66	33.35	264	285	306	327	343
	Male	105,374	294.40	36.55	249	272	296	318	340
Ethnicity	Asian	17,099	315.08	34.82	272	296	318	340	357
	Black	37,655	285.76	32.73	245	267	287	308	324
	Hispanic	45,309	287.87	32.65	249	270	289	308	327
	American Indian	1,019	289.53	34.34	249	270	292	311	333
	Multiracial	1,845	301.27	36.95	258	280	304	327	343
	Native Hawaiian or Other Pacific Islander	383	307.38	35.91	261	282	311	333	352
	White	102,490	306.94	34.37	264	287	311	330	347
NRC	New York City	68,732	295.21	34.29	252	275	296	318	336
	Big 4 Cities	7,590	272.86	37.32	223	249	275	299	318
	High Needs Urban/Suburban	14,810	285.67	34.41	241	267	287	308	327
	High Needs Rural	11,599	292.39	35.08	249	272	294	316	333
	Average Needs	59,859	304.24	33.78	261	285	306	327	343
	Low Needs	31,615	316.26	30.71	280	299	318	336	352
	Charter	5,394	295.69	27.48	261	280	296	313	330
	Private School	6,188	295.15	43.92	233	275	301	324	343
SWD	All Codes	30,831	264.29	34.45	223	245	267	287	306
SUA	All Codes	26,042	268.22	35.71	223	249	270	292	311
ELL	ELL=Y	10,132	255.63	31.96	215	237	258	277	292
SWD/SUA	SUA=504 plan codes	17,467	261.63	34.69	215	241	264	285	304
ELL/SUA	SUA & ELL codes	3,410	254.21	32.12	215	233	255	275	292

ELA Grade 8

Scale score statistics and N-counts of demographic groups for Grade 8 are presented in Table 37F. The population scale score mean was 299.52, with a standard deviation of 35.40. By gender, Females outperformed Males by about 9 scale score points. Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the population mean scale score, as did students from Low Needs and Average Needs districts and Private Schools. The students from the Big 4 Cities scored below their peers from schools with other NRC designations, and it was a bit less than a standard deviation below the population mean. The SWD and SUA subgroups scored approximately one standard deviation below the mean scale score for the population. The ELL subgroup, which had a scale score mean of more than 45 scale score points below the population mean, was the lowest performing group analyzed. At the 50th percentile, the following groups exceeded that of the population (301): Female (306), Asian (316), Multiracial (303), Native Hawaiian or Other Pacific Islander (306), White (311), Low Needs districts (319), Average Needs districts (308), and Private Schools (308).

Table 37F. ELA Scale Score Distribution by Subgroup, Grade 8

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	214,090	299.52	35.40	256	279	301	322	343
Gender	Female	104,876	303.86	33.83	261	284	306	325	343
	Male	109,214	295.35	36.37	249	275	298	319	339
Ethnicity	Asian	16,850	312.76	36.23	267	294	316	335	353
	Black	39,575	284.50	33.12	243	267	286	306	325
	Hispanic	46,975	287.13	33.39	246	269	289	308	328
	American Indian	1,044	289.75	33.10	246	269	291	313	328
	Multiracial	1,839	301.95	35.89	259	279	303	325	343
	Native Hawaiian or Other Pacific Islander	392	302.04	35.50	259	282	306	325	343
	White	107,415	308.43	33.24	267	289	311	331	348
NRC	New York City	68,108	291.94	35.18	249	272	294	316	335
	Big 4 Cities	7,373	271.07	37.87	226	249	272	296	316
	High Needs Urban/Suburban	14,551	287.33	33.93	246	267	289	311	328
	High Needs Rural	11,632	294.30	34.45	249	275	296	316	335
	Average Needs	59,860	305.11	33.34	264	286	308	328	343
	Low Needs	31,455	317.60	30.19	282	301	319	339	353
	Charter	4,424	295.02	27.06	261	277	296	313	328
	Private School	16,649	304.51	33.99	264	286	308	328	343
SWD	All Codes	30,367	263.46	34.97	218	243	267	286	306
SUA	All Codes	24,449	269.10	36.06	226	249	272	294	311
ELL	ELL=Y	10,964	254.19	32.00	211	235	259	277	291
SWD/SUA	SUA=504 plan codes	16,591	262.98	35.15	218	243	267	286	306
ELL/SUA	SUA & ELL codes	2,840	251.67	32.43	203	235	256	275	289

Mathematics Scale Score Distributions

Table 38 shows some key statistics characterizing the distribution of mathematics scale scores, while Table 39 summarizes the mathematics subscores derived from the test in each grade. Tables 40A-40F break down the scale scores by selected subgroups. Some general observations from the mathematics data are as follows: Females and Males performed fairly consistently; Asians scored considerably higher than other 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 often outperforming other school types. Students taking the Chinese and Korean translations met or exceeded the population at every reported percentile, whereas the other translation subgroups (Haitian-Creole, Spanish, and Russian) were below the population scale score at each percentile; and ELLs, SWDs, and/or SUAs achieved below the State aggregate (All Students) in most percentiles. This pattern of achievement was consistent across all grades.

Table 38. Mathematics Scale Score Distribution Summary

Grade	N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
3	204,504	299.78	35.13	256	279	300	321	345
4	215,394	299.84	35.15	255	278	300	324	344
5	202,579	299.77	35.25	255	278	300	323	343
6	213,147	299.56	35.11	256	279	301	322	342
7	207,251	299.72	35.28	253	280	302	324	342
8	214,932	299.43	35.23	255	280	302	324	341

Table 39. Mathematics Subscore Summary

Grade	N-count	Subscore	Max	Mean	SD
3	204504	Operations and Algebraic Thinking	27	16.53	6.42
		Number and Operations—Fractions	12	5.65	2.96
		Measurement and Data	12	7.93	2.78
4	215394	Operations and Algebraic Thinking	11	5.22	2.68
		Number and Operations in Base Ten	17	9.40	3.77
		Number and Operations—Fractions	18	10.90	4.45
5	202579	Number and Operations in Base Ten	17	10.54	3.93
		Number and Operations—Fractions	26	11.46	5.78
		Measurement and Data	9	5.27	2.40
6	213147	Ratios and Proportional Relationships	19	12.25	4.47
		The Number System	11	6.66	2.52
		Expressions and Equations	28	14.99	6.43
7	207251	Ratios and Proportional Relationships	20	10.21	4.98
		The Number System	14	6.88	3.68
		Expressions and Equations	22	8.80	4.66
8	214932	Expressions and Equations	30	14.67	6.88
		Functions	11	5.68	2.50
		Geometry	14	8.50	3.24

Mathematics Grade 3

Scale score statistics and N-counts of demographic groups for Grade 3 are presented in Table 40A. The population scale score mean was 299.78, with a standard deviation of 35.13. Students of both genders performed similarly, with a mean difference of 0.3 of a scale score point. Asian, Multiracial, Native Hawaiian or Other Pacific Islander, and White students had scale score means that exceeded the State mean scale score on the test, as did students from Low Needs and Average Needs districts, and the Charter Schools. The lowest performing NRC subgroup was the Big 4 Cities, with a mean of 272.24, and the lowest performing ethnic subgroup was Black (mean scale score of 286.08). At the 50th percentile, the scale scores on translated forms range from 269 (Haitian-Creole) to 327 (Korean), a difference that exceeds a standard deviation. The subgroup that used the Haitian-Creole translation had a scale score mean of 35 scale score points below the population mean, which was the lowest performing group analyzed. However, it should be noted that the sample size of this subgroup was rather small. At the 50th percentile, the following groups exceeded that of the population (300): Asian (321),

Multiracial (302), Native Hawaiian or Other Pacific Islander (308), White (306), Average Needs (302), Low Needs (314), Charter Schools (306), and ELL students who used the Chinese (310) or Korean (327) translation.

Table 40A. Mathematics Scale Score Distribution by Subgroup, Grade 3

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	204,504	299.78	35.13	256	279	300	321	345
Gender	Female	99,775	299.94	33.76	259	279	300	321	340
	Male	104,729	299.62	36.39	253	277	300	321	345
Ethnicity	Asian	17,939	321.76	35.15	279	300	321	345	365
	Black	36,064	286.08	33.39	245	265	287	308	327
	Hispanic	49,162	289.49	32.49	249	270	291	310	330
	American Indian	1,102	291.56	34.66	249	270	292	314	333
	Multiracial	2,794	301.18	36.21	256	279	302	324	350
	Native Hawaiian or Other Pacific Islander	445	305.88	34.85	265	283	308	327	350
	White	96,998	306.05	33.38	265	287	306	327	345
NRC	New York City	72,758	299.21	36.12	256	277	298	321	345
	Big 4 Cities	7,980	272.24	34.72	228	249	273	296	314
	High Needs Urban/Suburban	15,366	285.64	32.76	245	265	287	306	324
	High Needs Rural	11,031	290.48	31.89	249	272	291	310	330
	Average Needs	56,089	302.33	32.76	262	283	302	321	340
	Low Needs	27,281	315.24	32.07	277	296	314	336	357
	Charter	7,452	306.48	30.44	270	287	306	324	345
	Private School	6,538	294.59	33.89	253	274	296	317	333
SWD	All Codes	28,832	272.31	35.24	228	249	272	294	314
SUA	All Codes	33,593	277.23	34.88	234	256	279	300	319

Table 40A. Mathematics Scale Score Distribution by Subgroup, Grade 3 (cont.)

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
ELL	ELL=Y	19,516	278.50	33.13	234	259	279	300	319
SWD/SUA	SUA=504 plan codes	17,938	268.05	33.77	219	245	270	291	310
ELL/SUA	SUA & ELL codes	10,145	276.48	32.96	234	256	279	298	317
ELL Test Language	English	15,884	278.97	32.67	240	259	281	300	319
	Chinese	575	309.30	32.27	270	291	310	330	345
	Haitian-Creole	44	264.93	33.68	219	249	269	286	306
	Korean	45	327.33	34.12	300	314	327	340	378
	Russian	73	286.77	38.60	240	265	289	312	330
	Spanish	2,895	269.03	30.76	228	249	272	291	306
	All Translations	3,632	276.43	35.02	234	253	277	298	319

Mathematics Grade 4

Scale score statistics and N-counts of demographic groups for Grade 4 are presented in Table 40B. The population scale score mean was 299.84, with a standard deviation of 35.15. Students of both genders performed the same, with a mean difference of 0.16 of a scale score point. Asian, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the population mean scale score on the test, as did students from Low Needs and Average Needs districts, and Charter Schools. The lowest performing NRC subgroup was the Big 4 Cities, with a mean of 272.74, about three-fourths of a standard deviation below the population mean. SWD, SUA, and ELL subgroups scored consistently below the State-wide percentile scale score rankings. The Haitian-Creole translation subgroup had means one standard deviation below the population, and was the lowest performing group analyzed. At the 50th percentile, the following groups exceeded that of the population (300): Asian (324), Native Hawaiian or Other Pacific Islander (305), White (308), Average Needs (304), Low Needs (318), Charter Schools (304), and ELL students who used the Chinese (315) or Korean (344) translation.

Table 40B. Mathematics Scale Score Distribution by Subgroup, Grade 4

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	215,394	299.84	35.15	255	278	300	324	344
Gender	Female	105,936	299.76	33.70	259	278	300	322	341
	Male	109,458	299.92	36.49	255	276	300	324	347
Ethnicity	Asian	17,924	322.89	34.98	278	302	324	347	368
	Black	37,219	285.28	33.07	242	265	285	308	326
	Hispanic	50,239	289.57	32.70	247	271	289	312	331
	American Indian	1,140	294.13	34.11	251	273	294	316	335
	Multiracial	2,537	299.78	36.92	255	276	300	324	347
	Native Hawaiian or Other Pacific Islander	476	304.02	32.95	265	283	305	324	347
	White	105,859	305.98	33.34	265	285	308	328	347

Table 40B. Mathematics Scale Score Distribution by Subgroup, Grade 4 (cont.)

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
NRC	New York City	70,981	299.70	36.69	255	276	300	324	347
	Big 4 Cities	8,030	272.74	34.10	231	251	273	296	316
	High Needs Urban/Suburban	15,566	283.97	32.93	242	262	285	306	324
	High Needs Rural	11,027	290.72	32.46	251	271	292	312	331
	Average Needs	57,933	302.56	32.44	262	283	304	324	344
	Low Needs	28,173	317.09	31.61	278	298	318	338	358
	Charter	5,806	304.62	29.75	268	285	304	324	341
	Private School	17,866	294.52	31.92	255	276	296	316	333
SWD	All Codes	31,161	270.17	34.62	224	247	271	292	314
SUA	All Codes	30,895	272.67	34.06	231	251	273	296	316
ELL	ELL=Y	16,834	274.38	33.90	231	255	276	296	316
SWD/SUA	SUA=504 plan codes	18,082	265.57	33.33	224	242	265	287	308
ELL/SUA	SUA=ELL codes	6,608	268.46	32.92	224	247	271	289	308
ELL Test Language	English	13,476	274.65	33.08	231	255	276	296	316
	Chinese	538	314.85	31.25	278	296	315	333	358
	Haitian-Creole	52	255.12	33.25	208	240	259	280	296
	Korean	31	330.45	38.32	273	304	344	358	368
	Russian	76	289.64	36.82	237	265	293	318	331
	Spanish	2,661	264.11	31.09	224	247	265	285	302
	All Translations	3,358	273.29	36.98	231	251	273	296	320

Mathematics Grade 5

Grade 5 demographic group N-counts and scale score statistics are presented in Table 40C. The population scale score mean was 299.77, with a standard deviation of 35.25. Students of both genders performed very similarly, with a mean difference of just over one scale score point. Asian, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the population mean scale score on the test, as did students from Low Needs and Average Needs districts, and Charter Schools. The lowest performing NRC subgroup was the Big 4 Cities with a mean of 272.36, which were about four-fifths of a standard deviation below the population mean. SWD, SUA, and ELL subgroups scored consistently at or below the State-wide percentile scale score rankings. The Haitian-Creole translation subgroup's scale score mean (258.81) was approximately 41 points below the population mean, and was the lowest performing group analyzed. The Korean translation subgroup was the highest performing group analyzed, with a scale score mean of 335.39, about one standard deviation above the population mean. At the 50th percentile, the following groups exceeded that of the population (300): Female (302), Asian (327), Native Hawaiian or Other Pacific Islander (306), White (308), Average Needs (304), Low Needs (317), and ELL students who used the Chinese (319) or Korean (333) translation.

Table 40C. Mathematics Scale Score Distribution by Subgroup, Grade 5

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	202,579	299.77	35.25	255	278	300	323	343
Gender	Female	99,557	300.38	33.51	259	281	302	321	340
	Male	103,022	299.17	36.85	255	278	300	323	343
Ethnicity	Asian	17,003	325.18	35.33	283	304	327	349	372
	Black	36,425	285.78	33.10	245	266	288	308	325
	Hispanic	47,043	289.20	32.83	245	269	292	311	329
	American Indian	1,122	291.45	35.25	245	269	292	313	335
	Multiracial	2,130	298.35	36.94	250	278	300	323	343
	Native Hawaiian or Other Pacific Islander	345	306.84	32.75	269	288	306	327	346
	White	98,511	305.69	33.16	266	288	308	327	346
NRC	New York City	69,104	299.41	36.63	255	278	300	323	346
	Big 4 Cities	7,816	272.36	35.52	222	250	275	296	315
	High Needs Urban/Suburban	14,992	284.43	32.88	239	266	288	306	323
	High Needs Rural	11,042	290.89	32.16	250	272	294	311	329
	Average Needs	57,641	302.20	31.97	263	283	304	323	340
	Low Needs	28,355	316.87	31.26	281	298	317	335	357
	Charter	7,079	300.01	30.17	263	283	300	319	338
	Private School	6,531	290.62	37.18	245	269	294	315	333
SWD	All Codes	31,675	269.27	35.05	222	245	272	292	311
SUA	All Codes	32,552	274.02	35.31	222	255	275	296	317
ELL	ELL=Y	16,410	274.16	34.35	232	255	275	296	315
SWD/SUA	SUA=504 plan codes	19,571	266.07	34.32	214	245	269	290	308
ELL/SUA	SUA=ELL codes	7,523	273.23	34.82	222	255	275	294	315
ELL Test Language	English	12,890	274.60	33.37	232	255	278	296	313
	Chinese	494	318.13	29.07	283	296	319	338	357
	Haitian-Creole	75	258.81	27.82	214	245	263	281	290
	Korean	31	335.39	30.82	302	311	333	366	372
	Russian	84	291.44	38.74	255	269	288	319	333
	Spanish	2,836	263.73	32.30	214	245	268	285	302
	All Translations	3,520	272.55	37.69	222	250	275	296	319

Mathematics Grade 6

Grade 6 scale score statistics and N-counts of demographic groups are presented in Table 40D. The population scale score mean was 299.56, with a standard deviation of 35.11. Students of both genders performed very similarly, with a mean difference of less than two scale score points. Asian, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the population mean scale score, as did students from Low Needs and Average Needs districts, and Charter Schools. The lowest performing NRC subgroup was the Big 4 Cities, with a

mean of 274.36. SWD, SUA, and ELL subgroups scored consistently below the Statewide percentile scale score rankings. The Haitian-Creole translation subgroup, whose scale score mean (251.91)—over 47 points below the population mean—was the lowest performing group analyzed. At the 50th percentile, the following groups exceeded that of the population (301): Asian (325), Native Hawaiian or Other Pacific Islander (308), White (308), Average Needs (305), Low Needs (318), and ELL students who used the Chinese (316) or Korean (335) translation.

Table 40D. Mathematics Scale Score Distribution by Subgroup, Grade 6

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	213,147	299.56	35.11	256	279	301	322	342
Gender	Female	104,829	300.46	34.15	259	279	301	322	342
	Male	108,318	298.68	35.99	253	277	299	322	342
Ethnicity	Asian	17,218	324.61	36.41	279	303	325	348	372
	Black	39,107	284.73	33.20	241	265	286	306	325
	Hispanic	49,176	288.78	32.64	246	270	290	310	327
	American Indian	1,092	290.22	33.03	253	270	290	310	329
	Multiracial	2,027	299.13	35.95	256	277	299	322	345
	Native Hawaiian or Other Pacific Islander	424	307.76	36.05	268	286	308	333	352
	White	104,103	306.15	32.55	265	288	308	327	345
NRC	New York City	68,465	297.88	37.48	253	275	298	322	345
	Big 4 Cities	7,933	274.36	35.76	231	253	275	298	318
	High Needs Urban/Suburban	14,762	285.79	32.25	246	268	288	306	323
	High Needs Rural	11,435	292.30	30.87	253	275	294	313	329
	Average Needs	59,186	302.36	31.96	262	284	305	323	340
	Low Needs	28,777	317.45	31.66	279	299	318	337	355
	Charter	6,701	300.87	30.10	265	282	301	320	337
	Private School	15,863	294.10	31.66	256	277	296	315	331
SWD	All Codes	31,841	266.44	34.01	224	246	268	290	308
SUA	All Codes	32,323	270.11	34.41	224	249	273	294	311
ELL	ELL=Y	13,063	268.65	36.14	224	246	270	292	313
SWD/SUA	SUA=504 plan codes	20,628	263.68	33.44	216	241	265	286	305
ELL/SUA	SUA=ELL codes	5,607	263.10	35.25	216	241	265	286	305
ELL Test Language	English	9,301	268.17	34.69	224	246	270	290	310
	Chinese	642	315.33	32.50	275	294	316	337	355
	Haitian-Creole	143	251.91	34.16	200	231	256	273	294
	Korean	47	327.32	37.15	277	303	335	355	365
	Russian	95	288.44	27.94	253	275	294	303	318
	Spanish	2,835	258.86	32.39	216	237	262	282	299
	All Translations	3,762	269.83	39.47	216	246	270	296	320

Mathematics Grade 7

N-counts and scale score statistics of demographic groups for Grade 7 are presented in Table 40E. The population scale score mean was 299.72, with a standard deviation of 35.28. Students of both genders performed very similarly, with a mean difference of less than two scale score points. Asian, Multiracial, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the population mean scale score. The lowest performing NRC subgroup, Big 4 Cities, had a scale score mean of 271.56, while the Low Needs subgroup's scale score mean was 317.84. SWD, SUA, and ELL subgroups scored consistently below the Statewide percentile scale score rankings, and had means nearly one standard deviation below the population mean. The Haitian-Creole translation was the lowest performing group analyzed, with a scale score mean of 252.14. At the 50th percentile, the following groups exceeded that of the population (302): Females (303), Asian (327), Native Hawaiian or Other Pacific Islander (311), White (311), Average Needs (306), Low Needs (320), and ELL students who used the Chinese (317) or Korean (325) translation.

Table 40E. Mathematics Scale Score Distribution by Subgroup, Grade 7

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	207,251	299.72	35.28	253	280	302	324	342
Gender	Female	101,198	300.42	34.23	259	280	303	324	342
	Male	106,053	299.06	36.24	253	278	302	324	344
Ethnicity	Asian	17,553	323.74	34.98	280	305	327	348	364
	Black	37,786	281.94	33.92	234	264	285	305	324
	Hispanic	46,251	287.51	33.06	245	268	289	310	327
	American Indian	1,027	289.36	33.80	245	271	293	313	332
	Multiracial	1,836	300.78	35.24	259	279	302	324	344
	Native Hawaiian or Other Pacific Islander	398	309.29	41.03	259	283	311	336	364
	White	102,400	307.73	31.65	268	291	311	329	344
NRC	New York City	70,020	296.37	37.25	245	275	297	322	344
	Big 4 Cities	7,716	271.56	34.96	219	253	275	295	314
	High Needs Urban/Suburban	14,955	283.67	33.32	234	264	287	306	324
	High Needs Rural	11,648	292.70	31.17	253	275	297	314	329
	Average Needs	59,652	303.79	31.34	264	287	306	325	340
	Low Needs	31,525	317.84	29.98	280	302	320	338	352
	Charter	5,355	298.56	31.03	259	280	300	319	336
	Private School	6,368	294.56	36.33	245	275	298	319	336
SWD	All Codes	31,379	266.73	33.67	219	245	268	289	306
SUA	All Codes	26,091	271.53	34.67	226	253	275	295	314
ELL	ELL=Y	12,307	266.90	35.64	219	245	268	289	311
SWD/SUA	SUA=504 plan codes	17,390	265.65	33.03	219	245	268	289	305
ELL/SUA	SUA=ELL codes	3,805	263.16	36.21	219	234	264	285	308

Table 40E. Mathematics Scale Score Distribution by Subgroup, Grade 7 (cont.)

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
ELL Test Language	English	8,372	264.97	34.22	219	245	268	287	306
	Chinese	703	316.08	28.47	278	300	317	336	350
	Haitian-Creole	157	252.14	32.04	203	226	259	280	289
	Korean	36	327.03	31.25	295	304	325	345	367
	Russian	129	283.57	36.72	234	264	285	305	330
	Spanish	2,910	259.91	31.14	219	234	264	280	297
	All Translations	3,935	271.03	38.17	219	245	271	295	322

Mathematics Grade 8

Scale score statistics and N-counts of demographic groups for Grade 8 are presented in Table 40F. The population scale score mean was 299.43, with a standard deviation of 35.23. Students of both genders performed similarly, with a difference of approximately three points. Asian, Native Hawaiian or Other Pacific Islander, and White students' scale score means exceeded the population mean, as did students from Low Needs, Average Needs, and Charter Schools. Students from the Big 4 Cities achieved a scale score mean of about one standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, slightly less than one standard deviation below the mean scale score for the population. The SWD and SWD/SUA subgroups had a scale score mean approximately 34 scale score points below the population mean, and this subgroup was the lowest performing group analyzed. At the 50th percentile, the following groups exceeded that of the population (302): Female (303), Asian (328), Native Hawaiian or Other Pacific Islander (307), White (308), Average Needs districts (307), Low Needs districts (319), Charter Schools (307), and ELL students who used the Chinese (320) or Korean (326) translation.

Table 40F. Mathematics Scale Score Distribution by Subgroup, Grade 8

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
State	All Students	214,932	299.43	35.23	255	280	302	324	341
Gender	Female	105,389	300.97	33.70	259	282	303	324	341
	Male	109,543	297.95	36.58	250	275	300	322	341
Ethnicity	Asian	17,282	324.54	35.48	280	303	328	349	366
	Black	39,633	282.51	33.83	239	263	284	305	324
	Hispanic	47,763	288.47	33.55	245	269	291	312	329
	American Indian	1,038	290.52	32.33	250	272	292	312	329
	Multiracial	1,817	298.33	36.21	255	277	302	322	341
	Native Hawaiian or Other Pacific Islander	410	305.65	37.08	263	284	307	329	351
	White	106,989	306.61	31.81	266	289	308	328	344

Table 40F. Mathematics Scale Score Distribution by Subgroup, Grade 8 (cont.)

Demographic Category (Subgroup)		N-count	SS Mean	SS SD	10th %tile	25th %tile	50th %tile	75th %tile	90th %tile
NRC	New York City	69,217	297.02	37.44	250	272	298	322	344
	Big 4 Cities	7,492	266.38	35.78	215	245	269	291	310
	High Needs Urban/Suburban	14,651	284.63	32.99	245	266	287	307	324
	High Needs Rural	11,707	291.14	31.36	250	275	294	312	328
	Average Needs	59,339	303.46	31.61	263	287	307	324	339
	Low Needs	31,249	317.03	29.58	282	302	319	335	351
	Charter	4,405	304.99	29.63	269	287	307	326	341
	Private School	16,836	294.49	32.93	255	277	296	315	333
SWD	All Codes	30,749	265.65	34.47	215	245	269	289	307
SUA	All Codes	25,996	271.68	35.74	223	250	275	296	315
ELL	ELL=Y	13,148	271.24	36.15	223	250	272	294	317
SWD/SUA	SUA=504 plan codes	17,166	265.40	34.29	215	245	269	289	307
ELL/SUA	SUA=ELL codes	3,882	266.58	38.08	215	245	266	291	317
ELL Test Language	English	8,879	269.65	34.81	223	250	272	292	312
	Chinese	776	316.73	32.20	272	298	320	339	354
	Haitian-Creole	165	267.85	33.56	223	250	269	294	310
	Korean	33	320.12	32.61	275	296	326	341	354
	Russian	109	291.53	36.99	245	266	296	314	339
	Spanish	3,186	263.57	32.24	215	245	266	287	303
	All Translations	4,269	274.55	38.58	223	250	275	300	326

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. Tables 41 and 42 show the ELA and mathematics cut scores, respectively, used for classification of students into the four performance-level categories in 2013. 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 Appendix P: "Standard-setting Technical Report" for details). While vertical articulation helps to build consistent meaning to the performance levels, the very nature of grade-specific content, differing performance expectations, and panel set cut scores result in cut score differences across grades.

Table 41. ELA Performance Level Cut Scores

Performance Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level II	291	287	289	283	287	284
NYS Level III	320	320	320	320	318	316
NYS Level IV	358	343	346	338	347	343

Table 42. Mathematics Performance Level Cut Scores

Performance Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level II	285	283	294	284	293	287
NYS Level III	314	314	319	318	322	322
NYS Level IV	340	341	346	340	348	349

ELA Test Performance Level Distributions

Table 43 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 ethnic groups. Consistent with the pattern shown in scale score distribution across the group, 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 43. ELA Test Performance Level Distributions

Grade	N-count	Percentage of NYS Student Population in Performance Level				
		Level I	Level II	Level III	Level IV	Levels III & IV
3	203,041	35.61	33.18	27.47	3.75	31.22
4	213,673	29.90	39.35	21.22	9.52	30.74
5	201,117	34.15	35.54	21.78	8.53	30.31
6	211,594	28.28	41.78	16.24	13.70	29.94
7	205,800	31.97	36.57	23.42	8.04	31.46
8	214,090	28.88	36.83	23.81	10.47	34.29

ELA Grade 3

Performance level distributions and N-counts of demographic groups for ELA Grade 3 are presented in Table 44A. Statewide, 31% of third-graders were Level III and Level IV. Among Male students, 40% were Level I, as compared to 31% of Female students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroups. About 51% of Low Needs district students and about 50% of Asian students were classified in Levels III and IV; whereas the Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 81%–89% of students who were in Level I or Level II. At least two-thirds of students with ELL, SWD, or SUA status were in Level I, and less than 1% was in Level IV. The following groups had had a higher percentage of students in Levels III and IV than the general population: Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, White, Average Needs districts, Low Needs districts, and Private Schools.

Table 44A. ELA Performance Level Distribution by Subgroup, Grade 3

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	203,041	35.61	33.18	27.47	3.75	31.22
Gender	Female	99,173	30.79	33.66	30.72	4.83	35.54
	Male	103,868	40.21	32.71	24.37	2.72	27.09
Ethnicity	Asian	17,411	20.04	30.37	40.66	8.93	49.60
	Black	35,947	49.71	31.96	16.99	1.34	18.33
	Hispanic	48,391	46.86	34.11	17.85	1.19	19.03
	American Indian	1,099	44.31	33.48	20.38	1.82	22.20
	Multiracial	2,793	33.73	30.90	29.97	5.41	35.37
	Native Hawaiian or Other Pacific Islander	439	29.38	32.35	33.26	5.01	38.27
	White	96,961	27.54	33.73	33.77	4.96	38.72
NRC	New York City	71,676	38.53	33.30	24.59	3.58	28.18
	Big 4 Cities	7,921	64.92	24.20	10.38	0.50	10.88
	High Needs Urban/Suburban	15,208	49.99	31.99	16.87	1.16	18.02
	High Needs Rural	11,007	45.02	33.69	19.65	1.64	21.30
	Average Needs	56,084	30.97	34.88	30.30	3.85	34.15
	Low Needs	27,192	17.60	31.66	43.21	7.53	50.74
	Charter	7,368	33.05	37.08	27.93	1.94	29.87
	Private School	6,585	36.36	31.92	27.29	4.43	31.72
SWD	All Codes	28,051	73.40	19.75	6.41	0.44	6.85
SUA	All Codes	27,870	67.44	23.76	8.39	0.42	8.80
SWD/SUA	SUA=504 Plan Codes	15,750	78.71	16.63	4.46	0.20	4.66
ELL/SUA	SUA=ELL Codes	7,028	70.11	24.99	4.84	0.07	4.91
ELL	ELL status = Y	17,494	67.01	26.66	6.18	0.15	6.33

ELA Grade 4

Performance level distributions and N-counts of demographic groups for ELA Grade 4 are presented in Table 44B. As was seen in Grade 3, approximately 31% of fourth-grade students statewide were in Levels III and IV. The Low Needs subgroup had the highest percentage of students in Levels III and IV (50%), and the ELL/SUA subgroup had the lowest (2%). Fewer than 20% of Black and Hispanic students were classified in Levels III and IV. Approximately twice as many Big 4 Cities students were in Level I than the state population. Over two-thirds of the students with ELL, SWD, or SUA status were in Level I, and no more than 1% were in Level IV. The following groups had a higher percentage of students in Levels III and IV than the general population: Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, White, Average Needs districts, Low Needs districts, and Private Schools.

Table 44B. ELA Performance Level Distribution by Subgroup, Grade 4

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	213,673	29.90	39.35	21.22	9.52	30.74
Gender	Female	105,061	25.40	39.07	23.50	12.02	35.52
	Male	108,612	34.26	39.63	19.02	7.10	26.11
Ethnicity	Asian	17,408	15.69	35.28	29.06	19.98	49.03
	Black	37,085	42.35	39.69	13.84	4.12	17.96
	Hispanic	49,439	40.90	40.71	14.21	4.17	18.38
	American Indian	1,135	40.26	37.80	16.39	5.55	21.94
	Multiracial	2,544	29.36	38.60	21.78	10.26	32.04
	Native Hawaiian or Other Pacific Islander	477	26.62	40.67	25.58	7.13	32.70
	White	105,585	22.64	39.30	25.82	12.23	38.05
NRC	New York City	69,892	33.35	39.36	18.49	8.81	27.30
	Big 4 Cities	7,878	59.46	30.21	8.00	2.34	10.33
	High Needs Urban/Suburban	15,432	45.41	38.20	12.87	3.53	16.39
	High Needs Rural	11,026	40.14	39.29	16.36	4.21	20.57
	Average Needs	57,920	25.61	41.38	23.41	9.60	33.01
	Low Needs	28,128	13.13	36.77	32.48	17.62	50.10
	Charter	5,732	26.45	45.45	21.56	6.54	28.11
	Private School	17,660	25.05	40.00	23.06	11.89	34.95
SWD	All Codes	30,278	70.77	23.79	4.44	1.00	5.44
SUA	All Codes	26,444	65.01	28.16	5.84	0.99	6.83
SWD/SUA	SUA=504 Plan Codes	15,755	76.12	20.37	3.11	0.40	3.51
ELL/SUA	SUA=ELL Codes	4,687	74.57	23.21	1.98	0.23	2.22
ELL	ELL status = Y	14,634	70.18	26.47	2.98	0.37	3.35

ELA Grade 5

Performance level distributions and N-counts of demographic groups for ELA Grade 5 are presented in Table 44C. As was seen in Grades 3 and 4, about 30% of the Grade 5 students were in Levels III and IV. Asian students had the highest percentage of students in Levels III and IV (52%). Fewer than 20% of Black and Hispanic students were classified in Levels III and IV, approximately 20% less than other ethnic subgroups. Almost two times as many Big 4 Cities students were in Level I than the population. Over two-thirds of the students with ELL, SWD, or SUA status were in Level I, and at most 1% in Level IV. The following groups had a higher percentage of students in Levels III and IV than the general population: Female, Asian, Native Hawaiian or Other Pacific Islander, White, Average Needs districts, Low Needs districts, and Private Schools.

Table 44C. ELA Performance Level Distribution by Subgroup, Grade 5

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	201,117	34.15	35.54	21.78	8.53	30.31
Gender	Female	98,798	29.80	36.15	23.80	10.25	34.05
	Male	102,319	38.36	34.94	19.83	6.87	26.70
Ethnicity	Asian	16,571	17.14	30.41	31.76	20.68	52.44
	Black	36,258	47.54	35.48	13.66	3.32	16.98
	Hispanic	46,141	44.46	36.96	15.09	3.49	18.58
	American Indian	1,111	43.83	34.92	17.55	3.69	21.24
	Multiracial	2,114	35.24	34.48	20.34	9.93	30.27
	Native Hawaiian or Other Pacific Islander	336	27.68	33.93	27.68	10.71	38.39
	White	98,586	27.15	35.79	26.28	10.78	37.06
NRC	New York City	68,029	35.65	35.62	20.13	8.60	28.73
	Big 4 Cities	7,697	64.34	26.15	7.47	2.04	9.51
	High Needs Urban/Suburban	14,849	49.55	34.69	12.79	2.98	15.77
	High Needs Rural	11,022	44.54	35.25	16.10	4.11	20.21
	Average Needs	57,619	30.40	37.44	23.76	8.40	32.17
	Low Needs	28,319	16.16	34.28	33.52	16.05	49.57
	Charter	6,942	38.09	41.37	16.74	3.80	20.54
	Private School	6,634	37.37	30.77	22.78	9.09	31.87
SWD	All Codes	30,841	74.45	20.39	4.35	0.81	5.16
SUA	All Codes	28,230	68.88	24.25	5.82	1.04	6.86
SWD/SUA	SUA=504 Plan Codes	17,495	78.85	17.69	3.07	0.40	3.47
ELL/SUA	SUA=ELL Codes	5,299	75.17	21.55	2.93	0.36	3.28
ELL	ELL status = Y	14,247	73.12	23.25	3.28	0.34	3.63

ELA Grade 6

Performance level distributions and N-counts of demographic groups for ELA Grade 6 are presented in Table 44D. Statewide, about 30% of Grade 6 students were classified in Levels III and IV. As was seen in other grades, the Low Needs subgroup and students in the Asian subgroup had the most students classified in these two proficiency levels (52% and 48%, respectively), and the ELL, SWD, and SUA subgroups had the fewest. Fewer than 20% of Black, Hispanic, and American Indian students were classified in Levels III and IV. Students from Low Needs districts outperformed students in all other subgroups, across demographic categories as in the previous grades. The majority of students with ELL, SWD, and/or SUA status were in Level I, and fewer than 6% were in Level IV. The following groups had a higher percentage of students in Levels III and IV than the general population: Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, White, Average Needs districts, Low Needs districts, and Private Schools.

Table 44D. ELA Performance Level Distribution by Subgroup, Grade 6

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	211,594	28.28	41.78	16.24	13.70	29.94
Gender	Female	103,943	23.67	42.07	17.74	16.51	34.25
	Male	107,651	32.72	41.50	14.79	10.99	25.78
Ethnicity	Asian	16,720	15.84	36.51	21.68	25.97	47.66
	Black	39,005	43.12	42.56	9.50	4.82	14.32
	Hispanic	48,240	39.92	44.24	10.58	5.27	15.85
	American Indian	1,086	38.40	43.28	10.87	7.46	18.32
	Multiracial	2,023	27.68	39.84	16.07	16.41	32.48
	Native Hawaiian or Other Pacific Islander	414	21.74	38.16	18.12	21.98	40.10
	White	104,106	19.25	41.24	20.56	18.95	39.51
NRC	New York City	67,443	35.06	41.54	12.82	10.58	23.40
	Big 4 Cities	7,784	56.41	33.05	6.69	3.84	10.53
	High Needs Urban/Suburban	14,797	40.58	42.84	10.68	5.91	16.58
	High Needs Rural	11,320	32.22	43.90	15.05	8.83	23.89
	Average Needs	59,256	22.05	43.36	18.80	15.80	34.60
	Low Needs	28,795	11.26	37.16	24.86	26.72	51.58
	Charter	6,498	34.29	50.15	10.96	4.60	15.56
	Private School	15,688	22.90	43.77	18.45	14.88	33.33
SWD	All Codes	31,025	70.80	25.42	2.76	1.02	3.78
SUA	All Codes	28,080	63.93	30.41	4.02	1.63	5.66
SWD/SUA	SUA=504 Plan Codes	18,308	72.47	24.32	2.46	0.76	3.22
ELL/SUA	SUA=ELL Codes	3,588	82.02	16.64	0.92	0.42	1.34
ELL	ELL status = Y	10,694	80.53	18.07	1.05	0.36	1.40

ELA Grade 7

Performance level distributions and N-counts of demographic groups for ELA Grade 7 are presented in Table 44E. In Grade 7, approximately 31% of the students were in Levels III and IV. The Low Needs subgroup and students in the Asian subgroup had the most students classified in these two proficiency levels (51% each), and the ELL, SWD, and SUA subgroups had the fewest. About 10% more Female than Male students were classified in these two proficiency levels. About 89% of Big 4 Cities students were in Levels I and II. The ELL, SWD, and SUA subgroups tended to have more students in the lower performance levels than in the general population, with around 68–86% of those students in Level I. The following groups had a higher percentage of students in Levels III and IV than the general population: Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, White, Average Needs districts, Low Needs districts, and Private Schools.

Table 44E. ELA Performance Level Distribution by Subgroup, Grade 7

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	205,800	31.97	36.57	23.42	8.04	31.46
Gender	Female	100,426	26.35	37.41	26.44	9.80	36.24
	Male	105,374	37.33	35.77	20.54	6.36	26.90
Ethnicity	Asian	17,099	17.15	31.41	33.73	17.70	51.44
	Black	37,655	47.54	36.88	13.30	2.28	15.58
	Hispanic	45,309	44.17	38.80	14.40	2.62	17.03
	American Indian	1,019	42.69	37.29	15.80	4.22	20.02
	Multiracial	1,845	31.54	33.66	25.15	9.65	34.80
	Native Hawaiian or Other Pacific Islander	383	27.94	27.42	30.55	14.10	44.65
	White	102,490	23.25	36.41	29.42	10.92	40.34
NRC	New York City	68,732	37.29	37.13	19.28	6.30	25.58
	Big 4 Cities	7,590	62.06	27.18	9.20	1.57	10.76
	High Needs Urban/Suburban	14,810	47.19	35.92	14.08	2.81	16.89
	High Needs Rural	11,599	39.26	37.01	18.88	4.85	23.73
	Average Needs	59,859	25.95	37.71	27.15	9.18	36.33
	Low Needs	31,615	13.94	34.99	35.48	15.59	51.07
	Charter	5,394	34.26	44.14	19.08	2.52	21.60
	Private School	6,188	34.20	33.06	23.88	8.86	32.74
SWD	All Codes	30,831	73.07	22.52	4.00	0.42	4.41
SUA	All Codes	26,042	67.91	25.17	6.08	0.85	6.93
SWD/SUA	SUA=504 Plan Codes	17,467	75.41	20.89	3.39	0.30	3.70
ELL/SUA	SUA=ELL Codes	3,410	85.51	12.55	1.82	0.12	1.94
ELL	ELL status = Y	10,132	84.20	14.33	1.38	0.09	1.47

ELA Grade 8

Performance level distributions and N-counts of demographic groups for ELA Grade 8 are presented in Table 44F. In Grade 8, 34% of the students were in Levels III and IV. The Low Needs subgroup had the most students classified in Levels III and IV (56%), while the Big 4 Cities only had 11% of the students classified in these two proficiency levels. About 8% more Female students than Male students were in Levels III or IV. Over 80% of Black and Hispanic students were in Levels I and II. The ELL, SWD, and SUA subgroups were well below the performance of the general population, with over 64% of those students in Level I. The following groups had a higher percentage of students in Levels III and IV than the general population: Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, White, Average Needs districts, Low Needs districts, and Private Schools.

Table 44F. ELA Performance Level Distribution by Subgroup, Grade 8

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	214,090	28.88	36.83	23.81	10.47	34.29
Gender	Female	104,876	24.63	37.04	25.91	12.43	38.34
	Male	109,214	32.97	36.64	21.80	8.59	30.40
Ethnicity	Asian	16,850	17.03	30.72	31.93	20.31	52.24
	Black	39,575	45.10	38.12	13.77	3.01	16.78
	Hispanic	46,975	41.57	39.28	15.31	3.84	19.15
	American Indian	1,044	40.33	35.15	19.73	4.79	24.52
	Multiracial	1,839	27.30	36.27	24.09	12.34	36.43
	Native Hawaiian or Other Pacific Islander	392	26.79	36.48	25.26	11.48	36.73
	White	107,415	19.14	36.28	29.99	14.60	44.59
NRC	New York City	68,108	37.53	36.93	18.48	7.05	25.54
	Big 4 Cities	7,373	61.26	27.67	8.46	2.60	11.07
	High Needs Urban/Suburban	14,551	41.75	38.28	15.77	4.20	19.97
	High Needs Rural	11,632	33.10	39.26	20.78	6.86	27.64
	Average Needs	59,860	22.28	37.66	27.70	12.36	40.06
	Low Needs	31,455	11.04	33.07	35.24	20.65	55.89
	Charter	4,424	31.08	46.75	18.08	4.09	22.17
	Private School	16,649	21.67	39.08	27.59	11.65	39.25
SWD	All Codes	30,367	70.54	24.32	4.40	0.75	5.15
SUA	All Codes	24,449	63.54	28.24	6.77	1.45	8.23
SWD/SUA	SUA=504 Plan Codes	16,591	70.91	24.26	4.16	0.68	4.83
ELL/SUA	SUA=ELL Codes	2,840	84.44	14.51	0.95	0.11	1.06
ELL	ELL status = Y	10,964	82.35	16.28	1.23	0.14	1.37

Mathematics Test Performance Level Distributions

Table 45 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 45. Mathematics Test Performance Level Distributions

Grade	N-count	Percentage of NYS Student Population in Performance Level				
		Level I	Level II	Level III	Level IV	Level III & IV
3	204,504	30.42	35.54	21.82	12.22	34.04
4	215,394	29.17	35.12	23.23	12.48	35.70
5	202,579	40.02	30.32	20.88	8.77	29.65
6	213,147	29.03	40.98	17.93	12.06	29.99
7	207,251	37.97	34.42	20.27	7.34	27.61
8	214,932	31.37	41.73	19.90	7.00	26.91

Mathematics Grade 3

Performance level summaries and N-counts of demographic groups for Grade 3 are presented in Table 46A. Statewide, 34% of the third-grade population was placed in Levels III and IV. Unlike the pattern shown in ELA, the percentages of Female and Male students in each proficiency level did not differ much. However, student achievement varied widely by ethnicity, NRC, and test translation subgroups. Students in the Black, Hispanic, and American Indian subgroups had a lower percentage of students in Levels III and IV than the rest of the population, while the percentage of Asian, Native Hawaiian or Other Pacific Islander, and White students in Levels III and IV exceeded the overall population. About 52% of students from Low Needs districts were in Levels III or IV, but only 11% of Big 4 Cities students were in these two proficiency levels. The ELL, SWD, and SUA subgroups were well below the performance of the general population as well, with the majority of the students in Levels I and II. For ELL students who used translated test forms, those who used Spanish or Haitian-Creole had the lowest percentages of students classified in Levels III and IV (6% and 7%, respectively); those who used Korean translation had the highest percentage of students classified in Levels III and IV (82%). The following subgroups had a higher percentage of students meeting standards for Levels III and IV than the state population: Male, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, White, Average Needs, Low Needs, Charter school, Chinese translation, and Korean translation.

Table 46A. Mathematics Performance Level Distribution by Subgroup, Grade 3

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	204,504	30.42	35.54	21.82	12.22	34.04
Gender	Female	99,775	29.83	36.72	21.97	11.48	33.45
	Male	104,729	30.99	34.41	21.67	12.94	34.60

Table 46A. Mathematics Performance Level Distribution by Subgroup, Grade 3 (cont.)

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
Ethnicity	Asian	17,939	12.48	26.88	30.24	30.40	60.64
	Black	36,064	45.63	34.87	14.49	5.00	19.49
	Hispanic	49,162	40.83	37.45	16.14	5.59	21.72
	American Indian	1,102	39.47	35.21	17.70	7.62	25.32
	Multiracial	2,794	30.31	34.29	20.79	14.60	35.40
	Native Hawaiian or Other Pacific Islander	445	25.17	33.71	26.52	14.61	41.12
	White	96,998	22.74	36.46	25.92	14.88	40.80
NRC	New York City	72,758	32.25	34.52	20.35	12.88	33.23
	Big 4 Cities	7,980	61.99	26.89	8.55	2.57	11.12
	High Needs Urban/Suburban	15,366	46.32	35.09	13.89	4.69	18.59
	High Needs Rural	11,031	39.65	37.77	16.83	5.75	22.58
	Average Needs	56,089	26.00	38.08	23.96	11.95	35.92
	Low Needs	27,281	14.23	33.37	30.94	21.46	52.40
	Charter	7,452	21.18	39.21	26.50	13.11	39.61
	Private School	6,538	34.66	37.69	19.55	8.11	27.65
SWD	All Codes	28,832	62.34	26.62	8.14	2.90	11.04
SUA	All Codes	33,593	56.57	29.68	10.15	3.60	13.75
SWD/SUA	SUA=504 Plan Codes	17,938	67.44	24.63	6.25	1.68	7.93
ELL/SUA	SUA=ELL Codes	10,145	58.07	30.11	9.14	2.68	11.82
ELL	ELL status = Y	19,516	55.72	30.85	10.29	3.15	13.44
ELL Test Language	English	15,884	55.26	31.31	10.42	3.01	13.43
	Chinese	575	19.48	34.78	29.91	15.83	45.74
	Haitian-Creole	44	70.45	22.73	4.55	2.27	6.82
	Korean	45	6.67	11.11	53.33	28.89	82.22
	Russian	73	42.47	34.25	13.70	9.59	23.29
	Spanish	2,895	66.32	27.84	5.01	0.83	5.84
	All Translations	3,632	57.74	28.80	9.72	3.74	13.46

Mathematics Grade 4

Performance level summaries and N-counts of demographic groups for Grade 4 are presented in Table 46B. Statewide, 36% of the fourth-grade population was placed in Levels III and IV. As seen in Grade 3, there was little performance differentiation by gender. However, student achievement varied widely by ethnicity, NRC, and test translation subgroups. Students in the Black, Hispanic, and American Indian subgroups had a lower percentage of students in Levels III and IV than the rest of the population, while the percentage of Asian and White students in Levels III and IV exceeded the overall population. About 57% of students from Low Needs districts were in Levels III or IV, but only 12% of Big 4 Cities students were in Levels III and IV. The ELL, SWD, and SUA subgroups were well below the performance of the general

population as well, with the majority of the students in Levels I and II. For ELL students who used translated test forms, those who used Haitian-Creole or Spanish had the lowest percentages of students classified in Levels III and IV (4% and 5%, respectively); those who used Korean and Chinese translations had the highest percentages of students classified in Levels III and IV, with Korean students having the greatest percentage: more than 68%. The following subgroups had a higher percentage of students meeting standards for Levels III and IV than the state population: Male, Asian, Native Hawaiian or Other Pacific Islander, White, Average Needs, Low Needs, Charter, English form, Chinese translation, and Korean translation.

Table 46B. Mathematics Performance Level Distribution by Subgroup, Grade 4

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	215,394	29.17	35.12	23.23	12.48	35.70
Gender	Female	105,936	28.39	36.90	23.35	11.36	34.71
	Male	109,458	29.93	33.40	23.10	13.56	36.66
Ethnicity	Asian	17,924	11.41	24.97	31.28	32.34	63.62
	Black	37,219	45.03	35.34	14.82	4.80	19.62
	Hispanic	50,239	39.15	37.72	17.32	5.81	23.13
	American Indian	1,140	33.16	39.82	18.77	8.25	27.02
	Multiracial	2,537	30.39	34.37	21.68	13.56	35.24
	Native Hawaiian or Other Pacific Islander	476	24.16	36.34	25.00	14.50	39.50
	White	105,859	21.82	35.49	27.70	14.99	42.69
NRC	New York City	70,981	30.91	33.75	21.35	13.99	35.34
	Big 4 Cities	8,030	60.32	28.01	9.29	2.38	11.67
	High Needs Urban/Suburban	15,566	46.22	35.51	14.17	4.10	18.27
	High Needs Rural	11,027	36.99	38.31	19.06	5.64	24.70
	Average Needs	57,933	24.52	37.26	26.29	11.93	38.22
	Low Needs	28,173	12.13	31.25	33.05	23.57	56.61
	Charter	5,806	21.58	39.24	27.42	11.76	39.18
	Private School	17,866	32.98	39.34	20.64	7.05	27.68
SWD	All Codes	31,161	63.84	25.67	8.03	2.46	10.49
SUA	All Codes	30,895	60.39	28.23	9.10	2.28	11.38
SWD/SUA	SUA=504 Plan Codes	18,082	69.05	23.32	6.34	1.29	7.63
ELL/SUA	SUA=ELL Codes	6,608	66.04	26.20	6.11	1.65	7.76
ELL	ELL status = Y	16,834	58.74	29.46	8.99	2.81	11.80
ELL Test Language	English	13,476	58.26	30.49	8.79	2.46	11.25
	Chinese	538	12.08	33.83	34.76	19.33	54.09
	Haitian-Creole	52	78.85	17.31	3.85	0.00	3.85
	Korean	31	16.13	16.13	12.90	54.84	67.74
	Russian	76	34.21	34.21	26.32	5.26	31.58
	Spanish	2,661	71.40	23.60	4.36	0.64	5.00
	All Translations	3,358	60.66	25.31	9.80	4.23	14.03

Mathematics Grade 5

Performance level summaries and N-counts of demographic groups for Grade 5 are presented in Table 46C. Statewide, 30% of the fifth-grade population was placed in Levels III and IV. As seen in Grades 3 and 4, there was little performance differentiation by gender. However, there were marked differences across ethnic, NRC, and test translation subgroups. Students in the Black, Hispanic, and American Indian subgroups had a lower percentage of students in Levels III and IV than the rest of the population, while the percentage of Asian and White students in Levels III and IV exceeded the overall population. About 49% of students from Low Needs districts were in Levels III or IV, but only 9% of Big 4 Cities students were in Levels III and IV. The ELL, SWD, and SUA subgroups were well below the performance of the general population as well, with the majority of the students in Levels I and II. For ELL students who used translated test forms, 93% of those who used the Haitian-Creole translation were classified in Level I, and no student was classified in Level III or Level IV; those who used the Korean translation had the highest percentage of students classified in Levels III and IV (71%). Again, these subgroups had a very small number of students. The following subgroups had a higher percentage of students meeting standards for Levels III and IV than the state population: Male, Asian, Native Hawaiian or Other Pacific Islander, White, New York City, Average Needs, Low Needs, English form, Chinese translation, and Korean translation.

Table 46C. Mathematics Performance Level Distribution by Subgroup, Grade 5

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	202,579	40.02	30.32	20.88	8.77	29.65
Gender	Female	99,557	38.92	31.97	21.11	8.00	29.11
	Male	103,022	41.09	28.74	20.66	9.51	30.18
Ethnicity	Asian	17,003	16.37	23.54	31.99	28.10	60.09
	Black	36,425	57.18	27.78	12.12	2.92	15.04
	Hispanic	47,043	52.38	29.95	14.28	3.39	17.67
	American Indian	1,122	50.71	28.61	14.62	6.06	20.68
	Multiracial	2,130	41.13	29.77	20.28	8.83	29.11
	Native Hawaiian or Other Pacific Islander	345	31.88	32.75	23.48	11.88	35.36
	White	98,511	31.74	32.64	25.44	10.18	35.62
NRC	New York City	69,104	42.07	28.27	19.62	10.04	29.66
	Big 4 Cities	7,816	71.92	19.11	7.28	1.69	8.97
	High Needs Urban/Suburban	14,992	58.45	28.13	11.07	2.35	13.42
	High Needs Rural	11,042	49.68	31.79	15.41	3.12	18.53
	Average Needs	57,641	35.53	33.69	23.17	7.60	30.77
	Low Needs	28,355	19.83	31.60	31.57	17.00	48.58
	Charter	7,079	39.06	34.52	20.30	6.12	26.42
	Private School	6,531	49.75	28.23	16.44	5.57	22.02
SWD	All Codes	31,675	75.62	17.52	5.55	1.31	6.86
SUA	All Codes	32,552	70.70	20.21	7.18	1.91	9.09
SWD/SUA	SUA=504 Plan Codes	19,571	78.93	15.95	4.28	0.84	5.11

Table 46C. Mathematics Performance Level Distribution by Subgroup, Grade 5 (cont.)

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
ELL/SUA	SUA=ELL Codes	7,523	73.00	18.66	6.21	2.13	8.33
ELL	ELL status = Y	16,410	71.86	19.65	6.54	1.94	8.49
ELL Test Language	English	12,890	71.65	20.45	6.29	1.61	7.90
	Chinese	494	20.85	27.73	33.40	18.02	51.42
	Haitian-Creole	75	93.33	6.67	0.00	0.00	0.00
	Korean	31	6.45	22.58	29.03	41.94	70.97
	Russian	84	58.33	15.48	19.05	7.14	26.19
	Spanish	2,836	82.26	15.02	2.57	0.14	2.72
	All Translations	3,520	72.64	16.70	7.47	3.18	10.65

Mathematics Grade 6

Performance level summaries and N-counts of demographic groups for Grade 6 are presented in Table 46D. Statewide, 30% of the fourth-grade population was placed in Levels III and IV. As seen in previous grades, there was little performance differentiation by gender. However, student achievement varied widely by ethnicity, NRC, and test translation subgroups. Students in the Black, Hispanic, and American Indian subgroups had a lower percentage of students in Levels III and IV than the rest of the population, while the percentage of Asian and White students in Levels III and IV exceeded the overall population. About 51% of students from Low Needs districts were in Levels III or IV, but only 10% of Big 4 Cities students were in Levels III and IV. The ELL, SWD, and SUA subgroups were well below the performance of the general population as well, with the majority of the students in Levels I and II. For ELL students who used translated test forms, those who used Haitian-Creole or Spanish had the lowest percentages of students classified in Levels III and IV (1% and 2%, respectively); those who used Korean translation had the highest percentage of students classified in Levels III and IV (68%). The following subgroups had a higher percentage of students meeting standards for Levels III and IV than the state population: Female, Asian, Native Hawaiian or Other Pacific Islander, White, Average Needs, Low Needs, English form, Chinese translation, and Korean translation.

Table 46D. Mathematics Performance Level Distribution by Subgroup, Grade 6

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	213,147	29.03	40.98	17.93	12.06	29.99
Gender	Female	104,829	27.94	41.68	18.41	11.97	30.38
	Male	108,318	30.08	40.30	17.48	12.14	29.62
Ethnicity	Asian	17,218	11.32	28.88	24.29	35.51	59.80
	Black	39,107	45.62	39.29	10.79	4.30	15.09
	Hispanic	49,176	39.61	43.02	12.15	5.22	17.37
	American Indian	1,092	39.10	42.49	12.09	6.32	18.41
	Multiracial	2,027	29.90	40.36	16.97	12.78	29.75
	Native Hawaiian or Other Pacific Islander	424	21.23	40.33	18.40	20.05	38.44
	White	104,103	20.64	42.65	22.38	14.34	36.71

Table 46D. Mathematics Performance Level Distribution by Subgroup, Grade 6 (cont.)

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
NRC	New York City	68,465	33.19	37.89	15.44	13.48	28.92
	Big 4 Cities	7,933	58.17	31.73	6.91	3.19	10.10
	High Needs Urban/Suburban	14,762	42.85	42.05	11.34	3.77	15.11
	High Needs Rural	11,435	33.42	46.86	14.81	4.91	19.72
	Average Needs	59,186	23.92	44.41	20.58	11.09	31.67
	Low Needs	28,777	11.66	36.92	27.53	23.90	51.42
	Charter	6,701	25.92	45.86	18.43	9.79	28.22
	Private School	15,863	32.28	46.25	15.15	6.32	21.47
SWD	All Codes	31,841	67.81	26.60	4.27	1.32	5.59
SUA	All Codes	32,323	63.00	29.90	5.31	1.79	7.10
SWD/SUA	SUA=504 Plan Codes	20,628	70.96	24.60	3.64	0.80	4.45
ELL/SUA	SUA=ELL Codes	5,607	72.12	22.45	3.50	1.93	5.42
ELL	ELL status = Y	13,063	65.74	26.33	5.05	2.87	7.92
ELL Test Language	English	9,301	66.07	27.35	4.45	2.13	6.58
	Chinese	642	15.89	35.67	25.08	23.36	48.44
	Haitian-Creole	143	86.71	12.59	0.70	0.00	0.70
	Korean	47	10.64	21.28	25.53	42.55	68.09
	Russian	95	33.68	55.79	9.47	1.05	10.53
	Spanish	2,835	76.90	20.67	2.22	0.21	2.43
	All Translations	3,762	64.94	23.82	6.54	4.70	11.24

Mathematics Grade 7

Performance level summaries and N-counts of demographic groups for Grade 7 are presented in Table 46E. Statewide, 28% of the seventh-grade population was placed in Levels III and IV. As seen in previous grades, there was little performance differentiation by gender. However, student achievement varied widely by ethnicity, NRC, and test translation subgroups. Students in the Black, Hispanic, and American Indian subgroups had a lower percentage of students in Levels III and IV than the rest of the population, while the percentage of Asian and White students in Levels III and IV exceeded the overall population. About 48% of students from Low Needs districts were in Levels III or IV, but only 6% of Big 4 Cities students were in Levels III and IV. The ELL, SWD, and SUA subgroups were well below the performance of the general population as well, with the majority of the students in Levels I and II. For ELL students who used translated test forms, 93% of students who used the Haitian-Creole translation were classified in Level I, and no student was classified in Level III or Level IV; those who used the Korean translation had the highest percentage of students classified in Levels III and IV (64%). The following subgroups had a higher percentage of students meeting standards for Levels III and IV than the state population: Female, Asian, Multiracial, Native Hawaiian or Other Pacific Islander, White, Average Needs, Low Needs, English form, Chinese translation, and Korean translation.

Table 46E. Mathematics Performance Level Distribution by Subgroup, Grade 7

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	207,251	37.97	34.42	20.27	7.34	27.61
Gender	Female	101,198	36.66	35.72	20.85	6.78	27.62
	Male	106,053	39.22	33.17	19.73	7.88	27.60
Ethnicity	Asian	17,553	16.09	27.22	30.72	25.97	56.69
	Black	37,786	60.10	28.90	9.29	1.70	10.99
	Hispanic	46,251	52.81	32.94	12.06	2.19	14.25
	American Indian	1,027	49.95	33.50	13.73	2.82	16.55
	Multiracial	1,836	38.29	33.33	19.61	8.77	28.38
	Native Hawaiian or Other Pacific Islander	398	31.91	28.39	22.86	16.83	39.70
	White	102,400	26.75	38.40	26.31	8.54	34.85
NRC	New York City	70,020	44.42	30.49	16.70	8.38	25.09
	Big 4 Cities	7,716	71.42	22.08	5.52	0.97	6.49
	High Needs Urban/Suburban	14,955	57.35	31.09	10.04	1.52	11.56
	High Needs Rural	11,648	44.30	39.19	14.59	1.92	16.51
	Average Needs	59,652	30.94	39.47	23.45	6.15	29.59
	Low Needs	31,525	16.46	35.18	33.78	14.58	48.36
	Charter	5,355	38.51	38.99	17.95	4.56	22.50
	Private School	6,368	41.19	36.73	17.27	4.81	22.08
SWD	All Codes	31,379	77.70	18.21	3.47	0.62	4.09
SUA	All Codes	26,091	72.11	21.57	5.18	1.14	6.32
SWD/SUA	SUA=504 Plan Codes	17,390	78.92	17.72	3.00	0.36	3.36
ELL/SUA	SUA=ELL Codes	3,805	81.13	12.72	4.57	1.58	6.15
ELL	ELL status = Y	12,307	77.74	16.10	4.67	1.50	6.17
ELL Test Language	English	8,372	80.05	15.66	3.34	0.94	4.29
	Chinese	703	18.49	36.27	33.00	12.23	45.23
	Haitian-Creole	157	92.99	7.01	0.00	0.00	0.00
	Korean	36	8.33	27.78	38.89	25.00	63.89
	Russian	129	58.14	27.13	10.85	3.88	14.73
	Spanish	2,910	86.29	12.34	1.20	0.17	1.37
	All Translations	3,935	72.81	17.03	7.50	2.67	10.17

Mathematics Grade 8

Performance level summaries and N-counts of demographic groups for Grade 8 are presented in Table 46F. Statewide, 27% of the eighth-grade population was placed in Levels III and IV. As seen in previous grades, there was little performance differentiation by gender. However, student achievement varied widely by ethnicity, NRC, and test translation subgroups. Students in the Black, Hispanic, and American Indian subgroups had a lower percentage of students in Levels III and IV than the rest of the population, while the percentage of Asian and White students in Levels III and IV exceeded the overall population. About 46% of students from Low Needs

districts were in Levels III or IV, but only 6% of Big 4 Cities students were in Levels III and IV. The ELL, SWD, and SUA subgroups were well below the performance of the general population as well, with the majority of the students in Levels I and II. For ELL students who used translated test forms, those who used Haitian-Creole or Spanish had the lowest percentages of students classified in Levels III and IV (2% and 3%, respectively); those who used Korean translation had the highest percentages of students classified in Levels III and IV (55%). The following subgroups had a higher percentage of students meeting standards for Levels III and IV than the state population: Female, Asian, Native Hawaiian or Other Pacific Islander, White, Average Needs, Low Needs, Charter, English form, Chinese translation, and Korean translation.

Table 46F. Mathematics Performance Level Distribution, by Subgroup Grade 8

Demographic Category (Subgroup)		N-count	Level I %	Level II %	Level III %	Level IV %	Level III & IV %
State	All Students	214,932	31.37	41.73	19.90	7.00	26.91
Gender	Female	105,389	29.15	43.55	20.51	6.79	27.30
	Male	109,543	33.50	39.97	19.32	7.21	26.53
Ethnicity	Asian	17,282	12.57	29.93	31.19	26.31	57.50
	Black	39,633	50.97	37.53	9.73	1.77	11.51
	Hispanic	47,763	43.38	41.28	12.64	2.70	15.34
	American Indian	1,038	42.87	41.14	12.43	3.56	15.99
	Multiracial	1,817	32.91	41.50	18.77	6.82	25.59
	Native Hawaiian or Other Pacific Islander	410	26.59	41.71	20.00	11.71	31.71
	White	106,989	21.66	45.39	25.18	7.76	32.94
NRC	New York City	69,217	36.85	37.29	17.21	8.65	25.85
	Big 4 Cities	7,492	70.72	23.68	4.50	1.11	5.61
	High Needs Urban/Suburban	14,651	47.88	40.33	9.90	1.89	11.79
	High Needs Rural	11,707	37.66	47.54	12.93	1.86	14.79
	Average Needs	59,339	24.23	47.00	22.86	5.91	28.77
	Low Needs	31,249	12.09	41.79	33.22	12.90	46.12
	Charter	4,405	24.54	45.36	23.52	6.58	30.10
	Private School	16,836	35.16	45.56	15.36	3.93	19.29
SWD	All Codes	30,749	71.06	24.69	3.65	0.60	4.25
SUA	All Codes	25,996	63.71	29.33	5.69	1.28	6.97
SWD/SUA	SUA=504 Plan Codes	17,166	70.95	25.34	3.24	0.46	3.70
ELL/SUA	SUA=ELL Codes	3,882	70.71	21.28	5.72	2.29	8.01
ELL	ELL status = Y	13,148	65.74	26.32	6.05	1.88	7.93
ELL Test Language	English	8,879	67.47	26.50	4.75	1.27	6.03
	Chinese	776	15.85	34.92	34.15	15.08	49.23
	Haitian-Creole	165	70.30	27.27	1.82	0.61	2.42
	Korean	33	18.18	27.27	33.33	21.21	54.55
	Russian	109	38.53	42.20	15.60	3.67	19.27
	Spanish	3,186	74.26	23.13	2.45	0.16	2.61
	All Translations	4,269	62.15	25.95	8.76	3.14	11.90

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	0	0	7
		3	0	0	4	0	4
	3	4	0	0	6	0	6
	Total		**31**	**6**	**10**	**0**	**47**
4	1	1	24	6	0	0	30
	2	2	7	0	0	0	7
		3	0	0	4	0	4
	3	4	0	0	6	0	6
	Total		**31**	**6**	**10**	**0**	**47**
5	1	1	28	14	0	0	42
	2	2	14	7	0	0	21
		3	0	0	4	0	4
	3	4	0	0	6	0	6
	Total		**42**	**21**	**10**	**0**	**73**
6*	1	1	28	14	0	0	42
	2	2	14	7	0	0	21
		3	0	0	4	0	4
	3	4	0	0	6	0	6
	Total		**42**	**21**	**10**	**0**	**73**
7	1	1	28	14	0	0	42
	2	2	14	7	0	0	21
		3	0	0	4	0	4
	3	4	0	0	6	0	6
	Total		**42**	**21**	**10**	**0**	**73**
8	1	1	28	14	0	0	42
	2	2	14	7	0	0	21
		3	0	0	4	0	4
	3	4	0	0	6	0	6
	Total		**42**	**21**	**10**	**0**	**73**

*For grade 6, one item was excluded from the analysis and scoring due to poor fit to the item response theory (IRT) model.

Table A2. Mathematics Test Configuration

Grade	Day	Book	Number of Items				
			Multiple-Choice		Constructed-Response		Total
			Operational	Embedded	Operational	Embedded	
3	1	1	20	10	0	0	30
	2	2	21	10	0	0	31
	3	3	0	0	8	0	8
	Total		**41**	**20**	**8**	**0**	**69**
4	1	1	20	10	0	0	30
	2	2	22	10	0	0	32
	3	3	0	0	10	0	10
	Total		**42**	**20**	**10**	**0**	**72**
5	1	1	20	10	0	0	30
	2	2	22	10	0	0	32
	3	3	0	0	10	0	10
	Total		**42**	**20**	**10**	**0**	**72**
6	1	1	24	10	0	0	34
	2	2	24	10	0	0	34
	3	3	0	0	10	0	10
	Total		**48**	**20**	**10**	**0**	**78**
7	1	1	24	10	0	0	34
	2	2	24	10	0	0	34
	3	3	0	0	10	0	10
	Total		**48**	**20**	**10**	**0**	**78**
8	1	1	24	10	0	0	34
	2	2	24	10	0	0	34
	3	3	0	0	10	0	10
	Total		**48**	**20**	**10**	**0**	**78**

Table A3. ELA Testing Times

Grade	Day	Book	Estimated Time on Task	Session Time
3	1	1	50	70
	2	2	50	70
		3		
	3	4	50	70
	Total		**150**	**210**
4	1	1	50	70
	2	2	50	70
		3		
	3	4	50	70
	Total		**150**	**210**
5	1	1	70	90
	2	2	70	90
		3		
	3	4	50	90
	Total		**190**	**270**
6	1	1	70	90
	2	2	70	90
		3		
	3	4	50	90
	Total		**190**	**270**
7	1	1	70	90
	2	2	70	90
		3		
	3	4	50	90
	Total		**190**	**270**
8	1	1	70	90
	2	2	70	90
		3		
	3	4	50	90
	Total		**190**	**270**

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

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

Table A4. Math Testing Times

Grade	Day	Book	Estimated Time on Task	Session Time
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	70	90
	Total		**170**	**230**
5	1	1	50	90
	2	2	50	90
	3	3	70	90
	Total		**170**	**270**
6	1	1	50	90
	2	2	50	90
	3	3	70	90
	Total		**170**	**270**
7	1	1	50	90
	2	2	50	90
	3	3	70	90
	Total		**170**	**270**
8	1	1	50	90
	2	2	50	90
	3	3	70	90
	Total		**170**	**270**

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

- Average time to respond to a multiple-choice question- 1.5 minutes
- Average time to respond to a two-point constructed response question- 5 minutes
- Average time to responds 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 2013 *Teacher's Directions* and the School Administrator's Manual, which are accessible online.

- *2013 Common Core ELA Teacher's Directions*
  - http://www.p12.nysed.gov/assessment/ei/2013/td-ela35-13.pdf
  - http://www.p12.nysed.gov/assessment/ei/2013/td-ela68-13.pdf
- *2013 Common Core Mathematics Teacher's Directions*
  - http://www.p12.nysed.gov/assessment/ei/2013/td-math35-13.pdf
  - http://www.p12.nysed.gov/assessment/ei/2013/td-math68-13.pdf
- *2013 Common Core ELA and Mathematics Tests School Administrator's Manual*
  - http://www.p12.nysed.gov/assessment/sam/ei-samcc13rev.pdf

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	24	33%–80%	44%
		Information	18–44	28	33%–80%	51%
		Language	1–4	3	2%–7%	5%
4	55	Literature	18–44	26	33%–80%	47%
		Information	18–44	27	33%–80%	49%
		Language	1–4	2	2%–7%	4%
5	66	Literature	18–51	32	27%–77%	48%
		Information	18–51	32	27%–77%	48%
		Language	1–4	2	2%–6%	3%
6	65	Literature	11–44	24	17%–67%	37%
		Information	25–58	40	38%–88%	62%
		Language	1–4	1	2%–6%	2%
7	66	Literature	11–44	24	17%–67%	36%
		Information	25–58	40	38%–88%	61%
		Language	1–4	2	2%–6%	3%
8	66	Literature	11–44	17	17%–67%	26%
		Information	25–58	46	38%–88%	70%
		Language	1–4	3	2%–6%	5%

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	17	23%–32%	26%
		Number and Operations – Fractions	22–28	26	33%–42%	39%
		Measurement and Data	13–21	17	20%–32%	26%
		Geometry*	1–3	2	2%–5%	3%
6	72	Ratios and Proportional Relationships	16–20	19	22%–28%	26%
		The Number System	13–19	15	18%–26%	21%
		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	17	19%–28%	24%
		Geometry	16–22	19	22%–31%	26%
		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 2013 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 CCLS. While many of the expectations outlined in the CCLS align with previous versions of the NYS Learning Standards for ELA, the CCLS do represent some shifts in emphasis with direct implications for assessment development. In particular the CCLS 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 the workplace lies in the texts with which they interact. By the time they graduate, students should be prepared to successfully read and analyze the types of complex texts they will encounter after high school. Selecting passages of appropriate type and complexity for use in assessment is integral to this preparation.

One of the major shifts of the CCLS 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 they will encounter in college and 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 CCLS 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 CCLS 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 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 2013 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 2013 item selection included 2012 embedded and stand-alone field test items and items field-tested in New York State in 2012.

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 groups. 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 2012 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 spring 2012 operational test as a reference.

Appendix G: Operational Item Maps

The following are the operational item maps for the 2013 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. ELA Grade 3 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.3.1
2	MC	1	CCSS.ELA-Literacy.L.3.4a
3	MC	1	CCSS.ELA-Literacy.RI.3.8
4	MC	1	CCSS.ELA-Literacy.RI.3.4
5	MC	1	CCSS.ELA-Literacy.RI.3.1
6	MC	1	CCSS.ELA-Literacy.RI.3.2
7	MC	1	CCSS.ELA-Literacy.L.3.4a
8	MC	1	CCSS.ELA-Literacy.RL.3.4
9	MC	1	CCSS.ELA-Literacy.RL.3.3
10	MC	1	CCSS.ELA-Literacy.RL.3.1
11	MC	1	CCSS.ELA-Literacy.RL.3.3
12	MC	1	CCSS.ELA-Literacy.RL.3.3
13	MC	1	CCSS.ELA-Literacy.L.3.5a
14	MC	1	CCSS.ELA-Literacy.RL.3.1
15	MC	1	CCSS.ELA-Literacy.RL.3.1
16	MC	1	CCSS.ELA-Literacy.RL.3.3
17	MC	1	CCSS.ELA-Literacy.RL.3.2
18	MC	1	CCSS.ELA-Literacy.RL.3.2
19	MC	1	CCSS.ELA-Literacy.RI.3.5
20	MC	1	CCSS.ELA-Literacy.RI.3.3
21	MC	1	CCSS.ELA-Literacy.RI.3.1
22	MC	1	CCSS.ELA-Literacy.RI.3.5
23	MC	1	CCSS.ELA-Literacy.RI.3.1
24	MC	1	CCSS.ELA-Literacy.RI.3.2
25	MC	1	CCSS.ELA-Literacy.RI.3.3
26	MC	1	CCSS.ELA-Literacy.RI.3.4
27	MC	1	CCSS.ELA-Literacy.RI.3.8
28	MC	1	CCSS.ELA-Literacy.RI.3.1
29	MC	1	CCSS.ELA-Literacy.RI.3.3
30	MC	1	CCSS.ELA-Literacy.RI.3.7
31	MC	1	CCSS.ELA-Literacy.RI.3.1
32	CR	2	CCSS.ELA-Literacy.RI.3.8
33	CR	2	CCSS.ELA-Literacy.RI.3.2
34	CR	2	CCSS.ELA-Literacy.RL.3.4
35	CR	4	CCSS.ELA-Literacy.RL.3.3

Table G1. ELA Grade 3 Operational Item Map (cont.)

Item	Type	Points	Standard
36	CR	2	CCSS.ELA-Literacy.RL.3.3
37	CR	2	CCSS.ELA-Literacy.RL.3.3
38	CR	2	CCSS.ELA-Literacy.RL.3.3
39	CR	2	CCSS.ELA-Literacy.RL.3.2
40	CR	2	CCSS.ELA-Literacy.RI.3.1
41	CR	4	CCSS.ELA-Literacy.RI.3.2

Table G2. ELA Grade 4 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.4.1
2	MC	1	CCSS.ELA-Literacy.RL.4.1
3	MC	1	CCSS.ELA-Literacy.RL.4.2
4	MC	1	CCSS.ELA-Literacy.RL.4.3
5	MC	1	CCSS.ELA-Literacy.RL.4.2
6	MC	1	CCSS.ELA-Literacy.RL.4.6
7	MC	1	CCSS.ELA-Literacy.RL.4.3
8	MC	1	CCSS.ELA-Literacy.RL.4.1
9	MC	1	CCSS.ELA-Literacy.RL.4.5
10	MC	1	CCSS.ELA-Literacy.RL.4.3
11	MC	1	CCSS.ELA-Literacy.RL.4.4
12	MC	1	CCSS.ELA-Literacy.RL.4.2
13	MC	1	CCSS.ELA-Literacy.L.4.4c
14	MC	1	CCSS.ELA-Literacy.RI.4.1
15	MC	1	CCSS.ELA-Literacy.RI.4.2
16	MC	1	CCSS.ELA-Literacy.RI.4.8
17	MC	1	CCSS.ELA-Literacy.RI.4.2
18	MC	1	CCSS.ELA-Literacy.RI.4.5
19	MC	1	CCSS.ELA-Literacy.RI.4.8
20	MC	1	CCSS.ELA-Literacy.RI.4.1
21	MC	1	CCSS.ELA-Literacy.RI.4.3
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.4.8
25	MC	1	CCSS.ELA-Literacy.RI.4.4
26	MC	1	CCSS.ELA-Literacy.RI.4.1
27	MC	1	CCSS.ELA-Literacy.RI.4.2
28	MC	1	CCSS.ELA-Literacy.RI.4.1
29	MC	1	CCSS.ELA-Literacy.RI.4.3
30	MC	1	CCSS.ELA-Literacy.RI.4.7
31	MC	1	CCSS.ELA-Literacy.L.4.4a

Table G2. ELA Grade 4 Operational Item Map (cont.)

Item	Type	Points	Standard
32	CR	2	CCSS.ELA-Literacy.RI.4.8
33	CR	2	CCSS.ELA-Literacy.RI.4.5
34	CR	2	CCSS.ELA-Literacy.RI.4.1
35	CR	4	CCSS.ELA-Literacy.RI.4.9
36	CR	2	CCSS.ELA-Literacy.RL.4.3
37	CR	2	CCSS.ELA-Literacy.RL.4.3
38	CR	2	CCSS.ELA-Literacy.RL.4.1
39	CR	2	CCSS.ELA-Literacy.RL.4.2
40	CR	2	CCSS.ELA-Literacy.RL.4.3
41	CR	4	CCSS.ELA-Literacy.RL.4.3

Table G3. ELA Grade 5 Operational Item Map

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

Table G3. ELA Grade 5 Operational Item Map (cont.)

Item	Type	Points	Standard
30	MC	1	CCSS.ELA-Literacy.RI.5.7
31	MC	1	CCSS.ELA-Literacy.RI.5.4
32	MC	1	CCSS.ELA-Literacy.RI.5.1
33	MC	1	CCSS.ELA-Literacy.RI.5.3
34	MC	1	CCSS.ELA-Literacy.RI.5.8
35	MC	1	CCSS.ELA-Literacy.RI.5.3
36	MC	1	CCSS.ELA-Literacy.RL.5.3
37	MC	1	CCSS.ELA-Literacy.RL.5.1
38	MC	1	CCSS.ELA-Literacy.RL.5.3
39	MC	1	CCSS.ELA-Literacy.L.5.4a
40	MC	1	CCSS.ELA-Literacy.RL.5.3
41	MC	1	CCSS.ELA-Literacy.RL.5.6
42	MC	1	CCSS.ELA-Literacy.RL.5.2
43	CR	2	CCSS.ELA-Literacy.RL.5.3
44	CR	2	CCSS.ELA-Literacy.RL.5.1
45	CR	2	CCSS.ELA-Literacy.RI.5.8
46	CR	4	CCSS.ELA-Literacy.RI.5.2
47	CR	2	CCSS.ELA-Literacy.RI.5.2
48	CR	2	CCSS.ELA-Literacy.RI.5.8
49	CR	2	CCSS.ELA-Literacy.RI.5.2
50	CR	2	CCSS.ELA-Literacy.RL.5.1
51	CR	2	CCSS.ELA-Literacy.RL.5.2
52	CR	4	CCSS.ELA-Literacy.RL.5.9

Table G4. ELA Grade 6 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.6.2
2	MC	1	CCSS.ELA-Literacy.RI.6.7
3	MC	1	CCSS.ELA-Literacy.RI.6.8
4	MC	1	CCSS.ELA-Literacy.RI.6.4
5	MC	1	CCSS.ELA-Literacy.RI.6.1
6	MC	1	CCSS.ELA-Literacy.RI.6.1
7	MC	1	CCSS.ELA-Literacy.RI.6.5
8	MC	1	CCSS.ELA-Literacy.RI.6.1
9	MC	1	CCSS.ELA-Literacy.RI.6.1
10	MC	1	CCSS.ELA-Literacy.RI.6.1
11	MC	1	CCSS.ELA-Literacy.RI.6.5
12	MC	1	CCSS.ELA-Literacy.RI.6.8
13	MC	1	CCSS.ELA-Literacy.RI.6.2
14	MC	1	CCSS.ELA-Literacy.RI.6.5
15	MC	1	CCSS.ELA-Literacy.RL.6.4
16	MC	1	CCSS.ELA-Literacy.RL.6.2

Table G4. ELA Grade 6 Operational Item Map (cont.)

Item	Type	Points	Standard
17	MC	1	CCSS.ELA-Literacy.RL.6.1
18	MC	1	CCSS.ELA-Literacy.RL.6.5
19	MC	1	CCSS.ELA-Literacy.RL.6.4
20	MC	1	CCSS.ELA-Literacy.RL.6.2
21	MC	1	CCSS.ELA-Literacy.RI.6.3
22	MC	1	CCSS.ELA-Literacy.RI.6.7
23	MC	1	CCSS.ELA-Literacy.RI.6.1
24	MC	1	CCSS.ELA-Literacy.RI.6.5
25	MC	1	CCSS.ELA-Literacy.RI.6.1
26	MC	1	CCSS.ELA-Literacy.RI.6.6
27	MC	1	CCSS.ELA-Literacy.RI.6.6
28	MC	1	CCSS.ELA-Literacy.RI.6.5
29	MC	1	CCSS.ELA-Literacy.RI.6.7
30	MC	1	CCSS.ELA-Literacy.RI.6.3
31	MC	1	CCSS.ELA-Literacy.RI.6.8
32	MC	1	CCSS.ELA-Literacy.L.6.4a
33	MC	1	CCSS.ELA-Literacy.RI.6.2
34	MC	1	CCSS.ELA-Literacy.RL.6.1
35	MC	1	CCSS.ELA-Literacy.RL.6.3
36	MC	1	CCSS.ELA-Literacy.RL.6.4
37	MC	1	CCSS.ELA-Literacy.RL.6.4
38	MC	1	CCSS.ELA-Literacy.RL.6.5
39	MC	1	CCSS.ELA-Literacy.RL.6.5
40	MC	1	CCSS.ELA-Literacy.RL.6.3
41	MC	1	CCSS.ELA-Literacy.RL.6.6
42	CR	2	CCSS.ELA-Literacy.RI.6.3
43	CR	2	CCSS.ELA-Literacy.RI.6.5
44	CR	2	CCSS.ELA-Literacy.RL.6.2
45	CR	4	CCSS.ELA-Literacy.RL.6.5
46	CR	2	CCSS.ELA-Literacy.RL.6.1
47	CR	2	CCSS.ELA-Literacy.RL.6.1
48	CR	2	CCSS.ELA-Literacy.RI.6.2
49	CR	2	CCSS.ELA-Literacy.RI.6.4
50	CR	2	CCSS.ELA-Literacy.RI.6.1
51	CR	4	CCSS.ELA-Literacy.RI.6.9

Table G5. ELA Grade 7 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.7.6
2	MC	1	CCSS.ELA-Literacy.RL.7.1
3	MC	1	CCSS.ELA-Literacy.RL.7.4
4	MC	1	CCSS.ELA-Literacy.RL.7.1
5	MC	1	CCSS.ELA-Literacy.RL.7.4
6	MC	1	CCSS.ELA-Literacy.RL.7.2
7	MC	1	CCSS.ELA-Literacy.RL.7.1
8	MC	1	CCSS.ELA-Literacy.RI.7.3
9	MC	1	CCSS.ELA-Literacy.RI.7.8
10	MC	1	CCSS.ELA-Literacy.RI.7.3
11	MC	1	CCSS.ELA-Literacy.RI.7.2
12	MC	1	CCSS.ELA-Literacy.RI.7.5
13	MC	1	CCSS.ELA-Literacy.RI.7.2
14	MC	1	CCSS.ELA-Literacy.RL.7.3
15	MC	1	CCSS.ELA-Literacy.RL.7.4
16	MC	1	CCSS.ELA-Literacy.RL.7.1
17	MC	1	CCSS.ELA-Literacy.RL.7.3
18	MC	1	CCSS.ELA-Literacy.RL.7.6
19	MC	1	CCSS.ELA-Literacy.RL.7.6
20	MC	1	CCSS.ELA-Literacy.RL.7.2
21	MC	1	CCSS.ELA-Literacy.RI.7.3
22	MC	1	CCSS.ELA-Literacy.RI.7.4
23	MC	1	CCSS.ELA-Literacy.RI.7.3
24	MC	1	CCSS.ELA-Literacy.RI.7.1
25	MC	1	CCSS.ELA-Literacy.RI.7.1
26	MC	1	CCSS.ELA-Literacy.RI.7.6
27	MC	1	CCSS.ELA-Literacy.RI.7.5
28	MC	1	CCSS.ELA-Literacy.RI.7.2
29	MC	1	CCSS.ELA-Literacy.RI.7.5
30	MC	1	CCSS.ELA-Literacy.RI.7.3
31	MC	1	CCSS.ELA-Literacy.RI.7.8
32	MC	1	CCSS.ELA-Literacy.RI.7.5
33	MC	1	CCSS.ELA-Literacy.RI.7.1
34	MC	1	CCSS.ELA-Literacy.L.7.4
35	MC	1	CCSS.ELA-Literacy.RI.7.1
36	MC	1	CCSS.ELA-Literacy.RL.7.4
37	MC	1	CCSS.ELA-Literacy.L.7.4
38	MC	1	CCSS.ELA-Literacy.RL.7.1
39	MC	1	CCSS.ELA-Literacy.RL.7.2
40	MC	1	CCSS.ELA-Literacy.RL.7.3
41	MC	1	CCSS.ELA-Literacy.RL.7.3
42	MC	1	CCSS.ELA-Literacy.RL.7.5

Table G5. ELA Grade 7 Operational Item Map (cont.)

Item	Type	Points	Standard
43	CR	2	CCSS.ELA-Literacy.RL.7.1
44	CR	2	CCSS.ELA-Literacy.RL.7.1
45	CR	2	CCSS.ELA-Literacy.RI.7.3
46	CR	4	CCSS.ELA-Literacy.RI.7.3
47	CR	2	CCSS.ELA-Literacy.RI.7.5
48	CR	2	CCSS.ELA-Literacy.RI.7.1
49	CR	2	CCSS.ELA-Literacy.RI.7.2
50	CR	2	CCSS.ELA-Literacy.RI.7.5
51	CR	2	CCSS.ELA-Literacy.RI.7.1
52	CR	4	CCSS.ELA-Literacy.RI.7.9

Table G6. ELA Grade 8 Operational Item Map

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

Table G6. ELA Grade 8 Operational Item Map (cont.)

Item	Type	Points	Standard
30	MC	1	CCSS.ELA-Literacy.RI.8.1
31	MC	1	CCSS.ELA-Literacy.RI.8.3
32	MC	1	CCSS.ELA-Literacy.RI.8.8
33	MC	1	CCSS.ELA-Literacy.RI.8.3
34	MC	1	CCSS.ELA-Literacy.RI.8.1
35	MC	1	CCSS.ELA-Literacy.RI.8.2
36	MC	1	CCSS.ELA-Literacy.RL.8.4
37	MC	1	CCSS.ELA-Literacy.RL.8.3
38	MC	1	CCSS.ELA-Literacy.RL.8.4
39	MC	1	CCSS.ELA-Literacy.RL.8.2
40	MC	1	CCSS.ELA-Literacy.RL.8.2
41	MC	1	CCSS.ELA-Literacy.RL.8.2
42	MC	1	CCSS.ELA-Literacy.RL.8.6
43	CR	2	CCSS.ELA-Literacy.RL.8.3
44	CR	2	CCSS.ELA-Literacy.RL.8.4
45	CR	2	CCSS.ELA-Literacy.RI.8.3
46	CR	4	CCSS.ELA-Literacy.RI.8.2
47	CR	2	CCSS.ELA-Literacy.RI.8.4
48	CR	2	CCSS.ELA-Literacy.RI.8.5
49	CR	2	CCSS.ELA-Literacy.RI.8.1
50	CR	2	CCSS.ELA-Literacy.RI.8.8
51	CR	2	CCSS.ELA-Literacy.RI.8.1
52	CR	4	CCSS.ELA-Literacy.RI.8.3

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

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

Item	Type	Points	Standard
42	CR	2	CCSS.Math.Content.3.OA.D.8
43	CR	2	CCSS.Math.Content.3.MD.B.3
44	CR	2	CCSS.Math.Content.3.OA.D.9
45	CR	2	CCSS.Math.Content.3.NF.A.1
46	CR	2	CCSS.Math.Content.3.G.A.2
47	CR	3	CCSS.Math.Content.3.OA.A.3
48	CR	3	CCSS.Math.Content.3.MD.C.7a
49	CR	3	CCSS.Math.Content.3.NF.A.3b

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.4.G.A.1
3	MC	1	CCSS.Math.Content.4.OA.A.2
4	MC	1	CCSS.Math.Content.4.NBT.A.2
5	MC	1	CCSS.Math.Content.4.G.A.1
6	MC	1	CCSS.Math.Content.4.OA.A.2
7	MC	1	CCSS.Math.Content.4.NBT.A.1
8	MC	1	CCSS.Math.Content.4.MD.B.4
9	MC	1	CCSS.Math.Content.4.NBT.A.3
10	MC	1	CCSS.Math.Content.4.NF.A.1
11	MC	1	CCSS.Math.Content.4.NBT.B.5
12	MC	1	CCSS.Math.Content.4.NF.B.3c
13	MC	1	CCSS.Math.Content.4.NBT.B.5
14	MC	1	CCSS.Math.Content.4.NF.A.2
15	MC	1	CCSS.Math.Content.3.MD.D.8
16	MC	1	CCSS.Math.Content.4.NF.A.2
17	MC	1	CCSS.Math.Content.4.NBT.B.5
18	MC	1	CCSS.Math.Content.4.NF.A.2
19	MC	1	CCSS.Math.Content.4.G.A.3
20	MC	1	CCSS.Math.Content.4.NBT.B.6
21	MC	1	CCSS.Math.Content.4.NF.A.1
22	MC	1	CCSS.Math.Content.4.OA.B.4
23	MC	1	CCSS.Math.Content.4.NF.B.4a
24	MC	1	CCSS.Math.Content.4.NBT.B.6
25	MC	1	CCSS.Math.Content.4.MD.A.3
26	MC	1	CCSS.Math.Content.4.NBT.A.2
27	MC	1	CCSS.Math.Content.3.MD.D.8
28	MC	1	CCSS.Math.Content.4.OA.C.5
29	MC	1	CCSS.Math.Content.4.NF.A.1
30	MC	1	CCSS.Math.Content.4.OA.A.2
31	MC	1	CCSS.Math.Content.4.MD.A.3

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

Item	Type	Points	Standard
32	MC	1	CCSS.Math.Content.4.G.A.1
33	MC	1	CCSS.Math.Content.4.NF.A.1
34	MC	1	CCSS.Math.Content.4.NBT.A.2
35	MC	1	CCSS.Math.Content.4.NF.B.3d
36	MC	1	CCSS.Math.Content.4.NF.A.2
37	MC	1	CCSS.Math.Content.4.OA.A.1
38	MC	1	CCSS.Math.Content.4.OA.A.3
39	MC	1	CCSS.Math.Content.4.MD.A.3
40	MC	1	CCSS.Math.Content.4.MD.C.5b
41	MC	1	CCSS.Math.Content.4.NBT.A.1
42	MC	1	CCSS.Math.Content.4.OA.A.2
43	CR	2	CCSS.Math.Content.4.G.A.2
44	CR	2	CCSS.Math.Content.4.NF.B.4c
45	CR	2	CCSS.Math.Content.4.MD.C.6
46	CR	2	CCSS.Math.Content.4.OA.A.2
47	CR	2	CCSS.Math.Content.4.NBT.B.6
48	CR	2	CCSS.Math.Content.4.NF.B.4c
49	CR	3	CCSS.Math.Content.4.NF.B.3d
50	CR	3	CCSS.Math.Content.4.OA.A.3
51	CR	3	CCSS.Math.Content.4.NBT.B.5
52	CR	3	CCSS.Math.Content.4.MD.C.7

Table G9. Mathematics Grade 5 Operational Item Map

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

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

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

Table G10. Mathematics Grade 6 Operational Item Map

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

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

Item	Type	Points	Standard
44	MC	1	CCSS.Math.Content.6.NS.C.8
45	MC	1	CCSS.Math.Content.6.RP.A.3b
46	MC	1	CCSS.Math.Content.6.EE.A.3
47	MC	1	CCSS.Math.Content.6.G.A.1
48	MC	1	CCSS.Math.Content.6.RP.A.3c
49	CR	2	CCSS.Math.Content.5.OA.B.3
50	CR	2	CCSS.Math.Content.6.EE.B.6
51	CR	2	CCSS.Math.Content.6.NS.B.4
52	CR	2	CCSS.Math.Content.6.RP.A.3a
53	CR	2	CCSS.Math.Content.6.EE.A.3
54	CR	2	CCSS.Math.Content.6.NS.A.1
55	CR	3	CCSS.Math.Content.6.RP.A.2
56	CR	3	CCSS.Math.Content.6.EE.B.7
57	CR	3	CCSS.Math.Content.6.RP.A.3a
58	CR	3	CCSS.Math.Content.6.EE.C.9

Table G11. Mathematics Grade 7 Operational Item Map

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

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

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

Table G12. Mathematics Grade 8 Operational Item Map

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

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

Item	Type	Points	Standard
44	MC	1	CCSS.Math.Content.8.SP.A.4
45	MC	1	CCSS.Math.Content.8.EE.B.6
46	MC	1	CCSS.Math.Content.8.F.A.2
47	MC	1	CCSS.Math.Content.8.EE.C.7b
48	MC	1	CCSS.Math.Content.8.G.A.1b
49	CR	2	CCSS.Math.Content.8.EE.C.7b
50	CR	2	CCSS.Math.Content.8.F.A.3
51	CR	2	CCSS.Math.Content.8.G.A.4
52	CR	2	CCSS.Math.Content.8.F.B.4
53	CR	2	CCSS.Math.Content.8.EE.C.8a
54	CR	2	CCSS.Math.Content.8.G.A.2
55	CR	3	CCSS.Math.Content.8.EE.C.8b
56	CR	3	CCSS.Math.Content.8.F.B.4
57	CR	3	CCSS.Math.Content.8.G.C.9
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 or 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

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 selected subpopulations: 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. Factor Analysis Results for ELA Tests (Selected Subpopulations)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
3	ELL	**1**	**6.30**	**15.36**	**15.36**
		2	1.57	3.83	19.19
		3	1.12	2.74	21.93
		4	1.12	2.73	24.66
		5	1.08	2.63	27.29
		6	1.06	2.59	29.88
		7	1.04	2.54	32.42
		8	1.01	2.47	34.89
		9	1.01	2.46	37.35
	SWD	**1**	**7.37**	**17.99**	**17.99**
		2	1.60	3.89	21.88
		3	1.14	2.79	24.67
		4	1.10	2.68	27.35
		5	1.06	2.60	29.95
		6	1.05	2.57	32.52
		7	1.02	2.50	35.02
		8	1.01	2.47	37.49
	SUA	**1**	**7.48**	**18.24**	**18.24**
		2	1.56	3.81	22.05
		3	1.15	2.80	24.85
		4	1.09	2.65	27.50
		5	1.05	2.57	30.07
		6	1.03	2.52	32.59
		7	1.01	2.46	35.05
		8	1.00	2.44	37.49
	SWD & SUA	**1**	**6.86**	**16.73**	**16.73**
		2	1.59	3.88	20.61
		3	1.15	2.81	23.42
		4	1.12	2.73	26.15
		5	1.08	2.63	28.78
		6	1.06	2.58	31.36
		7	1.04	2.53	33.89
		8	1.03	2.51	36.40

Table L1. Factor Analysis Results for ELA Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
3	ELL & SUA	**1**	**6.06**	**14.79**	**14.79**
		2	1.51	3.67	18.46
		3	1.17	2.84	21.3
		4	1.12	2.73	24.03
		5	1.10	2.68	26.71
		6	1.05	2.55	29.26
		7	1.04	2.54	31.80
		8	1.03	2.51	34.31
		9	1.02	2.49	36.80
		10	1.01	2.48	39.28
4	ELL	**1**	**6.34**	**15.47**	**15.47**
		2	1.53	3.72	19.19
		3	1.15	2.80	21.99
		4	1.08	2.64	24.63
		5	1.07	2.60	27.23
		6	1.04	2.54	29.77
		7	1.03	2.51	32.28
		8	1.02	2.49	34.77
		9	1.01	2.45	37.22
	SWD	**1**	**7.65**	**18.66**	**18.66**
		2	1.56	3.81	22.47
		3	1.16	2.84	25.31
		4	1.05	2.57	27.88
		5	1.01	2.47	30.35
		6	1.00	2.45	32.80
	SUA	**1**	**7.79**	**19.01**	**19.01**
		2	1.55	3.79	22.80
		3	1.17	2.85	25.65
		4	1.03	2.51	28.16
		5	1.01	2.47	30.63
	SWD & SUA	**1**	**7.21**	**17.59**	**17.59**
		2	1.53	3.73	21.32
		3	1.20	2.93	24.25
		4	1.05	2.56	26.81
		5	1.05	2.55	29.36
		6	1.03	2.51	31.87
		7	1.02	2.48	34.35

Table L1. Factor Analysis Results for ELA Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
4	ELL & SUA	**1**	**6.09**	**14.86**	**14.86**
		2	1.42	3.45	18.31
		3	1.18	2.89	21.20
		4	1.13	2.77	23.97
		5	1.07	2.62	26.59
		6	1.07	2.61	29.20
		7	1.06	2.58	31.78
		8	1.05	2.57	34.35
		9	1.04	2.54	36.89
		10	1.03	2.51	39.40
		11	1.02	2.48	41.88
5	ELL	**1**	**6.43**	**12.37**	**12.37**
		2	1.46	2.81	15.18
		3	1.23	2.37	17.55
		4	1.14	2.2	19.75
		5	1.11	2.14	21.89
		6	1.11	2.13	24.02
		7	1.07	2.05	26.07
		8	1.06	2.03	28.10
		9	1.04	2.00	30.10
		10	1.04	2.00	32.10
		11	1.02	1.97	34.07
		12	1.01	1.95	36.02
		13	1.01	1.93	37.95
	SWD	**1**	**7.55**	**14.51**	**14.51**
		2	1.52	2.93	17.44
		3	1.16	2.23	19.67
		4	1.11	2.14	21.81
		5	1.06	2.04	23.85
		6	1.06	2.03	25.88
		7	1.03	1.99	27.87
		8	1.02	1.97	29.84
		9	1.02	1.96	31.80
		10	1.01	1.94	33.74
		11	1.00	1.93	35.67

Table L1. Factor Analysis Results for ELA Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
5	SUA	**1**	**7.99**	**15.37**	**15.37**
		2	1.48	2.85	18.22
		3	1.17	2.24	20.46
		4	1.10	2.12	22.58
		5	1.07	2.06	24.64
		6	1.06	2.03	26.67
		7	1.03	1.99	28.66
		8	1.02	1.96	30.62
		9	1.01	1.94	32.56
		10	1.00	1.92	34.48
	SWD & SUA	**1**	**7.12**	**13.68**	**13.68**
		2	1.52	2.93	16.61
		3	1.17	2.25	18.86
		4	1.11	2.14	21.00
		5	1.09	2.10	23.10
		6	1.08	2.07	25.17
		7	1.05	2.02	27.19
		8	1.04	2.00	29.19
		9	1.03	1.99	31.18
		10	1.03	1.98	33.16
		11	1.02	1.95	35.11
		12	1.00	1.93	37.04
	ELL & SUA	**1**	**6.34**	**12.19**	**12.19**
		2	1.43	2.76	14.95
		3	1.27	2.43	17.38
		4	1.16	2.24	19.62
		5	1.15	2.21	21.83
		6	1.13	2.17	24.00
		7	1.11	2.13	26.13
		8	1.09	2.09	28.22
		9	1.06	2.04	30.26
		10	1.06	2.03	32.29
		11	1.05	2.01	34.30
		12	1.03	1.98	36.28
		13	1.03	1.97	38.25
		14	1.02	1.96	40.21
		15	1.00	1.93	42.14

Table L1. Factor Analysis Results for ELA Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
6	ELL	**1**	**6.29**	**12.33**	**12.33**
		2	1.60	3.13	15.46
		3	1.19	2.34	17.80
		4	1.15	2.26	20.06
		5	1.11	2.18	22.24
		6	1.10	2.16	24.40
		7	1.08	2.12	26.52
		8	1.06	2.08	28.60
		9	1.05	2.06	30.66
		10	1.03	2.02	32.68
		11	1.02	2.01	34.69
		12	1.02	1.99	36.68
		13	1.01	1.97	38.65
		14	1.00	1.97	40.62
	SWD	**1**	**7.79**	**15.28**	**15.28**
		2	1.76	3.45	18.73
		3	1.17	2.30	21.03
		4	1.09	2.14	23.17
		5	1.07	2.11	25.28
		6	1.04	2.04	27.32
		7	1.04	2.03	29.35
		8	1.02	1.99	31.34
		9	1.00	1.97	33.31
	SUA	**1**	**8.4**	**16.47**	**16.47**
		2	1.75	3.43	19.90
		3	1.17	2.30	22.20
		4	1.07	2.11	24.31
		5	1.05	2.06	26.37
		6	1.05	2.05	28.42
		7	1.03	2.03	30.45
		8	1.02	2.01	32.46
	SWD & SUA	**1**	**7.74**	**15.18**	**15.18**
		2	1.72	3.38	18.56
		3	1.18	2.31	20.87
		4	1.10	2.16	23.03
		5	1.08	2.12	25.15
		6	1.05	2.07	27.22
		7	1.04	2.04	29.26
		8	1.03	2.01	31.27
		9	1.02	2.00	33.27
		10	1.01	1.98	35.25
		11	1.00	1.96	37.21

Table L1. Factor Analysis Results for ELA Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
6	ELL & SUA	**1**	**6.21**	**12.19**	**12.19**
		2	1.55	3.03	15.22
		3	1.24	2.43	17.65
		4	1.20	2.35	20.00
		5	1.15	2.26	22.26
		6	1.14	2.24	24.50
		7	1.12	2.20	26.70
		8	1.10	2.16	28.86
		9	1.09	2.14	31.00
		10	1.08	2.11	33.11
		11	1.06	2.09	35.20
		12	1.05	2.06	37.26
		13	1.05	2.05	39.31
		14	1.03	2.02	41.33
		15	1.01	1.99	43.32
		16	1.01	1.98	45.30
7	ELL	**1**	**6.35**	**12.21**	**12.21**
		2	1.46	2.81	15.02
		3	1.26	2.43	17.45
		4	1.16	2.24	19.69
		5	1.14	2.20	21.89
		6	1.12	2.15	24.04
		7	1.10	2.12	26.16
		8	1.09	2.09	28.25
		9	1.08	2.07	30.32
		10	1.05	2.01	32.33
		11	1.04	2.00	34.33
		12	1.03	1.98	36.31
		13	1.02	1.97	38.28
		14	1.01	1.95	40.23
		15	1.01	1.94	42.17
		16	1.00	1.93	44.10

Table L1. Factor Analysis Results for ELA Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
7	SWD	**1**	**7.73**	**14.86**	**14.86**
		2	1.56	3.00	17.86
		3	1.24	2.39	20.25
		4	1.15	2.22	22.47
		5	1.14	2.18	24.65
		6	1.06	2.04	26.69
		7	1.05	2.01	28.70
		8	1.04	2.00	30.70
		9	1.02	1.97	32.67
		10	1.01	1.95	34.62
		11	1.01	1.93	36.55
	SUA	**1**	**8.38**	**16.12**	**16.12**
		2	1.58	3.05	19.17
		3	1.25	2.40	21.57
		4	1.12	2.16	23.73
		5	1.09	2.11	25.84
		6	1.06	2.03	27.87
		7	1.04	1.99	29.86
		8	1.03	1.97	31.83
		9	1.01	1.94	33.77
	SWD & SUA	**1**	**7.62**	**14.64**	**14.64**
		2	1.56	3.00	17.64
		3	1.25	2.39	20.03
		4	1.16	2.23	22.26
		5	1.11	2.14	24.40
		6	1.08	2.07	26.47
		7	1.07	2.05	28.52
		8	1.05	2.01	30.53
		9	1.04	1.99	32.52
		10	1.02	1.96	34.48
		11	1.01	1.95	36.43
		12	1.01	1.93	38.36
		13	1.00	1.93	40.29

Table L1. Factor Analysis Results for ELA Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
7	ELL & SUA	**1**	**6.40**	**12.30**	**12.30**
		2	1.42	2.73	15.03
		3	1.30	2.51	17.54
		4	1.24	2.38	19.92
		5	1.22	2.35	22.27
		6	1.15	2.21	24.48
		7	1.14	2.18	26.66
		8	1.12	2.16	28.82
		9	1.11	2.14	30.96
		10	1.09	2.09	33.05
		11	1.07	2.05	35.10
		12	1.06	2.04	37.14
		13	1.05	2.03	39.17
		14	1.04	1.99	41.16
		15	1.03	1.98	43.14
		16	1.03	1.97	45.11
		17	1.02	1.96	47.07
8	ELL	**1**	**5.99**	**11.51**	**11.51**
		2	1.66	3.19	14.70
		3	1.18	2.27	16.97
		4	1.13	2.18	19.15
		5	1.12	2.15	21.30
		6	1.10	2.12	23.42
		7	1.08	2.08	25.50
		8	1.08	2.07	27.57
		9	1.06	2.05	29.62
		10	1.05	2.03	31.65
		11	1.05	2.01	33.66
		12	1.04	1.99	35.65
		13	1.03	1.98	37.63
		14	1.02	1.97	39.60
		15	1.02	1.95	41.55
		16	1.01	1.94	43.49

Table L1. Factor Analysis Results for ELA Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
8	SWD	**1**	**7.66**	**14.72**	**14.72**
		2	1.76	3.39	18.11
		3	1.20	2.32	20.43
		4	1.10	2.12	22.55
		5	1.08	2.08	24.63
		6	1.05	2.02	26.65
		7	1.03	1.98	28.63
		8	1.02	1.97	30.60
		9	1.01	1.95	32.55
		10	1.00	1.92	34.47
	SUA	**1**	**8.34**	**16.04**	**16.04**
		2	1.82	3.49	19.53
		3	1.19	2.28	21.81
		4	1.12	2.15	23.96
		5	1.08	2.07	26.03
		6	1.03	1.98	28.01
		7	1.01	1.95	29.96
		8	1.01	1.95	31.91
		9	1.00	1.93	33.84
	SWD & SUA	**1**	**7.67**	**14.76**	**14.76**
		2	1.82	3.50	18.26
		3	1.23	2.37	20.63
		4	1.11	2.14	22.77
		5	1.09	2.10	24.87
		6	1.05	2.03	26.90
		7	1.04	2.00	28.90
		8	1.03	1.99	30.89
		9	1.02	1.95	32.84
		10	1.01	1.95	34.79

Table L1. Factor Analysis Results for ELA Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
8	ELL & SUA	**1**	**5.96**	**11.46**	**11.46**
		2	1.58	3.05	14.51
		3	1.24	2.38	16.89
		4	1.22	2.34	19.23
		5	1.19	2.28	21.51
		6	1.16	2.23	23.74
		7	1.13	2.18	25.92
		8	1.13	2.16	28.08
		9	1.11	2.13	30.21
		10	1.10	2.12	32.33
		11	1.09	2.09	34.42
		12	1.07	2.06	36.48
		13	1.06	2.04	38.52
		14	1.05	2.02	40.54
		15	1.04	1.99	42.53
		16	1.03	1.98	44.51
		17	1.01	1.94	46.45
		18	1.00	1.92	48.37

Table L2. Factor Analysis Results for Mathematics Tests (Selected Subpopulations)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
3	ELL	**1**	**10.72**	**21.87**	**21.87**
		2	1.83	3.73	25.60
		3	1.27	2.60	28.2
		4	1.17	2.38	30.58
		5	1.06	2.17	32.75
		6	1.01	2.07	34.82
	SWD	**1**	**10.82**	**22.09**	**22.09**
		2	1.77	3.61	25.70
		3	1.36	2.78	28.48
		4	1.18	2.41	30.89
		5	1.04	2.13	33.02
		6	1.02	2.09	35.11
	SUA	**1**	**11.07**	**22.60**	**22.60**
		2	1.77	3.61	26.21
		3	1.32	2.69	28.90
		4	1.18	2.40	31.30
		5	1.05	2.15	33.45
	SWD & SUA	**1**	**9.86**	**20.12**	**20.12**
		2	1.73	3.52	23.64
		3	1.41	2.87	26.51
		4	1.19	2.42	28.93
		5	1.07	2.17	31.10
		6	1.04	2.13	33.23
		7	1.00	2.05	35.28
	ELL & SUA	**1**	**10.42**	**21.27**	**21.27**
		2	1.80	3.68	24.95
		3	1.27	2.59	27.54
		4	1.17	2.39	29.93
		5	1.11	2.26	32.19
		6	1.04	2.12	34.31
4	ELL	**1**	**9.81**	**18.86**	**18.86**
		2	2.02	3.89	22.75
		3	1.31	2.51	25.26
		4	1.16	2.24	27.50
		5	1.12	2.15	29.65
		6	1.06	2.05	31.70
		7	1.04	2.00	33.70
		8	1.00	1.93	35.63

Table L2. Factor Analysis Results for Mathematics Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
4	SWD	**1**	**10.03**	**19.29**	**19.29**
		2	1.86	3.57	22.86
		3	1.34	2.57	25.43
		4	1.14	2.18	27.61
		5	1.12	2.15	29.76
		6	1.07	2.05	31.81
		7	1.05	2.02	33.83
		8	1.01	1.93	35.76
	SUA	**1**	**9.90**	**19.03**	**19.03**
		2	1.84	3.54	22.57
		3	1.35	2.60	25.17
		4	1.14	2.18	27.35
		5	1.12	2.16	29.51
		6	1.09	2.09	31.60
		7	1.06	2.04	33.64
		8	1.01	1.94	35.58
	SWD & SUA	**1**	**9.10**	**17.50**	**17.50**
		2	1.75	3.37	20.87
		3	1.36	2.62	23.49
		4	1.17	2.24	25.73
		5	1.15	2.21	27.94
		6	1.10	2.12	30.06
		7	1.07	2.06	32.12
		8	1.03	1.99	34.11
		9	1.01	1.94	36.05
	ELL & SUA	**1**	**9.02**	**17.35**	**17.35**
		2	1.93	3.72	21.07
		3	1.36	2.61	23.68
		4	1.18	2.27	25.95
		5	1.14	2.18	28.13
		6	1.09	2.10	30.23
		7	1.07	2.06	32.29
		8	1.06	2.04	34.33
		9	1.02	1.96	36.29
		10	1.02	1.96	38.25

Table L2. Factor Analysis Results for Mathematics Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
5	ELL	**1**	**9.51**	**18.29**	**18.29**
		2	2.00	3.84	22.13
		3	1.24	2.38	24.51
		4	1.10	2.12	26.63
		5	1.08	2.08	28.71
		6	1.05	2.02	30.73
		7	1.03	1.98	32.71
		8	1.01	1.93	34.64
	SWD	**1**	**9.37**	**18.01**	**18.01**
		2	1.84	3.53	21.54
		3	1.28	2.47	24.01
		4	1.12	2.15	26.16
		5	1.09	2.10	28.26
		6	1.01	1.94	30.20
		7	1.01	1.94	32.14
	SUA	**1**	**9.87**	**18.98**	**18.98**
		2	1.90	3.65	22.63
		3	1.27	2.44	25.07
		4	1.10	2.11	27.18
		5	1.08	2.08	29.26
		6	1.01	1.94	31.20
	SWD & SUA	**1**	**8.71**	**16.75**	**16.75**
		2	1.74	3.35	20.10
		3	1.29	2.49	22.59
		4	1.14	2.18	24.77
		5	1.12	2.15	26.92
		6	1.03	1.98	28.90
		7	1.02	1.97	30.87
		8	1.01	1.95	32.82
	ELL & SUA	**1**	**9.70**	**18.66**	**18.66**
		2	1.99	3.82	22.48
		3	1.23	2.36	24.84
		4	1.09	2.10	26.94
		5	1.08	2.08	29.02
		6	1.06	2.03	31.05
		7	1.02	1.96	33.01
		8	1.02	1.96	34.97
		9	1.00	1.93	36.90

Table L2. Factor Analysis Results for Mathematics Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
6	ELL	**1**	**10.80**	**18.63**	**18.63**
		2	2.00	3.45	22.08
		3	1.31	2.26	24.34
		4	1.08	1.86	26.20
		5	1.07	1.84	28.04
		6	1.03	1.78	29.82
		7	1.01	1.75	31.57
		8	1.00	1.73	33.30
	SWD	**1**	**9.57**	**16.50**	**16.50**
		2	1.84	3.16	19.66
		3	1.39	2.40	22.06
		4	1.08	1.86	23.92
		5	1.07	1.84	25.76
		6	1.05	1.81	27.57
		7	1.02	1.76	29.33
		8	1.02	1.75	31.08
		9	1.01	1.73	32.81
	SUA	**1**	**10.06**	**17.34**	**17.34**
		2	1.91	3.29	20.63
		3	1.38	2.38	23.01
		4	1.07	1.85	24.86
		5	1.06	1.83	26.69
		6	1.03	1.78	28.47
		7	1.02	1.75	30.22
		8	1.00	1.73	31.95
	SWD & SUA	**1**	**8.99**	**15.51**	**15.51**
		2	1.78	3.06	18.57
		3	1.40	2.42	20.99
		4	1.08	1.87	22.86
		5	1.08	1.87	24.73
		6	1.07	1.84	26.57
		7	1.04	1.80	28.37
		8	1.03	1.77	30.14
		9	1.02	1.76	31.90
		10	1.00	1.73	33.63

Table L2. Factor Analysis Results for Mathematics Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
6	ELL & SUA	**1**	**9.84**	**16.97**	**16.97**
		2	1.95	3.37	20.34
		3	1.32	2.27	22.61
		4	1.11	1.91	24.52
		5	1.10	1.90	26.42
		6	1.07	1.85	28.27
		7	1.06	1.82	30.09
		8	1.05	1.80	31.89
		9	1.03	1.78	33.67
		10	1.02	1.76	35.43
		11	1.02	1.75	37.18
7	ELL	**1**	**9.50**	**16.38**	**16.38**
		2	1.57	2.71	19.09
		3	1.31	2.27	21.36
		4	1.18	2.04	23.40
		5	1.13	1.94	25.34
		6	1.11	1.91	27.25
		7	1.06	1.83	29.08
		8	1.04	1.79	30.87
		9	1.03	1.78	32.65
		10	1.02	1.76	34.41
		11	1.01	1.75	36.16
		12	1.01	1.74	37.90
	SWD	**1**	**8.23**	**14.19**	**14.19**
		2	1.55	2.68	16.87
		3	1.35	2.33	19.20
		4	1.16	2.00	21.20
		5	1.11	1.91	23.11
		6	1.07	1.84	24.95
		7	1.06	1.83	26.78
		8	1.04	1.80	28.58
		9	1.04	1.79	30.37
		10	1.02	1.77	32.14
		11	1.02	1.76	33.90
		12	1.02	1.75	35.65
		13	1.01	1.74	37.39

Table L2. Factor Analysis Results for Mathematics Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
7	SUA	**1**	**9.24**	**15.93**	**15.93**
		2	1.60	2.75	18.68
		3	1.37	2.37	21.05
		4	1.15	1.98	23.03
		5	1.09	1.89	24.92
		6	1.07	1.85	26.77
		7	1.05	1.81	28.58
		8	1.03	1.78	30.36
		9	1.02	1.76	32.12
		10	1.01	1.74	33.86
		11	1.00	1.73	35.59
	SWD & SUA	**1**	**7.72**	**13.31**	**13.31**
		2	1.52	2.62	15.93
		3	1.39	2.40	18.33
		4	1.17	2.02	20.35
		5	1.12	1.93	22.28
		6	1.10	1.89	24.17
		7	1.07	1.85	26.02
		8	1.06	1.82	27.84
		9	1.05	1.81	29.65
		10	1.04	1.79	31.44
		11	1.03	1.78	33.22
		12	1.03	1.77	34.99
		13	1.01	1.75	36.74
		14	1.01	1.73	38.47
	ELL & SUA	**1**	**9.78**	**16.86**	**16.86**
		2	1.62	2.79	19.65
		3	1.31	2.25	21.90
		4	1.19	2.06	23.96
		5	1.17	2.01	25.97
		6	1.14	1.96	27.93
		7	1.10	1.90	29.83
		8	1.09	1.88	31.71
		9	1.08	1.86	33.57
		10	1.07	1.85	35.42
		11	1.05	1.81	37.23
		12	1.03	1.78	39.01
		13	1.02	1.76	40.77
		14	1.01	1.74	42.51

Table L2. Factor Analysis Results for Mathematics Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
8	ELL	**1**	**10.13**	**17.46**	**17.46**
		2	1.82	3.14	20.60
		3	1.22	2.10	22.70
		4	1.18	2.03	24.73
		5	1.09	1.88	26.61
		6	1.07	1.84	28.45
		7	1.02	1.76	30.21
		8	1.01	1.75	31.96
		9	1.01	1.73	33.69
	SWD	**1**	**8.40**	**14.48**	**14.48**
		2	1.78	3.07	17.55
		3	1.28	2.20	19.75
		4	1.09	1.88	21.63
		5	1.06	1.82	23.45
		6	1.05	1.81	25.26
		7	1.05	1.80	27.06
		8	1.03	1.78	28.84
		9	1.02	1.75	30.59
		10	1.01	1.74	32.33
	SUA	**1**	**9.54**	**16.45**	**16.45**
		2	1.93	3.32	19.77
		3	1.28	2.21	21.98
		4	1.08	1.86	23.84
		5	1.05	1.81	25.65
		6	1.04	1.79	27.44
		7	1.02	1.77	29.21
		8	1.00	1.73	30.94
		9	1.00	1.73	32.67
	SWD & SUA	**1**	**8.17**	**14.09**	**14.09**
		2	1.75	3.01	17.10
		3	1.30	2.24	19.34
		4	1.11	1.91	21.25
		5	1.07	1.84	23.09
		6	1.06	1.83	24.92
		7	1.06	1.83	26.75
		8	1.04	1.79	28.54
		9	1.02	1.77	30.31
		10	1.02	1.76	32.07
		11	1.01	1.74	33.81

Table L2. Factor Analysis Results for Mathematics Tests (Selected Subpopulations, cont.)

		Initial Eigenvalues			
Grade	Subgroups	Component	Total	% of Variance	Cumulative %
8	ELL & SUA	**1**	**10.82**	**18.65**	**18.65**
		2	1.95	3.35	22.00
		3	1.20	2.07	24.07
		4	1.14	1.96	26.03
		5	1.11	1.92	27.95
		6	1.09	1.88	29.83
		7	1.07	1.85	31.68
		8	1.04	1.79	33.47
		9	1.04	1.78	35.25
		10	1.01	1.74	36.99
		11	1.00	1.73	38.72

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. ELA Classical Item Analysis, Grade 3

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	197,294	0.73	0.03	0.48
2	MC	197,065	0.31	0.10	0.27
3	MC	196,974	0.37	0.13	0.30
4	MC	196,960	0.70	0.12	0.52
5	MC	196,959	0.63	0.15	0.38
6	MC	196,981	0.49	0.13	0.22
7	MC	196,653	0.45	0.29	0.35
8	MC	196,575	0.60	0.32	0.43
9	MC	196,375	0.45	0.43	0.35
10	MC	196,445	0.64	0.42	0.36
11	MC	196,407	0.30	0.45	0.19
12	MC	196,221	0.36	0.51	0.17
13	MC	195,289	0.57	1.02	0.36
14	MC	195,093	0.64	1.12	0.48
15	MC	194,928	0.65	1.19	0.50
16	MC	194,601	0.64	1.33	0.53
17	MC	194,171	0.41	1.55	0.24
18	MC	193,974	0.71	1.66	0.50
19	MC	192,397	0.37	2.49	0.35
20	MC	191,782	0.60	2.77	0.45
21	MC	191,135	0.34	3.09	0.33
22	MC	190,778	0.50	3.29	0.43
23	MC	190,444	0.46	3.48	0.43
24	MC	189,902	0.68	3.78	0.48
25	MC	197,233	0.68	0.07	0.49
26	MC	197,160	0.81	0.08	0.45
27	MC	196,985	0.66	0.13	0.36
28	MC	196,997	0.66	0.13	0.45
29	MC	197,059	0.33	0.13	0.17
30	MC	196,912	0.62	0.17	0.49
31	MC	196,684	0.77	0.35	0.49

Table M1. ELA Classical Item Analysis, Grade 3 (cont.)

Item	Item Type	N-count	P-value	% Omit	Pbis Key
32	CR	195,276	0.42	1.08	
33	CR	194,117	0.52	1.67	
34	CR	192,787	0.49	2.34	
35	CR	190,488	0.43	3.50	
36	CR	196,563	0.56	0.43	
37	CR	195,825	0.50	0.80	
38	CR	194,601	0.57	1.42	
39	CR	194,531	0.55	1.46	
40	CR	194,544	0.59	1.45	
41	CR	192,041	0.43	2.72	

Table M2. ELA Classical Item Analysis, Grade 4

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	204,448	0.64	0.04	0.35
2	MC	204,387	0.78	0.06	0.43
3	MC	204,372	0.77	0.06	0.38
4	MC	204,115	0.50	0.15	0.28
5	MC	204,258	0.70	0.09	0.42
6	MC	204,293	0.65	0.10	0.45
7	MC	204,146	0.79	0.15	0.39
8	MC	204,065	0.63	0.21	0.46
9	MC	204,013	0.66	0.22	0.42
10	MC	203,979	0.51	0.25	0.45
11	MC	203,970	0.81	0.27	0.53
12	MC	203,870	0.76	0.31	0.49
13	MC	203,287	0.53	0.58	0.41
14	MC	203,171	0.63	0.63	0.44
15	MC	203,001	0.45	0.72	0.38
16	MC	202,727	0.55	0.85	0.45
17	MC	202,490	0.73	0.96	0.33
18	MC	202,277	0.67	1.09	0.45
19	MC	201,124	0.46	1.66	0.40
20	MC	200,700	0.56	1.85	0.48
21	MC	200,233	0.51	2.06	0.42
22	MC	199,949	0.56	2.20	0.41
23	MC	199,586	0.52	2.41	0.37
24	MC	199,240	0.42	2.59	0.37
25	MC	204,447	0.64	0.04	0.30
26	MC	204,341	0.59	0.07	0.38

Table M2. ELA Classical Item Analysis, Grade 4 (cont.)

Item	Item Type	N-count	P-value	% Omit	Pbis Key
27	MC	204,315	0.43	0.07	0.18
28	MC	204,324	0.59	0.09	0.23
29	MC	204,367	0.71	0.06	0.44
30	MC	204,273	0.41	0.11	0.34
31	MC	204,134	0.60	0.19	0.50
32	CR	203,497	0.55	0.52	
33	CR	202,374	0.57	1.07	
34	CR	202,870	0.62	0.82	
35	CR	200,821	0.45	1.83	
36	CR	203,807	0.50	0.37	
37	CR	203,514	0.61	0.51	
38	CR	203,548	0.65	0.49	
39	CR	203,023	0.58	0.75	
40	CR	201,787	0.39	1.35	
41	CR	201,166	0.48	1.66	

Table M3. ELA Classical Item Analysis, Grade 5

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	194,817	0.73	0.03	0.28
2	MC	194,742	0.77	0.06	0.46
3	MC	194,665	0.62	0.07	0.41
4	MC	194,655	0.44	0.08	0.28
5	MC	194,641	0.54	0.10	0.44
6	MC	194,616	0.68	0.11	0.35
7	MC	194,598	0.43	0.12	0.27
8	MC	194,594	0.70	0.12	0.48
9	MC	194,567	0.42	0.14	0.15
10	MC	194,597	0.64	0.11	0.38
11	MC	194,534	0.33	0.15	0.24
12	MC	194,447	0.46	0.18	0.20
13	MC	194,495	0.46	0.17	0.31
14	MC	194,560	0.53	0.15	0.42
15	MC	194,422	0.57	0.22	0.38
16	MC	194,250	0.50	0.30	0.42
17	MC	194,252	0.52	0.30	0.45
18	MC	194,152	0.54	0.34	0.36
19	MC	194,015	0.48	0.40	0.39
20	MC	193,990	0.64	0.42	0.50
21	MC	193,927	0.60	0.45	0.46
22	MC	191,498	0.67	1.71	0.46

Table M3. ELA Classical Item Analysis, Grade 5 (cont.)

Item	Item Type	N-count	P-value	% Omit	Pbis Key
23	MC	191,115	0.57	1.88	0.39
24	MC	190,669	0.56	2.12	0.43
25	MC	190,372	0.27	2.27	0.20
26	MC	190,200	0.66	2.37	0.52
27	MC	189,965	0.55	2.50	0.34
28	MC	189,836	0.58	2.58	0.43
29	MC	194,795	0.75	0.04	0.25
30	MC	194,777	0.56	0.04	0.37
31	MC	194,747	0.67	0.05	0.36
32	MC	194,700	0.70	0.06	0.32
33	MC	194,651	0.69	0.06	0.28
34	MC	194,676	0.53	0.08	0.17
35	MC	194,718	0.76	0.06	0.32
36	MC	194,591	0.59	0.12	0.38
37	MC	194,571	0.65	0.12	0.44
38	MC	194,485	0.71	0.17	0.45
39	MC	194,482	0.65	0.16	0.42
40	MC	194,449	0.43	0.17	0.39
41	MC	194,479	0.42	0.18	0.25
42	MC	194,451	0.70	0.21	0.42
43	CR	194,145	0.72	0.38	
44	CR	193,738	0.65	0.58	
45	CR	191,241	0.53	1.87	
46	CR	189,614	0.45	2.70	
47	CR	194,175	0.62	0.36	
48	CR	193,943	0.49	0.48	
49	CR	190,676	0.54	2.16	
50	CR	193,751	0.55	0.58	
51	CR	192,600	0.52	1.17	
52	CR	191,721	0.49	1.62	

Table M4. ELA Classical Item Analysis, Grade 6

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	201,872	0.60	0.07	0.30
2	MC	201,843	0.60	0.08	0.32
3	MC	201,845	0.74	0.07	0.45
4	MC	201,800	0.70	0.08	0.37
5	MC	201,868	0.80	0.06	0.47
6	MC	201,824	0.90	0.05	0.34
7	MC	201,812	0.77	0.09	0.46
8	MC	201,716	0.64	0.13	0.35
9	MC	201,704	0.61	0.13	0.49
10	MC	201,532	0.35	0.22	0.34
11	MC	201,706	0.75	0.14	0.49
12	MC	201,649	0.64	0.16	0.40
13	MC	201,531	0.49	0.21	0.40
14	MC	201,434	0.45	0.27	0.24
15	MC	201,504	0.39	0.23	0.33
16	MC	201,655	0.75	0.16	0.49
17	MC	201,555	0.81	0.19	0.52
18	MC	201,439	0.57	0.27	0.37
19	MC	201,477	0.78	0.24	0.44
20	MC	201,396	0.57	0.28	0.33
21	MC	199,840	0.42	1.05	0.31
22	MC	199,471	0.56	1.22	0.37
23	MC	199,311	0.53	1.30	0.32
24	MC	199,082	0.42	1.41	0.28
25	MC	199,009	0.66	1.45	0.54
26	MC	198,522	0.41	1.70	0.39
27	MC	198,559	0.56	1.69	0.47
28	MC	201,903	0.28	0.05	0.27
29	MC	201,789	0.35	0.10	0.24
30	MC	201,739	0.53	0.12	0.42
31	MC	201,759	0.77	0.11	0.44
32	MC	201,773	0.65	0.10	0.40
33	MC	201,815	0.66	0.09	0.36
34	MC	201,518	0.61	0.23	0.38
35	MC	201,685	0.61	0.14	0.41
36	MC	201,602	0.54	0.18	0.39
37	MC	201,655	0.79	0.16	0.41
38	MC	201,595	0.78	0.18	0.41
39	MC	201,531	0.61	0.21	0.40
40	MC	201,408	0.54	0.28	0.39
41	MC	201,416	0.70	0.28	0.46

Table M4. ELA Classical Item Analysis, Grade 6 cont.)

Item	Item Type	N-count	P-value	% Omit	Pbis Key
42	CR	200,403	0.65	0.80	
43	CR	199,795	0.62	1.10	
44	CR	196,345	0.64	2.81	
45	CR	192,074	0.48	4.92	
46	CR	201,318	0.64	0.35	
47	CR	201,075	0.68	0.47	
48	CR	201,289	0.56	0.36	
49	CR	200,232	0.72	0.88	
50	CR	198,850	0.64	1.57	
51	CR	199,364	0.53	1.31	

Table M5. ELA Classical Item Analysis, Grade 7

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	200,915	0.89	0.03	0.33
2	MC	200,856	0.68	0.06	0.32
3	MC	200,681	0.62	0.13	0.25
4	MC	200,854	0.87	0.05	0.41
5	MC	200,691	0.55	0.12	0.32
6	MC	200,774	0.56	0.09	0.19
7	MC	200,706	0.51	0.12	0.35
8	MC	200,715	0.67	0.12	0.33
9	MC	200,645	0.60	0.15	0.27
10	MC	200,635	0.38	0.15	0.36
11	MC	200,532	0.48	0.21	0.37
12	MC	200,612	0.49	0.16	0.14
13	MC	200,615	0.63	0.16	0.37
14	MC	200,565	0.59	0.19	0.32
15	MC	200,406	0.47	0.28	0.34
16	MC	200,530	0.68	0.21	0.54
17	MC	200,403	0.52	0.26	0.39
18	MC	200,317	0.50	0.30	0.32
19	MC	200,281	0.70	0.33	0.37
20	MC	200,295	0.76	0.32	0.44
21	MC	198,481	0.79	1.22	0.47
22	MC	198,254	0.51	1.32	0.46
23	MC	198,002	0.48	1.44	0.43
24	MC	197,893	0.77	1.50	0.50
25	MC	197,505	0.40	1.70	0.30
26	MC	197,268	0.76	1.83	0.51
27	MC	197,008	0.49	1.96	0.36

Table M5. ELA Classical Item Analysis, Grade 7 (cont.)

Item	Item Type	N-count	P-value	% Omit	Pbis Key
28	MC	196,751	0.47	2.10	0.31
29	MC	200,842	0.83	0.07	0.46
30	MC	200,760	0.64	0.10	0.23
31	MC	200,556	0.61	0.19	0.27
32	MC	200,659	0.79	0.14	0.39
33	MC	200,721	0.77	0.10	0.49
34	MC	200,640	0.46	0.15	0.17
35	MC	200,707	0.71	0.13	0.41
36	MC	200,626	0.69	0.17	0.44
37	MC	200,503	0.30	0.21	0.28
38	MC	200,578	0.80	0.18	0.43
39	MC	200,449	0.59	0.24	0.32
40	MC	200,380	0.39	0.28	0.34
41	MC	200,455	0.74	0.24	0.40
42	MC	200,336	0.48	0.31	0.46
43	CR	199,334	0.61	0.82	
44	CR	198,406	0.65	1.28	
45	CR	196,254	0.67	2.35	
46	CR	194,214	0.56	3.37	
47	CR	199,832	0.54	0.57	
48	CR	200,126	0.67	0.43	
49	CR	198,438	0.73	1.27	
50	CR	199,778	0.72	0.60	
51	CR	197,328	0.65	1.82	
52	CR	197,626	0.58	1.67	

Table M6. ELA Classical Item Analysis, Grade 8

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	207,339	0.90	0.06	0.31
2	MC	207,196	0.66	0.12	0.35
3	MC	207,218	0.74	0.11	0.32
4	MC	207,254	0.69	0.09	0.43
5	MC	207,202	0.55	0.12	0.44
6	MC	207,241	0.38	0.10	0.25
7	MC	207,056	0.63	0.19	0.38
8	MC	207,224	0.45	0.10	0.34
9	MC	207,142	0.69	0.14	0.47
10	MC	207,164	0.65	0.14	0.30
11	MC	207,082	0.60	0.17	0.28
12	MC	207,109	0.63	0.15	0.33

Table M6. ELA Classical Item Analysis, Grade 8 (cont.)

Item	Item Type	N-count	P-value	% Omit	Pbis Key
13	MC	207,088	0.73	0.17	0.33
14	MC	207,079	0.65	0.18	0.33
15	MC	207,030	0.60	0.20	0.42
16	MC	206,833	0.46	0.30	0.41
17	MC	207,032	0.68	0.20	0.39
18	MC	206,907	0.76	0.25	0.24
19	MC	206,972	0.66	0.23	0.51
20	MC	206,817	0.66	0.30	0.38
21	MC	206,814	0.71	0.31	0.43
22	MC	204,841	0.44	1.24	0.37
23	MC	204,914	0.60	1.21	0.28
24	MC	204,561	0.40	1.36	0.33
25	MC	204,491	0.50	1.41	0.37
26	MC	204,653	0.85	1.33	0.39
27	MC	204,261	0.49	1.53	0.30
28	MC	204,108	0.69	1.61	0.49
29	MC	207,121	0.32	0.17	0.22
30	MC	207,188	0.77	0.12	0.57
31	MC	207,025	0.49	0.21	0.29
32	MC	207,034	0.56	0.19	0.36
33	MC	206,828	0.34	0.29	0.26
34	MC	207,157	0.62	0.13	0.32
35	MC	207,054	0.33	0.19	0.30
36	MC	206,924	0.57	0.25	0.35
37	MC	206,844	0.74	0.29	0.32
38	MC	206,915	0.74	0.25	0.32
39	MC	206,823	0.71	0.29	0.45
40	MC	206,735	0.58	0.34	0.39
41	MC	206,672	0.45	0.38	0.30
42	MC	206,644	0.70	0.39	0.44
43	CR	206,150	0.77	0.64	
44	CR	201,615	0.54	2.82	
45	CR	199,357	0.68	3.91	
46	CR	197,080	0.59	5.01	
47	CR	206,647	0.75	0.40	
48	CR	206,136	0.72	0.65	
49	CR	206,396	0.79	0.52	
50	CR	203,211	0.67	2.06	
51	CR	202,750	0.71	2.28	
52	CR	204,372	0.66	1.50	

Table M7. Mathematics Classical Item Analysis, Grade 3

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	202,003	0.89	0.03	0.30
2	MC	201,907	0.88	0.05	0.34
3	MC	201,570	0.77	0.10	0.35
4	MC	201,784	0.88	0.08	0.36
5	MC	201,687	0.78	0.10	0.48
6	MC	201,707	0.71	0.09	0.58
7	MC	201,552	0.53	0.17	0.54
8	MC	201,377	0.37	0.29	0.32
9	MC	201,586	0.55	0.16	0.42
10	MC	201,592	0.60	0.16	0.53
11	MC	201,576	0.74	0.16	0.51
12	MC	201,583	0.63	0.19	0.49
13	MC	201,728	0.54	0.14	0.42
14	MC	201,693	0.57	0.16	0.54
15	MC	201,513	0.89	0.19	0.33
16	MC	201,264	0.54	0.34	0.48
17	MC	201,305	0.71	0.32	0.52
18	MC	201,282	0.57	0.36	0.62
19	MC	201,084	0.55	0.41	0.54
20	MC	200,896	0.72	0.52	0.61
21	MC	202,017	0.89	0.03	0.28
22	MC	201,912	0.83	0.06	0.44
23	MC	201,533	0.79	0.10	0.46
24	MC	201,774	0.69	0.10	0.51
25	MC	201,639	0.76	0.13	0.48
26	MC	201,460	0.44	0.22	0.24
27	MC	201,714	0.79	0.12	0.55
28	MC	201,845	0.63	0.09	0.44
29	MC	201,559	0.69	0.17	0.57
30	MC	201,501	0.71	0.18	0.57
31	MC	201,347	0.42	0.29	0.20
32	MC	201,637	0.56	0.15	0.45
33	MC	201,674	0.56	0.18	0.43
34	MC	201,670	0.52	0.19	0.45
35	MC	201,630	0.67	0.17	0.47
36	MC	201,398	0.62	0.25	0.46
37	MC	201,202	0.40	0.42	0.54
38	MC	201,351	0.45	0.34	0.57
39	MC	201,176	0.47	0.39	0.52
40	MC	200,976	0.37	0.49	0.30
41	MC	200,921	0.79	0.51	0.52

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

Item	Item Type	N-count	P-value	% Omit	Pbis Key
42	CR	201,860	0.63	0.12	
43	CR	201,791	0.74	0.15	
44	CR	201,692	0.67	0.20	
45	CR	200,755	0.31	0.66	
46	CR	201,625	0.58	0.23	
47	CR	201,433	0.44	0.33	
48	CR	201,592	0.52	0.25	
49	CR	200,611	0.34	0.73	

Table M8. Mathematics Classical Item Analysis, Grade 4

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	210,195	0.96	0.02	0.19
2	MC	210,147	0.92	0.04	0.28
3	MC	210,062	0.92	0.05	0.35
4	MC	210,072	0.88	0.05	0.37
5	MC	209,887	0.91	0.06	0.32
6	MC	209,900	0.91	0.08	0.34
7	MC	209,965	0.41	0.09	0.39
8	MC	209,887	0.46	0.13	0.41
9	MC	209,869	0.62	0.12	0.44
10	MC	210,009	0.66	0.07	0.57
11	MC	209,856	0.80	0.14	0.48
12	MC	209,915	0.87	0.11	0.35
13	MC	209,956	0.75	0.12	0.47
14	MC	210,037	0.59	0.08	0.56
15	MC	209,828	0.59	0.16	0.49
16	MC	209,880	0.57	0.15	0.54
17	MC	209,581	0.60	0.28	0.32
18	MC	209,716	0.57	0.22	0.38
19	MC	209,677	0.60	0.22	0.47
20	MC	209,174	0.61	0.48	0.54
21	MC	210,140	0.84	0.05	0.40
22	MC	210,098	0.80	0.05	0.47
23	MC	210,031	0.87	0.07	0.33
24	MC	209,929	0.77	0.11	0.54
25	MC	209,882	0.57	0.14	0.55
26	MC	209,971	0.50	0.08	0.44
27	MC	209,870	0.51	0.14	0.43
28	MC	209,832	0.49	0.17	0.52

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

Item	Item Type	N-count	P-value	% Omit	Pbis Key
29	MC	209,962	0.63	0.09	0.55
30	MC	209,782	0.36	0.14	0.45
31	MC	208,886	0.39	0.61	0.26
32	MC	209,995	0.64	0.09	0.40
33	MC	210,010	0.72	0.10	0.53
34	MC	209,811	0.49	0.19	0.48
35	MC	209,980	0.68	0.11	0.39
36	MC	209,685	0.54	0.21	0.52
37	MC	209,685	0.47	0.24	0.42
38	MC	209,595	0.40	0.28	0.49
39	MC	208,416	0.37	0.81	0.15
40	MC	209,087	0.51	0.47	0.38
41	MC	208,995	0.37	0.52	0.48
42	MC	208,299	0.25	0.87	0.27
43	CR	209,829	0.69	0.20	
44	CR	209,775	0.34	0.23	
45	CR	209,835	0.57	0.20	
46	CR	209,318	0.27	0.45	
47	CR	209,647	0.27	0.29	
48	CR	209,342	0.36	0.44	
49	CR	209,673	0.68	0.28	
50	CR	209,585	0.47	0.32	
51	CR	209,320	0.38	0.45	
52	CR	208,028	0.30	1.06	

Table M9. Mathematics Classical Item Analysis, Grade 5

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	199,969	0.95	0.03	0.23
2	MC	199,909	0.89	0.05	0.23
3	MC	199,717	0.52	0.12	0.43
4	MC	199,795	0.66	0.08	0.53
5	MC	199,830	0.67	0.06	0.39
6	MC	199,712	0.57	0.12	0.43
7	MC	199,614	0.43	0.16	0.49
8	MC	199,802	0.67	0.09	0.49
9	MC	199,533	0.63	0.22	0.48
10	MC	199,759	0.60	0.09	0.46
11	MC	199,357	0.26	0.30	0.09
12	MC	199,807	0.76	0.08	0.52
13	MC	199,768	0.56	0.12	0.39
14	MC	199,823	0.65	0.09	0.50

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

Item	Item Type	N-count	P-value	% Omit	Pbis Key
15	MC	199,670	0.63	0.15	0.54
16	MC	199,457	0.76	0.25	0.47
17	MC	199,587	0.73	0.16	0.47
18	MC	199,663	0.60	0.16	0.42
19	MC	199,233	0.37	0.36	0.47
20	MC	199,487	0.42	0.23	0.35
21	MC	199,907	0.86	0.05	0.43
22	MC	199,871	0.70	0.07	0.40
23	MC	199,922	0.89	0.04	0.35
24	MC	199,722	0.85	0.10	0.42
25	MC	199,888	0.73	0.05	0.40
26	MC	199,486	0.44	0.26	0.45
27	MC	199,887	0.79	0.05	0.47
28	MC	199,870	0.90	0.07	0.42
29	MC	199,799	0.59	0.09	0.34
30	MC	199,610	0.61	0.18	0.54
31	MC	199,822	0.82	0.07	0.50
32	MC	199,636	0.57	0.18	0.29
33	MC	199,702	0.40	0.15	0.52
34	MC	199,846	0.55	0.08	0.22
35	MC	199,818	0.56	0.08	0.62
36	MC	199,674	0.31	0.14	0.29
37	MC	199,696	0.50	0.15	0.44
38	MC	199,758	0.38	0.12	0.37
39	MC	199,743	0.47	0.11	0.35
40	MC	199,323	0.33	0.31	0.40
41	MC	199,557	0.87	0.14	0.37
42	MC	199,605	0.73	0.17	0.49
43	CR	199,629	0.60	0.20	
44	CR	199,553	0.54	0.24	
45	CR	198,932	0.47	0.55	
46	CR	199,124	0.26	0.46	
47	CR	196,792	0.25	1.62	
48	CR	199,256	0.34	0.39	
49	CR	199,160	0.44	0.44	
50	CR	199,142	0.49	0.45	
51	CR	198,747	0.31	0.64	
52	CR	198,432	0.20	0.80	

Table M10. Mathematics Classical Item Analysis, Grade 6

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	207,122	0.81	0.07	0.41
2	MC	207,055	0.62	0.09	0.38
3	MC	207,127	0.73	0.05	0.50
4	MC	207,113	0.77	0.04	0.49
5	MC	207,070	0.88	0.06	0.46
6	MC	206,931	0.73	0.12	0.39
7	MC	206,902	0.58	0.15	0.45
8	MC	207,014	0.80	0.08	0.37
9	MC	206,775	0.55	0.22	0.34
10	MC	206,717	0.43	0.23	0.27
11	MC	206,726	0.44	0.23	0.32
12	MC	206,948	0.70	0.12	0.42
13	MC	207,074	0.52	0.08	0.38
14	MC	207,026	0.41	0.09	0.42
15	MC	206,633	0.42	0.28	0.59
16	MC	207,087	0.84	0.08	0.37
17	MC	206,934	0.61	0.12	0.34
18	MC	206,851	0.58	0.16	0.36
19	MC	206,810	0.60	0.18	0.43
20	MC	206,641	0.55	0.28	0.40
21	MC	206,852	0.69	0.17	0.51
22	MC	206,866	0.49	0.16	0.44
23	MC	206,151	0.41	0.50	0.42
24	MC	206,671	0.44	0.24	0.46
25	MC	207,176	0.89	0.04	0.37
26	MC	207,169	0.86	0.02	0.35
27	MC	206,973	0.76	0.13	0.40
28	MC	207,130	0.77	0.04	0.46
29	MC	207,040	0.72	0.08	0.49
30	MC	206,762	0.38	0.21	0.33
31	MC	207,077	0.49	0.07	0.50
32	MC	207,045	0.80	0.09	0.46
33	MC	206,825	0.49	0.18	0.59
34	MC	206,869	0.53	0.16	0.31
35	MC	206,238	0.47	0.07	0.50
36	MC	206,768	0.37	0.21	0.53
37	MC	207,106	0.72	0.06	0.48
38	MC	206,881	0.54	0.15	0.46
39	MC	207,026	0.67	0.10	0.46

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

Item	Item Type	N-count	P-value	% Omit	Pbis Key
40	MC	207,115	0.49	0.07	0.49
41	MC	206,959	0.74	0.12	0.46
42	MC	206,771	0.43	0.21	0.36
43	MC	206,763	0.24	0.22	0.27
44	MC	206,881	0.63	0.17	0.46
45	MC	206,905	0.56	0.16	0.56
46	MC	207,025	0.38	0.10	0.45
47	MC	206,454	0.41	0.35	0.32
48	MC	206,676	0.62	0.24	0.57
49	CR	206,894	0.78	0.18	
50	CR	206,246	0.44	0.49	
51	CR	206,750	0.50	0.25	
52	CR	205,428	0.36	0.89	
53	CR	206,416	0.40	0.41	
54	CR	205,258	0.34	0.97	
55	CR	206,547	0.81	0.35	
56	CR	206,060	0.56	0.58	
57	CR	205,062	0.47	1.07	
58	CR	205,749	0.30	0.73	

Table M11. Mathematics Classical Item Analysis, Grade 7

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	204,188	0.74	0.12	0.38
2	MC	204,251	0.72	0.08	0.43
3	MC	204,209	0.76	0.09	0.32
4	MC	204,058	0.67	0.17	0.42
5	MC	204,146	0.68	0.11	0.52
6	MC	204,078	0.41	0.15	0.23
7	MC	204,042	0.61	0.17	0.47
8	MC	204,043	0.35	0.17	0.26
9	MC	204,133	0.48	0.14	0.31
10	MC	204,156	0.46	0.12	0.52
11	MC	204,111	0.55	0.14	0.38
12	MC	204,150	0.27	0.13	0.37
13	MC	203,723	0.47	0.33	0.32
14	MC	204,051	0.52	0.16	0.36
15	MC	203,808	0.38	0.29	0.54
16	MC	204,196	0.34	0.11	0.41
17	MC	204,010	0.49	0.18	0.55
18	MC	203,910	0.43	0.22	0.52

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

Item	Item Type	N-count	P-value	% Omit	Pbis Key
19	MC	203,611	0.48	0.37	0.36
20	MC	203,689	0.39	0.34	0.35
21	MC	203,684	0.31	0.34	0.14
22	MC	203,734	0.36	0.32	0.12
23	MC	203,530	0.56	0.41	0.51
24	MC	203,657	0.52	0.36	0.37
25	MC	204,331	0.79	0.06	0.48
26	MC	204,299	0.69	0.06	0.39
27	MC	204,186	0.63	0.11	0.52
28	MC	204,304	0.91	0.06	0.29
29	MC	204,055	0.45	0.17	0.30
30	MC	204,114	0.23	0.14	0.16
31	MC	204,070	0.52	0.16	0.56
32	MC	204,090	0.64	0.15	0.36
33	MC	204,085	0.46	0.15	0.56
34	MC	203,940	0.60	0.23	0.51
35	MC	204,199	0.32	0.09	0.43
36	MC	204,115	0.56	0.13	0.36
37	MC	204,140	0.22	0.12	0.29
38	MC	203,927	0.36	0.23	0.30
39	MC	204,014	0.58	0.20	0.42
40	MC	204,052	0.29	0.18	0.24
41	MC	203,944	0.53	0.22	0.52
42	MC	203,961	0.56	0.21	0.50
43	MC	204,113	0.43	0.14	0.62
44	MC	204,048	0.71	0.18	0.43
45	MC	204,183	0.52	0.11	0.38
46	MC	203,929	0.47	0.23	0.34
47	MC	204,010	0.52	0.18	0.45
48	MC	203,865	0.51	0.25	0.49
49	CR	203,316	0.70	0.55	
50	CR	203,242	0.55	0.59	
51	CR	202,901	0.46	0.76	
52	CR	200,522	0.13	1.92	
53	CR	202,089	0.36	1.15	
54	CR	202,196	0.28	1.10	
55	CR	203,130	0.58	0.64	
56	CR	202,717	0.47	0.85	
57	CR	202,390	0.43	1.01	
58	CR	201,842	0.41	1.27	

Table M12. Mathematics Classical Item Analysis, Grade 8

Item	Item Type	N-count	P-value	% Omit	Pbis Key
1	MC	208,210	0.91	0.07	0.31
2	MC	208,254	0.77	0.04	0.39
3	MC	208,002	0.74	0.16	0.43
4	MC	207,999	0.76	0.14	0.40
5	MC	208,153	0.79	0.09	0.35
6	MC	208,131	0.39	0.08	0.44
7	MC	207,913	0.27	0.18	0.47
8	MC	207,870	0.54	0.22	0.51
9	MC	207,998	0.43	0.16	0.39
10	MC	207,848	0.31	0.23	0.45
11	MC	208,170	0.83	0.07	0.31
12	MC	208,057	0.71	0.13	0.52
13	MC	207,994	0.31	0.16	0.43
14	MC	208,092	0.48	0.11	0.53
15	MC	208,093	0.65	0.12	0.41
16	MC	208,013	0.31	0.16	0.49
17	MC	207,106	0.71	0.16	0.34
18	MC	208,043	0.29	0.14	0.40
19	MC	208,038	0.55	0.13	0.42
20	MC	207,993	0.61	0.16	0.35
21	MC	207,965	0.54	0.17	0.45
22	MC	208,117	0.81	0.10	0.37
23	MC	207,855	0.34	0.23	0.32
24	MC	208,154	0.76	0.08	0.41
25	MC	208,227	0.82	0.06	0.38
26	MC	207,900	0.58	0.21	0.48
27	MC	208,180	0.66	0.08	0.38
28	MC	208,027	0.75	0.15	0.34
29	MC	208,107	0.45	0.11	0.25
30	MC	208,008	0.44	0.15	0.54
31	MC	208,033	0.66	0.12	0.47
32	MC	208,103	0.34	0.09	0.19
33	MC	208,181	0.53	0.07	0.39
34	MC	208,099	0.56	0.10	0.44
35	MC	208,043	0.62	0.14	0.51
36	MC	207,915	0.30	0.20	0.41
37	MC	208,036	0.60	0.13	0.36
38	MC	208,102	0.51	0.11	0.39
39	MC	208,118	0.46	0.10	0.43

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

Item	Item Type	N-count	P-value	% Omit	Pbis Key
40	MC	208,136	0.59	0.10	0.56
41	MC	207,984	0.47	0.16	0.35
42	MC	207,934	0.45	0.19	0.34
43	MC	207,987	0.55	0.15	0.50
44	MC	208,059	0.44	0.13	0.27
45	MC	208,087	0.37	0.12	0.10
46	MC	208,053	0.18	0.13	0.25
47	MC	207,932	0.66	0.18	0.50
48	MC	207,888	0.54	0.20	0.46
49	CR	205,514	0.50	1.37	
50	CR	205,860	0.63	1.21	
51	CR	199,802	0.43	4.11	
52	CR	206,166	0.63	1.06	
53	CR	204,461	0.54	1.88	
54	CR	205,491	0.53	1.38	
55	CR	198,791	0.20	4.60	
56	CR	204,723	0.42	1.75	
57	CR	202,926	0.51	2.61	
58	CR	206,393	0.68	0.95	

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–N4 show items flagged by the SMD, or Mantel-Haenszel methods. Note that positive values of SMD and Delta in Tables N1–N4 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 Classical DIF Item Flags—MC

Grade	Item #	Subgroup	DIF	Alpha	MH	Delta
4	4	Female	Against	1.72	2134.10	-1.27
4	9	High Needs	Against	1.63	2125.38	-1.14
4	9	Black	Against	1.58	1116.24	-1.08
4	9	Hispanic	Against	1.93	2850.02	-1.54
4	25	Hispanic	Against	1.57	1237.80	-1.07
5	17	Female	Against	1.85	3433.43	-1.45
6	18	Black	Against	1.63	1255.57	-1.15
6	18	Hispanic	Against	1.72	1812.44	-1.27
6	21	High Needs	Against	1.59	1718.13	-1.09
6	21	Black	Against	1.56	1009.20	-1.05
6	21	Hispanic	Against	1.92	2608.97	-1.53
6	31	Black	Against	1.62	688.71	-1.13
6	31	Hispanic	Against	1.63	805.22	-1.15
7	17	Female	Against	1.72	2685.12	-1.28
7	21	Female	Against	1.54	1260.21	-1.02
7	24	Female	Against	1.64	2024.08	-1.17
7	25	Female	Against	1.66	2435.58	-1.20
8	20	Hispanic	Against	1.55	1008.98	-1.02
8	27	Hispanic	Against	1.53	1105.24	-1.01
8	28	Female	Against	1.91	3575.75	-1.52
8	31	Female	Against	1.56	801.44	-1.04
8	32	Female	Against	1.58	2338.31	-1.08
8	34	Female	Against	1.61	1272.25	-1.12

Table N2. ELA Classical DIF Item Flags—CR

Grade	Item #	Subgroup	DIF	SMD	Effect Size
5	40	Female	In Favor	0.1191	0.187
5	41	Female	In Favor	0.1099	0.182
6	36	Female	In Favor	0.1102	0.171
8	50	High Needs	Against	-0.2006	-0.176
8	50	Black	Against	-0.2116	-0.192

Table N3. Mathematics Classical DIF Item Flags—MC

Grade	Item #	Subgroup	DIF	Alpha	MH	Delta
3	17	Female	Against	1.74	2102.39	-1.31
3	22	Black	In Favor	0.65	692.04	1.02
4	31	Hispanic	Against	1.60	1304.56	-1.10
5	41	Black	Against	1.97	1779.47	-1.60
5	41	Hispanic	Against	1.72	1406.94	-1.28
6	15	Female	Against	1.86	2441.60	-1.46
6	15	Black	Against	1.72	1105.83	-1.28
6	15	Hispanic	Against	1.69	1176.37	-1.23
7	9	Female	Against	1.56	1502.51	-1.05
7	23	Hispanic	Against	1.58	1164.09	-1.07
8	3	Female	Against	1.80	2866.94	-1.38

Table N4. Mathematics Classical DIF Item Flags—CR

Grade	Item #	Subgroup	DIF	SMD	Effect Size
3	36	High Needs	Against	-0.1308	-0.177
3	36	Hispanic	Against	-0.1570	-0.212
5	23	Black	In Favor	0.1497	0.200
8	30	High Needs	Against	-0.1419	-0.171
8	38	High Needs	Against	-0.2252	-0.213

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. ELA Grade 3 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	184.31	7	47.39	173.65	Y
2	3PL	117.69	7	29.58	173.56	Y
3	3PL	120.07	7	30.22	173.51	Y
4	3PL	181.94	7	46.75	173.53	Y
5	3PL	114.92	7	28.84	173.45	Y
6	3PL	43.42	7	9.73	173.47	Y
7	3PL	233.79	7	60.61	173.24	Y
8	3PL	120.28	7	30.28	173.17	Y
9	3PL	106.65	7	26.63	172.96	Y
10	3PL	123.39	7	31.11	172.97	Y
11	3PL	20.66	7	3.65	172.89	Y
12	3PL	33.55	7	7.10	172.81	Y
13	3PL	120.97	7	30.46	171.93	Y
14	3PL	194.33	7	50.07	171.73	Y
15	3PL	515.67	7	135.95	171.62	Y
16	3PL	256.74	7	66.75	171.37	Y
17	3PL	70.81	7	17.05	170.99	Y
18	3PL	223.69	7	57.91	170.80	Y
19	3PL	264.21	7	68.74	169.32	Y
20	3PL	154.37	7	39.39	168.76	Y
21	3PL	199.54	7	51.46	168.27	Y
22	3PL	143.40	7	36.45	167.97	Y
23	3PL	189.95	7	48.90	167.65	Y
24	3PL	116.86	7	29.36	167.11	Y
25	3PL	148.93	7	37.93	173.58	Y
26	3PL	195.43	7	50.36	173.57	Y
27	3PL	110.49	7	27.66	173.49	Y
28	3PL	140.77	7	35.75	173.47	Y
29	3PL	159.41	7	40.73	173.46	Y
30	3PL	144.24	7	36.68	173.42	Y
31	3PL	180.09	7	46.26	173.09	Y
32	2PPC	534.29	16	91.62	171.90	Y
33	2PPC	612.65	16	105.47	170.93	Y
34	2PPC	349.19	16	58.90	169.72	Y
35	2PPC	970.33	34	113.55	167.60	Y
36	2PPC	1276.15	16	222.77	173.03	N
37	2PPC	320.43	16	53.82	172.37	Y
38	2PPC	677.68	16	116.97	171.35	Y
39	2PPC	1221.35	16	213.08	171.31	N
40	2PPC	571.70	16	98.23	171.25	Y
41	2PPC	943.61	34	110.31	168.97	Y

Table O2. ELA Grade 4 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	76.97	7	18.70	179.93	Y
2	3PL	115.16	7	28.91	179.89	Y
3	3PL	97.78	7	24.26	179.90	Y
4	3PL	73.12	7	17.67	179.75	Y
5	3PL	90.09	7	22.21	179.85	Y
6	3PL	145.51	7	37.02	179.82	Y
7	3PL	93.33	7	23.07	179.71	Y
8	3PL	339.66	7	88.91	179.65	Y
9	3PL	111.34	7	27.89	179.64	Y
10	3PL	160.99	7	41.16	179.59	Y
11	3PL	204.20	7	52.71	179.53	Y
12	3PL	150.97	7	38.48	179.49	Y
13	3PL	135.84	7	34.44	178.95	Y
14	3PL	105.34	7	26.28	178.91	Y
15	3PL	136.57	7	34.63	178.73	Y
16	3PL	213.67	7	55.24	178.52	Y
17	3PL	62.85	7	14.93	178.31	Y
18	3PL	257.76	7	67.02	178.04	Y
19	3PL	240.13	7	62.31	177.00	Y
20	3PL	195.17	7	50.29	176.66	Y
21	3PL	213.61	7	55.22	176.26	Y
22	3PL	215.79	7	55.80	176.07	Y
23	3PL	112.59	7	28.22	175.68	Y
24	3PL	194.38	7	50.08	175.34	Y
25	3PL	57.90	7	13.60	179.94	Y
26	3PL	112.23	7	28.12	179.90	Y
27	3PL	30.24	7	6.21	179.90	Y
28	3PL	104.73	7	26.12	179.86	Y
29	3PL	127.04	7	32.08	179.90	Y
30	3PL	142.63	7	36.25	179.81	Y
31	3PL	192.11	7	49.47	179.70	Y
32	2PPC	369.77	16	62.54	179.03	Y
33	2PPC	491.51	16	84.06	178.08	Y
34	2PPC	1162.91	16	202.75	178.50	N
35	2PPC	1079.48	34	126.78	176.65	Y
36	2PPC	278.78	16	46.45	179.39	Y
37	2PPC	433.36	16	73.78	179.08	Y
38	2PPC	1565.19	16	273.86	179.09	N
39	2PPC	562.31	16	96.58	178.66	Y
40	2PPC	525.11	16	90.00	177.62	Y
41	2PPC	1025.30	34	120.21	177.03	Y

Table O3. ELA Grade 5 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	117.43	7	29.51	171.46	Y
2	3PL	111.37	7	27.89	171.39	Y
3	3PL	115.61	7	29.03	171.37	Y
4	3PL	48.17	7	11.00	171.34	Y
5	3PL	81.86	7	20.01	171.32	Y
6	3PL	78.46	7	19.10	171.28	Y
7	3PL	61.32	7	14.52	171.27	Y
8	3PL	119.61	7	30.10	171.31	Y
9	3PL	58.72	7	13.82	171.26	Y
10	3PL	83.74	7	20.51	171.31	Y
11	3PL	73.07	7	17.66	171.27	Y
12	3PL	40.39	7	8.92	171.19	Y
13	3PL	71.59	7	17.26	171.22	Y
14	3PL	99.91	7	24.83	171.26	Y
15	3PL	64.86	7	15.46	171.15	Y
16	3PL	263.83	7	68.64	170.98	Y
17	3PL	132.48	7	33.54	171.01	Y
18	3PL	78.98	7	19.24	170.95	Y
19	3PL	124.78	7	31.48	170.83	Y
20	3PL	145.35	7	36.98	170.80	Y
21	3PL	110.55	7	27.68	170.74	Y
22	3PL	110.27	7	27.60	168.55	Y
23	3PL	94.01	7	23.25	168.29	Y
24	3PL	130.15	7	32.91	167.91	Y
25	3PL	87.74	7	21.58	167.66	Y
26	3PL	155.05	7	39.57	167.52	Y
27	3PL	82.59	7	20.20	167.25	Y
28	3PL	106.74	7	26.66	167.17	Y
29	3PL	51.43	7	11.87	171.44	Y
30	3PL	92.14	7	22.75	171.44	Y
31	3PL	84.60	7	20.74	171.42	Y
32	3PL	80.23	7	19.57	171.41	Y
33	3PL	45.02	7	10.16	171.40	Y
34	3PL	371.70	7	97.47	171.35	Y
35	3PL	229.22	7	59.39	171.40	Y
36	3PL	74.71	7	18.10	171.34	Y
37	3PL	100.71	7	25.04	171.34	Y
38	3PL	143.90	7	36.59	171.24	Y
39	3PL	98.49	7	24.45	171.27	Y
40	3PL	173.02	7	44.37	171.23	Y
41	3PL	51.62	7	11.92	171.22	Y

Table O3. ELA Grade 5 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
42	3PL	228.34	7	59.15	171.19	Y
43	2PPC	562.61	16	96.63	170.85	Y
44	2PPC	525.59	16	90.08	170.53	Y
45	2PPC	382.43	16	64.78	168.45	Y
46	2PPC	888.88	34	103.67	167.04	Y
47	2PPC	360.32	16	60.87	170.89	Y
48	2PPC	646.43	16	111.44	170.69	Y
49	2PPC	292.44	16	48.87	167.93	Y
50	2PPC	343.18	16	57.84	170.50	Y
51	2PPC	307.72	16	51.57	169.51	Y
52	2PPC	1066.23	34	125.18	168.80	Y

Table O4. ELA Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	81.55	7	19.92	177.67	Y
2	3PL	51.37	7	11.86	177.65	Y
3	3PL	109.01	7	27.26	177.64	Y
4	3PL	115.22	7	28.92	177.63	Y
5	3PL	134.00	7	33.94	177.65	Y
6	3PL	77.90	7	18.95	177.69	Y
7	3PL	88.89	7	21.89	177.62	Y
8	3PL	56.39	7	13.20	177.57	Y
9	3PL	150.20	7	38.27	177.56	Y
10	3PL	276.11	7	71.92	177.43	Y
11	3PL	158.26	7	40.43	177.53	Y
12	3PL	97.87	7	24.29	177.50	Y
13	3PL	160.15	7	40.93	177.39	Y
14	3PL	35.44	7	7.60	177.31	Y
15	3PL	151.74	7	38.68	177.36	Y
16	3PL	110.00	7	27.53	177.50	Y
17	3PL	131.58	7	33.30	177.44	Y
18	3PL	112.06	7	28.08	177.35	Y
19	3PL	94.33	7	23.34	177.34	Y
20	3PL	66.66	7	15.94	177.30	Y
21	3PL	134.01	7	33.94	175.85	Y
22	3PL	80.16	7	19.55	175.53	Y
23	3PL	78.12	7	19.01	175.44	Y
24	3PL	54.66	7	12.74	175.25	Y
25	3PL	163.62	7	41.86	175.21	Y

Table O4. ELA Grade 6 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
26	3PL	190.62	7	49.07	174.77	Y
27	3PL	182.49	7	46.90	174.73	Y
28	3PL	164.54	7	42.11	177.68	Y
29	3PL	99.51	7	24.72	177.62	Y
30	3PL	112.83	7	28.28	177.55	Y
31	3PL	85.09	7	20.87	177.62	Y
32	3PL	65.85	7	15.73	177.63	Y
33	3PL	61.42	7	14.54	177.65	Y
34	3PL	62.48	7	14.83	177.37	Y
35	3PL	117.03	7	29.41	177.55	Y
36	3PL	96.52	7	23.93	177.47	Y
37	3PL	113.51	7	28.47	177.51	Y
38	3PL	436.14	7	114.69	177.51	Y
39	3PL	89.31	7	22.00	177.43	Y
40	3PL	335.10	7	87.69	177.28	Y
41	3PL	100.68	7	25.04	177.26	Y
42	2PPC	457.79	16	78.10	176.34	Y
43	2PPC	512.22	16	87.72	175.80	Y
44	2PPC	608.40	16	104.72	172.82	Y
45	2PPC	792.05	34	91.93	169.04	Y
46	2PPC	305.00	16	51.09	177.17	Y
47	2PPC	257.72	16	42.73	176.94	Y
48	2PPC	297.07	16	49.69	177.11	Y
49	2PPC	335.12	16	56.41	176.16	Y
50	2PPC	301.65	16	50.50	175.09	Y
51	2PPC	721.27	34	83.34	175.58	Y

Table O5. ELA Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	100.30	7	24.93	176.81	Y
2	3PL	58.56	7	13.78	176.77	Y
3	3PL	66.47	7	15.89	176.65	Y
4	3PL	79.47	7	19.37	176.78	Y
5	3PL	80.54	7	19.65	176.65	Y
6	3PL	75.84	7	18.40	176.74	Y
7	3PL	98.59	7	24.48	176.67	Y
8	3PL	73.17	7	17.69	176.62	Y
9	3PL	92.93	7	22.97	176.60	Y
10	3PL	95.42	7	23.63	176.58	Y
11	3PL	132.40	7	33.51	176.47	Y

Table O5. ELA Grade 7 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
12	3PL	63.06	7	14.98	176.56	Y
13	3PL	119.54	7	30.08	176.54	Y
14	3PL	57.20	7	13.42	176.54	Y
15	3PL	126.20	7	31.86	176.36	Y
16	3PL	166.15	7	42.53	176.47	Y
17	3PL	128.63	7	32.51	176.41	Y
18	3PL	74.18	7	17.95	176.36	Y
19	3PL	69.72	7	16.76	176.28	Y
20	3PL	195.75	7	50.45	176.30	Y
21	3PL	87.88	7	21.62	174.63	Y
22	3PL	220.08	7	56.95	174.49	Y
23	3PL	324.53	7	84.86	174.28	Y
24	3PL	129.77	7	32.81	174.17	Y
25	3PL	170.79	7	43.77	173.80	Y
26	3PL	152.76	7	38.96	173.60	Y
27	3PL	73.13	7	17.67	173.35	Y
28	3PL	99.80	7	24.80	173.07	Y
29	3PL	125.82	7	31.76	176.71	Y
30	3PL	523.37	7	138.00	176.67	Y
31	3PL	58.21	7	13.69	176.47	Y
32	3PL	101.51	7	25.26	176.59	Y
33	3PL	102.80	7	25.60	176.65	Y
34	3PL	90.10	7	22.21	176.56	Y
35	3PL	140.10	7	35.57	176.62	Y
36	3PL	112.80	7	28.28	176.53	Y
37	3PL	97.04	7	24.06	176.45	Y
38	3PL	107.54	7	26.87	176.51	Y
39	3PL	75.95	7	18.43	176.40	Y
40	3PL	323.79	7	84.67	176.33	Y
41	3PL	219.41	7	56.77	176.40	Y
42	3PL	278.06	7	72.44	176.28	Y
43	2PPC	467.62	16	79.84	175.36	Y
44	2PPC	594.44	16	102.26	174.51	Y
45	2PPC	258.58	16	42.88	172.64	Y
46	2PPC	998.93	34	117.01	170.94	Y
47	2PPC	258.78	16	42.92	175.91	Y
48	2PPC	312.57	16	52.43	176.10	Y
49	2PPC	292.09	16	48.81	174.64	Y
50	2PPC	730.25	16	126.26	175.70	Y
51	2PPC	399.37	16	67.77	173.44	Y
52	2PPC	771.26	34	89.41	173.85	Y

Table O6. ELA Grade 8 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	68.68	7	16.49	182.47	Y
2	3PL	66.24	7	15.83	182.37	Y
3	3PL	75.89	7	18.41	182.40	Y
4	3PL	97.06	7	24.07	182.43	Y
5	3PL	136.25	7	34.54	182.37	Y
6	3PL	54.02	7	12.57	182.42	Y
7	3PL	103.49	7	25.79	182.28	Y
8	3PL	182.51	7	46.91	182.40	Y
9	3PL	145.77	7	37.09	182.35	Y
10	3PL	162.29	7	41.50	182.32	Y
11	3PL	86.01	7	21.12	182.24	Y
12	3PL	85.44	7	20.96	182.28	Y
13	3PL	75.00	7	18.17	182.26	Y
14	3PL	74.06	7	17.92	182.24	Y
15	3PL	182.67	7	46.95	182.23	Y
16	3PL	185.10	7	47.60	182.04	Y
17	3PL	109.71	7	27.45	182.21	Y
18	3PL	802.20	7	212.53	182.13	N
19	3PL	200.99	7	51.85	182.13	Y
20	3PL	101.66	7	25.30	182.01	Y
21	3PL	119.25	7	30.00	181.98	Y
22	3PL	134.60	7	34.10	180.26	Y
23	3PL	93.74	7	23.18	180.39	Y
24	3PL	198.93	7	51.30	180.10	Y
25	3PL	188.11	7	48.40	180.01	Y
26	3PL	357.95	7	93.80	180.15	Y
27	3PL	107.95	7	26.98	179.82	Y
28	3PL	174.02	7	44.64	179.67	Y
29	3PL	54.02	7	12.57	182.27	Y
30	3PL	246.56	7	64.02	182.36	Y
31	3PL	151.38	7	38.59	182.22	Y
32	3PL	108.83	7	27.22	182.21	Y
33	3PL	89.57	7	22.07	182.01	Y
34	3PL	155.77	7	39.76	182.35	Y
35	3PL	191.43	7	49.29	182.22	Y
36	3PL	117.91	7	29.64	182.09	Y
37	3PL	47.26	7	10.76	182.08	Y
38	3PL	115.73	7	29.06	182.10	Y
39	3PL	139.52	7	35.42	182.03	Y
40	3PL	107.99	7	26.99	181.94	Y
41	3PL	144.87	7	36.85	181.89	Y
42	3PL	138.81	7	35.23	181.89	Y

Table O6. ELA Grade 8 Item Fit Statistics (cont.)

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
43	2PPC	350.55	16	59.14	181.45	Y
44	2PPC	652.68	16	112.55	177.39	Y
45	2PPC	391.54	16	66.39	175.64	Y
46	2PPC	1308.61	34	154.57	173.49	Y
47	2PPC	302.13	16	50.58	181.85	Y
48	2PPC	508.33	16	87.03	181.41	Y
49	2PPC	343.67	16	57.92	181.66	Y
50	2PPC	421.76	16	71.73	178.87	Y
51	2PPC	384.04	16	65.06	178.41	Y
52	2PPC	882.52	34	102.90	179.78	Y

Table O7. Mathematics Grade 3 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	270.16	7	70.33	177.81	Y
2	3PL	193.53	7	49.85	177.75	Y
3	3PL	92.69	7	22.90	177.67	Y
4	3PL	131.56	7	33.29	177.71	Y
5	3PL	154.00	7	39.29	177.66	Y
6	3PL	157.09	7	40.11	177.69	Y
7	3PL	98.69	7	24.51	177.53	Y
8	3PL	219.16	7	56.70	177.30	Y
9	3PL	77.70	7	18.90	177.60	Y
10	3PL	146.60	7	37.31	177.57	Y
11	3PL	127.85	7	32.30	177.56	Y
12	3PL	137.60	7	34.90	177.54	Y
13	3PL	173.42	7	44.48	177.64	Y
14	3PL	121.83	7	30.69	177.58	Y
15	3PL	61.76	7	14.63	177.56	Y
16	3PL	162.48	7	41.55	177.30	Y
17	3PL	142.86	7	36.31	177.32	Y
18	3PL	148.45	7	37.80	177.18	Y
19	3PL	128.19	7	32.39	177.10	Y
20	3PL	148.93	7	37.93	176.94	Y
21	3PL	58.03	7	13.64	177.79	Y
22	3PL	123.00	7	31.00	177.75	Y
23	3PL	98.36	7	24.42	177.69	Y
24	3PL	70.18	7	16.89	177.66	Y
25	3PL	90.06	7	22.20	177.59	Y
26	3PL	151.11	7	38.51	177.42	Y
27	3PL	211.24	7	54.59	177.62	Y

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

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
28	3PL	85.23	7	20.91	177.68	Y
29	3PL	208.33	7	53.81	177.55	Y
30	3PL	142.31	7	36.16	177.50	Y
31	3PL	98.56	7	24.47	177.32	Y
32	3PL	129.19	7	32.66	177.58	Y
33	3PL	83.79	7	20.52	177.54	Y
34	3PL	254.20	7	66.07	177.50	Y
35	3PL	139.77	7	35.48	177.55	Y
36	3PL	112.11	7	28.09	177.31	Y
37	3PL	266.31	7	69.30	177.06	Y
38	3PL	187.91	7	48.35	177.24	Y
39	3PL	129.27	7	32.68	177.10	Y
40	3PL	118.83	7	29.89	176.95	Y
41	3PL	133.98	7	33.94	176.94	Y
42	2PPC	479.98	16	82.02	177.61	Y
43	2PPC	246.24	16	40.70	177.57	Y
44	2PPC	655.97	16	113.13	177.47	Y
45	2PPC	4297.65	16	756.90	176.65	N
46	2PPC	685.09	16	118.28	177.40	Y
47	2PPC	1674.38	25	233.26	177.22	N
48	2PPC	532.45	25	71.76	177.39	Y
49	2PPC	519.38	25	69.92	176.43	Y

Table O8. Mathematics Grade 4 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	64.48	7	15.36	185.00	Y
2	3PL	53.52	7	12.43	184.95	Y
3	3PL	166.51	7	42.63	184.94	Y
4	3PL	123.12	7	31.03	184.95	Y
5	3PL	52.91	7	12.27	184.93	Y
6	3PL	165.90	7	42.47	184.86	Y
7	3PL	112.94	7	28.31	184.85	Y
8	3PL	122.85	7	30.96	184.75	Y
9	3PL	75.34	7	18.27	184.82	Y
10	3PL	168.30	7	43.11	184.90	Y
11	3PL	344.76	7	90.27	184.74	Y
12	3PL	235.89	7	61.17	184.78	Y
13	3PL	124.59	7	31.43	184.81	Y
14	3PL	126.02	7	31.81	184.87	Y
15	3PL	79.23	7	19.31	184.72	Y

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

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
16	3PL	123.76	7	31.21	184.74	Y
17	3PL	82.44	7	20.16	184.45	Y
18	3PL	116.39	7	29.24	184.62	Y
19	3PL	69.02	7	16.57	184.60	Y
20	3PL	115.43	7	28.98	184.15	Y
21	3PL	86.40	7	21.22	184.93	Y
22	3PL	265.02	7	68.96	184.94	Y
23	3PL	66.40	7	15.87	184.87	Y
24	3PL	291.25	7	75.97	184.83	Y
25	3PL	179.00	7	45.97	184.77	Y
26	3PL	71.52	7	17.24	184.87	Y
27	3PL	60.84	7	14.39	184.77	Y
28	3PL	196.14	7	50.55	184.72	Y
29	3PL	141.55	7	35.96	184.86	Y
30	3PL	93.94	7	23.24	184.78	Y
31	3PL	71.03	7	17.11	183.97	Y
32	3PL	86.52	7	21.25	184.89	Y
33	3PL	121.62	7	30.63	184.85	Y
34	3PL	73.25	7	17.71	184.69	Y
35	3PL	59.66	7	14.07	184.82	Y
36	3PL	95.87	7	23.75	184.67	Y
37	3PL	72.27	7	17.45	184.56	Y
38	3PL	274.72	7	71.55	184.55	Y
39	3PL	51.55	7	11.91	183.52	Y
40	3PL	57.51	7	13.50	184.14	Y
41	3PL	134.37	7	34.04	184.05	Y
42	3PL	1192.04	7	316.71	183.48	N
43	2PPC	381.97	16	64.70	184.63	Y
44	2PPC	446.87	16	76.17	184.58	Y
45	2PPC	1861.48	16	326.24	184.63	N
46	2PPC	207.37	16	33.83	184.23	Y
47	2PPC	637.92	16	109.94	184.54	Y
48	2PPC	1323.22	16	231.09	184.18	N
49	2PPC	456.80	25	61.07	184.55	Y
50	2PPC	508.64	25	68.40	184.42	Y
51	2PPC	818.80	25	112.26	184.19	Y
52	2PPC	346.42	25	45.46	183.08	Y

Table O9. Mathematics Grade 5 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	204.34	7	52.74	175.96	Y
2	3PL	70.13	7	16.87	175.94	Y
3	3PL	118.43	7	29.78	175.79	Y
4	3PL	159.85	7	40.85	175.89	Y
5	3PL	98.60	7	24.48	175.91	Y
6	3PL	54.62	7	12.73	175.81	Y
7	3PL	167.73	7	42.96	175.73	Y
8	3PL	63.98	7	15.23	175.87	Y
9	3PL	74.41	7	18.02	175.61	Y
10	3PL	49.49	7	11.36	175.83	Y
11	3PL	1544.91	7	411.02	175.52	N
12	3PL	119.51	7	30.07	175.87	Y
13	3PL	173.99	7	44.63	175.81	Y
14	3PL	80.82	7	19.73	175.84	Y
15	3PL	125.64	7	31.71	175.72	Y
16	3PL	207.62	7	53.62	175.61	Y
17	3PL	57.44	7	13.48	175.75	Y
18	3PL	70.62	7	17.00	175.78	Y
19	3PL	129.83	7	32.83	175.36	Y
20	3PL	49.50	7	11.36	175.63	Y
21	3PL	348.27	7	91.21	175.95	Y
22	3PL	102.08	7	25.41	175.90	Y
23	3PL	96.95	7	24.04	175.98	Y
24	3PL	261.04	7	67.90	175.83	Y
25	3PL	58.83	7	13.85	175.93	Y
26	3PL	103.39	7	25.76	175.57	Y
27	3PL	120.96	7	30.46	175.95	Y
28	3PL	274.79	7	71.57	175.93	Y
29	3PL	80.58	7	19.66	175.90	Y
30	3PL	83.05	7	20.33	175.70	Y
31	3PL	242.34	7	62.90	175.93	Y
32	3PL	62.47	7	14.83	175.73	Y
33	3PL	132.76	7	33.61	175.76	Y
34	3PL	311.21	7	81.30	175.86	Y
35	3PL	198.71	7	51.24	175.87	Y
36	3PL	27.57	7	5.50	175.75	Y
37	3PL	63.00	7	14.97	175.77	Y
38	3PL	89.86	7	22.15	175.83	Y
39	3PL	35.90	7	7.72	175.83	Y
40	3PL	206.10	7	53.21	175.47	Y
41	3PL	125.86	7	31.77	175.80	Y
42	3PL	66.47	7	15.89	175.75	Y

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

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
43	2PPC	1270.97	16	221.85	175.66	N
44	2PPC	759.03	16	131.35	175.62	Y
45	2PPC	633.85	16	109.22	175.05	Y
46	2PPC	565.41	16	97.12	175.24	Y
47	2PPC	268.36	16	44.61	173.18	Y
48	2PPC	1170.18	16	204.03	175.32	N
49	2PPC	1207.44	25	167.22	175.24	Y
50	2PPC	1102.73	25	152.41	175.29	Y
51	2PPC	1589.64	25	221.27	174.90	N
52	2PPC	728.96	25	99.56	174.63	Y

Table O10. Mathematics Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	48.68	7	11.14	182.29	Y
2	3PL	38.02	7	8.29	182.24	Y
3	3PL	161.65	7	41.33	182.31	Y
4	3PL	142.94	7	36.33	182.34	Y
5	3PL	187.56	7	48.26	182.30	Y
6	3PL	169.33	7	43.38	182.16	Y
7	3PL	92.60	7	22.88	182.13	Y
8	3PL	68.18	7	16.35	182.26	Y
9	3PL	32.64	7	6.85	182.00	Y
10	3PL	522.67	7	137.82	182.04	Y
11	3PL	40.37	7	8.92	181.99	Y
12	3PL	237.23	7	61.53	182.18	Y
13	3PL	47.16	7	10.73	182.26	Y
14	3PL	75.02	7	18.18	182.25	Y
15	3PL	114.57	7	28.75	181.90	Y
16	3PL	52.26	7	12.10	182.26	Y
17	3PL	197.79	7	50.99	182.19	Y
18	3PL	102.05	7	25.40	182.10	Y
19	3PL	87.15	7	21.42	182.04	Y
20	3PL	119.55	7	30.08	181.87	Y
21	3PL	178.25	7	45.77	182.08	Y
22	3PL	77.21	7	18.76	182.09	Y
23	3PL	86.70	7	21.30	181.44	Y
24	3PL	121.81	7	30.69	181.93	Y
25	3PL	131.12	7	33.17	182.33	Y
26	3PL	59.24	7	13.96	182.35	Y
27	3PL	48.81	7	11.17	182.17	Y

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

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
28	3PL	82.65	7	20.22	182.33	Y
29	3PL	124.84	7	31.50	182.23	Y
30	3PL	49.87	7	11.46	182.04	Y
31	3PL	126.07	7	31.82	182.26	Y
32	3PL	138.03	7	35.02	182.23	Y
33	3PL	109.70	7	27.45	182.08	Y
34	3PL	62.03	7	14.71	182.10	Y
35	3PL	90.75	7	22.38	181.53	Y
36	3PL	163.07	7	41.71	182.04	Y
37	3PL	117.91	7	29.64	182.26	Y
38	3PL	105.85	7	26.42	182.10	Y
39	3PL	99.54	7	24.73	182.20	Y
40	3PL	275.37	7	71.72	182.25	Y
41	3PL	75.13	7	18.21	182.16	Y
42	3PL	50.79	7	11.70	182.04	Y
43	3PL	198.16	7	51.09	182.00	Y
44	3PL	67.93	7	16.29	182.06	Y
45	3PL	82.94	7	20.29	182.11	Y
46	3PL	136.78	7	34.69	182.21	Y
47	3PL	109.69	7	27.45	181.79	Y
48	3PL	100.19	7	24.91	181.95	Y
49	2PPC	1698.21	16	297.37	182.10	N
50	2PPC	451.92	16	77.06	181.52	Y
51	2PPC	1302.63	16	227.45	181.93	N
52	2PPC	351.32	16	59.28	180.75	Y
53	2PPC	553.34	16	94.99	181.60	Y
54	2PPC	751.98	16	130.10	180.65	Y
55	2PPC	1036.21	25	143.01	181.81	Y
56	2PPC	1351.16	25	187.55	181.26	Y
57	2PPC	342.61	25	44.92	180.40	Y
58	2PPC	1018.69	25	140.53	181.03	Y

Table O11. Mathematics Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	162.74	7	41.62	179.73	Y
2	3PL	219.76	7	56.86	179.79	Y
3	3PL	230.22	7	59.66	179.75	Y
4	3PL	165.97	7	42.49	179.61	Y
5	3PL	209.50	7	54.12	179.72	Y
6	3PL	241.37	7	62.64	179.66	Y
7	3PL	233.26	7	60.47	179.65	Y
8	3PL	68.38	7	16.40	179.62	Y
9	3PL	41.05	7	9.10	179.71	Y
10	3PL	68.75	7	16.50	179.71	Y
11	3PL	47.99	7	10.95	179.66	Y
12	3PL	191.71	7	49.37	179.70	Y
13	3PL	194.99	7	50.24	179.34	Y
14	3PL	31.01	7	6.42	179.66	Y
15	3PL	110.75	7	27.73	179.41	Y
16	3PL	73.60	7	17.80	179.74	Y
17	3PL	55.71	7	13.02	179.60	Y
18	3PL	64.76	7	15.44	179.57	Y
19	3PL	36.16	7	7.79	179.27	Y
20	3PL	72.56	7	17.52	179.31	Y
21	3PL	321.51	7	84.06	179.33	Y
22	3PL	225.83	7	58.48	179.41	Y
23	3PL	63.86	7	15.20	179.22	Y
24	3PL	103.77	7	25.86	179.31	Y
25	3PL	192.76	7	49.65	179.83	Y
26	3PL	154.08	7	39.31	179.81	Y
27	3PL	127.77	7	32.28	179.74	Y
28	3PL	86.83	7	21.34	179.84	Y
29	3PL	34.12	7	7.25	179.60	Y
30	3PL	208.91	7	53.96	179.73	Y
31	3PL	59.84	7	14.12	179.64	Y
32	3PL	37.23	7	8.08	179.69	Y
33	3PL	120.21	7	30.26	179.64	Y
34	3PL	134.99	7	34.21	179.56	Y
35	3PL	45.14	7	10.19	179.79	Y
36	3PL	38.32	7	8.37	179.71	Y
37	3PL	264.11	7	68.72	179.73	Y
38	3PL	60.89	7	14.40	179.55	Y
39	3PL	33.89	7	7.19	179.57	Y
40	3PL	90.12	7	22.21	179.62	Y
41	3PL	89.12	7	21.95	179.55	Y
42	3PL	90.55	7	22.33	179.58	Y

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

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
43	3PL	137.10	7	34.77	179.67	Y
44	3PL	70.81	7	17.05	179.56	Y
45	3PL	57.59	7	13.52	179.73	Y
46	3PL	63.61	7	15.13	179.51	Y
47	3PL	77.24	7	18.77	179.60	Y
48	3PL	116.49	7	29.26	179.49	Y
49	2PPC	974.00	16	169.35	178.95	Y
50	2PPC	1189.06	16	207.37	178.87	N
51	2PPC	1095.63	16	190.85	178.62	N
52	2PPC	273.92	16	45.59	176.57	Y
53	2PPC	1129.98	16	196.93	177.91	N
54	2PPC	468.74	16	80.03	177.94	Y
55	2PPC	1378.65	25	191.43	178.73	N
56	2PPC	1103.47	25	152.52	178.38	Y
57	2PPC	664.66	25	90.46	178.14	Y
58	2PPC	748.14	25	102.27	177.64	Y

Table O12. Mathematics Grade 8 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	191.51	7	49.31	183.23	Y
2	3PL	80.90	7	19.75	183.28	Y
3	3PL	153.88	7	39.26	183.07	Y
4	3PL	219.61	7	56.82	183.15	Y
5	3PL	75.10	7	18.20	183.21	Y
6	3PL	169.96	7	43.55	183.22	Y
7	3PL	268.01	7	69.76	183.03	Y
8	3PL	65.15	7	15.54	182.98	Y
9	3PL	67.25	7	16.10	183.11	Y
10	3PL	177.24	7	45.50	182.97	Y
11	3PL	128.32	7	32.42	183.22	Y
12	3PL	129.36	7	32.70	183.13	Y
13	3PL	195.74	7	50.44	183.06	Y
14	3PL	80.76	7	19.71	183.17	Y
15	3PL	59.29	7	13.98	183.18	Y
16	3PL	106.99	7	26.72	183.05	Y
17	3PL	88.05	7	21.66	182.27	Y
18	3PL	187.66	7	48.28	183.15	Y
19	3PL	53.65	7	12.47	183.15	Y
20	3PL	47.44	7	10.81	183.07	Y

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

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
21	3PL	95.32	7	23.60	183.02	Y
22	3PL	280.71	7	73.15	183.17	Y
23	3PL	120.02	7	30.21	182.94	Y
24	3PL	117.89	7	29.64	183.21	Y
25	3PL	183.96	7	47.29	183.23	Y
26	3PL	110.80	7	27.74	182.98	Y
27	3PL	135.89	7	34.45	183.21	Y
28	3PL	51.82	7	11.98	183.06	Y
29	3PL	29.34	7	5.97	183.13	Y
30	3PL	85.24	7	20.91	183.10	Y
31	3PL	113.30	7	28.41	183.12	Y
32	3PL	92.01	7	22.72	183.17	Y
33	3PL	76.75	7	18.64	183.23	Y
34	3PL	76.41	7	18.55	183.15	Y
35	3PL	85.90	7	21.09	183.07	Y
36	3PL	74.85	7	18.13	182.99	Y
37	3PL	140.20	7	35.60	183.09	Y
38	3PL	40.66	7	9.00	183.15	Y
39	3PL	55.67	7	13.01	183.15	Y
40	3PL	154.43	7	39.40	183.17	Y
41	3PL	46.81	7	10.64	183.03	Y
42	3PL	63.06	7	14.98	182.99	Y
43	3PL	127.70	7	32.26	183.06	Y
44	3PL	197.95	7	51.03	183.10	Y
45	3PL	176.22	7	45.23	183.11	Y
46	3PL	352.64	7	92.38	183.09	Y
47	3PL	90.49	7	22.31	183.00	Y
48	3PL	106.19	7	26.51	182.97	Y
49	2PPC	414.09	16	70.37	180.81	Y
50	2PPC	325.83	16	54.77	181.18	Y
51	2PPC	349.81	16	59.01	175.56	Y
52	2PPC	607.59	16	104.58	181.39	Y
53	2PPC	334.98	16	56.39	179.80	Y
54	2PPC	597.95	16	102.87	180.83	Y
55	2PPC	840.47	25	115.33	174.92	Y
56	2PPC	719.40	25	98.20	180.19	Y
57	2PPC	1473.52	25	204.85	178.56	N
58	2PPC	839.61	25	115.20	181.60	Y

Table O13. ELA Operational Item Parameter Estimates, Grade 3

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.934	-0.666	0.166		
2	1	0.566	1.617	0.132		
3	1	0.816	1.293	0.201		
4	1	1.265	-0.320	0.238		
5	1	0.615	-0.234	0.167		
6	1	0.244	0.302	0.038		
7	1	1.028	0.920	0.265		
8	1	0.930	0.171	0.263		
9	1	0.877	0.897	0.234		
10	1	0.478	-0.707	0.037		
11	1	0.430	2.528	0.175		
12	1	0.192	2.536	0.076		
13	1	0.590	0.160	0.187		
14	1	0.896	-0.232	0.166		
15	1	0.901	-0.334	0.120		
16	1	1.183	-0.159	0.188		
17	1	0.413	1.585	0.190		
18	1	1.088	-0.394	0.218		
19	1	1.419	1.083	0.226		
20	1	0.871	0.023	0.197		
21	1	0.978	1.217	0.176		
22	1	0.899	0.470	0.193		
23	1	1.085	0.627	0.200		
24	1	1.004	-0.224	0.246		
25	1	1.017	-0.297	0.205		
26	1	0.932	-1.124	0.154		
27	1	0.477	-0.820	0.038		
28	1	0.977	-0.115	0.265		
29	1	0.890	2.000	0.257		
30	1	0.999	-0.133	0.174		
31	1	1.140	-0.695	0.230		
32	2	1.221	-0.563	1.571		
33	2	1.015	-1.415	1.154		
34	2	0.859	-1.229	1.397		
35	4	1.353	-1.675	-0.259	1.232	2.859
36	2	0.986	-1.478	0.859		
37	2	1.102	-1.156	1.195		
38	2	1.401	-2.407	1.354		
39	2	1.236	-1.975	1.331		
40	2	1.027	-2.172	0.938		
41	4	1.183	-1.745	-0.210	1.199	2.418

Table O14. ELA Operational Item Parameter Estimates, Grade 4

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.510	-0.327	0.166		
2	1	0.933	-0.606	0.328		
3	1	0.646	-0.939	0.245		
4	1	0.602	1.008	0.280		
5	1	0.694	-0.527	0.190		
6	1	0.763	-0.318	0.143		
7	1	0.683	-1.046	0.235		
8	1	0.822	-0.164	0.171		
9	1	0.685	-0.372	0.167		
10	1	0.866	0.324	0.150		
11	1	1.245	-0.877	0.189		
12	1	1.125	-0.563	0.270		
13	1	0.761	0.320	0.170		
14	1	0.806	-0.110	0.194		
15	1	0.726	0.691	0.159		
16	1	1.131	0.345	0.239		
17	1	0.477	-0.957	0.177		
18	1	0.766	-0.391	0.162		
19	1	0.959	0.705	0.200		
20	1	0.977	0.158	0.167		
21	1	0.977	0.520	0.218		
22	1	0.707	0.177	0.175		
23	1	0.636	0.392	0.166		
24	1	0.657	0.791	0.123		
25	1	0.354	-0.977	0.020		
26	1	0.797	0.301	0.272		
27	1	0.181	1.102	0.025		
28	1	0.245	-0.818	0.009		
29	1	0.776	-0.559	0.180		
30	1	0.821	0.989	0.201		
31	1	1.295	0.149	0.246		
32	2	1.181	-1.583	1.011		
33	2	1.315	-1.749	0.920		
34	2	1.484	-2.933	1.008		
35	4	1.179	-1.903	-0.319	1.073	2.408
36	2	0.993	-1.199	1.238		
37	2	1.305	-2.379	0.870		
38	2	1.271	-2.589	0.511		
39	2	1.706	-2.035	0.992		
40	2	0.983	-0.404	1.671		
41	4	1.329	-2.426	-0.634	1.050	2.603

Table O15. ELA Operational Item Parameter Estimates, Grade 5

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.376	-1.647	0.025		
2	1	1.071	-0.579	0.320		
3	1	0.758	0.011	0.230		
4	1	0.536	1.096	0.204		
5	1	0.781	0.174	0.150		
6	1	0.592	-0.347	0.248		
7	1	0.546	1.256	0.221		
8	1	0.997	-0.351	0.230		
9	1	0.140	1.610	0.018		
10	1	0.590	-0.301	0.163		
11	1	0.585	1.797	0.178		
12	1	0.212	0.624	0.018		
13	1	0.429	0.594	0.091		
14	1	0.720	0.228	0.146		
15	1	0.671	0.178	0.200		
16	1	0.713	0.299	0.121		
17	1	0.963	0.383	0.196		
18	1	0.589	0.309	0.182		
19	1	0.908	0.656	0.224		
20	1	1.007	-0.191	0.173		
21	1	0.945	0.040	0.214		
22	1	0.891	-0.274	0.212		
23	1	0.733	0.273	0.229		
24	1	0.834	0.214	0.191		
25	1	0.661	2.170	0.170		
26	1	1.148	-0.191	0.210		
27	1	0.511	0.209	0.162		
28	1	0.765	0.101	0.182		
29	1	0.489	-0.179	0.461		
30	1	0.604	0.222	0.184		
31	1	0.582	-0.350	0.221		
32	1	0.457	-0.798	0.166		
33	1	0.538	0.087	0.407		
34	1	0.168	-0.336	0.010		
35	1	0.448	-1.673	0.012		
36	1	0.787	0.272	0.271		
37	1	0.878	-0.104	0.245		
38	1	0.841	-0.527	0.194		
39	1	0.642	-0.477	0.102		
40	1	0.685	0.682	0.122		
41	1	0.444	1.431	0.199		
42	1	0.599	-0.961	0.018		

Table O15. ELA Operational Item Parameter Estimates, Grade 5 (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
43	2	1.167	-2.871	-0.088		
44	2	1.196	-2.388	0.432		
45	2	1.392	-1.484	1.204		
46	4	1.279	-1.985	-0.350	1.135	2.773
47	2	1.308	-2.332	0.666		
48	2	1.267	-1.007	1.147		
49	2	0.967	-0.976	0.630		
50	2	1.245	-0.803	0.312		
51	2	1.296	-1.739	1.434		
52	4	1.359	-1.811	-0.596	0.738	2.091

Table O16. ELA Operational Item Parameter Estimates, Grade 6

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.350	-0.701	0.016		
2	1	0.463	-0.048	0.179		
3	1	0.827	-0.711	0.174		
4	1	0.600	-0.616	0.190		
5	1	0.903	-1.132	0.120		
6	1	0.872	-1.494	0.399		
7	1	0.956	-0.680	0.253		
8	1	0.567	-0.200	0.209		
9	1	0.977	-0.094	0.171		
10	1	1.035	1.151	0.176		
11	1	1.002	-0.690	0.184		
12	1	0.941	0.149	0.320		
13	1	0.814	0.509	0.183		
14	1	0.392	1.311	0.201		
15	1	0.838	1.075	0.191		
16	1	1.014	-0.654	0.212		
17	1	1.322	-0.889	0.212		
18	1	0.684	0.318	0.238		
19	1	0.871	-0.836	0.230		
20	1	0.461	-0.046	0.122		
21	1	0.635	1.038	0.189		
22	1	0.774	0.441	0.265		
23	1	0.471	0.374	0.163		
24	1	0.549	1.220	0.198		
25	1	1.437	-0.134	0.236		
26	1	1.080	0.865	0.193		
27	1	0.953	0.158	0.175		
28	1	0.904	1.595	0.158		

Table O16. ELA Operational Item Parameter Estimates, Grade 6 (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
29	1	0.693	1.664	0.209		
30	1	0.755	0.283	0.162		
31	1	0.914	-0.660	0.280		
32	1	0.716	-0.164	0.237		
33	1	0.526	-0.606	0.111		
34	1	0.652	0.011	0.214		
35	1	0.620	-0.285	0.103		
36	1	0.706	0.294	0.188		
37	1	0.721	-1.201	0.130		
38	1	0.630	-1.398	0.007		
39	1	0.627	-0.151	0.153		
40	1	0.501	-0.132	0.029		
41	1	0.783	-0.620	0.131		
42	2	1.319	-1.715	0.084		
43	2	1.149	-2.015	0.569		
44	2	1.388	-2.385	0.512		
45	4	1.076	-1.663	-0.410	0.826	1.901
46	2	1.169	-2.373	0.578		
47	2	1.176	-2.375	0.175		
48	2	1.100	-1.573	0.937		
49	2	1.369	-3.435	0.092		
50	2	1.403	-2.278	0.462		
51	4	1.201	-2.622	-0.830	0.650	1.787

Table O17. ELA Operational Item Parameter Estimates, Grade 7

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.624	-2.267	0.026		
2	1	0.401	-1.187	0.009		
3	1	0.620	0.789	0.422		
4	1	0.822	-1.681	0.083		
5	1	0.575	0.496	0.241		
6	1	0.660	1.540	0.452		
7	1	0.620	0.516	0.194		
8	1	0.460	-0.735	0.109		
9	1	0.463	0.433	0.294		
10	1	0.657	0.959	0.117		
11	1	0.745	0.636	0.196		
12	1	0.128	0.581	0.047		
13	1	0.515	-0.423	0.108		
14	1	0.703	0.513	0.331		
15	1	0.824	0.871	0.249		
16	1	1.289	-0.261	0.208		
17	1	0.805	0.494	0.215		
18	1	0.622	0.748	0.240		
19	1	0.555	-0.795	0.137		
20	1	0.772	-0.944	0.135		
21	1	1.082	-0.697	0.295		
22	1	1.144	0.443	0.206		
23	1	0.907	0.498	0.174		
24	1	1.140	-0.630	0.244		
25	1	0.880	1.206	0.235		
26	1	1.207	-0.581	0.260		
27	1	0.661	0.577	0.185		
28	1	0.604	0.881	0.218		
29	1	1.023	-1.019	0.254		
30	1	0.256	-1.374	0.006		
31	1	0.596	0.634	0.374		
32	1	0.616	-1.417	0.070		
33	1	0.967	-0.851	0.154		
34	1	0.171	0.702	0.013		
35	1	0.777	-0.428	0.261		
36	1	1.070	-0.064	0.339		
37	1	0.708	1.546	0.145		
38	1	0.740	-1.277	0.085		
39	1	0.417	-0.283	0.093		
40	1	1.342	1.045	0.230		
41	1	0.782	-0.465	0.307		
42	1	1.627	0.580	0.240		

Table O17. ELA Operational Item Parameter Estimates, Grade 7 (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
43	2	1.224	-0.556	-0.423		
44	2	1.205	-1.740	0.072		
45	2	1.316	-2.164	0.096		
46	4	1.273	-2.343	-0.923	0.417	1.528
47	2	1.106	-0.933	0.556		
48	2	1.262	-2.383	0.228		
49	2	1.448	-2.797	-0.249		
50	2	1.222	-2.819	-0.079		
51	2	1.537	-2.356	0.388		
52	4	1.257	-2.720	-1.309	0.306	1.597

Table O18. ELA Operational Item Parameter Estimates, Grade 8

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.580	-2.553	0.025		
2	1	0.501	-0.562	0.145		
3	1	0.434	-1.494	0.022		
4	1	0.830	-0.327	0.252		
5	1	0.699	0.001	0.090		
6	1	0.377	1.378	0.108		
7	1	0.716	0.048	0.262		
8	1	0.418	0.322	0.006		
9	1	0.964	-0.311	0.228		
10	1	0.354	-1.107	0.010		
11	1	0.353	-0.351	0.104		
12	1	0.426	-0.619	0.073		
13	1	0.471	-1.148	0.129		
14	1	0.638	0.119	0.326		
15	1	0.786	0.020	0.200		
16	1	0.881	0.588	0.165		
17	1	0.697	-0.331	0.230		
18	1	0.289	-2.396	0.007		
19	1	1.141	-0.185	0.217		
20	1	0.787	-0.010	0.309		
21	1	0.720	-0.638	0.156		
22	1	0.676	0.692	0.132		
23	1	0.326	-0.622	0.044		
24	1	0.805	1.029	0.189		
25	1	0.773	0.571	0.209		
26	1	0.663	-1.860	0.010		
27	1	0.429	0.545	0.134		

Table O18. ELA Operational Item Parameter Estimates, Grade 8 (cont.)

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
28	1	1.206	-0.192	0.286		
29	1	0.426	2.021	0.144		
30	1	1.429	-0.709	0.169		
31	1	0.406	0.529	0.126		
32	1	0.624	0.248	0.197		
33	1	0.453	1.616	0.118		
34	1	0.505	-0.020	0.226		
35	1	0.583	1.380	0.117		
36	1	0.543	0.091	0.156		
37	1	0.441	-1.316	0.109		
38	1	0.430	-1.476	0.040		
39	1	0.914	-0.385	0.247		
40	1	0.723	0.215	0.224		
41	1	0.844	1.062	0.266		
42	1	1.000	-0.198	0.303		
43	2	1.001	-2.446	-0.570		
44	2	0.878	-0.834	0.551		
45	2	1.169	-1.834	-0.118		
46	4	1.081	-1.898	-1.018	0.091	1.188
47	2	1.234	-3.128	-0.334		
48	2	1.244	-2.512	-0.176		
49	2	1.288	-2.526	-0.928		
50	2	1.413	-1.690	-0.129		
51	2	1.527	-2.976	-0.023		
52	4	1.288	-3.234	-1.692	-0.267	1.091

Table O19. Mathematics Operational Item Parameter Estimates, Grade 3

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.622	-2.381	0.014	
2	1	0.675	-2.124	0.010	
3	1	0.679	-0.745	0.326	
4	1	0.750	-1.996	0.010	
5	1	0.961	-0.968	0.132	
6	1	1.151	-0.659	0.040	
7	1	1.134	0.118	0.122	
8	1	1.178	1.220	0.235	
9	1	0.831	0.311	0.219	
10	1	1.152	-0.043	0.166	
11	1	0.986	-0.784	0.095	
12	1	1.307	0.059	0.283	
13	1	1.062	0.452	0.265	

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

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
14	1	1.251	0.060	0.164	
15	1	0.790	-1.567	0.362	
16	1	0.832	0.033	0.098	
17	1	0.980	-0.669	0.097	
18	1	1.411	-0.068	0.078	
19	1	1.061	0.034	0.109	
20	1	1.641	-0.510	0.135	
21	1	0.604	-1.824	0.358	
22	1	0.886	-1.330	0.086	
23	1	0.936	-0.916	0.186	
24	1	1.030	-0.429	0.171	
25	1	1.123	-0.564	0.285	
26	1	1.067	1.361	0.330	
27	1	1.299	-0.943	0.082	
28	1	0.801	-0.147	0.199	
29	1	1.103	-0.639	0.024	
30	1	1.237	-0.563	0.113	
31	1	0.704	1.802	0.321	
32	1	1.058	0.296	0.246	
33	1	0.755	0.092	0.159	
34	1	0.844	0.269	0.157	
35	1	0.947	-0.275	0.224	
36	1	0.991	0.053	0.250	
37	1	1.673	0.558	0.119	
38	1	1.506	0.387	0.109	
39	1	0.984	0.311	0.087	
40	1	0.656	1.302	0.169	
41	1	1.140	-0.953	0.116	
42	2	1.078	-0.812	-0.343	
43	2	0.907	1.316	-2.946	
44	2	1.141	-1.103	-0.533	
45	2	0.686	0.583	0.942	
46	2	1.090	-0.764	0.027	
47	3	0.918	0.263	0.781	-0.445
48	3	1.030	0.422	-0.950	0.470
49	3	1.117	0.704	0.643	0.633

Table O20. Mathematics Operational Item Parameter Estimates, Grade 4

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.639	-3.379	0.032	
2	1	0.791	-1.625	0.473	
3	1	1.027	-1.867	0.117	
4	1	0.847	-1.792	0.050	
5	1	0.964	-1.374	0.487	
6	1	0.867	-2.035	0.019	
7	1	1.110	0.929	0.214	
8	1	1.467	0.793	0.270	
9	1	0.901	0.033	0.244	
10	1	1.141	-0.453	0.063	
11	1	1.045	-1.097	0.066	
12	1	0.733	-1.912	0.013	
13	1	0.995	-0.714	0.197	
14	1	1.132	-0.134	0.102	
15	1	1.174	0.152	0.252	
16	1	1.312	0.129	0.203	
17	1	1.028	0.716	0.423	
18	1	0.529	-0.163	0.070	
19	1	0.747	-0.251	0.083	
20	1	1.197	-0.107	0.164	
21	1	1.015	-0.982	0.349	
22	1	0.939	-1.199	0.011	
23	1	0.711	-1.595	0.242	
24	1	1.341	-0.828	0.108	
25	1	1.007	-0.140	0.071	
26	1	0.841	0.385	0.168	
27	1	0.801	0.369	0.168	
28	1	0.965	0.239	0.093	
29	1	1.121	-0.277	0.102	
30	1	0.997	0.863	0.117	
31	1	1.324	1.402	0.296	
32	1	0.671	-0.235	0.187	
33	1	1.269	-0.498	0.198	
34	1	0.810	0.263	0.094	
35	1	0.793	-0.133	0.312	
36	1	1.256	0.247	0.189	
37	1	0.920	0.646	0.200	
38	1	0.864	0.559	0.063	
39	1	1.335	1.960	0.330	
40	1	0.767	0.568	0.230	
41	1	1.275	0.817	0.136	
42	1	1.614	1.493	0.159	

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

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
43	2	0.457	-0.039	-1.144	
44	2	1.173	1.215	0.130	
45	2	1.430	-1.131	0.398	
46	2	1.126	1.548	0.403	
47	2	1.458	1.014	1.522	
48	2	1.396	-0.320	2.228	
49	3	1.040	-0.162	-0.999	-0.749
50	3	1.297	-0.355	-0.121	1.112
51	3	1.160	-0.514	0.585	1.972
52	3	0.792	2.382	-0.310	-0.300

Table O21. Mathematics Operational Item Parameter Estimates, Grade 5

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.677	-2.954	0.027	
2	1	0.916	-0.115	0.760	
3	1	0.926	0.459	0.221	
4	1	1.142	-0.321	0.160	
5	1	0.574	-0.689	0.082	
6	1	1.032	0.327	0.279	
7	1	1.909	0.676	0.211	
8	1	1.120	-0.216	0.233	
9	1	0.961	-0.150	0.186	
10	1	0.925	0.033	0.205	
11	1	2.754	1.783	0.225	
12	1	1.189	-0.745	0.143	
13	1	1.538	0.622	0.376	
14	1	1.035	-0.248	0.191	
15	1	1.144	-0.212	0.149	
16	1	0.919	-0.945	0.085	
17	1	1.158	-0.370	0.309	
18	1	0.943	0.195	0.283	
19	1	1.356	0.823	0.154	
20	1	0.730	0.931	0.189	
21	1	1.056	-1.509	0.011	
22	1	1.307	0.178	0.465	
23	1	0.977	-1.371	0.382	
24	1	0.973	-1.457	0.058	
25	1	0.918	-0.245	0.373	
26	1	1.573	0.751	0.239	
27	1	1.111	-0.892	0.187	
28	1	1.355	-1.580	0.062	

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

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
29	1	1.821	0.749	0.448	
30	1	1.323	-0.046	0.189	
31	1	1.279	-1.113	0.073	
32	1	0.526	0.541	0.286	
33	1	1.440	0.634	0.147	
34	1	1.069	1.386	0.464	
35	1	1.575	-0.023	0.099	
36	1	0.770	1.578	0.175	
37	1	0.944	0.464	0.207	
38	1	1.228	1.058	0.221	
39	1	0.851	0.867	0.263	
40	1	1.561	1.123	0.186	
41	1	0.789	-1.801	0.036	
42	1	1.285	-0.363	0.291	
43	2	0.760	-0.579	-0.223	
44	2	1.366	-0.153	-0.234	
45	2	1.861	-0.309	0.735	
46	2	1.696	0.309	3.707	
47	2	1.403	1.476	1.059	
48	2	0.936	0.350	1.140	
49	3	1.457	-0.326	0.225	1.129
50	3	1.427	-1.083	0.498	0.700
51	3	0.814	0.939	1.203	-0.467
52	3	1.246	0.390	2.264	2.778

Table O22. Mathematics Operational Item Parameter Estimates, Grade 6

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.919	-0.887	0.306	
2	1	0.806	0.220	0.321	
3	1	0.933	-0.794	0.065	
4	1	0.996	-0.960	0.080	
5	1	1.294	-1.451	0.060	
6	1	0.588	-1.172	0.008	
7	1	0.899	0.116	0.206	
8	1	0.671	-1.360	0.119	
9	1	0.634	0.493	0.255	
10	1	2.065	1.185	0.329	
11	1	1.112	1.124	0.293	
12	1	0.664	-0.825	0.071	
13	1	0.950	0.637	0.280	
14	1	1.315	0.857	0.209	

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

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
15	1	1.596	0.471	0.103	
16	1	0.912	-0.847	0.422	
17	1	0.434	-0.665	0.007	
18	1	1.539	0.727	0.420	
19	1	0.739	-0.084	0.168	
20	1	1.285	0.597	0.335	
21	1	0.939	-0.613	0.079	
22	1	0.923	0.486	0.187	
23	1	1.642	0.870	0.228	
24	1	1.357	0.677	0.208	
25	1	0.869	-1.789	0.099	
26	1	0.764	-1.437	0.288	
27	1	0.756	-0.820	0.213	
28	1	1.041	-0.638	0.267	
29	1	0.918	-0.697	0.111	
30	1	1.120	1.173	0.236	
31	1	1.378	0.483	0.207	
32	1	1.166	-0.733	0.305	
33	1	1.455	0.245	0.119	
34	1	0.740	0.864	0.319	
35	1	1.071	0.411	0.140	
36	1	1.495	0.681	0.115	
37	1	0.870	-0.787	0.068	
38	1	1.068	0.325	0.228	
39	1	0.828	-0.391	0.162	
40	1	1.442	0.486	0.216	
41	1	0.880	-0.770	0.144	
42	1	0.834	0.924	0.213	
43	1	1.344	1.632	0.153	
44	1	0.813	-0.258	0.148	
45	1	1.460	0.127	0.178	
46	1	0.924	0.782	0.110	
47	1	1.236	1.162	0.270	
48	1	1.479	-0.093	0.188	
49	2	1.128	-2.033	-0.971	
50	2	1.282	0.062	0.543	
51	2	1.316	-0.914	0.895	
52	2	1.946	0.469	1.390	
53	2	1.194	-0.633	1.889	
54	2	1.027	1.582	-0.360	
55	3	1.060	-1.123	-0.869	-1.825
56	3	1.225	-1.062	0.955	-0.874
57	3	1.164	-0.861	-0.053	1.543
58	3	1.085	0.914	0.513	1.100

Table O23. Mathematics Operational Item Parameter Estimates, Grade 7

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.635	-0.956	0.149	
2	1	0.714	-0.880	0.070	
3	1	0.478	-1.575	0.030	
4	1	0.873	-0.121	0.294	
5	1	1.030	-0.500	0.117	
6	1	0.775	1.656	0.299	
7	1	1.364	0.275	0.332	
8	1	1.468	1.490	0.270	
9	1	0.589	0.893	0.233	
10	1	1.134	0.450	0.136	
11	1	0.784	0.499	0.271	
12	1	1.784	1.294	0.152	
13	1	1.783	1.071	0.346	
14	1	0.729	0.594	0.249	
15	1	1.238	0.659	0.099	
16	1	1.303	1.093	0.169	
17	1	1.183	0.294	0.129	
18	1	1.346	0.592	0.152	
19	1	1.280	0.934	0.318	
20	1	1.106	1.162	0.241	
21	1	2.915	1.882	0.283	
22	1	1.590	1.989	0.328	
23	1	1.227	0.245	0.225	
24	1	0.855	0.681	0.279	
25	1	1.192	-0.875	0.186	
26	1	0.617	-0.763	0.106	
27	1	1.136	-0.106	0.195	
28	1	0.948	-1.146	0.581	
29	1	1.481	1.186	0.340	
30	1	2.027	1.853	0.194	
31	1	1.353	0.284	0.165	
32	1	0.757	0.207	0.338	
33	1	1.375	0.416	0.131	
34	1	1.629	0.239	0.301	
35	1	0.791	1.026	0.073	
36	1	0.944	0.667	0.349	
37	1	2.425	1.490	0.142	
38	1	1.487	1.357	0.265	
39	1	1.860	0.599	0.394	
40	1	1.849	1.594	0.222	
41	1	1.560	0.401	0.249	
42	1	1.302	0.292	0.244	

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

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
43	1	1.352	0.377	0.057	
44	1	1.054	-0.132	0.365	
45	1	0.634	0.379	0.176	
46	1	0.713	0.904	0.242	
47	1	1.064	0.450	0.235	
48	1	1.644	0.545	0.270	
49	2	1.537	-1.630	-0.685	
50	2	0.992	0.393	-0.724	
51	2	1.937	-0.315	0.905	
52	2	1.475	2.281	2.461	
53	2	1.157	1.397	-0.211	
54	2	1.970	0.153	3.975	
55	3	1.148	-0.357	-0.508	-0.052
56	3	1.666	-0.358	-0.166	1.191
57	3	1.657	-0.450	0.405	1.359
58	3	1.830	-0.515	0.129	2.801

Table O24. Mathematics Operational Item Parameter Estimates, Grade 8

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.762	-2.125	0.097	
2	1	0.766	-0.855	0.248	
3	1	0.969	-0.479	0.302	
4	1	0.660	-1.257	0.009	
5	1	0.841	-0.450	0.442	
6	1	1.314	0.819	0.180	
7	1	1.989	1.047	0.108	
8	1	1.389	0.338	0.228	
9	1	1.424	0.904	0.261	
10	1	1.388	1.009	0.125	
11	1	0.531	-2.023	0.023	
12	1	1.775	-0.162	0.332	
13	1	1.163	1.029	0.116	
14	1	1.240	0.378	0.151	
15	1	0.716	-0.247	0.213	
16	1	1.596	0.920	0.117	
17	1	0.507	-1.053	0.085	
18	1	1.362	1.160	0.134	
19	1	0.815	0.282	0.210	
20	1	0.932	0.515	0.379	
21	1	1.238	0.468	0.274	
22	1	0.642	-1.593	0.035	

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

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
23	1	1.867	1.214	0.234	
24	1	0.727	-0.954	0.160	
25	1	0.688	-1.551	0.056	
26	1	0.868	-0.044	0.125	
27	1	0.518	-0.826	0.025	
28	1	1.520	0.333	0.581	
29	1	0.564	1.399	0.279	
30	1	1.338	0.479	0.133	
31	1	1.243	-0.017	0.303	
32	1	1.120	1.809	0.279	
33	1	0.634	0.261	0.155	
34	1	1.258	0.434	0.297	
35	1	1.121	-0.030	0.213	
36	1	1.210	1.115	0.129	
37	1	0.700	0.237	0.282	
38	1	0.814	0.539	0.227	
39	1	1.054	0.688	0.210	
40	1	1.298	-0.039	0.147	
41	1	0.712	0.773	0.216	
42	1	0.941	0.984	0.267	
43	1	0.927	0.070	0.118	
44	1	0.319	0.513	0.004	
45	1	1.604	2.036	0.343	
46	1	1.776	1.656	0.118	
47	1	1.244	-0.095	0.261	
48	1	1.316	0.464	0.275	
49	2	1.607	-0.267	0.355	
50	2	1.303	-1.189	-0.107	
51	2	1.216	-0.307	1.098	
52	2	0.968	-0.921	-0.225	
53	2	1.430	0.412	-0.684	
54	2	1.344	-0.991	0.686	
55	3	1.326	0.982	2.408	1.347
56	3	1.227	-0.184	0.404	1.034
57	3	1.294	-0.382	-0.684	1.328
58	3	0.880	-0.741	0.022	-1.199

Appendix P: Standard-setting Technical Report

Overview of Standard-setting Meetings

The standard-setting activities occurred over a five-day period from June 29 to July 3, 2013. There were two components: grade-level committees and vertical-articulation committees. Altogether, ninety-five[1] committee members from across the state participated in the standard-setting committee meetings.

The standard-setting methodology employed was designed in collaboration with NYSED and its Technical Advisory Committee (TAC). An evidence-based, standard-setting process was used that included both external benchmark data to empirically define college readiness, as well as a content-driven standard-setting methodology called the Bookmark Method.

The grade-level committee meetings were held from June 29 to July 2, 2013. There were a total of four committees, three grades in a given subject area:

- Grades 3–5 ELA
- Grades 3–5 Mathematics
- Grades 6–8 ELA
- Grades 6–8 Mathematics

For each individual grade, the committee participants went through a total of four rounds of review to recommend the performance standards.

The vertical-articulation committee meetings occurred on July 3, 2013. This committee was comprised of representatives from each of the grade-level committees that met on the four days before the vertical-articulation meeting. The task of the vertical-articulation committee was to make standards recommendations across grades. The committee members participated in a total of two rounds to make final recommendations for performance standards across all six grades for a given subject (i.e., ELA and mathematics).

Participants

NYSED went through a thorough nomination and recruitment process to invite participants for the standard-setting meetings. Various stakeholder groups such as School Administrators Association of New York State (SAANYS), New York State United Teachers (NYSUT), New York State Council of School Superintendents (NYSCOSS), Association of Mathematics Teachers of New York State (AMTNYS), New York State English Council (NYSEC), Staff and Curriculum Development Network (SCDN), Big Five Cities, Institutes of Higher Education, special education directors, district superintendents, assistant superintendents, and superintendents of schools were asked to nominate individuals to serve as panelists at the standard-setting. The section labeled "Standard-setting Nomination Letters" provides the nomination letters sent along with a document explaining the standard-setting process to the people while considering the nominations. The final list of panelists of ninety-five teachers,

[1] One panelist had to leave the meeting early on Day 4 due to a family emergency.

administrators, and individuals from institutes of higher education were selected to be representatives of New York State.

A total of ninety-five participants attended the standard-setting meetings. The individuals who participated in the meetings have held a variety of positions in education with a wide array of experiences. Many of the participants have had experiences educating students in special populations, such as English Language Learners (ELL) and special education students. Committee members also represented a range of district sizes from small to large with geographic representation across the state. The following tables provide a breakout of the demographic information, geographic locations, size and location of the districts, and experiences of the panelists. As can be observed from the tables, NYSED made a great effort to get a representative standard-setting panel using a variety of demographic variables.

Table P1. Demographic Composition of the Panels, Field, and Gender

	K-12 Educators	Higher Ed	Male	Female
ELA Grades 3–5	87%	13%	8%	92%
ELA Grades 6–8	81%	19%	41%	59%
Math Grades 3–5	96%	4%	26%	74%
Math Grades 6–8	79%	21%	37%	63%

Table P2. Demographic Composition of the Panel, Ethnicity

	African American	Hispanic	White	Other
ELA Grades 3–5	8%	8%	67%	17%
ELA Grades 6–8	14%	5%	77%	4%
Math Grades 3–5	17%	9%	74%	0%
Math Grades 6–8	12%	8%	76%	4%

Table P3. Demographic Composition of the Panel, Years of Teaching, and Special Education and ELL Experience

	Years of Experience Mean	Years of Experience SD	Special Ed	ELL	Both Special Ed and ELL
ELA Grades 3–5	20	8	25%	8%	42%
ELA Grades 6–8	18	10	18%	18%	41%
Math Grades 3–5	18	10	17%	4%	48%
Math Grades 6–8	18	9	44%	4%	28%

Table P4. Demographic Composition of the Panel, Size and Location of the District

	Size of the District			Location of the District			
Subject & Grades	Small	Medium	Large	Rural	Suburban	Urban	Urban/Suburban
ELA Grades 3–5	21%	29%	50%	21%	42%	38%	0%
ELA Grades 6–8	23%	36%	41%	14%	45%	36%	5%
Math Grades 3–5	20%	27%	55%	9%	55%	32%	5%
Math Grades 6–8	20%	24%	56%	12%	36%	48%	4%

Table P5. Demographic Composition of the Panel, Geographic Location of the District

Subject & Grades	North Country	Long Island	NYC	Low/Mid Hudson	Capital Region	Central NY	Western NY	Big Four
ELA Grades 3–5	13%	16%	17%	4%	8%	8%	17%	17%
ELA Grades 6–8	5%	9%	27%	9%	14%	5%	14%	17%
Math Grades 3–5	0%	13%	13%	13%	9%	13%	9%	30%
Math Grades 6–8	4%	20%	20%	8%	12%	8%	16%	12%

Standard-setting Materials

A variety of materials were provided to participants to support the standard-setting process. The Pearson psychometric team, in conjunction with the NYSED, prepared these materials. Two TAC members and one external consultant with extensive standard-setting experiences reviewed the standard-setting material during the preparation phase. The materials included:

- Performance Level Descriptors (PLDs)—the specific knowledge, skills, and practices associated with each performance level for a given grade and content area
- Ordered Item Book (OIB)—a compilation of the 2013 operational test items and selected embedded field-test items ordered from the easiest item to the most difficult item; the OIB is one of the key materials used for the Bookmark standard-setting process
- external benchmark results—using the New York State adopted empirical definition for college readiness, along with the 2011 NAEP test results
- participant rating sheets—the document upon which the participants record recommendations
- practice materials—a sample OIB (5 items) provided for participants to practice the rating task before making actual standard-setting recommendations
- readiness survey—a brief questionnaire provided to participants before each round of the standard-setting process in which participants are asked to verify that they understand the task and are ready to move forward

In addition to these materials, participants were provided with (or given access to) a variety of supplemental documents intended to inform their recommendations, including content standards, test specifications, and operational test books.

In the section labeled "Standard-setting Materials," selected materials from the standard-setting were included.

Standard-setting Activities

Large Group Training

On the morning of Day 1, all participants were convened together for introductions and training. The Commissioner, Dr. John B. King, welcomed the participants and introduced the staff, including the NYSED and the Regents' Research Fund. The Commissioner described the new assessment system, the intended rigor of the Common Core State Standards, the empirical definition of college readiness, and New York State's reform agenda. Next, Pearson described the purpose of standard-setting and provided a general overview, which included the rationale and the context for setting standards and the method to be used.

Breakdown of the Group

After the large-group session, participants were broken into grade-level groups and assigned to separate meeting rooms for the remainder of the standard-setting activities. As participants arrived in their rooms, they were directed to one of five pre-assigned tables. To help ensure diversity across tables, participants were assigned in consideration of gender, ethnicity, years of teaching, school district, and current educational role (e.g., teacher, administrator, higher education, etc.). In addition, each table had one "table leader" who had been previously selected by the NYSED. Table leaders were expected to keep track of the table-level discussion and represent their table's point of view during the vertical-articulation meeting on the final day, Day 5. Table leaders were briefly trained about their roles and responsibilities over lunch on Day 1.

Each committee recommended standards for a total of three grades. Each committee started with the highest grade in its grade cluster, moved to the middle grade, and finally completed the lowest grade. The same process was used to recommend the standards for all three grades and is described below.

Standard-setting Process

Discussion of PLDs and Development of Threshold Descriptions

After the completion of confidentiality forms and participant introductions, the facilitator initiated a discussion about PLDs. The facilitator described the role of the specific PLDs (i.e., to describe the expectations for student performance at each performance level given the grade-level demands of the Common Core standards) and reiterated that the PLDs were developed in consideration of the Common Core State Standards and the associated expectations.

Next, the facilitator introduced the concept of the threshold and what qualities a student's performance at this level possesses. This is a student who has just enough knowledge, skills, and practices to achieve a given performance level. The facilitator noted that there is a range of achievement represented by the students within each performance level, and the facilitator asked

the participants to think about the ways in which a student's achievement at the threshold of a given level may differ from a student's achievement in the middle or the top of a given level.

After group discussion of this concept, each table was asked to generate several statements that best characterized its expectations for a typical student's achievement at the threshold of each performance level given the provided PLDs. Since PLDs were defined by content standard, each table was assigned a few content standards to consider during the process. Participants at each table began their discussion by thinking about a typical student's achievement at the threshold of the New York State Level 3 performance level. This same task was then followed by students at the New York State Level 4 and New York State Level 2 threshold achievement. Each tables' drafts of threshold descriptions were reviewed by the entire group, and a consensus was reached. The threshold descriptors were then made available for the duration of the standard-setting meeting.

Review Test Items

Test books used in the 2013 operational assessments were provided to the participants. Participants were also provided a focus list (those items that the committee was asked to specifically review) that was selected by content experts to be representative in content and rigor of the full assessment. About 50% of the items were identified on the focus list, and participants were asked to review these items first. If there was time remaining, they were asked to review the remainder of the items in the test books. Through this exercise, the participants were able to experience the items on the assessment and to see the rigor and design of the test items.

Articulation of the Expectations

External-benchmark data were then presented to the participants. These external-benchmark data were conducted by researchers at the College Board. The empirical definition of college readiness, based on NYSED's policy decision, is described as:

- ELA: 0.75 probability of a B- or higher in college level credit bearing courses; and
- Math: 0.60 probability of a C+ or higher in college level credit bearing courses in math.

Using both SAT and PSAT data, along with the validity data based from College Board (freshmen GPA in 150 plus colleges across the country), the researchers conducted an analysis to identify, using both national and New York State student samples from the 2010 cohort, what the "cut scores" on SAT and PSAT were, based on the empirical definition of college readiness. Next, the percentage of students at or above these "cut scores" was identified. Figures F1 and F2 present the results from the external-benchmark studies.

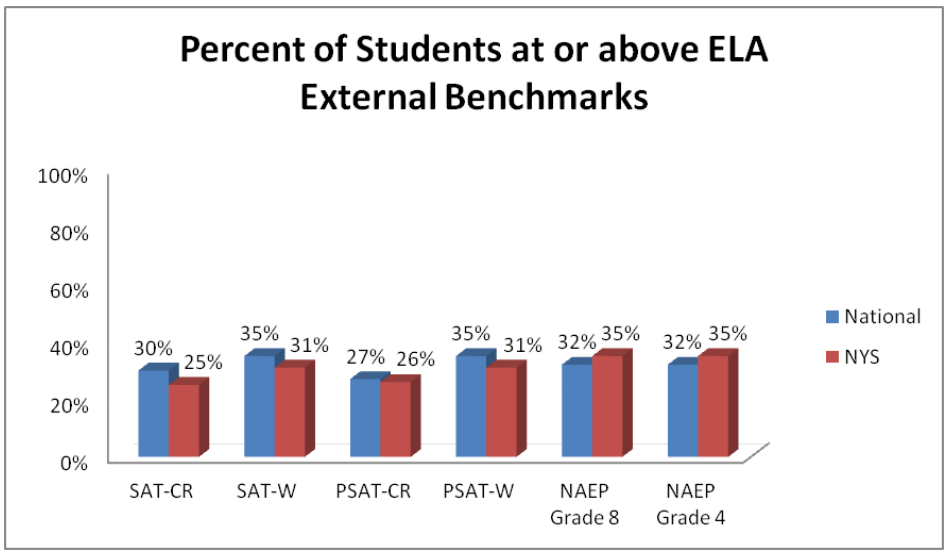

Figure O1. External-Benchmark Study Results for ELA

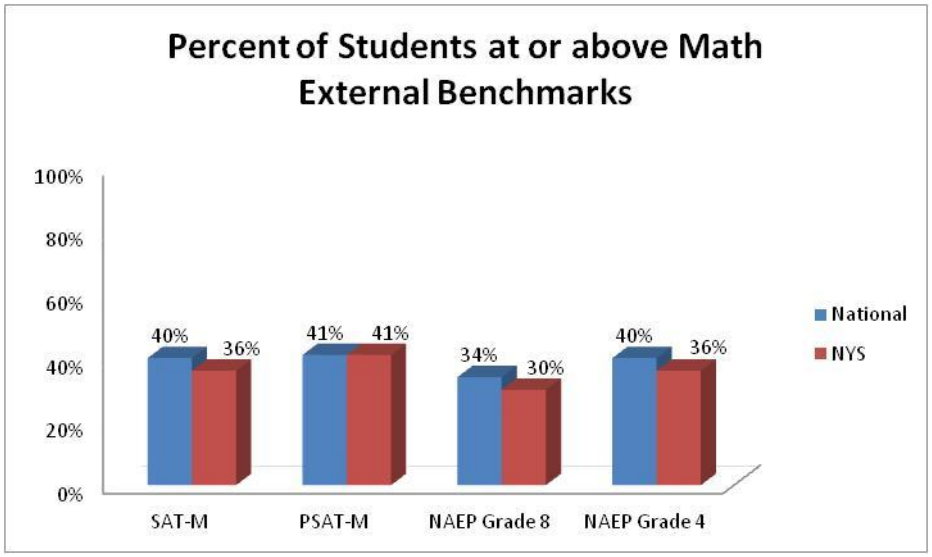

Figure O2. External-Benchmark Study Results for Mathematics

The facilitator described the attributes of the data, including their limitations. Panelists were asked to articulate performance expectations for the students given the PLDs, the threshold descriptions, the review of test items, and external benchmarks. This is a modified version of the Hofstee method where panelists articulated expectations without looking at the achievement test data. The rating sheets were collected and the results were summarized. See the section labeled "Standard-setting Materials" for the rating sheet, and the section labeled "Full Results from Standard-setting" for the results for the articulation of expectation activity. These results were then presented to the participants, along with impact data, to provide an additional artifact of evidence to be used in the standard-setting activities.

Standard-setting Training and Practice
The facilitator reiterated and expanded on the standard-setting training provided during the morning's large-group session. The complete training session included the following:

- a recap of the purpose and goal of the standard-setting meeting
- a detailed description of the format and content of the OIB
- a detailed description of the standard-setting methodology–the bookmark method
- a description of the process and strategy to use when making recommendations
- instructions on how to record ratings onto the rating sheet

After training, participants were provided with a mock OIB to practice implementing the standard-setting procedure. The practice set allowed participants to conceptualize, practice, and discuss the standard-setting process before the actual standard-setting.

Bookmark Rounds
After the training and practice exercises, the participants began the recommendation process using the bookmark methodology.

The participants completed Round 1 recommendations for the highest grade in their grade band by the end of Day 1. At the beginning of the second day, participants were provided with table-level feedback on their Round 1 recommendations, including the minimum, maximum, mean, and median recommendation associated with each level. Each table was instructed to discuss its Round 1 recommendations with the goal of identifying major sources of variance among participants. Understanding, rather than consensus, was the ultimate goal of the discussion. The participants started the discussion with the New York State Level 3, then moved to New York State Level 4, and finally moved to New York State Level 2.

After the table-level discussion, each participant was also presented with a p-value report (a report that indicates the difficulty of items), which provided the percentage of students who answered each item correctly. The facilitator informed the committee that the intent of the p-value report was to help participants validate their conceptions around the difficulty of items (when needed) and should only be used as a reference.

After this discussion, the impact data based on the Round 1 recommendations were presented to the participants. The facilitator assisted the participants with interpreting the data. In addition, the following comparisons were made:

- impact data based on Round 1 and external benchmarks; and
- impact data based on Round 1 and performance expectation articulation (completed before the standard-setting rounds were presented).

Participants were given time to react to the data and have a group discussion.

Finally, the facilitator asked the participants to complete a readiness survey for Round 2 and then complete Round 2 ratings.

A similar process following Round 1 was followed once Round 2 ratings were available:

- table-level discussion on Round 2 ratings
- impact data provided based on the Round 2 ratings

- comparison between the impact data, external benchmarks, and performance expectations
- readiness survey for the next round

Once participants completed Round 3, they moved on to the middle grade in their grade band (i.e., Grades 7 and 4), and went through the entire process again completing three rounds. The participants then completed the same process for the lowest grade in their assigned grade band (i.e., Grades 6 and 3).

After all three rounds for the three grades assigned, participants were shown the Round 3 results across all Grades 3–8 for mathematics or ELA. The participants were then given the opportunity to discuss the impact data associated with these recommendations for all grades in a given subject area. Panelists then completed the readiness survey before they made their final recommendations at Round 4.

Exit Survey
The participants were asked to complete an exit survey asking them about their specific experience about the standard-setting, as well as their comfort level with the final recommendations.

Vertical Articulation
A subset of the participants (17–18 people for each subject) who attended the first four days of standard-setting participated in the vertical-articulation meeting on Day 5. The purpose of the vertical-articulation meeting was to review the impact data associated with the recommended cut scores across all grades. The participants were asked to determine if the recommended cut scores were reasonable given the set of expectations outlined at each grade, the test-taking population, the external benchmarks, and the skills/tasks presented on the assessments. There was a vertical-articulation committee for mathematics, and a separate committee for ELA.

The steps in the vertical-articulation process were as follows:
1. Participants reviewed the PLDs associated with all grades within the assessment.
2. As a group, the participants discussed their expectations for impact across the grade levels in relationship to the PLDs and the content assessed in each grade.
3. The group reviewed the impact data associated with the recommended cut scores at Round 4 at Grades 3–8, and then discussed the extent to which the data matched its expectations.
4. As a group, the committee discussed if the cut scores should be adjusted to provide for impact more consistent with its expectations.
5. If adjustments were deemed necessary, participants were provided constraints on how much they could move the cut scores (4 raw-score points, which was the rounded overall test's standard error of measurement).
6. Participants made independent recommendations on raw-score cuts across grades. Participants were reminded that the goal was to make a recommendation that considered *both* the content-based ratings (from Round 4) and their expectations. They were also instructed to consult on the mapping between raw scores, and their corresponding page numbers on the OIB, when considering their recommendations.

7. Recommendations were processed, and the results were presented to the participants for discussions.
8. The participants were asked to discuss whether the median of the cut recommendations contained appropriately represented expectations for the test-taking population. The associated impact to the percent of students placed in each performance category on the 2013 operational assessment was also presented and discussed. Participants were asked to make a consensus recommendation after this discussion.
9. Participants completed evaluations of the process during the vertical-articulation meeting.

Final Recommendations Resulting from the Standard-setting Process

Panelists recommended scores on page numbers in the OIB during the first four days of standard-setting activities. The underlying theta value (using a response probability of 0.67) corresponding to the page number was identified, in addition to the theta value corresponding to the page number immediately after it. The median of the two theta values (the recommendation page and its immediate higher page number) was obtained and considered the cut score theta from the panel recommendation. Raw scores associated with the median theta (round up rule is used) and their associated impact are presented in Table P6 and Table P7 based on Round 4 results. Table P8 and Table H9 present the consensus results from vertical articulation. The section labeled "Full Results from Standard-setting" provides all the results across all rounds.

Table P6. Raw Score Cuts Associated with Committee Recommendations—Round 4

Subject	Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
ELA	Level 2	28	27	34	35	36	37
	Level 3	35	38	45	47	49	49
	Level 4	46	45	54	53	57	55
Math	Level 2	29	28	33	31	28	29
	Level 3	44	43	44	50	46	49
	Level 4	52	55	56	61	60	62

Table P7. Percentage of Students in Each Performance Level Based on Standard-setting Recommendations—Round 4

Subject	Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
ELA	Level 1	45	33	39	36	35	34
	Level 2	24	36	30	34	37	35
	Level 3	27	21	22	16	20	18
	Level 4	4	9	8	14	8	13
Math	Level 1	30	29	45	29	38	31
	Level 2	35	35	25	40	34	41
	Level 3	20	24	21	18	20	20
	Level 4	15	13	9	12	7	7

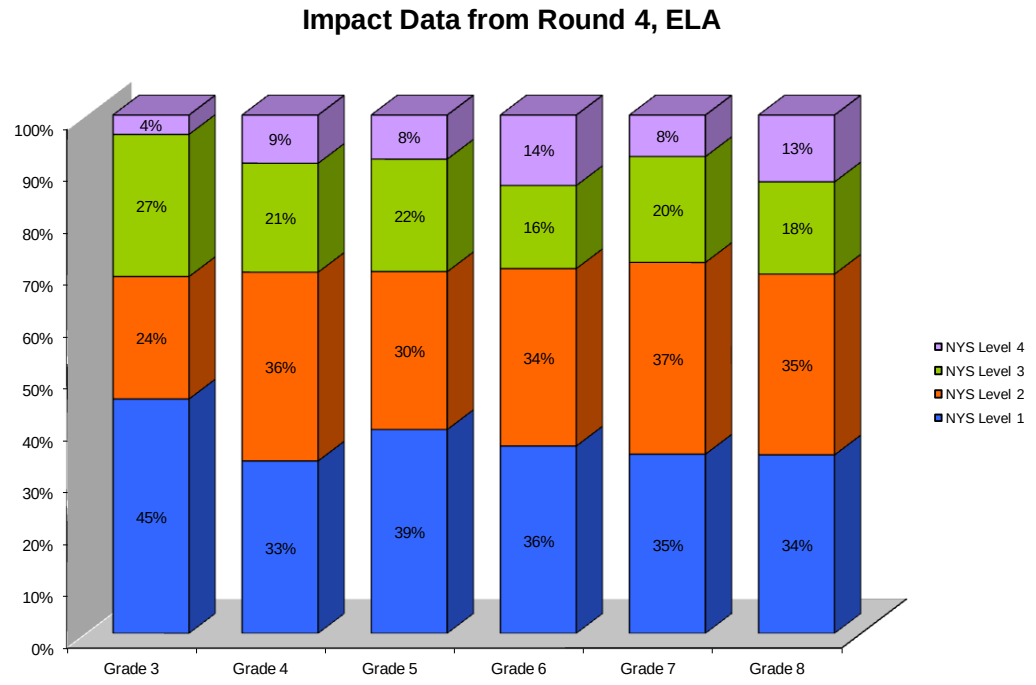

Figure O3. Percentage of Students in Each Performance Level Based on Standard-setting Recommendations —ELA (Round 4)

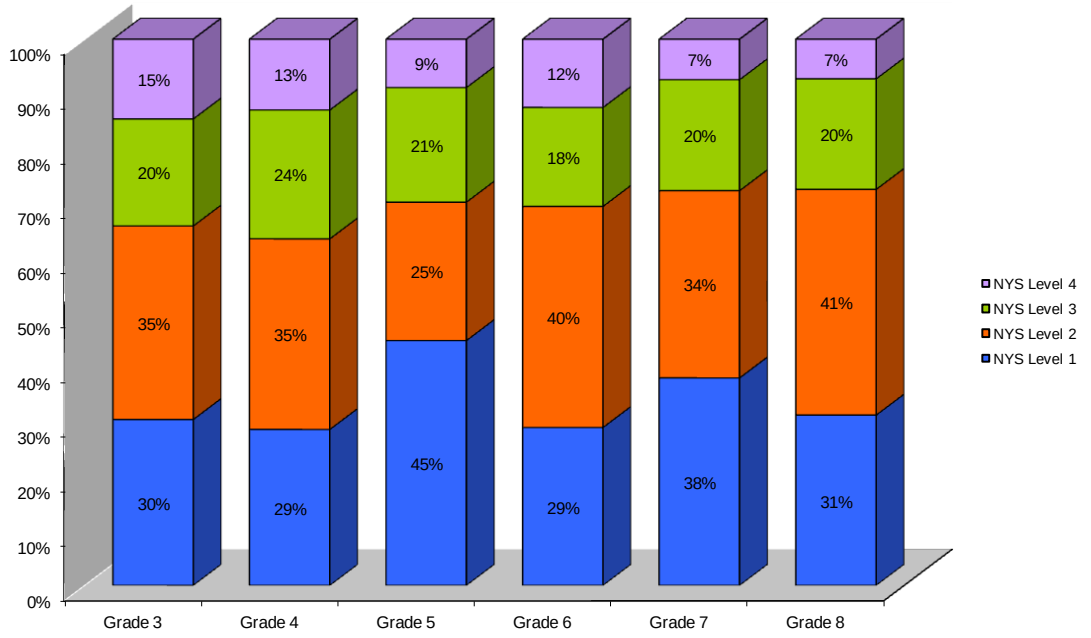

Figure O4. Percentage of Students in Each Performance Level Based on Standard-setting Recommendations —Math (Round 4)

Table P8. Raw Score Cuts Associated with Committee Recommendations—Vertical Articulation

Subject	Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
ELA	Level 2	25	26	32	32	35	35
	Level 3	35	38	45	47	48	48
	Level 4	46	45	54	53	57	56
Math	Level 2	29	28	31	31	28	29
	Level 3	44	43	44	50	46	49
	Level 4	53	55	56	61	60	62

Table P9. Percentage of Students in Each Performance Level Based on Standard-setting Recommendations —Vertical Articulation

Subject	Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
ELA	Level 1	36	30	34	29	32	30
	Level 2	33	39	36	42	37	37
	Level 3	27	21	22	16	23	23
	Level 4	4	9	8	14	8	10
Math	Level 1	30	29	40	29	38	31
	Level 2	35	35	30	40	34	41
	Level 3	22	24	21	18	20	20
	Level 4	12	13	9	12	7	7

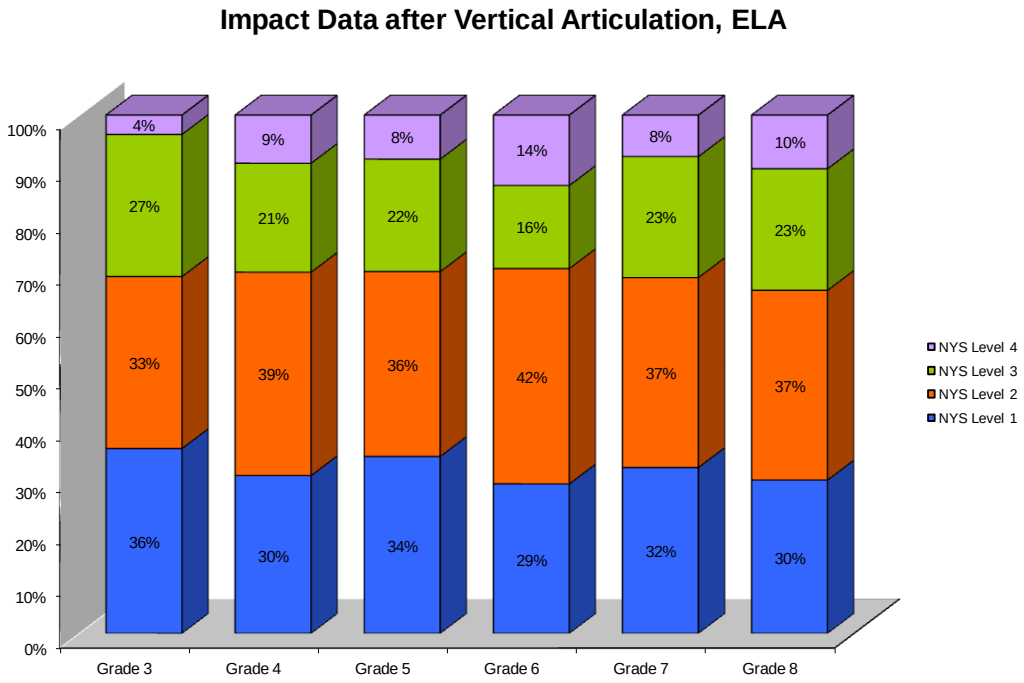

Figure O5. Percentage of Students in Each Performance Level Based on Standard-setting Recommendations —ELA (Vertical Articulation)

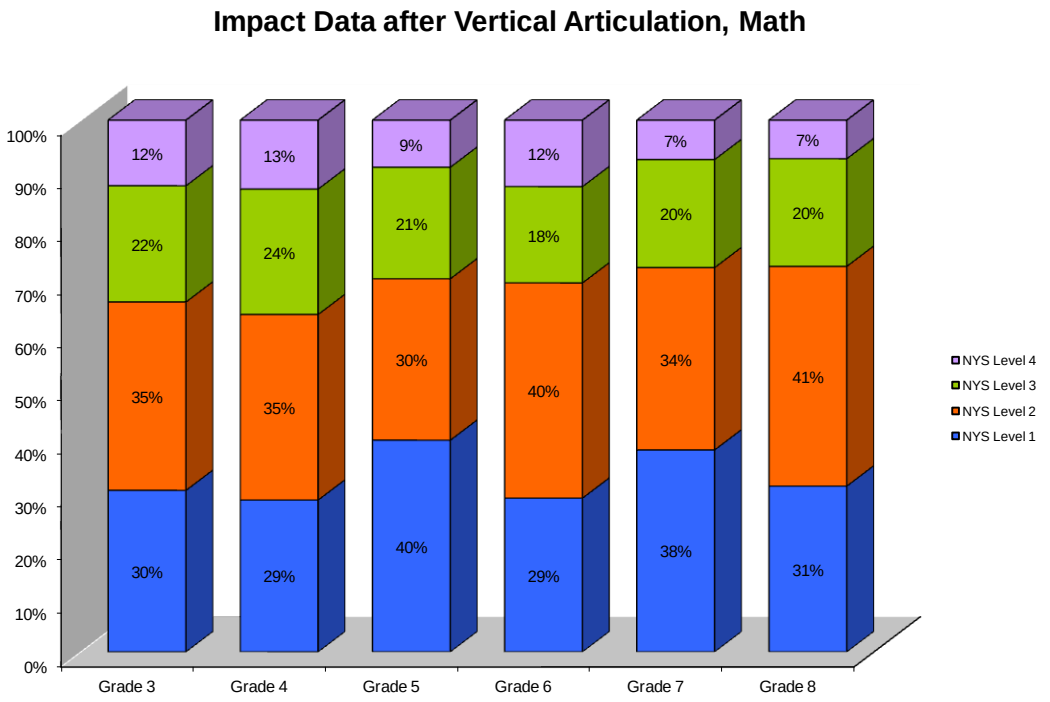

Figure O6. Percentage of Students in Each Performance Level Based on Standard-setting Recommendations —Math (Vertical Articulation)

Exit Survey

Panelists were asked to fill out an evaluation on the basis of the portion(s) of the conference they attended. Tables 10–15 present evaluation summaries for all the committee meetings. Looking at the evaluation results in general, panelists appeared to have been comfortable with the process, understood its purpose, and had the ability to follow its steps. Feedback and facilitation of the meetings were helpful in guiding them through recommendations. Participants' opinions were valued, and they generally felt comfortable expressing their opinions and sharing their perspectives. Participants generally felt comfortable with the final recommendations. We also received many additional comments where people expressed their appreciation for participating in the standard-setting activities.

Table P10. Evaluation Results, ELA Grades 3–5

Survey Question	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.
I clearly understood the purpose of the meeting.	18	4	1	0	0
The meeting facilitators clearly explained the tasks I needed to complete.	9	13	0	0	0
The achievement-level descriptions were clear to me.	16	7	0	0	0
The activities in which we clarified the achievement-level descriptors helped me to recommend cut scores.	10	12	1	0	0
The activities in which we developed threshold descriptions helped me to recommend cut scores.	9	14	1	0	0
The examples and exercises helped me to understand the process.	14	7	1	1	0
I was able to understand the external benchmark study data presented.	11	8	4	1	0
The panelists' discussions after the rounds for the three grades were helpful to me.	18	5	1	0	0
The panelists' agreement data presented after the rounds for the three grades were helpful to me.	17	6	1	0	0
The impact data (percentages of students at or above the suggested cut scores) presented based on the results for all three grades were helpful to me.	18	4	2	0	0
I felt comfortable expressing my opinions.	18	5	1	0	0
Everyone was given the opportunity to express his/her opinions.	19	4	0	0	0
I could clearly distinguish between levels of achievement.	11	11	2	0	0
Group discussions were helpful and relevant.	19	2	1	0	0

Table P10. Evaluation Results, ELA Grades 3–5 (cont.)

Survey Question	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.
The pace of the workshop was appropriate (not too fast or too slow).	5	9	8	1	0
I believe the standard-setting workshop was fair and unbiased.	8	13	2	0	0
I would defend the standards recommended by our committee.	11	9	3	0	0
The final group-recommended NYS Level 2 cut score fairly represents the minimal level of achievement for students at NYS Level 2.	6	11	5	0	0
If you answered Disagree or Strongly Disagree to question 18, do you believe the final group-recommended cut score is (check one):	Too High: 4	Too Low: 1	0	0	0
The final group-recommended NYS Level 3 cut score fairly represents the minimal level of achievement for students at NYS Level 3.	10	11	2	0	0
If you answered Disagree or Strongly Disagree to question 20, do you believe the final group-recommended cut score is (check one):	Too High: 1	Too Low: 1	0	0	0
The final group-recommended NYS Level 4 cut score fairly represents the minimal level of achievement for students at NYS Level 4.	6	13	4	0	0
If you answered Disagree or Strongly Disagree to question 22, do you believe the final group-recommended cut score is (check one):	Too High: 3	Too Low: 1	0	0	0

Table P11. Evaluation Results, ELA Grades 6–8

Survey Question	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.
I clearly understood the purpose of the meeting.	20	1	0	0	0
The meeting facilitators clearly explained the tasks I needed to complete.	18	3	0	0	0
The achievement-level descriptions were clear to me.	11	10	0	0	0
The activities in which we clarified the achievement-level descriptors helped me to recommend cut scores.	8	12	1	0	0
The activities in which we developed threshold descriptions helped me to recommend cut scores.	15	5	1	0	0
The examples and exercises helped me to understand the process.	15	6	0	0	0
I was able to understand the external benchmark study data presented.	16	5	0	0	0
The panelists' discussions after the rounds for the three grades were helpful to me.	16	5	0	0	0
The panelists' agreement data presented after the rounds for the three grades were helpful to me.	12	8	0	0	0
The impact data (percentages of students at or above the suggested cut scores) presented, based on the results for all three grades, were helpful to me.	15	5	1	0	0
I felt comfortable expressing my opinions.	15	6	0	0	0
Everyone was given the opportunity to express his/her opinions.	16	5	0	0	0
I could clearly distinguish between levels of achievement.	9	11	1	0	0
Group discussions were helpful and relevant.	15	6	0	0	0

Table P11. Evaluation Results, ELA Grades 6–8 (cont.)

Survey Question	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.
The pace of the workshop was appropriate (not too fast or too slow).	9	10	1	0	0
I believe the standard-setting workshop was fair and unbiased.	8	11	1	1	0
I would defend the standards recommended by our committee.	8	10	3	0	0
The final group-recommended NYS Level 2 cut score fairly represents the minimal level of achievement for students at NYS Level 2.	5	10	5	1	0
If you answered Disagree or Strongly Disagree to question 18, do you believe the final group-recommended cut score is (check one):	Too High: 5	Too Low: 0	0	0	0
The final group-recommended NYS Level 3 cut score fairly represents the minimal level of achievement for students at NYS Level 3.	6	11	4	0	0
If you answered Disagree or Strongly Disagree to question 20, do you believe the final group-recommended cut score is (check one):	Too High: 3	Too Low: 0	0	0	0
The final group-recommended NYS Level 4 cut score fairly represents the minimal level of achievement for students at NYS Level 4.	6	11	3	1	0
If you answered Disagree or Strongly Disagree to question 22, do you believe the final group-recommended cut score is (check one):	Too High: 2	Too Low: 2	0	0	0

Table P12. Evaluation Results, Math Grades 3–5

Survey Question	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.
I clearly understood the purpose of the meeting.	20	4	0	0	0
The meeting facilitators clearly explained the tasks I needed to complete.	23	1	0	0	0
The achievement-level descriptions were clear to me.	15	8	1	0	0
The activities in which we clarified the achievement-level descriptors helped me to recommend cut scores.	15	9	0	0	0
The activities in which we developed threshold descriptions helped me to recommend cut scores.	21	3	0	0	0
The examples and exercises helped me to understand the process.	23	1	0	0	0
I was able to understand the external benchmark study data presented.	15	8	0	0	0
The panelists' discussions after the rounds for the three grades were helpful to me.	20	3	1	0	0
The panelists' agreement data presented after the rounds for the three grades were helpful to me.	18	6	0	0	0
The impact data (percentages of students at or above the suggested cut scores) presented, based on the results for all three grades were helpful to me.	12	10	1	0	0
I felt comfortable expressing my opinions.	22	2	0	0	0
Everyone was given the opportunity to express his/her opinions.	14	7	3	0	0
I could clearly distinguish between the levels of achievement.	15	9	0	0	0
Group discussions were helpful and relevant.	19	5	0	0	0

Table P12. Evaluation Results, Math Grades 3–5 (cont.)

Survey Question	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.
The pace of the workshop was appropriate (not too fast or too slow).	13	6	5	0	0
I believe the standard-setting workshop was fair and unbiased.	16	7	1	0	0
I would defend the standards recommended by our committee.	19	5	0	0	0
The final group-recommended NYS Level 2 cut score fairly represents the minimal level of achievement for students at NYS Level 2.	19	4	1	0	0
If you answered Disagree or Strongly Disagree to question 18, do you believe the final group-recommended cut score is (check one):	Too High: 1	Too Low: 0	0	0	1
The final group-recommended NYS Level 3 cut score fairly represents the minimal level of achievement for students at NYS Level 3.	19	5	0	0	0
If you answered Disagree or Strongly Disagree to question 20, do you believe the final group-recommended cut score is (check one):	Too High: 0	Too Low: 0	0	0	1
The final group-recommended NYS Level 4 cut score fairly represents the minimal level of achievement for students at NYS Level 4.	18	5	0	1	0
If you answered Disagree or Strongly Disagree to question 22, do you believe the final group-recommended cut score is (check one):	Too High: 0	Too Low: 0	0	0	1

Table P13. Evaluation Results, Math Grades 6–8

Survey Question	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.
I clearly understood the purpose of the meeting.	23	4	0	0	0
The meeting facilitators clearly explained the tasks I needed to complete.	23	4	0	0	0
The achievement-level descriptions were clear to me.	11	14	1	1	0
The activities in which we clarified the achievement-level descriptors helped me to recommend cut scores.	15	11	1	0	0
The activities in which we developed threshold descriptions helped me to recommend cut scores.	19	8	0	0	0
The examples and exercises helped me to understand the process.	21	6	0	0	0
I was able to understand the external benchmark study data presented.	17	8	0	2	0
The panelists' discussions after the rounds for the three grades were helpful to me.	19	7	0	1	0
The panelists' agreement data presented after the rounds for the three grades were helpful to me.	19	7	1	0	0
The impact data (percentages of students at or above the suggested cut scores) presented, based on the results for all three grades, were helpful to me.	14	11	1	1	0
I felt comfortable expressing my opinions.	24	3	0	0	0
Everyone was given the opportunity to express his/her opinions.	22	4	0	1	0
I could clearly distinguish between the levels of achievement.	11	14	2	0	0
Group discussions were helpful and relevant.	22	4	1	0	0

Table P13. Evaluation Results, Math Grades 6–8 (cont.)

Survey Question	Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.
The pace of the workshop was appropriate (not too fast or too slow).	13	10	3	1	0
I believe the standard-setting workshop was fair and unbiased.	18	9	0	0	0
I would defend the standards recommended by our committee.	17	9	1	0	0
The final group-recommended NYS Level 2 cut score fairly represents the minimal level of achievement for students at NYS Level 2.	14	10	3	0	0
If you answered Disagree or Strongly Disagree to question 18, do you believe the final group-recommended cut score is (check one):	Too High: 3	Too Low: 1	0	0	0
The final group-recommended NYS Level 3 cut score fairly represents the minimal level of achievement for students at NYS Level 3.	12	15	0	0	0
If you answered Disagree or Strongly Disagree to question 20, do you believe the final group-recommended cut score is (check one):	Too High: 2	Too Low: 1	0	0	0
The final group-recommended NYS Level 4 cut score fairly represents the minimal level of achievement for students at NYS Level 4.	12	14	0	0	0
If you answered Disagree or Strongly Disagree to question 22, do you believe the final group-recommended cut score is (check one):	Too High: 2	Too Low: 1	0	0	0

Table P14. Evaluation Results, Vertical Articulation, ELA

Survey Question	Too Little Time		About Right		Too Much Time
1. To what extent was the length of this meeting appropriate for completing the vertical articulation?	3	3	10	1	0
	Not at all Accurate		Somewhat Accurate		Extremely Accurate
2. To what extent do you believe the Round 4 impact accurately reflected the percentage of students that should be classified in each level across grades?	0	4	6	6	0
	Not at all Comfortable		Somewhat Comfortable		Extremely Comfortable
3. What was your level of comfort with the vertical articulation impact rating task?	1	4	4	6	2
	Not at all Comfortable		Somewhat Comfortable		Extremely Comfortable
4. How comfortable are you with the final group-level recommendations?	0	4	3	7	1
5. How influential were the following factors in determining your impact recommendations?	Not at all Influential		Somewhat Influential		Very Influential
A. The Round 4 impact data	0	3	6	5	3
B. Other panelists' comments/group discussion	1	1	3	11	1
C. Other data points (external benchmark study results)	1	3	4	8	0
D. Vertical scale alignment of recommended cut scores	0	1	6	10	0
E. My professional experience	0	1	1	8	7
F. The expected rigor of the common core standards	1	0	3	8	5

Table P15. Evaluation Results, Vertical Articulation, Math

Survey Question	Too Little Time		About Right		Too Much Time
1. To what extent was the length of this meeting appropriate for completing the vertical articulation?	0	1	11	3	2
	Not at all Accurate		Somewhat Accurate		Extremely Accurate
2. To what extent do you believe the Round 4 impact accurately reflected the percentage of students that should be classified in each level across grades?	1	0	3	9	3
	Not at all Comfortable		Somewhat Comfortable		Extremely Comfortable
3. What was your level of comfort with the vertical articulation impact rating task?	1	0	3	7	6
	Not at all Comfortable		Somewhat Comfortable		Extremely Comfortable
4. How comfortable are you with the final group-level recommendations?	1	0	3	6	7
5. How influential were the following factors in determining your impact recommendations?	Not at all Influential		Somewhat Influential		Very Influential
A. The Round 4 impact data	0	1	5	5	6
B. Other panelists' comments/group discussion	0	3	7	3	4
C. Other data points (external benchmark study results)	0	2	7	4	4
D. Vertical scale alignment of recommended cut scores	0	1	5	5	5
E. My professional experience	0	1	2	7	7
F. The expected rigor of the common core standards	0	0	2	3	12

Guidance from the Technical Advisory Committee

Throughout the standard-setting process, representatives from the New York State Technical Advisory Committee (TAC) were involved in the various activities. In the following, we describe the different activities that had the various levels of TAC involvement:

- In November 2011, the standard-setting plan was presented to the entire TAC to obtain their feedback and suggestions.
- TAC provided feedback related to the standard-setting process, as well as the use of external-benchmark data to inform performance standards. The feedback was incorporated into an updated standard-setting plan.
- The updated plan was again presented to TAC for additional feedback and updates. Standard-setting was a TAC agenda topic for all TAC meetings between November 2011 and March 2013.
- A subcommittee of the TAC members, consisting of well-known scholars, as well as representatives from CUNY, were invited to have regular conversions with the NYSED psychometric team and the Pearson psychometric team to hone in on the details of the external-benchmark data, and how that can be used to inform performance standards.
- Two TAC members and an external consultant from Center for Assessment also were involved in carefully reviewing all the standard-setting materials developed for the meeting. Their feedback was incorporated into the final standard-setting materials used for the meeting.
- Additionally, two TAC members were invited to attend the standard-setting meeting. One of them attended the entire five days of the standard-setting meeting, and the other TAC member attended the first two days of the standard-setting meeting. The two TAC members provided the following statement regarding the standard-setting process:

  > *"In observing the training for the NY State Grades 3–8 ELA and Math Common Core Tests Standard-setting on June 28, 2013, we were comfortable that the facilitators were following best practices in implementing research-based procedures. After observing a full standard-setting session, we are confident that the recommended cut scores were derived using a well-implemented process that followed the plan presented to the NY technical advisory committee (TAC)."*
  >
  > Marianne Perie and Michael Rodriguez
  > New York State TAC

Final Adopted Performance Standards

The Commissioner of Education in New York State was presented with the standard-setting panel's Round 4 recommendations, the vertical-articulation panel recommendations, external-benchmark study results, and the first three round results. After discussion with his advisers, Commissioner King made the final decision of adopting the recommendations from the vertical-articulation panel. These performance standards were presented to the Board of Regents in July 2013, and were also formally approved at the Board of Regents meeting.

Standard-setting Nomination Letters

January 4, 2013

Dear Colleagues:

The New York State Education Department (NYSED) is seeking nominations for teachers to serve on the Grades 3-8 Common Core English Language Arts (ELA) or the Mathematics Tests Standard Setting Committee. The Committees will meet on Saturday, June 29 through Tuesday, July 2, 2013, from 8:00 a.m. to 5:00 p.m. each day, at a hotel in the Albany area. A subset of the standard setting committees will be identified as Table Leaders. Table Leaders will also be required to participate in the vertical articulation process on Wednesday, July 3, 2013 from 8:00 a.m. to 3:00 p.m.

The purpose of each Standard Setting Committee is to discuss the performance standards and to provide recommendations for the performance level cut scores for the Grades 3-8 Common Core ELA and Mathematics Tests. The Committee members will follow a standard setting process, discuss performance expectations for students in each performance level, review secure test items, and make recommendations to the Commissioner. Further information regarding the standard setting process is attached. The nominees will serve the Department in an advisory capacity. The final decision regarding performance standards for the New York State Common Core Tests will be determined by the Commissioner of Education, taking into consideration the recommendations from the Standard Setting Committees.

The nominees must have content expertise and current teaching experience in grades 3-8 teaching math and/or ELA, English language learners, or students with disabilities, or be a curriculum specialist or teach in higher education. We need teachers with a deep knowledge of the Common Core standards and with experience teaching different student populations. Please make every effort to ensure that nominees represent the diversity of New York State, and include minorities, urban/rural/suburban schools, and the various geographic areas of the state.

In an effort to recruit a balanced, representative committee, not all nominees will be selected. Those nominees not selected for the standards setting event will be included in a pool from which NYSED staff may select individuals for participation in future events related to the New York State Testing Program. Nominees **must** attend **all** days of the meeting and commit to the work. To ensure security and confidentiality of any discussions and materials shared at this meeting, Committee members must sign non-disclosure agreements.

Please submit the names of **ten** nominees to Candace Shyer, Director, via e-mail (cshyer@mail.nysed.gov) by **January 18, 2013**. In addition, the nominees **must** fill out an on-line recruitment form posted at https://www.research.net/s/C22X6N7.

Participants will be notified in early February if they are selected to participate in the Standard Setting meeting. Once selected, the Department's contractor, Pearson, will contact participants who live more than 50 miles from the meeting site, to make arrangements for lodging. Participants will be reimbursed for appropriate expenses, and the vendor will provide breakfast, lunch, and snacks during the meetings. For each of the days attended, Pearson will pay participants an honorarium of $175.

If you have any questions regarding nominations or the meetings, please contact Candace Shyer at cshyer@mail.nysed.gov or (518) 474-5900. Thank you for your help with this important effort on behalf of New York State's students.

Sincerely,

Ken Wagner

Attachment

January 8, 2013

Dear Colleagues:

I recently sent you a letter asking for nominations for teachers to serve on the Grades 3-8 Common Core English Language Arts (ELA) or the Mathematics Tests Standard Setting Committees. I am now asking you for nominations for administrators to assist the Department with the vertical articulation process of the Grades 3-8 Math and ELA standard setting. The vertical articulation meeting will be held on July 3, 2013, 8:00 a.m. to 3:00 p.m., at a hotel in the Albany area. A subset of the standard-setting committees of teachers will be identified as Table Leaders. Table Leaders will also be required to participate in the vertical articulation process on July 3, 2013 along with the administrators.

The purpose of vertical articulation is to review the impact of the recommended cut scores on student performance across all grades to see if this outcome makes sense given the performance expectations outlined at each grade, the test-taking population, the rigor associated with the content at the different grades, the skills/tasks presented on the various assessments, as well as the results of external benchmark analyses (e.g., results on other related measures, such as college readiness). The nominees will serve the Department in an advisory capacity. The final decision regarding performance standards for the New York State Common Core Tests will be determined by the Commissioner of Education, taking into consideration the recommendations from the Standard Setting Committees and from the vertical articulation meeting.

The nominees must be current grades 3-8 administrators with a deep knowledge of the Common Core standards. Please make every effort to ensure that nominees represent the diversity of New York State, and include minorities, urban/rural/suburban schools, and the various geographic areas of the state.

Not all nominees will be selected, however those nominees not selected will be included in a pool from which NYSED staff may select individuals for participation in future events related to the New York State Testing Program. Nominees **must** attend the full day of the meeting and commit to the work. To ensure security and confidentiality of any discussions and materials shared at this meeting, members must sign non-disclosure agreements.

Please submit the names of **three** nominees to Candace Shyer, Director, via e-mail (cshyer@mail.nysed.gov) by **January 22, 2013**. In addition, the nominees **must** fill out an on-line recruitment form posted at: https://www.research.net/s/C22X6N7.

The administrators will be notified in early February if they are selected to participate in Standard Setting meeting. Once selected, the Department's contractor, Pearson, will contact participants who live more than 50 miles from the meeting site, to make arrangements for lodging. Participants will be reimbursed for appropriate expenses, and the vendor will provide breakfast, lunch, and snacks during the meetings.

If you have any questions regarding nominations or the meetings, please contact Candace Shyer at cshyer@mail.nysed.gov or (518) 474-5900. Thank you for your help with this important effort on behalf of New York State's students.

Sincerely,

Ken Wagner

Attachment

The Standard Setting Process in Educational Assessments

Educational assessments are used to make a range of decisions that categorize test takers based on their performance. This may involve classifying the test takers as "pass" or "fail" or may place them into one of several achievement levels. For example, students taking New York State Grades 3-8 ELA and Mathematics tests are currently classified as Level I (Below Standards), Level II (Meets Basic Standards), Level III (Meets Proficiency Standards), or Level IV (Exceeds Proficiency Standards) to establish and communicate achievement goals. The achievement level descriptions define what students should know and be able to do when they have reached each level. Standard setting is the process where the descriptions of these achievement levels and the minimum scores necessary to be classified into each level (i.e., cutscores) are established.

In New York State, standard setting panels are made up of certified teachers who represent the state both geographically and in diversity. Additionally, teachers with diverse roles (e.g., teaching students with disabilities or English language learners) are sought to ensure that all student interests are represented. The panelists of teachers are tasked with recommending to the Commissioner of Education the minimally acceptable student performances for classification into each achievement level. These cutscores, or *standards*, serve as the minimum threshold for classifying students into a specific performance level rather than the next lowest level. Throughout the standard setting process, all procedures for establishing cutscores and decisions are documented as required by the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999). This includes information on panelist qualifications, procedures followed for setting performance standards, and the impact of the proposed cutscores on student achievement. This documentation, along with standardized procedures and expert panelists, provides evidence for the defensibility of the final cutscores.

While there are numerous methods for setting performance standards, all include a combination of technical considerations and expert judgment. The standard setting process allows panelists to incorporate their professional knowledge and experience in accordance with the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999). Panelists receive training on the standard setting process, their specific task, and the materials that they will be using. Multiple rounds of discussion allow panelists to revisit the standards they have previously set and the panelists' judgments are informed using empirical data to give feedback (e.g., percentage of students classified into each achievement level given a proposed cutscore and minimum, maximum, mean, and median cutscores proposed by the group). Once the predetermined number of rounds is completed, the final cutscores recommendations are established. Based on the recommendations from the standard setting panel and technical advisors, the Commissioner of Education establishes the final cutscores for the operational examinations.

Reference:

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Standard-setting Materials

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Agenda

Day 1: Saturday, June 29

7:30–8:00	Check-in and Breakfast
8:00–8:15	Welcome and Introductions (NYSED)
8:15–9:15	Purpose and Goal of the Standard-setting Meeting (NYSED)
9:15–9:45	General Overview of the Standard-setting Process
9:45–10:00	Assignment to Subject/Grade Groups and Move to Separate Rooms
10:00–10:15	Panelist Introductions/Overview of the Agenda
10:15–12:00	Review Performance-Level Descriptors and Develop Threshold Descriptions (Highest Grade)
12:00–1:00	LUNCH (table leader training)
1:00–1:30	Continued: Review Performance-Level Descriptors and Develop Threshold Descriptions (Highest Grade)
1:30–2:30	Panelists Review the Test (Highest Grade)
2:30–3:15	Discussion of Expectations
3:15–4:00	Training on and Practice with the Bookmark Procedure
4:00–5:00	Round 1 of Standard-setting (Highest Grade)
5:00–5:15	Check-in of Secure Materials

Day 2: Sunday, June 30

7:30–8:00	Check-in and Breakfast
8:00–10:00	Review and Discuss Round 1 Feedback in Small Groups
10:00–10:30	Round 2 Ratings (Highest Grade)
10:30–10:45	Break/Data Entry
10:45–12:00	Review and Discuss Round 2 Feedback in Small Groups
12:00–12:30	Round 3 Ratings (Highest Grade)
12:30–1:00	LUNCH
1:00–2:45	Review Performance-Level Descriptors and Develop Threshold Descriptions (Middle Grade)
2:45–3:45	Panelists Review the Test (Middle Grade)
3:45–4:15	Discussion of Expectations
4:15–5:00	Round 1 Ratings (Middle Grade)

Day 3: Monday, July 1

7:30–8:00	Check-in and Breakfast
8:00–10:00	Review and Discuss Round 1 Feedback in Small Groups
10:00–10:30	Round 2 Ratings (Middle Grade)
10:30–10:45	Break/Data Entry
10:45–12:00	Review and Discuss Round 2 Feedback in Small Groups
12:00–12:30	Round 3 Ratings (Middle Grade)
12:30–1:00	LUNCH
1:00–2:45	Review Performance-Level Descriptors and Develop Threshold Descriptions (Lowest Grade)
2:45–3:45	Panelists Review the Test (Lowest Grade)
3:45–4:15	Discussion of Expectations
4:15–5:00	Round 1 Ratings (Lowest Grade)

Day 4: Tuesday, July 2

7:30–8:00	Check-in and Breakfast
8:00–9:30	Review and Discuss Round 1 Feedback in Small Groups
9:30–10:30	Round 2 Ratings (Lowest Grade)
10:30–10:45	Break/Data Entry
10:45–12:00	Review and Discuss Round 2 Feedback in Small Groups
12:00–12:30	Round 3 Ratings (Lowest Grade)
12:30–1:30	LUNCH
1:30–2:30	Review Round 3 Recommendations for All Grades
2:30–3:00	Round 4 Ratings for All Three Grades
3:00–3:30	Break/Data Entry
3:30–4:00	Evaluation Forms and Secure Materials Check-In

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Panelist Information Sheet

Name: _____

Subject: _____ **Grade Band: _____**

Please provide the following demographic information that will be used to describe the general characteristics of the panelists who are attending the standard-setting for the New York State Grades 3–8 ELA & Math Common Core Tests.

Your Current Position:

_____

_____

Courses/Grades Taught:

_____

_____

_____

_____

Are You from Institutes of Higher Education? Yes No

Gender (circle one): Male Female

Which of the following describes your race/ethnicity?

- □ Black
- □ Hispanic
- □ Native American/Alaskan Native
- □ White
- □ Asian/Pacific Islander
- □ Other

Years of Teaching/Relevant Educational Experience:

_____

Experience with Special Populations (circle all that apply):

Special Education ELL

Compared to other school districts/colleges in New York State, how would you describe the <u>size</u> of your school district/college (circle one)?

Small Medium Large

Compared to other school districts/colleges in New York State, how would you describe the <u>location</u> of your school district/college (circle one)?

Urban Suburban Rural

Which best describes the geographic location of your school district or college (circle one)?

North Country Long Island NYC Lower and Mid-Hudson

Capital Region Central NY Western NY Big Four City (Buffalo, Yonkers, Syracuse, or Rochester)

Have you been asked to stay on the fifth day to participate in the vertical articulation panel?

Yes No

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Bookmark Method Steps

For any given performance level:

1. Read the item.
2. Identify the skills, knowledge, and practices needed for a correct response.
3. Evaluate why each item is more difficult than the preceding one.
4. Review PLDs and related threshold descriptions for the performance level.
5. Ask yourself: "**Should** threshold students who are *minimally* at the performance level have ***a 67% chance or better*** of answering this question correctly?" In other words, at least 67 out of 100 students whom you consider just at the performance level **should** answer correctly.
6. Mark the "zone" or "neighborhood"—the first "no" followed by a "yes" and the first "no" followed by only "no's."
7. Identify the last page at which a student just at the performance level should have at least a 67% probability of answering the last item correctly (the last "yes" item).
8. Mark the pages with flags.
9. Record the pages on the rating sheet.
10. Repeat for the next performance level.

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Expectation Rating Sheet

Directions: Given the external benchmark data presented, as well as your review of the test items, please provide your personal expectations in terms of the percentage of students reaching and/or surpassing NYS Level 3, the proficiency level linked with interpretations with college readiness.

Panelist ID_____ Table Number_____

Subject_____ Grade_____

Given the rigor of the Common Core State Standards, I would be surprised if

- less than _____ percent of students reach proficiency
- more than _____ percent of students reach proficiency

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Bookmark Recommendation Sheet
Rounds 1–3[2]

*Directions: For each performance level, write down the **page number** corresponding to the **last YES** item. No cells should be left blank within a given round.*

Panelist ID_____ Table Number_____

Subject_____ Grade_____

	Page Number of LAST YES Item		
	NYS Level 2	NYS Level 3	NYS Level 4
Round 1			
Round 2			
Round 3			

[2] We will use a different rating sheet for Round 4.

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Bookmark Recommendation Sheet

*Directions: For each performance level and each grade, write down the **page number** corresponding to the **last YES** item. No cells should be left blank within a given round.*

Panelist ID_____ Table Number_____

Subject_____

	Page Number of LAST YES Item		
	Grade 3	Grade 4	Grade 5
NYS Level 2			
NYS Level 3			
NYS Level 4			

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Readiness Survey

Panelist ID: _____

Subject: _____ Grade: _____

Instructions: Please circle your response to the following statements.

Round 1		
I understand my task for Round 1.	No	Yes
I am ready to begin Round 1.	No	Yes

Round 2		
I understand my task for Round 2.	No	Yes
I understand the data that were presented from Round 1.	No	Yes
I am ready to begin Round 2.	No	Yes

Round 3		
I understand my task for Round 3.	No	Yes
I understand the data that were presented from Round 2.	No	Yes
I am ready to begin Round 3.	No	Yes

Round 4		
I understand my task for Round 4.	No	Yes
I understand the data that were presented from Round 3.	No	Yes
I am ready to begin Round 4.	No	Yes

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Standard-setting Evaluation Form

Subject: _____ Grade Band: _____

Please circle the response following each statement that best indicates your experience in the standard-setting meeting.

1. I clearly understood the purpose of the meeting.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

2. The meeting facilitators clearly explained the tasks I needed to complete.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

3. The performance-level descriptors were clear to me.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

4. The activities in which we developed threshold descriptors helped me to recommend cut scores.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

5. The examples and exercises helped me to understand the process.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

6. The explanations about how to set the cut score helped me to understand what I needed to do in the rounds.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

7. I was able to understand the external benchmark study data presented.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

8. The panelist discussions after the rounds for the three grades were helpful to me.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

9. The panelist agreement data presented after the rounds for the three grades were helpful to me.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

10. The impact data (percentages of students at or above the suggested cut scores) presented based on the results for all three grades were helpful to me.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

11. I felt comfortable expressing my opinions.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

12. Everyone was given the opportunity to express his/her opinions.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

13. I could clearly distinguish between levels of achievement.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

14. Group discussions were helpful and relevant.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

15. The pace of the workshop was appropriate (not too fast or too slow).

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

16. I believe the standard-setting workshop was fair and unbiased.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

17. I would defend the standards recommended by our committee.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

18. The final group-recommended NYS Level 2 cut score fairly represents the minimal level of achievement for students at NYS Level 2.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

19. If you answered Disagree or Strongly Disagree to question 18, do you believe the final group-recommended cut score is (check one):

Too High Too Low

20. The final group-recommended NYS Level 3 cut score fairly represents the minimal level of achievement for students at NYS Level 3.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

21. If you answered Disagree or Strongly Disagree to question 20, do you believe the final group-recommended cut score is (check one):

Too High Too Low

22. The final group-recommended NYS Level 4 cut score fairly represents the minimal level of achievement for NYS Level 4.

Strongly Agree	Moderately Agree	Moderately Disagree	Strongly Disagree	N.A.

23. If you answered Disagree or Strongly Disagree to question 22, do you believe the final group-recommended cut score is (check one):

Too High Too Low

Use the remaining space on this page to provide any additional comments you have regarding the standard-setting process, facilitator, etc.

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Vertical Articulation Agenda

Wednesday, July 3

8:30–9:00	Introduction to Vertical Articulation
9:00–9:30	Review PLDs for All Grade Levels
9:30–10:00	Discuss Impact Expectations across Grades
10:00–11:30	Review/Discuss Impact Associated with Recommendations
11:30–12:00	Make Impact Recommendations (Panelists)
12:00–1:00	LUNCH
1:00–1:45	Present and Discuss Round 1 Feedback; Establish Group Consensus
1:45–2:00	Evaluations

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Vertical Articulation Readiness Survey

Panelist ID: _____ Subject: _____

Instructions: Please circle your response to the following statements.

Rating Task		
I understand the vertical articulation rating task.	No	Yes
I understand that I am to consider the raw score cuts and associated impact from Round 4 recommendations, the external benchmark study results, the vertical scaling data, and today's large group discussions when making my ratings.	No	Yes
I am ready to begin the vertical articulation rating task.	No	Yes

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Vertical Articulation Rating Sheet

Subject: _____

Directions: Your task is to recommend the raw score cuts for each performance level based on the various data and discussions from this morning. As you recommend the raw score cuts, please refer to the OIBs and be aware of how much you are moving the page numbers from the Round 4 recommendations.

Performance Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level 2						
NYS Level 3						
NYS Level 4						

New York State
Grades 3–8 ELA & Math Common Core Tests
Standard-setting

Vertical Articulation Evaluation Form

Subject: _____

1.	To what extent was the length of this meeting appropriate for completing the vertical articulation?	Too Little Time		About Right		Too Much Time
		1	2	3	4	5
2.	To what extent do you believe the **Round 4** impact accurately reflected the percentage of students that should be classified in each level across grades?	Not at all Accurate		Somewhat Accurate		Extremely Accurate
		1	2	3	4	5
3.	What was your level of comfort with the vertical articulation rating task?	Not at all Comfortable		Somewhat Comfortable		Extremely Comfortable
		1	2	3	4	5
4.	How comfortable are you with the final group-level recommendations?	Not at all Comfortable		Somewhat Comfortable		Extremely Comfortable
		1	2	3	4	5

5. How influential were the following factors in determining your recommendations?

	Not at all Influential		Somewhat Influential		Very Influential
A. The Round 4 impact data	1	2	3	4	5
B. Other panelists' comments/group discussion	1	2	3	4	5
C. Other data points (external benchmark study results)	1	2	3	4	5
D. Vertical scale alignment of recommended cut scores	1	2	3	4	5
E. My professional experience	1	2	3	4	5
F. The expected rigor of the common core standards	1	2	3	4	5

6. Please use the back of this page to provide any additional comments you may have about the vertical articulation process.

Thank you for your hard work and valuable feedback!

Full Results from Standard-setting

Table P16. Summary of the Articulated Expectations (Hofstee Method)—Minimum and Maximum Percentage of Students at Level 3 and above

Grade	ELA		Math	
	Minimum	Maximum	Minimum	Maximum
3	30	40	30	48
4	25	40	25	43
5	25	40	30	45
6	25	40	23	35
7	28	40	20	36
8	30	45	22	45

Table P17. Percentage of Students in Each Performance Level Based on Standard-setting Recommendations for All Rounds

Subject & Grade	Level	Round 1	Round 2	Round 3	Round 4
ELA Grade 3	Level 1	45	45	45	45
	Level 2	44	30	24	24
	Level 3	8	21	27	27
	Level 4	3	4	4	4
ELA Grade 4	Level 1	33	39	33	33
	Level 2	36	30	36	36
	Level 3	21	21	21	21
	Level 4	9	9	9	9
ELA Grade 5	Level 1	37	39	45	39
	Level 2	39	36	25	30
	Level 3	19	19	23	22
	Level 4	5	5	7	8
ELA Grade 6	Level 1	36	36	36	36
	Level 2	34	34	34	34
	Level 3	16	16	16	16
	Level 4	14	14	14	14
ELA Grade 7	Level 1	35	35	35	35
	Level 2	40	37	37	37
	Level 3	18	20	18	20
	Level 4	8	8	10	8
ELA Grade 8	Level 1	42	48	48	34
	Level 2	27	21	21	35
	Level 3	9	15	18	18
	Level 4	21	16	13	13

Table P17. Percentage of Students in Each Performance Level Based on Standard-Setting Recommendations for All Rounds (cont.)

Subject & Grade	Level	Round 1	Round 2	Round 3	Round 4
Math Grade 3	Level 1	30	30	30	30
	Level 2	35	35	35	35
	Level 3	20	15	20	20
	Level 4	15	19	15	15
Math Grade 4	Level 1	26	29	29	29
	Level 2	37	35	35	35
	Level 3	22	22	24	24
	Level 4	15	15	13	13
Math Grade 5	Level 1	45	40	42	45
	Level 2	35	30	28	25
	Level 3	14	21	21	21
	Level 4	6	9	9	9
Math Grade 6	Level 1	29	29	29	29
	Level 2	40	42	40	40
	Level 3	18	16	18	18
	Level 4	12	12	12	12
Math Grade 7	Level 1	38	38	38	38
	Level 2	33	33	34	34
	Level 3	22	22	20	20
	Level 4	7	7	7	7
Math Grade 8	Level 1	40	31	31	31
	Level 2	35	41	41	41
	Level 3	20	19	20	20
	Level 4	6	8	7	7

Table P18. Raw Score Cuts Based on Committee Recommendations for All Rounds

Subject & Grade	Level	Raw Score Cuts				Total Score
		Round 1	Round 2	Round 3	Round 4	
ELA Grade 3	Level 2	28	28	28	28	55
	Level 3	42	37	35	35	55
	Level 4	47	46	46	46	55
ELA Grade 4	Level 2	27	29	27	27	55
	Level 3	38	38	38	38	55
	Level 4	45	45	45	45	55
ELA Grade 5	Level 2	33	34	36	34	66
	Level 3	47	47	45	45	66
	Level 4	56	56	55	54	66
ELA Grade 6	Level 2	35	35	35	35	65
	Level 3	47	47	47	47	65
	Level 4	53	53	53	53	65
ELA Grade 7	Level 2	36	36	36	36	66
	Level 3	50	49	49	49	66
	Level 4	57	57	56	57	66

Table P18. Raw Score Cuts Based on Committee Recommendations for All Rounds (cont.)

Subject & Grade	Level	Raw Score Cuts				Total Score
		Round 1	Round 2	Round 3	Round 4	
ELA Grade 8	Level 2	40	42	42	37	66
	Level 3	49	49	49	49	66
	Level 4	52	54	55	55	66
Math Grade 3	Level 2	29	29	29	29	60
	Level 3	44	44	44	44	60
	Level 4	52	50	52	52	60
Math Grade 4	Level 2	27	28	28	28	66
	Level 3	43	43	43	43	66
	Level 4	54	54	55	55	66
Math Grade 5	Level 2	33	31	32	33	66
	Level 3	49	44	44	44	66
	Level 4	58	56	56	56	66
Math Grade 6	Level 2	31	31	31	31	72
	Level 3	50	51	50	50	72
	Level 4	61	61	61	61	72
Math Grade 7	Level 2	28	28	28	28	72
	Level 3	45	45	46	46	72
	Level 4	60	60	60	60	72
Math Grade 8	Level 2	33	29	29	29	72
	Level 3	50	49	49	49	72
	Level 4	63	61	62	62	72

Table P19. Raw Score Cuts Associated with Committee Recommendations—Vertical Articulation Round 1

Subject	Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
ELA	Level 2	25	26	32	32	35	35
	Level 3	35	38	45	47	48	48
	Level 4	46	45	54	53	57	56
Math	Level 2	29	28	31	31	28	29
	Level 3	44	43	44	50	46	49
	Level 4	52	55	56	61	60	62

Table P20. Percentage of Students in Each Performance Level Based on Standard-Setting Recommendations—Vertical Articulation Round 1

Subject	Level	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
ELA	Level 1	36	30	34	29	32	30
	Level 2	33	39	36	42	37	37
	Level 3	27	21	22	16	23	23
	Level 4	4	9	8	14	8	10
Math	Level 1	30	29	40	29	38	31
	Level 2	35	35	30	40	34	41
	Level 3	20	24	21	18	20	20
	Level 4	15	13	9	12	7	7

Appendix Q: Derivation and Estimation of Classification Consistency and Agreement

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 h^{th} 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

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

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

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

The agreement index (classification consistency) can be computed as

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

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

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

Then, Kappa (Cohen, 1960) is

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

Classification Agreement

Let Γ_w denote true category. When an examinee has an observed score, $x \in L_h$ ($h = 1, 2, \ldots, \mathrm{H}$), and a latent score , $\theta \in \Gamma_w$ ($w = 1, 2, \ldots, \mathrm{H}$), 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 agreement 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 R: Raw Score-to-Scale Score and Scale Score Frequency Tables

Tables R1–R12 show the raw score-to-scale score conversion tables, while Tables R13–R24 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 R1. ELA Raw Score-to-Scale Score Table, Grade 3

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	148	57	28	300	9
1	156	50	29	303	9
2	164	44	30	305	9
3	172	39	31	308	9
4	180	34	32	311	9
5	188	30	33	314	9
6	196	27	34	317	9
7	204	24	35	320	9
8	212	21	36	322	9
9	220	19	37	326	9
10	228	17	38	329	9
11	235	16	39	332	9
12	242	15	40	335	9
13	248	14	41	338	9
14	253	13	42	342	9
15	258	12	43	346	10
16	262	11	44	349	10
17	266	11	45	354	11
18	269	11	46	358	11
19	273	10	47	363	12
20	276	10	48	368	12
21	279	10	49	375	13
22	283	9	50	382	15
23	286	9	51	391	17
24	288	9	52	399	19
25	291	9	53	407	21
26	294	9	54	415	24
27	297	9	55	423	27

Table R2. ELA Raw Score-to-Scale Score Table, Grade 4

Raw Score	Scale Score	Standard Error
0	139	64
1	147	56
2	155	49
3	163	43
4	171	37
5	179	32
6	187	28
7	195	25
8	203	22
9	211	19
10	220	17
11	228	15
12	234	14
13	240	13
14	245	13
15	249	12
16	254	11
17	258	11
18	261	11
19	265	10
20	268	10
21	272	10
22	275	10
23	278	9
24	281	9
25	284	9
26	287	9
27	290	9

Raw Score	Scale Score	Standard Error
28	292	9
29	295	9
30	298	9
31	301	9
32	303	8
33	306	8
34	309	8
35	311	8
36	314	8
37	317	9
38	320	9
39	323	9
40	326	9
41	329	9
42	332	9
43	336	9
44	339	10
45	343	10
46	347	11
47	352	11
48	357	12
49	363	13
50	369	14
51	377	15
52	388	18
53	396	21
54	404	24
55	412	27

Table R3. ELA Raw Score-to-Scale Score Table, Grade 5

Raw Score	Scale Score	Standard Error
0	116	82
1	124	73
2	132	64
3	140	57
4	148	50
5	156	44
6	164	39
7	172	34
8	180	30
9	188	27
10	196	24
11	204	21
12	212	19
13	220	17
14	228	16
15	234	14
16	239	14
17	244	13
18	248	12
19	252	12
20	256	11
21	259	11
22	263	10
23	266	10
24	269	10
25	271	10
26	274	9
27	277	9
28	279	9
29	282	9
30	284	9
31	287	8
32	289	8
33	292	8
34	294	8
35	296	8
36	298	8
37	301	8
38	303	8
39	305	8
40	308	8
41	310	8
42	312	8
43	315	8
44	317	8
45	320	8
46	322	8
47	325	9
48	327	9
49	330	9
50	333	9
51	336	9
52	339	9
53	343	10
54	346	10
55	350	11
56	354	11
57	359	12
58	364	12
59	369	13
60	376	14
61	383	16
62	393	18
63	401	20
64	409	23
65	417	26
66	425	29

Table R4. ELA Raw Score-to-Scale Score Table, Grade 6

Raw Score	Scale Score	Standard Error
0	112	88
1	120	76
2	128	66
3	136	57
4	144	49
5	152	43
6	160	37
7	168	32
8	176	28
9	184	25
10	192	22
11	200	19
12	208	17
13	216	16
14	224	14
15	230	13
16	234	13
17	239	12
18	243	11
19	246	11
20	250	11
21	253	10
22	256	10
23	259	10
24	262	9
25	265	9
26	268	9
27	270	9
28	273	9
29	276	9
30	278	9
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	305	8
42	307	8
43	309	8
44	312	8
45	315	8
46	317	9
47	320	9
48	323	9
49	325	9
50	328	9
51	331	9
52	335	9
53	338	10
54	341	10
55	345	10
56	349	11
57	354	11
58	359	12
59	364	13
60	370	14
61	378	15
62	387	18
63	396	20
64	404	23
65	412	26

Table R5. ELA Raw Score-to-Scale Score Table, Grade 7

Raw Score	Scale Score	Standard Error
0	103	98
1	111	86
2	119	76
3	127	66
4	135	58
5	143	50
6	151	44
7	159	38
8	167	33
9	175	29
10	183	25
11	191	22
12	199	19
13	207	17
14	215	15
15	223	14
16	228	13
17	233	12
18	237	12
19	241	11
20	245	11
21	249	10
22	252	10
23	255	10
24	258	10
25	261	9
26	264	9
27	267	9
28	270	9
29	272	9
30	275	9
31	277	8
32	280	8
33	282	8
34	285	8
35	287	8
36	289	8
37	292	8
38	294	8
39	296	8
40	299	8
41	301	8
42	304	8
43	306	8
44	308	8
45	311	8
46	313	8
47	316	8
48	318	8
49	321	8
50	324	9
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	378	16
63	389	20
64	397	22
65	405	25
66	413	29

Table R6. ELA Raw Score-to-Scale Score Table, Grade 8

Raw Score	Scale Score	Standard Error
0	100	79
1	108	70
2	116	61
3	124	54
4	131	48
5	139	42
6	147	37
7	155	33
8	163	29
9	171	25
10	179	23
11	187	20
12	195	18
13	203	16
14	211	15
15	218	14
16	226	13
17	231	12
18	235	12
19	239	11
20	243	11
21	246	11
22	249	11
23	253	10
24	256	10
25	259	10
26	261	10
27	264	9
28	267	9
29	269	9
30	272	9
31	275	9
32	277	9
33	279	9
34	282	8
35	284	8
36	286	8
37	289	8
38	291	8
39	294	8
40	296	8
41	298	8
42	301	8
43	303	8
44	306	8
45	308	9
46	311	9
47	313	9
48	316	9
49	319	9
50	322	9
51	325	10
52	328	10
53	331	10
54	335	10
55	339	11
56	343	11
57	348	12
58	353	13
59	358	13
60	365	15
61	372	16
62	381	18
63	393	21
64	401	24
65	409	27
66	417	30

Table R7. Mathematics Raw Score-to-Scale Score Table, Grade 3

Raw Score	Scale Score	Standard Error
0	139	77
1	147	66
2	155	57
3	163	50
4	171	43
5	179	38
6	187	33
7	195	29
8	203	26
9	211	22
10	219	20
11	228	17
12	234	15
13	240	13
14	245	12
15	249	11
16	253	10
17	256	10
18	259	9
19	262	9
20	265	9
21	267	8
22	270	8
23	272	8
24	274	8
25	277	7
26	279	7
27	281	7
28	283	7
29	285	7
30	287	7

Raw Score	Scale Score	Standard Error
31	289	7
32	291	7
33	293	7
34	294	7
35	296	7
36	298	7
37	300	7
38	302	7
39	304	7
40	306	7
41	308	7
42	310	7
43	312	7
44	314	7
45	317	7
46	319	7
47	321	7
48	324	8
49	327	8
50	330	8
51	333	9
52	336	9
53	340	10
54	345	11
55	350	12
56	357	13
57	365	16
58	378	20
59	386	23
60	394	27

Table R8. Mathematics Raw Score-to-Scale Score Table, Grade 4

Raw Score	Scale Score	Standard Error
0	126	82
1	134	72
2	142	63
3	150	55
4	159	48
5	167	42
6	175	37
7	183	32
8	191	28
9	200	24
10	208	21
11	216	19
12	224	16
13	231	15
14	237	13
15	242	13
16	247	12
17	251	11
18	255	10
19	259	10
20	262	9
21	265	9
22	268	9
23	271	9
24	273	8
25	276	8
26	278	8
27	281	8
28	283	8
29	285	8
30	287	7
31	289	7
32	292	7
33	294	7
34	296	7
35	298	7
36	300	7
37	302	7
38	304	7
39	306	7
40	308	7
41	310	7
42	312	7
43	314	7
44	316	7
45	318	7
46	320	7
47	322	7
48	324	7
49	326	7
50	328	7
51	331	7
52	333	8
53	335	8
54	338	8
55	341	8
56	344	8
57	347	9
58	350	9
59	354	9
60	358	10
61	363	11
62	368	12
63	375	13
64	386	17
65	394	20
66	402	24

Table R9. Mathematics Raw Score-to-Scale Score Table, Grade 5

Raw Score	Scale Score	Standard Error
0	126	113
1	134	97
2	142	84
3	150	72
4	158	63
5	166	54
6	174	47
7	182	41
8	190	36
9	198	31
10	206	27
11	214	23
12	222	19
13	232	16
14	239	14
15	245	12
16	250	11
17	255	11
18	259	10
19	263	10
20	266	9
21	269	9
22	272	9
23	275	8
24	278	8
25	281	8
26	283	8
27	285	8
28	288	7
29	290	7
30	292	7
31	294	7
32	296	7
33	298	7
34	300	7
35	302	7
36	304	7
37	306	7
38	308	6
39	309	6
40	311	6
41	313	6
42	315	6
43	317	6
44	319	6
45	321	6
46	323	6
47	325	6
48	327	6
49	329	7
50	331	7
51	333	7
52	335	7
53	338	7
54	340	7
55	343	8
56	346	8
57	349	8
58	353	8
59	357	9
60	361	9
61	366	10
62	372	11
63	379	14
64	390	17
65	398	21
66	406	24

Table R10. Mathematics Raw Score-to-Scale Score Table, Grade 6

Raw Score	Scale Score	Standard Error
0	119	128
1	127	111
2	135	96
3	143	83
4	151	72
5	159	62
6	168	53
7	176	45
8	184	38
9	192	33
10	200	28
11	208	23
12	216	20
13	224	17
14	231	14
15	237	13
16	241	12
17	246	11
18	249	11
19	253	10
20	256	10
21	259	9
22	262	9
23	265	9
24	268	9
25	270	8
26	273	8
27	275	8
28	277	8
29	279	8
30	282	8
31	284	8
32	286	7
33	288	7
34	290	7
35	292	7
36	294	7
37	296	7
38	298	7
39	299	7
40	301	7
41	303	6
42	305	6
43	306	6
44	308	6
45	310	6
46	311	6
47	313	6
48	315	6
49	316	6
50	318	6
51	320	6
52	322	6
53	323	6
54	325	6
55	327	6
56	329	6
57	331	6
58	333	7
59	335	7
60	337	7
61	340	7
62	342	7
63	345	7
64	348	8
65	352	8
66	355	9
67	360	10
68	365	11
69	372	13
70	382	16
71	390	19
72	399	24

Table R11. Mathematics Raw Score-to-Scale Score Table, Grade 7

Raw Score	Scale Score	Standard Error
0	133	180
1	140	158
2	148	135
3	156	116
4	164	99
5	172	85
6	179	74
7	187	62
8	195	53
9	203	45
10	211	37
11	219	31
12	226	27
13	234	22
14	245	17
15	253	14
16	259	12
17	264	11
18	268	10
19	271	10
20	275	9
21	278	8
22	280	8
23	283	8
24	285	7
25	287	7
26	289	7
27	291	7
28	293	7
29	295	6
30	297	6
31	298	6
32	300	6
33	302	6
34	303	6
35	305	6
36	306	6
37	308	6
38	310	6
39	311	6
40	313	6
41	314	6
42	316	6
43	317	6
44	319	6
45	320	6
46	322	6
47	324	6
48	325	6
49	327	6
50	329	6
51	330	6
52	332	6
53	334	6
54	336	6
55	338	6
56	340	6
57	342	6
58	344	6
59	346	6
60	348	6
61	350	6
62	352	6
63	355	6
64	358	6
65	360	6
66	364	7
67	367	7
68	371	8
69	377	10
70	385	13
71	393	16
72	401	21

Table R12. Mathematics Raw Score-to-Scale Score Table, Grade 8

Raw Score	Scale Score	Standard Error
0	119	120
1	127	103
2	135	89
3	143	76
4	151	66
5	159	57
6	167	49
7	175	42
8	183	36
9	191	32
10	199	28
11	207	24
12	215	21
13	223	19
14	232	17
15	239	15
16	245	14
17	250	13
18	255	12
19	259	11
20	263	11
21	266	10
22	269	10
23	272	9
24	275	9
25	277	9
26	280	8
27	282	8
28	284	8
29	287	8
30	289	7
31	291	7
32	292	7
33	294	7
34	296	7
35	298	7
36	300	7
37	302	7
38	303	7
39	305	6
40	307	6
41	308	6
42	310	6
43	312	6
44	314	6
45	315	6
46	317	6
47	319	6
48	320	6
49	322	6
50	324	6
51	326	6
52	328	6
53	329	6
54	331	6
55	333	6
56	335	6
57	337	6
58	339	6
59	341	6
60	344	7
61	346	7
62	349	7
63	351	7
64	354	8
65	358	8
66	361	8
67	366	9
68	371	10
69	377	12
70	387	16
71	395	19
72	403	24

Table R13. Grade 3 ELA Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
148	17	0.01	17	0.01
156	23	0.01	40	0.02
164	53	0.03	93	0.05
172	100	0.05	193	0.10
180	250	0.12	443	0.22
188	505	0.25	948	0.47
196	866	0.43	1,814	0.89
204	1,164	0.57	2,978	1.47
212	1,592	0.78	4,570	2.25
220	1,911	0.94	6,481	3.19
228	2,247	1.11	8,728	4.30
235	2,634	1.30	11,362	5.60
242	2,936	1.45	14,298	7.04
248	3,220	1.59	17,518	8.63
253	3,650	1.80	21,168	10.43
258	3,895	1.92	25,063	12.34
262	4,197	2.07	29,260	14.41
266	4,579	2.26	33,839	16.67
269	4,901	2.41	38,740	19.08
273	5,124	2.52	43,864	21.60
276	5,391	2.66	49,255	24.26
279	5,401	2.66	54,656	26.92
283	5,796	2.85	60,452	29.77
286	5,785	2.85	66,237	32.62
288	6,062	2.99	72,299	35.61
291	6,435	3.17	78,734	38.78
294	6,405	3.15	85,139	41.93
297	6,563	3.23	91,702	45.16
300	6,717	3.31	98,419	48.47
303	6,818	3.36	105,237	51.83
305	6,851	3.37	112,088	55.20
308	6,876	3.39	118,964	58.59
311	7,088	3.49	126,052	62.08
314	6,925	3.41	132,977	65.49
317	6,681	3.29	139,658	68.78
320	6,599	3.25	146,257	72.03
322	6,471	3.19	152,728	75.22
326	6,403	3.15	159,131	78.37

Table R13. Grade 3 ELA Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
329	6,002	2.96	165,133	81.33
332	5,611	2.76	170,744	84.09
335	5,309	2.61	176,053	86.71
338	4,789	2.36	180,842	89.07
342	4,478	2.21	185,320	91.27
346	3,920	1.93	189,240	93.20
349	3,378	1.66	192,618	94.87
354	2,815	1.39	195,433	96.25
358	2,262	1.11	197,695	97.37
363	1,800	0.89	199,495	98.25
368	1,369	0.67	200,864	98.93
375	924	0.46	201,788	99.38
382	601	0.30	202,389	99.68
391	347	0.17	202,736	99.85
399	195	0.10	202,931	99.95
407	84	0.04	203,015	99.99
415	22	0.01	203,037	100.00
423	4	0.00	203,041	100.00

Table R14. Grade 4 ELA Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
139	10	0.00	10	0.00
147	22	0.01	32	0.01
155	36	0.02	68	0.03
163	118	0.06	186	0.09
171	225	0.11	411	0.19
179	440	0.21	851	0.40
187	671	0.31	1,522	0.71
195	1,007	0.47	2,529	1.18
203	1,320	0.62	3,849	1.80
211	1,544	0.72	5,393	2.52
220	1,771	0.83	7,164	3.35
228	2,049	0.96	9,213	4.31
234	2,075	0.97	11,288	5.28
240	2,344	1.10	13,632	6.38
245	2,607	1.22	16,239	7.60
249	2,943	1.38	19,182	8.98
254	2,997	1.40	22,179	10.38
258	3,368	1.58	25,547	11.96
261	3,746	1.75	29,293	13.71
265	4,017	1.88	33,310	15.59

Table R14. Grade 4 ELA Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
268	4,251	1.99	37,561	17.58
272	4,727	2.21	42,288	19.79
275	4,962	2.32	47,250	22.11
278	5,254	2.46	52,504	24.57
281	5,588	2.62	58,092	27.19
284	5,806	2.72	63,898	29.90
287	6,052	2.83	69,950	32.74
290	6,373	2.98	76,323	35.72
292	6,504	3.04	82,827	38.76
295	6,835	3.20	89,662	41.96
298	6,944	3.25	96,606	45.21
301	6,972	3.26	103,578	48.48
303	7,233	3.39	110,811	51.86
306	7,260	3.40	118,071	55.26
309	7,498	3.51	125,569	58.77
311	7,592	3.55	133,161	62.32
314	7,483	3.50	140,644	65.82
317	7,345	3.44	147,989	69.26
320	7,407	3.47	155,396	72.73
323	7,209	3.37	162,605	76.10
326	6,995	3.27	169,600	79.37
329	6,583	3.08	176,183	82.45
332	6,272	2.94	182,455	85.39
336	5,766	2.70	188,221	88.09
339	5,112	2.39	193,333	90.48
343	4,680	2.19	198,013	92.67
347	4,041	1.89	202,054	94.56
352	3,351	1.57	205,405	96.13
357	2,707	1.27	208,112	97.40
363	2,067	0.97	210,179	98.36
369	1,475	0.69	211,654	99.06
377	999	0.47	212,653	99.52
388	596	0.28	213,249	99.80
396	281	0.13	213,530	99.93
404	117	0.05	213,647	99.99
412	26	0.01	213,673	100.00

Table R15. Grade 5 ELA Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
116	3	0.00	3	0.00
124	7	0.00	10	0.00
132	13	0.01	23	0.01
140	8	0.00	31	0.02
148	24	0.01	55	0.03
156	53	0.03	108	0.05
164	108	0.05	216	0.11
172	179	0.09	395	0.20
180	319	0.16	714	0.36
188	565	0.28	1,279	0.64
196	725	0.36	2,004	1.00
204	895	0.45	2,899	1.44
212	1,270	0.63	4,169	2.07
220	1,436	0.71	5,605	2.79
228	1,695	0.84	7,300	3.63
234	1,984	0.99	9,284	4.62
239	2,216	1.10	11,500	5.72
244	2,306	1.15	13,806	6.86
248	2,719	1.35	16,525	8.22
252	2,907	1.45	19,432	9.66
256	3,165	1.57	22,597	11.24
259	3,268	1.62	25,865	12.86
263	3,495	1.74	29,360	14.60
266	3,662	1.82	33,022	16.42
269	3,917	1.95	36,939	18.37
271	4,058	2.02	40,997	20.38
274	4,240	2.11	45,237	22.49
277	4,347	2.16	49,584	24.65
279	4,565	2.27	54,149	26.92
282	4,707	2.34	58,856	29.26
284	4,849	2.41	63,705	31.68
287	4,981	2.48	68,686	34.15
289	5,123	2.55	73,809	36.70
292	5,262	2.62	79,071	39.32
294	5,242	2.61	84,313	41.92
296	5,341	2.66	89,654	44.58
298	5,462	2.72	95,116	47.29
301	5,459	2.71	100,575	50.01
303	5,574	2.77	106,149	52.78
305	5,565	2.77	111,714	55.55
308	5,670	2.82	117,384	58.37
310	5,612	2.79	122,996	61.16
312	5,799	2.88	128,795	64.04

Table R15. Grade 5 ELA Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
315	5,657	2.81	134,452	66.85
317	5,705	2.84	140,157	69.69
320	5,611	2.79	145,768	72.48
322	5,457	2.71	151,225	75.19
325	5,251	2.61	156,476	77.80
327	5,247	2.61	161,723	80.41
330	4,989	2.48	166,712	82.89
333	4,641	2.31	171,353	85.20
336	4,485	2.23	175,838	87.43
339	4,181	2.08	180,019	89.51
343	3,943	1.96	183,962	91.47
346	3,414	1.70	187,376	93.17
350	3,104	1.54	190,480	94.71
354	2,695	1.34	193,175	96.05
359	2,179	1.08	195,354	97.13
364	1,861	0.93	197,215	98.06
369	1,416	0.70	198,631	98.76
376	1,033	0.51	199,664	99.28
383	690	0.34	200,354	99.62
393	400	0.20	200,754	99.82
401	227	0.11	200,981	99.93
409	101	0.05	201,082	99.98
417	28	0.01	201,110	100.00
425	7	0.00	201,117	100.00

Table R16. Grade 6 ELA Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
112	3	0.00	3	0.00
120	6	0.00	9	0.00
128	6	0.00	15	0.01
136	7	0.00	22	0.01
144	30	0.01	52	0.02
152	48	0.02	100	0.05
160	95	0.04	195	0.09
168	187	0.09	382	0.18
176	298	0.14	680	0.32
184	428	0.20	1,108	0.52
192	582	0.28	1,690	0.80
200	853	0.40	2,543	1.20
208	988	0.47	3,531	1.67
216	1,160	0.55	4,691	2.22

Table R16. Grade 6 ELA Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
224	1,370	0.65	6,061	2.86
230	1,515	0.72	7,576	3.58
234	1,739	0.82	9,315	4.40
239	2,051	0.97	11,366	5.37
243	2,111	1.00	13,477	6.37
246	2,308	1.09	15,785	7.46
250	2,508	1.19	18,293	8.65
253	2,793	1.32	21,086	9.97
256	2,980	1.41	24,066	11.37
259	3,073	1.45	27,139	12.83
262	3,368	1.59	30,507	14.42
265	3,518	1.66	34,025	16.08
268	3,836	1.81	37,861	17.89
270	4,034	1.91	41,895	19.80
273	4,210	1.99	46,105	21.79
276	4,369	2.06	50,474	23.85
278	4,649	2.20	55,123	26.05
281	4,706	2.22	59,829	28.28
283	4,996	2.36	64,825	30.64
285	5,249	2.48	70,074	33.12
288	5,214	2.46	75,288	35.58
290	5,433	2.57	80,721	38.15
293	5,761	2.72	86,482	40.87
295	5,896	2.79	92,378	43.66
297	5,978	2.83	98,356	46.48
300	5,956	2.81	104,312	49.30
302	6,082	2.87	110,394	52.17
305	6,289	2.97	116,683	55.14
307	6,408	3.03	123,091	58.17
309	6,403	3.03	129,494	61.20
312	6,338	3.00	135,832	64.19
315	6,234	2.95	142,066	67.14
317	6,173	2.92	148,239	70.06
320	6,166	2.91	154,405	72.97
323	5,924	2.80	160,329	75.77
325	5,937	2.81	166,266	78.58
328	5,623	2.66	171,889	81.24
331	5,477	2.59	177,366	83.82
335	5,230	2.47	182,596	86.30
338	4,766	2.25	187,362	88.55
341	4,423	2.09	191,785	90.64
345	4,105	1.94	195,890	92.58
349	3,610	1.71	199,500	94.28

Table R16. Grade 6 ELA Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
354	3,120	1.47	202,620	95.76
359	2,637	1.25	205,257	97.01
364	2,140	1.01	207,397	98.02
370	1,659	0.78	209,056	98.80
378	1,145	0.54	210,201	99.34
387	744	0.35	210,945	99.69
396	416	0.20	211,361	99.89
404	198	0.09	211,559	99.98
412	35	0.02	211,594	100.00

Table R17. Grade 7 ELA Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
103	2	0.00	2	0.00
111	5	0.00	7	0.00
119	1	0.00	8	0.00
127	9	0.00	17	0.01
135	13	0.01	30	0.01
143	35	0.02	65	0.03
151	63	0.03	128	0.06
159	113	0.05	241	0.12
167	215	0.10	456	0.22
175	310	0.15	766	0.37
183	491	0.24	1,257	0.61
191	586	0.28	1,843	0.90
199	755	0.37	2,598	1.26
207	986	0.48	3,584	1.74
215	1,178	0.57	4,762	2.31
223	1,265	0.61	6,027	2.93
228	1,496	0.73	7,523	3.66
233	1,574	0.76	9,097	4.42
237	1,666	0.81	10,763	5.23
241	1,896	0.92	12,659	6.15
245	2,086	1.01	14,745	7.16
249	2,272	1.10	17,017	8.27
252	2,466	1.20	19,483	9.47
255	2,716	1.32	22,199	10.79
258	2,845	1.38	25,044	12.17
261	3,010	1.46	28,054	13.63
264	3,422	1.66	31,476	15.29
267	3,470	1.69	34,946	16.98
270	3,735	1.81	38,681	18.80

Table R17. Grade 7 ELA Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
272	4,014	1.95	42,695	20.75
275	4,281	2.08	46,976	22.83
277	4,377	2.13	51,353	24.95
280	4,660	2.26	56,013	27.22
282	4,716	2.29	60,729	29.51
285	5,070	2.46	65,799	31.97
287	5,243	2.55	71,042	34.52
289	5,281	2.57	76,323	37.09
292	5,402	2.62	81,725	39.71
294	5,638	2.74	87,363	42.45
296	5,794	2.82	93,157	45.27
299	5,835	2.84	98,992	48.10
301	5,961	2.90	104,953	51.00
304	5,890	2.86	110,843	53.86
306	6,058	2.94	116,901	56.80
308	6,111	2.97	123,012	59.77
311	5,978	2.90	128,990	62.68
313	6,012	2.92	135,002	65.60
316	6,059	2.94	141,061	68.54
318	6,030	2.93	147,091	71.47
321	5,789	2.81	152,880	74.29
324	5,783	2.81	158,663	77.10
327	5,828	2.83	164,491	79.93
330	5,439	2.64	169,930	82.57
333	5,162	2.51	175,092	85.08
336	4,981	2.42	180,073	87.50
340	4,688	2.28	184,761	89.78
343	4,498	2.19	189,259	91.96
347	3,935	1.91	193,194	93.87
352	3,413	1.66	196,607	95.53
357	2,902	1.41	199,509	96.94
363	2,339	1.14	201,848	98.08
370	1,704	0.83	203,552	98.91
378	1,121	0.54	204,673	99.45
389	702	0.34	205,375	99.79
397	285	0.14	205,660	99.93
405	120	0.06	205,780	99.99
413	20	0.01	205,800	100.00

Table R18. Grade 8 ELA Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
100	3	0.00	3	0.00
108	2	0.00	5	0.00
116	8	0.00	13	0.01
124	8	0.00	21	0.01
131	13	0.01	34	0.02
139	23	0.01	57	0.03
147	51	0.02	108	0.05
155	120	0.06	228	0.11
163	193	0.09	421	0.20
171	284	0.13	705	0.33
179	415	0.19	1,120	0.52
187	557	0.26	1,677	0.78
195	736	0.34	2,413	1.13
203	886	0.41	3,299	1.54
211	1,087	0.51	4,386	2.05
218	1,125	0.53	5,511	2.57
226	1,381	0.65	6,892	3.22
231	1,449	0.68	8,341	3.90
235	1,605	0.75	9,946	4.65
239	1,844	0.86	11,790	5.51
243	1,995	0.93	13,785	6.44
246	2,090	0.98	15,875	7.42
249	2,310	1.08	18,185	8.49
253	2,409	1.13	20,594	9.62
256	2,698	1.26	23,292	10.88
259	2,782	1.30	26,074	12.18
261	3,106	1.45	29,180	13.63
264	3,256	1.52	32,436	15.15
267	3,598	1.68	36,034	16.83
269	3,763	1.76	39,797	18.59
272	3,976	1.86	43,773	20.45
275	4,253	1.99	48,026	22.43
277	4,299	2.01	52,325	24.44
279	4,671	2.18	56,996	26.62
282	4,833	2.26	61,829	28.88
284	5,081	2.37	66,910	31.25
286	5,262	2.46	72,172	33.71
289	5,472	2.56	77,644	36.27
291	5,555	2.59	83,199	38.86
294	5,818	2.72	89,017	41.58
296	5,967	2.79	94,984	44.37
298	6,158	2.88	101,142	47.24
301	6,482	3.03	107,624	50.27

Table R18. Grade 8 ELA Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
303	6,374	2.98	113,998	53.25
306	6,630	3.10	120,628	56.34
308	6,547	3.06	127,175	59.40
311	6,709	3.13	133,884	62.54
313	6,802	3.18	140,686	65.71
316	6,744	3.15	147,430	68.86
319	6,734	3.15	154,164	72.01
322	6,678	3.12	160,842	75.13
325	6,720	3.14	167,562	78.27
328	6,312	2.95	173,874	81.22
331	6,427	3.00	180,301	84.22
335	5,863	2.74	186,164	86.96
339	5,505	2.57	191,669	89.53
343	5,042	2.36	196,711	91.88
348	4,539	2.12	201,250	94.00
353	3,860	1.80	205,110	95.81
358	3,106	1.45	208,216	97.26
365	2,364	1.10	210,580	98.36
372	1,655	0.77	212,235	99.13
381	1,037	0.48	213,272	99.62
393	558	0.26	213,830	99.88
401	205	0.10	214,035	99.97
409	49	0.02	214,084	100.00
417	6	0.00	214,090	100.00

Table R19. Grade 3 Mathematics Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
139	6	0.00	6	0.00
147	5	0.00	11	0.01
155	7	0.00	18	0.01
163	31	0.02	49	0.02
171	66	0.03	115	0.06
179	147	0.07	262	0.13
187	254	0.12	516	0.25
195	501	0.24	1,017	0.50
203	794	0.39	1,811	0.89
211	1,219	0.60	3,030	1.48
219	1,613	0.79	4,643	2.27
228	1,869	0.91	6,512	3.18
234	2,161	1.06	8,673	4.24
240	2,340	1.14	11,013	5.39

Table R19. Grade 3 Mathematics Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
245	2,562	1.25	13,575	6.64
249	2,686	1.31	16,261	7.95
253	2,801	1.37	19,062	9.32
256	2,968	1.45	22,030	10.77
259	3,136	1.53	25,166	12.31
262	3,205	1.57	28,371	13.87
265	3,330	1.63	31,701	15.50
267	3,402	1.66	35,103	17.16
270	3,543	1.73	38,646	18.90
272	3,665	1.79	42,311	20.69
274	3,850	1.88	46,161	22.57
277	3,848	1.88	50,009	24.45
279	3,974	1.94	53,983	26.40
281	4,063	1.99	58,046	28.38
283	4,173	2.04	62,219	30.42
285	4,341	2.12	66,560	32.55
287	4,322	2.11	70,882	34.66
289	4,531	2.22	75,413	36.88
291	4,590	2.24	80,003	39.12
293	4,648	2.27	84,651	41.39
294	4,694	2.30	89,345	43.69
296	4,879	2.39	94,224	46.07
298	5,060	2.47	99,284	48.55
300	5,033	2.46	104,317	51.01
302	4,939	2.42	109,256	53.42
304	5,101	2.49	114,357	55.92
306	5,119	2.50	119,476	58.42
308	5,116	2.50	124,592	60.92
310	5,098	2.49	129,690	63.42
312	5,202	2.54	134,892	65.96
314	5,140	2.51	140,032	68.47
317	5,158	2.52	145,190	71.00
319	4,998	2.44	150,188	73.44
321	5,100	2.49	155,288	75.93
324	5,100	2.49	160,388	78.43
327	4,926	2.41	165,314	80.84
330	4,777	2.34	170,091	83.17
333	4,825	2.36	174,916	85.53
336	4,591	2.24	179,507	87.78
340	4,291	2.10	183,798	89.88
345	4,149	2.03	187,947	91.90
350	4,025	1.97	191,972	93.87
357	3,524	1.72	195,496	95.60

Table R19. Grade 3 Mathematics Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
365	3,255	1.59	198,751	97.19
378	2,677	1.31	201,428	98.50
386	1,993	0.97	203,421	99.47
394	1,083	0.53	204,504	100.00

Table R20. Grade 4 Mathematics Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
126	7	0.00	7	0.00
134	1	0.00	8	0.00
142	3	0.00	11	0.01
150	5	0.00	16	0.01
159	21	0.01	37	0.02
167	56	0.03	93	0.04
175	120	0.06	213	0.10
183	221	0.10	434	0.20
191	436	0.20	870	0.40
200	689	0.32	1,559	0.72
208	1,003	0.47	2,562	1.19
216	1,375	0.64	3,937	1.83
224	1,692	0.79	5,629	2.61
231	2,038	0.95	7,667	3.56
237	2,406	1.12	10,073	4.68
242	2,708	1.26	12,781	5.93
247	2,954	1.37	15,735	7.31
251	3,372	1.57	19,107	8.87
255	3,551	1.65	22,658	10.52
259	3,816	1.77	26,474	12.29
262	3,936	1.83	30,410	14.12
265	4,189	1.94	34,599	16.06
268	4,386	2.04	38,985	18.10
271	4,481	2.08	43,466	20.18
273	4,757	2.21	48,223	22.39
276	4,804	2.23	53,027	24.62
278	4,906	2.28	57,933	26.90
281	4,905	2.28	62,838	29.17
283	5,064	2.35	67,902	31.52
285	5,111	2.37	73,013	33.90
287	5,315	2.47	78,328	36.36
289	5,178	2.40	83,506	38.77
292	5,220	2.42	88,726	41.19
294	5,129	2.38	93,855	43.57

Table R20. Grade 4 Mathematics Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
296	5,244	2.43	99,099	46.01
298	5,031	2.34	104,130	48.34
300	5,074	2.36	109,204	50.70
302	5,118	2.38	114,322	53.08
304	4,892	2.27	119,214	55.35
306	4,948	2.30	124,162	57.64
308	4,824	2.24	128,986	59.88
310	4,811	2.23	133,797	62.12
312	4,694	2.18	138,491	64.30
314	4,604	2.14	143,095	66.43
316	4,658	2.16	147,753	68.60
318	4,415	2.05	152,168	70.65
320	4,401	2.04	156,569	72.69
322	4,467	2.07	161,036	74.76
324	4,236	1.97	165,272	76.73
326	4,142	1.92	169,414	78.65
328	4,024	1.87	173,438	80.52
331	3,863	1.79	177,301	82.31
333	3,866	1.79	181,167	84.11
335	3,782	1.76	184,949	85.87
338	3,568	1.66	188,517	87.52
341	3,529	1.64	192,046	89.16
344	3,292	1.53	195,338	90.69
347	3,196	1.48	198,534	92.17
350	3,134	1.46	201,668	93.63
354	2,770	1.29	204,438	94.91
358	2,582	1.20	207,020	96.11
363	2,309	1.07	209,329	97.18
368	2,018	0.94	211,347	98.12
375	1,739	0.81	213,086	98.93
386	1,204	0.56	214,290	99.49
394	762	0.35	215,052	99.84
402	342	0.16	215,394	100.00

Table R21. Grade 5 Mathematics Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
126	4	0.00	4	0.00
134	3	0.00	7	0.00
142	5	0.00	12	0.01
150	14	0.01	26	0.01
158	35	0.02	61	0.03

Table R21. Grade 5 Mathematics Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
166	57	0.03	118	0.06
174	129	0.06	247	0.12
182	266	0.13	513	0.25
190	456	0.23	969	0.48
198	842	0.42	1,811	0.89
206	1,168	0.58	2,979	1.47
214	1,549	0.76	4,528	2.24
222	1,985	0.98	6,513	3.22
232	2,308	1.14	8,821	4.35
239	2,693	1.33	11,514	5.68
245	2,949	1.46	14,463	7.14
250	3,230	1.59	17,693	8.73
255	3,449	1.70	21,142	10.44
259	3,609	1.78	24,751	12.22
263	3,946	1.95	28,697	14.17
266	4,067	2.01	32,764	16.17
269	4,192	2.07	36,956	18.24
272	4,411	2.18	41,367	20.42
275	4,635	2.29	46,002	22.71
278	4,738	2.34	50,740	25.05
281	4,899	2.42	55,639	27.47
283	4,977	2.46	60,616	29.92
285	5,104	2.52	65,720	32.44
288	5,046	2.49	70,766	34.93
290	5,201	2.57	75,967	37.50
292	5,110	2.52	81,077	40.02
294	5,038	2.49	86,115	42.51
296	5,148	2.54	91,263	45.05
298	5,103	2.52	96,366	47.57
300	5,051	2.49	101,417	50.06
302	4,886	2.41	106,303	52.47
304	5,013	2.47	111,316	54.95
306	4,805	2.37	116,121	57.32
308	4,635	2.29	120,756	59.61
309	4,532	2.24	125,288	61.85
311	4,479	2.21	129,767	64.06
313	4,370	2.16	134,137	66.21
315	4,296	2.12	138,433	68.34
317	4,073	2.01	142,506	70.35
319	4,162	2.05	146,668	72.40
321	3,887	1.92	150,555	74.32
323	4,092	2.02	154,647	76.34
325	3,869	1.91	158,516	78.25

Table R21. Grade 5 Mathematics Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
327	3,726	1.84	162,242	80.09
329	3,461	1.71	165,703	81.80
331	3,516	1.74	169,219	83.53
333	3,382	1.67	172,601	85.20
335	3,238	1.60	175,839	86.80
338	3,077	1.52	178,916	88.32
340	2,952	1.46	181,868	89.78
343	2,945	1.45	184,813	91.23
346	2,597	1.28	187,410	92.51
349	2,520	1.24	189,930	93.76
353	2,357	1.16	192,287	94.92
357	2,049	1.01	194,336	95.93
361	1,907	0.94	196,243	96.87
366	1,698	0.84	197,941	97.71
372	1,496	0.74	199,437	98.45
379	1,220	0.60	200,657	99.05
390	899	0.44	201,556	99.50
398	687	0.34	202,243	99.83
406	336	0.17	202,579	100.00

Table R22. Grade 6 Mathematics Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
119	1	0.00	1	0.00
127	4	0.00	5	0.00
135	4	0.00	9	0.00
143	8	0.00	17	0.01
151	18	0.01	35	0.02
159	50	0.02	85	0.04
168	91	0.04	176	0.08
176	211	0.10	387	0.18
184	342	0.16	729	0.34
192	620	0.29	1,349	0.63
200	822	0.39	2,171	1.02
208	1,086	0.51	3,257	1.53
216	1,374	0.64	4,631	2.17
224	1,646	0.77	6,277	2.94
231	1,970	0.92	8,247	3.87
237	2,044	0.96	10,291	4.83
241	2,375	1.11	12,666	5.94
246	2,392	1.12	15,058	7.06
249	2,661	1.25	17,719	8.31

Table R22. Grade 6 Mathematics Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
253	2,778	1.30	20,497	9.62
256	2,982	1.40	23,479	11.02
259	2,990	1.40	26,469	12.42
262	3,355	1.57	29,824	13.99
265	3,454	1.62	33,278	15.61
268	3,570	1.67	36,848	17.29
270	3,775	1.77	40,623	19.06
273	3,876	1.82	44,499	20.88
275	4,068	1.91	48,567	22.79
277	4,279	2.01	52,846	24.79
279	4,554	2.14	57,400	26.93
282	4,474	2.10	61,874	29.03
284	4,643	2.18	66,517	31.21
286	4,700	2.21	71,217	33.41
288	4,797	2.25	76,014	35.66
290	4,740	2.22	80,754	37.89
292	4,851	2.28	85,605	40.16
294	4,996	2.34	90,601	42.51
296	4,800	2.25	95,401	44.76
298	4,759	2.23	100,160	46.99
299	4,958	2.33	105,118	49.32
301	4,829	2.27	109,947	51.58
303	4,647	2.18	114,594	53.76
305	4,534	2.13	119,128	55.89
306	4,590	2.15	123,718	58.04
308	4,491	2.11	128,209	60.15
310	4,364	2.05	132,573	62.20
311	4,259	2.00	136,832	64.20
313	4,251	1.99	141,083	66.19
315	4,087	1.92	145,170	68.11
316	4,051	1.90	149,221	70.01
318	4,015	1.88	153,236	71.89
320	3,898	1.83	157,134	73.72
322	3,685	1.73	160,819	75.45
323	3,676	1.72	164,495	77.17
325	3,515	1.65	168,010	78.82
327	3,507	1.65	171,517	80.47
329	3,403	1.60	174,920	82.07
331	3,358	1.58	178,278	83.64
333	3,120	1.46	181,398	85.10
335	3,071	1.44	184,469	86.55
337	2,977	1.40	187,446	87.94
340	2,915	1.37	190,361	89.31

Table R22. Grade 6 Mathematics Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
342	2,778	1.30	193,139	90.61
345	2,734	1.28	195,873	91.90
348	2,662	1.25	198,535	93.14
352	2,629	1.23	201,164	94.38
355	2,476	1.16	203,640	95.54
360	2,387	1.12	206,027	96.66
365	2,062	0.97	208,089	97.63
372	1,975	0.93	210,064	98.55
382	1,560	0.73	211,624	99.29
390	1,039	0.49	212,663	99.77
399	484	0.23	213,147	100.00

Table R23. Grade 7 Mathematics Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
133	5	0.00	5	0.00
140	8	0.00	13	0.01
148	12	0.01	25	0.01
156	12	0.01	37	0.02
164	28	0.01	65	0.03
172	77	0.04	142	0.07
179	182	0.09	324	0.16
187	386	0.19	710	0.34
195	632	0.30	1,342	0.65
203	1,123	0.54	2,465	1.19
211	1,723	0.83	4,188	2.02
219	2,389	1.15	6,577	3.17
226	2,988	1.44	9,565	4.62
234	3,615	1.74	13,180	6.36
245	4,113	1.98	17,293	8.34
253	4,566	2.20	21,859	10.55
259	4,705	2.27	26,564	12.82
264	5,029	2.43	31,593	15.24
268	5,073	2.45	36,666	17.69
271	4,965	2.40	41,631	20.09
275	4,857	2.34	46,488	22.43
278	4,831	2.33	51,319	24.76
280	4,737	2.29	56,056	27.05
283	4,682	2.26	60,738	29.31
285	4,606	2.22	65,344	31.53
287	4,502	2.17	69,846	33.70
289	4,433	2.14	74,279	35.84

Table R23. Grade 7 Mathematics Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
291	4,415	2.13	78,694	37.97
293	4,396	2.12	83,090	40.09
295	4,250	2.05	87,340	42.14
297	4,159	2.01	91,499	44.15
298	4,241	2.05	95,740	46.20
300	4,133	1.99	99,873	48.19
302	4,112	1.98	103,985	50.17
303	3,973	1.92	107,958	52.09
305	3,942	1.90	111,900	53.99
306	3,848	1.86	115,748	55.85
308	3,952	1.91	119,700	57.76
310	3,857	1.86	123,557	59.62
311	3,944	1.90	127,501	61.52
313	3,795	1.83	131,296	63.35
314	3,856	1.86	135,152	65.21
316	3,822	1.84	138,974	67.06
317	3,708	1.79	142,682	68.85
319	3,653	1.76	146,335	70.61
320	3,686	1.78	150,021	72.39
322	3,590	1.73	153,611	74.12
324	3,512	1.69	157,123	75.81
325	3,381	1.63	160,504	77.44
327	3,394	1.64	163,898	79.08
329	3,255	1.57	167,153	80.65
330	3,277	1.58	170,430	82.23
332	3,117	1.50	173,547	83.74
334	3,001	1.45	176,548	85.19
336	2,864	1.38	179,412	86.57
338	2,876	1.39	182,288	87.96
340	2,669	1.29	184,957	89.24
342	2,435	1.17	187,392	90.42
344	2,384	1.15	189,776	91.57
346	2,263	1.09	192,039	92.66
348	2,059	0.99	194,098	93.65
350	1,979	0.95	196,077	94.61
352	1,789	0.86	197,866	95.47
355	1,625	0.78	199,491	96.26
358	1,548	0.75	201,039	97.00
360	1,264	0.61	202,303	97.61
364	1,239	0.60	203,542	98.21
367	1,018	0.49	204,560	98.70
371	857	0.41	205,417	99.12
377	728	0.35	206,145	99.47

Table R23. Grade 7 Mathematics Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
385	550	0.27	206,695	99.73
393	391	0.19	207,086	99.92
401	165	0.08	207,251	100.00

Table R24. Grade 8 Mathematics Scale Score Frequency Distribution

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
119	6	0.00	6	0.00
127	7	0.00	13	0.01
135	4	0.00	17	0.01
143	8	0.00	25	0.01
151	10	0.00	35	0.02
159	56	0.03	91	0.04
167	97	0.05	188	0.09
175	179	0.08	367	0.17
183	341	0.16	708	0.33
191	622	0.29	1,330	0.62
199	903	0.42	2,233	1.04
207	1,344	0.63	3,577	1.66
215	1,845	0.86	5,422	2.52
223	2,219	1.03	7,641	3.56
232	2,628	1.22	10,269	4.78
239	2,909	1.35	13,178	6.13
245	3,329	1.55	16,507	7.68
250	3,591	1.67	20,098	9.35
255	3,782	1.76	23,880	11.11
259	3,969	1.85	27,849	12.96
263	4,024	1.87	31,873	14.83
266	4,176	1.94	36,049	16.77
269	4,293	2.00	40,342	18.77
272	4,470	2.08	44,812	20.85
275	4,387	2.04	49,199	22.89
277	4,465	2.08	53,664	24.97
280	4,494	2.09	58,158	27.06
282	4,553	2.12	62,711	29.18
284	4,709	2.19	67,420	31.37
287	4,571	2.13	71,991	33.49
289	4,563	2.12	76,554	35.62
291	4,570	2.13	81,124	37.74
292	4,636	2.16	85,760	39.90
294	4,770	2.22	90,530	42.12
296	4,577	2.13	95,107	44.25
298	4,666	2.17	99,773	46.42
300	4,648	2.16	104,421	48.58
302	4,709	2.19	109,130	50.77

Table R24. Grade 8 Mathematics Scale Score Frequency Distribution (cont.)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
303	4,659	2.17	113,789	52.94
305	4,584	2.13	118,373	55.07
307	4,535	2.11	122,908	57.18
308	4,451	2.07	127,359	59.26
310	4,440	2.07	131,799	61.32
312	4,299	2.00	136,098	63.32
314	4,241	1.97	140,339	65.29
315	4,248	1.98	144,587	67.27
317	4,288	2.00	148,875	69.27
319	4,169	1.94	153,044	71.21
320	4,058	1.89	157,102	73.09
322	3,880	1.81	160,982	74.90
324	3,870	1.80	164,852	76.70
326	3,805	1.77	168,657	78.47
328	3,629	1.69	172,286	80.16
329	3,548	1.65	175,834	81.81
331	3,462	1.61	179,296	83.42
333	3,332	1.55	182,628	84.97
335	3,216	1.50	185,844	86.47
337	3,099	1.44	188,943	87.91
339	2,859	1.33	191,802	89.24
341	2,876	1.34	194,678	90.58
344	2,639	1.23	197,317	91.80
346	2,562	1.19	199,879	93.00
349	2,353	1.09	202,232	94.09
351	2,217	1.03	204,449	95.12
354	2,026	0.94	206,475	96.07
358	1,767	0.82	208,242	96.89
361	1,641	0.76	209,883	97.65
366	1,414	0.66	211,297	98.31
371	1,226	0.57	212,523	98.88
377	1,008	0.47	213,531	99.35
387	762	0.35	214,293	99.70
395	466	0.22	214,759	99.92
403	173	0.08	214,932	100.00

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Psychological Association, Inc.

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.

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) *Agreement 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.

Thissen, D. (1991). *MULTILOG* [Computer program]. Chicago, IL: Scientific Software, Inc.

Thompson, S.J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal Design Report 44*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, http://education.umn.edu/NCEO/OnlinePubs/Sysnthesis44.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.

Wright, B.D. & J. M. Linacre (1992). *BIGSTEPS Rasch Analysis* [Computer program]. Chicago, IL: MESA Press.

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