



Pennsylvania
Department of Education

2024–2025 FIREFLY ASSESSMENT TECHNICAL REPORT
MATHEMATICS, ENGLISH LANGUAGE ARTS, AND SCIENCE

Provided by Data Recognition Corporation

This document has been formatted to be ADA compliant.

TABLE OF CONTENTS

Glossary of Common Terms5

Chapter One: Pennsylvania Firefly Assessments10

Chapter Two: Test Development11

 Background for the PSSA Assessment Anchors and Eligible Content11

 Background for the Keystone Assessment Anchors and Eligible Content.12

Chapter Three: Item Development Processes18

 Item Development Considerations18

 Item and Test Development Cycle20

 General Item and Test Development Process20

Chapter Four: Item Pools for 2024–202524

 Foundation of the Operational Item Pool24

 Embedded Field-Test Item Pool34

Chapter Five: Field Test36

 Administration of Field-Test Items36

 Classical Item Analysis37

 Differential Item Functioning40

 Criteria for Identifying Items45

 Review of Items with Data46

Chapter Six: Operational Item Statistics48

 Item-Level Statistics48

 Item Exposure49

Chapter Seven: Rasch Item Calibration54

 Description of the Rasch Model54

 Checking Rasch Assumptions55

 Rasch Item Statistics61

Chapter Eight: Pre-Equating and Scale Maintenance70

 Pre-Equating70

 Scale Maintenance and Equating70

 Results and Observations71

Chapter Nine: Scaling73

 Raw Scores to Rasch Ability Estimates73

 Rasch Ability Estimates to Scale Scores73

Chapter Ten: Benchmarks And Cut Scores	76
CDT Benchmarking Activities	76
CDT Benchmarks	76
Firefly Cut Scores	77
Validation of Firefly Cut Scores	78
Chapter Eleven: Operational Test Design and Computer Adaptive Test Configurations	80
Operational Test Design	80
Computer Adaptive Test Algorithm	83
Firefly CAT Configurations	86
Simulation Results	87
Chapter Twelve: Test Administration Procedures	98
Test Setup	98
Timing and Setting	98
Pennsylvania (PA) Online Assessments Software	99
Training and Customer Service Support	99
Chapter Thirteen: Scores and Score Reports	100
Report Development	100
Score Reports	100
Accessing Score Reports	102
Chapter Fourteen: Operational Administration 2024–2025	104
Data Sources and Filtering Criteria	104
Key Validation Data	104
Data Review	104
Final Data	104
Firefly Matched to Spring Summative	104
Administration Counts	106
Demographic Characteristics	107
Test Length	110
Student Performance	112
Student Progress and Progress Levels	117
Chapter Fifteen: Predicting Performance on Summative Tests	120
Evaluating 2024–2025 Predictions	120
2025–2026 Prediction Modeling	120
Interpreting Predictions	124

Chapter Sixteen: Reliability	126
Standard Error of Measurement	126
Rasch Conditional Standard Errors of Measurement	128
Results and Observations	129
Reliability Indices	136
Coefficient Alpha	137
Marginal Reliability	137
Results and Observations	139
Decision Consistency	139
Chapter Seventeen: Validity	142
Purposes and Intended Uses of Firefly	142
Evidence Based on Test Content	142
Evidence Based on Response Process	144
Evidence Based on Internal Structure	144
Evidence Based on Relationships with Other Variables	150
Evidence Based on Consequences of Tests	152
Evidence Related to Use of the Rasch Model	153
Validity Evidence Summary	153
Appendix A: Operational Item Pool Summary	154
Appendix B: Item Residual Correlations	156
Appendix C: Scale Maintenance	161
Appendix D: Subscore Reliability	237
Appendix E: Reliability by Student Group	240
Appendix F: Decision Consistency – Retest Classification	247
References	250

GLOSSARY OF COMMON TERMS

The following table contains some terms used in this technical report and their meanings. Some of these terms are used universally in the assessment community, and some of these terms are used commonly by psychometric professionals.

Term	Common Definition
Ability	In Rasch scaling, <i>ability</i> is a generic term indicating the level of an individual on the construct measured by an exam. As an example for Firefly, a student's mathematics ability is measured by how the student performed on the grade-specific Firefly mathematics test.
Average	<i>Average</i> is a measure of central tendency in a score distribution that usually refers to the arithmetic mean of a set of scores. In this case, it is determined by adding all the scores in a distribution and then dividing the obtained value by the total number of scores. Sometimes people use the word average to refer to other measures of central tendency such as the median (the score in the middle of a distribution) or mode (the score value with the greatest frequency).
Benchmark Cut	A <i>benchmark cut</i> marks a specified point on a score scale where scores at or above that point are interpreted differently from scores below that point (e.g., a score designated as the minimum level of performance needed to pass a competency test). A test can be divided into multiple proficiency levels by setting one or more cut scores. Methods for establishing cut scores vary. Firefly cut scores were based on CDT cut scores, and distinguish students into three performance levels: Support Needed, Near Target, and Prepared.
Benchmark Test	A <i>benchmark test</i> is a standardized assessment administered multiple times throughout a year to assess student's mastery of content. Firefly is an on-grade benchmark test that can be administered up to three times per school year. There are two types of benchmark tests: interim assessments and formative assessments. Firefly is considered an interim assessment.
Bias	In a statistical context, <i>bias</i> refers to any source of systematic error in the measurement of a test score. In discussing test fairness, bias may refer to construct-irrelevant components of test scores that differentially affect the performance of different groups of test takers (e.g., gender, ethnicity). Attempts are made to reduce bias by conducting item fairness reviews and various differential item functioning (DIF) analyses, detecting potential areas of concern, and either removing or revising the flagged test items prior to including them in the final operational pools (see also <i>Differential Item Functioning</i>).
Computer Adaptive Test (CAT)	A <i>computer adaptive test (CAT)</i> is a computer-based test with an item selection routine that adjusts (adapts) to a student's performance during the test. For this reason, it has also been called a tailored test. Rather than all students taking the same set of items (fixed form), each student's test is individually tailored with items selected from a large item pool based on the student's performance.
Constructed-Response Item	A <i>constructed-response item</i> —referred to by some as an open-ended response item—is an item format that requires examinees to create their own responses, which can be expressed in various forms. This format is in contrast to multiple-choice items, which require students to make a choice from a supplied set of answer options. There are no constructed-response items on Firefly.
Content Validity Evidence	<i>Content validity evidence</i> shows the extent to which an exam provides an appropriate sampling of a content domain of interest (e.g., assessable portions of a state's grade 6 mathematics curriculum in terms of the knowledge, skills, objectives, and processes sampled).
Criterion-Referenced Interpretation	The <i>criterion-referenced interpretation</i> is a measure of a student's performance against an expected level of mastery, educational objective, or standard. The types of resulting score interpretations provide information about what a student knows or can do in a given content area.
Decision Consistency	<i>Decision consistency</i> is the extent to which classifications based on test scores would match the decisions on students' proficiency levels based on scores from a second parallel form of the same test. It is often expressed as the proportion of examinees who are classified the same way from the two test administrations.
Reporting Category	A <i>reporting category</i> is a grouping of content used for reporting results on Firefly tests. Each Firefly test has three to six diagnostic categories, dependent on the content area. The reporting categories are based on the Pennsylvania Academic Standards (Mathematics, Reading, and Writing) or the Pennsylvania Science, Technology & Engineering, Environmental Literacy & Sustainability (STEELS) Standards (Science).

Term	Common Definition
Differential Item Functioning (DIF)	<i>Differential item functioning (DIF)</i> is a statistical property of a test item in which different groups of test takers (who have the same total test score) have different average item scores. In other words, students with the same ability level but different group memberships do not have the same probability of answering the item correctly (see also <i>Bias</i>).
Distractor	A <i>distractor</i> is an incorrect option in a multiple-choice item.
Equating	The process that results in scores that can be used interchangeably across different test forms and/or test administrations. Equated test scores are considered exchangeable and can be interpreted in similar ways. Consequently, the requirements for <i>equating</i> are strong and somewhat complex (equal construct and precision, equity, and invariance). In practical terms, it is often stated that students should perceive no differences regardless of the test form administered (see also <i>Scale Linking</i> , <i>Pre-equating</i> , and <i>Post-equating</i>).
Evidence-Based Selected-Response (EBSR) Item	A type of item that has two parts and requires the tester to select a response from a group of possible answer choices in Part One, one of which is the correct answer (or key) to the question posed, and to then select one or two responses from a group of possible answer choices in Part Two, which provide evidence to support the correct answer in Part One. <i>Evidence-Based Selected-Response (EBSR)</i> items are only present on Firefly ELA tests.
Field-Test item	A <i>field-test item</i> is a newly developed item that is ready to be tried out to determine its statistical properties (e.g., see <i>p</i> -value and Point-Biserial Correlation). Items are field tested prior to operational administration. Items with acceptable statistical properties in field-test form the pool of Firefly operational items.
Frequency	<i>Frequency</i> is the number of times that a certain value or range of values (score interval) occurs in a distribution of scores.
Frequency Distribution	<i>Frequency distribution</i> is a tabulation of scores from low to high or high to low with the number and/or percent of individuals who obtain each score or who fall within each score interval.
Infit/Outfit	<i>Infit</i> and <i>outfit</i> are statistical indicators of the agreement of the data and the measurement model. Infit and outfit are highly correlated, and they both are highly correlated with the point-biserial correlation. Underfit can be caused when low-ability students correctly answer difficult items (perhaps by guessing or atypical experience) or high-ability students incorrectly answer easy items (perhaps because of carelessness or gaps in instruction). Any model expects some level of variability, so overfit can occur when nearly all low-ability students miss an item while nearly all high-ability students get the item correct.
Interim Assessment	An interim assessment is a type of benchmark assessment and is designed to support educators in classroom instruction and policy decisions at the classroom, school, and district level. Results are often reported in aggregate to support decisions. Firefly is an interim assessment, as is DRC's Beacon.
Item Difficulty	For the Rasch model, the dichotomous <i>item difficulty</i> represents the point along the latent trait continuum where an examinee has a 0.50 probability of making a correct response.
Key	The <i>key</i> is the correct response option or answer to a test item.
Linking	A generic term referring to a number of processes by which scores from one or more tests are made comparable to some degree. Linking includes several classes of transformations (equating, scale alignment, prediction, etc.). Equating is associated with the strongest degree of comparability (exchangeable scores). Other linkages may be very strong but fail to meet one or more of the strict criteria required of equating (see also <i>Equating</i>). Firefly scores are equated.
Logit	In Rasch scaling, <i>logits</i> are units used to express both student ability and item difficulty. When expressing student ability, if two students take the same set of items, a student who answers more items correctly has a higher logit than a student who answers fewer items correctly. Logits are transformed into scale scores through a linear transformation. When expressing item difficulty, logits and <i>p</i> -values (see also <i>P-value</i>) tend to be inversely correlated, but represent different properties. A higher logit value represents a relatively harder item, while a lower logit value would represent a relatively easier item.
Mean	<i>Mean</i> is also referred to as the arithmetic mean of a set of scores. It is found by adding all the score values in a distribution and dividing by the total number of scores. For example, the mean of the set {66, 76, 85, and 97} is 81. The value of a mean can be influenced by extreme values in a score distribution.

Term	Common Definition
Measure	In Rasch scaling, <i>measure</i> generally refers to a specific estimate of an examinee's ability (often expressed as logits) or an item's difficulty (again, often expressed as logits). As an example, for Firefly, a student's Literature measure might be equal to 0.525 logit or, a Firefly Literature test item might have a logit equal to -0.905.
Median	The <i>median</i> is the middle point or score in a set of rank-ordered observations that divides the distribution into two equal parts; each part contains 50 percent of the total data set. More simply put, half of the scores are below the median value and half of the scores are above the median value. As an example, the median for the following ranked set of scores {2, 3, 6, 8, 9} is 6.
Multiple-Choice (MC) Item	A <i>multiple-choice (MC) item</i> is a type of item format that requires the test taker to select a response from a group of possible choices, one of which is the correct answer (or key) to the question posed. Most items on Firefly tests are multiple-choice items.
N-count	Sometimes designated as <i>N</i> or <i>n</i> , it is the number of observations (usually individuals or students) in a particular group. Some examples include the number of students tested, the number of students tested from a specific subpopulation (e.g., females), and the number of students who attained a specific score. In the following set {23, 32, 56, 65, 78, 87}, <i>n</i> = 6.
Operational Item	After initial item tryout (field test), all items with acceptable statistical properties form the pool of Firefly <i>operational items</i> . Students' tests are selected from the pool of operational items.
Percent Correct	When referring to an individual item, the <i>percent correct</i> is the item's <i>p</i> -value from the field test administration expressed as a percent (instead of a proportion). For computer adaptive tests, percent correct scores are not appropriate for individual items or students.
Percentile	<i>Percentile</i> is the score or point in a score distribution at or below which a given percentage of scores fall. It should be emphasized that it is a value on the score scale, not the associated percentage (although sometimes in casual usage this misinterpretation is made). For example, if 72 percent of the students score at or below a scale score of 450 on a given test, then the scale score of 450 would be considered the 72nd percentile. As another example, the median is the 50th percentile.
Percentile Rank	The <i>percentile rank</i> is the percentage of scores in a specified distribution that fall at/below a certain point on a score distribution. Percentile ranks range in value from 1 to 99. They indicate the status or relative standing of an individual within a specified group by indicating the percent of individuals in that group who obtained equal or lower scores. An individual's percentile rank can vary depending on which group is used to determine the ranking. As suggested above, percentiles and percentile ranks are sometimes used interchangeably; however, strictly speaking, a percentile is a value on the score scale.
Point-Biserial Correlation	In classical test theory, <i>point-biserial correlation</i> is an item discrimination index. It is the correlation between a dichotomously scored item and a continuous criterion, usually represented by the total test score (or the corrected total test score with the reference item removed). It reflects the extent to which an item differentiates between high-scoring and low-scoring examinees. This discrimination index ranges from -1.00 to +1.00. The higher the discrimination index (the closer to +1.00), the better the item is considered to be performing. For multiple-choice items scored as 0 or 1, it is rare for the value of this index to exceed 0.5.
Post-Equating	<i>Post-equating</i> refers to the method of utilizing data from the current administration for scale linking and equating. Post-equating relies heavily on collecting data from a representative sample, estimating new item parameters, linking the item parameters to the base sale, and estimating student ability based on the linked item parameters. In order to provide immediate results, Firefly utilizes pre-equating. Post-equating is conducted for field-test analyses and updating item parameters.
Pre-Equating	<i>Pre-equating</i> refers to the method of utilizing previously estimated and linked item parameters for equating. Because item parameters have already been linked to the base scale, pre-equated solutions are available immediately after a Firefly is completed.
P-value	A <i>p-value</i> is an index indicating an item's difficulty for some specified group (perhaps grade). It is calculated as the proportion (sometimes percent) of students in the group who answer an item correctly. <i>P</i> -values range from 0.0 to 1.0 on the proportion scale. Lower values correspond to more difficult items and higher values correspond to easier items. <i>P</i> -values are usually provided for multiple-choice items or other items worth one point. For multi-point items, difficulty on a <i>p</i> -value-like scale can be estimated by dividing the item mean score by the maximum number of points possible for the item (see also <i>Logit</i>).

Term	Common Definition
Raw Score	<i>Raw score</i> is an unadjusted score usually determined by tallying the number of questions answered correctly or by the sum of item scores (i.e., points). Raw scores typically have little or no meaning by themselves and require additional information like the number of items on the test and the difficulty of the test items. Under a computer adaptive test administration, where each student takes a unique set of items, raw scores are not comparable across students.
Reliability	<i>Reliability</i> is the expected degree to which test scores for a group of examinees are consistent over exchangeable replications of an assessment procedure and, therefore, considered dependable and repeatable for an individual examinee. A test that produces highly consistent, stable results (i.e., relatively free from random error) is said to be highly reliable. The reliability of a test is typically expressed as a reliability coefficient or by the standard error of measurement derived by that coefficient.
Reliability Coefficient	<i>Reliability coefficient</i> is a statistical index that reflects the degree to which scores are free from random measurement error. Theoretically, it expresses the consistency of test scores as the ratio of true score variance to total score variance (true score variance plus error variance). This statistic is often expressed as a correlation coefficient (e.g., correlation between two forms of a test) or with an index that resembles a correlation coefficient (e.g., calculation of a test's internal consistency using coefficient alpha). Expressed this way, the reliability coefficient is a "unitless" index. The higher the value of the index (closer to 1.0), the greater the reliability of the test (see also <i>Standard Error of Measurement</i>).
Scale Linking	The first step in any equating process in which independent item estimates are placed on the same scale of measurement (the logit scale). <i>Scale linking</i> results in item parameters that are on the same scale of measurement. Equating procedures can only be implemented once scale linking is achieved (see also <i>Equating</i>).
Scale Score	<i>Scale score</i> is a mathematical transformation of a Rasch ability estimate developed through a process called scaling. Scale scores are most useful when comparing test results over time. Several different methods of scaling exist, but each is intended to provide a continuous and meaningful score scale across different forms of a test.
Standard Deviation	<i>Standard deviation</i> is a statistic that measures the degree of spread or dispersion of a set of scores. The value of this statistic is always greater than or equal to zero. If all of the scores in a distribution are identical, the standard deviation is equal to zero. The further the scores are away from one another in value, the greater the standard deviation. This statistic is calculated using the information about the deviations (distances) between each score and the distribution's mean. It is equivalent to the square root of the variance statistic. The standard deviation is a commonly used method of examining a distribution's variability since the standard deviation is expressed in the same units as the data.
Standard Error of Measurement (SEM)	<i>Standard error of measurement (SEM)</i> is the amount an observed score is expected to fluctuate around the true score. As an example, across replications of a measurement procedure, the true score will not differ by more than plus or minus one standard error from the observed score about 68 percent of the time (assuming normally distributed errors). The SEM is frequently used to obtain an idea of the consistency of a person's score in actual score units, or to set a confidence band around a score in terms of the error of measurement. Often a single SEM value is calculated for all test scores. On other occasions, however, the value of the SEM can vary along a score scale. Conditional standard error of measurement (CSEM) also indicates the degree of measurement error in scale score units but varies as a function of a student's unique set of items and actual scale score.
Step Difficulty	<i>Step difficulty</i> is a parameter estimate for multi-point, polytomously scored items in Masters' Partial Credit Model (PCM). It represents the relative difficulty of each score step (e.g., going from a score of 1 to a score of 2). The higher the value of a particular step difficulty, the more difficult a particular step is relative to other score steps (e.g., is it harder to go from a 1 to a 2, or to go from a 2 to a 3).
Technical Advisory Committee (TAC)	The <i>Technical Advisory Committee (TAC)</i> is a group of individuals (most often professionals in the field of testing) that are either appointed or selected to make recommendations for and to guide the technical development of a given testing program.
Technology-Enhanced (TE) Items	<i>Technology-Enhanced (TE)</i> items are items that capitalize on computer-based interactions for collecting response data. Examples of TE items include drop-down menus, drag-and-drop functionality, text highlighting, and other interactions.

Term	Common Definition
Validity	<i>Validity</i> is the degree to which accumulated evidence and theory support specific interpretations of test scores entailed by the purpose of a test. There are various ways of gathering validity evidence.

CHAPTER ONE: PENNSYLVANIA FIREFLY ASSESSMENTS

In 2023, the Pennsylvania Department of Education launched an initiative to introduce high-quality benchmark assessments that would be designed for Pennsylvania, aligned to Pennsylvania standards, and created to measure the same knowledge and skills as the Pennsylvania System of Assessment (PSSA) and Keystone Exams. The benchmark exams were to be a new addition to PDE's *Pennsylvania System of Assessments* program, which already included the PSSA, the Keystone Exams, and the Classroom Diagnostic Tools (CDT). A key component of this initiative was that the resulting benchmark assessments would provide reliable predictions for student performance on the spring summative exams—PSSA and Keystone Exams.

Prior to the Fall 2024 launch of the benchmark assessments, DRC collaborated with PDE and Pennsylvania's Technical Advisory Committee (TAC) to define the test designs (test blueprints), reporting elements, and the predictive capabilities of the newly-named benchmark assessments—*Pennsylvania Firefly Student Progress Indicators (Firefly)*. Firefly was designed as a fully online, computer adaptive assessment to be administered on the same testing platform as the CDT, PSSA, and Keystone Exams.

The length of the Firefly tests was established to allow schools to administer the tests in a single class period, while at the same time including enough items at each Reporting Category to support the inclusion of sub-score reporting. The same Reporting Categories as the PSSA and Keystone Exams were included as sub-scores on student and school reports. Firefly was also designed to assess the same grades and content areas as the PSSA and Keystone Exams, including ELA and Mathematics for grades 3–8, Science for grades 5 & 8, and Algebra I, Biology, and Literature for high school. However, because PSSA Science and Keystone Biology were transitioning to new standards in 2025–2026, those subjects were not offered in the 2024–2025 Firefly administration.

To meet PDE's goal for the Firefly assessments to provide summative-assessment predictions by the second administration (2025–2026), the 2024–2025 administration was launched as an “operational pilot.” This approach was meant to encourage participation by providing high-level student reports to the LEAs. The expected outcome was for DRC to receive adequate representative participation data to develop prediction models for the 2025–2026 administration. Because Science and Biology were not offered in 2024–2025, reporting for those subjects would not include predictions in 2025–2026.

The 2024–2025 Firefly testing window opened on November 8, 2024, and closed at the beginning of the spring 2025 summative assessment windows. Testing was available throughout the entire window, and the window was defined by three distinct seasons—Fall, Winter, and Spring. Students were allowed to take an assessment a maximum of three times in the 2024–2025 administration, and schools were encouraged (but not required) to administer the assessments in each season. The seasonal test administration windows were introduced to produce more meaningful score interpretation relative to milestones throughout the school year (i.e., what a student should be expected to know at certain milestones throughout the school year). The seasonal data would be used by DRC to generate unique prediction models for each season.

CHAPTER TWO: TEST DEVELOPMENT

In AY 2024–2025, the Pennsylvania Firefly Benchmark Assessment was available for Mathematics and Algebra I, English Language Arts (3–8), and Literature (HS)¹. The assessments were administered online in a computer adaptive test (CAT) format.

The Pennsylvania Firefly consists of multiple-choice and technology-enhanced questions that align to the Pennsylvania Assessment Anchors and Eligible Content at grades 3 through high school for mathematics, reading, and writing, and the Keystone Assessment Anchors and Eligible Content for end-of-course Algebra I, and Literature, and evidence-based selected-response questions that align to the Pennsylvania Assessment Anchors and Eligible Content at grade 3 through 8 for reading. These Pennsylvania Assessment Anchors and Eligible Content were developed previously for the PSSA and Keystone Exams as described in the following sections.

BACKGROUND FOR THE PSSA ASSESSMENT ANCHORS AND ELIGIBLE CONTENT

The PSSA Assessment Anchor Content Standards and Eligible Content in Mathematics, Reading, and Writing are based on the Pennsylvania Core Standards. Although the Pennsylvania Core Standards and the Pennsylvania Academic Standards indicate what students should know and be able to do, educator concerns about the number and breadth of Academic Standards led to the Pennsylvania Department of Education (PDE) developing Assessment Anchor Content Standards (Assessment Anchors) to indicate which parts of the Academic Standards (Instructional Standards) would be assessed on the PSSA. Based on recommendations from Pennsylvania educators, the Assessment Anchors are a tool to improve the articulation of curricular, instructional, and assessment practices. The Assessment Anchors clarify what is expected across each grade span and focus the content of the standards into what is assessable on a large-scale test. The Assessment Anchor documents also serve to communicate Eligible Content, also called assessment limits, or the range of knowledge and skills from which the PSSA would be designed.

The Assessment Anchor's coding is read like an outline. The coding includes the content, grade level, Reporting Category, Assessment Anchor, descriptor (Sub-Assessment Anchor), and Eligible Content.

Each of the Assessment Anchors has one or more descriptors (Sub-Assessment Anchors) and Eligible Content varying to reflect grade-level appropriateness. The Assessment Anchors form the basis of the test design for the grades undergoing new test development. In turn, this hierarchy is the basis for organizing the total content scores (based on the core [common] sections).

With Pennsylvania's decision to adopt the Pennsylvania Core Standards based on the Common Core State Standards, committees of Pennsylvania educators met in October 2011 to write, review, and approve the Assessment Anchors and Eligible Content statements. To provide initial focus, each content and grade span committee was presented with materials specific to the content and grade span in question, including a basic blueprint structure, the Pennsylvania Academic Standards, the Pennsylvania Assessment Anchors and Eligible Content aligned to the Pennsylvania Academic Standards, the Common Core State Standards, and draft Eligible Content statements. Committees then completed an iterative process of reviewing and revising the draft Eligible Content statements followed by discussions across grade-span committees to ensure vertical articulation across the grades. The results from the committee work were evaluated by national, state, and local subject matter experts, and, following revisions, they were ultimately validated by another committee of Pennsylvania educators. Following committee approval, the Pennsylvania Core Standards-aligned Assessment Anchors and Eligible Content for English Language Arts and Mathematics were approved by the State Board of Education in September 2013.

The complete set of Assessment Anchors and Eligible Content can be referenced at PDE's website: www.education.pa.gov.

- Roll over 'Data and Reporting' in the bar across the top of the page.
- Select 'Assessment and Accountability.' Click on the link that reads 'PSSA – PA System of School Assessment.' Then click on Assessment Anchors/Eligible Content

¹ Starting in AY 2025–26, Firefly included Grade 5 Science, Grade 8 Science, and Biology; the selection of those operational items from the CDT bank and the development of additional field-test items will be addressed in future technical reports.

BACKGROUND FOR THE KEYSTONE ASSESSMENT ANCHORS AND ELIGIBLE CONTENT

The Keystone Test Blueprints—known as the Keystone Assessment Anchors and Eligible Content—are based on Pennsylvania Keystone Course Standards and the Common Core State Standards. Prior to the development of the Assessment Anchors, multiple groups of Pennsylvania educators convened to create a set of standards for each of the Keystone Exams. Derived from a review of existing standards, these Enhanced Standards (Course Standards) focus on what students need to know and be able to do in order to be college and career ready.

Although the Keystone Course Standards indicate what students should know and be able to do, Assessment Anchors are designed to indicate which parts of the Keystone Course Standards (Instructional Standards) will be assessed on the Keystone Exams. Based on recommendations from Pennsylvania educators, the Assessment Anchors were designed as a tool to improve the articulation of curricular, instructional, and assessment practices. The Assessment Anchors clarify what is expected and focus the content of the standards into what is assessable on a large-scale exam. The Assessment Anchor documents also serve to communicate Eligible Content, or the range of knowledge and skills from which the Keystone Exams are designed.

The Keystone Assessment Anchors and Eligible Content have been designed to hold together or *anchor* the state assessment system and curriculum/instructional practices in schools following these design parameters:

- **Clear:** The Assessment Anchors are easy to read and are user-friendly; they clearly detail which standards are assessed on the Keystone Exams.
- **Focused:** The Assessment Anchors identify a core set of standards that could be reasonably assessed on a large-scale assessment, which will keep educators from having to guess which standards are critical.
- **Rigorous:** The Assessment Anchors support the rigor of the state standards by assessing higher order and reasoning skills.
- **Manageable:** The Assessment Anchors define the standards in a way that can be easily incorporated into a course to prepare students for success.

The Assessment Anchors and Eligible Content are organized into cohesive blueprints, each structured with a common labeling system. This framework is organized first by Module (Reporting Category), then by Assessment Anchor, followed by Anchor Descriptor, and then finally, at the greatest level of detail, by an Eligible Content statement. The common format of this outline is followed across the Keystone Exams.

Here is a description of each level in the labeling system for the Keystone Exams.

- **Module:** The Assessment Anchors are organized into two thematic modules for each of the Keystone Exams, and these modules serve as the Reporting Categories for the Keystone Exams. The Module title appears at the top of each page in the Assessment Anchor document. The Module level is also important because the Keystone Exams are built using a Module format, with each of the Keystone Exams divided into two equally sized test modules. Each Module is made up of two or more Assessment Anchors.
- **Assessment Anchor:** The Assessment Anchor appears in the shaded bar across the top of each Assessment Anchor table in the Assessment Anchor document. The Assessment Anchors represent categories of subject matter that anchor the content of the Keystone Exams. Each Assessment Anchor is part of a Module and has one or more Anchor Descriptors unified under it.
- **Anchor Descriptor:** Below each Assessment Anchor in the Assessment Anchor document is a specific Anchor Descriptor. The Anchor Descriptor level provides further details that delineate the scope of content covered by the Assessment Anchor. Each Anchor Descriptor is part of an Assessment Anchor and has one or more Eligible Content statements unified under it.
- **Eligible Content:** The column to the right of the Anchor Descriptor in the Assessment Anchor document contains the Eligible Content statements. The Eligible Content is the most specific description of the content that is assessed on the Keystone Exams. This level is considered the assessment limit and helps educators identify the range of content covered on the Keystone Exams.

- **Enhanced Standard:** In the column to the right of each Eligible Content statement is a code representing one or more Enhanced Standards that correlate to the Eligible Content statement. Some Eligible Content statements include annotations that indicate certain clarifications about the scope of an Eligible Content.
- **Notes:** There are three types of notes included in the Assessment Anchor document:
 “e.g.” (“for example”)—sample approach, but not a limit to the Eligible Content
 “i.e.” (“that is”)—specific limit to the Eligible Content
 “Note”—content exclusions or definable range of the Eligible Content

The Assessment Anchor’s coding is read like an outline. The coding includes the Subject (Exam), Reporting Category/Module, Assessment Anchor, Anchor Descriptor, and Eligible Content. Each exam has two modules. Each Module has two or more Assessment Anchors. Each of the Assessment Anchors has one or more Anchor Descriptors, and each Anchor Descriptor has at least one Eligible Content statements (generally more than one). The Assessment Anchors form the basis of the test design for the exams undergoing test development. In turn, this hierarchy is the basis for organizing the total Module and exam scores.

Table 2–1. Sample Keystone Assessment Anchor Coding

Sample Code	Subject (Exam)	Reporting Category (Module)	Assessment Anchor (AA)	Anchor Descriptor (AD)	Eligible Content (EC)
A1.1.1.2.1	A1–Algebra I	1 – Operations and Linear Equations & Inequalities	1 – Linear Equations	2 – Write, solve, and/or graph linear equations using various methods.	1 – Write, solve, and/or apply a linear equation (including problem situations).
L.F.2.4.1	L –Literature	F – Fiction	2 – Analyzing and Interpreting Literature—Fiction	4 – Use appropriate strategies to interpret and analyze the universal significance of literary fiction.	1 – Interpret and analyze works from a variety of genres for literary, historical, and/or cultural significance.

The complete set of Assessment Anchors and Eligible Content can be referenced at *PDE’s Standards Aligned System (SAS) website* at <http://www.pdesas.org/Standard>. Assessment Anchors and Eligible Content for Grades 3–8 can be found by selecting “Download PSSA and PASA Anchors and Eligible Content” while Assessment Anchors and Eligible Content for high school courses can be found by selecting “Download Keystone Anchors.”

FIREFLY TEST DEVELOPMENT

Firefly assesses the same grade levels and content areas as Pennsylvania’s summative assessments, PSSA and Keystone, and assesses the same content standards. Student performance is disaggregated by similar reporting categories as the PSSA and Keystone Exams. The tables below outline the number of operational items by reporting category for each assessment. Because tests may include multi-point items, the number of items may differ from the total number of points on a test. In addition to the operational items, students may receive a small number of field-test items to help maintain a robust and diverse item bank.

The following tables summarize the Mathematics test designs.

Table 2–2. Math Grade 3 Test Design

Reporting Category	Number of Operational Items
A-T: Numbers and Operations in Base Ten	6
A-F: Numbers and Operations – Fractions	6
B-O: Operations and Algebraic Thinking	9
C-G: Geometry	6
D-M: Measurement and Data	9
Total	36

Table 2–3. Math Grade 4 Test Design

Reporting Category	Number of Operational Items
A-T: Numbers and Operations in Base Ten	7
A-F: Numbers and Operations – Fractions	8
B-O: Operations and Algebraic Thinking	9
C-G: Geometry	6
D-M: Measurement and Data	6
Total	36

Table 2–4. Math Grade 5 Test Design

Reporting Category	Number of Operational Items
A-T: Numbers and Operations in Base Ten	8
A-F: Numbers and Operations – Fractions	10
B-O: Operations and Algebraic Thinking	6
C-G: Geometry	6
D-M: Measurement and Data	6
Total	36

Table 2–5. Math Grade 6 Test Design

Reporting Category	Number of Operational Items
A-N: The Number System	7
A-R: Ratios and Proportional Relationships	6
B-E: Expressions and Equations	10
C-G: Geometry	6
D-S: Statistics and Probability	7
Total	36

Table 2–6. Math Grade 7 Test Design

Reporting Category	Number of Operational Items
A-N: The Number System	6
A-R: Ratios and Proportional Relationships	8
B-E: Expressions and Equations	9
C-G: Geometry	7
D-S: Statistics and Probability	6
Total	36

Table 2–7. Math Grade 8 Test Design

Reporting Category	Number of Operational Items
A-N: The Number System	6
B-E: Expressions and Equations	10
B-F: Functions	8
C-G: Geometry	6
D-S: Statistics and Probability	6
Total	36

Table 2–8. Algebra I Test Design

Reporting Category	Number of Operational Items
Operations and Linear Equations & Inequalities	--
A1.1.1 Operations with Real Numbers and Expressions	6
A1.1.2 Linear Equations	6
A1.1.3 Linear Inequalities	6
Linear Functions & Data Organizations	--
A1.2.1 Functions	6
A1.2.2 Coordinate Geometry	6
A1.2.3 Data Analysis	6
Total	36

The following tables summarize the English Language Arts (ELA) test designs.

Table 2–9. ELA Grades 3–8 Test Design

Reporting Category	Number of Operational Items
Reading	--
Literature Text	14–16
Informational Text	14–16
Reading sub-total	28–32
Writing	--
Conventions of Standard English (Writing)	8
Total	36–40

Table 2–10. Literature Test Design

Reporting Category	Number of Operational Items
Fiction	--
L.F.1 Reading for Meaning—Fiction	6–8
L.F.2 Analyzing and Interpreting Literature—Fiction	9–11
Non-fiction	--
L.N.1 Reading for Meaning—Nonfiction	6–8
L.N.2 Analyzing and Interpreting Literature—Nonfiction	9–11
Total	30–38

ITEM TYPES

The Firefly assessment employs several types of test questions, including standalone Multiple-Choice (MC) questions (all content areas), passage-based Multiple-Choice questions (Reading and Literature only), Evidence-Based Selected-Response (EBSR) questions (Reading only), and Technology-Enhanced (TE) questions (all content areas). Firefly does not contain Short-Answer (SA) questions or Text-Dependent Analysis (TDA) questions.

PASSAGE-BASED MULTIPLE-CHOICE ITEMS

Passage-based multiple-choice items measure how well students comprehend the overall meaning of a passage or make basic inferences about it. At times, asking students to choose a preferred answer is the best way to determine whether they have gleaned certain information from a story. Such information may include setting, central idea, or main events and their sequence. These multiple-choice items are aligned to Reporting Categories within the Reading Reporting Cluster.

Each reading multiple-choice item has four response options, only one of which is correct. The student is awarded one point for choosing the correct response. Incorrect response choices, or distractors, typically represent some kind of misinterpretation, predisposition, unsound reasoning, or casual reading of the item and/or stimuli.

STANDALONE MULTIPLE-CHOICE ITEMS

Writing

Standalone multiple-choice items require that a student demonstrate both passive (recognizing and identifying grammatical and mechanical errors in text, such as misspellings, errors in word choice, errors in verb tense, or pronoun usage) and active (choosing the appropriate correction of an embedded error, such as deleting an irrelevant detail, changing the sequence of details, or placing correct marks of punctuation) language skills related to conventions of standard English and knowledge of language.

All language multiple-choice items have four response options that include only one correct answer. The student is awarded one raw score point for choosing the correct response. Incorrect response choices, or distractors, typically represent some kind of misinterpretation or predisposition, unsound reasoning, or casual reading of the item and/or stimuli.

Math

As on the PSSA, the majority of the mathematics items included on the Firefly assessment are multiple-choice (selected-response) items. This item type is especially efficient for measuring a broad range of content. In the PSSA mathematics assessment, each multiple-choice item has four response options, only one of which is correct. The student is awarded one point for choosing the correct response. Distractors typically represent incorrect concepts, incorrect logic, incorrect application of an algorithm, or computational errors.

Multiple-choice items are used to assess a variety of skill levels, from short-term recall of facts to problem solving. Firefly items involving application emphasize the requirement to carry out some mathematical process to find an answer, rather than simply recalling information from memory.

TECHNOLOGY-ENHANCED ITEMS

A small number of technology-enhanced (TE) items are included on the Firefly assessment, in proportion to the number slated to appear on the PSSA and Keystone assessments. Each TE is designed to elicit evidence of a broad range of student understanding. A student interacts with the enhanced features of these computer-delivered, auto-scorable (machine scored) test items to show an understanding of standards, skills, and concepts taught. These item types use specific enhancements to augment the user interface, such as interactive tables and diagrams, on-screen manipulatives, and selective-response generators (such as drop-down lists, hot spot, number line and coordinate graphing, data displays, and matching matrices).

EVIDENCE-BASED SELECTED-RESPONSE ITEMS

Each two-part evidence-based selected-response (EBSR) question is designed to elicit an evidence-based response from a student who has read either a Literature or Informational Text passage. In Part One, which is similar to a multiple-choice question, the student analyzes a passage and chooses the best answer from four answer choices. In Part Two, the student elicits evidence from the passage to select one or more answers based on his/her response to Part One. Part Two is different from a multiple-choice question in there may be more than four answer options and more than one correct answer. Each EBSR test question is worth either two or three points, and students can receive partial credit for providing a correct response to Part One or for providing one or more correct responses in Part Two. The student is awarded one raw score point for choosing each correct response. Incorrect response choices, or distractors, in both Part One and Part Two typically represent some kind of misinterpretation, predisposition, unsound reasoning, or casual reading of the item and/or stimuli.

CHAPTER THREE: ITEM DEVELOPMENT PROCESSES

Because the 2024–2025 Firefly assessment was launched using existing content from the PA Classroom Diagnostic Tools (CDT) bank; from legacy PSSA and Keystone banks; and from the DRC Beacon bank, the following chapter describes *in general terms* the processes used to develop the items in the banks from which the Firefly pool was built¹. Processes for Pennsylvania-specific development align with industry-wide best practices, so the development of the DRC Beacon content was similar in sequence and rigor.

ITEM DEVELOPMENT CONSIDERATIONS

The primary considerations in the item development process were alignment to the client's state standards: PSSA and Keystone Assessment Anchors and Eligible Content grade- or course-level appropriateness (as specified by PDE) for items sourced from the Pennsylvania projects, Common Core State Standards for items sourced from DRC Beacon; depth of knowledge (DOK); item/task level of complexity; estimated difficulty level; relevance of context to student experience; style; accuracy; and correct terminology. The *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) and the *Principles of Universal Design* (Thompson, Johnstone, & Thurlow, 2002) guided the development process. In addition, DRC's *Bias, Fairness, and Sensitivity Guidelines* were used for developing items. All items were reviewed for fairness by bias and sensitivity committees and for content by Pennsylvania educators and field specialists.

BIAS, FAIRNESS, AND SENSITIVITY OVERVIEW

At every stage of the item and test development process, DRC employs procedures that are designed to ensure that items and tests meet Standard 7.4 of the Standards for Educational and Psychological Testing (AERA, APA, NCME, 1999).

Standard 7.4: Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain.

To meet Standard 7.4, DRC employs a series of internal quality steps. DRC provides specific training for test developers, item writers, and reviewers on how to write, review, revise, and edit items for issues of bias, fairness, and sensitivity (as well as for technical quality). Training also includes an awareness of and sensitivity to issues of cultural diversity. In addition to providing *internal* training in reviewing items to eliminate potential bias, DRC also provides *external* training to the review panels of minority experts, teachers, and other stakeholders.

DRC's guidelines for bias, fairness, and sensitivity include instruction in how to eliminate language, symbols, words, phrases, and content that might be considered offensive by members of racial, ethnic, gender, or other groups. Areas of bias that are specifically targeted include, but are not limited to, stereotyping, gender, regional/geographic, ethnic/cultural, socioeconomic/class, religious, experiential, and biases against a particular age group (ageism) or persons with disabilities. DRC catalogues topics that should be avoided and maintains balance in gender and ethnic emphasis within the pool of available items and passages.

UNIVERSAL DESIGN OVERVIEW

The Principles of Universal Design were incorporated throughout the item development process to allow the widest possible range of students to participate in the Firefly assessment. The following checklist was used as a guideline for writing and evaluating items:

- Items measure what they are intended to measure.
- Items respect the diversity of the assessment population.
- Items have a clear format for text.
- Stimuli and items have clear pictures and graphics.

¹ In 2025–26, the Firefly assessment included new field-test items developed specifically for this program, in Reading, Writing, Math, Algebra I, Science Grade 5, Science Grade 8, and Biology. The processes for this development will be detailed in a future technical report.

- Items have concise and readable text.
- The arrangement of the items on the test has an overall appearance that is clean and well organized.

DEPTH OF KNOWLEDGE (DOK) OVERVIEW

An important element in statewide assessments is the alignment between the overall assessment system and the state’s standards. A methodology developed by Norman Webb (1999, 2006) offers a comprehensive model that can be applied to a wide variety of contexts. With regard to the alignment between standards statements and the assessment instruments, Webb’s criteria include five categories, one of which deals with content. Within the content category is a useful set of levels for evaluating depth of knowledge (DOK). According to Webb (1999), “depth-of-knowledge consistency between standards and assessments indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards” (p. 7–8). The four levels of cognitive complexity (i.e., depths of knowledge) are as follows:

- Level 1: Recall
- Level 2: Application of Skill/Concept
- Level 3: Strategic Thinking
- Level 4: Extended Thinking

Depth-of-knowledge levels were incorporated in the item writing and review process, and items were coded with respect to the level each represented.

PASSAGE/SCENARIO READABILITY OVERVIEW

In addition to items, Firefly also comprises numerous passages (in Reading and Literature). Evaluating the readability of a passage or scenario is essentially a judgmental process by individuals familiar with the classroom context and what is linguistically appropriate. Although various readability indices were computed and reviewed when evaluating prospective Firefly texts, it is recognized that such methods measure different aspects of readability and are often fraught with particular interpretive liabilities. Thus, the commonly available readability formulas were not used in a rigid way, but more informally to provide several snapshots of a passage that senior test development staff considered along with experience-based judgments. In addition, passages and scenarios were reviewed by committees of educators, who evaluated them for readability and grade-level appropriateness, and bias-trained professionals, who analyzed the passages through a bias and sensitivity lens.

In addition to the best practices for passage-composition or selection identified above, passages were selected for Firefly use based on two further criteria specific to the project. First, they could not be as long as the longest PSSA and Keystone passages, as the Firefly test is intended to be administered in the span of a single class period. Second, they had to be either commissioned or in the public domain—that is, not copyrighted—as the expense of acquiring permissions for the number of “views” that Firefly passages would receive was not compatible with Pennsylvania’s and DRC’s budget goals and the commitment to providing a free benchmark assessment.

TEST ITEM READABILITY OVERVIEW

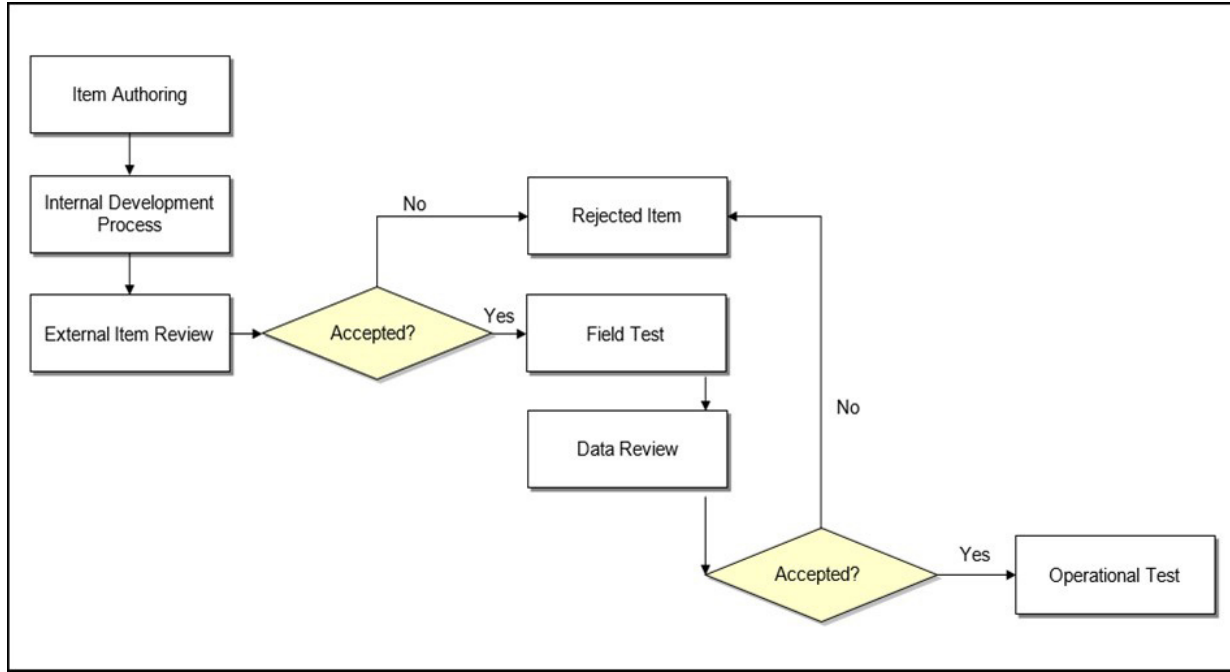
Careful attention was given to the readability of the items to ensure that the assessment focus of the item did not shift based on the difficulty of reading the item. Subject/course areas such as Mathematics and Algebra I contain many content-specific vocabulary terms. As a result, readability formulas were not used. However, wherever it was practicable and reasonable, every effort was made to keep the vocabulary at, or one level below, the grade or course level for non-Reading/Literature items. This was a conscious consideration made to ensure that each question was evaluating a student’s mastery of the course standards rather than the student’s reading ability. Resources used to verify the vocabulary level were the *EDL Core Vocabularies* and the *Children’s Writer’s Word Book*.

In addition, every test question was brought before committees of educators who are course-level/grade-level experts in the content area of the questions. They reviewed each question from the perspective of the students they teach, determined the accessibility of the vocabulary used, and worked with DRC to minimize the level of reading required.

ITEM AND TEST DEVELOPMENT CYCLE

The development process for items followed a logical cycle and timeline, which is outlined in the figure on the following page. On the front end of the schedule, tasks were generally completed with the goal of presenting field-test candidate items to committees of educators. On the back end of the schedule, all tasks lead to the field-test data review and operational test construction. This presentation represents a typical life cycle for a field-test event.

Figure 3–1. Item and Test Development Cycle



GENERAL ITEM AND TEST DEVELOPMENT PROCESS

ITEM WRITER TRAINING

All items, from each of the sources that make up the Firefly assessment, were the result of a similar process. Qualified writers were college graduates with teaching experience and a demonstrated base of knowledge in the content area. Many of these writers were content assessment specialists and curriculum specialists. The writers were trained individually and had previous experience in writing multiple-choice and technology-enhanced items. Prior to developing items for DRC item writers were trained in the following areas:

- PSSA and Keystone Assessment Anchors and Eligible Content or the Common Core State Standards
- Webb’s Levels of Cognitive Complexity, Depth of Knowledge
- Bias, Fairness, and Sensitivity Guidelines
- Principles of Universal Design
- Item Quality Technical Style Guidelines
- Reference Information
- Sample Items

READING/LITERATURE PASSAGE DEVELOPMENT

Passages selected for Firefly had been either developed internally or culled from public domain sources. The task of initially developing passages was conducted by DRC professionals with classroom experience in reading/English language arts. These professionals also underwent specialized training (provided by DRC) in the characteristics of quality passages. Guidelines for passage selection included appropriate length, text structure; text density; and vocabulary; each of these characteristics was evaluated in the context of the particular time limitations of the Firefly assessment. A judgment was also made about whether the reading level required by a particular passage was at the independent level—that is, where the average student should be able to read 90 percent of words in the text

independently. Passages were further selected to present a balance of fiction and informational content within each grade as well as a mix of genres within each mode.

In addition, DRC identified appropriate public domain passages to help cultivate a range of topics, tones, and difficulty levels in the Firefly passage pool. Regardless of the passages' origins, each had at some point undergone an internal review by several test development content experts to judge their merit with regard to the following criteria:

- Interest value for students
- Passages demonstrate the qualities of good writing and relevance to students' lives, without being overly timely.
- Use of appropriate grade-level vocabulary and language characteristics
- Freedom from bias, fairness, and sensitivity issues
- Representation of different cultures and lived experiences
- Sufficient richness to generate a variety of multiple-choice and technology-enhanced items across a range of eligible content standards
- Passages avoid dated subject matter unless a relevant historical context is provided.
- Passages should not require students to have extensive background knowledge in a certain discipline or area to understand a text.

Once through the internal review process, those passages deemed potentially acceptable were reviewed by the Reading Content Committee and Bias, Fairness, and Sensitivity Committee for final approval.

ITEM AUTHORING AND TRACKING

Initially, items were created to DRC orders and submitted via a secure transfer portal on Word documents with a locked template. The template provided a field for student-facing item content and numerous additional fields for essential metadata: subject area, grade, content categories, passage information (in the case of reading), item type, depth of knowledge, estimated difficulty, and answer key.

As items were submitted by the writers, content leads, working with TD Tech colleagues, uploaded them to the DRC Item Development and Educational Assessment System (IDEAS), a comprehensive, secure, online item banking system. It accommodates item writing, item viewing and reviewing, and item tracking and versioning. IDEAS manages the transition of an item from its developmental stage to its approval for use within a test form. It is at this phase that passage and item graphics, if needed, were requested and developed to client specifications by the DRC Media and Composition team. Additionally, any interactive elements of technology-enhanced items, such as drop-down menus, clickable charts or tables, or text-highlighting fields, were formally integrated into the items from sketches or verbal descriptions on the initial Word document submissions.

INTERNAL REVIEWS

As part of the development process, each item was further reviewed by internal groups at DRC. Content specialists evaluated each item to make sure that it measured the intended Eligible Content and Assessment Anchor or state standard, that it was appropriate for the intended grade, and that it provided only one correct answer. In addition, the difficulty level, depth of knowledge, graphics, language demand, quality of distractors, and, where relevant, functionality of tech enhancements were also evaluated. Other elements considered in this process include, but are not limited to, Universal Design, bias, source of challenge, grammar/punctuation, and adherence to project style.

Following these internal content reviews, items were evaluated by an internal Bias, Fairness, and Sensitivity expert and by the DRC Editorial Services team. After each additional review, DRC content leads evaluated the feedback, integrating or adapting it where deemed beneficial.

ITEM CONTENT REVIEWS

Prior to field testing, all newly developed test items were reviewed by content committees. The content committees consisted of educators—some with postsecondary university affiliations.

The primary responsibility of the content committee was to evaluate items for general quality and content classification, including grade-level or course appropriateness, estimated difficulty, depth of knowledge, and source of challenge. Ascribing source of challenge differs from gauging estimated difficulty or depth of knowledge in that it attempts to describe or anticipate the element of an item that may be responsible for its cognitive demand and evaluate whether that element is intrinsic to, or irrelevant to, the standard or the construct intended to be assessed. Specifically, items are identified where the cognitive demand is focused on an unintended content, concept, or skill (Webb, 2002). In addition, source of challenge may be attributed if the reason that an answer could be given results from a cultural bias embedded in the item; an inappropriate reading level in the item stimulus, stem, or options; a flawed or unnecessarily difficult to decode graphic in an item; or the possession of specialized, non-content related knowledge by the test-taker. Source of challenge could result in a student who has mastered the intended content or skill answering the item incorrectly or a student who has not mastered the intended content or skill answering the item correctly. Committee members were asked to note any items with a source of challenge and to suggest revisions to remove the source of challenge. They also suggested revisions and made recommendations for reclassification of items. The committee members also reviewed the items for adherence to the Principles of Universal Design, including language demand and issues of bias, fairness, and sensitivity.

Committee members, grouped by content area, received training by working through and reviewing a group of items for quality and content, as well as for the following categories:

- Assessment Anchor/Common Core State Standard alignment
- Content Limits
- Grade-Level (Course-Level) Appropriateness
- Difficulty Level
- Depth of Knowledge
- Appropriate Source of Challenge
- Correct Answer
- Quality of Distractors
- Appropriateness of Graphics
- Calibration of Language Demand
- Freedom from Bias

The members then received a binder containing items to independently review and provided their recommendation for the status of each item: Approved, Accepted with Revision, or Rejected. All comments were reviewed and addressed by DRC content staff.

Security was addressed by adhering to a strict set of procedures. All attendees were required to sign a confidentiality agreement. All materials not in use at any time were stored in a locked room. Secure materials that did not need to be retained after the meetings were deposited in secure barrels, the contents of which were shredded.

BIAS, FAIRNESS, AND SENSITIVITY REVIEWS

Prior to field testing, all newly developed test items were also submitted to a Bias, Fairness, and Sensitivity Committee for review. These reviews took place prior to the Item Content Review for each content area. The committee's primary responsibility was to evaluate items with regard to bias, fairness, and sensitivity issues. They also made recommendations for changes or deletion of items in order to remove the potential for issues of bias, fairness, and/or sensitivity. Included in the review were proposed reading passages. An expert, multi-ethnic committee composed of men and women was trained by a DRC test development lead to review items for bias, fairness, and sensitivity issues. Training materials included a manual developed by DRC (DRC, 2003–2013). Members of the committee also had expertise with special-needs students and English Language Learners. All items were read by a cross-section of committee members. Each member noted bias, fairness, and/or sensitivity comments on tracking sheets and on the item, if needed, for clarification. Committee members individually categorized any concerns as related to ageism, disability, ethnicity/culture, gender, region, religion, socioeconomics, or stereotypes. These categories were the framework through which recommendations for modification or rejection of items occurred during the subsequent committee consensus process. The committee discussed each of the issues as a group and came to a consensus as to which issues should represent the view of the committee. All consensus comments were then compiled, and the suggested actions on these items were recorded and submitted to DRC content staff. This review followed the same security procedures as outlined above.

CHAPTER FOUR: ITEM POOLS FOR 2024–2025

This chapter details the Firefly item pools used for the first year of administration in 2024–2025. Several decisions were made during the inception of Firefly to support a relatively quick launch and adoption of the benchmark test into Pennsylvania’s assessment system. The first decision was to lease items from Pennsylvania’s Classroom Diagnostic Tools (CDT) to develop the necessary foundation and infrastructure of the Firefly testing program. Using CDT as the foundation of Firefly was strategic in that it allowed for other CDT infrastructure to be utilized. Utilizing the CDT item pool for the launch of Firefly provided several opportunities.

First, all operational items in the CDT item pool have existing pre-equated item parameters. Leasing a subset of the CDT item pool to Firefly allowed for a speedy launch that did not require the rigorous and time-intensive process of new item development, new item review, and field-testing items prior to becoming operational. This alone provided many advantages by removing at least a year from the item cycle process. Moreover, by selecting previously operational CDT items, it allowed for score reporting in the first operational pilot year of Firefly in 2024–2025. While there were some limitations to score reporting in the first operational pilot year of Firefly, such as no scores for ELA writing (grades 3–8), the advantages outweighed the disadvantages.

Second, Firefly utilized previously estimated pre-equated item parameters which were on the baseline CDT vertical scale. The established CDT vertical scale is discussed in the CDT Technical Report (refer to Chapter Nine). One of the goals for Firefly was to develop and utilize a vertical scale so that students, educators, school administrators, and other stakeholders could make inferences about student performance within and across years within a single content area. Because the underlying Firefly (logit) scale is the same as the underlying CDT scale, item difficulties and student performance can be interpreted in similar ways.

Lastly, the CDT theta cut scores could be applied to Firefly since both programs utilize the same underlying baseline scale. No additional standard setting, standards validation, or benchmarking activity was necessary because the cut scores were previously established through a Benchmarking activity for CDT (refer to Chapter Ten)¹. Thus, scale transformations could be directly applied to the underlying CDT scale to a newly established Firefly reporting scale. Chapter Nine discusses the details for the Firefly reporting scale.

This chapter focuses on the development of the first operational item pool as well as development of the first field-test item pool. The first operational item pool used CDT item pools as the foundation due to its many advantages discussed above. Although CDT administers both on- and off-grade content to testers, Firefly was designed as an on-grade assessment. As such, operational and field-test item pools are defined for each content area and grade/level separately.

FOUNDATION OF THE OPERATIONAL ITEM POOL

The Firefly mathematics and ELA operational item pools were selected from the CDT mathematics and reading item pools. CDT items have been through a rigorous process that involved Pennsylvania educators, including item development, item review, bias review, field testing with Pennsylvania students, and data review. These steps ultimately support the use of CDT items to assess Pennsylvania standards. This section details the steps that were taken to select items from the CDT bank for Firefly for operational use in 2024–2025.

CDT ELIGIBLE ITEM POOL DEVELOPMENT

First, the CDT eligible item pools were established. The eligible item pool consisted of previously used operational CDT items in mathematics and reading grades 3 through 8 as well as high school content areas, Algebra I and Literature. These items had classical item statistics from a field-test event (refer to Chapter Five in the CDT Technical Report for additional information on field testing CDT items). Eligible item pools were refined using field-test item statistics to exclude:

- Items with p-values less than 0.05 and p-values greater than 0.95
- Items with correlations less than 0.15

¹ Following the first operational year, additional analyses were conducted to validate the CDT cut scores using Firefly data. No adjustments to the cut scores were required.

- Multi-point items with unordered or reversed step measures² (CDT reading only)
- Items with evidence of severe Differential Item Functioning³ (DIF)

For reading pools, additional filters were used to refine the eligible item pool to best mirror content on the PSSA and Keystone examinations. For example, stand-alone vocabulary items were removed from the eligible pool because PSSA only has passage-based vocabulary items. In addition, the eligible item pools were refined to exclude:

- Reading passages with less than 4 items in a passage
- Reading passages with more than two evidence-based selected-response (EBSR) items⁴

FIREFLY OPERATIONAL POOL DEVELOPMENT

After the eligible item pools were selected from the CDT bank, items and passages were selected for each grade/level separately, given Firefly is an on-grade computer adaptive test. The approaches differed for mathematics and ELA given the passage-based structure of reading tests.

For mathematics, CDT items were selected by reporting category using the general rule of thumb that healthy CAT item pools are 10 times the length of the test. The ideal number of items given that Firefly math is a fixed length test of 36 items was calculated by grade/level and reporting category. If there were enough items in the eligible pool of items (more than 10 times), then items were randomly sampled to reach the desired number of 10 times the number of items per grade/level and reporting category. If the number of items required was more than 90% of the number of eligible items or if there were fewer items in the eligible pool of items than 10 times the items per reporting category, then all items were selected for that reporting category (no sampling was performed). Meaning, all eligible items were selected from reporting categories with fewer items than the ideal quantity (10 times the number of items per test). In addition, we removed items from eligible content codes that were considered only assessable without calculators⁵.

For ELA, CDT reading items were selected in an iterative process to ensure reasonable representation of each eligible content code. The goal was to develop an operational item pool that was reflective of the PSSA and Keystone operational test blueprints. In some cases, passages and items were retained, despite their less-than-ideal statistics, to have the operational pool reflect PSSA and Keystone test blueprints. After each iteration of item/passage selection, the selected pool was evaluated in terms of passage genre, passage length, passage average item difficulty, item type, reporting category, and eligible content code.

RESULTS

Table 4–1 presents a summary of the operational item pool by content area and grade/level. As the table shows, there were between 350 and 490 items in each item pool. All of the items in the operational item pool originated from the CDT item bank. Mathematics pools consisted entirely of multiple-choice (MC) items, whereas ELA and Literature pools consisted of both MC and evidence-based selected-response (EBSR) items. The CDT item bank did not have any technology-enhanced (TE) items for mathematics, reading, or writing. Therefore, no TE items were included in the operational pools in 2024–2025.

² Only CDT reading item pools included multi-point items.

³ Refer to Chapter Five for details on DIF.

⁴ PSSA has no more than 2 evidence-based selected-response (EBSR) items per passage.

⁵ Items aligned to math grade 8 eligible content codes A-N.1.1.3 and B.E.1.1.1 were removed from the operational item pool.

Table 4–1. Operational Item Pool Summary

Content Area	Grade/Level	MC Item Count	EBSR Item Count	Total Count
Mathematics	3	467	0	467
Mathematics	4	438	0	438
Mathematics	5	394	0	394
Mathematics	6	442	0	442
Mathematics	7	397	0	397
Mathematics	8	384	0	384
Algebra I	HS	423	0	423
ELA	3	309	48	357
ELA	4	309	48	357
ELA	5	299	56	355
ELA	6	327	37	364
ELA	7	316	38	354
ELA	8	317	43	360
Literature	HS	452	40	492

Note. The 2024–2025 Firefly ELA operational item pools for grades 3 through 8 were comprised entirely of CDT reading items. In future administrations, the Firefly ELA operational item pools will include both reading and writing items.

Table 4–2 presents a summary of the operational reading passages in the ELA and Literature item pools. As previously discussed, passages with at least 3 items were selected from the CDT reading item pools. Overall, each passage has between 3 to 8 associated items. On average, there are approximately 5 items per passage, and the median item count per passage is 6 across all grades/levels.

Table 4–2. Operational Reading Passage Summary

Content Area	Grade/Level	N Passages	Min Item Ct	Max Item Ct	Average Item Ct	Median Item Ct
ELA	3	64	3	8	5.58	6
ELA	4	66	3	7	5.41	6
ELA	5	63	3	7	5.63	6
ELA	6	68	3	6	5.35	6
ELA	7	67	3	6	5.28	6
ELA	8	67	3	6	5.37	6
Literature	HS	88	3	7	5.59	6

Appendix A provides a summary of the operational item pool by reporting category and item type.

The final Firefly operational item pools were similar to the eligible grade-specific CDT item pools developed for selection in terms of range and distribution of item difficulty. Figure 4–1 shows the distribution of operational item difficulties for Firefly that served as the foundation of each content area and grade/level item pool. Additional information on operational item performance and item exposure is presented in Chapter Six.

Figure 4–1. Distribution of Operational Item Difficulties, by Content Area and Grade/Level

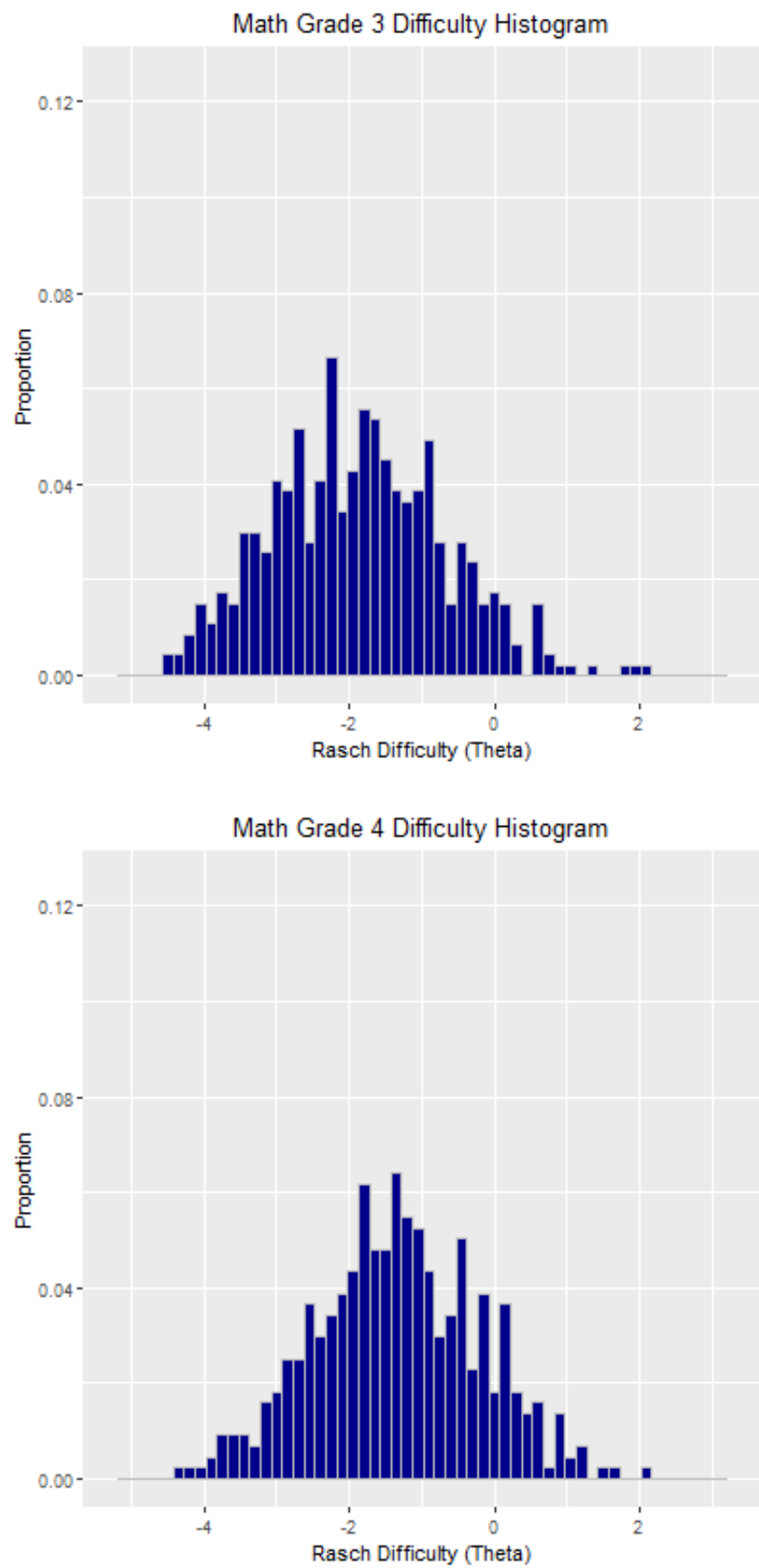


Figure 4–1 (continued). Distribution of Operational Item Difficulties, by Content Area and Grade/Level

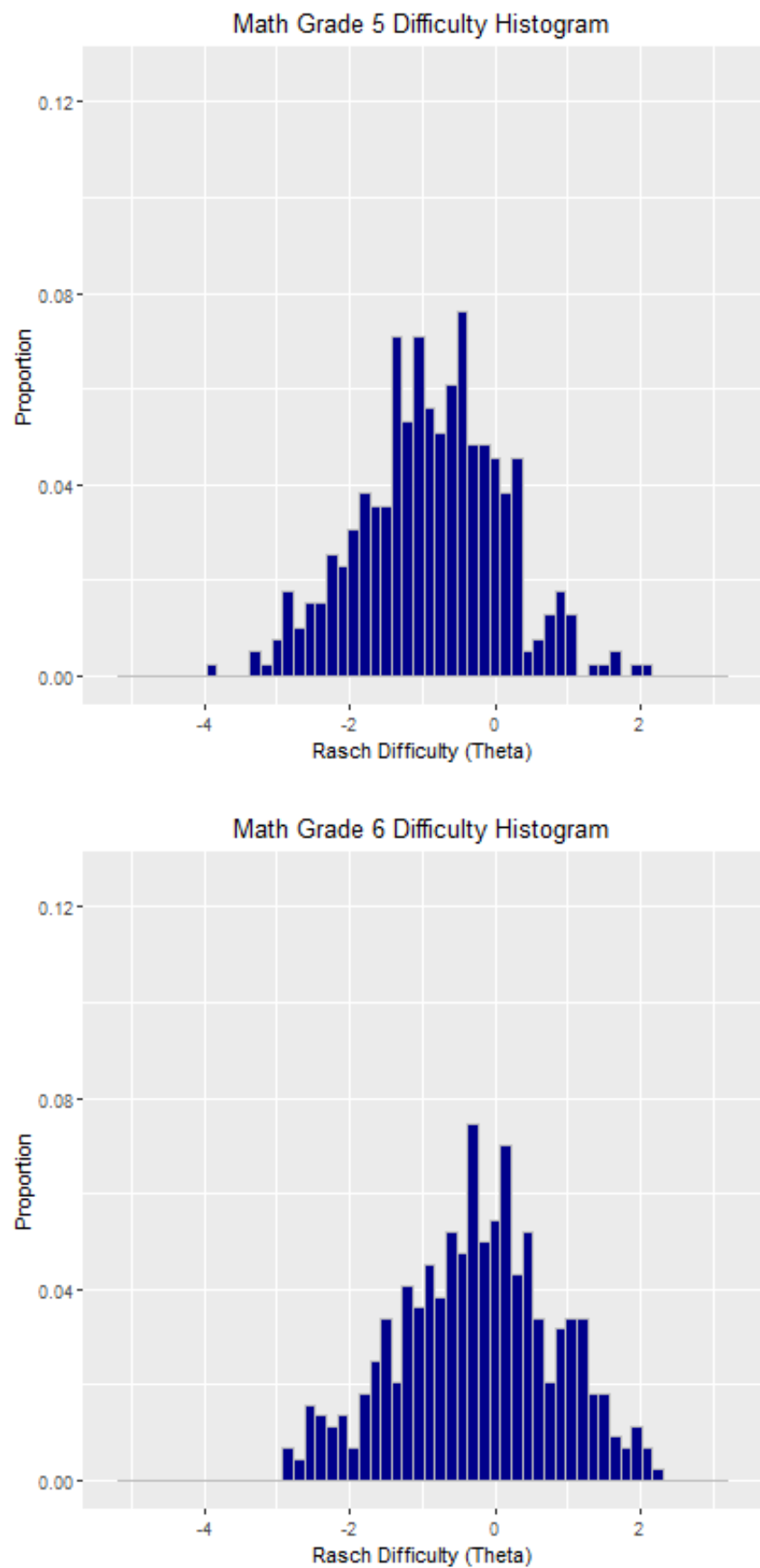


Figure 4–1 (continued). Distribution of Operational Item Difficulties, by Content Area and Grade/Level

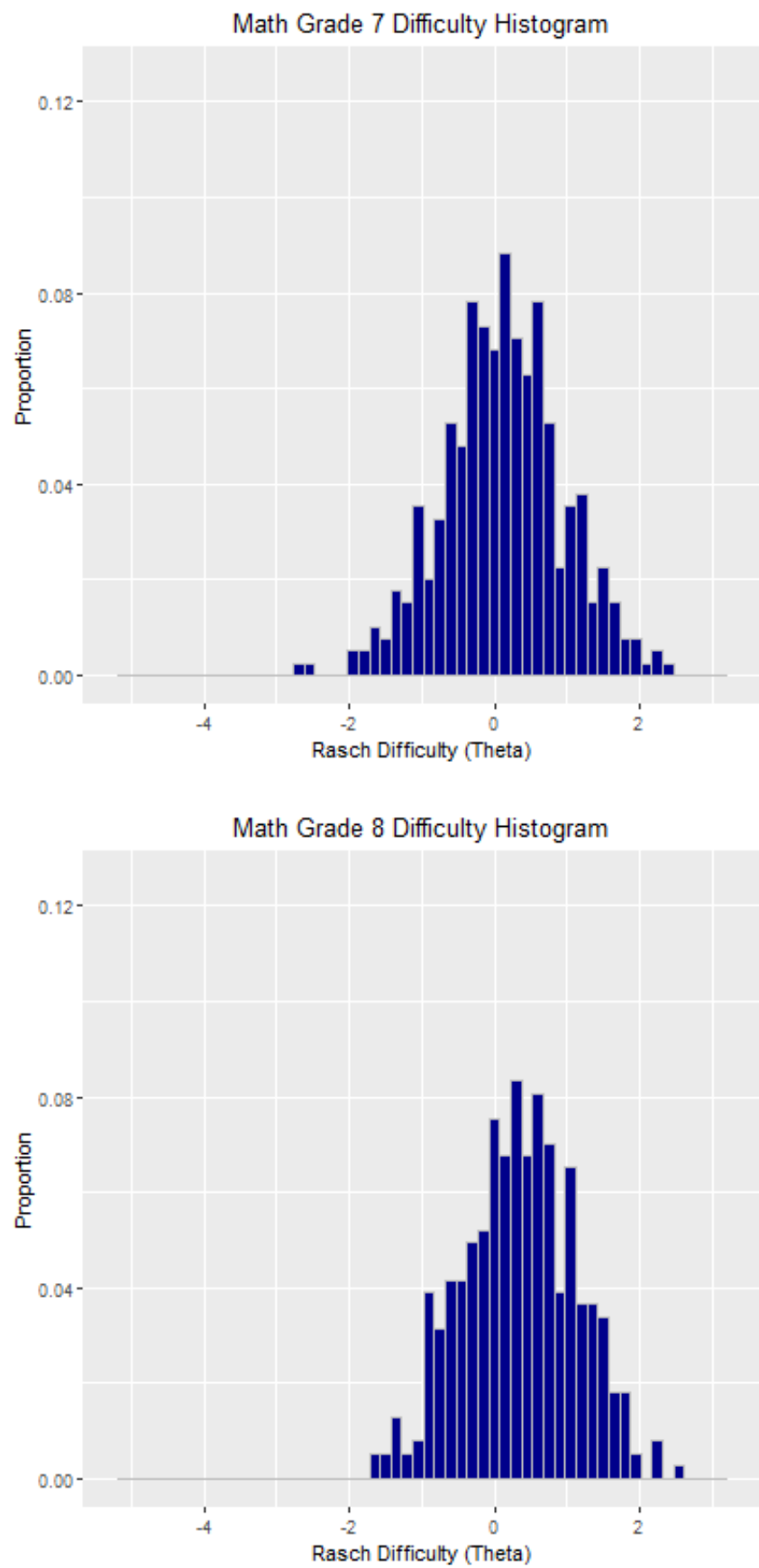


Figure 4–1 (continued). Distribution of Operational Item Difficulties, by Content Area and Grade/Level

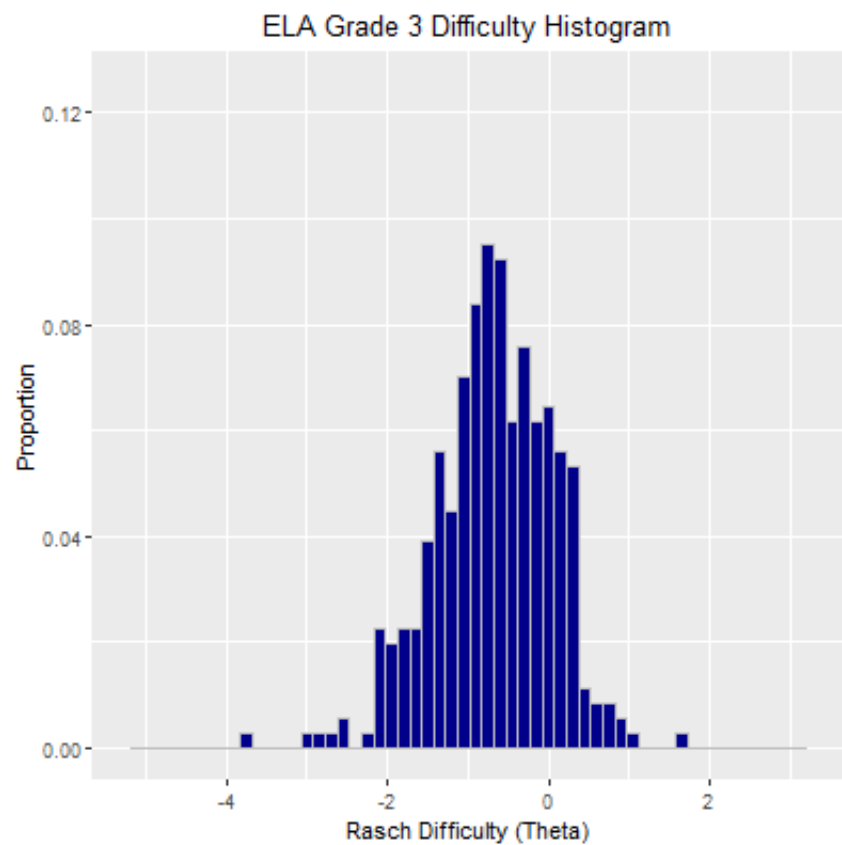
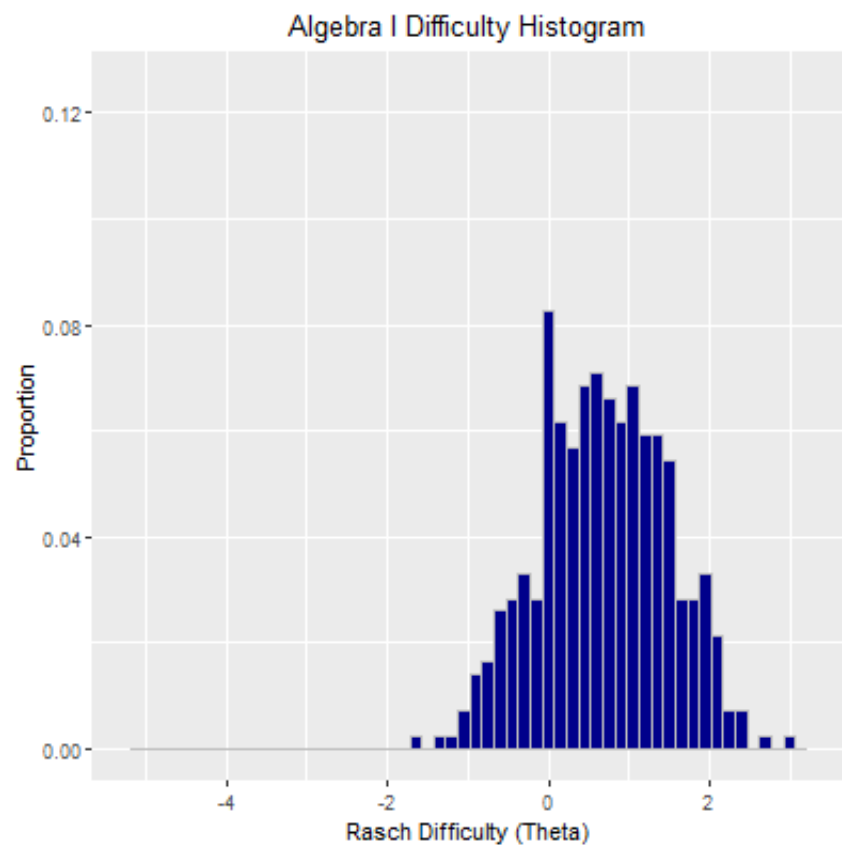


Figure 4–1 (continued). Distribution of Operational Item Difficulties, by Content Area and Grade/Level

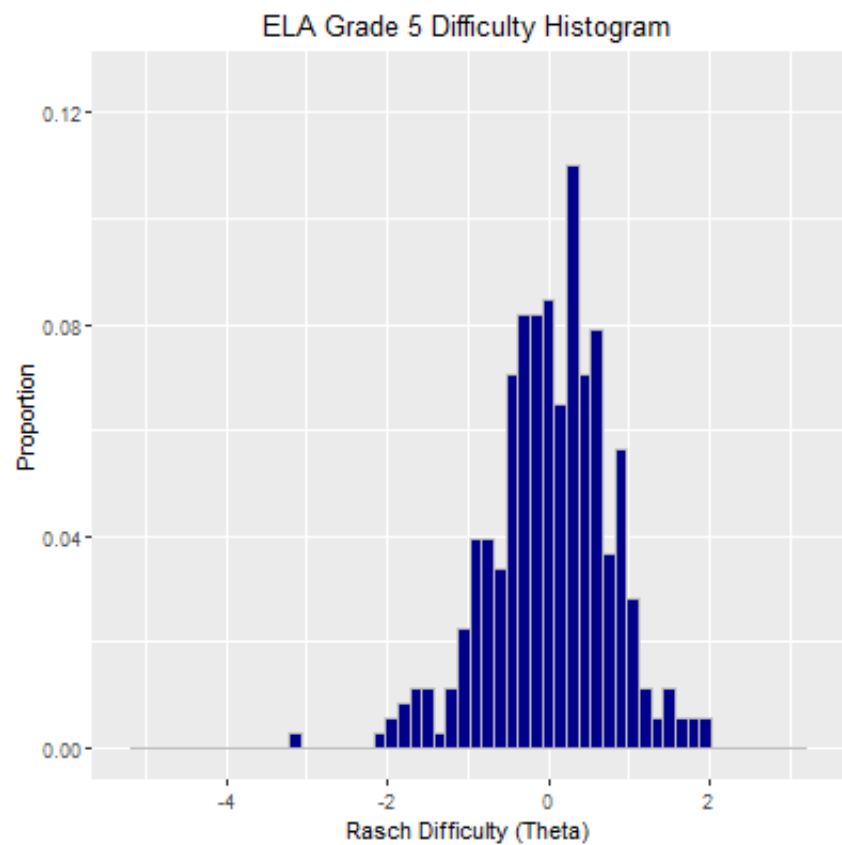
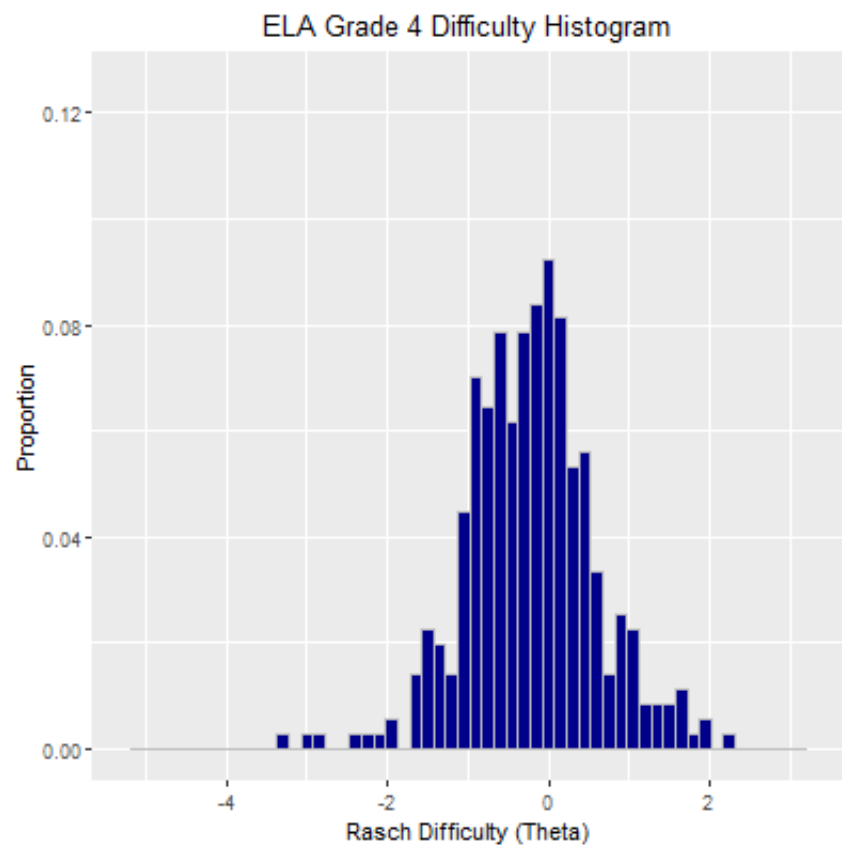


Figure 4–1 (continued). Distribution of Operational Item Difficulties, by Content Area and Grade/Level

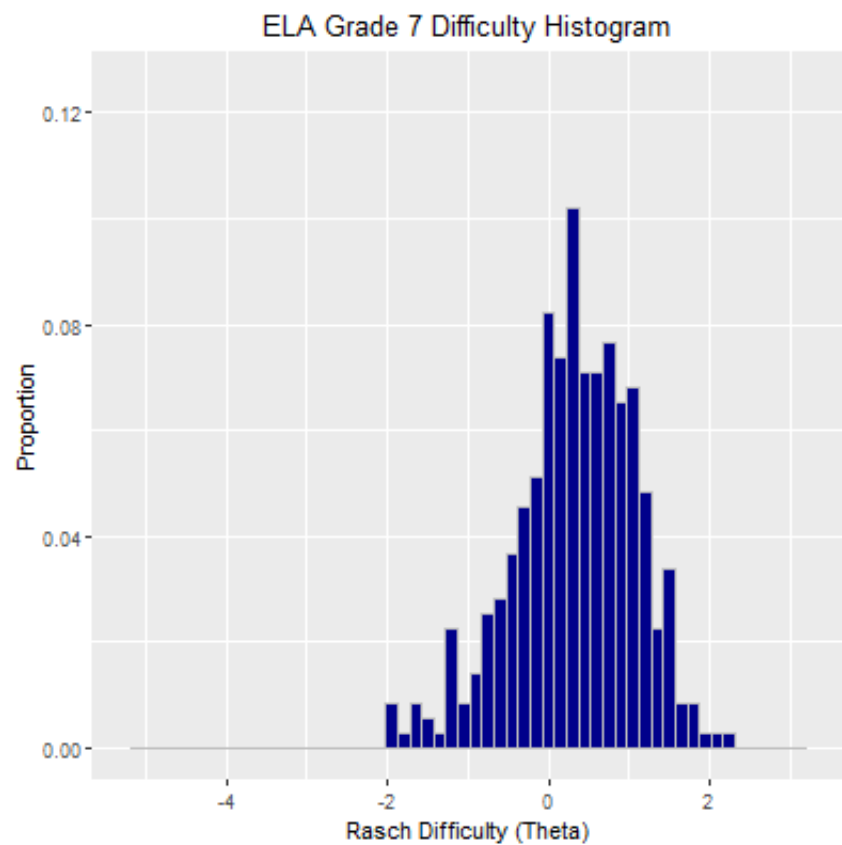
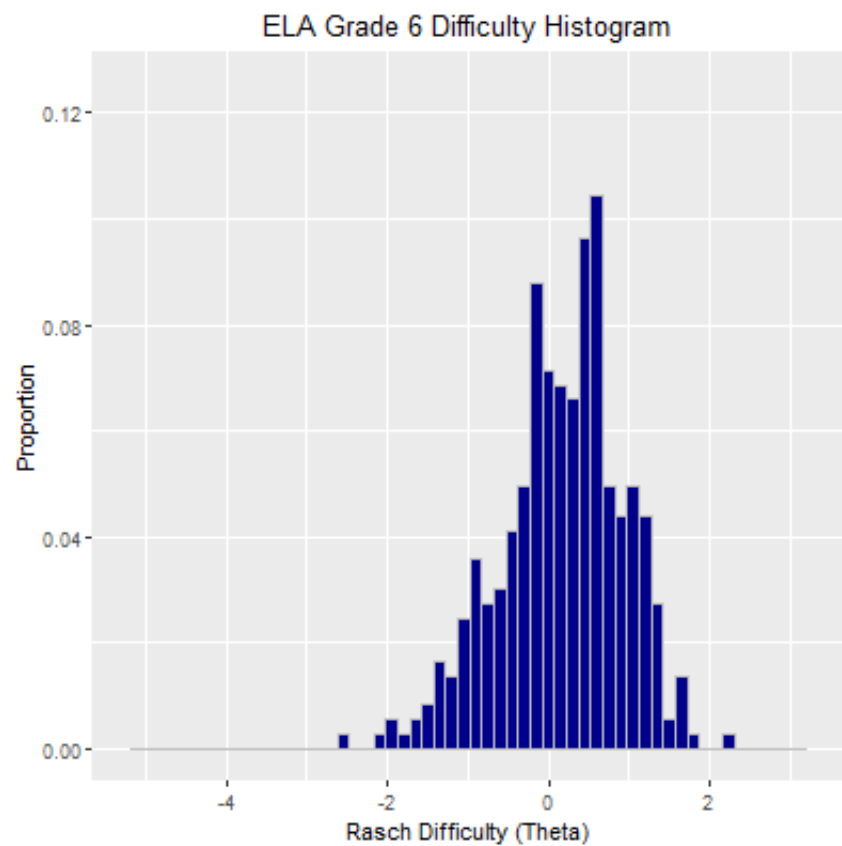
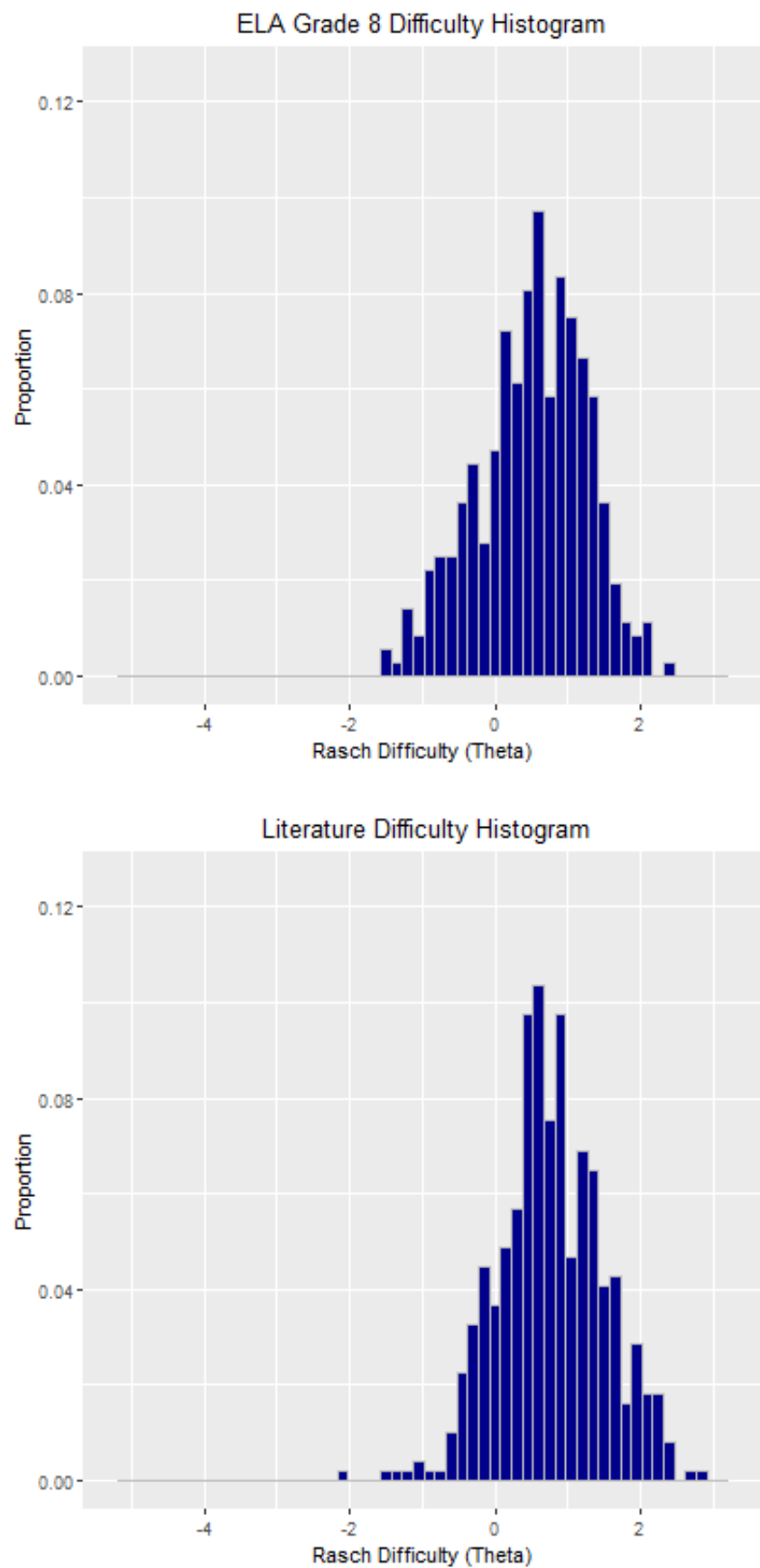


Figure 4–1 (continued). Distribution of Operational Item Difficulties, by Content Area and Grade/Level



EMBEDDED FIELD-TEST ITEM POOL

One goal of the first several years of Firefly is to bolster and expand the operational Firefly item pools. To do so, DRC culled multiple sources for the embedded field test in 2024–2025. As discussed in Chapter Five, the purpose of embedded field tests is to collect data to evaluate item quality and estimate item parameters for use in future operational item pools. There were three sources of field-test items for the 2024–2025 administration:

- CDT Writing items for Firefly ELA grades 3 through 8
- Previously administered items from PSSA and Keystone that had not been administered in over six years and were no longer eligible for future selection for PSSA or Keystone forms
- Technology-enhanced items from DRC Beacon Test (and passage sets associated with TE items)

Each of the three sources are discussed separately. Based on the large number of field-test items for each content area and grade/level as well as the modest goal participation rates of Firefly the first year (approximately 10,000 administrations per content and level), DRC preemptively split the field test item pools in half to ensure sufficient exposure and response data. Refer to Chapter Five for results from the 2024–2025 embedded field test.

CDT WRITING ITEM SELECTION

CDT writing items were selected in a similar manner to operational CDT mathematics items, using the same statistical filtering criteria outlined in the previous section. Across grades 3 to 8, there were enough items in the eligible pool of items (more than 10 times) to randomly sample to reach the desired number of 10 times the number of items. In addition, there were at least four items selected at each eligible content (EC) code within each grade. If less than four items were available for selection for an EC, then all eligible items were selected.

CDT has two separate tests for reading and writing; although the CDT writing items had pre-equated and banked parameters, the parameters were on a separate scale from CDT reading items. Thus, all CDT writing items needed to be field tested alongside Firefly reading items to ensure that item parameters were estimated on the same baseline scale, developed following the 2024–2025 Firefly administration.

PSSA AND KEYSTONE ITEM SELECTION

Previous PSSA and Keystone items were selected from their respective item banks. Multiple-choice or evidence-based selected-response items were eligible if they were last tested prior to 2017 for PSSA or field-tested prior to 2017 for Keystone, meaning items were more than six years old at the time of item pool development. Then, filtering criteria was applied to exclude items with the following characteristics:

- Items with p-values less than 0.05 and p-values greater than 0.95
- Items with correlations less than 0.15
- Items with mean-square infit values greater than 1.3, indicating poor fit with the Rasch model
- Items with evidence of severe Differential Item Functioning⁶ (DIF)
- Reading passages with fewer than three items
- Non-calculator based items in math grades 4 through 8⁷

Items selected according to the criteria above were provided to DRC Test Development specialists for review and final selection. For PSSA mathematics and Keystone Algebra I, nearly all the items were deemed appropriate for Firefly purposes. Items were previously written in alignment with PA core standards and met style and format requirements.

⁶ Refer to Chapter Five for details on DIF.

⁷ Firefly mathematics in grades 4 through 8 allows access to calculators for all items whereas PSSA mathematics in grades 4 through 8 has a small number of non-calculator items on each form. During the development of the item banks, non-calculator-based items were removed from mathematics grades 4 through 8.

There were additional considerations for items leased from PSSA ELA (both reading and writing domains) and Keystone Literature. To ensure rigor and quality, both PSSA and Keystone assessments use authentic passages as the basis for item sets; most of these are culled from published sources (e.g., magazines and novels) and so require paid permissions to administer. Acquiring the necessary rights for these passages for an indefinite term and for an unrestricted number of views would be prohibitively expensive, so the decision was made to avoid leasing these passages and passage-based items. The passages leased from PSSA and Keystone programs, then, are only those in the public domain, which do not require permission to administer. As more grade-appropriate public-domain passages exist for high-school readers than for middle or elementary school, more Keystone Literature than PSSA Reading content was ultimately selected for Firefly.

A further consideration was passage length. To support a full array of selected-response items, technology-enhanced, and constructed-response items, PSSA and Keystone passages must possess a substantiveness and richness that is often supported by passage length. However, that length, which is a virtue on PSSA and Keystone tests, can be at odds with the intended function of Firefly, which was developed to be completed in the span of a single class period. Passages in the public domain were thus further evaluated for length, and only those that supported the state's aims of a single-period assessment were selected.

Because of these two limiting considerations, a very small number of PSSA item sets, and only a small number of Keystone Literature item sets, were ultimately leased to Firefly. On the other hand, most PSSA ELA stand-alone writing items were deemed appropriate for Firefly purposes and leased to Firefly⁸.

DRC BEACON ITEM SELECTION

In order to anticipate the inclusion of technology-enhanced (TE) items in PSSA and Keystone assessments starting in 2026, it was decided that Firefly should be designed, developed, and launched with those item types. As no TE items had been written to reading, writing, or mathematics standards for CDT, PSSA, or Keystone at the inception of the Firefly program, DRC opted to draw such items from DRC Beacon, an existing interim assessment program administered at the state, district, or school level in several other states. Using these items allowed Firefly to launch with TE items available without the myriad challenges specific to their development. Helpfully, the goals of Beacon are like those of Firefly, meaning that issues of passage permissions and length were not applicable. However, the DRC Beacon content is aligned to the Common Core State Standards and adheres to its own style specifications; thus, additional rounds and types of review were integrated into the leasing process, described below.

To determine the items that could be leased to Firefly from the DRC Beacon item banks, test developers went through the following process.

1. Identified operational technology-enhanced items from DRC Beacon item bank. For mathematics, only single-point TE items were identified; for ELA, both single and multi-point TE items were identified. For ELA, when TE passage-based items were identified, the entire passage set was identified and reviewed in subsequent steps.
2. Realigned items to the Pennsylvania Core Standards. If items could not be realigned to PA Core Standards, they were removed from the potential pool of items.
3. Ensured that items met Pennsylvania style and format requirements. In some cases, edits were made to ensure consistency with PA requirements.

Items selected from DRC Beacon maintain their operational status in both Firefly and DRC Beacon item pools.

⁸ When items are leased from summative tests PSSA and Keystone to Firefly, items are retired in their existing bank so that they cannot be selected for future forms. On the other hand, when items are leased from CDT to Firefly, items can maintain their operational status in both item pools.

CHAPTER FIVE: FIELD TEST

The main purposes of field-testing items prior to future operational use are (a) to calculate classical item statistics, (b) to determine whether items meet the criteria with respect to statistical properties for future operational use, and (c) to obtain item parameters for pre-equating purposes, a requirement for any computer adaptive testing (CAT) program. The embedded field test approach allows item parameters to be used for future pre-equating purposes and is based on the assumption that students should be equally motivated to take the operational and embedded field-test items, especially when they are not aware of which item is a field-test item. In each Firefly test, field-test items were interspersed within the operational test.

Each year, field-test items may be embedded into Firefly operational forms. As discussed in Chapter Four, the first operational pilot year was in 2024–2025 which included both operational items and embedded field-test items for Mathematics, Algebra I, ELA, and Literature. Firefly did not include Science or Biology in the 2024–2025 school year. This chapter presents information about the 2024–2025 field test, including results from classical item and differential item functioning (DIF) analyses. In addition, this chapter details the process in which field-test items were identified for further review.

ADMINISTRATION OF FIELD-TEST ITEMS

The embedded field-test item pool in 2024–2025 was designed to deepen the operational CAT item pool for ELA, Mathematics, and high school Algebra I and Literature. Firefly utilizes several different item types: multiple-choice (MC) items for all content areas and levels, evidence-based selected-response (EBSR) items for ELA only, and technology-enhanced (TE) items for all content areas and levels. Table 5–1 provides a summary of the number of field-test items by content area, grade/level and item type. For each mathematics grade/level and Literature, approximately 85 to 110 items were field tested. The items represented all reporting categories. For ELA grades 3–8, items were field tested in both reading and writing domains¹. However, unlike reading, all the writing item positions were field-test positions. As such, the higher number of field-test items for ELA is a result of the large number of field-test writing items.

Table 5–1. Embedded Field-Test Item Pools 2024–2025

Content Area	Grade/Level	MC Count	EBSR Count	TE Count	Total Count
Mathematics	3	81	0	17	98
Mathematics	4	76	0	20	96
Mathematics	5	87	0	15	102
Mathematics	6	86	0	16	102
Mathematics	7	70	0	16	86
Mathematics	8	71	0	15	86
Algebra I	HS	89	0	2	91
ELA	3	178	17	29	224
ELA	4	173	22	25	220
ELA	5	169	28	30	227
ELA	6	177	25	24	226
ELA	7	191	22	17	230
ELA	8	176	16	10	202
Literature	HS	95	11	5	111

¹ As discussed in Chapter Four, CDT has two separate tests for reading and writing; although the CDT writing items had pre-equated and banked parameters, the parameters were on a separate scale from CDT reading items. Thus, all CDT writing items needed to be field tested alongside Firefly reading items to ensure that item parameters were estimated on the same baseline scale, developed following the 2024–2025 Firefly administration.

During the 2024–2025 administration year, all Firefly tests included embedded field-test items. The computer adaptive algorithm administered a different number of items based on content area. For mathematics grades 3 through 8 and Algebra I, six field-test items were administered per test. For ELA grades 3 through 8, both reading and writing field-test items were administered. Specifically, eight writing field-test items were administered per test, whereas reading field-test passages may have had three to eight associated field-test items. Literature consisted of five to eight passage-based field-test items with each passage. Table 5–2 summarizes this information. For all tests, students did not know which items were operational and which were field test. Field-test items did not count in calculation of total scores or reporting category subscores.

Table 5–2. Embedded Field Test Design in 2024–2025

Content Area	Grade/Level	Number of Embedded Field-Test Items
Mathematics	3	6
Mathematics	4	6
Mathematics	5	6
Mathematics	6	6
Mathematics	7	6
Mathematics	8	6
Algebra I	HS	6
ELA	3	1 reading passage with 3–8 items 8 writing items
ELA	4	1 reading passage with 3–8 items 8 writing items
ELA	5	1 reading passage with 3–8 items 8 writing items
ELA	6	1 reading passage with 3–8 items 8 writing items
ELA	7	1 reading passage with 3–8 items 8 writing items
ELA	8	1 reading passage with 3–8 items 8 writing items
Literature	HS	1 passage with 5–8 items

CLASSICAL ITEM ANALYSIS

Classical item analyses of field-test items are conducted to assess the quality of the field-test items and to identify items for data review. Specifically, item difficulty and item-total correlations (the relationship between answering an item correctly and total test score) are estimated for each item, for each option for multiple-choice (MC) and evidence-based selected-response (EBSR) item, and for each score point for multi-point items. Technology-enhanced (TE) items were also field tested in ELA, and mathematics. TE items may be dichotomously scored, like MC items, or polytomously scored.

ITEM DIFFICULTY

At the most general level, an item’s difficulty is indicated by its mean score in some specified group (e.g., grade/level).

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

In the mean score formula above, the individual item scores (x_i) are summed and then divided by the total number of students (n). For multiple-choice items, student scores are represented by 0s and 1s (0 = wrong, 1 = right). With 0–1 scoring, the equation above also represents the number of students correctly answering the item divided by the total number of students. Therefore, this is also the proportion correct for the item, or the p -value. In theory, p -values can range from 0.00² to 1.00 on the proportion-correct scale. For example, if an item has a p -value of 0.89, it means 89 percent of the students answered the item correctly. Additionally, this value might also suggest that the item was relatively easy and/or that the students who attempted the item were relatively high achievers. In other words, item difficulty and student ability are somewhat confounded.

For multi-point items, mean scores can range from the minimum possible score (usually zero points) to the maximum possible score (e.g., two or three points). Often times, a pseudo p -value is calculated for multi-point TE items by dividing the mean item score by the maximum possible item score. Psuedo p -values were calculated for all multi-point items for ease of comparability.

The minimum and maximum extremes of the difficulty scale are typically not seen in applied practice. However, understanding the extremes helps illustrate that relatively lower values correspond to more difficult items and that relatively higher values correspond to easier items. (As a result, some assert that this index would be more accurately referred to as the item’s easiness.)

ITEM DISCRIMINATION

At the most general level, item discrimination³ indicates an item’s ability to differentiate between high and low achievers. It is expected that students with high ability (i.e., those who perform well on Firefly overall) would be more likely to answer any given Firefly item correctly, while students with low ability (i.e., those who perform poorly on Firefly overall) would be less likely to answer the same item correctly. For the Firefly tests, Pearson’s product-moment correlation coefficient between item scores and test scores (student ability estimate) is used to indicate discrimination. The correlation coefficient can range from -1.0 to +1.0. If this expectation is met (i.e., high-ability students tend to answer the item correctly while low-ability students answer the item incorrectly), the correlation between the item score and the total test score will be both positive and noticeably large in its magnitude (i.e., well above zero), meaning the item is a good discriminator between high- and low-ability students. This should be the case for all Firefly field-test items prior to becoming operational test items.

In summary, the correlation will be positive in value when the mean test score of the students answering the item correctly is higher than the mean test score of the students answering the item incorrectly.⁴ In other words, the relationship between student test performance and item performance is expected to be consistent. However, an interaction can exist between item discrimination and item difficulty. Items answered correctly (or incorrectly) by a large proportion of examinees (i.e., the items have extreme p -values) can have reduced power to discriminate, and thus, can have lower correlations.

CLASSICAL ITEM ANALYSIS RESULTS

Table 5–3 provides the summary statistics for the difficulty and discrimination for the 2024–2025 field-test items with respect to content area, grade, and item type. There is a range of p -values across all content areas and grade levels, where mean p -values were between 0.41 and 0.51 for MC or EBSR items, and between 0.07 and 0.59 for TE items. The mean item-total correlations were between 0.30 and 0.47 for MC or EBSR items, and between 0.31 and 0.41 for TE items (see Table 5–3).

² For MC items with four response options, pure random guessing would lead to an expected p -value of 0.25.

³ As noted earlier, the discrimination index for dichotomous MC items is typically referred to as the point-biserial correlation coefficient.

⁴ It is legitimate to view the point-biserial correlation as a standardized mean difference. A positive value indicates that students who chose that response had a higher mean score than the average student; a negative value indicates that students who chose that response had a lower-than-average mean score.

Table 5–3. Summary Statistics of Difficulty and Discrimination by Content Area, Grade and Item Type

Content Area	Grade/ Level	Item Type	N	Mean <i>p</i> -val.	Min <i>p</i> -val.	Median <i>p</i> -val.	Max <i>p</i> -val.	Mean I-T Corr.	Min I-T Corr.	Median I-T Corr.	Max I-T Corr.
Mathematics	3	MC	81	0.51	0.11	0.52	0.90	0.37	0.12	0.39	0.59
Mathematics	3	TE	17	0.32	0.01	0.29	0.71	0.38	0.16	0.41	0.55
Mathematics	4	MC	76	0.47	0.14	0.46	0.83	0.35	0.11	0.37	0.54
Mathematics	4	TE	20	0.33	0.04	0.22	0.78	0.38	0.05	0.39	0.53
Mathematics	5	MC	87	0.42	0.20	0.41	0.71	0.34	0.00	0.35	0.51
Mathematics	5	TE	15	0.29	0.08	0.29	0.66	0.40	0.15	0.42	0.54
Mathematics	6	MC	86	0.47	0.16	0.44	0.81	0.31	0.03	0.32	0.51
Mathematics	6	TE	16	0.23	0.02	0.19	0.56	0.38	0.13	0.40	0.56
Mathematics	7	MC	70	0.41	0.13	0.41	0.73	0.32	0.08	0.32	0.56
Mathematics	7	TE	16	0.28	0.07	0.23	0.66	0.37	0.08	0.38	0.55
Mathematics	8	MC	71	0.42	0.19	0.39	0.83	0.33	0.13	0.34	0.52
Mathematics	8	TE	15	0.26	0.04	0.30	0.59	0.41	0.12	0.45	0.57
Algebra I	HS	MC	89	0.42	0.20	0.42	0.76	0.30	0.03	0.33	0.53
Algebra I	HS	TE	2	0.07	0.03	0.07	0.11	0.31	0.27	0.31	0.36
ELA	3	MC	178	0.48	0.13	0.47	0.89	0.30	-0.04	0.31	0.53
ELA	3	EBSR	17	0.42	0.33	0.40	0.56	0.44	0.24	0.42	0.63
ELA	3	TE	29	0.51	0.22	0.49	0.79	0.40	0.21	0.41	0.62
ELA	4	MC	173	0.53	0.16	0.53	0.95	0.32	0.02	0.32	0.52
ELA	4	EBSR	22	0.43	0.28	0.43	0.55	0.41	-0.06	0.43	0.58
ELA	4	TE	25	0.46	0.16	0.41	0.77	0.35	0.17	0.35	0.56
ELA	5	MC	169	0.55	0.14	0.55	0.94	0.34	0.01	0.36	0.58
ELA	5	EBSR	28	0.46	0.26	0.45	0.64	0.45	0.23	0.46	0.61
ELA	5	TE	30	0.48	0.20	0.49	0.84	0.37	0.14	0.37	0.56
ELA	6	MC	177	0.54	0.23	0.53	0.93	0.31	0.07	0.32	0.57
ELA	6	EBSR	25	0.49	0.29	0.51	0.67	0.44	0.15	0.45	0.61
ELA	6	TE	24	0.53	0.19	0.51	0.86	0.37	0.14	0.37	0.56
ELA	7	MC	191	0.53	0.18	0.53	0.85	0.34	0.03	0.34	0.57
ELA	7	EBSR	22	0.47	0.27	0.45	0.64	0.44	0.15	0.46	0.63
ELA	7	TE	17	0.59	0.31	0.60	0.79	0.37	0.12	0.37	0.55
ELA	8	MC	176	0.52	0.19	0.53	0.94	0.33	0.04	0.33	0.59
ELA	8	EBSR	16	0.49	0.35	0.51	0.70	0.47	0.33	0.47	0.59
ELA	8	TE	10	0.57	0.39	0.57	0.86	0.36	0.13	0.35	0.53
Literature	HS	MC	95	0.46	0.22	0.46	0.74	0.34	0.09	0.35	0.58
Literature	HS	EBSR	11	0.51	0.41	0.46	0.69	0.47	0.18	0.48	0.61
Literature	HS	TE	5	0.45	0.09	0.46	0.74	0.37	0.09	0.40	0.64

Note. I-T Corr. is the item-test score correlation.

DIFFERENTIAL ITEM FUNCTIONING

Differential item functioning (DIF) occurs when examinees with the same ability level but different group memberships do not have the same probability of answering an item correctly. When the probability differs, it is important for content experts to review the relevant items for any potential *item bias*. It is important to note that, as a statistical concept, DIF is different from item bias. DIF detects a difference in performance after controlling for student ability, whereas bias is a content issue that can arise in situations where something other than the intended construct of measurement affects the probability of a correct response for a particular group. For example, bias is likely present when an item presents negative group stereotypes that draw the attention of the examinee, uses non-construct-relevant language that is more familiar to one subpopulation than to another, or is presented in a non-construct-relevant format that disadvantages certain learning styles. While the source of item bias can be plain to trained judges, DIF may have no clear cause. In such cases, something other than bias, including construct-relevant content, may be explaining the differential performance on the item. Flagging items with DIF provides an opportunity for reviewers to assess and correct potential bias, but DIF does not necessarily mean that bias is present.

LIMITATIONS OF STATISTICAL DETECTION

No statistical procedure should be used as a substitute for rigorous, hands-on reviews by content and bias specialists. The statistical results can help organize the review so the effort is concentrated on the most problematic cases. Further, no items should be automatically rejected simply because a statistical method flagged them or accepted because they were not flagged.

Statistical detection of DIF is also not an exact science. There have been a variety of methods proposed for detecting DIF, but no single statistic can be considered either necessary or sufficient. Different methods are more or less successful but can also detect DIF at different rates. No analysis can guarantee that a test is free of bias, but thoughtful item development and field-test analysis can prevent potentially biased items from unfairly impacting student scores.

A fundamental shortcoming of all statistical methods used in DIF evaluation is that all are intrinsic to the test being evaluated. If a test is unbiased overall but contains one or two DIF items, any method can identify DIF. However, because all current methods use total test performance as the measure on which to control for group abilities, a test with all DIF items will not be able to separate DIF effects from differences in achievement on the test.

MANTEL-HAENSZEL PROCEDURE FOR DIFFERENTIAL ITEM FUNCTIONING

For dichotomously-scored MC and TE items, the Mantel-Haenszel procedure (Mantel & Haenszel, 1959) for detecting DIF is a commonly used technique in educational testing. It does not depend on the application or the fit of any specific measurement model. However, it does have significant philosophical overlap with the Rasch model since it uses a test's total score for the analysis.

The procedure as implemented by DRC contrasts a focal group with a reference group. While it makes no practical difference in the analysis which group is defined as the focal group, the group most apt to be disadvantaged by a biased measurement is typically defined as the focal group. In these analyses, the focal groups were female for gender-based DIF and Black and Hispanic for ethnicity-based DIF; reference groups were male and White, respectively. The Mantel-Haenszel (MH) statistic for each item is computed from a contingency table. It has two groups (focal and reference) and two outcomes (right and wrong). The ability groups are defined using the test score distribution for the total examinee population.

The basic MH statistic is a single degree of freedom chi-square that compares the observed number in each cell to the expected number. The expected counts are computed to ensure that the analysis is not confounded with differences in the achievement level of the two groups.

For polytomously-scored items, a comparable statistic is computed based on the standardized mean difference (SMD) (Dorans, Schmitt, & Bleistein, 1992), which is computed as the differences in mean scores for the focal and reference groups if both groups had the same score distribution.

To support better interpretation of DIF, the items are assigned both a severity code (A, B, or C) based on the magnitude of the MH statistic, as well as a direction (minus or plus) based on the direction of the MH statistic. Items classified as A+ or A- have little or no statistical indication of DIF. Items classified as B+ or B- have some indication of DIF but may be judged to be acceptable for future use. Items classified as C+ or C- have strong evidence of DIF and should be reviewed and possibly rejected from the eligible item pool. The plus sign indicates that the item favors the focal group and a minus sign indicates that the item favors the reference group.

RESULTS AND OBSERVATIONS

DIF analyses were conducted on the 2024–2025 field-test items. The number of items from each content area and grade that were assigned to each severity code is shown in Table 5–4 for MC items, Table 5–5 for EBSR items, and Table 5–6 for TE items. Results from DIF analyses are provided for gender (male/female) and ethnicity (white/black, and white/Hispanic). For MC and TE items, very few items were flagged based on gender, and a slightly higher number of items were flagged based on race/ethnicity. For EBSR, very few items were flagged based on gender or race/ethnicity. The total count may differ from the total number of items shown above if there were insufficient N counts to conduct DIF.

Table 5–4a. DIF Summary for Male/Female – MC Items

Content Area	Grade/Level	A+	A-	B+	B-	C+	C-	Total
Mathematics	3	39	38	1	3	0	0	81
Mathematics	4	37	35	0	3	0	1	76
Mathematics	5	33	53	1	0	0	0	87
Mathematics	6	41	45	0	0	0	0	86
Mathematics	7	41	26	1	1	0	1	70
Mathematics	8	41	26	1	3	0	0	71
Algebra I	HS	49	38	0	1	1	0	89
ELA	3	105	72	0	1	0	0	178
ELA	4	108	59	6	0	0	0	173
ELA	5	105	52	12	0	0	0	169
ELA	6	112	59	6	0	0	0	177
ELA	7	126	59	5	1	0	0	191
ELA	8	118	48	10	0	0	0	176
Literature	HS	61	31	2	1	0	0	95

Table 5–4b. DIF Summary for Ethnicity (White/Black) – MC Items

Content Area	Grade/Level	A+	A-	B+	B-	C+	C-	Total
Mathematics	3	21	53	0	6	0	1	81
Mathematics	4	24	42	1	7	0	2	76
Mathematics	5	19	57	3	6	0	2	87
Mathematics	6	21	59	1	3	0	2	86
Mathematics	7	18	48	1	3	0	0	70
Mathematics	8	29	40	0	2	0	0	71
Algebra I	HS	34	44	2	7	0	2	89
ELA	3	44	108	0	23	0	3	178
ELA	4	31	112	1	24	0	5	173
ELA	5	39	112	0	14	0	1	166
ELA	6	43	110	1	16	0	1	171
ELA	7	45	120	0	16	0	5	186
ELA	8	45	106	0	14	0	3	168
Literature	HS	27	62	1	4	0	1	95

Table 5–4c. DIF Summary for Ethnicity (White/Hispanic) – MC Items

Content Area	Grade/Level	A+	A-	B+	B-	C+	C-	Total
Mathematics	3	29	47	1	4	0	0	81
Mathematics	4	22	44	0	9	1	0	76
Mathematics	5	32	53	0	2	0	0	87
Mathematics	6	26	54	2	3	0	1	86
Mathematics	7	26	43	0	1	0	0	70
Mathematics	8	23	44	0	3	0	1	71
Algebra I	HS	43	42	0	3	0	1	89
ELA	3	45	98	0	27	0	8	178
ELA	4	36	104	1	23	0	9	173
ELA	5	29	123	1	9	0	4	166
ELA	6	45	111	1	13	0	5	175
ELA	7	55	122	1	8	0	0	186
ELA	8	37	114	0	16	0	1	168
Literature	HS	24	63	0	6	0	0	93

Table 5–5a. DIF Summary for Male/Female – EBSR Items

Content Area	Grade/Level	A+	A-	B+	B-	C+	C-	Total
ELA	3	11	6	0	0	0	0	17
ELA	4	12	10	0	0	0	0	22
ELA	5	19	9	0	0	0	0	28
ELA	6	15	10	0	0	0	0	25
ELA	7	13	8	0	1	0	0	22
ELA	8	9	7	0	0	0	0	16
Literature	HS	7	3	1	0	0	0	11

Table 5–5b. DIF Summary for Ethnicity (White/Black) – EBSR Items

Content Area	Grade/Level	A+	A-	B+	B-	C+	C-	Total
ELA	3	3	12	0	1	0	1	17
ELA	4	6	15	0	1	0	0	22
ELA	5	1	19	1	7	0	0	28
ELA	6	3	19	0	3	0	0	25
ELA	7	4	16	0	1	0	0	21
ELA	8	6	10	0	0	0	0	16
Literature	HS	4	7	0	0	0	0	11

Table 5–5c. DIF Summary for Ethnicity (White/Hispanic) – EBSR Items

Content Area	Grade/Level	A+	A-	B+	B-	C+	C-	Total
ELA	3	3	13	0	1	0	0	17
ELA	4	4	17	0	1	0	0	22
ELA	5	7	16	0	5	0	0	28
ELA	6	5	18	0	2	0	0	25
ELA	7	6	15	0	0	0	0	21
ELA	8	10	5	0	1	0	0	16
Literature	HS	4	5	0	0	0	0	9

Table 5–6a. DIF Summary for Male/Female – TE Items

Content Area	Grade/Level	A+	A-	B+	B-	C+	C-	Total
Mathematics	3	8	9	0	0	0	0	17
Mathematics	4	12	7	0	1	0	0	20
Mathematics	5	9	4	0	2	0	0	15
Mathematics	6	9	7	0	0	0	0	16
Mathematics	7	7	8	1	0	0	0	16
Mathematics	8	5	10	0	0	0	0	15
Algebra I	HS	1	1	0	0	0	0	2
ELA	3	17	12	0	0	0	0	29
ELA	4	19	5	0	0	1	0	25
ELA	5	21	9	0	0	0	0	30
ELA	6	13	9	0	1	1	0	24
ELA	7	11	6	0	0	0	0	17
ELA	8	6	3	1	0	0	0	10
Literature	HS	4	1	0	0	0	0	5

Table 5–6b. DIF Summary for Ethnicity (White/Black) – TE Items

Content Area	Grade/Level	A+	A-	B+	B-	C+	C-	Total
Mathematics	3	6	8	0	1	0	1	16
Mathematics	4	3	14	0	3	0	0	20
Mathematics	5	4	9	0	0	0	2	15
Mathematics	6	4	8	0	1	0	3	16
Mathematics	7	5	9	0	1	0	1	16
Mathematics	8	5	7	0	1	0	2	15
Algebra I	HS	0	2	0	0	0	0	2
ELA	3	2	14	1	8	0	4	29
ELA	4	8	12	0	4	0	1	25
ELA	5	6	19	0	5	0	0	30
ELA	6	4	15	0	3	0	1	23
ELA	7	2	10	0	2	0	2	16
ELA	8	3	6	0	0	0	1	10
Literature	HS	1	4	0	0	0	0	5

Table 5–6c. DIF Summary for Ethnicity (White/Hispanic) – TE Items

Content Area	Grade/Level	A+	A-	B+	B-	C+	C-	Total
Mathematics	3	4	9	1	1	0	1	16
Mathematics	4	4	14	0	1	0	1	20
Mathematics	5	5	8	0	1	0	1	15
Mathematics	6	3	10	0	2	0	1	16
Mathematics	7	3	13	0	0	0	0	16
Mathematics	8	4	10	0	0	0	1	15
Algebra I	HS	1	1	0	0	0	0	2
ELA	3	3	15	0	8	0	3	29
ELA	4	4	17	0	2	0	2	25
ELA	5	1	22	0	4	0	3	30
ELA	6	4	16	0	2	0	1	23
ELA	7	4	9	1	0	0	2	16
ELA	8	4	5	0	0	0	1	10
Literature	HS	2	2	0	0	0	0	4

CRITERIA FOR IDENTIFYING ITEMS

As previously discussed, all field-test items were analyzed statistically using conventional item-analysis methods. For MC and EBSR items, classical item statistics included the corrected point-measure correlation for the correct and incorrect responses (distractors), percent correct (p -value), and the percent responding to incorrect responses. For TE items, the statistical indices included the p -value, the total point-measure correlation, the correlation for each score point, and the percent of responses at each score point.

In general, more capable students are expected to respond correctly to easy items and less capable students are expected to respond incorrectly to difficult items. The primary way of detecting such conditions is through the correlation coefficient for items. In each case, the statistic will be positive if the total test mean score is higher for the students who respond correctly to MC items and negative when the reverse is true. If an item does not perform well, the item is flagged for additional review. Given that the 2024–2025 field-test items were previously written and administered operationally, items were flagged for internal review by DRC test development staff according to the following criteria.

The following set of criteria is used to identify items for additional review.

For an MC item to be flagged, the criteria included any of the following:

- Percent correct (p -value) less than 0.2 or greater than 0.9
- Point-measure correlation for the correct response less than 0.2
- Point-measure correlation for any incorrect response greater than 0.0
- Gender or ethnic DIF code of either C- or C+

For an EBSR item to be flagged, the criteria included any of the following:

- p -value less than 0.2 or greater than 0.9
- Point-measure correlation less than 0.2
- Gender or ethnic DIF code of either C- or C+
- Score proportion less than 0.05

For an MS or TE item to be flagged, the criteria included any of the following:

- p -value less than 0.2 or greater than 0.9
- Point-measure correlation less than 0.2
- Score proportion less than 0.05
- Gender or ethnic DIF code of either C- or C+

REVIEW OF ITEMS WITH DATA

In future administrations, committees of Pennsylvania educators will review newly developed, administered and flagged field-test items to determine the appropriateness of the item for future operational administration during a data review committee meeting.

Following the 2024–2025 administration, DRC test development staff reviewed flagged field-test items and determined whether the item should be moved to the operational item pool for the following year. Table 5–7 provides a summary of the items identified. Approximately 2,100 field-test items were administered in the 2024–2025 academic year. About 20% of the items were flagged for review based on item statistics. As expected, a larger number of items were flagged for ELA than mathematics because there were more items field tested for ELA. A total of 22 field-test items (5% of the flagged items) were rejected following internal review.

Table 5–7. Summary of 2024–2025 Internal Item Review

Content Area	Grade/ Level	N MC Flagged	N EBSR Flagged	N TE Flagged	N Total Flagged	N MC Rejected	N EBSR Rejected	N TE Rejected	N Total Rejected
Mathematics	3	18	0	8	26	4	0	0	4
Mathematics	4	18	0	10	28	6	0	0	6
Mathematics	5	15	0	6	21	0	0	0	0
Mathematics	6	18	0	8	26	1	0	0	1
Mathematics	7	20	0	7	27	2	0	0	2
Mathematics	8	18	0	7	25	1	0	0	1
Algebra I	HS	21	0	2	23	0	0	0	0
ELA	3	50	0	0	50	2	0	0	2
ELA	4	36	2	5	43	1	0	0	1
ELA	5	28	0	3	31	0	0	0	0
ELA	6	35	0	4	39	2	0	0	2
ELA	7	27	1	1	29	0	0	0	0
ELA	8	35	1	2	38	2	0	0	2
Literature	HS	26	1	1	28	0	0	1	1
Total	--	365	5	64	434	21	0	1	22

CHAPTER SIX: OPERATIONAL ITEM STATISTICS

This chapter provides an overview of operational item statistics from the first administration in 2024–2025. Recall that the operational administration of 2024–2025 consisted entirely of items previously tested on Pennsylvania’s Classroom Diagnostic Tools. In addition, for ELA grades 3 through 8, the operational administration consisted only of reading, passage-based items. Refer to Chapter Four for details about the operational item pool. In this chapter, results from classical item analyses are provided in addition to summaries of item exposure. A comprehensive description of these statistics is presented in Chapter Five; other statistics such as Rasch item parameters and statistics are discussed in Chapter Seven.

ITEM-LEVEL STATISTICS

As discussed in Chapter Five, classical item statistics based on field-test data provide information on how items perform, how plausible distractors are for multiple-choice items, and how the item differentiates between higher performing and lower performing students. Field-test items are dealt randomly to students, so the classical item statistics are a reliable representation of how a random sample of Firefly testers perform on each item. On a fixed form test, we expect operational items to perform similarly, but not identical, to when they were field tested. However, on a computer adaptive test, like Firefly and CDT, operational items are assigned and administered based on the student ability estimate and the item difficulty parameters. Thus, if the computer adaptive algorithm functions as expected, we assume that the average p -value should be close to 50%. The correlations are calculated based on the relationship between item score and total test score (student ability).

Table 6–1 presents a summary of the operational item pools, in terms of number of items, mean p -value, and mean correlation for each content area and grade/level. For each content area and grade/level, there were 340 to 467 items in each of the operational item pools in 2024–2025. The mean p -values ranged from 0.43 to 0.48 and the mean correlations were greater than 0.30. The mean p -values indicate that the computer adaptive algorithm is performing close to expected with assigning items in that the likelihood that students correctly answer items is approximately 50%. As a primarily passage-based test, ELA grades 3 through 8 and Literature may have slightly different (lower or higher) average p -values due to the adaptive algorithm used to select passages. For example, suppose a student with estimated ability of 0.33 is assigned a passage set with item difficulties of 0.2, 0.55, and -0.32. The student may perform as expected on the first two items (50% likelihood of correctly answering the item); however, on the third item the student has a higher likelihood of answering the third item correctly because the third item is of lower difficulty than their estimated ability. This pattern can work in either direction based on student ability estimate and the item difficulties within a passage. Refer to Chapter Eleven for additional information about the CAT configurations.

Table 6–1. Summary of Operational Item Statistics

Content Area	Grade/Level	N Items	Mean P-Value	Mean Correlation
Mathematics	3	467	0.47	0.34
Mathematics	4	438	0.47	0.33
Mathematics	5	394	0.48	0.35
Mathematics	6	442	0.46	0.33
Mathematics	7	397	0.45	0.33
Mathematics	8	384	0.47	0.34
Algebra I	HS	423	0.45	0.31
ELA	3	344	0.43	0.33
ELA	4	344	0.44	0.34
ELA	5	347	0.44	0.36
ELA	6	358	0.47	0.35
ELA	7	349	0.45	0.35
ELA	8	358	0.47	0.35
Literature	HS	454	0.46	0.34

ITEM EXPOSURE

For computer adaptive testing programs, it is important to monitor item exposure. Firefly, unlike CDT, is an on-grade assessment; therefore, evaluating item-level exposure relative to each content area and grade/level can provide important information about the depth and breadth of the operational item pool. Because Firefly was an operational pilot in 2024–2025, it was expected that item exposure would be higher than more well-established CAT programs. Table 6–2 provides a summary of item-level exposure, by content area, grade/level and item type. Based on the 2024–2025 administration data, operational items were administered between 200 and 9,500 times for mathematics and Algebra I, and between 1 and 12,500 times for ELA and Literature. Given that ELA and Literature are passage-based tests, it is expected that some passages may not have been selected because the large range of item difficulties across the associated items. For example, if a passage has three associated items with difficulties of -1.24, -0.88, and 0.95, then it is unlikely the passage would be selected given the requirement of administering at least 3 items per passage. The average exposure was similar across math and ELA, with mean *N* counts ranging from about 1,700 to 2,900. Moreover, exposure was similar across different item types.

Table 6–2. Summary of Operational Item Exposure

Content Area	Grade/Level	Item Type	Max Points	N Items	Min Exposure	Mean Exposure	Max Exposure
Mathematics	3	MC	1	467	166	2,437.59	10,130
Mathematics	4	MC	1	438	347	2,387.10	8,103
Mathematics	5	MC	1	394	213	2,891.51	9,315
Mathematics	6	MC	1	442	351	2,422.67	7,929
Mathematics	7	MC	1	397	808	2,410.37	6,977
Mathematics	8	MC	1	384	674	2,253.09	7,604
Algebra I	HS	MC	1	423	870	2,291.66	7,158
ELA	3	MC	1	299	1	2,474.61	12,377
ELA	3	ESR	2–3	45	74	1,718.20	6,853
ELA	4	MC	1	299	1	2,385.70	10,931
ELA	4	ESR	2–3	45	4	1,693.18	9,679
ELA	5	MC	1	292	1	2,709.72	12,037
ELA	5	ESR	2–3	55	3	1,526.24	9,362
ELA	6	MC	1	324	1	2,244.56	10,594
ELA	6	ESR	2–3	34	22	2,164.50	9,751
ELA	7	MC	1	312	1	2,054.68	7,616
ELA	7	ESR	2–3	37	22	1,935.65	6,305
ELA	8	MC	1	316	1	2,067.75	8,527
ELA	8	ESR	2–3	42	14	1,704.57	7,029
Literature	HS	MC	1	414	1	1,837.51	13,142
Literature	HS	ESR	2–3	40	102	1,737.63	12,586

It is important to contextualize the number of exposures presented in Table 6–2. Table 6–3 provides the total number of items and the percentage of items for five exposure rate categories: less than 0.1, between 0.1 and 0.2, between 0.2 and 0.3, between 0.3 and 0.4, and greater than 0.4. For example, an exposure rate for an item of [0.0, 0.1] means that less than 10 percent of Firefly testers were administered the item. For all content areas and grade/levels, most of items were administered to 0 to 10% testers; moreover, no more than 2% of items have an exposure rate greater than 40%. One of the main purposes of developing and field-testing new items routinely is to increase the depth and breadth of the item pool and decrease item exposure.

Table 6–3. Operational Item Exposure Rates by Content Area and Grade/Level

Content Area	Grade/Level	N	Exposure Rate [0.0,0.1]	Exposure Rate (0.1,0.2]	Exposure Rate (0.2,0.3]	Exposure Rate (0.3,0.4]	Exposure Rate > 0.4
Mathematics	3	467	83.3%	14.6%	1.9%	0.2%	0.0%
Mathematics	4	438	78.1%	18.9%	3.0%	0.0%	0.0%
Mathematics	5	394	69.5%	26.9%	3.6%	0.0%	0.0%
Mathematics	6	442	77.8%	19.5%	2.7%	0.0%	0.0%
Mathematics	7	397	66.5%	30.2%	3.3%	0.0%	0.0%
Mathematics	8	384	70.1%	26.0%	3.4%	0.5%	0.0%
Algebra I	HS	423	75.9%	20.6%	3.5%	0.0%	0.0%
ELA	3	344	67.2%	23.8%	4.9%	2.0%	2.0%
ELA	4	344	62.5%	28.5%	7.3%	0.3%	1.5%
ELA	5	347	70.3%	19.0%	6.1%	2.6%	2.0%
ELA	6	358	67.6%	23.2%	6.7%	2.5%	0.0%
ELA	7	349	67.6%	17.5%	12.0%	2.9%	0.0%
ELA	8	358	67.3%	20.9%	8.9%	2.8%	0.0%
Literature	HS	454	75.3%	14.3%	8.1%	0.9%	1.3%

Overall exposure rates provide aggregate information for each item pool. Yet, it is also necessary to evaluate the item exposure rates by reporting category as it can provide information on specific reporting categories that may need additional item development. Table 6–4 presents the same information as Table 6–3, disaggregated by reporting category. Across all content areas, grade/level, and reporting categories, exposure rates were relatively low. Over 84% of items were exposed to less than 20% of examinees.

Table 6–4. Operational Item Exposure Rates by Content Area, Grade/Level, and Reporting Category

Content Area	Grade/ Level	Reporting Category	N	Exposure Rate [0.0,0.1]	Exposure Rate (0.1,0.2]	Exposure Rate (0.2,0.3]	Exposure Rate (0.3,0.4]	Exposure Rate > 0.4
Mathematics	3	Numbers and Operations in Base Ten	70	67.1%	27.1%	4.3%	1.4%	0.0%
Mathematics	3	Numbers and Operations – Fractions	78	87.2%	9.0%	3.8%	0.0%	0.0%
Mathematics	3	Operations and Algebraic Thinking	113	79.6%	20.4%	0.0%	0.0%	0.0%
Mathematics	3	Geometry	74	87.8%	8.1%	4.1%	0.0%	0.0%
Mathematics	3	Measurement and Data	132	90.2%	9.8%	0.0%	0.0%	0.0%
Mathematics	4	Numbers and Operations in Base Ten	81	63.0%	32.1%	4.9%	0.0%	0.0%
Mathematics	4	Numbers and Operations – Fractions	99	83.8%	14.1%	2.0%	0.0%	0.0%
Mathematics	4	Operations and Algebraic Thinking	103	73.8%	22.3%	3.9%	0.0%	0.0%
Mathematics	4	Geometry	70	82.9%	12.9%	4.3%	0.0%	0.0%
Mathematics	4	Measurement and Data	85	87.1%	12.9%	0.0%	0.0%	0.0%
Mathematics	5	Numbers and Operations in Base Ten	100	75.0%	24.0%	1.0%	0.0%	0.0%
Mathematics	5	Numbers and Operations – Fractions	84	54.8%	39.3%	6.0%	0.0%	0.0%
Mathematics	5	Operations and Algebraic Thinking	70	75.7%	21.4%	2.9%	0.0%	0.0%
Mathematics	5	Geometry	70	65.7%	30.0%	4.3%	0.0%	0.0%
Mathematics	5	Measurement and Data	70	77.1%	18.6%	4.3%	0.0%	0.0%
Mathematics	6	The Number System	96	87.5%	11.5%	1.0%	0.0%	0.0%
Mathematics	6	Ratios and Proportional Relationships	78	89.7%	6.4%	3.8%	0.0%	0.0%
Mathematics	6	Expressions and Equations	117	75.2%	23.1%	1.7%	0.0%	0.0%
Mathematics	6	Geometry	71	67.6%	26.8%	5.6%	0.0%	0.0%
Mathematics	6	Statistics and Probability	80	67.5%	30.0%	2.5%	0.0%	0.0%
Mathematics	7	The Number System	51	35.3%	54.9%	9.8%	0.0%	0.0%
Mathematics	7	Ratios and Proportional Relationships	91	71.4%	25.3%	3.3%	0.0%	0.0%
Mathematics	7	Expressions and Equations	99	59.6%	38.4%	2.0%	0.0%	0.0%
Mathematics	7	Geometry	82	73.2%	25.6%	1.2%	0.0%	0.0%
Mathematics	7	Statistics and Probability	74	83.8%	13.5%	2.7%	0.0%	0.0%
Mathematics	8	The Number System	57	61.4%	33.3%	5.3%	0.0%	0.0%
Mathematics	8	Expressions and Equations	121	78.5%	19.8%	1.7%	0.0%	0.0%
Mathematics	8	Functions	56	37.5%	48.2%	10.7%	3.6%	0.0%
Mathematics	8	Geometry	79	73.4%	24.1%	2.5%	0.0%	0.0%
Mathematics	8	Statistics and Probability	71	84.5%	15.5%	0.0%	0.0%	0.0%
Algebra I	HS	Operations and Linear Equations & Inequalities	213	78.4%	16.9%	4.7%	0.0%	0.0%

Table 6–4 (continued). Operational Item Exposure Rates by Content Area, Grade/Level, and Reporting Category

Content Area	Grade/ Level	Reporting Category	N	Exposure Rate [0.0,0.1]	Exposure Rate (0.1,0.2]	Exposure Rate (0.2,0.3]	Exposure Rate (0.3,0.4]	Exposure Rate > 0.4
Algebra I	HS	Operations with Real Numbers and Expressions	73	79.5%	16.4%	4.1%	0.0%	0.0%
Algebra I	HS	Linear Equations	70	78.6%	14.3%	7.1%	0.0%	0.0%
Algebra I	HS	Linear Inequalities	70	77.1%	20.0%	2.9%	0.0%	0.0%
Algebra I	HS	Linear Functions & Data Organizations	210	73.3%	24.3%	2.4%	0.0%	0.0%
Algebra I	HS	Functions	70	72.9%	27.1%	0.0%	0.0%	0.0%
Algebra I	HS	Coordinate Geometry	70	75.7%	21.4%	2.9%	0.0%	0.0%
Algebra I	HS	Data Analysis	70	71.4%	24.3%	4.3%	0.0%	0.0%
ELA	3	Literature Text	163	67.5%	21.5%	5.5%	3.1%	2.5%
ELA	3	Informational Text	181	66.9%	26.0%	4.4%	1.1%	1.7%
ELA	4	Literature Text	172	52.9%	40.1%	7.0%	0.0%	0.0%
ELA	4	Informational Text	172	72.1%	16.9%	7.6%	0.6%	2.9%
ELA	5	Literature Text	171	73.1%	14.0%	6.4%	4.1%	2.3%
ELA	5	Informational Text	176	67.6%	23.9%	5.7%	1.1%	1.7%
ELA	6	Literature Text	180	65.6%	27.2%	7.2%	0.0%	0.0%
ELA	6	Informational Text	178	69.7%	19.1%	6.2%	5.1%	0.0%
ELA	7	Literature Text	176	68.8%	15.3%	14.2%	1.7%	0.0%
ELA	7	Informational Text	173	66.5%	19.7%	9.8%	4.0%	0.0%
ELA	8	Literature Text	177	62.7%	25.4%	11.9%	0.0%	0.0%
ELA	8	Informational Text	181	71.8%	16.6%	6.1%	5.5%	0.0%
Literature	HS	Fiction	219	77.6%	12.8%	6.8%	0.0%	2.7%
Literature	HS	Reading for Meaning—Fiction	80	76.3%	13.8%	7.5%	0.0%	2.5%
Literature	HS	Analyzing and Interpreting Literature—Fiction	139	78.4%	12.2%	6.5%	0.0%	2.9%
Literature	HS	Nonfiction	235	73.2%	15.7%	9.4%	1.7%	0.0%
Literature	HS	Reading for Meaning—Nonfiction	86	72.1%	15.1%	11.6%	1.2%	0.0%
Literature	HS	Analyzing and Interpreting Literature—Nonfiction	149	73.8%	16.1%	8.1%	2.0%	0.0%

As previously discussed in Chapter Five, the embedded field test in 2024–2025 as well as continued field testing will bolster the operational item pool and should decrease exposure rates in subsequent administrations.

CHAPTER SEVEN: RASCH ITEM CALIBRATION

The item response theory (IRT) model used for Firefly is based on the work of Georg Rasch. Rasch models have had a long-standing presence in applied testing programs and it has been the methodology used for other Pennsylvania assessment programs, including PSSA, Keystone, and Classroom Diagnostic Tools (CDT). IRT has several advantages over classical test theory and has become the standard procedure for analyzing item response data in large-scale assessments. However, IRT models make several strong assumptions related to dimensionality, local independence, model-data fit, and item parameter invariance. Resulting inferences derived from any application of IRT rests strongly on the degree to which the underlying assumptions are met.

This chapter briefly introduces the Rasch model, reports the results from evaluations of the adequacy of the Rasch assumptions, and summarizes the Rasch item statistics for Firefly tests. Additional Rasch procedures are discussed with respect to equating and scale maintenance in Chapter Eight.

DESCRIPTION OF THE RASCH MODEL

The Rasch partial credit model (RPCM; Wright & Masters, 1982) is used to calibrate Firefly data because both dichotomously scored items (i.e., multiple-choice [MC], technology-enhanced [TE] items) and polytomously scored items (i.e., multi-point Evidence-Based Constructed-Response [EBSR], multiple-select [MS], and TE) are in the operational and field-test item pools. The RPCM extends the Rasch model (Rasch, 1960) for dichotomously scored (0, 1) items so that it accommodates the polytomous item data. Under the RPCM, for a given item i with m_i score categories, the probability of person n scoring x ($x = 0, 1, 2, \dots, m_i$) is given by:

$$P_{ni}(X = x) = \frac{\exp \sum_{j=0}^x (\theta_n - D_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^k (\theta_n - D_{ij})},$$

where θ_n represents a student's proficiency (ability) level, and D_{ij} is the step difficulty of the j^{th} step on item i . For dichotomous MC items, the RPCM reduces to the standard Rasch model and the single step difficulty is referred to as the item's difficulty. The Rasch model predicts the probability of person n getting item i correct as follows:

$$P_{ni}(X = 1) = \frac{\exp(\theta_n - D_{ij})}{1 + \exp(\theta_n - D_{ij})}.$$

The Rasch model places both student ability and item difficulty (estimated in terms of log-odds or logits) on the same continuum. When the model assumptions are met, the Rasch model provides estimates of a person's ability which are independent of the items employed in the assessment, and conversely, estimates item difficulty independently of the sample of examinees.

SOFTWARE AND ESTIMATION ALGORITHM

Item calibration was implemented via WINSTEPS 4.2 computer program (Linacre, 2019), which employs unconditional (UCON), joint-maximum-likelihood estimation (JMLE).

DESCRIPTION OF CALIBRATION DATASET

The datasets used in the calibration included students who completed the test and excluded students who tested off-grade (i.e., 5th grade student who took a 4th grade Firefly test), duplicated students (defined as those who re-tested within a two-week period). The characteristics of students in the calibration dataset are reported in Chapter Fourteen.

CHECKING RASCH ASSUMPTIONS

Since the Rasch model was the basis of all calibration, scoring, and scaling analyses associated with Firefly, the validity of the inferences from these results depends on the degree to which the assumptions of the model were met and how well the model fits the test data. Therefore, it is important to check these assumptions. This section evaluates the dimensionality of the data, local item independence, and item fit. It should be noted that only operational items were analyzed since they are the basis of student scores.

UNIDIMENSIONALITY

Rasch models assume that one dominant dimension determines the difference among students' performances. Both Principal Components Analysis (PCA) and Confirmatory Factor Analysis (CFA) can be used to assess the unidimensionality assumption. The purpose of these analyses is to verify that the internal structure supports one dominant component or factor. If any other dimensions are found, the unidimensionality assumption would be violated.

In computer adaptive testing, there are many limitations to the direct application of PCA and CFA. First, these methods rely on a matrix of all item responses for each student. For CAT programs, this would require a matrix of N students by M items for each content and grade/level, of which most data would be missing as students only respond to 36 to 42 items in a pool of 400 to 600 items. The matrix for a fixed form test is much more meaningful because each student answers every operational item on the test. Second, for well-targeted CAT programs, items are selected and administered that have a 50% likelihood of a correct response—thus responses are randomly distributed for dichotomously scored item types and the item covariance is low. Historically, some CAT programs conduct factor analysis or PCA based on fixed forms administered during pilot testing. However, Firefly used a CAT since the inception and launch of the program. Thus, alternative approaches for examining dimensionality need to be considered for Firefly.

Wang et al. (2012) proposed and conducted an alternate approach for dimensionality analysis for a CAT program. The approach utilized the cluster or reporting category scores (on the reporting scale metric) to conduct Confirmatory Factor Analysis (CFA). Results across multiple state programs using a benchmark test supported the methodology.

Using reporting category subscores (on the scale score metric), both PCA and CFA were conducted for Firefly. The purpose of the PCA was to identify the proportion of the variance explained by the first dimension. Table 7–1 shows the PCA results for each Firefly content area and grade/level. The results include the eigenvalues and the percentage of variance explained for each component. As Table 7–1 shows, between 57% and 65% of mathematics scores and between 78% and 88% of ELA scores are explained by the first dimension. Furthermore, no secondary dimensions had eigenvalues greater than 1. Meaning the results from the PCA suggest that there is one clear dominant dimension for all Mathematics, ELA and Literature tests.

Table 7–1. Principal Component Analysis – Eigenvalues and Variance Explained by Component Extracted, Content, and Grade/Level

Content Area	Grade/Level	Component	Eigenvalue	Percent of Variance Explained
Mathematics	3	1	3.3	65.2%
Mathematics	3	2	0.5	11.0%
Mathematics	3	3	0.5	10.4%
Mathematics	3	4	0.4	7.2%
Mathematics	3	5	0.3	6.2%
Mathematics	4	1	3.1	61.9%
Mathematics	4	2	0.7	13.0%
Mathematics	4	3	0.5	9.5%
Mathematics	4	4	0.4	8.7%
Mathematics	4	5	0.3	6.9%
Mathematics	5	1	3.2	63.6%
Mathematics	5	2	0.6	11.1%
Mathematics	5	3	0.5	9.9%
Mathematics	5	4	0.4	8.3%
Mathematics	5	5	0.4	7.1%
Mathematics	6	1	2.9	58.5%
Mathematics	6	2	0.6	12.7%
Mathematics	6	3	0.6	11.5%
Mathematics	6	4	0.5	9.8%
Mathematics	6	5	0.4	7.5%
Mathematics	7	1	3.0	60.7%
Mathematics	7	2	0.6	12.2%
Mathematics	7	3	0.6	11.2%
Mathematics	7	4	0.4	8.3%
Mathematics	7	5	0.4	7.6%
Mathematics	8	1	3.0	61.0%
Mathematics	8	2	0.6	11.4%
Mathematics	8	3	0.5	10.8%
Mathematics	8	4	0.5	10.0%
Mathematics	8	5	0.3	6.9%
Algebra I	HS	1	3.4	56.6%
Algebra I	HS	2	0.6	10.0%
Algebra I	HS	3	0.6	9.3%

Table 7–1 (continued). Principal Component Analysis – Eigenvalues and Variance Explained by Component Extracted, Content, and Grade/Level

Content Area	Grade/Level	Component	Eigenvalue	Percent of Variance Explained
Algebra I	HS	4	0.5	8.7%
Algebra I	HS	5	0.5	8.1%
Algebra I	HS	6	0.4	7.3%
ELA	3	1	1.7	86.1%
ELA	3	2	0.3	13.9%
ELA	4	1	1.8	87.7%
ELA	4	2	0.2	12.3%
ELA	5	1	1.8	88.0%
ELA	5	2	0.2	12.0%
ELA	6	1	1.8	88.0%
ELA	6	2	0.2	12.0%
ELA	7	1	1.8	88.0%
ELA	7	2	0.2	12.0%
ELA	8	1	1.8	88.0%
ELA	8	2	0.2	12.0%
Literature	HS	1	3.1	77.5%
Literature	HS	2	0.4	11.2%
Literature	HS	3	0.3	8.6%
Literature	HS	4	0.1	2.7%

In addition, CFA was conducted to examine the fit of the unidimensional model to the 2024–2025 Firefly data. Due to the reporting category structure of ELA grades 3–8, chi-square and several other indices could not be reported based on the 2024–2025 administration data. This will be re-examined in subsequent years, when there will be 3 reporting categories for ELA grades 3–8. Table 7–2 presents the number of testers, chi-square statistics, the associated p-values, the standardized root mean square residual (SRMR), as well as goodness of fit indices. The goodness of fit indices include the comparative fit index (CFI), the Tucker-Lewis fit index (TLFI), and the standardized root mean square error of approximation (RMSEA). Although chi-square statistics and p-values are reported in Table 7–2, there are limitations to interpreting these statistics. Specifically, chi-square statistics are often significant for large enough samples. As such, it is not surprising to see significant findings based on the chi-square index. In contrast, Hu and Bentler (1999) recommended a criteria greater than 0.95 for CFI, GFI, and NFI as well as a criterion less than 0.06 for RMSEA and a criterion less than 0.08 for SRMR. The Tucker-Lewis fit index (TLFI) can be interpreted similarly to other fit indices. According to this, all math grades and Algebra I meet the criterion. However, Literature does not meet the criteria for RMSEA but does meet the criteria for CFI and SRMR.

Table 7–2. Confirmatory Factor Analysis Results by Content Area and Grade/Level

Content Area	Grade/ Level	N	Chi-Sq	PVal	SRMR	CFI	TFI	RMSEA
Mathematics	3	31621	235.70	0.00	1.00	0.99	0.04	0.01
Mathematics	4	29045	283.91	0.00	1.00	0.99	0.04	0.01
Mathematics	5	31646	66.18	0.00	1.00	1.00	0.02	0.01
Mathematics	6	29745	148.73	0.00	1.00	0.99	0.03	0.01
Mathematics	7	26581	133.50	0.00	1.00	0.99	0.03	0.01
Mathematics	8	24033	85.96	0.00	1.00	1.00	0.03	0.01
Algebra I	HS	26927	37.39	0.00	1.00	1.00	0.01	0.00
ELA	3	26345	NA	NA	NA	NA	0.00	0.00
ELA	4	25691	NA	NA	NA	NA	0.00	0.00
ELA	5	28078	NA	NA	NA	NA	0.00	0.00
ELA	6	26547	NA	NA	NA	NA	0.00	0.00
ELA	7	24015	NA	NA	NA	NA	0.00	0.00
ELA	8	23941	NA	NA	NA	NA	0.00	0.00
Literature	HS	23820	2068.11	0.00	0.97	0.91	0.21	0.04

Note. ELA grades 3–8 had only 2 operational reporting categories in 2024–2025. Results are limited.

ITEM INDEPENDENCE

Local independence (LI) is a fundamental assumption of IRT. No relationship should exist between examinees' responses to different items after accounting for the abilities measured by a test. In formal statistical terms, a test X that is comprised of items $X_1, X_2, \dots, X_n$ is locally independent with respect to the latent variable θ if, for all $x = (x_1, x_2, \dots, x_n)$ and θ ,

$$P(\mathbf{X} = \mathbf{x} | \theta) = \prod_{i=1}^I P(X_i = x_i | \theta).$$

This formula essentially states that the probability of any pattern of responses across all items ($\mathbf{x}$), after conditioning on the abilities (θ) measured by the test, should be equal to the product of the conditional probabilities across each item (cf. the multiplication rule for independent events where the joint probabilities are equal to the product of the associated marginal probabilities).

The equation above shows the condition after satisfying the strong form of local independence. A weak form of local independence (LI) was proposed by McDonald (1979). The distinction is important as many indicators of local dependency are framed by LI. The requirement would be for the conditional covariances of all pairs of item responses, conditioned on the abilities, to be equal to zero. When this assumption is met, the joint probability of responses to an item pair, conditioned on abilities, is the product of the probabilities of responses to these two items, as shown below. (This is a weaker form because higher-order dependencies among items are allowed.) Based on the LI, the following expression can be derived:

$$P(X_i = x_i, X_j = x_j | \theta) = P(X_i = x_i | \theta) P(X_j = x_j | \theta).$$

Marais and Andrich (2008) pointed out that local item dependence in the Rasch model can occur in two ways that some may not distinguish. The first way occurs when the assumption of unidimensionality is violated. Here, other nuisance dimensions besides a dominant dimension determine student performance (this can be called “trait dependence”). The second violation occurs when responses to an item depend on responses to another. This is a violation of statistical independence and can be called response dependence. Many people treat the assumptions of unidimensionality and local independence as one phenomenon and believe that once unidimensionality holds, that local independence also holds. By distinguishing the two sources of local dependence, one can see that while local independence can be related to unidimensionality, the two are different assumptions and therefore, require different tests.

Residual item correlations estimated in WINSTEPS for each item pair were used to assess the local dependence among the Firefly items. In general, these residuals are computed as follows. First, expected item performance based on the Rasch model is determined using ability and item parameter estimates. Next, deviations (residuals) between the examinees’ expected and observed performance is determined for each item. Finally, for each item pair, a correlation between the respective deviations is computed.

Three types of residual correlations are available in WINSTEPS: raw, standardized, and logit. It should be noted that the raw score residual correlation essentially corresponds to Yen’s Q_3 index, a popular LI statistic. The expected value for the Q_3 statistic is approximately $-1/(k-1)$ when no local dependence exists, where k is test length (Yen, 1993). Thus, the expected Q_3 values should be approximately -0.03 for Firefly since most Firefly tests have close to 40 operational items. Index values that are greater than 0.20 indicate a degree of local dependence that probably should be examined by test developers (Chen & Thissen, 1997).

Since the three residual correlations are very similar, the standardized residual correlations from WINSTEPS are reported. Table 7–3 shows the summary statistics—mean, SD, minimum, maximum, and several percentiles (P_{10} , P_{25} , P_{50} , P_{75} , P_{90}) — for all the residual correlations for each test. The total number of item pairs (N) and the number of pairs with residual correlations greater than 0.20 are also reported in this table. The mean residual correlations were close to Q_3 index. Three item pairs in mathematics and 114 item pairs in ELA showed residual correlations greater than 0.2. For the three flagged item pairs in mathematics, the items tested the same eligible content standard. For ELA and Literature, the flagged item pairs shared a passage. It can be expected that items which share a passage may have larger absolute stronger correlations than stand-alone items or items that do not have shared passages. Additional details about the items flagged based on residual correlations is presented in Appendix B.

Table 7–3. Summary of Item Residual Correlations by Content Area and Grade/Level

Content Area	Grade/Level	N	Mean	SD	Min	P ₁₀	P ₂₅	P ₅₀	P ₇₅	P ₉₀	Max	> 1.20I
Mathematics	3	108811	0.00	0.01	-0.10	-0.01	-0.01	0.00	0.00	0.01	0.13	0
Mathematics	4	95703	0.00	0.01	-0.08	-0.01	-0.01	0.00	0.00	0.01	0.16	0
Mathematics	5	77421	0.00	0.01	-0.07	-0.01	-0.01	0.00	0.00	0.01	0.12	0
Mathematics	6	97461	0.00	0.01	-0.13	-0.01	-0.01	0.00	0.00	0.01	0.20	1
Mathematics	7	78606	0.00	0.01	-0.09	-0.01	-0.01	0.00	0.00	0.01	0.12	0
Mathematics	8	73536	0.00	0.01	-0.08	-0.01	-0.01	0.00	0.00	0.01	0.14	0
Algebra I	HS	89253	0.00	0.01	-0.08	-0.01	-0.01	0.00	0.00	0.01	0.24	2
ELA	3	58996	0.00	0.01	-0.91	-0.01	-0.01	0.00	0.00	0.01	1.00	12
ELA	4	58996	0.00	0.01	-1.00	-0.01	-0.01	0.00	0.00	0.00	1.00	9
ELA	5	60031	0.00	0.01	-0.74	-0.01	-0.01	0.00	0.00	0.00	1.00	19
ELA	6	63903	0.00	0.01	-0.55	-0.01	0.00	0.00	0.00	0.00	0.51	21
ELA	7	60726	0.00	0.01	-0.30	-0.01	-0.01	0.00	0.00	0.01	0.35	11
ELA	8	63903	0.00	0.01	-0.56	-0.01	0.00	0.00	0.00	0.00	1.00	18
Literature	HS	102831	0.00	0.01	-0.97	-0.01	0.00	0.00	0.00	0.01	1.00	24

ITEM FIT

Additional evidence of validity related to the internal test structure is obtained through an ongoing evaluation of item fit, person fit, and test summary statistics. The item fit of the Rasch Model is routinely evaluated within field testing as well as following each operational administration year, as a part of scale maintenance (see Chapter Eight). WINSTEPS provides two item fit statistics (infit and outfit) for evaluating the degree to which the Rasch model predicts the observed item responses. Each fit statistic can be expressed as a mean square (MnSq) statistic or on a standardized metric (Zstd with mean = 0 and variance = 1). MnSq values are more oriented toward practical significance, while Zstd values are more oriented toward statistical significance. Though both are informative, the Zstd values are often too sensitive for larger sample sizes. In this situation it is recommended that the Zstd values be ignored if the MnSq values are acceptable (Linacre, 2014).

Both infit and outfit MnSq are the average of standardized residual variance (the difference between the observed score and the Rasch estimated score divided by the square root of the Rasch model variance). The difference is that the outfit statistic gives all examinees equal weight in computing the fit and tends to be affected more by unexpected responses far from the person, item, or rating scale category measure (i.e., it is more sensitive to outlying, off-target, low-information responses). The infit statistic is weighted by the examinee locations relative to item difficulty and tends to be affected more by unexpected responses close to the person, item, or rating scale category measure (i.e., informative, on-target responses). Some feel that extreme infit values are a greater threat to the measurement process than extreme outfit since most tests intend to measure the on-target population rather than extreme outliers.

The expected MnSq value is 1.0 and can range from 0 to infinity. Large deviations from the expected value can be interpreted as noise or lack of fit between the items and the model. Values lower than the expected value can be interpreted as item redundancy or overfitting items (too predictable, too much redundancy), and values greater than the expected value indicate underfitting items (too unpredictable, too much noise). Rules of thumb regarding “practically significant” MnSq values vary. More conservative users might prefer items with MnSq values that range from 0.8 to 1.2. Others believe reasonable test results can be achieved with values from 0.5 to 1.5. The results shown in this section highlight values outside of a range of 0.7 to 1.3 given their practical importance.

Table 7–4 presents the summary statistics of infit and outfit mean square statistics for Firefly, including the mean, SD, and minimum and maximum values. The percentage of items with infit and outfit within the desired range of [0.7, 1.3] is also reported. The mean values for both infit and outfit statistics were close to 1.00 across all content areas and grades/levels. Over 98% of the mathematics items had infit values falling in the range of [0.7, 1.3], suggesting reasonable model infit. Whereas approximately 90% of ELA and Literature items had infit values in the desired range. Values above 1.3 can sometimes suggest higher than normal guessing or careless mistake patterns on items, and values below 0.7 can suggest an item over fits. It should be noted that some items with low *N* counts showed very low infit mean-square values.

Table 7–4. Summary of Item Infit and Outfit Mean Square Statistics by Content Area and Grade/Level

Content Area	Grade/Level	Infit Mean	Infit SD	Infit Min	Infit Max	Infit [0.7,1.3] (%)	Outfit Mean	Outfit SD	Outfit Min	Outfit Max	Outfit [0.7,1.3] (%)
Mathematics	3	1.00	0.10	0.84	1.69	98.5%	1.05	0.36	0.79	6.76	93.8%
Mathematics	4	1.00	0.09	0.83	1.54	98.9%	1.04	0.18	0.78	2.87	95.0%
Mathematics	5	1.00	0.08	0.86	1.34	99.7%	1.03	0.19	0.81	3.48	95.4%
Mathematics	6	0.99	0.07	0.86	1.38	99.3%	1.00	0.10	0.84	1.64	98.0%
Mathematics	7	1.00	0.09	0.84	1.86	99.2%	1.03	0.18	0.80	3.21	94.7%
Mathematics	8	0.99	0.08	0.83	1.53	99.7%	1.01	0.15	0.78	2.83	96.9%
Algebra I	HS	0.99	0.06	0.85	1.24	100.0%	1.00	0.08	0.83	1.38	99.5%
ELA	3	1.03	0.30	0.14	4.83	91.0%	1.05	0.50	0.14	8.88	86.6%
ELA	4	1.08	0.55	0.18	9.90	89.2%	1.10	0.58	0.18	9.90	87.8%
ELA	5	1.03	0.22	0.47	2.73	91.6%	1.04	0.25	0.47	2.82	88.5%
ELA	6	1.04	0.28	0.09	3.66	89.7%	1.06	0.38	0.09	4.48	88.0%
ELA	7	1.06	0.28	0.04	3.08	89.1%	1.07	0.36	0.04	3.55	86.8%
ELA	8	1.05	0.27	0.14	3.09	89.9%	1.06	0.38	0.14	5.52	86.9%
Literature	HS	1.05	0.30	0.11	4.86	91.4%	1.06	0.36	0.11	5.67	86.3%

RASCH ITEM STATISTICS

As noted earlier, the Rasch model expresses item difficulty (and student ability) in units referred to as logits, rather than on the percent-correct metric. In the simplest case, a logit is a transformed p-value with the average p-value becoming a logit of zero. In this form, logits resemble z-scores or standard normal deviates; a very difficult item might have a logit of +4.0 and a very easy item might have a logit of –4.0. However, they have no formal relationship to the normal distribution.

The logit metric has several mathematical and statistical advantages over p-values. Logits have an interval scale, meaning that two items with logits of 0.0 and +1.0, respectively, are the same distance apart as two items with logits of +3.0 and +4.0. Logits are not dependent on the ability level of the students. For example, a test form can have a mean logit of zero, whether the average item p-value for the student sample is 0.8 or 0.3.

The standard, unanchored Rasch calibration procedure arbitrarily sets the mean difficulty of the items in any calibration at zero. As previously discussed in Chapter Four, the first operational Firefly item pools were leased from the Pennsylvania’s Classroom Diagnostic Tools (CDT) item bank to report operational student scores in the first year. For Firefly, the existing operational item parameters from CDT were used to establish the baseline scale in 2024–2025. As such, even in the first year of Firefly, item difficulties (Rasch item parameters) were not centered at zero.

One of the advantages of relying on and utilizing the CDT item bank to construct the foundation of the Firefly item pools was the ability to use the previously developed vertical scale. The vertical scale is evidenced by Table 7–5 by the increasing means across grades/levels within each content area. Specifically, the increasing trend in mean item difficulties for grades 3 through 8 and high school exemplifies the vertical scale.

Table 7–5. Summary of Operational Rasch Item Difficulties by Content Area and Grade/Level

Content Area	Grade/Level	N	Mean	SD	Min	Max
Mathematics	3	467	-1.87	1.19	-4.57	2.16
Mathematics	4	438	-1.35	1.14	-5.24	2.06
Mathematics	5	394	-0.88	0.98	-3.83	2.14
Mathematics	6	442	-0.26	1.07	-2.90	2.32
Mathematics	7	397	0.11	0.82	-2.69	2.39
Mathematics	8	384	0.34	0.78	-1.70	2.48
Algebra I	HS	423	0.68	0.79	-1.62	2.96
ELA	3	357	-0.69	0.73	-3.76	1.68
ELA	4	357	-0.22	0.79	-3.24	2.18
ELA	5	355	0.03	0.73	-3.20	1.91
ELA	6	364	0.17	0.76	-2.49	2.23
ELA	7	354	0.33	0.76	-2.02	3.26
ELA	8	360	0.51	0.74	-1.51	2.40
Literature	HS	492	0.77	0.73	-2.03	2.86

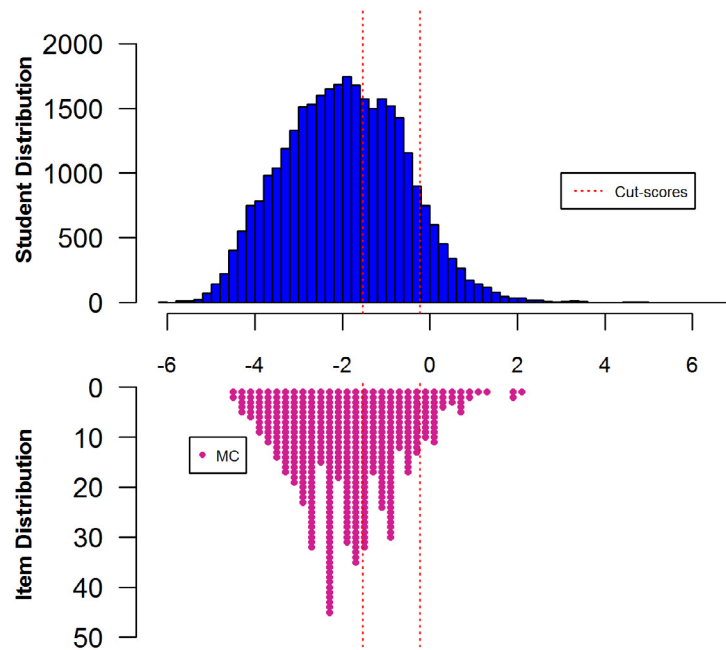
Note. The baseline vertical scale from CDT was used as the foundation for Firefly, item difficulties are not centered at zero.

ITEM DIFFICULTY/STUDENT ABILITY WRIGHT MAPS

The distributions of the Rasch item logits (item difficulty estimates) are shown on the item difficulty/student ability maps presented in Figure 7–1. In each item-student map, the top bar graph displays the student distribution on the logit scale, and the bottom displays markers of item difficulty parameter estimates. MC items are represented by a circle (all content areas and levels), EBSR and MS items are represented by a square. As noted earlier, the Rasch model enables placement of both items and students on the same scale. Consequently, one can easily visualize information regarding the relationship between the distributions of item difficulty and student ability. The vertical red lines show the cut-points for each performance level (Support Needed, Near Target, and Prepared). On the top plot, the logit represents lower student abilities (negative values) to higher abilities (positive values), whereas on the bottom plot the logit represents easier items (negative values) to harder items (positive values). For a CAT program, the goal is to have a deep and wide item bank that reflects similar properties to that of the population.

Figure 7–1. Wright Maps

Mathematics Grade 3



Mathematics Grade 4

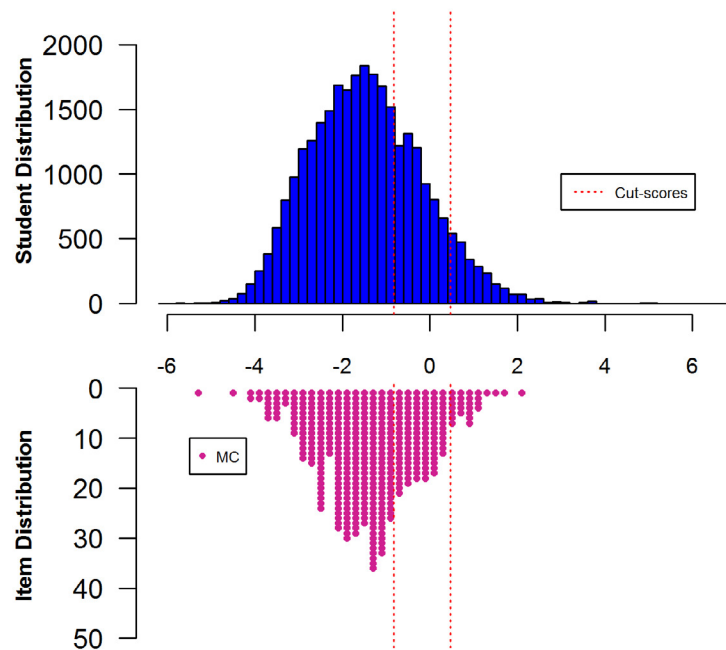
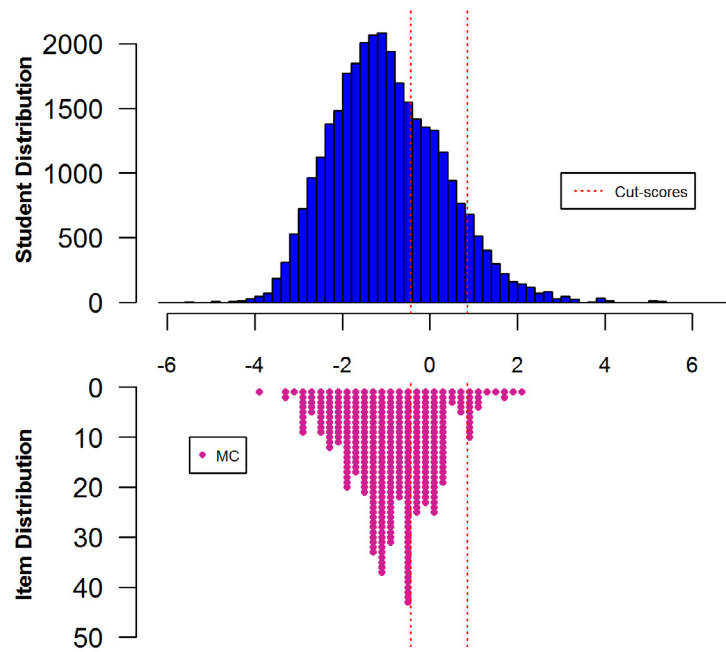


Figure 7–1 (continued). Wright Maps

Mathematics Grade 5



Mathematics Grade 6

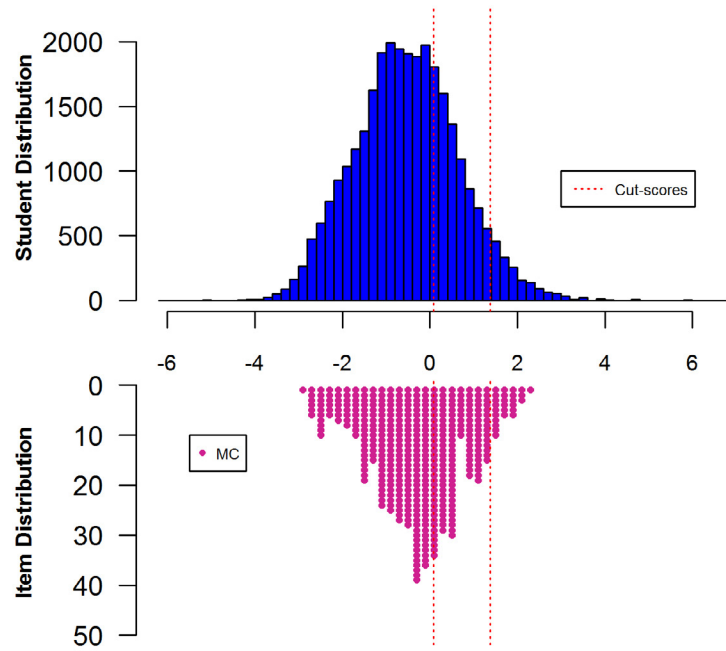
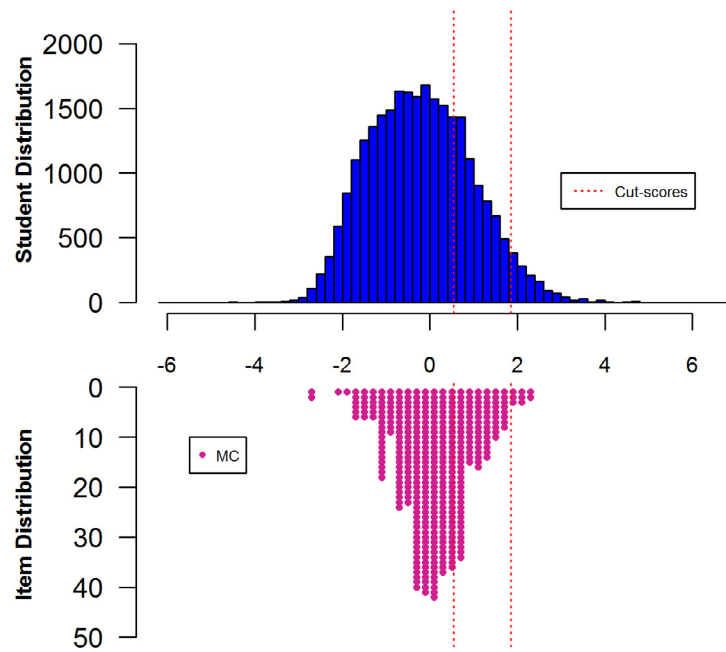


Figure 7–1 (continued). Wright Maps

Mathematics Grade 7



Mathematics Grade 8

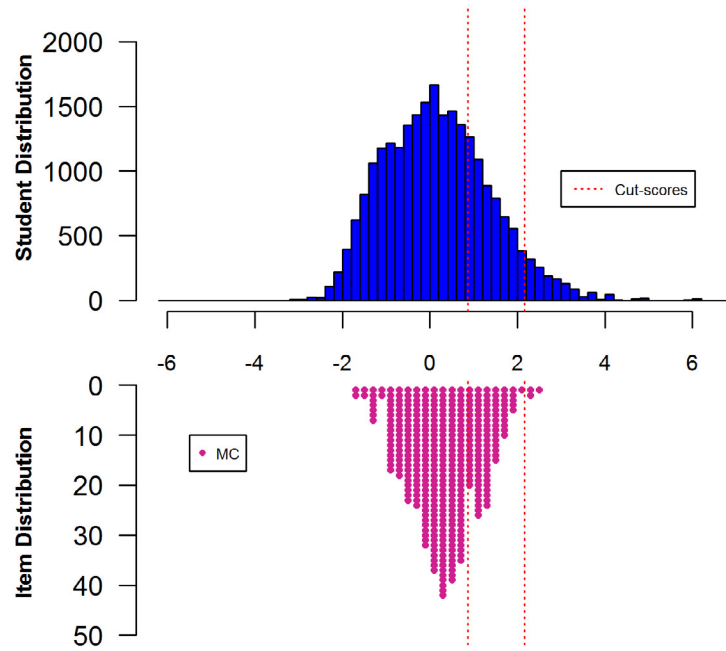
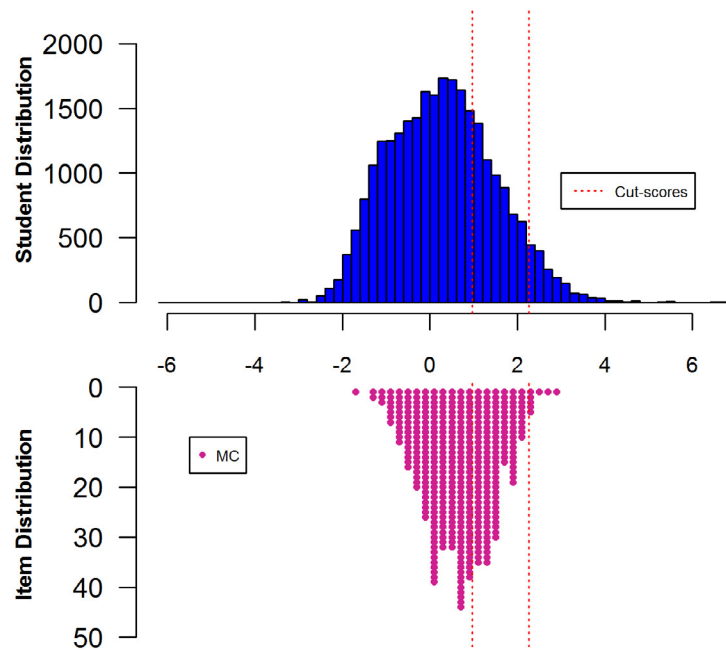


Figure 7–1 (continued). Wright Maps

Algebra I



ELA Grade 3

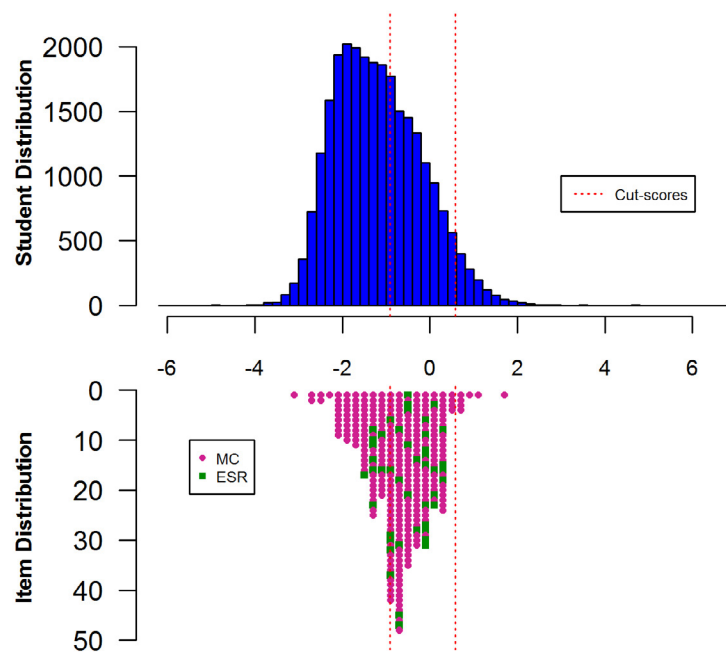
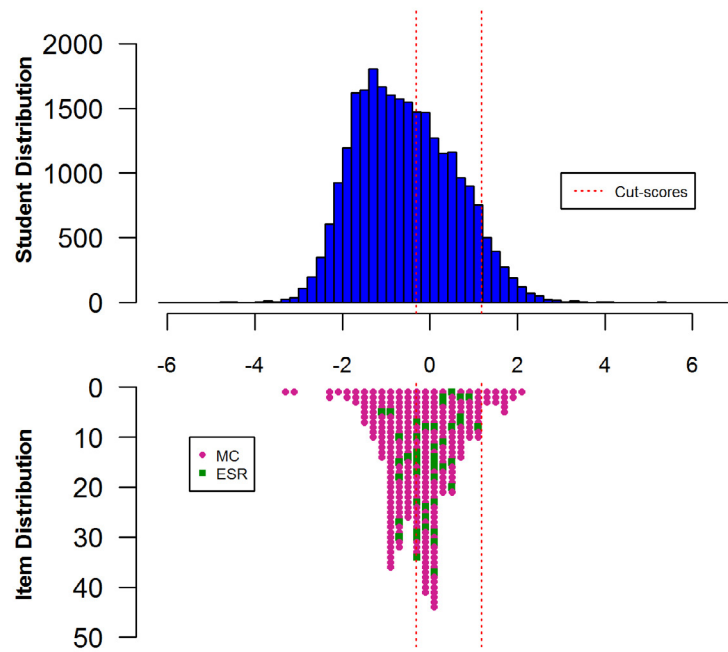


Figure 7–1 (continued). Wright Maps

ELA Grade 4



ELA Grade 5

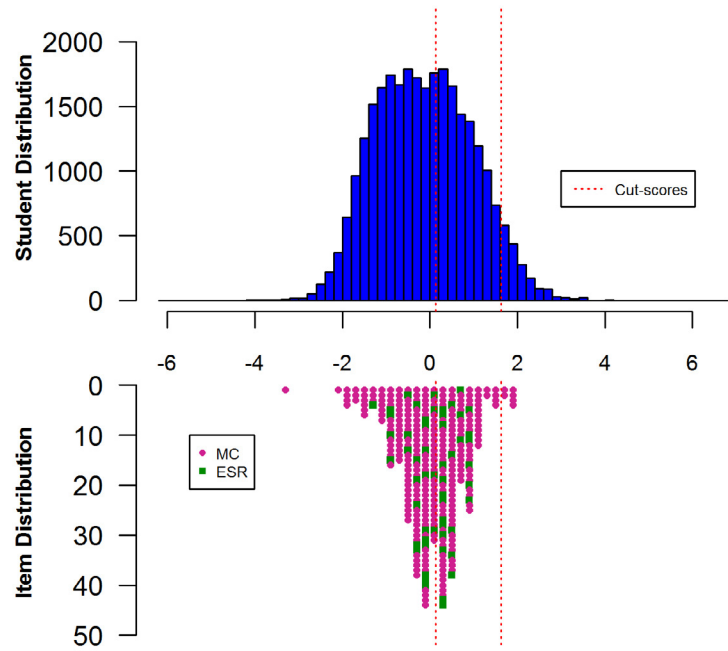
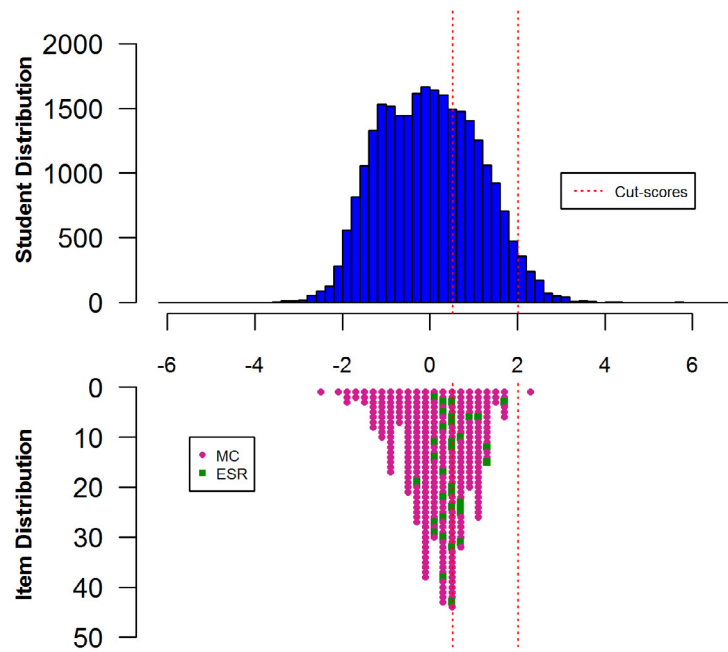


Figure 7–1 (continued). Wright Maps

ELA Grade 6



ELA Grade 7

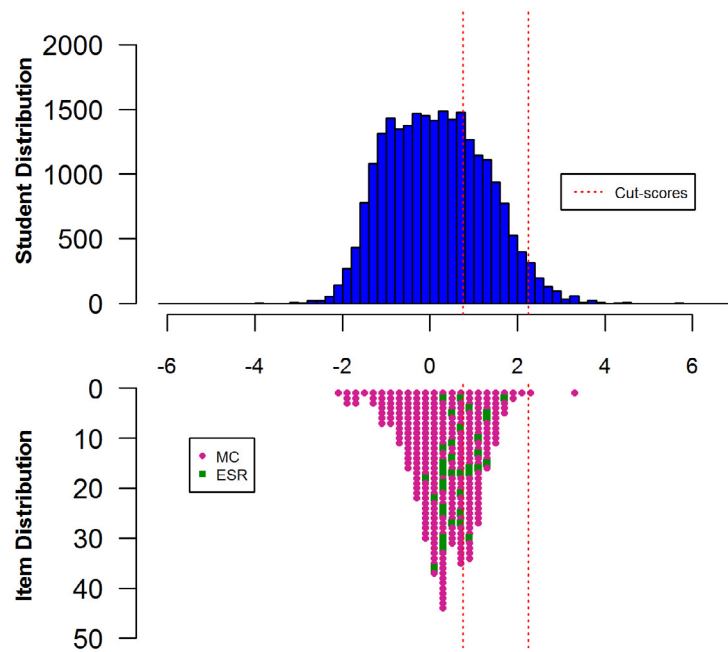
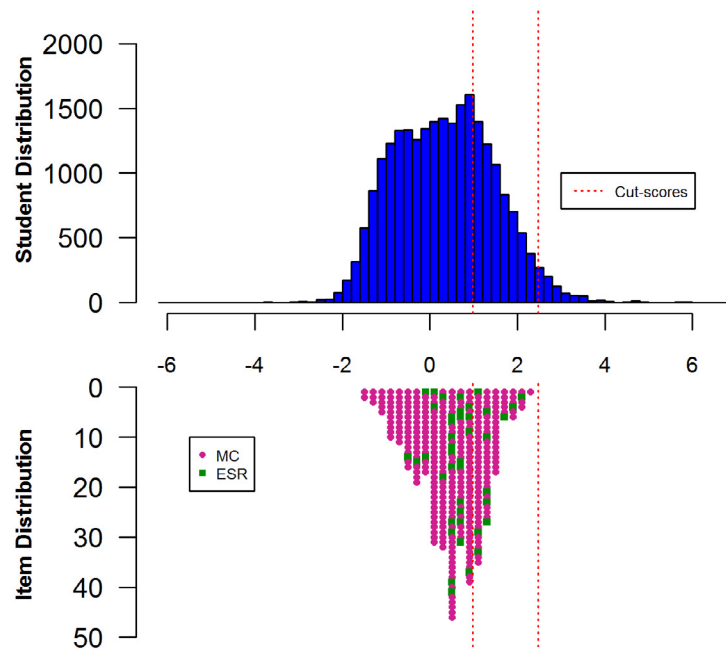
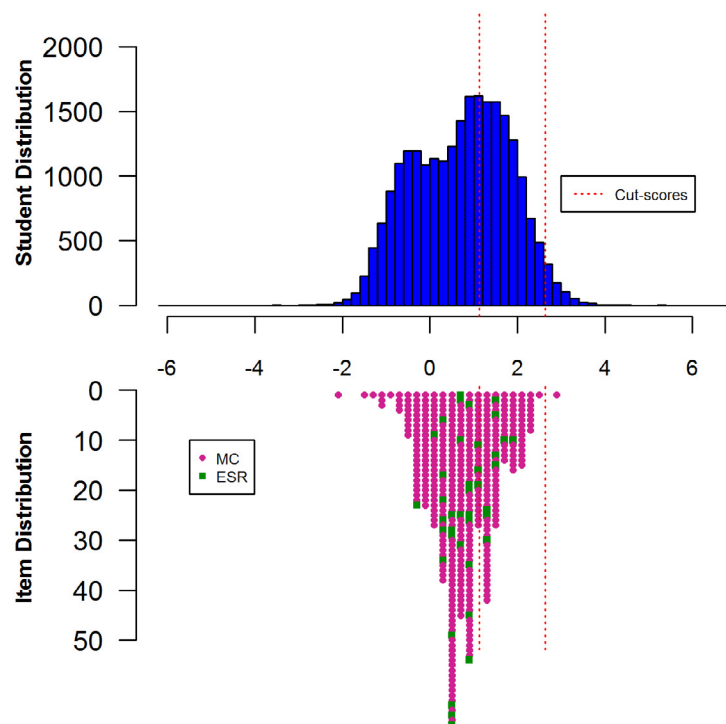


Figure 7–1 (continued). Wright Maps
ELA Grade 8



Literature



CHAPTER EIGHT: PRE-EQUATING AND SCALE MAINTENANCE

Equating is a statistical process that is used to adjust scores on test forms so that scores on the forms can be used interchangeably, even though the test forms consist of different items (Kolen & Brennan, 2004). To make meaningful comparisons of test scores across administrations, various equating models, designs, and procedures have been developed. For example, the testing model can be classified as either classical test theory based equating model or modern test theory (e.g., Rasch model or item response theory) based equating model. There are many different equating designs discussed in the literature, most commonly used are the randomly equivalent groups design and common-item non-equivalent groups design. In terms of when the equating is conducted in the assessment cycle, the model can be classified as pre-equating or post-equating.

Computer adaptive tests (CAT) like Firefly and Classroom Diagnostic Tools (CDT) administer a unique set of items to each student, depending on their estimated ability (theta). Thus, all items must have existing item parameters for the computer adaptive test algorithm to select appropriate, well-targeted items. The requirements of adaptive testing and immediate score reporting necessitate pre-equating within an item response theory framework; as such, this chapter discusses the application of pre-equating to Firefly.

PRE-EQUATING

Like other Pennsylvania assessment programs, Firefly uses the Rasch model to guide test design, calibration, scaling, and equating. The key element of pre-equating under the Rasch model is to place or estimate the item parameters on the same scale prior to each administration. Once the field-test item difficulties are placed or estimated on the base scale or common metric, one should not expect the Rasch item difficulties to change once they are administered on an operational test, except within a reasonable range of measurement error, providing the Rasch model fits the data. Based on this theoretical advantage of using Rasch models, field-test data is used to estimate item parameters¹. This statistical procedure is referred to as pre-equating. For fixed form, pre-equated tests, like PSSA and Keystone, raw scores are associated with Rasch ability estimates and then converted to scale scores. For CATs, like Firefly and CDT, Rasch ability estimates are calculated iteratively after each item or each passage set, as students progress through the test. At the end of each CAT, the final Rasch ability estimate is transformed into a scale score for comparisons within and across years. All Rasch ability estimates are on the same scale (and can be used interchangeably) because they are based on the same pool of items, with parameters on the same scale.

SCALE MAINTENANCE AND EQUATING

Following each year long administration, DRC conducts scale maintenance and equating. Scale maintenance is a key step in any pre-equated program, especially CAT programs, as it involves evaluating the item pool using recent administration data and addressing any issues that may impact data-model fit. There are two approaches to scale maintenance used for Firefly. From a test development perspective, the scale maintenance process allows for a systematic review of potential items that may have been identified during the administration year, reviewing their most recent statistics, and making decisions on whether to keep the items in the item pool for the following administration or retiring them from use. If items are identified for retirement, they are removed from the dataset prior to any calibration. From a psychometric perspective, the scale maintenance process involves identifying misfitting items or items with unstable item parameters and subsequently updating and re-estimating item parameters. Consistent with procedures for other Pennsylvania assessments, the criteria of displacement greater than $|0.5|^2$ or infit mean-square values greater than 1.3 is used to identify items³.

From a Rasch perspective, equating is the process in which newly tested item parameters are estimated on the baseline scale. To do so, a concurrent calibration approach is used in which all operational item parameters are anchored to either the existing pre-equated parameters or parameters updated during scale maintenance and field-test item parameters are estimated in a partially-anchored design. The resultant item parameters are on the same, baseline scale and are banked for use in the following administration.

¹ Scale maintenance addresses issues in which the item parameters estimated from a field-test event are unstable or misfitting the data. See following section for additional information on scale maintenance.

² In the first operational year, a criteria of 0.4 was used for displacement. For future administrations, the criteria for displacement will be 0.5, consistent with other PA assessment programs.

³ This is often an iterative process until all items meet the specified criteria.

The following scale maintenance and calibration procedures were used for Firefly in 2024–2025. While the baseline scale is vertical, meaning scores can be compared across-grades, Firefly is an on-grade assessment. Thus, calibrations are conducted separately for each content area and grade/level. Calibration was conducted using WINSTEPS (Linacre, 2019).

1. Remove items identified for content-related reasons from the dataset as these items are considered ineligible for future administration.
2. In the first operational year, identify operational items with CDT parameters that showed instability (displacement $> |0.5|$) and had sufficient N counts⁴. Drop identified items from the operational anchor set and run partially anchored calibration. The result is updated parameters for items that showed unstable item parameters based on CDT data.
3. Evaluate the stability of all other operational item parameters. Based on current administration Firefly data, conduct scale maintenance.
 - a. Identify and drop items from the anchor set if they meet either of the following criteria:
 - i. Displacement $> |0.4|$, or
 - ii. Mean-square item infit > 1.3
 - b. Use item parameters from step 2 and run partially-anchored calibration, allowing re-estimation of parameters for items identified.
4. Estimate field-test item parameters in a partially-anchored calibration using item parameters estimated in step 3.
 - a. The result is used for item banking and future pre-equating.

RESULTS AND OBSERVATIONS

Scale maintenance is an important aspect of any pre-equated program. It was especially important for Firefly because the CDT items were administered under different conditions than Firefly. As previously discussed, there are several differences between CDT and Firefly, and each program has a distinct purpose. In comparison to CDT, Firefly was designed as a shorter test that reflects the summative (PSSA and Keystone) test blueprints. In addition, Firefly is considered a benchmark or interim assessment to be administered up to three times annually; whereas CDT is a diagnostic tool for educators to identify students' relative strengths and weaknesses to support classroom instruction designed to be administered up to 5 times. In addition, CDT can be administered in a secure or non-secure browser. Thus, evaluating the scale is an important component for Firefly, especially in the first year.

In 2024–2025, items were identified in two ways: (1) for displacement based on previous CDT data, or (2) for misfit or displacement based on current Firefly data. Table 8–1 shows the number of total operational items evaluated during this process (N OP Items), the number of items flagged for each of the three criteria, the total number of items flagged, and the percentage of items flagged. It is important to note that items that did not have any exposure were excluded from this table given they could not be evaluated. Moreover, if items did not have sufficient exposure necessary for parameter re-estimation (i.e., less than 100 administrations), item parameters were not updated.

As shown in Table 8–1, between 23% and 53% of mathematics items and 21% to 29% of ELA and Literature items were flagged during scale maintenance. Most items were flagged due to displacement in either CDT or Firefly. A smaller number of items were flagged for infit issues, of which most were for ELA and Literature. It was not surprising that a larger percentage of math grade 3 items were flagged for re-estimation given that the CDT math grade 3 test allows calculator for all items. This is one of the main differences between Firefly and CDT in that Firefly does not allow calculator use for math grade 3 for consistency with PSSA math grade 3 testing procedures.

⁴ For re-estimation of operational item parameters, items were identified based on results from concurrent, cross-level calibration results from CDT. Item-level N counts greater than 100 were considered sufficient.

Table 8–1. Results from Operational Item Calibration and Scale Maintenance

Subject	Grade/Level	N Items	N Displace Items (CDT)	N Displace Items (Firefly)	N Infit Items (Firefly)	Total Flagged	Flagged %
Mathematics	3	467	160	94	6	247	52.9
Mathematics	4	438	127	93	2	208	47.5
Mathematics	5	394	55	96	1	147	37.3
Mathematics	6	442	88	104	1	183	41.4
Mathematics	7	397	62	79	0	131	33.0
Mathematics	8	384	47	84	1	129	33.6
Algebra I	HS	423	37	62	0	96	22.7
ELA	3	344	28	86	26	101	29.4
ELA	4	344	30	74	32	89	25.9
ELA	5	347	12	81	23	87	25.1
ELA	6	358	22	71	25	76	21.2
ELA	7	349	28	82	31	89	25.5
ELA	8	358	20	77	19	86	24.0
Literature	HS	454	32	102	27	109	24.0

Items flagged due to statistical reasons in the scale maintenance process were removed from the anchor set during calibration to allow item parameters to be re-estimated based on the most recent Firefly data. The Rasch statistics (difficulty, mean-square infit, and displacement value) for the administration and the re-estimated Rasch statistics are presented in Appendix C.

CHAPTER NINE: SCALING

Scaling is used to transform test score values onto a scale that can be easily understood and interpreted by users. Raw scores cannot be used to compare students' achievement on Firefly because they depend on the difficulty of the selected and administered test items by the computer adaptive test (CAT) algorithm. Given the adaptive nature of Firefly, each student receives test items targeted at his or her level of achievement. Therefore, two students may have taken very different sets of items in terms of difficulty but have the same raw score. This makes the use of raw scores for comparison across students or administrations, or to a specific standard (cut score) meaningless. Rasch ability estimates in the logit metric rely on the difficulty of the items administered. Although Rasch ability estimates can be used to make comparisons across students, scale scores are used to report results because they are easier to understand and interpret than logits.

The underlying baseline scale defines item difficulties in the logit metric on the same scale. To develop and launch Firefly in a relatively short timeframe, the underlying baseline scale for Firefly was inherited from the baseline scale of the Classroom Diagnostic Tools (CDT). In doing so, the Firefly baseline scale relied heavily on the previously built CDT infrastructure, including the previously field-tested items and estimated Rasch item difficulties, underlying baseline scale, as well as the cut scores established for CDT. Upon utilizing an existing and well-constructed underlying baseline scale from CDT, Firefly scale scores are derived using a two-step process. First, there is a nonlinear transformation that converts an individual raw score on a unique set of items to Rasch ability (in logits). Second, a linear transformation is used to convert the Rasch abilities (in logits) to scale scores.

RAW SCORES TO RASCH ABILITY ESTIMATES

For each Firefly test, the calibrated item difficulties associated with the unique set of items administered were used to obtain Rasch person ability estimates and asymptotic standard errors of measurement for the overall test, as well as each reporting category. As previously discussed, calibrated item difficulties were based on CDT field tests and vertical linking. Refer to Chapters Eight and Nine in the Classroom Diagnostic Tools Technical Report for additional information on how the CDT vertical scale was established.

Raw scores (total and by reporting category) on the unique set of items that make up an individual Firefly test were mapped to Rasch ability estimates using unconditional, joint-maximum likelihood estimation. In the case of zero or perfect raw scores, a fractional raw score (a value less than one) was added to zero scores and subtracted from perfect scores to determine the corresponding logit values for these extreme scores. The Rasch ability estimates were then transformed to scale scores as discussed in the next section.

RASCH ABILITY ESTIMATES TO SCALE SCORES

Generally, scale scores are preferred over Rasch ability estimates for reporting purposes because they are easier to understand and interpret. Specifically, Rasch ability estimates in the logit metric are on a scale that includes negative and decimal values (i.e., -4.5 to +4.1). By transforming the Rasch ability estimates to scale scores, all reported values become positive integers, which are more easily interpretable by teachers, parents, and students. Since Rasch ability estimates are comparative, the transformed scale scores can be interpreted in the same way across administrations. Given that Firefly can be administered as an interim assessment, up to three times annually, it is critical that scale scores can be interpreted in the same way within and across years.

Scale scores are usually obtained using a linear transformation of Rasch ability estimates. Before the linear equation was established for each content area, a few points were considered:

- Avoid scales that might be confused with scores for other assessments; for example:
 - Scale scores ranging from 0 to 100 (because this might be confused with percent correct scores or percentile ranks).
 - Scale scores ranging from 200 to 800 (because this might be confused with SAT scores).
 - Scale scores with similar ranges as the ones for the Pennsylvania System of School Assessment (PSSA), Keystone Exams, or CDT.
- Avoid scales similar to raw scores.

- Avoid scales that might suggest the scores are more precise than they actually are (in other words, suggesting more precision than can be supported by the test scores).
- Avoid scales with negative numbers and decimals.

In terms of industry standard practice, a common perspective is that scale scores should facilitate score interpretation while at the same time minimize misinterpretation and unwarranted inferences. Often this is done by attaching meaning to the scale scores¹ (Peterson, Kolen, and Hoover, 1989). The incorporation of content meaning is one way to facilitate score interpretation. For other Pennsylvania assessment programs (PSSA, Keystone, and CDT), content meaning is defined by performance-level cut scores on the scaled score metric. Chapter Ten discusses cut scores used to establish performance levels for Firefly.

Like the CDT, the Firefly reporting scale must be sufficiently large to cover the entire vertical scale. As a result, an initial scale score range of 180 to 980 was established for each content area (mathematics and ELA). To aide in interpretability, the lowest and highest observable scale scores (LOSS and HOSS, respectively) were set by grade or level so that the range for each grade or level represents 680 scale score points.

The following processes of linearly transforming Rasch ability estimates to scale scores and assigning a lowest or highest obtainable scale score were conducted for both the total score as well as scores for each reporting category.

LINEAR TRANSFORMATION FORMULAS

The scale scores for the CDT for each content area are obtained through a linear transformation of the Rasch ability estimates ($\hat{\beta}$). Specifically,

$$SS = m\hat{\beta} + b,$$

where m is the slope and b is the intercept. The linear transformation for each Firefly content area was derived by setting the slope for each content area, which influences the variability of the scale scores². It is important that the slopes are sufficiently large to cover the full range of the vertical scale. The scaling equations produce scale score distributions with standard deviations of approximately 150 scale score points and cover logit ranges of approximately -6.0 to 5.5. The final slopes and intercepts for deriving scale scores for Firefly are provided in Table 9–1.

Table 9–1. Scaling Constants by Content Area

Content Area	Slope	Intercept
Mathematics	100	550
ELA	70	600

ROUNDING

The linearly transformed scale scores are rounded to the nearest integer value for reporting purposes. Values greater than or equal to 0.50 are rounded up. Values less than 0.50 are rounded down.

¹ Not everyone agrees with this sentiment. Some have argued the opposite point—that is, any attempt to add meaning to test scores actually predisposes the scores to be misinterpreted (see Angoff, 1984).

² Using cut scores for two-point anchoring was also explored, but was ultimately not supported.

LOWEST AND HIGHEST OBTAINABLE SCALE SCORES

Each Firefly test has a lowest obtainable scale score (LOSS) and a highest obtainable scale score (HOSS), which varies by grade or level to best represent the on-grade nature of Firefly. Any derived scale score less than LOSS is truncated to this minimum value and any scale score greater than the HOSS is assigned the maximum value. In addition, any raw scores of 0 are assigned the LOSS and any perfect scores are assigned the HOSS, regardless of the items administered. The selection of the LOSS and HOSS is mainly based on two considerations:

- 1. Extremely low or high scale scores may have an impact on the average of the scale scores if Firefly data is summarized at school, district, or state level.
- 2. Score truncation makes sense from a score precision perspective given measurement errors at the extremes are large.

Table 9–2 shows the lowest and highest obtainable scale scores for each content area and grade/level. As shown in the table, the scale score range for each Firefly test is 680 points. Moreover, the LOSS and HOSS are adjusted incrementally by 20 scale score points for each adjacent grade/level.

Table 9–2. Lowest and Highest Obtainable Scale Scores

Content Area	Grade/Level	LOSS	HOSS
Mathematics	3	180	860
Mathematics	4	200	880
Mathematics	5	220	900
Mathematics	6	240	920
Mathematics	7	260	940
Mathematics	8	280	960
Algebra I	HS	300	980
ELA	3	180	860
ELA	4	200	880
ELA	5	220	900
ELA	6	240	920
ELA	7	260	940
ELA	8	280	960
Literature	HS	300	980

CHAPTER TEN: BENCHMARKS AND CUT SCORES

As previously discussed, the development and launch of Firefly relied heavily on the established infrastructure for the Classroom Diagnostics Tools (CDT). Because the underlying baseline scale for CDT was used for Firefly, we can make similar inferences about the student scores in the Rasch ability logit metric. The cut scores used for Firefly were the same as the cut scores established through CDT's Benchmarking Activities discussed in Chapters Ten and Nineteen of the CDT Technical Report. In this chapter, the CDT benchmarking activities are summarized, and the transformation and definition of the Firefly cut scores are discussed.

CDT BENCHMARKING ACTIVITIES

CDT scores are placed along a continuum from “Areas of Need” to “Strengths to Build On,” which are represented with colors red, green, and blue, respectively. “Areas of Need” are depicted in the red range, while “Strengths to Build On” are depicted in the green and blue ranges. The center of the green range is the point that separates students into two categories: solidly ready for the next grade or course and not solidly ready for the next grade or course. In each content area, the center of the green range for grades 5 and above was established by panels of Pennsylvania educators during benchmarking activities¹. Each benchmarking event followed the initial stand-alone field-test event for that content area. As operational data became available, cut scores were routinely evaluated and potentially revised. Chapter Nineteen in the CDT Technical Report includes details including the benchmark cut scores that were implemented for Firefly.

The CDT Benchmarking activity utilized the Randomly Ordered Item Booklet (ROIB) Angoff (Yes/No) method to set benchmark cut points. Multiple panels of educators worked in groups to establish cut points within each content area and grade level (5 to 8), as well as for content area courses Algebra I, Literature, and Biology. After a training session describing the process and definition of roles, a discussion was held in which panelists were asked to define “solidly ready for the next grade or course.” Then, panelists were asked to review approximately 40 randomly ordered test questions and make individual yes/no judgments as to whether a “solidly ready” student would be successful in answering each question. The judgments were made over two iterations or rounds with a sequence of Round 1 judgments, show and verification of Round 1 results, group discussion, and Round 2 judgments.

After cut points were set for each grade and course within a content area, the vertical articulation of cut points across grades and courses was reviewed. Given that each content area is vertically scaled, it was expected that cut points would increase as grade increased. For example, the grade 8 cut point would be higher than the grade 7 cut point on the vertical scale. In some cases, post-smoothing was required to ensure increasing cut points across grades/courses and smooth transitions.

CDT BENCHMARKS

EXTRAPOLATION OF BENCHMARK CUTS FOR GRADES 2 THROUGH 4

The introduction of CDT tests for students in grades 3 through 5 in spring 2014 required benchmark cuts for grades 2 through 4. For each content area, the benchmark cuts in place for the 2013–2014 school year in grades 5 and above were used to extrapolate cuts in grades 2 through 4. See Chapter Nineteen of the 2013–2014 CDT technical report for details.

REVISION OF BENCHMARK CUTS BASED ON CHANGES TO PSSA

In spring 2015, changes were made to PSSA test designs and cut points in mathematics and ELA. Considering these changes, CDT benchmark cuts were evaluated and analyzed using matched data sets between operational CDT with PSSA and Keystone. The new PSSA cut points approved in July 2015 represented higher, more rigorous content standards. Therefore, CDT benchmark cuts in mathematics, reading, and writing were increased to allow for more consistent performance-level classifications between CDT and PSSA. See Chapter Nineteen of the 2015–2016 CDT technical report for details.

Table 10–1 shows the benchmark cut points used for CDT student reporting during the 2023–2024 and 2024–2025 school year in the logit metric for each content area. The cut points presented are in the logit metric and reflect all revisions that occurred prior to 2025. The green cut point and blue cut point were established using 1 standard

¹ The center of the green range for grades 2 through 4 was extrapolated from grades 5 and above prior to the launch of each CDT for students in grades 3 through 5 in spring of 2014. See Chapter Nineteen of the CDT technical report for details.

error of measurement above and below the Center of Green cut point, respectively. The CDT cut points in the logit metric were directly applied to Firefly.

Table 10–1. CDT Benchmark Cut Points

Content Area	Grade/Level	Green Logit Cut Score	Blue Logit Cut Score
Mathematics	3	-1.533	-0.233
Mathematics	4	-0.830	0.470
Mathematics	5	-0.442	0.858
Mathematics	6	0.076	1.376
Mathematics	7	0.545	1.845
Mathematics	8	0.863	2.163
Algebra I	HS	0.963	2.263
Reading	3	-0.917	0.583
Reading	4	-0.321	1.179
Reading	5	0.129	1.629
Reading	6	0.515	2.015
Reading	7	0.749	2.249
Reading	8	0.975	2.475
Literature	HS	1.132	2.632
Science	5	-1.101	0.199
Science	8	-0.011	1.289
Biology	HS	0.462	1.762

Note. Only a subset of cut points are presented as they relate directly to Firefly content areas and grades/levels.

FIREFLY CUT SCORES

As previously discussed, the CDT logit cut scores could be directly applied to Firefly because they share the same underlying baseline scale. CDT Reading cut scores were used for ELA and CDT mathematics cut scores were directly related to Firefly mathematics in 2024–2025. As presented in Table 10–1, the green logit cut score was implemented as the Near Target cut score in Firefly and the blue logit cut score was implemented as the Prepared cut score in Firefly. While CDT performance levels were defined as colors, three performance levels were defined as Support Needed, Near Target, and Prepared. Students earning a Support Needed performance level need additional foundational skills and knowledge to support student learning. Students earning a Near Target performance level are close to on-grade expectations. Students earning a Prepared performance level exceed on-grade expectations.

The CDT cut points presented in Table 10–1 were transformed to the Firefly scale using the scale transformation constants presented in Chapter Nine. Table 10–2 shows the Firefly scale score ranges for each performance level, content area and grade/level.

Table 10–2. Firefly Score Ranges

Content Area	Grade/Level	Support Needed	Near Target	Prepared
Mathematics	3	180–492	493–583	584–860
Mathematics	4	200–541	542–632	633–880
Mathematics	5	220–568	569–659	660–900
Mathematics	6	240–604	605–696	697–920
Mathematics	7	260–638	639–729	730–940
Mathematics	8	280–659	660–751	752–960
Algebra I	HS	300–667	668–758	759–980
ELA	3	180–458	459–608	609–860
ELA	4	200–517	518–668	669–880
ELA	5	220–562	563–712	713–900
ELA	6	240–600	601–751	752–920
ELA	7	260–624	625–775	776–940
ELA	8	280–647	648–797	798–960
Literature	HS	300–662	663–813	814–980

VALIDATION OF FIREFLY CUT SCORES

Following the 2024–2025 Firefly administration and 2025 PSSA and Spring Keystone administrations, the validity of the Firefly cut scores was analyzed. Firefly records were matched with PSSA and Keystone data based on Pennsylvania student identification number relative to each content area. The process and inclusion criteria for matching is discussed in Chapter Fourteen.

DRC analyzed the degree to which PSSA and Keystone performance-level classifications were comparable to the Firefly performance-level classifications in a similar manner to the analyses conducted previously for CDT. It is expected that students should earn similar, comparable performance-level classifications when Firefly is taken within a short timeframe with PSSA and Keystone. PSSA and Keystone performance levels were considered comparable if students who earned Below Basic or Basic on PSSA and Keystone also earned Support Needed on Firefly, students who earned Proficient on PSSA and Keystone and also earned Near Target on Firefly, and students who earned Advanced on PSSA and Keystone earned Prepared on Firefly. The percent of students with comparable performance levels between the two assessments (Firefly and PSSA, and Firefly and Keystone) was calculated based on the spring testing window (March 1 through the start of summative tests). The spring Firefly data was used because tests that measure similar constructs have the strongest relationship when administered in a similar timeframe.

Table 10–3 shows the N count of students that were matched between PSSA and Keystone with Firefly, and percent of students earning comparable performance levels between Firefly and PSSA and Keystone. As shown in the table, 72% to 81% of Firefly math and Algebra I testers earned similar performance levels and 62% to 76% of Firefly ELA and high school Literature testers earned similar performance levels. As Table 10–3 shows, higher similarity indices were observed for mathematics than ELA. This may be somewhat expected given a high proportion of the ELA and Literature tests are comprised of writing prompts (text-dependent analysis items for ELA and constructed-response items for Literature), which may be harder to measure on a test comprised predominantly of multiple-choice, or selected-response items.

Table 10–3. Performance-Level Comparability between Firefly and PSSA and between Firefly and Keystone

Program	Content	Grade/Level	Spring N	Similarity Index (%)
PSSA	Mathematics	3	11,168	74.19
PSSA	Mathematics	4	9,864	73.91
PSSA	Mathematics	5	10,617	76.53
PSSA	Mathematics	6	10,543	76.71
PSSA	Mathematics	7	9,059	79.28
PSSA	Mathematics	8	7,842	81.11
Keystone	Algebra I	HS	8,731	71.90
PSSA	ELA	3	9,316	76.22
PSSA	ELA	4	9,003	74.44
PSSA	ELA	5	9,531	75.55
PSSA	ELA	6	8,207	68.22
PSSA	ELA	7	6,673	68.38
PSSA	ELA	8	6,968	62.36
Keystone	Literature	HS	5,587	66.14

Notes. 1. The similarity index represents the percentage of students earning similar/comparable performance levels on Firefly and PSSA or Keystone based on the Firefly spring testing window.
2. Firefly Science was not administered in 2024–2025.

CHAPTER ELEVEN: OPERATIONAL TEST DESIGN AND COMPUTER ADAPTIVE TEST CONFIGURATIONS

Firefly was designed and developed as a benchmark assessment that is predictive of performance on end-of-year PSSA and end-of-course Keystone statewide summative tests. The Firefly test designs and blueprints were developed to reflect that of the statewide summative assessments. In 2024, Firefly was launched in the following content areas and grades/levels: Mathematics grades 3 through 8, ELA grades 3 through 8, Algebra I, and high school Literature¹. Firefly is available to all districts and schools in Pennsylvania and is voluntary. Firefly is administered in DRC’s INSIGHT testing platform using a computer adaptive test (CAT) model.

This chapter details the operational Firefly test design and configuration of the CAT algorithm. Test design elements include the number of reporting categories, the number of operational items to administer per reporting category, the number of multi-point items administered and the number of embedded field-test items. CAT algorithm elements include entry point, item selection criteria, test navigation, and termination.

OPERATIONAL TEST DESIGN

TEST BLUEPRINTS

The test design and blueprints were developed to reflect that of the statewide summative PSSA and Keystone Exams, with similar reporting category structures. Firefly item pools include single-point multiple-choice (MC) and technology-enhanced (TE) items as well as multi-point TE and evidence-based selected-response (EBSR) items. MC and TE items are included on all Firefly item pools, whereas EBSR items are only included on ELA and Literature item pools². As discussed in Chapter Four, all Firefly items are aligned to the Pennsylvania Core Standards. Tables 11–1 through 11–9 display the number of operational items administered for each reporting category and the total number of operational items. It is important to note that the number of operational points may exceed the number of operational items for ELA and Literature because of the inclusion of multi-point items.

Table 11–1. Mathematics Grade 3 Test Blueprint

Reporting Category	Number of Operational Items
Numbers and Operations in Base Ten	6
Numbers and Operations – Fractions	6
Operations and Algebraic Thinking	9
Geometry	6
Measurement and Data	9
Total	36

Table 11–2. Mathematics Grade 4 Test Blueprint

Reporting Category	Number of Operational Items
Numbers and Operations in Base Ten	7
Numbers and Operations – Fractions	8
Operations and Algebraic Thinking	9
Geometry	6
Measurement and Data	6
Total	36

¹ In the 2025–26 administration year, science grades 5 and 8 as well as high school Biology will be incorporated into Firefly as well.
² In 2024–25, all TE items in mathematics and ELA were field-test items.

Table 11–3. Mathematics Grade 5 Test Blueprint

Reporting Category	Number of Operational Items
Numbers and Operations in Base Ten	8
Numbers and Operations – Fractions	10
Operations and Algebraic Thinking	6
Geometry	6
Measurement and Data	6
Total	36

Table 11–4. Mathematics Grade 6 Test Blueprint

Reporting Category	Number of Operational Items
The Number System	7
Ratios and Proportional Relationships	6
Expressions and Equations	10
Geometry	6
Statistics and Probability	7
Total	36

Table 11–5. Mathematics Grade 7 Test Blueprint

Reporting Category	Number of Operational Items
The Number System	6
Ratios and Proportional Relationships	8
Expressions and Equations	9
Geometry	7
Statistics and Probability	6
Total	36

Table 11–6. Mathematics Grade 8 Test Blueprint

Reporting Category	Number of Operational Items
The Number System	6
Expressions and Equations	10
Functions	8
Geometry	6
Statistics and Probability	6
Total	36

Table 11–7. Algebra I Test Blueprint

Reporting Category	Number of Operational Items
Operations and Linear Equations & Inequalities	18
Operations with Real Numbers and Expressions	6
Linear Equations	6
Linear Inequalities	6
Linear Functions & Data Organizations	18
Functions	6
Coordinate Geometry	6
Data Analysis	6
Total	36

Table 11–8. ELA Grades 3–8 Test Blueprint

Reporting Category	Number of Operational Items
Reading	28–32
Literature Text	14–16
Informational Text	14–16
Writing	0
Conventions of Standard English (Writing)	0
Total	28–32

Note. In 2024–2025, only field-test writing items were included on Firefly ELA.

Table 11–9. Literature Test Blueprint

Reporting Category	Number of Operational Items
Fiction	15–19
Reading for Meaning—Fiction	6–8
Analyzing and Interpreting Literature—Fiction	9–11
Non-Fiction	15–19
Reading for Meaning—Nonfiction	6–8
Analyzing and Interpreting Literature—Nonfiction	9–11
Total	30–38

NUMBER OF ITEMS PER REPORTING CATEGORY

Various factors were considered when determining the number of operational items to administer per reporting category. One goal of Firefly is to assess student knowledge and skills on the same constructs measured by the end-of-year or end-of-course summative assessments in a single class period. As such, Firefly was envisioned as a short test that could be administered in a single class period (approximately 60 minutes) that could provide reporting category subscores. To reliably estimate subscores with low error, at least six items must be administered for each reporting category. These were the main considerations in determining the test length.

EMBEDDED FIELD-TEST ITEMS

As discussed in Chapter Five, embedded field-test items allow for evaluating item performance and estimating item parameters for inclusion in the operational item pool for future administrations. Embedding field-test items within operational tests is advantageous for two reasons. First, sufficient item level data can be gathered without the time and expense of a separate stand-alone administration. Second, it allows the new items to be placed on the existing operational scale.

For each embedded field-test event, the factors considered when determining the number of field-test items to embed included the number of items to be field tested, the expected number of students testing, and the desired N count per item to estimate item-level analyses. For mathematics field-test items and writing field-test items embedded within ELA tests, field-test items were randomly assigned to fixed positions spread throughout the operational test. In ELA and Literature, a field-test passage was randomly assigned near the middle of the test and students were administered all or most of the items associated with the passage. In all content areas, the positions of field-test items were unknown to students. Moreover, field-test items were not clustered at the end of the test to avoid any fatigue effect when estimating item parameters on the baseline scale.

In 2024–2025, field-test items were embedded in operational tests to bolster the operational item pool for Firefly. ELA consisted of field-test positions in both reading and writing subdomains. One passage with three to eight reading field-test items and eight writing items were presented on each ELA test. Writing items had previously estimated item parameters on a different underlying baseline scale than the reading items, therefore all writing items needed to be re-field tested for operational use in Firefly. Thus, in 2024–2025, writing items were not included in student scores. For Firefly Literature tests, one passage with approximately five to eight field-test items were presented on each test. For Firefly mathematics and Algebra I, six field-test items were presented on each test.

COMPUTER ADAPTIVE TEST ALGORITHM

Computer adaptive testing uses an algorithm that ensures each test a student is administered meets the content specifications and blueprint while concurrently selecting items matched to the student's estimated ability. Within each test event, the algorithm selects subsequent items by considering both the elements of the test blueprint (i.e., reporting categories) that have not been met yet and the student's performance on prior items administered. In other words, a student's prior performance on items determines the difficulty of the remaining content that is administered. At the end of the assessment, the student has been administered an assessment that met content and blueprint requirements and was optimally constructed for their demonstrated ability. The algorithm essentially defines the set of rules that define the adaptive administration of the test. This section covers elements of the CAT algorithm including entry point, item selection criteria, test navigation, and termination.

ENTRY POINT

Every computer adaptive test administered includes a different collection of items tailored to a student's estimated ability. Items are selected based on performance on prior items administered. Yet the entry or starting point for tests must be predetermined and specified in the algorithm because student performance (initial ability) is often unknown at the beginning of a computer adaptive test.

If a student has not previously taken Firefly within the same content area and grade/level, there is no prior information available about student performance. The entry point for Firefly is a small locator section in which a small number of items of average difficulty are administered. For Firefly math tests, one item of average difficulty is selected and administered for each reporting category. For Firefly ELA tests, students are assigned a passage with items of average difficulty.

If a student has previously taken Firefly within the same content area and grade/level, the previous ability estimate is used as a starting point in the algorithm. For Firefly math, if a student has previously taken Firefly math test in the same grade, then the first item selected in each reporting category in the next administration will be of similar difficulty to the student's prior ability estimate. Whereas if a student has previously taken Firefly ELA or Literature, then the first passage is selected based on the student's prior total score.

The CAT algorithm includes a randomization component when selecting items to control item exposure. That is, one item is selected from among a set of items that are near the targeted item difficulty. This is especially important at the entry point of Firefly when no prior information is available and items are selected from average difficulty.

Randomization of items and reporting categories helps to ensure that all students will not see the same set of items in the same order even when the assigned items of average difficulty.

ITEM SELECTION CRITERIA

Once the initial set of items has been administered, the CAT algorithm is designed to administer items targeted for the individual student based on performance. In targeting items, the CAT algorithm uses Rasch ability estimates from the current test session and considers a number of factors including test blueprint, response probability, item pool refinement, and passage-related concerns. Each of these is discussed in detail on the following pages.

RASCH ABILITY ESTIMATES

As described in Chapters Seven and Eight, Firefly item pools are calibrated and scaled using the Rasch partial credit model (Wright & Masters, 1982) and were vertically linked across grades and levels. The CAT algorithm has access to all item parameters in the item pool. After each item response, Rasch ability estimates and standard errors are calculated via maximum likelihood estimation (MLE) for the total test. In the case of zero (all items incorrect) and perfect (all items correct) scores, a correction factor is applied before computing the relevant maximum likelihood estimates. A fractional value is added to a zero score and subtracted from a perfect score before estimation.

After the locator section of Firefly, the total Rasch ability estimate is used in item selection. Using the total estimate rather than reporting category estimates prevents students from experiencing extreme fluctuations in the difficulty of items.

TEST BLUEPRINT

The CAT algorithm closely resembles a modified constrained CAT (MCCAT) design (Leung, Chang, & Hau, 2003). The CAT algorithm is configured with upper and lower bounds which specify the minimum and maximum numbers of items that will be administered to students for both total and reporting categories. The algorithm keeps track of what parts of the blueprint have been filled and what remains relative to the blueprint configurations as it proceeds through the assessment. Moreover, for Firefly mathematics assessments, the CAT algorithm further selects and administers content across assessment anchors where possible to ensure that a breadth of knowledge, skills and standards are assessed. For example, for Firefly mathematics grade 7, the Data and Statistics reporting category has three associated assessment anchors. To ensure that items are selected in a proportional way to the number of standards within each assessment anchor, at least one item must be selected from both D.S.1 and D.S.2, and at least 3 items must be selected from D.S.3 because the first two assessment anchors (D.S.1 and D.S.2) have fewer associated eligible content standards than the latter (D.S.3).

RESPONSE PROBABILITY

The Firefly CAT algorithm targets items where the student has a specific response probability (RP) of answering correctly, based on the ability estimate and item parameters associated with the item using the Rasch Model as discussed in Chapters Seven and Eight. Theoretically, the RP of 0.5 is the most efficient value to use within adaptive testing because item information is maximized at this probability. That is, selecting an item where the student has a 50 percent chance of getting the item correct will produce the smallest standard error for any given number of items. A response probability of 0.5 was used for all Firefly content areas and grades/levels.

ITEM POOL REFINEMENT

DRC's CAT algorithm has configurable elements that allow for refinement of the item pool used in item selection. Two configurable elements are:

- **Restrict pool**—The ability to restrict the available item pool by grade/course at various points in the test.
- **Favor items**—The ability to favor items that are close to the student's grade when evaluating items near a student's estimated score.

The difference between restricting the pool and favoring items is that when the pool is restricted, some items may NOT be selected. With favoring, all non-restricted items are eligible for administration, but they are made more or less likely to be selected based on closeness to student grade.

Firefly was designed as an on-grade assessment. As such, only on-grade items are included in the eligible item pool. Meaning, students will only be administered items that assess eligible content standards in their same grade level. This is a key differentiation between Firefly and the Classroom Diagnostic Tools (CDT) assessment in that CDT allows for the administration of off-grade content.

PASSAGE RELATED CONCERNS

As previously mentioned, Firefly ELA and Literature tests consist of passage-based items. Passages may have between three and twelve associated items. The CAT algorithm does not require that all items associated with a passage be administered. Instead, it evaluates all possible combinations of items within a passage. Item sequencing within a passage is preserved when items are presented to the student. For example, if a six-item passage is selected and items 1 and 4 are NOT administered, then the items administered in order will be 2, 3, 5, and 6.

The configurable elements of passage-based CAT include:

- **Passage minimum percent**—Define the minimum percentage of the items associated with a passage to be used.

For example, if the passage minimum percent is set at 60, then the selection routine will consider combinations such as 3 of 5 (60%), 3 of 4 (75%), 4 of 5 (80%), 5 of 6 (83%), and 6 of 6 (100%). It will not consider combinations such as 2 of 4 (50%), or 2 of 5 (40%), etc. Near the end of a test, the passage minimum percent constraint may need to be loosened to meet content constraints such as number of items per diagnostic category. Firefly utilizes a minimum of 60 percent for the administration of passages and passage-based items.
- **Passage evaluation criteria**—Multiple factors are considered when evaluating and ranking each passage combination to determine the best combination to administer to a student. They include:
 - Percent of items associated with the passage used; the higher the percent, the higher the combination is ranked
 - Number of items associated with the passage used; the higher the number, the higher the combination is ranked
 - Distance between items' difficulties and student's estimated score; the smaller the distance, the higher the combination is ranked

Different weights may be assigned to each of the factors. For example, if all of the weight is put on number of items used, then the algorithm will select the passages with the most associated items and administer all of them until the maximum number of items is reached.

TEST NAVIGATION

Many versions of computer adaptive tests do not allow students to skip items in the test or back up to previously answered items and change answers due to some complicating factors.

Currently, all Firefly mathematics tests do not permit skipping items, navigating between items nor changing answers. However, because Firefly ELA grades 3 through 8 tests include both passage-based items and stand-alone items, different rules apply. For the writing segment within Firefly ELA tests, students cannot skip or navigate between items like Firefly mathematics. For the Firefly Literature test and the reading segment of Firefly ELA tests, students may skip items and navigate between items within a passage. However, to move onto a new passage, a student must respond to each item presented within the passage. For example, when presented with a passage and five associated items, the student does not have to answer questions one through five in the order presented, rather a student can navigate between items within the passage. However, when the student attempts to advance to the next passage without answering all the items associated with a passage, the test engine prompts and requires the student to answer all items before moving on. Once the student advances to the next passage, students may no longer navigate to previously answered passage-based items.

TERMINATION

CAT algorithms can be configured for both fixed- or variable-length tests. For fixed-length tests, the test ends when a student has taken a predefined number of items in total and in each reporting category. While multiple methods can be used as the stopping rule for variable length tests, Firefly ELA and Literature tests are terminated once a student is administered the predetermined maximum number of items in total and within each reporting category. Note that with both fixed-and variable-length tests, there is no requirement that the predefined number of items in reporting categories be equal.

Firefly mathematics tests are fixed length with 36 operational items and 6 field-test items. Firefly ELA and Literature tests vary in length due to the nature of passage-based reading tests. Firefly ELA tests consist of 28 to 32 operational items and 3 to 8 field-test items. Firefly Literature tests consist of 30 to 38 operational items and 5 to 8 field-test items.

FIREFLY CAT CONFIGURATIONS

Elements of the CAT configurations that support the Firefly administration are reported below. An extensive set of simulations were run for each configuration and results are provided in the subsequent section.

FIREFLY MATHEMATICS GRADES 3–8 CONFIGURATIONS

Each grade of Firefly mathematics is configured with respect to five reporting categories (refer to Tables 11–1 through 11–9). The reporting categories differ by grade level and reflect the same reporting categories as PSSA mathematics. Each student is administered between six and ten items for each reporting category, which reflects the PSSA mathematics test blueprints with respect to each grade level. Additional details previously discussed include:

- Firefly mathematics is a fixed-length test.
- There is a locator section at the beginning of the test which consists of one item per reporting category.
- There are six field-test items administered throughout the test.

FIREFLY ALGEBRA I CONFIGURATION

The Firefly Algebra I test is configured with respect to six reporting categories, that represent assessment anchors. The reporting categories reflect the same reporting categories as Keystone Algebra I. Students are administered six items in each reporting category, which reflects similar proportions as the Keystone Algebra I test blueprint.

- Firefly Algebra I is a fixed-length test.
- There is a locator section at the beginning of the test which consists of one item per reporting category.
- There are six field-test items administered throughout the test.

FIREFLY ELA GRADES 3–8 CONFIGURATIONS

Each grade of Firefly ELA is configured with respect to three reporting categories (refer to Tables 11–1 through 11–9). The reporting categories are the same across grade levels and are similar to the reporting categories for PSSA ELA. The key difference between the blueprints between Firefly ELA and PSSA ELA is that PSSA ELA grades 4 through 8 includes a text-dependent analysis item (writing prompt) that is reported in a separate reporting category. Because Firefly ELA is a selected-response assessment with immediate score reporting, no text-dependent analysis or constructed-response items are included on Firefly ELA. Each student is administered between 14 to 16 items for each reading reporting category. Additional details previously discussed include:

- Firefly ELA is a variable-length test, consisting of 28 to 32 total items (14 to 16 per reading reporting category).
- There is no locator section.
- No operational writing items were administered in 2024–2025. There were eight field-test writing items on each Firefly ELA test in 2024–2025.
- One field-test passage is administered that consists of three to eight passage-based field-test items.

- Passage minimum is set to 60%. Whenever possible, only passage combinations that use 60% or more of the associated items are used. (Near the end of a test, the passage minimum percent constraint may need to be loosened to meet content constraints.)

FIREFLY LITERATURE CONFIGURATION

The Firefly Literature test is configured with respect to four reporting categories, which represent assessment anchors. The reporting categories reflect the same reporting categories as Keystone Literature. Students are administered 6 to 11 items for each reporting category, which reflects the Keystone Literature test blueprint.

- Firefly Literature is a variable-length test, consisting of 30 to 38 total items (6 to 11 per reporting category).
- There is no locator section.
- One field-test passage is administered that consists of five to eight passage-based field-test items.
- Passage minimum is set to 60%. Whenever possible, only passage combinations that use 60% or more of the associated items are used. (Near the end of a test, the passage minimum percent constraint may need to be loosened to meet content constraints.)

SIMULATION RESULTS

Once all content and psychometric specifications were configured for Firefly, extensive simulated test administrations were conducted prior to the administration launch. The simulation was designed to mimic actual test administrations using 5,000 to 10,000 students with known abilities. A representative sample of the CDT distribution of student abilities for each grade or level was selected as a basis for the simulation work in absence of Firefly operational data. The simulations tested the functionality of the adaptive testing system and item pool across the full range of student proficiency in the following areas: test blueprint coverage, item exposure, student ability estimation, and the standard errors of measurement. Simulation results are provided for each Firefly test. Overall, the results are as expected and meet the acceptable psychometric requirements given the available item pool.

TEST DESIGN AND BLUEPRINT COVERAGE

An essential requirement of adaptive testing is that each test administration must meet the test blueprint and associated test specifications to assure the comparability of student scores. In addition to verifying that student administrations are aligned to the test blueprint, we use the simulation tool to ensure item-order patterns and item sets were appropriate and expected. Table 11–10 summarizes the test design components from the simulation and indicates that every student received the number of items as configured.

Table 11–10. Simulation Summary: Test Design Results by Content Area and Grade/Level

Content Area	Grade/ Level	N	OP Item Ct Minimum	OP Item Ct Maximum	OP Item Ct Mean	OP Passage Ct Minimum	OP Passage Ct Max	FT Item Ct Minimum	FT Item Ct Maximum	FT Passage Ct
Mathematics	3	5000	36	36	36	0	0	6	6	0
Mathematics	4	5000	36	36	36	0	0	6	6	0
Mathematics	5	5000	36	36	36	0	0	6	6	0
Mathematics	6	5000	36	36	36	0	0	6	6	0
Mathematics	7	5000	36	36	36	0	0	6	6	0
Mathematics	8	5000	36	36	36	0	0	6	6	0
Algebra I	HS	10000	36	36	36	0	0	6	6	0
ELA	3	5000	28	32	30.45	6	7	12	16	1
ELA	4	5000	28	32	30.04	6	7	13	16	1
ELA	5	5000	28	32	30.27	6	8	11	16	1
ELA	6	5000	28	32	29.77	6	8	11	16	1
ELA	7	5000	28	32	29.48	6	8	11	16	1
ELA	8	5000	28	32	30.09	6	8	12	16	1
Literature	HS	10000	30	38	35.19	6	8	5	8	1

EVALUATING STUDENT ABILITY ESTIMATION

To estimate the examinee ability in the simulation study, maximum likelihood estimation (MLE) was utilized. To provide a limit to the score range for extreme values, the test scoring algorithm used the highest and lowest obtainable scale scores (i.e., HOSS and LOSS) that were derived during the creation of the vertical scale. Statistics from the simulations are organized in three general areas: bias of the ability estimates, magnitude of standard errors, and reliability.

Statistics for simulations that focused on ability estimation included the following:

- The statistical bias of the estimated student ability as a measure of whether scores systematically overestimate or underestimate student ability. This is a test of the assumption that error is randomly distributed around true ability.
- The Root Mean-Squared Error (MSE) is the square root of the magnitude of difference between true and estimated theta.
- The significance of the bias is an indicator of statistical significance.
- The average Standard Error (SE) of the estimated theta is the marginal reliability for the simulated population.
- Percentage of students' estimated theta falling outside the 95% and 99% confidence intervals.

BIAS

The relevant computational details for the statistics used to summarize the simulations are provided below and are followed by summary tables for each CAT configuration as well as some associated plots to facilitate the interpretability of the results.

At the test and reporting category levels, the bias is a difference between actual, θ_j , and estimated ability, $\hat{\theta}_j$, for the student.

$$Bias_j = \theta_j - \hat{\theta}_j ,$$

And the average bias over examinees, $\overline{Bias}$, is defined as

$$\overline{Bias} = \frac{\sum_{j=1}^N (\theta_j - \hat{\theta}_j)}{N} ,$$

where N is the number of total students. The standard deviation of the estimated bias is

$$SD_{Bias} = \sqrt{\frac{\sum_{j=1}^N (\theta_j - \bar{\hat{\theta}})^2}{N - 1}} ,$$

where $\bar{\hat{\theta}}$ is the average of estimated ability, $\hat{\theta}_j$, and the standard error of the estimated bias is

$$SE_{Bias} = \frac{SD_{Bias}}{\sqrt{N}}$$

The average bias is tested for statistical significance with a z-test,

$$z_{Bias} = \frac{\overline{Bias}}{SE_{Bias}} .$$

The z-statistic follow standard normal distribution, $N(0,1)$, and the p-value for a two-tailed test is reported. The mean squared error (MSE) is average of squared bias,

$$MSE = \frac{\sum_{j=1}^N (\theta_j - \hat{\theta}_j)^2}{N} .$$

At a student level, the degree of the deviation of the estimated ability from the actual ability is assessed with z-test

$$z_{Bias_j} = \frac{\theta_j - \hat{\theta}_j}{SE_{\hat{\theta}_j}} ,$$

where $SE_{\hat{\theta}_j}$ is the standard error of the estimated score for the student. The percentage of students who are outside of 95 percent and 99 percent of confidence interval are computed and reported. The critical values used for the confidence interval are 1.96 and 2.58 for the 95th confidence interval and 99th confidence interval, respectively.

Table 11–11 provides the estimated bias indices discussed above. The bias in the overall total scores tends to be small and non-significant for most content areas and grades/levels. For Mathematics grades 6, 7, 8 and Algebra I, the bias tends to be higher and significant. This will continue to be monitored in future administrations.

Table 11–11. Simulation Summary: Bias by Content Area and Grade/Level

Content Area	Grade/Level	Mean Bias	SE of Mean Bias	P-value Bias	RMSE	95% CI Miss Rate	99% CI Miss Rate
Mathematics	3	0.00	0.01	0.98	0.40	6.90	1.54
Mathematics	4	0.01	0.01	0.82	0.42	8.46	1.94
Mathematics	5	-0.04	0.01	0.22	0.45	9.56	2.94
Mathematics	6	-0.05	0.01	0.05	0.41	7.32	1.96
Mathematics	7	-0.05	0.01	0.02	0.43	6.64	1.92
Mathematics	8	-0.05	0.01	0.02	0.48	10.58	3.58
Algebra I	HS	-0.11	0.00	0.00	0.46	9.81	3.85
ELA	3	0.05	0.01	0.10	0.46	5.72	0.94
ELA	4	0.04	0.01	0.07	0.47	5.58	0.92
ELA	5	0.02	0.01	0.28	0.44	4.76	0.74
ELA	6	0.02	0.01	0.13	0.42	5.44	1.12
ELA	7	0.00	0.01	0.86	0.42	4.62	0.78
ELA	8	0.03	0.01	0.09	0.42	4.60	0.70
Literature	HS	-0.02	0.00	0.05	0.43	7.91	2.04

Figure 11–1 shows scatterplots of true ability versus estimate ability based on the simulated data. The identity line is shown as a reference. If all points were plotted on the identity line, it would indicate a perfect relationship between the true ability and estimated ability. In practice, the scatterplots show a reasonable relationship, as evidenced by the R^2 values between the two abilities, ranging from 0.87 to 0.92 for mathematics and 0.85 to 0.89 for ELA and Literature. In other words, the true ability accounts for 85% to 92% of the variance in the student ability estimates based on the simulation results.

Figure 11–1. Simulation Summary: True Ability Versus Estimated Student Ability

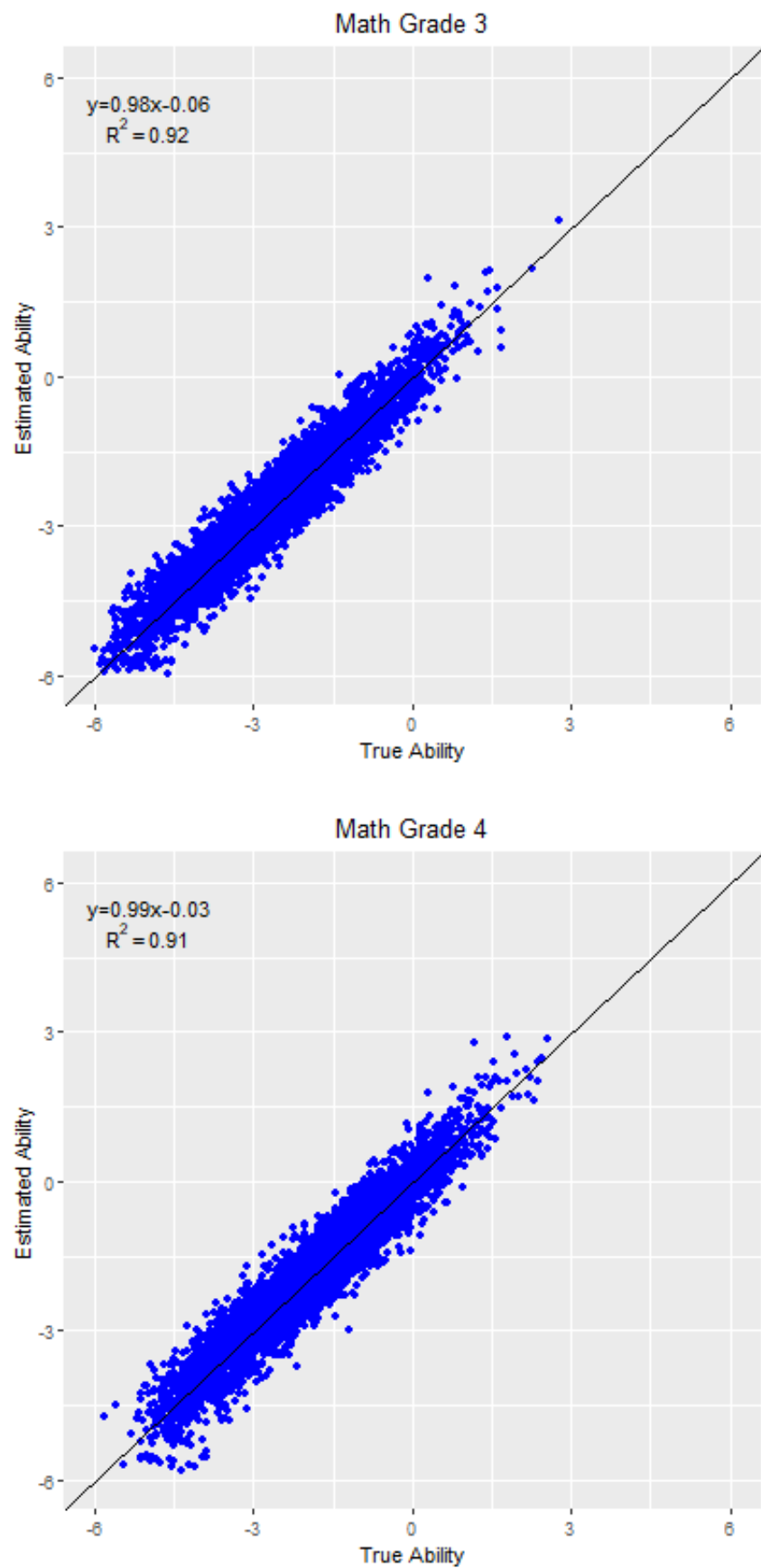


Figure 11–1 (continued). Simulation Summary: True Ability Versus Estimated Student Ability

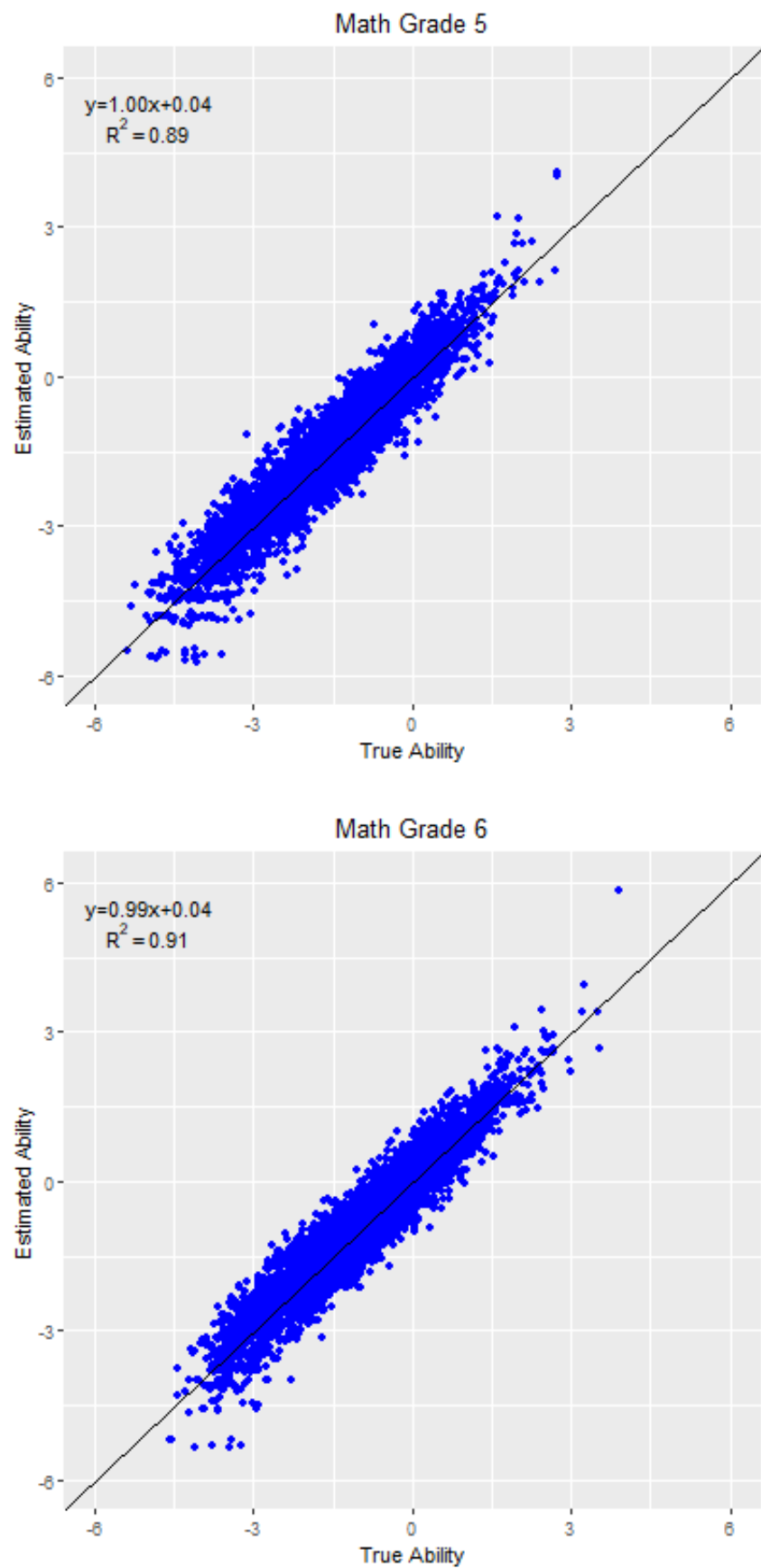


Figure 11–1 (continued). Simulation Summary: True Ability Versus Estimated Student Ability

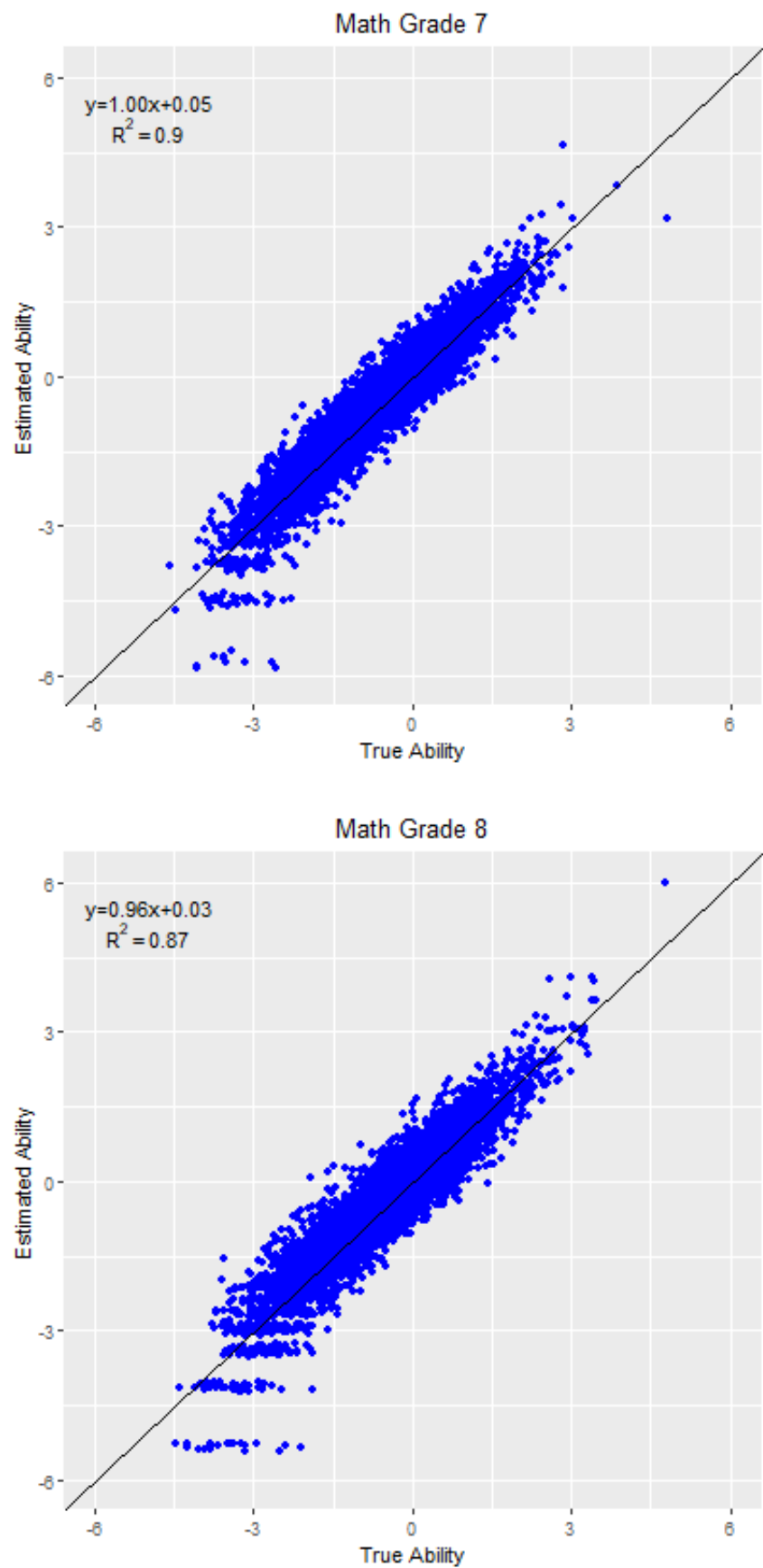


Figure 11–1 (continued). Simulation Summary: True Ability Versus Estimated Student Ability

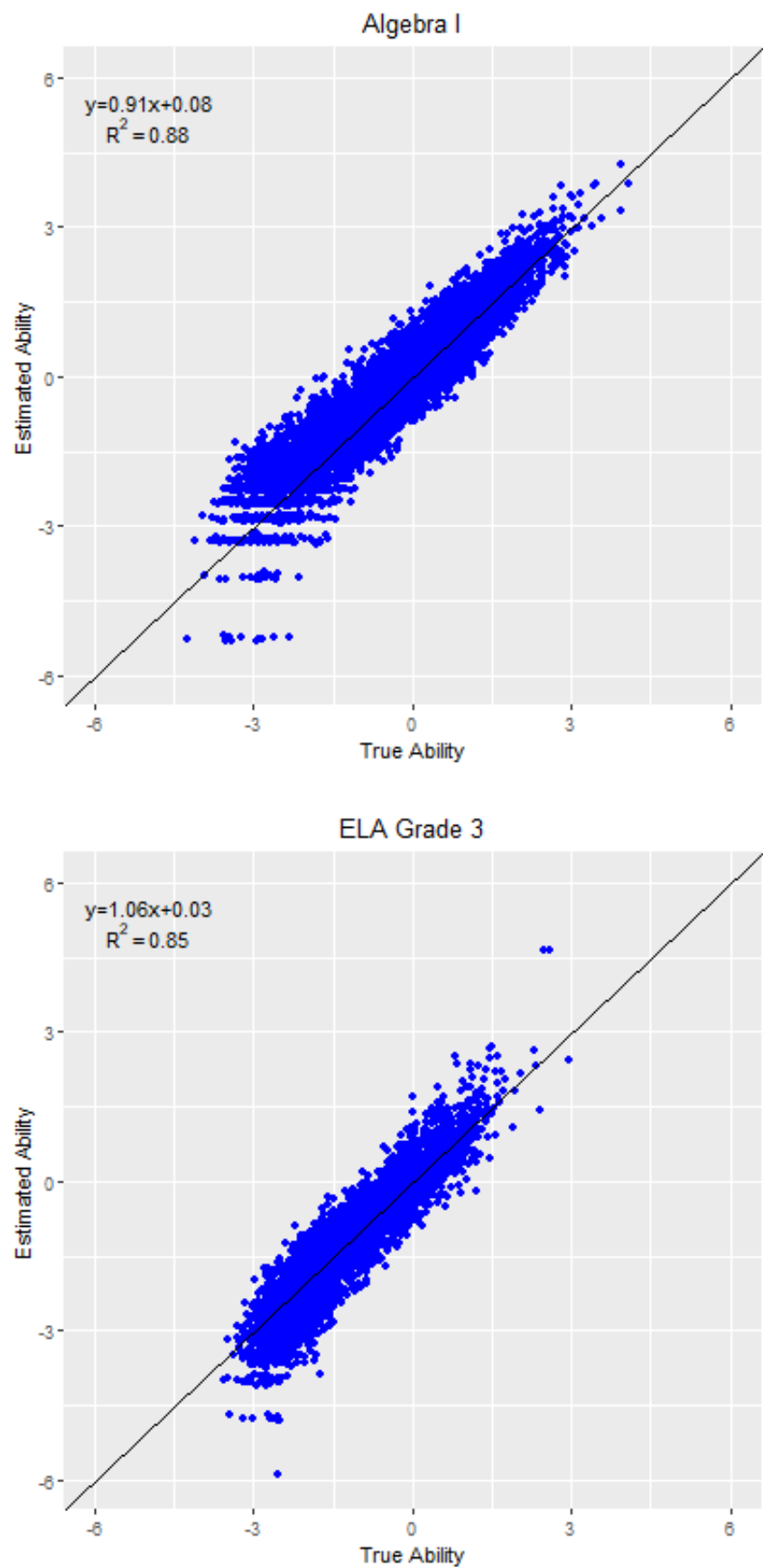


Figure 11–1 (continued). Simulation Summary: True Ability Versus Estimated Student Ability

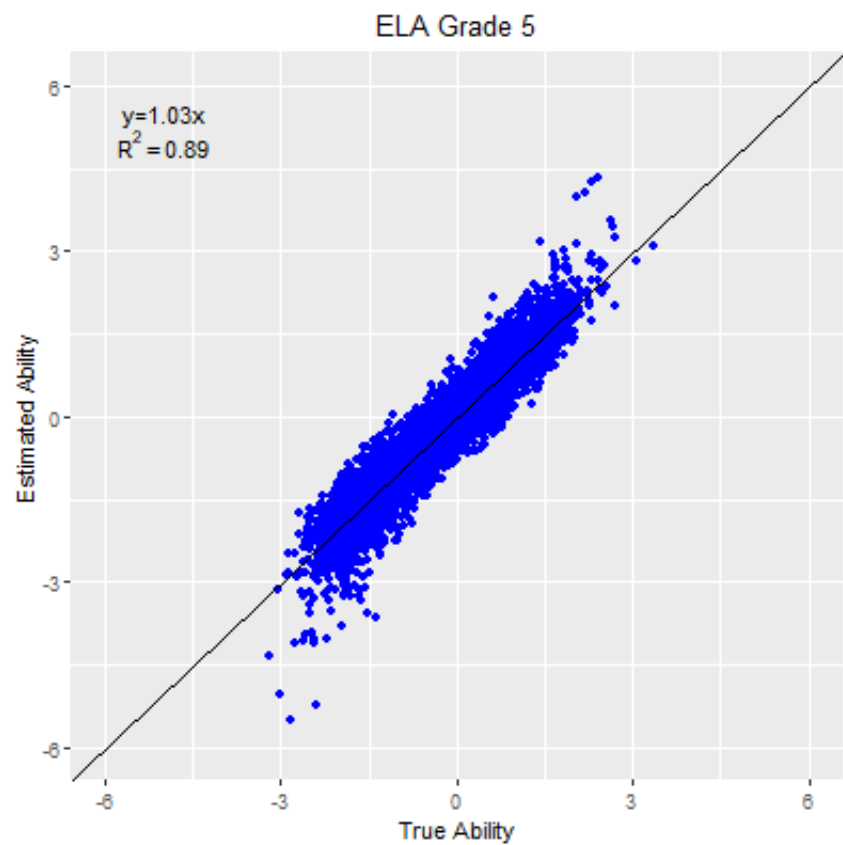
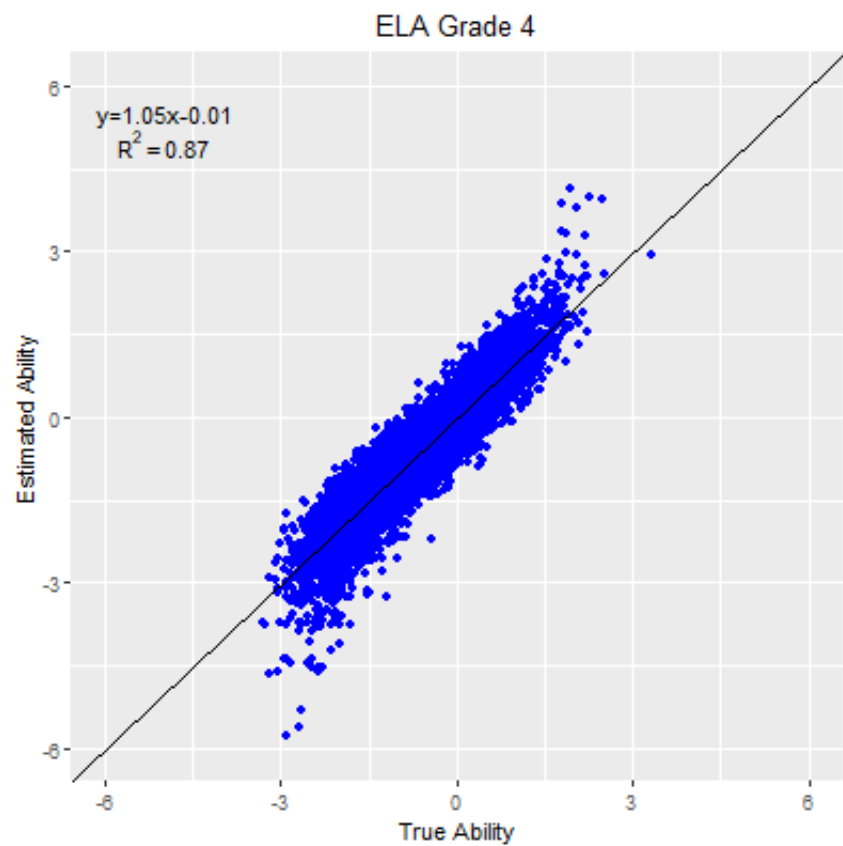


Figure 11–1 (continued). Simulation Summary: True Ability Versus Estimated Student Ability

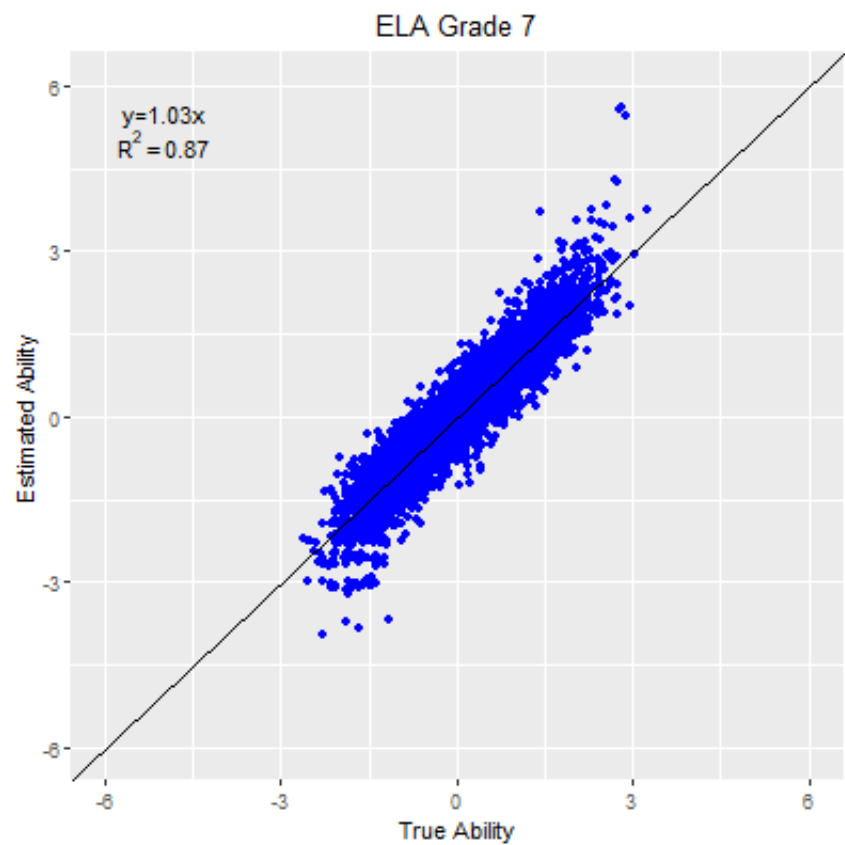
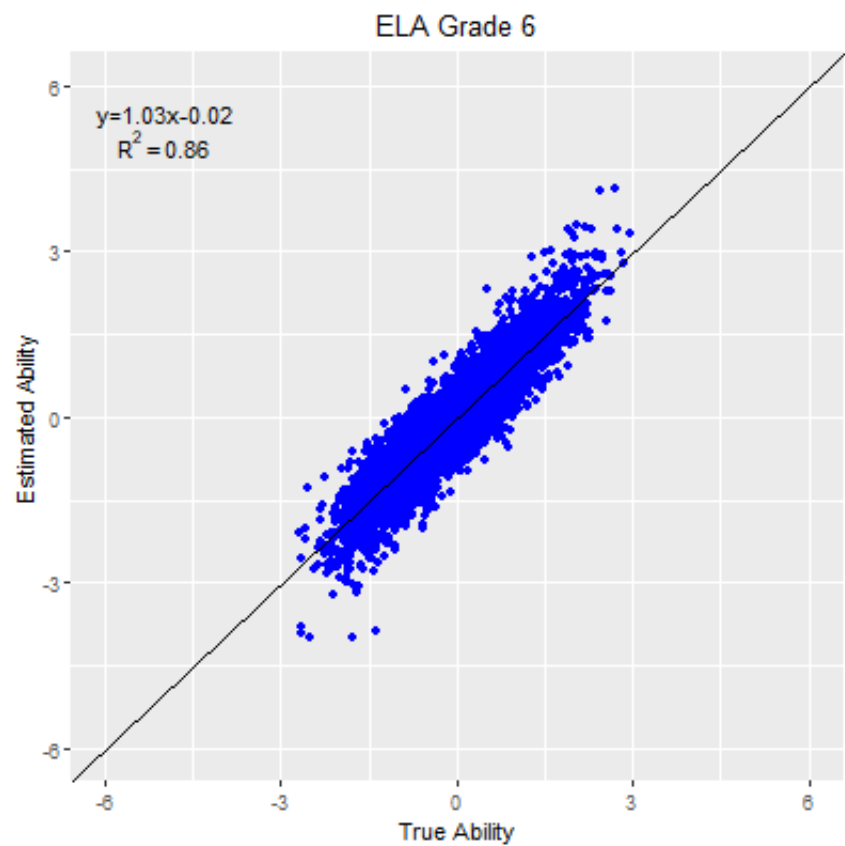
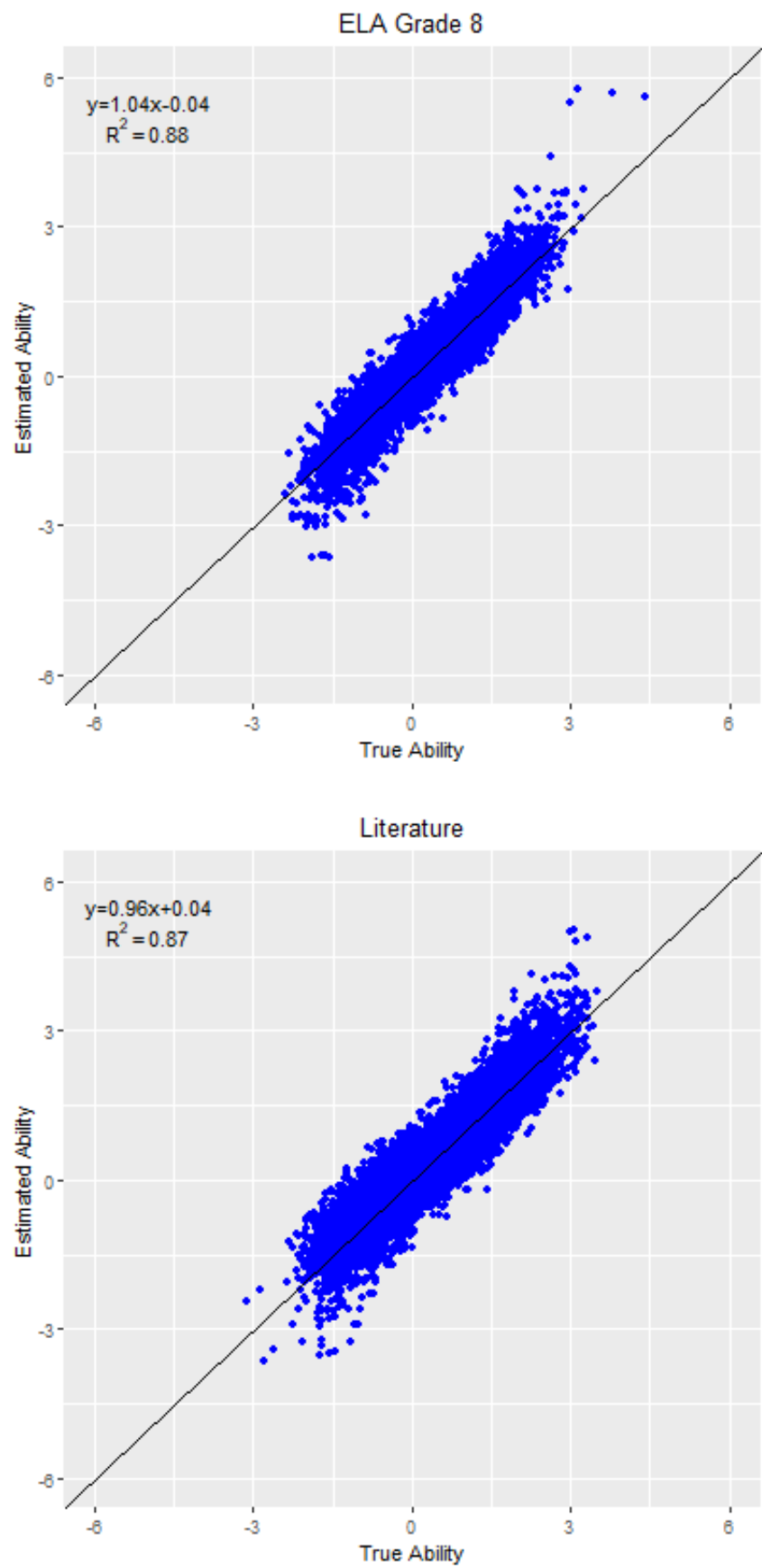


Figure 11–1 (continued). Simulation Summary: True Ability Versus Estimated Student Ability



CHAPTER TWELVE: TEST ADMINISTRATION PROCEDURES

TEST SETUP

The process to set up students to take a Firefly assessment is accomplished through an online interface located on the DRC INSIGHT Portal (<https://www.drcdirect.com/all/eca-portal-ui/welcome/PA>). The DRC INSIGHT Portal is a permission-based site that enables districts to assign users different roles and permissions depending on their role in the setup process. Each district can set up users with as much or as little permission as deemed necessary. A user's role and permission may be modified at any time.

Student records can be imported into the DRC INSIGHT Portal at any time during the Test Setup window (which opens approximately two weeks prior to the start of the test window and remains open through the entire test window). Once the data is imported, users organize students into test sessions and rosters. Test sessions and rosters can be created by class, grade, school, or any other variation.

Rosters are not required, but they are necessary for teachers to access student data. Student data is available to District and School Assessment Coordinators with or without rosters, but teachers can only access data for a roster that is assigned to them.

Note: Results from the 2024–2025 Firefly administration were available to District and School Assessment Coordinators only; therefore, the roster functionality was not available for the operational pilot administration.

Assessment coordinators are instructed to assign a specific teacher to each roster, and the roster should include the same students that are included in the test session for that teacher. Students may be assigned to multiple rosters. This allows districts/schools the ability to allow multiple users to view the data by class, grade, or even school. Rosters may be created and modified at any time during the administration window.

Test sessions are generated to create test tickets that are distributed to students prior to testing. A test ticket contains the student's full name, username, password, and the assessment he/she will be taking. The test session, like the roster, may also be created by class, grade, or school. Each time an assessment is administered, a new test session must be created. Test sessions can be copied to simplify the administration of a Firefly assessment to the same students multiple times each year.

TIMING AND SETTING

Firefly is an untimed test, with estimated testing times of approximately 50–60 minutes per test. Schools are instructed to administer each Firefly test in a single class period. Since not all students are expected to finish the test at the same time, test administrators are advised to use the flexibility of the time limits so that students do not feel rushed while they are taking the test, and no student is penalized because he or she works slowly.

School Assessment Coordinators and Test Administrators are advised to provide scheduled extended time for students with special requirements and/or abilities (i.e., physical, visual, auditory, or learning disabilities as defined by their IEP or service contracts) and for students who just work slowly. All students are allowed to request extended time if they indicate that they have not completed the task, and such requests are granted if the test administrator finds the request to be educationally valid. As a general guideline, when all students indicate that they have finished the test, that session is closed. Each student's test ticket locks overnight on the day the test was administered, regardless of whether the student completed the test.

Schools are required to administer Firefly tests in regular classroom settings or other settings that are approved for administering the PSSA or Keystone Exams. This requirement, in conjunction with administering the tests in a single classroom session, is meant to ensure the administration of Firefly closely mirrors the administration of the PSSA or Keystone Exams. The purpose is to strengthen the accuracy of using Firefly performance to predict student performance on the spring summative assessments. In other words, ensuring that Firefly is administered within the same timing and setting parameters as the summative assessments eliminates other variables that could affect the relationship between the two tests, such as students taking a Firefly test across multiple days or in a remote setting without close supervision of a test administrator.

Ultimately, schools have the final authority to modify the timing recommendations if students cannot complete the test in a single classroom period, especially in schools that have short class periods. In these scenarios, the recommendation is for students to complete the test in a separate class period on the same day. If such an outcome is not feasible, schools are able to unlock test tickets for students to complete the assessment on a subsequent day. In no cases, however, are schools approved to have students take Firefly tests outside of proctored, in-school settings (i.e., remote/home testing is not available or allowable).

PENNSYLVANIA (PA) ONLINE ASSESSMENTS SOFTWARE

Prior to testing, each student testing device needs to have the PA Online Assessments software installed. The testing software downloads are located on the DRC INSIGHT Portal. The installer is an MSI file that can be pushed out across a server to expedite the installation process. Once the software is installed, users also have access to the PA Assessment Online Student Tutorials and the PA Assessment Online Tools Training (OTT).

The PA Assessment Online Tools Training (OTT) is designed to provide an introductory experience using the online assessment software in preparation for taking Firefly. The purpose of the OTT is for students to observe and experiment with the features of the online assessment software prior to the actual assessment. The OTT is NOT designed to demonstrate complete coverage of the tested content, and it is not scored. Rather, sample items have been chosen to demonstrate online assessment features and uses.

In addition to student preparation, schools are also strongly encouraged to use the OTT prior to testing to allow technology staff to confirm the school is prepared to support online testing. Because the OTT interacts with DRC servers the same as a live Firefly test, students accessing the online training will provide a good indication that the software is installed correctly, and that the school's technology is adequate and configured properly.

The web-based PA Online Assessment Student Tutorials are available for each operational assessment and are designed to be used by students at all grade levels. They use pictures, motion, and sound to present visual and verbal descriptions of the features and functionality of the PA Online Assessment system. Schools are instructed to allow a minimum of 20 minutes to view the tutorials. Tutorials may be reviewed as often as needed.

TRAINING AND CUSTOMER SERVICE SUPPORT

DRC provides annual training for Technology Coordinators through a live web conference each fall. A recording of the technology training is available throughout the school year on the DRC INSIGHT Portal. Technology Coordinator training focuses on all technical aspects required for the setup of Firefly. Detailed installation instructions for the PA Online Assessments Software and Central Office Services – Service Device (COS-SD) are provided. (The COS-SD is a local network server that helps mitigate internet traffic by allowing student machines to retrieve items from the COS-SD rather than from DRC servers.) Firefly requires an internet connection at all times during testing, and schools using one-to-one student devices are required to use a COS-SD.

Test Administration training is also available to district and school staff via the Pennsylvania Assessment training site (<https://sites.google.com/datarecognitioncorp.com/psa-assessments/home>). The training site is available year-round and provides training materials and videos specific to Firefly. Test coordinators and administrators are directed to access the site prior to administering Firefly to gain valuable information about processes that need to be followed before, during, and after test administration. The DRC INSIGHT Portal also hosts a wide array of Firefly documents that are designed to help users understand the purpose and functionality of the PA Firefly assessments.

In addition to the training and materials that are readily available, DRC has a live customer service support team that is available to assessment coordinators during regular school hours. Assessment and technology coordinators are encouraged to call or email DRC with any questions or issues that cannot be resolved locally. DRC employs a hierarchal system that will escalate an issue to support engineers or developers if a problem cannot be resolved via a customer service representative. If an issue requires more research, DRC will contact the caller daily to provide an update.

CHAPTER THIRTEEN: SCORES AND SCORE REPORTS

REPORT DEVELOPMENT

Several months prior to the first release of reports for Firefly, DRC collaborated with PDE to design an Individual Student Report (ISR) that would clearly convey the most important aspects of the benchmark tests. The collaborative effort included multiple rounds of revision and final approval by PDE's Communication Bureau. The draft ISRs were also reviewed by Pennsylvania's Technical Advisory Committee and PDE's Core Team for CDT and Firefly oversight. In addition, DRC developed Report Interpretation Guides that include definitions of each element of a Firefly ISR. Instructions to access the Report Interpretation Guides are referenced on the ISR and the guides are posted on PDE's website for access by parents, educators, and the public.

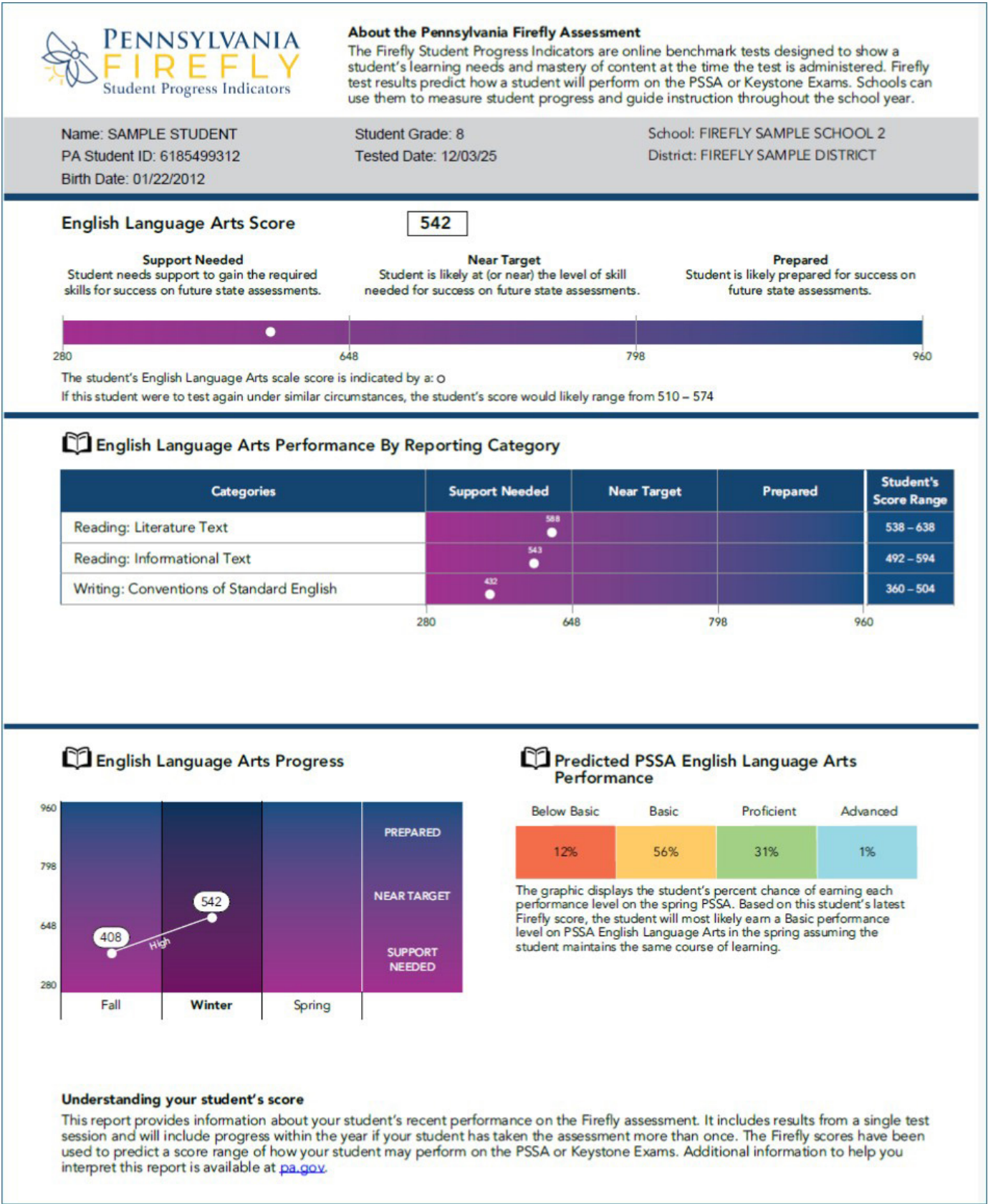
SCORE REPORTS

The following score reports were available to schools and districts for the 2024–2025 Firefly administration.

- Individual Student Reports
- School Roster Reports

The Individual Student Reports (ISRs) display easy-to-comprehend total score and sub-score information, as well as a separate section for within-year progress (across seasons) and a separate section for predicted performance on the summative assessments. The Firefly ISRs were designed to ensure the color scheme would not convey a direct correlation between Firefly performance levels and the PSSA and Keystone Exams performance levels, which are displayed in red, yellow, green, and blue for Below Basic, Basic, Proficient, and Advanced. Accordingly, a gradient shade of purple was selected for the Firefly ISRs and moves from lighter-to-darker across the three Firefly performance levels of Support Needed, Near Target, and Prepared. The ISRs also include text to help students, parents, and educators understand the overall content and important information about interpreting progress and predictions.

Figure 13–1. Firefly Individual Student Report



Note. The ISRs for 2024–2025 did not display data for Progress or Predictions. These sections were first populated on the ISRs for the 2025–2026 Firefly administration.

The School Roster Report was available to all District and School Assessment Coordinators for LEAs that participated in the 2024–2025 Firefly administration. The downloadable and exportable report is formatted to provide the same level of detail that is included on the ISRs. In addition, the roster report includes a record for each Firefly test taken by a student, which allows schools to track student participation and performance across the school year. The interactive roster report can also be manipulated by users to select the columns displayed or to filter fields to focus on specific data elements.

Figure 13–2. School Roster Report

School Name ¹ ▼	Student Name ¹ ▼	PA Student ID ¹ ▼	Birth Date ▼	Grade ¹ ▼	Subject ¹ ▼	Test Date ¹ ▼	Overall Scale Score ▼	Overall Performance Level ▼	Reporting Category 1 Name ▼	Reporting Category Scale Score ▼
SAMPLE SCHOOL 1	BADSTUDENTID_MEKHI	8160033605	04/02/2009	3	Mathematics	10/28/2024	586	Prepared	Numbers and Operations in Base Ten	481
SAMPLE SCHOOL 1	ADAMS_VIRGINIA_G	4812727391	10/03/2003	4	English Language Arts	11/15/2024	686	Prepared	Reading: Literature Text	662
SAMPLE SCHOOL 1	ADAMS_VIRGINIA_G	4812727391	10/03/2003	4	Mathematics	11/15/2024	771	Prepared	Numbers and Operations in Base Ten	665
SAMPLE SCHOOL 1	ALEXANDER_CALVIN_F	6290988166	10/20/2003	4	Mathematics	10/28/2024	601	Near Target	Numbers and Operations in Base Ten	702
SAMPLE SCHOOL 1	AUSTIN_VANESSA_E	4473460843	11/12/2003	4	English Language Arts	10/28/2024	549	Near Target	Reading: Literature Text	535

Note. The 2024–2025 roster reports were not available to DRC INSIGHT Portal users with Teacher-level access. The roster reports were first made available to Teachers in the 2025–2026 administration (in conjunction with the release of the rostering functionality within Test Setup for that administration). In addition, the roster reports for 2024–2025 did not display data for Progress or Predictions. These sections were first populated on the roster reports starting with the 2025–2026 Firefly administration.

FUTURE SCORE REPORTS

Reporting for the 2024–2025 operational pilot was limited to ISRs and School Roster Reports to allow DRC sufficient time to gather data and requirements to expand Firefly reporting to a full suite of interactive reports. The following additional reports will be released for the 2025–2026 Firefly administration.

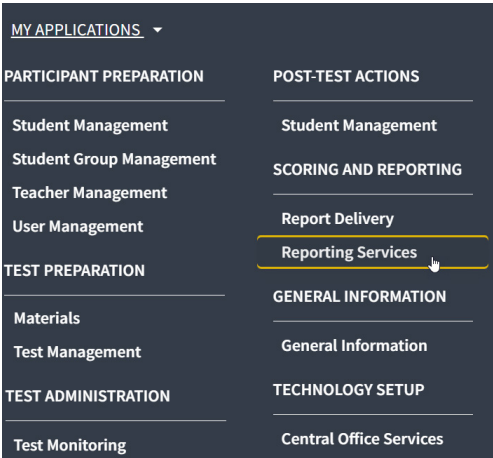
- Instructional Supports
- Test Status
- Student Status
- Subject Comparison
- Reporting Category Comparison
- Group Performance
- Longitudinal Roster
- Group Progress

ACCESSING SCORE REPORTS

Firefly reports are available in the DRC INSIGHT Portal within one hour of a student submitting a Firefly test. All Portal users with district or school-level permissions for the 2024–2025 Firefly administration had access to the available reports via Reporting Services. DRC also provides a comprehensive Firefly Report Guide that includes easy-to-follow instructions for Portal users to access the Firefly reports. The intuitive nature of the DRC’s Reporting Services functionality and the accompanying support materials ensured that all districts and schools had full access to student results from the 2024–2025 Firefly administration.

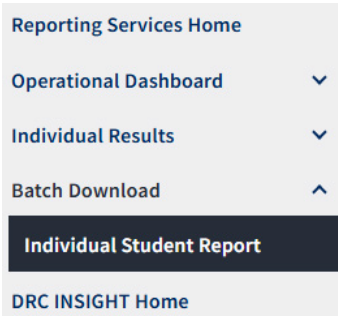
Once a user is logged into the Portal, Reporting Services can be selected under MY APPLICATIONS, at the top of the screen.

Figure 13–3. My Applications dropdown on the DRC INSIGHT Portal



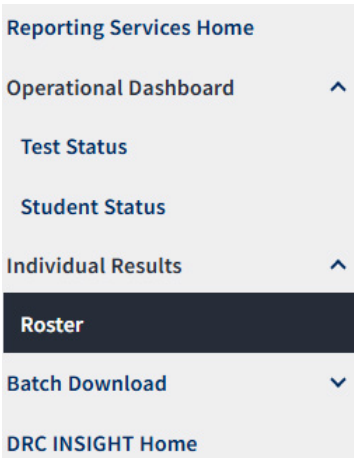
To access the Individual Student Reports, a user would next select Individual Student Report from the Batch Download dropdown.

Figure 13–4. Batch Download and Individual Student Report



Displays the list of applications in the DRC INSIGHT Portal. It highlights selecting Reporting Services link then to access the School Roster Report, a user would select Roster from the Individual Results dropdown.

Figure 13–5. Individual Results and Roster



A user is then required to add District and School filters for the 2024–2025 Firefly administration to produce the applicable reports. For the ISRs, users are able to add Subject and Grade filters to generate a more specific group of reports for downloading or printing.

CHAPTER FOURTEEN: OPERATIONAL ADMINISTRATION 2024–2025

This chapter contains summary information about the operational administration of Firefly during the 2024–2025 school year.

DATA SOURCES AND FILTERING CRITERIA

This section describes the datasets used for different analyses and results discussed throughout the Technical Report. Unlike statewide summative tests, Firefly participation is voluntary and students can take Firefly up to three times per year. In the design of Firefly as an interim assessment, Pennsylvania Department of Education and DRC provided guidance that students in grades 3 through 8 should take Firefly once per season (fall, winter, and spring) to serve as checkpoints in meeting end-of-year expectations. For high school students, or students enrolled in high school courses (Algebra I and Literature), the guidance of once per season is not as valid due to differences in student schedules (i.e., block scheduling, quarter or semester courses, etc.).

The criteria for inclusion in student datasets are:

- The student must complete the test
- The student grade must match the test grade
- The first test administration if the student completed the same test within two weeks

For post-administration analyses (e.g., predictions and evaluation), additional criteria were implemented such that if a student completed multiple tests within the same testing window (fall, winter, and spring), the most recent administration was retained.

KEY VALIDATION DATA

These data are only mentioned for the sake of completeness, as no formal results from this data are provided in this technical document. For the first operational administration, an analysis of all operational items was conducted once sufficient item *N* counts were available to confirm that the items are performing as expected. This is an important quality check. Key validation data represented students that met the criteria by late January 2025.

DATA REVIEW

Data used for item review included students who met the criteria above by April 1. No sampling was undertaken in this data.

FINAL DATA

The final student data represents tests administered throughout the year if the inclusion criteria were met. No sampling was undertaken in this data. The final dataset was used for classical item analyses and calibration for item statistics banking, technical reporting, as well as matching for the predictive analyses. The final dataset is used as the primary source of data presented in this technical report. The final dataset includes students who were administered the same Firefly test multiple times, and the data presented in this report summarizes all student records, including repeat testers, unless otherwise specified.

FIREFLY MATCHED TO SPRING SUMMATIVE

There are several analyses conducted for Firefly which rely on merging and matching data across Pennsylvania assessments. Matched datasets are required for prediction model evaluation and generation, as well as evaluation and validation of performance-level consistency. The following inclusion criteria was used for the PSSA and Keystone datasets.

PSSA DATA INCLUSION CRITERIA

- No duplicated student identification numbers within a subject/grade level
- Students that completed the test

KEYSTONE DATA INCLUSION CRITERIA

- No duplicated student identification numbers within a content area
- No retesters
- Students that completed both modules of the test
- Students who were administered the same form for both modules
- Students who completed and received scores (i.e., no invalidation or exclusion codes)
- Students who used the video sign language (VSL) form were excluded from analyses
- Students in grades 6 and above

Firefly records were matched with PSSA and Keystone data based on Pennsylvania student identification number relative to each content area. For Firefly, the technical reporting filtering criteria was applied (completed test, no subsequent tests within two weeks, and matched student grade level and test grade level). PSSA and Firefly records were matched by content area and grade for students in grades 3 through 8. Keystone Algebra I and Literature were matched to Firefly Algebra I and Literature, respectively based solely on student identification number.

The *N* counts for all data sources are provided in Table 14–1.

Table 14–1. Data Source *N* Counts

Content Area	Grade/Level	Key Validation	Data Review	Final	Matched with EOY/ EOC Summative
Mathematics	3	10,477	27,110	31,621	30,966
Mathematics	4	10,205	25,524	29,043	28,452
Mathematics	5	10,768	27,583	31,646	30,894
Mathematics	6	10,170	26,444	29,745	29,005
Mathematics	7	10,274	23,245	26,581	25,760
Mathematics	8	8,787	21,291	24,033	23,182
Algebra I	HS	10,215	21,704	26,927	21,281
ELA	3	9,196	23,086	26,345	25,772
ELA	4	9,129	22,690	25,689	25,118
ELA	5	9,625	25,064	28,077	27,421
ELA	6	8,705	23,833	26,547	25,876
ELA	7	9,604	21,633	24,015	23,327
ELA	8	10,039	21,200	23,941	23,179
Literature	HS	9,115	20,601	23,820	16,078

ADMINISTRATION COUNTS

Tables 14–2 through 14–4 present information related to the number of Firefly administrations during the 2024–2025 school year. Firefly was designed as an interim assessment that could be administered to Pennsylvania students up to three times per year. The following tables present information about the number of administrations by season and by test number. Table 14–2 presents the administration counts by season of administration – Fall, Winter, and Spring. Table 14–3 presents the administration counts by test number (1 to 3). The student count with one test event represent the unique, non-duplicated student counts.

As Table 14–2 shows, the highest volume of administrations occurred in the winter, which lasted from December 1 through the end of February. The lowest volume of administrations occurred in the fall season, due to the relatively late launch in November through the end of November. Overall, Firefly mathematics was administered slightly more than Firefly ELA or Literature in the first operational year.

Table 14–2. Number of Firefly Administrations by Content Area, Grade/Level, and Season

Content Area	Grade/Level	Fall	Winter	Spring	Total
Mathematics	3	1,960	18,318	11,343	31,621
Mathematics	4	2,556	16,459	10,028	29,043
Mathematics	5	2,672	18,152	10,822	31,646
Mathematics	6	2,186	16,816	10,743	29,745
Mathematics	7	1,686	15,619	9,276	26,581
Mathematics	8	1,746	14,238	8,049	24,033
Algebra I	HS	2,196	14,306	10,425	26,927
ELA	3	2,062	14,788	9,495	26,345
ELA	4	2,324	14,192	9,173	25,689
ELA	5	2,560	15,783	9,734	28,077
ELA	6	1,757	16,426	8,364	26,547
ELA	7	1,913	15,256	6,846	24,015
ELA	8	2,029	14,767	7,145	23,941
Literature	HS	1,684	14,056	8,080	23,820

Firefly was designed as an interim assessment that could be administered to students up to three times within a content area each year. Table 14–3 shows the number of Firefly administrations by test number. As the table shows, over 20,000 students took Firefly within each content area and grade/level. For each content area and grade/level, more than 3,500 students took Firefly twice, and less than 1,000 students took Firefly three times. Given that 2024–2025 was the first operational year, the number of repeat administrations is expected to grow in subsequent administrations.

Table 14–3. Number of Firefly Administrations by Content Area, Grade/Level, and Test Number

Content Area	Grade/Level	Test No.1	Test No.2	Test No.3	Total
Mathematics	3	23,232	7,427	962	31,621
Mathematics	4	21,358	6,991	694	29,043
Mathematics	5	23,155	7,799	692	31,646
Mathematics	6	23,188	6,302	255	29,745
Mathematics	7	21,199	5,278	104	26,581
Mathematics	8	19,618	4,292	123	24,033
Algebra I	HS	21,821	4,800	306	26,927
ELA	3	19,546	6,040	759	26,345
ELA	4	18,937	6,155	597	25,689
ELA	5	20,932	6,500	645	28,077
ELA	6	21,699	4,658	190	26,547
ELA	7	20,436	3,501	78	24,015
ELA	8	20,002	3,849	90	23,941
Literature	HS	18,794	4,864	162	23,820

While Firefly mathematics and ELA grades 3 through 8 are on-grade assessments, Algebra I and Literature can be administered to students in grades 6 through 12 because high school courses and Keystone Exams can be offered to students in middle school grades. Table 14–4 presents the number of administrations by grade for high school Firefly tests, Algebra I and Literature. As Table 14–4 shows, most Firefly Algebra I tests were administered to students in grades 8, 9, and 10 while most Firefly Literature tests were administered to students in grades 9 and 10. These trends are somewhat consistent with Keystone first-time tester patterns in the spring.

Table 14–4. Number of Firefly High School Administrations by Content Area and Grade

Content Area	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12
Algebra I	23	881	5,110	14,643	4,899	1,232	139
Literature	22	220	367	6,559	15,928	685	39

DEMOGRAPHIC CHARACTERISTICS

Given that schools and teachers are encouraged to administer Firefly multiple times throughout the year, the following tables rely on the first administration of Firefly to report demographic and background characteristics. Demographic data were obtained primarily from Pennsylvania Information Management System (PIMS) and subsequently transmitted to DRC. However, Firefly can be administered to students without demographic data in PIMS at the time of testing, which may result in incomplete demographic information.

Frequency data for various demographic and background characteristics are presented in Tables 14–5M and 14–5E for mathematics and ELA, respectively. The unique student count is shown as the first row of the table of which the percentages are based. Percentages in some categories may sum to a quantity below 100 percent due to missing data. Caution should be used in making conclusions about demographic data given participation in Firefly is voluntary and complete demographic data via PIMS is not required prior to testing.

As shown in the tables below, the racial/ethnic backgrounds of Firefly testers differ from that of the state population. Specifically, a higher percentage of Firefly testers are white students and lower percentages of testers are Black and Hispanic compared to statewide summative assessment data. In addition, the percentage of economically disadvantaged students taking Firefly is also lower than that of the population.

Table 14–5M. Demographic Characteristics of Students Taking Firefly Mathematics

Demographic or Background Characteristic	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8	Algebra I
Number of Students	23232	21358	23155	23188	21199	19618	21821
Female (N)	11534	10380	11232	11351	10336	9530	10256
Female (Pct)	49.65	48.60	48.51	48.95	48.76	48.58	47.00
Male (N)	11698	10978	11923	11837	10863	10088	11565
Male (Pct)	50.35	51.40	51.49	51.05	51.24	51.42	53.00
American Indian or Alaskan Native (N)	43	43	36	312	28	54	67
American Indian or Alaskan Native (Pct)	0.19	0.20	0.16	1.35	0.13	0.28	0.31
Asian non-Hispanic (N)	815	598	922	846	810	783	1159
Asian non-Hispanic (Pct)	3.51	2.80	3.98	3.65	3.82	3.99	5.31
Black/African American non-Hispanic (N)	2569	2188	2199	2419	2057	1785	2416
Black/African American non-Hispanic (Pct)	11.06	10.24	9.50	10.43	9.70	9.10	11.07
Hispanic (N)	2935	2242	2287	2385	1954	1864	2496
Hispanic (Pct)	12.63	10.50	9.88	10.29	9.22	9.50	11.44
Multi-Racial non-Hispanic (N)	1410	1227	1288	1218	1072	954	1050
Multi-Racial non-Hispanic (Pct)	6.07	5.74	5.56	5.25	5.06	4.86	4.81
Native Hawaiian or Pacific Islander (N)	229	180	185	189	153	200	18
Native Hawaiian or Pacific Islander (Pct)	0.99	0.84	0.80	0.82	0.72	1.02	0.08
White/Caucasian non-Hispanic (N)	15231	14880	16238	15819	15125	13978	14615
White/Caucasian non-Hispanic (Pct)	65.56	69.67	70.13	68.22	71.35	71.25	66.98
IEP (N)	3135	3182	3303	2958	2538	2513	2776
IEP (Pct)	13.49	14.90	14.26	12.76	11.97	12.81	12.72
English Learner (N)	1391	992	859	758	559	531	975
English Learner (Pct)	5.99	4.64	3.71	3.27	2.64	2.71	4.47
Economically Disadvantaged (N)	7895	6801	7096	6819	6219	5358	5876
Economically Disadvantaged (Pct)	33.98	31.84	30.65	29.41	29.34	27.31	26.93

Table 14–5E. Demographic Characteristics of Students Taking Firefly ELA and Literature

Demographic or Background Characteristic	Gr. 3	Gr. 4	Gr. 5	Gr. 6	Gr. 7	Gr. 8	Literature
Number of Students	19546	18937	20932	21699	20436	20002	18794
Female (N)	9597	9208	10157	10645	9964	9791	8981
Female (Pct)	49.10	48.62	48.52	49.06	48.76	48.95	47.79
Male (N)	9949	9729	10775	11054	10472	10211	9813
Male (Pct)	50.90	51.38	51.48	50.94	51.24	51.05	52.21
American Indian or Alaskan Native (N)	39	38	29	317	21	46	50
American Indian or Alaskan Native (Pct)	0.20	0.20	0.14	1.46	0.10	0.23	0.27
Asian non-Hispanic (N)	689	476	764	799	997	811	787
Asian non-Hispanic (Pct)	3.53	2.51	3.65	3.68	4.88	4.05	4.19
Black/African American non-Hispanic (N)	2285	1961	2081	2256	2062	1759	1855
Black/African American non-Hispanic (Pct)	11.69	10.36	9.94	10.40	10.09	8.79	9.87
Hispanic (N)	1994	1863	1942	2277	1816	1780	1761
Hispanic (Pct)	10.20	9.84	9.28	10.49	8.89	8.90	9.37
Multi-Racial non-Hispanic (N)	1186	1124	1198	1167	1047	998	839
Multi-Racial non-Hispanic (Pct)	6.07	5.94	5.72	5.38	5.12	4.99	4.46
Native Hawaiian or Pacific Islander (N)	219	184	183	186	153	205	12
Native Hawaiian or Pacific Islander (Pct)	1.12	0.97	0.87	0.86	0.75	1.02	0.06
White/Caucasian non-Hispanic (N)	13134	13291	14735	14697	14340	14403	13490
White/Caucasian non-Hispanic (Pct)	67.20	70.19	70.39	67.73	70.17	72.01	71.78
IEP (N)	2596	2809	2823	2673	2350	2546	2217
IEP (Pct)	13.28	14.83	13.49	12.32	11.50	12.73	11.80
English Learner (N)	778	774	691	617	502	523	514
English Learner (Pct)	3.98	4.09	3.30	2.84	2.46	2.61	2.73
Economically Disadvantaged (N)	6363	5921	6174	6640	5944	5557	5194
Economically Disadvantaged (Pct)	32.55	31.27	29.50	30.60	29.09	27.78	27.64

TEST LENGTH

Test length and testing time was an important consideration in the design and development of Firefly. One of the goals for Firefly administrations was for educators to be able to administer Firefly in a single class period. Thus, Firefly was designed to be a relatively short test, but one that would also allow for reporting subscores. Tables 14–6 and 14–7 present descriptive statistics for the test length for each content area and grade/level for operational items and field-test items, respectively. Summary statistics include minimum, maximum, quartiles 1 (Q1) and 3 (Q3), median, mean and standard deviation (SD). For each content area and grade/level, the table shows the number of student records, the number of items (and the number of points in parentheses). For mathematics, the number of points and number of items is the same for all grades/levels because only 1-point items are in the mathematics item pools. Moreover, because mathematics are fixed-length tests, the data indicates that 36 items and points were administered in every test. However, for ELA and Literature, multi-point technology-enhanced (TE) and evidence-based selected-response (EBSR) items are in the item pool and the total number of item points administered may be greater than the number of items administered. For ELA grades 3 through 8, on average students answered 30 to 31 operational items, worth 34 to 35 points. For Literature, on average students answered 35 operational items, worth 39 points.

Table 14–6. Test Length Summary Statistics of Operational Items (Points)

Content Area	Grade/Level	N Students	Minimum	Q1	Median	Q3	Maximum	Mean
Mathematics	3	31621	36(36)	36(36)	36(36)	36(36)	36(36)	36.0(36.0)
Mathematics	4	29043	36(36)	36(36)	36(36)	36(36)	36(36)	36.0(36.0)
Mathematics	5	31646	36(36)	36(36)	36(36)	36(36)	36(36)	36.0(36.0)
Mathematics	6	29745	36(36)	36(36)	36(36)	36(36)	36(36)	36.0(36.0)
Mathematics	7	26581	36(36)	36(36)	36(36)	36(36)	36(36)	36.0(36.0)
Mathematics	8	24033	36(36)	36(36)	36(36)	36(36)	36(36)	36.0(36.0)
Algebra I	HS	26927	36(36)	36(36)	36(36)	36(36)	36(36)	36.0(36.0)
ELA	3	26345	28(31)	31(35)	31(35)	31(36)	32(38)	31.0(35.3)
ELA	4	25689	28(31)	30(34)	31(34)	31(35)	32(38)	30.7(34.4)
ELA	5	28077	28(31)	31(35)	31(36)	32(36)	32(38)	31.2(35.7)
ELA	6	26547	28(29)	30(33)	30(34)	31(35)	32(38)	30.2(34.0)
ELA	7	24015	28(30)	29(34)	30(35)	30(35)	32(38)	29.7(34.5)
ELA	8	23941	28(31)	30(34)	30(35)	31(36)	32(38)	30.3(34.7)
Literature	HS	23820	31(33)	34(38)	35(39)	36(40)	38(44)	34.9(39.0)

Table 14–7 presents the test length summary statistics for field-test items and points. For every mathematics test, students were administered 6 field-test items worth a total of 6 points. For each ELA grade 3 to 8 test, students were administered 1 field-test passage set, which could vary in length and 9 field-test writing items. Therefore, ELA grades 3 through 8 testers on average answered 14 to 15 field-test items. Whereas Literature testers were administered one field-test passage set representing an average of 8 field-test items.

Table 14–7. Test Length Summary Statistics of Field-Test Items (Points)

Content Area	Grade/Level	N Students	Minimum	Q1	Median	Q3	Maximum	Mean
Mathematics	3	31621	6(6)	6(6)	6(6)	6(6)	6(6)	6.0(6.0)
Mathematics	4	29043	6(6)	6(6)	6(6)	6(6)	6(6)	6.0(6.0)
Mathematics	5	31646	6(6)	6(6)	6(6)	6(6)	6(6)	6.0(6.0)
Mathematics	6	29745	6(6)	6(6)	6(6)	6(6)	6(6)	6.0(6.0)
Mathematics	7	26581	6(6)	6(6)	6(6)	6(6)	6(6)	6.0(6.0)
Mathematics	8	24033	6(6)	6(6)	6(6)	6(6)	6(6)	6.0(6.0)
Algebra I	HS	26927	6(6)	6(6)	6(6)	6(6)	6(6)	6.0(6.0)
ELA	3	26345	12(13)	13(15)	15(16)	15(18)	16(24)	14.2(16.4)
ELA	4	25689	13(14)	14(17)	15(18)	16(19)	16(24)	15.1(17.9)
ELA	5	28077	11(11)	15(18)	16(19)	16(20)	16(24)	15.3(18.7)
ELA	6	26547	11(12)	16(17)	16(19)	16(20)	16(23)	15.3(18.3)
ELA	7	24015	11(11)	15(16)	15(17)	16(20)	16(25)	15.0(17.7)
ELA	8	23941	12(12)	15(16)	16(18)	16(19)	16(23)	15.3(17.4)
Literature	HS	23820	5(7)	8(8)	8(8)	8(10)	8(13)	7.8(8.8)

Tables 14–8 and 14–9 summarize information about reading passages for ELA grades 3–8 and Literature for operational use and field-test use, respectively. While the total number of items is controlled by the CAT algorithm and configuration, the number of operational reading passages is not. The table reports the minimum, median, and maximum of operational reading passages by grade/level as well as the minimum, median and maximum number of items per passage. Across all grades/levels, 6 to 8 operational reading passages were presented in each test. Overall, students were presented with 3 to 7 items with each reading passage, in which the median items per passage was 5 for grades 3 to 8 and 6 for high school Literature.

Table 14–8. Operational Reading Passage Summary

Content Area	Grade/Level	Min Passages	Med Passages	Max Passages	Min Item Ct/ Passage	Med Item Ct/Passage	Max Item Ct/Passage
ELA	3	6	6	7	3	5	6
ELA	4	6	6	7	3	5	6
ELA	5	6	6	7	3	5	6
ELA	6	6	6	7	3	5	6
ELA	7	6	6	8	3	5	6
ELA	8	6	6	7	3	5	6
Literature	HS	6	6	8	3	6	7

Table 14–9 presents similar information as Table 14–8 for field-test reading passages. The table reports the number of field-test reading passages by grade/level, as defined by the CAT algorithm, as well as the minimum, median, and maximum number of field-test items per passage. As the table shows, one field-test passage was administered during each test which consisted of 3 to 8 associated items. The median number of passage-based field-test items students were presented with was 8.

Table 14–9. Field-Test Reading Passage Summary

Content Area	Grade/Level	Passage Ct	Min Item Ct/ Passage	Med Item Ct/Passage	Max Item Ct/Passage
ELA	3	1	4	8	8
ELA	4	1	5	8	8
ELA	5	1	3	8	8
ELA	6	1	3	8	8
ELA	7	1	3	8	8
ELA	8	1	4	8	8
Literature	HS	1	5	8	8

STUDENT PERFORMANCE

Table 14–10 presents student performance on each Firefly test. Because within each content area, Firefly is vertically scaled, one can compare scale scores across grade levels. Within each content area, the mean scale scores increase across grades. As expected, there are smaller increases in mean scale scores between mathematics grade 8 and Algebra I. The majority of testers across all content areas and grade levels earned the lowest performance level classification of Support Needed. Specifically, about 68% to 74% of Firefly math testers in grades 4 through Algebra I were classified as Support Needed; whereas only 61% of math grade 3 testers earned a performance level of Support Needed. For ELA and Literature, there was a more wide range of testers classified in the lowest performance level (59% to 70%). A relatively small percentage of students earned a Near Target or Prepared classification.

Table 14–10. Student Performance Summary by Content Area and Grade/Level

Content Area	Grade/ Level	N	Mean Scale Score	SD Scale Score	Support Needed (%)	Near Target (%)	Prepared (%)
Mathematics	3	31621	463.61	95.28	61.25	28.52	10.23
Mathematics	4	29043	498.49	89.41	69.70	22.65	7.65
Mathematics	5	31646	532.36	89.03	67.80	23.84	8.36
Mathematics	6	29745	562.91	82.57	69.64	24.76	5.60
Mathematics	7	26581	585.58	81.86	73.50	22.02	4.49
Mathematics	8	24033	611.51	84.60	72.64	21.62	5.74
Algebra I	HS	26927	619.44	84.32	71.78	22.58	5.64
ELA	3	26345	431.44	98.73	62.66	32.69	4.65
ELA	4	25689	492.63	110.98	60.27	33.20	6.53
ELA	5	28077	536.95	109.45	58.70	35.54	5.76
ELA	6	26547	548.85	112.24	66.40	30.12	3.48
ELA	7	24015	565.43	111.48	69.11	27.80	3.09
ELA	8	23941	583.24	112.56	70.17	26.93	2.90
Literature	HS	23820	621.03	109.67	60.73	36.62	2.64

Note. Percent of students classified in each performance level: Needs Support, Near Target, and Prepared.

Table 14–11 presents the same information as Table 14–12, but with respect to season. It is expected that as students progress through the curriculum during the school year, that student performance would improve given more opportunity to learn. For all content areas and grade/levels except ELA grades 6, 7, and 8, mean scale scores and percentage of students earning performance-level classifications of Near Target or Prepared increased across the seasons. For ELA grades 6–8, there were increases from fall to winter, but modest decreases from the winter administration to the spring administration.

Table 14–11. Student Performance Summary by Content Area, Grade/Level, and Season

Content Area	Grade/Level	Season	N	Mean Scale Score	SD Scale Score	Support Needed (%)	Near Target (%)	Prepared (%)
Mathematics	3	Fall	1960	445.17	86.57	69.95	25.10	4.95
Mathematics	3	Winter	18318	456.31	92.69	64.18	27.66	8.16
Mathematics	3	Spring	11343	478.59	98.84	55.01	30.49	14.49
Mathematics	4	Fall	2556	482.40	81.78	78.17	17.88	3.95
Mathematics	4	Winter	16459	491.74	85.17	72.91	21.22	5.86
Mathematics	4	Spring	10028	513.68	95.75	62.28	26.20	11.53
Mathematics	5	Fall	2672	511.89	77.97	78.41	18.19	3.41
Mathematics	5	Winter	18152	526.71	85.62	70.16	22.99	6.85
Mathematics	5	Spring	10822	546.88	94.94	61.22	26.65	12.13
Mathematics	6	Fall	2186	549.17	72.82	77.36	20.40	2.24
Mathematics	6	Winter	16816	557.15	80.72	72.45	22.98	4.57
Mathematics	6	Spring	10743	574.73	85.84	63.66	28.44	7.90
Mathematics	7	Fall	1686	574.09	78.01	80.37	16.25	3.38
Mathematics	7	Winter	15619	582.53	80.08	75.25	21.03	3.72
Mathematics	7	Spring	9276	592.81	84.91	69.30	24.72	5.98
Mathematics	8	Fall	1746	583.11	71.40	85.85	12.54	1.60
Mathematics	8	Winter	14238	610.18	83.54	73.39	21.14	5.47
Mathematics	8	Spring	8049	620.04	87.57	68.46	24.44	7.11
Algebra I	HS	Fall	2196	595.18	72.14	83.47	15.39	1.14
Algebra I	HS	Winter	14306	614.94	81.33	74.40	21.17	4.43
Algebra I	HS	Spring	10425	630.72	88.95	65.72	26.02	8.26
ELA	3	Fall	2062	424.93	92.58	66.25	30.60	3.15
ELA	3	Winter	14788	430.08	98.30	63.30	32.22	4.48
ELA	3	Spring	9495	434.97	100.55	60.90	33.86	5.24
ELA	4	Fall	2324	488.64	110.55	62.01	32.01	5.98
ELA	4	Winter	14192	492.90	110.16	59.96	33.62	6.42
ELA	4	Spring	9173	493.24	112.33	60.30	32.86	6.85
ELA	5	Fall	2560	533.77	107.75	60.16	34.57	5.27
ELA	5	Winter	15783	536.06	108.10	59.08	35.56	5.37

Table 14–11 (continued). Student Performance Summary by Content Area, Grade/Level, and Season

Content Area	Grade/Level	Season	N	Mean Scale Score	SD Scale Score	Support Needed (%)	Near Target (%)	Prepared (%)
ELA	5	Spring	9734	539.24	112.01	57.72	35.76	6.52
ELA	6	Fall	1757	542.02	105.02	69.61	28.46	1.94
ELA	6	Winter	16426	551.19	111.92	65.57	30.87	3.56
ELA	6	Spring	8364	545.70	114.19	67.36	28.99	3.65
ELA	7	Fall	1913	564.59	112.05	68.69	27.97	3.35
ELA	7	Winter	15256	569.85	110.93	68.01	28.70	3.28
ELA	7	Spring	6846	555.82	111.92	71.68	25.74	2.59
ELA	8	Fall	2029	569.02	108.04	74.47	23.61	1.92
ELA	8	Winter	14767	585.38	112.76	69.43	27.63	2.95
ELA	8	Spring	7145	582.85	113.11	70.50	26.42	3.08
Literature	HS	Fall	1684	601.86	102.67	68.29	30.52	1.19
Literature	HS	Winter	14056	619.13	108.67	61.81	35.70	2.49
Literature	HS	Spring	8080	628.33	112.16	57.29	39.49	3.22

Note. Percent of students classified in each performance level: Needs Support, Near Target, and Prepared.

SCALE SCORES AND CONDITIONAL STANDARD ERROR

Table 14–12 shows the descriptive statistics for scale scores.

Table 14–12. Scale Score Summary by Content Area and Grade/Level

Content Area	Grade/Level	N	Minimum	Q1	Median	Q3	Maximum	Mean	SD
Mathematics	3	31621	180	394	463	532	860	463.61	95.28
Mathematics	4	29043	200	434	494	558	880	498.49	89.41
Mathematics	5	31646	220	469	525	591	900	532.36	89.03
Mathematics	6	29745	240	507	562	617	920	562.91	82.57
Mathematics	7	26581	282	524	583	642	931	585.58	81.86
Mathematics	8	24033	386	548	607	665	960	611.51	84.60
Algebra I	HS	26927	367	556	618	676	980	619.44	84.32
ELA	3	26345	180	355	423	500	860	431.44	98.73
ELA	4	25689	200	407	484	572	880	492.63	110.98
ELA	5	28077	220	452	534	616	900	536.95	109.45
ELA	6	26547	240	461	546	632	920	548.85	112.24
ELA	7	24015	260	477	561	646	940	565.43	111.48
ELA	8	23941	280	494	583	663	960	583.24	112.56
Literature	HS	23820	300	532	631	705	980	621.03	109.67

Figure 14–1 shows the scale score distributions for Firefly mathematics and ELA. The cut scores are shown by vertical dashed lines.

Figure 14–1. Score Distribution by Content Area, and Grade/Level

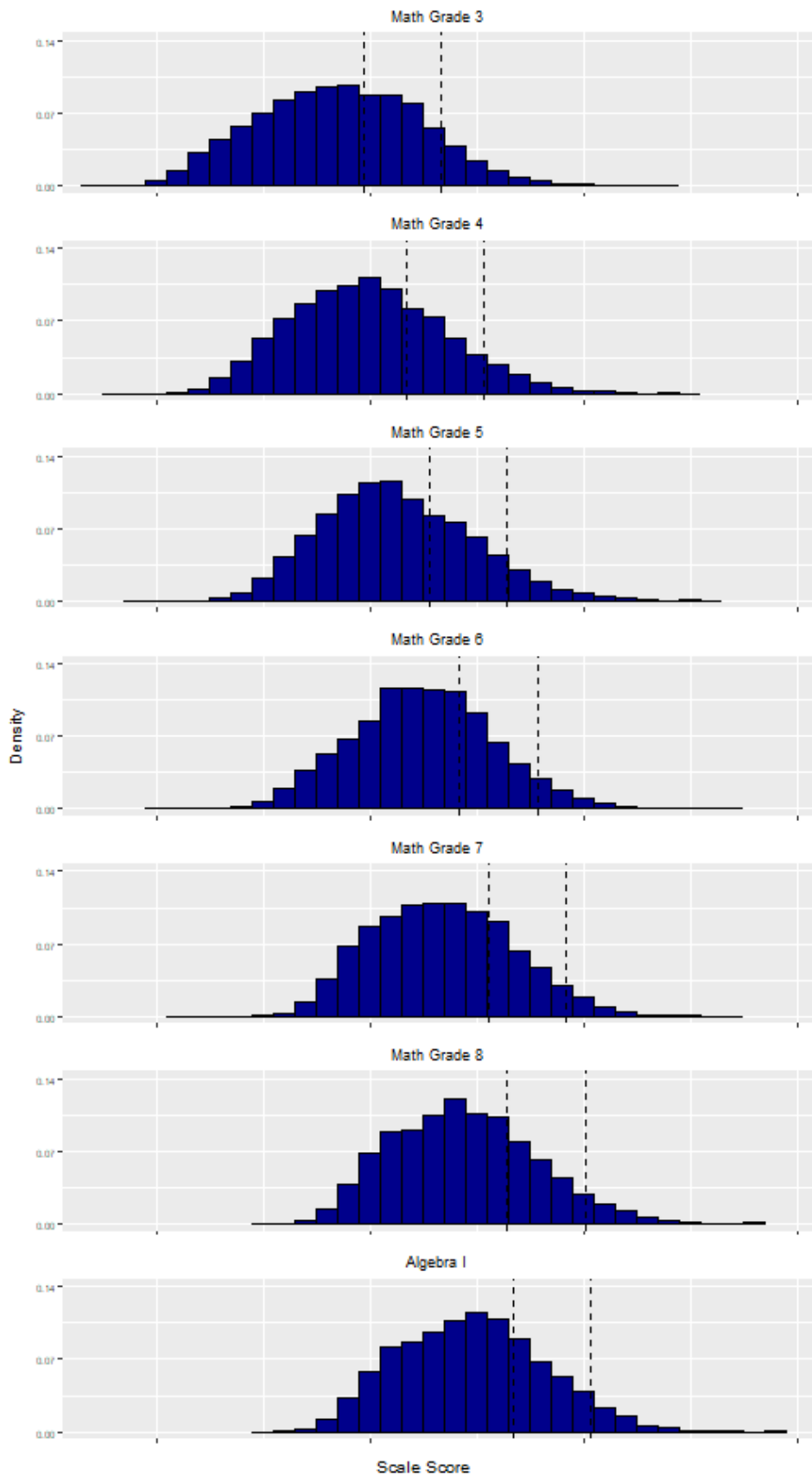


Figure 14–1 (continued). Score Distribution by Content Area, and Grade/Level

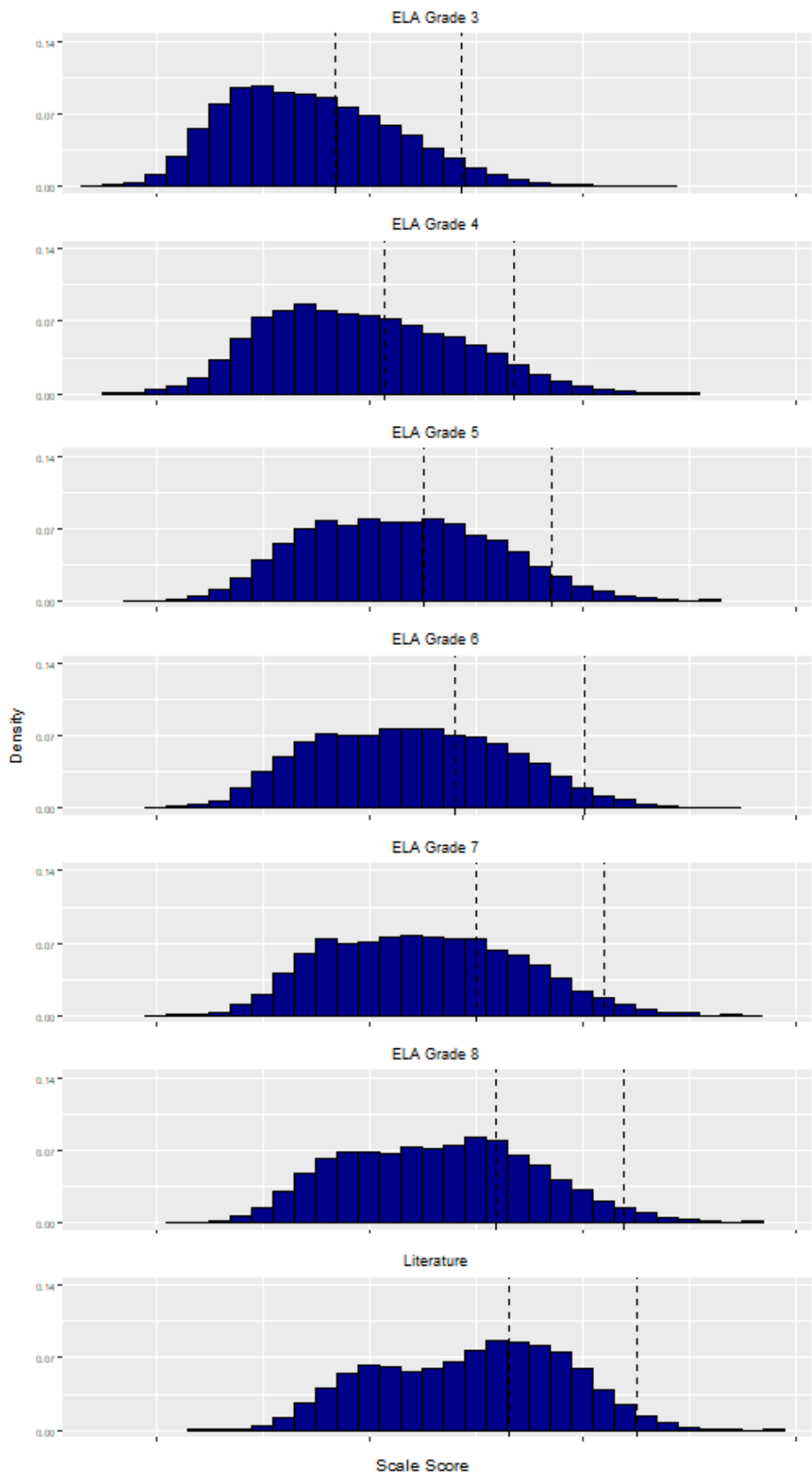


Table 14–13 shows the summary statistics for the conditional standard errors of measurement (CSEM) in the scale score metric. CSEM provides information about the precision of scale scores. For Firefly mathematics, the average and median CSEM is about 24 to 25 scale score points. For ELA and Literature, the average and median CSEM is about 34 to 38 scale score points. The CSEM based on the scale scores cannot be directly compared between ELA and mathematics because they do not have the same underlying nor reporting scale properties.

Table 14–13. Conditional Standard Error Summary by Content Area and Grade/Level

Content Area	Grade/Level	N	Minimum	Q1	Median	Q3	Maximum	Mean	SD
Mathematics	3	31621	23	24	25	26	92	25.48	2.82
Mathematics	4	29043	23	24	25	26	86	25.35	2.65
Mathematics	5	31646	23	24	25	26	90	25.53	3.60
Mathematics	6	29745	23	24	24	25	71	25.00	2.09
Mathematics	7	26581	23	24	24	26	71	25.31	2.44
Mathematics	8	24033	23	24	24	26	88	25.51	3.43
Algebra I	HS	26927	23	24	24	25	74	25.06	2.39
ELA	3	26345	33	35	36	38	91	37.47	3.43
ELA	4	25689	33	35	37	40	78	38.24	4.31
ELA	5	28077	33	35	36	38	80	37.09	4.12
ELA	6	26547	33	35	37	40	82	38.28	4.37
ELA	7	24015	33	35	36	39	86	37.85	4.29
ELA	8	23941	33	35	36	39	87	37.53	4.13
Literature	HS	23820	31	33	34	36	74	35.33	3.44

STUDENT PROGRESS AND PROGRESS LEVELS

For Firefly, student progress is defined as the difference in scale scores between two timepoints. For grades 3 through 8, the timepoints are defined by season – Fall, Winter, and Spring. For Firefly Algebra I and Literature, student progress is defined relative to two test events, rather than seasons, due to differences in schedules (i.e., traditional, block or alternate) which often occur in high schools. Progress, calculated as the difference in scale score points between two timepoints, can be classified into levels to provide meaning and magnitude. Specifically, there are four progress level classifications in Firefly: Not Yet Progressing, Low Progress, Medium Progress, and High Progress. The following definitions are used to classify the progress between two timepoints.

- Not Yet is defined as one of two scenarios. The scale score difference is negative, indicating a lower scale score at the second timepoint compared to the first timepoint. In addition, scale score differences less than 0.5 times the standard error of measurement (SEM) of the first timepoint.
- Low Progress is defined as scale score differences greater than or equal to 0.5 times the SEM and less than 1.5 times the SEM.
- Medium Progress is defined as scale score differences greater than or equal to 1.5 times the SEM and less than 2.5 times the SEM.
- High Progress is defined as scale score differences greater than or equal to 2.5 times the SEM.

Progress and Progress Levels were not defined or presented on score reports in 2024–2025. However, as a part of the post-administration analyses, DRC conducted analyses following the 2024–2025 administration to support progress levels for subsequent years.

Table 14–14 presents a summary of the progress levels based on the 2024–2025 data. The table provides the number of students who tested between two subsequent timepoints (i.e., fall and winter for grades 3–8, and test event 1 and 2 for high school), along with the percentage of those students who classified into each progress level. The students shown in for fall and spring did not have a test event in the winter. For mathematics, there is reasonable representation in each of the progress levels. For higher mathematics grade levels (7, 8 and Algebra I), a higher percentage of students are classified in “Not Yet Progressing.” For ELA and Literature, a higher percentage of students are classified in “Not Yet Progressing” than mathematics, and there is a relatively small percentage of students classified into “High Progress” (often less than 15%).

Table 14–14. Progress Level Summary

Content Area	Grade/ Level	Timepoint 1	Timepoint 2	N	Not Yet (%)	Low (%)	Medium (%)	High (%)
Mathematics	3	Fall	Winter	1024	35.25	20.51	16.50	27.73
Mathematics	3	Winter	Spring	5496	33.82	17.38	18.72	30.08
Mathematics	3	Fall	Spring	586	25.77	15.53	19.80	38.91
Mathematics	4	Fall	Winter	1062	36.82	19.49	20.53	23.16
Mathematics	4	Winter	Spring	4471	32.92	19.08	19.08	28.92
Mathematics	4	Fall	Spring	1030	22.14	14.27	18.35	45.24
Mathematics	5	Fall	Winter	979	45.97	21.86	15.53	16.65
Mathematics	5	Winter	Spring	5435	38.42	20.17	18.22	23.20
Mathematics	5	Fall	Spring	1221	29.16	17.36	18.26	35.22
Mathematics	6	Fall	Winter	534	38.95	22.47	19.29	19.29
Mathematics	6	Winter	Spring	4165	39.93	20.77	17.41	21.90
Mathematics	6	Fall	Spring	913	28.70	19.50	20.70	31.11
Mathematics	7	Fall	Winter	292	51.03	20.21	16.10	12.67
Mathematics	7	Winter	Spring	3784	45.24	22.67	15.99	16.09
Mathematics	7	Fall	Spring	636	45.44	20.44	18.40	15.72
Mathematics	8	Fall	Winter	236	50.00	18.64	16.10	15.25
Mathematics	8	Winter	Spring	2896	41.16	20.03	16.71	22.10
Mathematics	8	Fall	Spring	648	42.13	19.75	13.27	24.85
Algebra I	HS	Test Event 1	Test Event 2	4800	48.79	17.54	13.96	19.71
Algebra I	HS	Test Event 2	Test Event 3	305	47.21	19.02	14.10	19.67
ELA	3	Fall	Winter	878	50.11	23.69	16.06	10.14
ELA	3	Winter	Spring	4166	52.38	21.32	15.07	11.23
ELA	3	Fall	Spring	751	41.54	22.77	19.97	15.71
ELA	4	Fall	Winter	1037	53.91	17.84	15.43	12.83
ELA	4	Winter	Spring	3901	52.50	20.61	14.30	12.59
ELA	4	Fall	Spring	895	47.82	21.90	13.97	16.31
ELA	5	Fall	Winter	1222	55.48	20.13	13.18	11.21
ELA	5	Winter	Spring	4487	57.79	19.50	12.81	9.90

Table 14–14 (continued). Progress Level Summary

Content Area	Grade/ Level	Timepoint 1	Timepoint 2	<i>N</i>	Not Yet (%)	Low (%)	Medium (%)	High (%)
ELA	5	Fall	Spring	783	52.36	20.69	15.84	11.11
ELA	6	Fall	Winter	720	58.75	18.06	13.19	10.00
ELA	6	Winter	Spring	3107	57.52	20.66	13.04	8.79
ELA	6	Fall	Spring	367	54.50	17.71	13.62	14.17
ELA	7	Fall	Winter	361	64.82	18.56	9.97	6.65
ELA	7	Winter	Spring	2312	58.82	17.95	13.24	9.99
ELA	7	Fall	Spring	401	55.86	21.45	11.72	10.97
ELA	8	Fall	Winter	419	53.70	15.99	16.71	13.60
ELA	8	Winter	Spring	2380	56.47	19.75	12.94	10.84
ELA	8	Fall	Spring	404	56.93	19.31	14.36	9.41
Literature	HS	Test Event 1	Test Event 2	4864	61.02	17.66	11.23	10.09
Literature	HS	Test Event 2	Test Event 3	152	40.13	19.08	14.47	26.32

Note. A small number of Algebra I and Literature records were excluded from this table in which students records were flagged as duplicates due to taking multiple test events in quick succession.

CHAPTER FIFTEEN: PREDICTING PERFORMANCE ON SUMMATIVE TESTS

One of the main purposes in the design and development of Firefly Student Progress Indicators is to provide a benchmark test that is predictive of the end-of-year PSSA and end-of-course Keystone summative assessments. Many design features and decisions for Firefly were made to support the strongly desired predictive element. Specifically, the test blueprints and reporting category structures for Firefly closely reflect that of PSSA and Keystone. Firefly and PSSA as well as Firefly and Keystone are purported to measure similar constructs, with items aligned to the Pennsylvania Core Standards. Moreover, apart from constructed-response items or writing prompts, Firefly item types for each content area reflect those on PSSA and Keystone. Lastly, there are similar test administration conditions. The online tests are administered using the same test platform, DRC’s INSIGHT, and have the same functionality (highlighter, scrolling features, calculators, etc.). Many elements and features of Firefly were purposefully designed to support the predictability of the benchmark test onto the end-of-year PSSA and end-of-course Keystone assessments.

This chapter summarizes two sets of analyses with respect to predictions: the evaluation of the predictions reported in the current year, and the development of the prediction models developed for the subsequent year. Lastly, this chapter provides information on how one can interpret predictions. In 2024–2025, Firefly was administered in mathematics, Algebra I, ELA, and Literature. Although no predictions were reported in 2024–2025, the data from the 2024–2025 Firefly was used to develop predictions for the 2025–2026 Firefly administration.

EVALUATING 2024–2025 PREDICTIONS

In future administrations, this section will provide results from the analyses and evaluation of the prediction models developed for the current year. However, the academic year of 2024–2025 was the first year of Firefly and no predictions were reported.

2025–2026 PREDICTION MODELING

As discussed, prediction models were developed following the 2024–2025 Firefly administration to support reporting predictions in the 2025–2026 administration. Operational Firefly data was required to develop predictions.

Prediction models were developed for each content area and grade/level separately. In order maximize precision and accuracy in the prediction models, seasonal prediction models were built for each content area and grade/level. As students progress through the academic year and curriculum, there is more opportunity to learn. Therefore, prediction models were defined to best reflect the learning that has occurred thus far in the academic year based on three timepoints. In other words, because student performance on Firefly is expected to improve throughout the academic year, the prediction to the end-of-year or end-of-course summative assessment must also change. To incorporate this progression in student learning, seasonal prediction models were built (see Table 15–1 for the seasonal definitions used for 2024–2025). It is important to note that December 1 and March 1 will continue to be the start of the season for the winter and spring, respectively. However, the starting point (i.e., launch of Firefly) and the ending point (conclusion of Firefly) may change across years.

Table 15–1. Firefly Season Definition

Season	Testing Window
Fall	November 18 to November 30
Winter	December 1 to February 28
Spring	PSSA: March 1 to April 18 Keystone: March 1 to May 9

DATA

Data was gathered from three separate sources: Firefly, PSSA, and Keystone. The final Firefly data consisted of the entire testing window, with data disaggregated by season. The PSSA and Keystone data used is detailed in Chapter Fourteen. For all datasets, the technical reporting criteria was used to refine the datasets for each content area and grade/level separately. Then, data was merged by content area, grade/level, and Pennsylvania student identification number. The details from the data merge and matching are presented in Chapter Fourteen.

Once the datasets were defined (three per content area and grade/level), the demographic and background characteristics were compared between the matched Firefly data and total population data (PSSA and Keystone). To examine representativeness of the matched Firefly data to the total population PSSA and Keystone data, the following variables were considered: gender, race/ethnicity, student group (economically disadvantaged students, English Learners [EL], and students with Individualized Education Plans [IEPs]), and locale (city, suburb, rural, and town). The results of this comparison of key demographics showed that Firefly testers tended to differ from the statewide summative population.

There were several concerns with relying solely on available matched data for development of prediction models. The primary concern is that Black and Hispanic students, students from city settings, and economically disadvantaged students were underrepresented in Firefly participation in 2024–2025. Yet, the prediction models used for the following year will be applied to all Firefly testers. As a voluntary benchmark test, it remains unknown the degree to which districts, schools, and students will participate in Firefly in subsequent years. For this reason, DRC utilized Propensity Score Matching (PSM) to develop more comparable groups to build the prediction models for 2025–2026. Variables used in PSM to develop comparable groups included gender, race/ethnicity, student group (economically disadvantaged, EL, students with IEPs), and locale. Table 15–2 shows the season N counts as well as the N counts after applying PSM for each content area, grade/level, and season. The total summative N count is shown for reference.

Table 15–2. Season Total and Propensity Score Matched N Counts by Content Area and Grade/Level

Content Area	Grade/Level	Fall Total	Fall Matched (PSM)	Winter Total	Winter Matched (PSM)	Spring Total	Spring Matched (PSM)	Summative Spring Population
Mathematics	3	1,917	1,009	17,881	3,804	11,168	3,110	120,647
Mathematics	4	2,495	1,207	16,093	3,522	9,864	2,980	115,897
Mathematics	5	2,596	1,234	17,681	3,471	10,617	2,788	119,130
Mathematics	6	2,116	966	16,346	3,233	10,543	2,947	118,382
Mathematics	7	1,624	901	15,077	3,057	9,059	2,568	117,728
Mathematics	8	1,668	839	13,672	2,976	7,842	2,376	117,781
Algebra I	HS	1,507	876	11,043	3,039	8,731	2,643	116,179
ELA	3	2,016	980	14,440	2,882	9,316	2,519	120,329
ELA	4	2,271	1,131	13,844	3,199	9,003	2,862	115,634
ELA	5	2,501	1,203	15,389	3,388	9,531	2,755	119,192
ELA	6	1,705	843	15,964	3,406	8,207	2,629	118,649
ELA	7	1,846	948	14,808	3,291	6,673	2,331	118,180
ELA	8	1,964	951	14,247	3,280	6,968	2,424	118,375
Literature	HS	819	394	9,672	2,419	5,587	1,867	106,917

Note. The Summative Spring Population represents the N counts PSSA and Keystone, after filtering criteria were applied.

For prediction modeling, the matched records using PSM were weighted to sum the total population.

ANALYSES

Once PSM matched and weighted datasets were developed for each content area, grade/level, and season, the prediction models were developed using isotonic regression (Robertson, Wright, and Dykstra, 1988). PSSA and Keystone scale scores were derived using isotonic regression of the summative test conditional mean score given a Firefly scale score. The summative test conditional mean score $\bar{Y}_i$ is calculated as follows:

$$\bar{Y}_i = E(Y|X = x_i)$$

Where x_i is the Firefly score and Y is the PSSA or Keystone summative scale score. In summary, the isotonic regression finds the fitted value $\hat{z}_i$ that minimizes the sum $\sum_i (Y_i - \hat{Z}_i)^2$ under the order constraints: if $x_i \geq x_j$, then $\hat{z}_i \geq \hat{z}_j$.

Then, the standard error of projection (SEP) is calculated as follows:

$$SEP = \frac{\sqrt{\sum_i (Y_i - \hat{Z}_i)^2}}{N}$$

Where Y_i is the PSSA or Keystone summative scale score and Z_i is the predicted score. The SEP equation represents the square root of the sum of the squared residuals across all observations divided by the number of students.

REPORTING PREDICTIONS

For each Firefly scale score, there is a predicted scale score and associated error. Assuming a normal distribution and using the SEP, probabilities are calculated to reflect the likelihood of each predicted scale score falling into each summative performance level. Then, the predicted performance level is identified as the performance level with the highest probability or likelihood. If two performance levels have the same maximum probability, the lower performance level is used as the predicted performance level. For example, if a student has a 15% likelihood of earning a Below Basic performance level, a 40% likelihood of earning a Basic performance level, a 40% likelihood of earning a Proficient performance level, and a 5% likelihood of earning an Advanced performance level, the student will receive a predicted performance level of Basic. The policy rationale is to err on the side of caution and predict a lower performance level.

There are many ways that predictions can be presented on individual student reports and in the suite of interactive reporting. Several options were considered and discussed. The Standards of Educational and Psychological Assessment (2014) state that “Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use” (page 42). Under this principle, reporting a level of error or precision for the prediction is critical. The policy decision to present the probabilities of earning each performance level was supported by assessment best practices in that it integrates the precision by showing a likelihood of each performance level that was estimated using the standard error of prediction. Refer to Chapter Thirteen for additional information about Firefly scores and score reports.

Additional rules are applied to the probabilities displayed on score reports and in the suite of interactive reporting.

- Probabilities are rounded to the nearest integer and expressed with a percent (%) sign.
- In cases where the rounded calculated probability is 0%, a value of < 1% is displayed.
- In cases where the rounded calculated probability is 100%, > 99% is displayed.

PREDICTION RESULTS

The results from prediction analyses following the 2024–2025 Firefly administration will be utilized the following year (2025–2026). Following each Firefly annual administration, predictions will be evaluated and updated based on prior year’s data. Table 15–3 presents the predicted performance levels on the end-of-year PSSA and end-of-course Keystone exams based on Firefly performance by season. As expected, increasing trends are noted across seasons within each content area and grade/level. These predictions are presented on the 2025–2026 student score reports.

Table 15–3. Firefly Scale Scores and Predicted Performance Level on PSSA and Keystone

Content Area	Grade/Level	Season	Assessment	Below Basic	Basic	Proficient	Advanced
Mathematics	3	Fall	PSSA	180–352	353–402	403–497	498–860
Mathematics	3	Winter	PSSA	180–366	367–419	420–513	514–860
Mathematics	3	Spring	PSSA	180–398	399–443	444–547	548–860
Mathematics	4	Fall	PSSA	200–395	396–462	463–523	524–880
Mathematics	4	Winter	PSSA	200–417	418–474	475–546	547–880
Mathematics	4	Spring	PSSA	200–442	443–513	514–581	582–880
Mathematics	5	Fall	PSSA	220–431	432–512	513–586	587–900
Mathematics	5	Winter	PSSA	220–447	448–518	519–600	601–900
Mathematics	5	Spring	PSSA	220–471	472–545	546–637	638–900
Mathematics	6	Fall	PSSA	240–491	492–565	566–628	629–920
Mathematics	6	Winter	PSSA	240–498	499–567	568–651	652–920
Mathematics	6	Spring	PSSA	240–521	522–596	597–690	691–920
Mathematics	7	Fall	PSSA	260–528	529–594	595–653	654–940
Mathematics	7	Winter	PSSA	260–552	553–600	601–668	669–940
Mathematics	7	Spring	PSSA	260–554	555–606	607–701	702–940
Mathematics	8	Fall	PSSA	280–575	576–612	613–707	708–960
Mathematics	8	Winter	PSSA	280–595	596–657	658–729	730–960
Mathematics	8	Spring	PSSA	280–598	599–667	668–758	759–960
Algebra I	HS	Fall	Keystone	300–544	545–635	636–676	677–980
Algebra I	HS	Winter	Keystone	300–524	525–633	634–695	696–980
Algebra I	HS	Spring	Keystone	300–557	558–658	659–724	725–980
ELA	3	Fall	PSSA	180–311	312–436	437–583	584–860
ELA	3	Winter	PSSA	180–323	324–419	420–569	570–860
ELA	3	Spring	PSSA	180–340	341–447	448–623	624–860
ELA	4	Fall	PSSA	200–363	364–509	510–647	648–880
ELA	4	Winter	PSSA	200–362	363–510	511–640	641–880
ELA	4	Spring	PSSA	200–369	370–522	523–659	660–880
ELA	5	Fall	PSSA	220–427	428–527	528–687	688–900

Table 15–3 (continued). Firefly Scale Scores and Predicted Performance Level on PSSA and Keystone

Content Area	Grade/Level	Season	Assessment	Below Basic	Basic	Proficient	Advanced
ELA	5	Winter	PSSA	220–427	428–555	556–741	742–900
ELA	5	Spring	PSSA	220–418	419–548	549–740	741–900
ELA	6	Fall	PSSA	240–240	241–520	521–677	678–920
ELA	6	Winter	PSSA	240–343	344–537	538–693	694–920
ELA	6	Spring	PSSA	240–354	355–549	550–688	689–920
ELA	7	Fall	PSSA	260–368	369–563	564–657	658–940
ELA	7	Winter	PSSA	260–304	305–573	574–712	713–940
ELA	7	Spring	PSSA	260–260	261–566	567–716	717–940
ELA	8	Fall	PSSA	280–401	402–581	582–706	707–960
ELA	8	Winter	PSSA	280–437	438–586	587–715	716–960
ELA	8	Spring	PSSA	280–331	332–589	590–730	731–960
Literature	HS	Fall	Keystone	300–438	439–572	573–800	801–980
Literature	HS	Winter	Keystone	300–442	443–611	612–822	823–980
Literature	HS	Spring	Keystone	300–413	414–569	570–828	829–980

INTERPRETING PREDICTIONS

The information presented in the previous sections can help support interpretations about predictions. For each Firefly assessment, the prediction reflects the expected performance level on the spring summative test (PSSA or Keystone) assuming the student maintains the same learning trajectory. For example, DRC used student performance on Firefly between December 1, 2024 and February 28, 2025 along with PSSA performance in spring 2025 to develop the Winter prediction model. As such, the prediction reflects expected performance in the **spring**. The prediction does **not** reflect how the student would perform on the summative assessment if they were to test the same day.

Across the academic year, predictions will change based on the season of administration. If a student earns the same Firefly scale score in fall and winter, one can conclude that the student did not show progress between the two seasons. One would also observe that the same earned scale score on Firefly did not map to the same predicted performance level. For example, a student who earns a scale score of 605 on Firefly mathematics grade 7 in fall, winter, and spring would be predicted to earn a Proficient performance level on PSSA based on the fall administration, but a Basic performance level based on their winter and spring administration (refer to Table 15–3). For most content areas and levels, the earned Firefly scale score must increase to maintain the same predicted performance level across seasons. This trend is expected because students must demonstrate increased mastery of knowledge and skills throughout the year.

In addition to providing a single predicted performance level, the probability that a student will earn each performance level is also presented. These probabilities reflect a percent likelihood and integrate the precision of the prediction. For example, a student may show probabilities of 5%, 40%, 45%, and 10% for performance levels of Below Basic, Basic, Proficient, and Advanced, respectively. The student is predicted to earn a performance level of Proficient with the maximum probability of 45%. However, one should also consider that the likelihood of earning a Basic performance level is very close (within 5%) to the likelihood of a Proficient performance level. The student is almost as likely to earn a Basic performance level as a Proficient. Therefore, when students, teachers, or parents interpret the predicted performance level, they should also review and consider the probabilities of adjacent performance levels.

KEYSTONE PREDICTIONS

There are additional considerations for predictions for the high-school Keystone Exams. As previously discussed, the data used to develop the prediction models for Algebra I and Literature Keystone Exams included first-time testers only from the spring Keystone administration. The data may have included students who were on alternative or block scheduling if they took Keystone in the spring; however, the data did not include first-time testers from the winter or summer Keystone administrations.

The applicability of predictions for students on block scheduling who plan to take the Winter or Spring Keystone is not yet known. One should exercise caution in utilizing or overinterpreting predictions for students who are on block scheduling. For example, a student on a block schedule who takes a corresponding Firefly test multiple times in the fall season will have the same prediction model applied to each attempt. If the student's score improves within the season (which would be expected), the prediction could overestimate the student's preparedness for the spring assessment. The higher score would cause the prediction model to assign a higher success probability based on student expectations for a full-year course (i.e., the amount of content instruction), which is not necessarily representative of expectations for students on a block schedule. The accuracy of predictions for students on alternate schedules will be evaluated and analyzed once scheduling data is available following the first year of predictions.

Moreover, repeat testers also tend to have different learning trajectories. The predictions were based on data from students who progress in a typical scope and sequence of curriculum (learning trajectory); as such, retesters were excluded from prediction model development because there is too much variation in when they were first enrolled in the associated Keystone course, the time elapsed between first Keystone exam and subsequent retest opportunities, and whether they were enrolled previously or currently in a remedial class. By selecting only first-time testers in the development of the prediction model, the variance and the error in the model was controlled and minimized.

CHAPTER SIXTEEN: RELIABILITY

This chapter addresses the reliability of Firefly test scores. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), the general notion of reliability/precision refers to

the consistency of scores across replications of a testing procedure, regardless of how this consistency is estimated or reported (p.33).

Frisbie (2005) highlighted several elements of reliability. First, reliability is a property of test scores, not of a test itself. Many may appreciate this distinction, but in casual usage, individuals frequently make reference to a “reliable test.” While reliability concerns test scores (and not the test specifically), it’s important to appreciate the fact that test scores can be affected by characteristics of the instrument. For example, all other things being equal, tests with more items/points tend to produce more reliable scores than tests with fewer items/points. Second, reliability coefficients are group specific. Reliabilities tend to be higher in populations that are more heterogeneous and lower in populations that are more homogeneous. Consequently, both test length and population heterogeneity should be considered when evaluating reliability.

There are other reliability considerations that may be less evident from the *Standards’* definition yet are still important for test users to understand. While freedom from measurement error is very important, reliability is specifically concerned with random sources of error. Indeed, the degree of inconsistency due to random error sources is what determines reliability: less consistency is associated with lower reliability and more consistency is associated with higher reliability. Of course, systematic error sources also exist. These can artificially increase reliability and decrease validity. Validity is further discussed in Chapter Seventeen.

Another noteworthy issue is that multiple sources of error exist (e.g., the day of testing, the items administered). However, most widely used reliability indices only reflect a single type of error. Consequently, it is important for test users to understand what specific type of error is being considered in a reliability study, and equally, if not more importantly, what types are not.

Understanding the distinction between relative error and absolute error is also important, as many reliability indices only reflect relative error. Relative error is of interest whenever the relative ordering of individuals with respect to their test performance is of interest. Understanding examinee rank-order stability is important; however, such stability might be well achieved even when the specific score values are considerably different. When specific score values are considered important (e.g., if cut scores are used), then absolute error is of interest, too. Generally, there is more error variance when considering the absolute scores of examinees, which, in turn, suggests lower reliability.

As the above discussion suggests, reliability is a complex, nonunitary notion that cannot be adequately represented by a single number. There are several reliability indices available, and these may not provide the same results (Frisbie, 2005). The remainder of this chapter covers the following:

- Unconditional and conditional standard errors of measurement (SEMs and CSEMs)
- Reliability coefficients and their interpretation
- Decision consistency

STANDARD ERROR OF MEASUREMENT

The standard error of measurement (SEM) is an important indicator of test score precision that can also support and help users understand and interpret test scores. The SEM is well-suited for determining the effect of measurement inconsistencies on the scores obtained by individual examinees. There are two types of standard error discussed in this section: traditional SEM and conditional SEM.

TRADITIONAL STANDARD ERROR OF MEASUREMENT

A precise, theoretical interpretation of the SEM is somewhat unwieldy. A beginning point for understanding the concept is as follows. If everyone being tested had the same true score¹, there would still be some variation in observed scores due to imperfections in the measurement process, such as random differences in attention during instruction or concentration during testing, the sampling of test items, etc. The standard error is defined as the standard deviation² of the distribution of observed scores for students with identical true scores. Because the SEM is an index of the random variability in test scores in actual score units, it represents very important information for test score users.

Traditional SEM depends on both the reliability coefficient (discussed in the following section) and the standard deviation of test scores. If the reliability were equal to 0.00 (the lowest possible value), the SEM would be equal to the standard deviation of the test scores. If test reliability were equal to 1.00 (the highest possible value), the SEM would be 0.0. In other words, a perfectly reliable test has no measurement error (Harvill, 1991). Additionally, the value of the SEM takes the group variation (i.e., score standard deviation) into account.

The SEM is an index of the random variability in test scores in actual score units, which is why it has such great utility for test score users. SEMs allow statements regarding the precision of individual tests scores. SEMs help place reasonable limits (Gulliksen, 1950) around observed scores through construction of an approximate score band. Often referred to as confidence intervals, these bands are constructed by taking the observed scores, X , and adding and subtracting a multiplicative factor of the SEM. As an example, students with a given true score will have observed scores that fall between ± 1 SEM about two-thirds of the time.³ For ± 2 SEM confidence intervals, the percentage increases to about 95 percent.

FURTHER INTERPRETATIONS

ONE SEM FOR ALL TEST SCORES

The SEM approach described above provides a single numerical estimate for constructing the confidence intervals for examinees regardless of their score levels. However, in reality, such confidence intervals vary according to one's score, especially within item response theory or Rasch psychometric frameworks. Consequently, care should be taken when using the traditional SEM for students with extreme scores. An alternate approach is described in the next section that conditions the SEM on a student's score estimate.

GROUP SPECIFIC

As noted in the introduction, reliabilities are group specific. The same is true for SEMs because both score reliabilities and score standard deviations vary across groups.

SCALE SCORE METRIC

The SEM approach is calculated using scale scores, and as such, the resulting confidence interval bands are in the scale score metric.

TYPE OF ERROR REFLECTED

The interpretation of the SEM should be driven by the type of score reliability that underpins it. So, the Firefly SEMs involve the same source of error relevant to internal consistency indices. As noted earlier, a precise technical explanation of the SEM (and resulting confidence intervals) can be unwieldy. Because of this, score users are often provided with less complex interpretations.

One simpler description sometimes used is that a confidence interval represents the possible score range that one would observe if a student could be tested twice with the same instrument. Taking the same test on a different day implies the only source of random error being considered is related to the occasion of testing—such as a student might be sleepier one day than another, might be sick, or might not have eaten a good breakfast. There is a reliability index that captures this source of random error and it is referred to as the test-retest reliability coefficient.

¹ True score is the score the person would receive if the measurement process were perfect.

² The standard deviation of a distribution is a measure of the dispersion of the observations. For the normal distribution, about 16 percent of the observations are more than one standard deviation above the mean.

³ Some prefer the following interpretation: if a student were tested an infinite number of times, the ± 1 SEM confidence intervals constructed for each score would capture the student's true score 68 percent of the time.

This is not the type of reliability computed for Firefly. When internal consistency reliability estimates are used, such an explanation blurs the fact that random error based on the occasion of testing is not considered.

When SEMs are derived from internal consistency reliability estimates, a better approach is to describe the confidence interval as providing reasonable bounds for the range of scores that a student might receive if he or she took an equivalent version of the test. That is, the student took a test that covered exactly the same content but included a different set of items. As an example, if the Algebra I score was 725 and the SEM band was 701 to 749 based on 1 SEM, then a student would be likely to receive a score somewhere between 701 and 749 if he or she took a different version of the test without additional instruction.

RASCH CONDITIONAL STANDARD ERRORS OF MEASUREMENT

The conditional SEM (CSEM) also indicates the degree of measurement error in scale score units but varies as a function of a student's earned scale score. Therefore, the CSEM may be especially useful in characterizing measurement precision in the neighborhood of a score level used for decision-making—such as cut scores for identifying students who meet a performance standard.

Technically, when a Rasch model is applied, the CSEM at any given point on the ability continuum is defined as the reciprocal of the square root of the test information function derived from the Rasch scaling model:

$$CSEM(\hat{\beta}_n) = \frac{1}{\sqrt{I(\hat{\beta}_n)}}$$

where $CSEM(\hat{\beta}_n)$ is the conditional standard error of measurement and $I(\hat{\beta}_n)$ is the test information function. Test information depends on the sum of the corresponding information functions for the test items. Item information depends on each item's difficulty and conditional item score variance. The formula above utilizes the Rasch ability ($\hat{\beta}_n$) metric. The conditional standard error on the scale score (SS) metric is determined simply by multiplying the $CSEM(\hat{\beta}_n)$ by the slope (multiplicative constant, m) of the linear transformation equation used to convert the Rasch ability estimates to scale scores:

$$CSEM(SS) = CSEM(\hat{\beta}_n) * m$$

Chapter Nine provides the linear transformation formulas for each Firefly content area.

RASCH CSEM CONFIDENCE INTERVALS

Conditional Standard Errors of Measurement also allow statements regarding the precision of individual test scores. And like SEMs, they help place reasonable limits around observed scale scores through construction of an approximate score band. The confidence intervals are constructed by adding and subtracting a multiplicative factor of the CSEM and may be interpreted as described in the earlier section.

FURTHER INTERPRETATIONS

DIFFERENT CSEMS FOR DIFFERENT TEST SCORES

The CSEM approach provides different numerical estimates for constructing the confidence intervals for examinees depending on their specific score. On fixed form tests, the magnitude of the CSEM values is often “U” shaped, with larger CSEM values associated with extremely lower or higher scores. With a fixed set of items, there is less information for students scoring at the extremes, and CSEM is inversely related to the information function (the more information, the lower the CSEM). When there is a deep and wide item pool for a CAT, the “U” shape may be less pronounced as students are presented with items targeted at their level. For newly developed and emerging CAT item pools that cover a smaller range of difficulties, there may not be sufficient items with extreme high or extreme low difficulties, so the “U” shape may persist. However, one of the goals for CAT programs is to have a deep item pool that covers the range of item difficulties in order to be able to select and administer items tailored to student abilities. When this is the case, adaptive tests allow for greater information and, therefore, lower CSEMs across a wider range of the vertical scale.

GROUP SPECIFIC

Assuming reasonable model-data fit—as explored in Chapter Seven—the Rasch-based CSEMs (conditioned on score level) should not vary across groups.

SCALE SCORE METRIC

The CSEM and associated confidence interval bands are in the scale score metric.

TYPE OF ERROR REFLECTED

The CSEMs reported in the reporting suite are based on the Rasch-based conditional standard errors of measurement described above. Score report content is considered in greater detail in Chapter Twelve.

RESULTS AND OBSERVATIONS

Chapter Fourteen provides descriptive statistics of the CSEM. Figure 16–1 displays the median CSEMs across the range of scale scores based on operational data from the 2024–2025 school year. Figure 16–1 shows the median CSEM across the scale score range for each content area and grade/level. The vertical black dotted lines represent the cut scores for Near Target and Prepared performance-level classifications. Each grade/level has a different lowest obtainable scale score and highest obtainable scale score. However, the scale score range is constant across all grades/levels for consistency of comparisons. For most Firefly tests, the minimum SEM value is close to the Near Target cut score, which is ideal. As the operational Firefly item pools grow, it is expected that the typical “U” shape may decrease in magnitude as more items are incorporated in the Prepared performance-level ranges.

Figure 16–1. Median Standard Error of Measurement by Content Area and Grade/Level

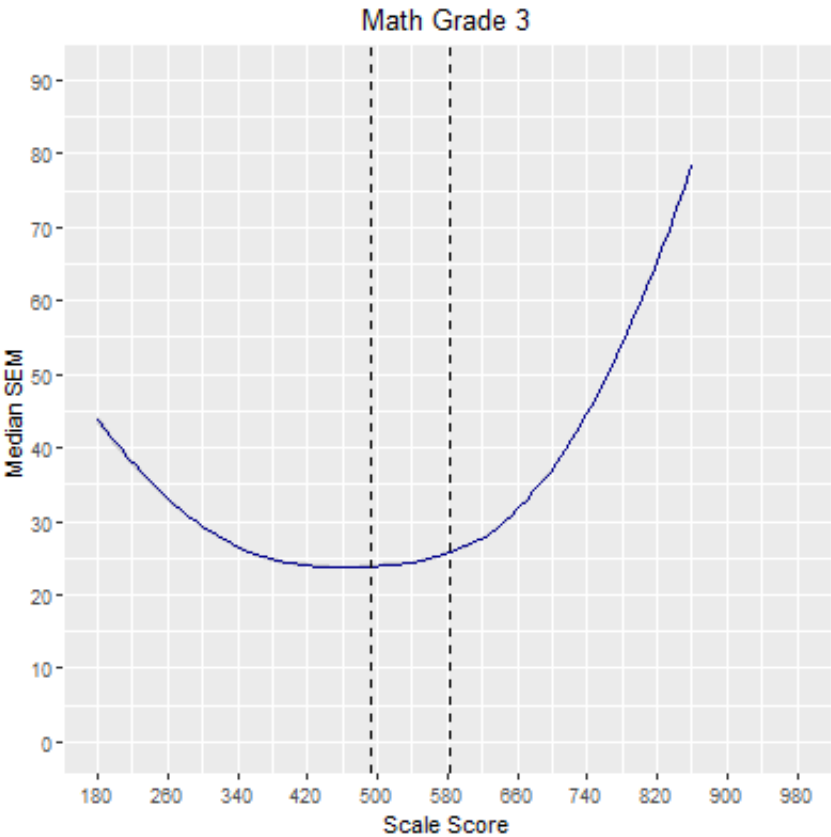


Figure 16–1 (continued). Median Standard Error of Measurement by Content Area and Grade/Level

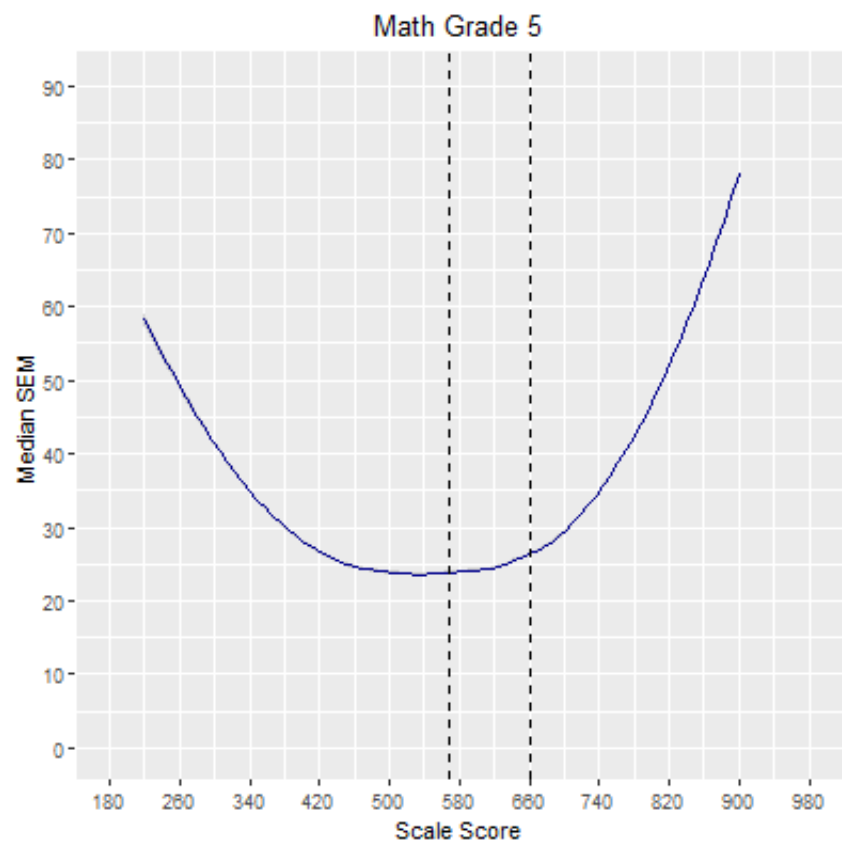
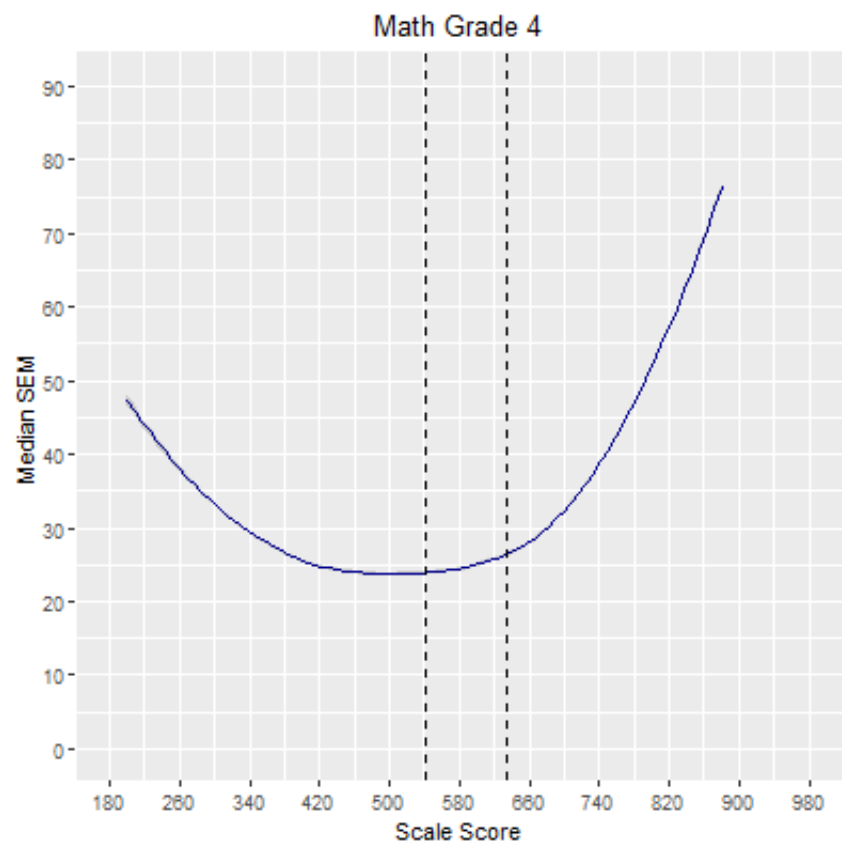


Figure 16–1 (continued). Median Standard Error of Measurement by Content Area and Grade/Level

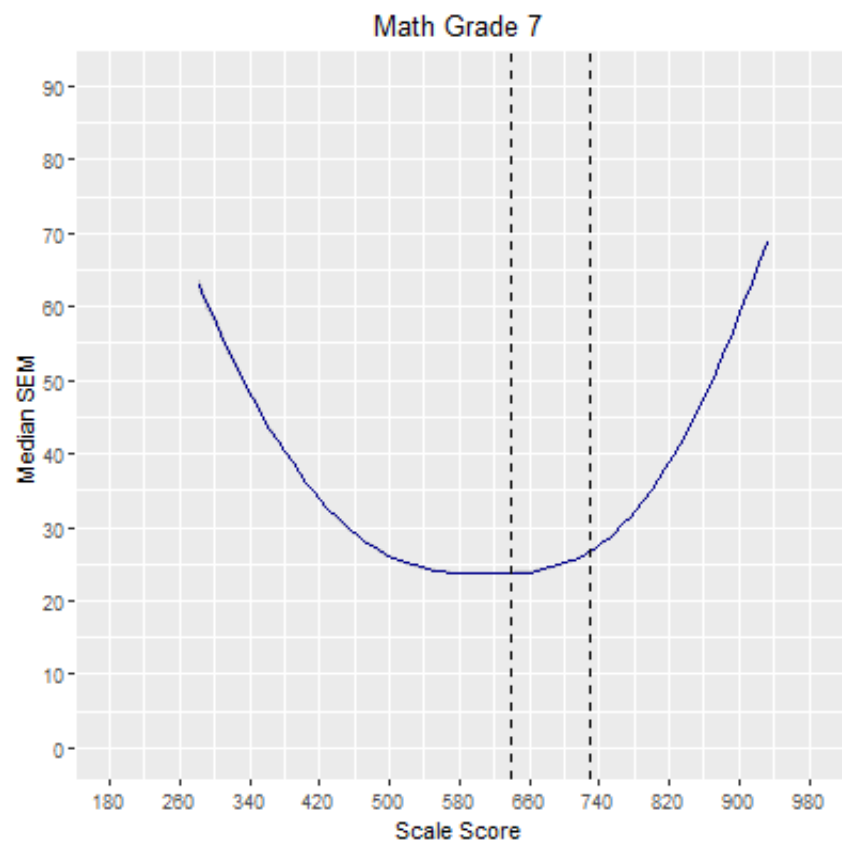
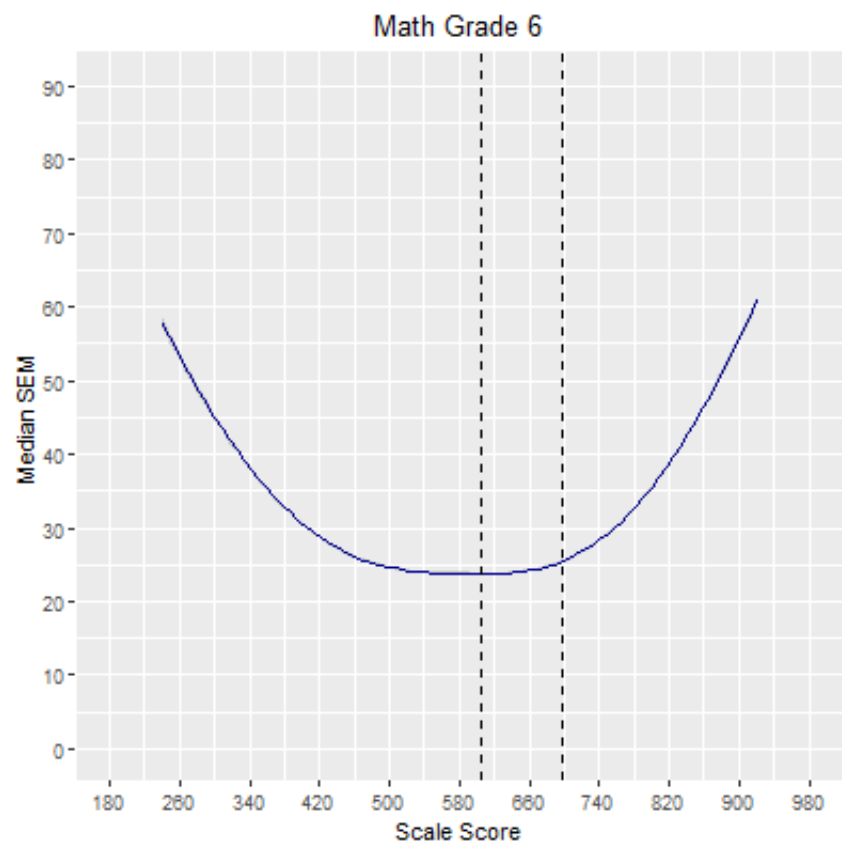


Figure 16–1 (continued). Median Standard Error of Measurement by Content Area and Grade/Level

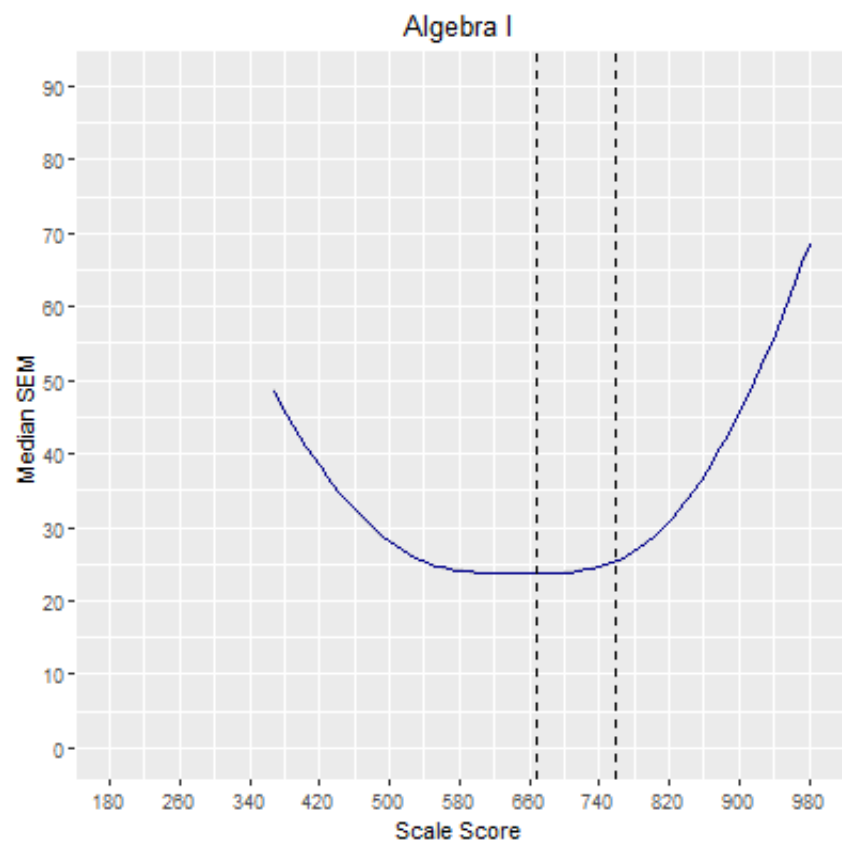
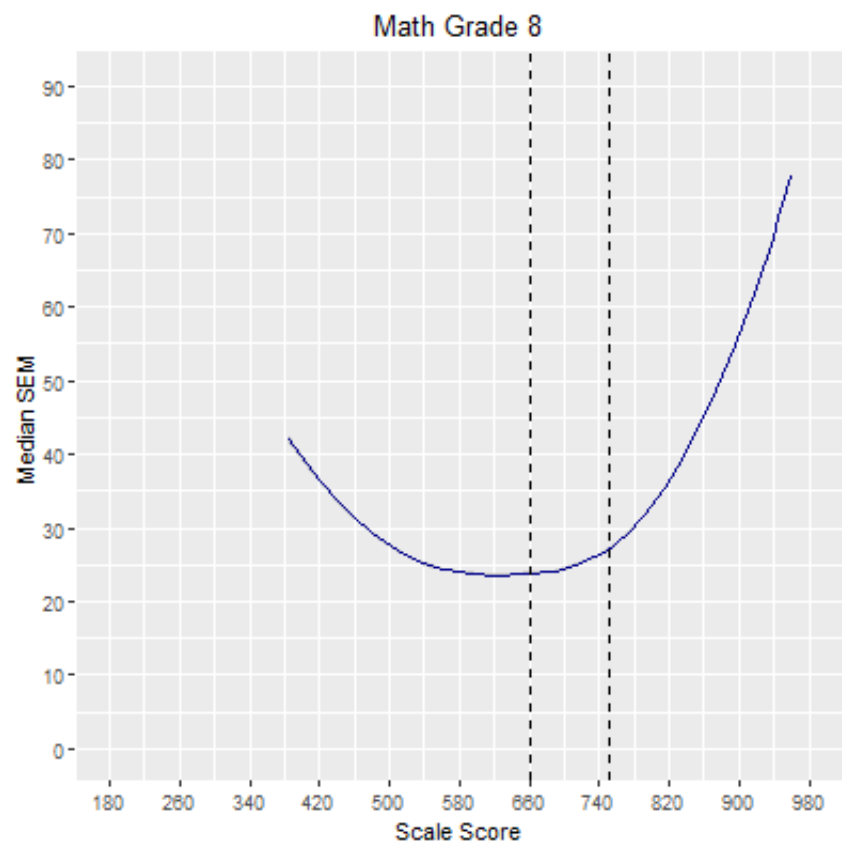


Figure 16–1 (continued). Median Standard Error of Measurement by Content Area and Grade/Level

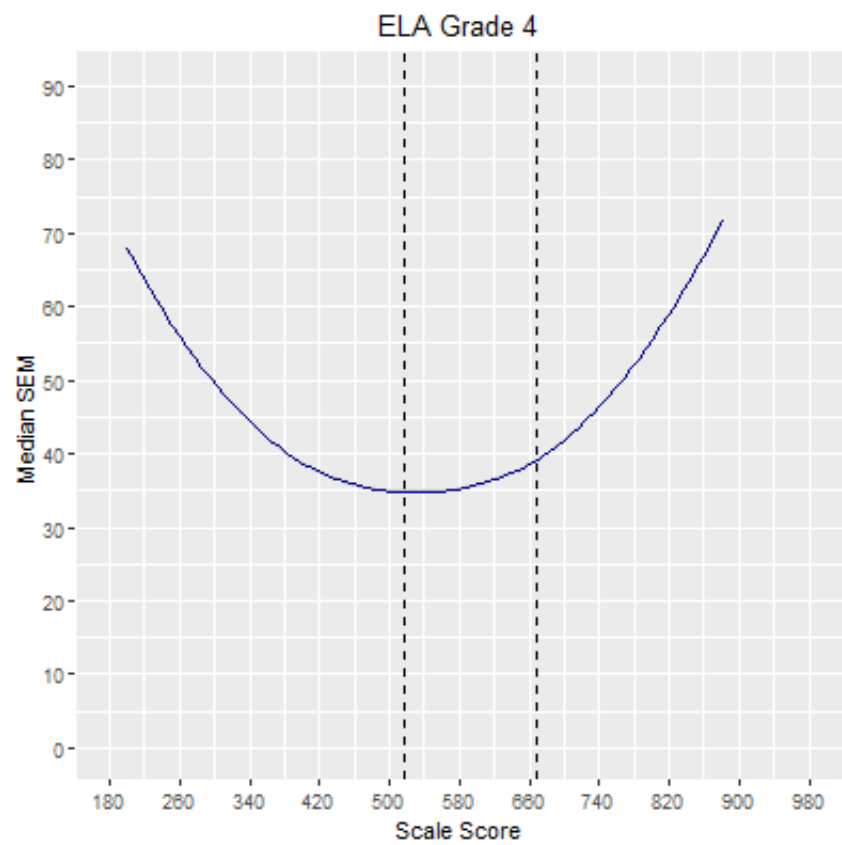
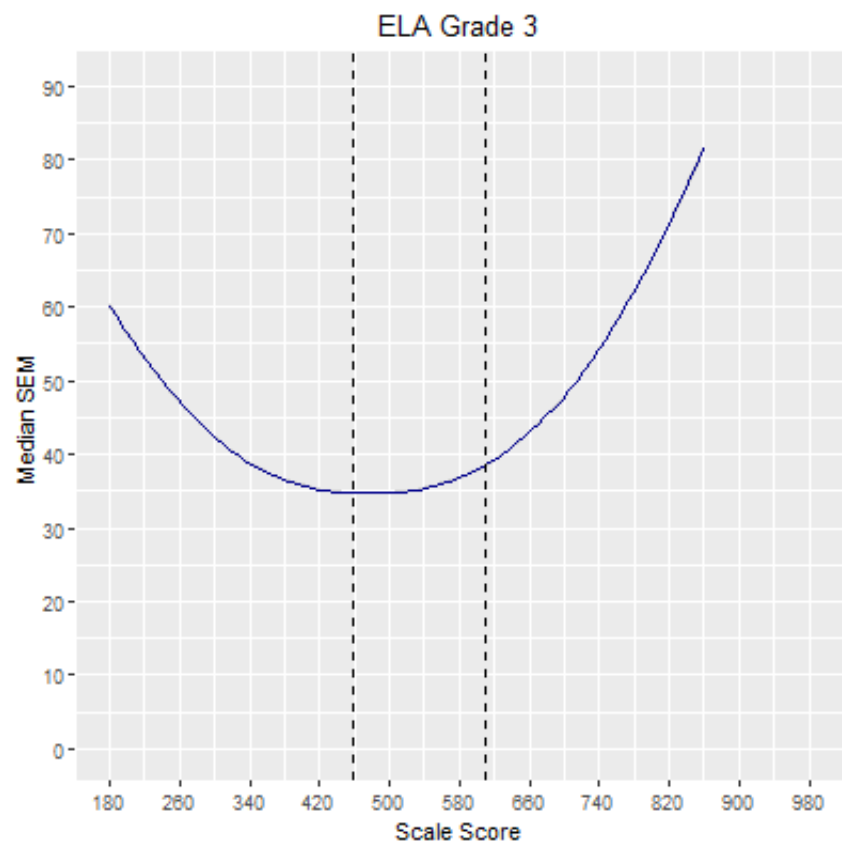


Figure 16–1 (continued). Median Standard Error of Measurement by Content Area and Grade/Level

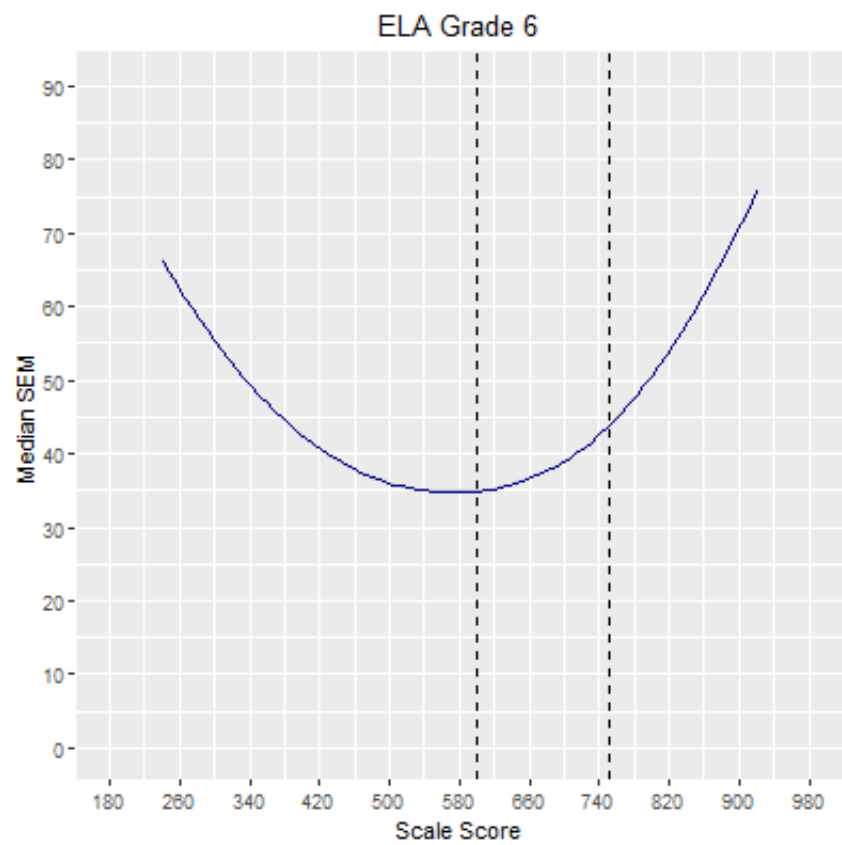
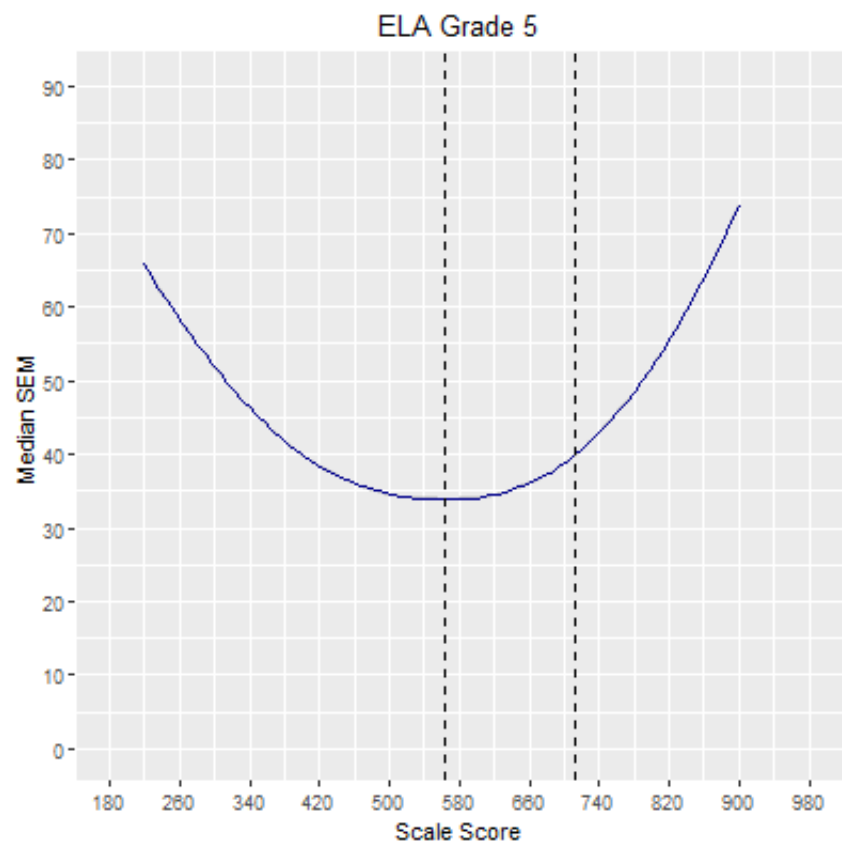


Figure 16–1 (continued). Median Standard Error of Measurement by Content Area and Grade/Level

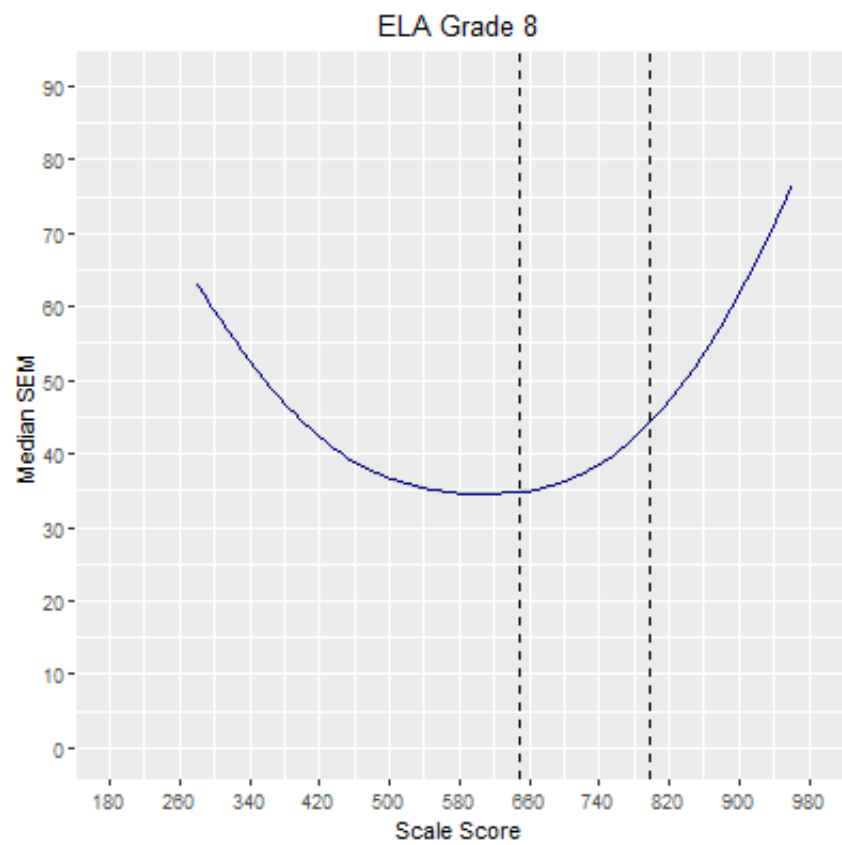
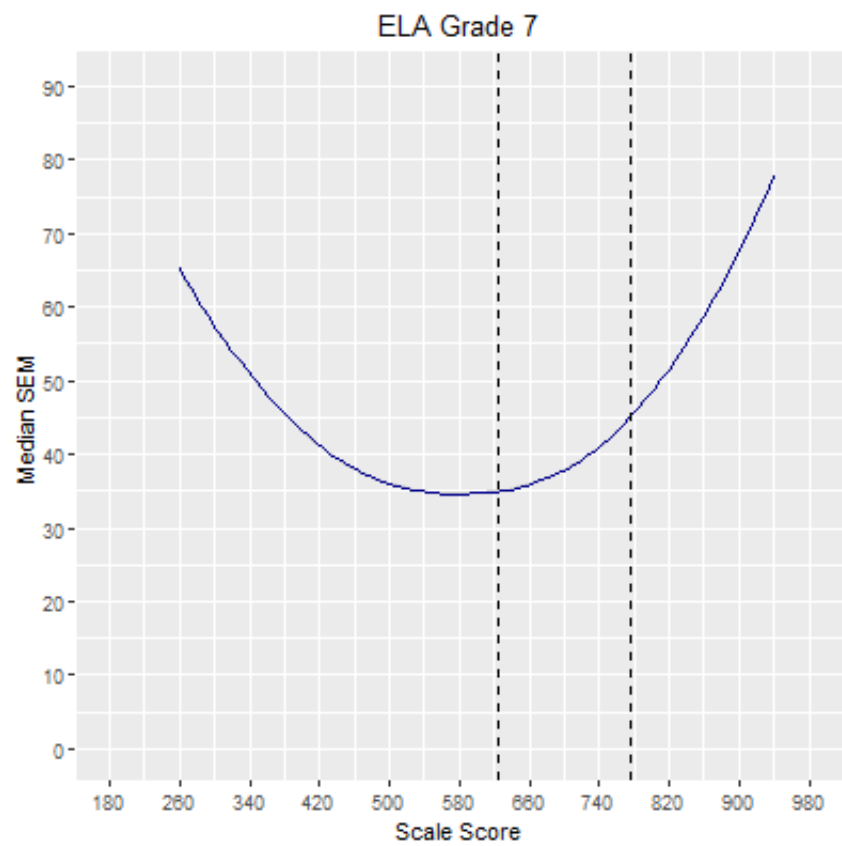
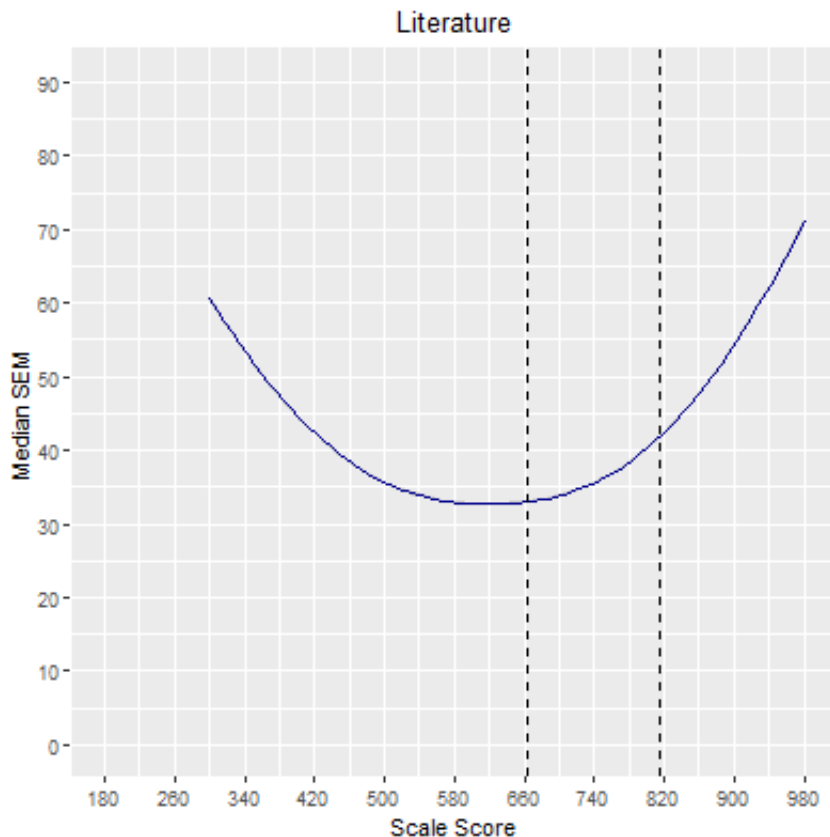


Figure 16–1 (continued). Median Standard Error of Measurement by Content Area and Grade/Level



RELIABILITY INDICES

As shown below, the reliability coefficient expresses the consistency of test scores as the ratio of true score variance to total score variance. The total variance contains two components: 1) variance in true scores and 2) variance due to the imperfections in the measurement process. Put differently, total variance equals true score variance plus error variance.⁴

$$\rho_x^2 = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

Reliability coefficients indicate the degree to which differences in test scores reflect true differences in the attribute being tested rather than random fluctuations. Total test score variance (i.e., individual differences) is partly due to real differences in the attribute (true variance) and partly due to random error in the measurement process (error variance).

Reliability coefficients can range from 0.0 to 1.0. If all test score variance were true, the index would equal 1.0. The index would be 0.0 if none of the test score variance were true. Such scores would be pure random noise (i.e., all measurement error). If the index had a value of 1.0, scores would be perfectly consistent (i.e., contain no measurement error). Although values of 1.0 are never achieved in practice, it is clear that larger coefficients are more desirable, as they indicate that test scores are less influenced by random error. “How big is big enough?” and “how small is too small?” are issues considered in a later section.

As noted earlier in this chapter, there are several different indices that can be used to estimate this ratio. One approach is referred to as internal consistency, which is derived from analyzing the performance consistency of individuals over the items within a test. As discussed below, these internal consistency indices do not take into account other sources of error, such as day-to-day variations (student health, testing environment, etc.).

⁴ A covariance term is not required, as true scores and error are assumed to be uncorrelated in classical test theory.

COEFFICIENT ALPHA

Although a number of reliability indices exist, one of the most frequently reported for achievement tests is coefficient alpha. For example, both PSSA and Keystone programs report coefficient alpha.

FORMULA FOR ALPHA

Consider the following data matrix representing the scores of persons (rows) on items (columns):

Table 16–1. Person × Item Score (X_{pi}) Infinite (Population-Universe) Matrix

Person	Item 1	Item 2	...	Item i	...	Item k
Person 1	Y_{11}	Y_{12}	...	Y_{1i}	...	X_{1k}
Person 2	Y_{21}	Y_{22}	...	Y_{2i}	...	X_{2k}
....	...	...	...	...	...	...
Person p	Y_{p1}	Y_{p2}	...	Y_{pi}	...	X_{pk}
....	...	...	...	...	...	...
Person N	Y_{N1}	Y_{N2}	...	Y_{Ni}	...	X_{Nk}

Note. Adapted from Cronbach and Shavelson (2004).

Then, a general computational formula for alpha is as follows:

$$\alpha = \frac{N}{N-1} \left(1 - \frac{\sum_{i=1}^N \sigma_{Y_i}^2}{\sigma_X^2} \right),$$

where N is the number of parts (items or testlets), σ_X^2 is the variance of the observed total test scores, and $\sigma_{Y_i}^2$ is the variance of part i .

Examination of the formula for alpha indicates why the coefficient is not appropriate for computer adaptive tests like Firefly and CDT. In Firefly, each student takes a unique set of test items rather than the same fixed form. An item score matrix for Firefly analogous to Table 16–1 would include all items in the operational item pool (up to 500), in which each student is administered only a small subset of items (36–45) from the available pool. Summing the variance of more than 500 item scores and dividing by the variance of test scores based on 36–45 items is not appropriate. Therefore, a measure of reliability other than coefficient alpha is required for Firefly.

MARGINAL RELIABILITY

Marginal reliability indices are often reported in computer adaptive tests because they can be used when students take different forms (i.e., sets of items). In computer adaptive tests, the set of administered items differs for each administration and is tailored to examinee ability; a CSEM is estimated for each test event. Using the known relationship between traditional standard error and reliability, one can infer reliability given the known SEM of a test.

$$SEM = sd_x \sqrt{1 - r}$$

Where sd_x represents the standard deviation of scale scores, and r represents the estimated reliability index.

Using this relationship, one can estimate marginal reliability using the mean squared error (MSE) in place of the SEM of test scores.

$$r = 1 - \frac{MSE}{sd_x^2}$$

Where MSE is calculated as

$$MSE = \frac{\sum_{i=1}^N SEM(\theta_i)}{N}$$

Marginal reliability is estimated for each Firefly content area and grade level as a measure of internal consistency. Marginal reliability can be interpreted in the same way as other reliability indices discussed.

FURTHER INTERPRETATIONS

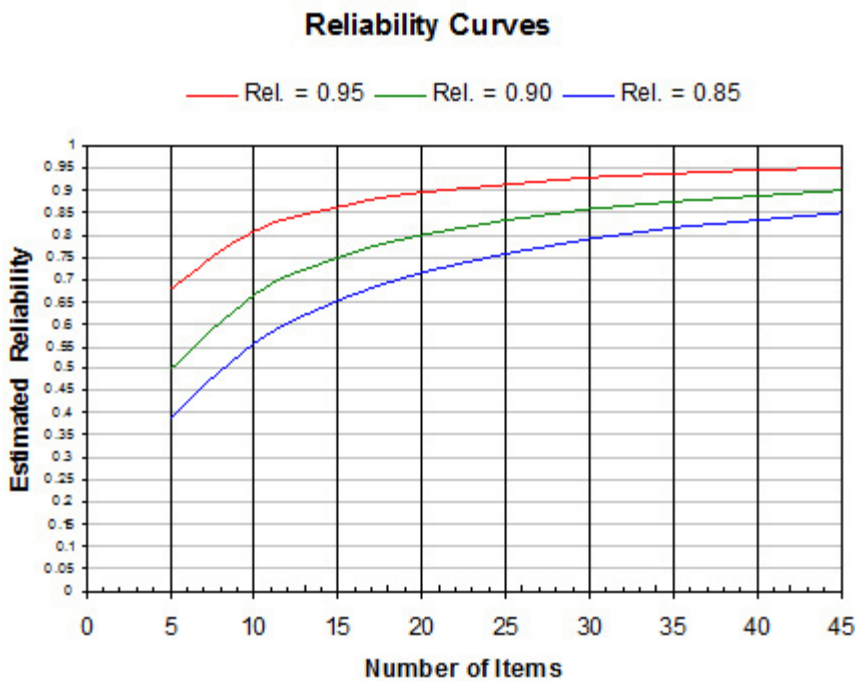
What reliability value is considered high enough? What values are considered too low? Although frequently asked for, any rules of thumb for interpreting the magnitude of reliability indices are mostly arbitrary. One approach is to research the reliabilities from similar testing instruments to see what values are commonly observed, keeping in mind that higher reliability indices are often associated with longer tests. For 2025 PSSA tests in Mathematics and English Language Arts (ELA), reliability coefficients ranged from 0.89 to 0.93. For Spring 2025 Keystone exams in Algebra I and Literature reliability coefficients were 0.92. For CDT administered during the 2024–2025 academic year, the reliability coefficients ranged from 0.93 to 0.94 for mathematics and reading. For many other state assessment programs, reliabilities in the low 0.90s are usually the highest observed, and reliabilities in the high 0.80s are very common.

The lower a given reliability coefficient, the greater the potential for over-interpretation of the associated results. As suggested earlier, there is no firm guideline regarding how low is too low. However, as an informative point of reference, a reliability coefficient of 0.50 would mean that there is as much error variance as true-score variance in the scores.

REPORTING CATEGORY SCORE RELIABILITY

As noted in the introduction, reliabilities tend to be higher with an increase in test length and lower with a decrease in test length. Figure 16–2 illustrates this relationship for a hypothetical 45-item test with three total score reliabilities: 0.95, 0.90, and 0.85. As an example, the curve for reliability equal to 0.90 suggests that a 10-item diagnostic category score would be expected to have a score reliability of just over 0.65. The use of the Spearman-Brown prediction formula assumes all items are exchangeable, which, in practice, they may not be. While such a chart may not perfectly model actual diagnostic category reliability, the intent is to illustrate the substantial impact that limited numbers of items can have on reporting-category score reliability.

Figure 16–2. Example of the Relationship between Test Length and Reliability.



RESULTS AND OBSERVATIONS

Table 16–2 presents information about the reliability of Firefly based on total scores, by content area and grade/level. The mean CSEM and marginal reliabilities are presented for each content area and grade/level. For Firefly mathematics, tests have a fixed number of items and points, whereas Firefly ELA and HS Literature range in length and points (see Table 16–2). As such, information about the minimum, maximum, and mean number of operational items administered is presented. Values were derived using the operational data from the 2024–2025 academic year.

As shown in the table, all reliability coefficients support the internal consistency of Firefly tests. The Firefly mathematics reliabilities range from 0.90 to 0.92 and tend to be slightly higher than ELA and Literature, which range from 0.85 to 0.90.

Table 16–2. Firefly Reliability Coefficients

Content Area	Grade/Level	<i>N</i> Students	<i>N</i> Operational Items Min	<i>N</i> Operational Items Max	<i>N</i> Operational Items Mean	Mean Scale Score	SD Scale Score	Mean CSEM	Reliability
Mathematics	3	31621	36	36	36.00	463.61	95.28	25.48	0.93
Mathematics	4	29043	36	36	36.00	498.49	89.41	25.35	0.92
Mathematics	5	31646	36	36	36.00	532.36	89.03	25.53	0.92
Mathematics	6	29745	36	36	36.00	562.91	82.57	25.00	0.91
Mathematics	7	26581	36	36	36.00	585.58	81.86	25.31	0.90
Mathematics	8	24033	36	36	36.00	611.51	84.60	25.51	0.91
Algebra I	HS	26927	36	36	36.00	619.44	84.32	25.06	0.91
ELA	3	26345	28	32	31.02	431.44	98.73	37.47	0.85
ELA	4	25689	28	32	30.73	492.63	110.98	38.24	0.88
ELA	5	28077	28	32	31.17	536.95	109.45	37.09	0.88
ELA	6	26547	28	32	30.17	548.85	112.24	38.28	0.88
ELA	7	24015	28	32	29.68	565.43	111.48	37.85	0.88
ELA	8	23941	28	32	30.28	583.24	112.56	37.53	0.89
Literature	HS	23820	31	38	34.85	621.03	109.67	35.33	0.90

Reliability can also be estimated for reporting category subscores. As previously discussed, it is expected that the reliability of subscores is lower than that of the total, given that standard errors for a shorter test tend to be lower than standard errors of a longer test. When SE and reliability are based on a smaller number of items, items, such as in subscore reporting, the reliability tends to be lower. Appendix D provides the reliabilities by reporting category subscore.

Calculating reliability coefficients by student group provides evidence that reliability of scores for different groups of students are comparable. Appendix E presents reliabilities by student group based on gender, race/ethnicity, student group (economically disadvantaged students, English Learners, and students with Individualized Education Plans [IEPs]). As shown in the appendix, reliabilities across different groups are comparable.

DECISION CONSISTENCY

Classification decision consistency refers to the degree to which the performance level for each student can be replicated upon retesting using an equivalent form (Huynh, 1976). While Firefly is designed to be administered multiple times in the school year to gauge progress following instruction, retesting in the context of decision consistency refers to retesting shortly after testing without additional instruction.

In criterion-referenced testing programs, there is great interest in knowing how accurately students are classified into performance levels. In contrast to reliability, which is concerned with the relative rank-ordering of students, it is the absolute values of student scores that are important in decision consistency.

Decision consistency answers the question “What is the agreement between performance level classifications based on two non-overlapping, equally difficult forms of the test?” If two parallel forms of the test were given to the same students without additional instruction, the consistency of the measure would be reflected by the extent to which the classification decisions made based on the first set of test scores matched the decisions based on the second set of test scores. Consider Table 16–3:

Table 16–3. Pseudo-Decision Table for Three Hypothetical Categories

Test Level	Test One – Level I	Test One – Level II	Test One – Level III	Test One – Marginal
Test Two – Level I	ϕ_{11}	ϕ_{12}	ϕ_{13}	$\phi_{1\bullet}$
Test Two – Level II	ϕ_{21}	ϕ_{22}	ϕ_{23}	$\phi_{2\bullet}$
Test Two – Level III	ϕ_{31}	ϕ_{32}	ϕ_{33}	$\phi_{3\bullet}$
Test Two – Marginal	$\phi_{\bullet 1}$	$\phi_{\bullet 2}$	$\phi_{\bullet 3}$	1

If a student is classified as in one category based on Test One’s score, how probable would it be that the student would be reclassified in the same category if he or she took Test Two (a non-overlapping, equally difficult form of the test)?

The proportions of correct decisions, ϕ , for three categories is computed as:

$$\tilde{O} = \phi_{11} + \phi_{22} + \phi_{33}$$

The sum of the diagonal entries, the proportion of students classified into the same performance level, signifies the overall decision consistency.

Since it is not feasible to repeat Firefly tests one right after the other with no additional instruction in order to estimate the proportion of students who would be reclassified in the same performance levels, a statistical model can be used to project the consistency of classifications solely using data from the available administration (Hambleton and Novick, 1973). Two well-known methods were developed by Hanson and Brennan (1990) and Livingston and Lewis (1995) utilizing specific true-score models. While both measures are reported for PSSA and Keystone Exams, the statistical models imposed on the data depend upon a beta binomial distribution of raw scores. Given that Firefly is adaptive (i.e., raw scores using a response probability of 0.5 are generally equal to one-half of test length), these measures are not appropriate for Firefly. Instead, decision consistency measures in this section are a Rasch-based index that relies on conditional standard errors (CSEMs). In addition, Cohen’s Kappa is also calculated (Cohen, 1960). The same procedure is used for calculating decision consistency for CDT.

The decision consistency measures reported in the section are based on the Rasch model and conditional standard errors (Stearns and Smith, 2007). Following each administration, each student earns a scale score, with an associated CSEM, and a performance-level classification based on established benchmark cut scores in the scale score metric. Given these three pieces of information, the assumption of a normal distribution of measurement error allows one to calculate the probability that a student would earn each performance level upon retesting. Using cumulative normal probabilities, the probability that a retest would produce the same performance-level classification and the probability of a different performance-level classification were calculated. The process was repeated for each cut score, which results in a probability of classification in each of the performance levels. The decision consistency index for the entire sample is the average of the probabilities of the same classification on retesting. The values can be interpreted in similar ways.

Stearns and Smith (2007) point out that one advantage of this method is that each student can understand how likely it is that he or she would be classified in the same or different performance levels if the student retook the test without additional instruction. For example, a student scoring right at the cut score will have a lower rate of consistent classification than a student scoring in the middle of a performance-level band.

Table 16–4 reports the decision consistency and kappa indices by content area and grade level. The data shows high agreement rates ranging from 0.89 to 0.90 for mathematics and 0.88 to 0.90 for ELA and Literature. If students were to retest within a short timeframe, with no additional instruction, and under similar conditions, they can expect to earn the same performance-level classification.

The classification-probability contingency tables are presented in Appendix F.

Table 16–4. Decision Consistency and Kappa by Content Area and Grade/Level

Content Area	Grade/Level	N	Decision Consistency	Kappa
Mathematics	3	31621	0.89	0.79
Mathematics	4	29043	0.90	0.79
Mathematics	5	31646	0.90	0.79
Mathematics	6	29745	0.89	0.77
Mathematics	7	26581	0.90	0.77
Mathematics	8	24033	0.90	0.77
Algebra I	HS	26927	0.90	0.77
ELA	3	26345	0.88	0.76
ELA	4	25689	0.88	0.77
ELA	5	28077	0.88	0.77
ELA	6	26547	0.89	0.77
ELA	7	24015	0.90	0.77
ELA	8	23941	0.89	0.76
Literature	HS	23820	0.89	0.77

CHAPTER SEVENTEEN: VALIDITY

As defined in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), validity refers to “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (p. 11). The *Standards* provides a framework for describing the sources of evidence that should be considered when evaluating validity. These sources include evidence based on 1) test content, 2) response processes, 3) the internal structure of the test, 4) the relationships between test scores and other variables, and 5) the consequences of testing. In addition, when Item Response Theory (IRT) models are used to analyze assessment data, validity considerations related to those processes should also be explored.

The validity process involves the collection of a variety of evidence to support the proposed test score interpretations and uses. The entire technical report describes the technical aspects of Firefly Student Progress Indicators in support of its score interpretations and uses. Each of the previous chapters contributes important evidence components that pertain to score validation: test development, test administration, test scoring, item analysis, Rasch calibration, scaling, equating, score reporting, and reliability. This chapter is used to summarize and synthesize the evidence based on the framework of the *Standards*. The purposes and intended use of Firefly is reviewed first, and then each type of validity evidence is addressed in turn.

PURPOSES AND INTENDED USES OF FIREFLY

The *Standards* emphasize that validity pertains to how test scores are used. To help contextualize the evidence that will be presented below, the purposes of Firefly will be reviewed first. Firefly was developed to provide a state-sponsored on-grade interim assessment to districts and schools that students can take multiple times a year to assess progress relative to end-of-year or end-of-course goals. One of the main objectives in the development of Firefly was to provide predictions to the statewide summative tests, PSSA for grades 3 through 8 and Keystone for high school courses. As discussed previously, many design and development decisions for Firefly were made to support predictions. For example, test blueprints for each Firefly grade/level reflect that of PSSA and Keystone. Furthermore, the item types included on Firefly are like those on PSSA and Keystone¹. In other words, Firefly purports to measure similar constructs in a similar way to PSSA and Keystone.

Firefly was designed to be administered in a single class period, so the length of the test was a guiding factor. The test was designed to be the minimum length required in order to report subscores for reporting categories, acknowledging that subscore reliability tends to be lower than total score reliability, as evidenced in Chapter Seventeen and Appendix D.

The assessment is administered completely online using a computer adaptive test (CAT) model, and participation is voluntary. Firefly scores are available immediately after testing in the dynamic reporting suite. In addition to the scores, this suite includes links to instructional resources. Firefly was designed to provide instructional strategies to support student learning. Instructional supports were developed with respect to performance bands to support classroom learning. Specifically, the instructional supports provide educators with the necessary information to identify students’ academic strengths and areas of need and to group students using performance banks. Then, instructional resources were developed as a tool for classroom learning.

EVIDENCE BASED ON TEST CONTENT

Test content validity evidence for Firefly rests greatly on establishing a link between each piece of the assessment (i.e., the items) and what students should know and be able to do as prescribed by Pennsylvania Core Standards. Firefly is intended to measure the knowledge and skills described in the Assessment Anchors and Eligible Content for grades 3 through 8 in mathematics and English Language Arts as well as in high school courses Algebra I and Literature.

Lane (1999) suggests taking several steps to support the content validity of an assessment. In the case of the operational Firefly, these steps include

- evaluating the degree to which the test specifications represent and align with the knowledge and skills described in the corresponding Assessment Anchors and Eligible Content.

¹ As a computer adaptive test, Firefly provides immediate score reporting. Thus, open-ended or constructed-response item types are not included on Firefly. This is one of the main differences between PSSA, Keystone, and Firefly tests.

- evaluating the alignment between the Firefly items and test specifications to ensure representativeness.
- conducting content reviews of the Firefly items using a panel of content experts to see whether items measure the intended construct or are the sources of construct-irrelevant variance.
- conducting fairness reviews of the items to avoid issues related to a specific subpopulation.
- evaluating procedures for administration and scoring such as the appropriateness of instructions to examinees, practice/training with online tools and tests, and time limits for the assessments.

Chapters Two through Five of this report present considerable information on test content. In summary, Firefly is composed of content from three sources: Pennsylvania CDT; legacy PSSA and Keystone; and, for technology-enhanced items, DRC Beacon, a shelf product with content aligned to the Common Core State Standards in Math and English Language Arts. As described in these earlier chapters, all items were developed to or, in the case of the Beacon-sourced content, aligned by DRC content specialists with the Assessment Anchors and Eligible Content. Regardless of their source, all items were written and reviewed by content specialists; reviewed by external committees for content accuracy, grade-appropriateness, and other qualitative criteria; and reviewed by committees of experts in bias, sensitivity, and fairness. After field testing, all items were reviewed with respect to their statistical properties and, if deemed successful, moved to operational pools in their respective banks. In other words, although Firefly is woven from content from various sources, the development of that content was similar in process and rigor.

Some of the efforts made to ensure content validity are further described below.

- DRC used Webb’s (1999) Depth of Knowledge (DOK) model to ensure that the items aligned with the Assessment Anchors and Eligible Content and the Academic Content Standards in terms of both content and cognitive levels.
- DRC established detailed test and item/passage development specifications and ensured the items were sufficient in number and adequately distributed across content, levels of cognitive complexity, and levels of difficulty.
- DRC selected qualified item writers and provided training to help ensure they wrote high-quality items.
- All newly developed items were first reviewed by content specialists and editors at DRC to make sure they measured the intended standard or Assessment Anchors and Eligible Content. Appropriateness for the intended students was also considered, as well as depth of knowledge, graphics, grammar/punctuation, language demand, and distractor reasonableness.
- Prior to field testing, the test items were submitted to content committees (composed of educators and content experts) for review using, but not limited to, the following categories:
 - Overall quality and clarity
 - Standard or Assessment Anchor and Eligible Content alignment
 - Grade-level appropriateness
 - Difficulty level
 - Depth of knowledge
 - Appropriate sources of challenge (e.g., unintended content and skills)
 - Correct answer
 - Quality of distractors
 - Accessibility, clarity, and overall quality of graphics
 - Appropriate language demand
 - Freedom from bias
- The items were also submitted to a Bias, Fairness, and Sensitivity Committee for review. This committee reviewed items for issues related to diversity, gender, and other pertinent factors.

- Items passing all prior hurdles were tried out in an embedded field-test event. Several statistical analyses were conducted on the field-test data including classical item analyses, distractor analyses, and differential item functioning (DIF) analyses. Items were again carefully reviewed by DRC staff and a committee of teachers with respect to their statistical characteristics. DIF was used to detect test items that might bias test scores for particular groups. Empirical investigation of DIF strengthens the validity evidence related to score interpretations for students in particular groups by eliminating potential sources of construct-irrelevant variance.
- Firefly was administered according to standardized procedures with allowable accommodations. Students were given ample time to complete the tests.

EVIDENCE BASED ON RESPONSE PROCESS

Response-process evidence is used to examine the extent to which the cognitive skills and processes employed by students match those identified in the test developer’s defined and intended construct domains for all students and for each subgroup. Think-aloud procedures or “cognitive labs” can be used to collect this type of evidence.

For the operational 2024–2025 Firefly, no cognitive lab studies were conducted to collect the response process evidence.

EVIDENCE BASED ON INTERNAL STRUCTURE

As described in the *Standards* (2014), internal-structure evidence refers to the degree to which the relationships among test items and test components conform to the construct on which the proposed test interpretations are based. Several dimensionality studies were conducted to provide internal-structure evidence relating to the use of scores.

ITEM-TEST CORRELATIONS

Item-test correlations are discussed in Chapter Seven. Items in the final operational pools had acceptable values for correlations.

DIMENSIONALITY

Dimensionality analyses were conducted for Firefly to provide evidence of the internal structure. Both Principal Components Analyses (PCA) and Confirmatory Factor Analyses (CFA) were conducted to examine the fit of the unidimensional model to the Firefly data. Results are presented in Chapter Seven. The results from both PCA and CFA support the assumption that each Firefly test is essentially unidimensional, supporting the validity of using the total scores to estimate a student’s overall ability. Moreover taken as a whole, the internal structure evidence presented generally indicates that related elements of each of the Firefly tests are correlated in the intended manner, which further supports using a total score to report students’ performance.

REPORTING CATEGORY CORRELATIONS

Correlations and disattenuated correlations between reporting category scores within each content area and grade/level are presented below. This data can also provide information on score dimensionality that is part of internal-structure validity evidence. As noted in Chapter Eleven, Firefly mathematics in grades 3 through 8 have five reporting categories per grade and ELA grades 3 through 8 have three reporting categories². Lastly, Firefly Algebra I has six reporting categories (in addition to two module scores), and Literature has four reporting categories (in addition to the two module scores).

For each content area and grade/level, Pearson’s correlation coefficients between the reporting categories are reported in Tables 17–1a through 17–1i. The intercorrelations between the reporting categories within the content areas are positive and generally range from moderate to high in value, and correlations between strands across content areas are generally slightly lower, providing contrasting evidence of convergent and discriminant validity.

² Only the two reading reporting categories were operational in 2024–2025 in Firefly ELA.

Table 17–1a. Correlations among Mathematics Grade 3 Reporting Categories

Reporting Category	A–T	A–F	B–O	C–G	D–M
A–T	-				
A–F	0.51	-			
B–O	0.67	0.53	-		
C–G	0.51	0.47	0.55	-	
D–M	0.64	0.52	0.68	0.55	-

Table 17–1b. Correlations among Mathematics Grade 4 Reporting Categories

Reporting Category	A–T	A–F	B–O	C–G	D–M
A–T	-				
A–F	0.56	-			
B–O	0.65	0.61	-		
C–G	0.41	0.44	0.45	-	
D–M	0.54	0.55	0.57	0.42	-

Table 17–1c. Correlations among Mathematics Grade 5 Reporting Categories

Reporting Category	A–T	A–F	B–O	C–G	D–M
A–T	-				
A–F	0.64	-			
B–O	0.58	0.56	-		
C–G	0.53	0.51	0.47	-	
D–M	0.58	0.59	0.51	0.47	-

Table 17–1d. Correlations among Mathematics Grade 6 Reporting Categories

Reporting Category	A–N	A–R	B–E	C–G	D–S
A–N	-				
A–R	0.54	-			
B–E	0.62	0.53	-		
C–G	0.43	0.42	0.45	-	
D–S	0.48	0.44	0.49	0.38	-

Table 17–1e. Correlations among Mathematics Grade 7 Reporting Categories

Reporting Category	A–N	A–R	B–E	C–G	D–S
A–N	-				
A–R	0.59	-			
B–E	0.58	0.61	-		
C–G	0.44	0.47	0.50	-	
D–S	0.46	0.49	0.52	0.40	-

Table 17–1f. Correlations among Mathematics Grade 8 Reporting Categories

Reporting Category	A–N	B–E	B–F	C–G	D–S
A–N	-				
B–E	0.59	-			
B–F	0.48	0.61	-		
C–G	0.44	0.55	0.46	-	
D–S	0.47	0.56	0.50	0.44	-

Table 17–1g. Correlations among Algebra I Reporting Categories

Reporting Category	A1.1.1	A1.1.2	A1.1.3	A1.2.1	A1.2.2	A1.2.3
A1.1.1	-					
A1.1.2	0.45	-				
A1.1.3	0.41	0.48	-			
A1.2.1	0.44	0.52	0.47	-		
A1.2.2	0.47	0.53	0.49	0.56	-	
A1.2.3	0.43	0.48	0.44	0.49	0.51	-

Table 17–1h. Correlations among ELA Grades 3–8 Reporting Categories

Grade	Reporting Category	A	B
3	A	-	
3	B	0.72	-
4	A	-	
4	B	0.75	-
5	A	-	
5	B	0.76	-
6	A	-	
6	B	0.76	-
7	A	-	
7	B	0.76	-
8	A	-	
8	B	0.76	-

Table 17–1i. Correlations among Literature Reporting Categories

Reporting Category	L.F.1	L.F.2	L.N.1	L.N.2
L.F.1	-			
L.F.2	0.65	-		
L.N.1	0.61	0.64	-	
L.N.2	0.63	0.67	0.66	-

The correlations in Tables 17–1a through 17–1i are based on the observed reporting category scores. These observed-score correlations are weakened by the existing measurement error contained within each reporting category. As a result, disattenuating the observed correlations can provide an estimate of the relationships between reporting categories if there were no measurement error. (An important caveat is provided further below.) The disattenuated correlation coefficients (R_{xy}) can be computed by using the formula (Spearman 1904, 1910) below:

$$R_{xy} = \frac{r_{xy}}{\sqrt{r_{xx} r_{yy}}},$$

where r_{xy} is the observed correlation, and r_{xx} and r_{yy} are the reliabilities for strand X and strand Y. Disattenuated correlations very near 1.00 might suggest that the same or very similar constructs are being measured. Values somewhat less than 1.00 might suggest that different reporting categories are measuring slightly different aspects of the same construct. Values markedly less than 1.00 might suggest the reporting categories reflect different constructs.

Tables 17–2a through 17–2i show the corresponding disattenuated correlations for Firefly for each grade and level. The disattenuated correlations are higher than their observed score counterparts, given that reporting category scores do not have perfect reliabilities (see Chapter Sixteen).

Table 17–2a. Disattenuated Correlations among Mathematics Grade 3 Reporting Categories

Reporting Category	A–T	A–F	B–O	C–G	D–M
A–T	-				
A–F	0.73	-			
B–O	0.89	0.75	-		
C–G	0.72	0.70	0.77	-	
D–M	0.89	0.77	0.94	0.79	-

Table 17–2b. Disattenuated Correlations among Mathematics Grade 4 Reporting Categories

Reporting Category	A–T	A–F	B–O	C–G	D–M
A–T	-				
A–F	0.78	-			
B–O	0.91	0.84	-		
C–G	0.63	0.67	0.69	-	
D–M	0.82	0.83	0.87	0.70	-

Table 17–2c. Disattenuated Correlations among Mathematics Grade 5 Reporting Categories

Reporting Category	A–T	A–F	B–O	C–G	D–M
A–T	-				
A–F	0.88	-			
B–O	0.87	0.83	-		
C–G	0.81	0.76	0.78	-	
D–M	0.87	0.86	0.81	0.77	-

Table 17–2d. Disattenuated Correlations among Mathematics Grade 6 Reporting Categories

Reporting Category	A–N	A–R	B–E	C–G	D–S
A–N	-				
A–R	0.82	-			
B–E	0.91	0.78	-		
C–G	0.67	0.65	0.67	-	
D–S	0.76	0.70	0.76	0.61	-

Table 17–2e. Disattenuated Correlations among Mathematics Grade 7 Reporting Categories

Reporting Category	A–N	A–R	B–E	C–G	D–S
A–N	-				
A–R	0.89	-			
B–E	0.87	0.88	-		
C–G	0.72	0.75	0.80	-	
D–S	0.75	0.77	0.81	0.69	-

Table 17–2f. Disattenuated Correlations among Mathematics Grade 8 Reporting Categories

Reporting Category	A–N	B–E	B–F	C–G	D–S
A–N	-				
B–E	0.83	-			
B–F	0.75	0.91	-		
C–G	0.70	0.83	0.76	-	
D–S	0.73	0.82	0.81	0.73	-

Table 17–2g. Disattenuated Correlations among Algebra I Reporting Categories

Reporting Category	A1.1.1	A1.1.2	A1.1.3	A1.2.1	A1.2.2	A1.2.3
A1.1.1	-					
A1.1.2	0.74	-				
A1.1.3	0.67	0.79	-			
A1.2.1	0.72	0.84	0.77	-		
A1.2.2	0.74	0.85	0.79	0.88	-	
A1.2.3	0.70	0.79	0.73	0.80	0.81	-

Table 17–2h. Disattenuated Correlations among ELA Grades 3–8 Reporting Categories

ELA Grade	Reporting Category	A	B
3	A	-	
3	B	0.97	-
4	A	-	
4	B	0.96	-
5	A	-	
5	B	0.96	-
6	A	-	
6	B	0.96	-
7	A	-	
7	B	0.96	-
8	A	-	
8	B	0.95	-

Table 17–2i. Disattenuated Correlations among Literature Reporting Categories

Reporting Category	L.F.1	L.F.2	L.N.1	L.N.2
L.F.1	-			
L.F.2	1.02	-		
L.N.1	1.00	0.97	-	
L.N.2	0.98	0.95	1.00	-

For ELA and Literature, most disattenuated correlations are very high (e.g., above 0.95), suggesting that the reporting categories appear to be measuring essentially the same construct. This, in turn, suggests that some reporting category scores might not provide entirely unique information about the strengths or weaknesses of students. On the other hand, disattenuated correlations for mathematics and Algebra I tend to range between 0.7 and 0.9, indicating that scores are strongly related, but measure slightly different constructs. Note that, while the theoretical maximum for observed correlations is 1.00, disattenuated correlations can exceed this value when high observed correlations are combined with high reliabilities.

EVIDENCE BASED ON RELATIONSHIPS WITH OTHER VARIABLES

As described in the *Standards* (AERA, APA, & NCME, 2014), “. . . Evidence based on relationships with other variables provides evidence about the degree to which these relationships are consistent with the construct underlying the proposed test score interpretations” (p. 16). This category of evidence refers to “external structure evidence” and has been classified as three types of evidence: *convergent*, *discriminant*, and *criterion-related*. *Convergent evidence* is provided by relationships among students’ performances on different assessments intended to measure a similar construct. *Discriminant evidence* is provided by relationships among students’ performances on different tests intended to measure different constructs. *Criterion-related evidence*, either predictive or concurrent, is provided by relationships between students’ test scores and their performances on a criterion measure (Cronbach, 1971; Messick, 1989).

Correlations and disattenuated correlations among students’ test scores across Firefly content areas provide evidence of discriminant validity (see Table 17–3). Firefly mathematics and ELA tests as well as Firefly Algebra I and Literature tests are designed to measure different constructs, so the correlations across content areas were not

expected to be very high. The correlations ranged from 0.53 to 0.71 and the disattenuated correlations ranged from 0.58 to 0.78, thus supporting this expectation.

Table 17–3. Correlations across Firefly Mathematics and ELA Tests

Grade/Level	N	Correlation	Disattenuated Correlation
3	17725	0.69	0.77
4	17215	0.67	0.74
5	18352	0.71	0.78
6	18414	0.69	0.77
7	16195	0.68	0.76
8	14016	0.67	0.75
HS	4768	0.53	0.58

Some students may take high school Firefly Algebra I or Literature at the same time as mathematics or ELA grade-specific tests if they are expected to take both PSSA and Keystone in the spring. For example, an eighth-grade student who is enrolled in Algebra I may take both Keystone Algebra I as well as the required PSSA math grade 8. As a result, students will likely take both Firefly Algebra I and Firefly math grades 7 or 8. Thus, evaluating the relationship of performance within content areas provides evidence of convergent validity. While the Firefly Algebra I and the Firefly math grade 8 have different test blueprints, it is expected that there is a positive relationship between student scores on both tests. Table 17–4 presents the N count, as well as the correlation and disattenuated correlation of student scores within content areas. As shown in Table 17–4, there is a positive relationship between scores on grade-specific Firefly tests and high school Firefly tests. The correlations between mathematics and Algebra I is somewhat greater than that of ELA and Literature; however, there are fewer students who take both Firefly Literature and Firefly ELA tests within the same year.

Table 17–4. Correlations within Firefly Content Areas

Content Area (HS)	Content Area and Grade	N	Correlation	Disattenuated Correlation
Algebra I	Mathematics Grade 7	109	0.80	0.88
Algebra I	Mathematics Grade 8	1091	0.69	0.76
Literature	ELA Grade 7	54	0.58	0.65
Literature	ELA Grade 8	158	0.78	0.88

External evidence for Firefly can be examined by evaluating the relationship between Firefly and student scores on PSSA and Keystone Examinations as external criteria. One of the main reasons for the development of Firefly was to predict performance on summative examinations. Table 17–5 summarizes the relationship between Firefly and PSSA for grades 3 through 8 and Firefly and Keystone Exams. The correlations for math grades 3 to 8 were above 0.80; whereas the correlations for ELA grades 3 to 8 were slightly lower, ranging from 0.74 to 0.78. Correlations between Firefly and Keystone were slightly lower than those between Firefly and PSSA.

Table 17–5. Correlations among Firefly and Spring Summative Tests, PSSA, and Keystone

Firefly Test	Spring Summative Test	N	Correlation
Mathematics Grade 3	PSSA Mathematics Grade 3	30966	0.83
Mathematics Grade 4	PSSA Mathematics Grade 4	28452	0.82
Mathematics Grade 5	PSSA Mathematics Grade 5	30894	0.82
Mathematics Grade 6	PSSA Mathematics Grade 6	29005	0.81
Mathematics Grade 7	PSSA Mathematics Grade 7	25760	0.83
Mathematics Grade 8	PSSA Mathematics Grade 8	23182	0.81
Algebra I	Keystone Algebra I	21281	0.76
ELA Grade 3	PSSA ELA Grade 3	25772	0.76
ELA Grade 4	PSSA ELA Grade 4	25118	0.76
ELA Grade 5	PSSA ELA Grade 5	27421	0.78
ELA Grade 6	PSSA ELA Grade 6	25876	0.76
ELA Grade 7	PSSA ELA Grade 7	23327	0.75
ELA Grade 8	PSSA ELA Grade 8	23179	0.74
Literature	Keystone Literature	16078	0.67

The aggregation of the results provides internal and external evidence in support of Firefly as a valid measure of student achievement. The collection of external evidence related to Firefly is an ongoing process. As more Firefly data become available, other criterion-related evidence will be evaluated. In addition to examining the relationship between Firefly and PSSA or Keystone Exams, other criterion variables such as Scholastic Aptitude Test (SAT) scores, American College Test (ACT) scores, or student grade point average (GPA) may be considered.

EVIDENCE BASED ON CONSEQUENCES OF TESTS

According to the *Standards* (AERA, APA, & NCME, 2014), evidence of the consequences of implementing an assessment program is an additional source of validity information. Both positive and negative (intended and unintended) consequences of score-based inferences must be investigated to fully evaluate the pool of validity evidence.

Lane and Stone (2002) summarized the general *intended* consequences for state assessments and accountability programs:

- Student, teacher, and administrator motivation and effort
- Curriculum and instruction practices (including content and strategies)
- Improved learning for all students
- Content and format of classroom assessments
- Professional development support
- Use and nature of test preparation activities
- Student, teacher, administrator, and public awareness and beliefs about the assessment, criteria for judging performance, and the use of assessment results

Evidence for the progress of student learning can be seen by looking at differences in student performance for students who took Firefly multiple times. One of the goals for any interim or benchmark assessment is for students to show progress between test events. Chapter Fourteen summarizes differences between test events. Evidence

of student progress supports the intended use and interpretation of student performance on Firefly across different test events. Refer to Chapter Fourteen, Table 14–13 for additional information about student progress.

In addition, evidence for predictive models is presented in Chapter Fifteen. Student performance on Firefly was designed to be predictive of end-of-year and end-of-course summative assessments. The evidence presented in Chapter Fifteen can support the intended use and interpretation of predictions. Refer to Chapter Fifteen for additional information about predictions.

Lane and Stone (2002) also summarized the possible *unintended* outcomes:

- Narrowing of curriculum and instruction to focus only on the specific standards assessed and ignoring the broader construct reflected in the specified standards
- Use of test preparation materials that are closely linked to the assessment without making changes to instruction
- Use of unethical test preparation materials or administration procedures
- Differential performance gains for subgroups of students
- Inappropriate or unfair uses of test scores, such as questionable practices in reassignment of teachers or principals
- For some students, decreased confidence and motivation to learn and to perform well on the assessment because of past experiences with assessments

As noted above, one important piece of consequential evidence pertains to the use of assessment results. As discussed in Chapter Thirteen, the suite of Firefly reports will be available starting in the 2025–2026 academic year. The extent to which various groups of users (e.g., students and teachers) interpret these reports appropriately impacts the validity of subsequent uses of these results. As noted in Chapter Thirteen, there are report training scenarios for each content area and grade/level. The intent is that the scenarios will support the utility of the reports and help users avoid unintended uses and interpretations of Firefly results.

EVIDENCE RELATED TO USE OF THE RASCH MODEL

Since the Rasch model is the basis of all calibration, scaling, and equating analyses associated with Firefly, the validity of the inferences from these results depends on the degree to which the assumptions of the model are met, as well as the fit between the model and the test data. As discussed in Chapter Seven, the underlying assumptions of Rasch models were essentially met for all Firefly data, supporting the use of the Rasch model to analyze Firefly data.

VALIDITY EVIDENCE SUMMARY

Validity evidence related to test content was reviewed earlier in this chapter. Overall, the early chapters of this technical report show that a strong link can be established between Firefly items and associated Pennsylvania Core Standards and Eligible Content. Detailed information regarding educator reviews is presented in Chapters Two and Three.

Evidence of convergent and divergent validity was presented in this chapter. This included the relationship among reporting category subscores, the relationship among scores within a content area, and the relationship between scores on Firefly and scores on PSSA and Keystone. The data shows favorable evidence of discriminant and convergent validity.

Validity of score inferences is bolstered when test scores are consistent and when decision consistency is high. Here, the reliabilities of the total test scores (presented in Chapter Sixteen) were very good, with many in the high 0.80s and low 0.90s. Moreover, the consistency of decisions was presented in Chapter Sixteen.

Reported in Chapters Five and Six, differential item functioning (DIF) with respect to gender and ethnicity helps address construct-irrelevant variance, which represents an important threat to the validity of inferences made from achievement test scores. As noted in that chapter, field-test items are screened and reviewed for DIF.

APPENDIX A: OPERATIONAL ITEM POOL SUMMARY

As discussed in Chapter Four, the operational item pools were developed to ensure adequate coverage across each reporting category. The table below shows item counts by item type and reporting category for each content area and grade/level.

Content Area	Grade/Level	Reporting Category	MC Count	EBSR Count	Total Count
Mathematics	3	A-F	78	0	78
Mathematics	3	A-T	70	0	70
Mathematics	3	B-O	113	0	113
Mathematics	3	C-G	74	0	74
Mathematics	3	D-M	132	0	132
Mathematics	4	A-F	99	0	99
Mathematics	4	A-T	81	0	81
Mathematics	4	B-O	103	0	103
Mathematics	4	C-G	70	0	70
Mathematics	4	D-M	85	0	85
Mathematics	5	A-F	84	0	84
Mathematics	5	A-T	100	0	100
Mathematics	5	B-O	70	0	70
Mathematics	5	C-G	70	0	70
Mathematics	5	D-M	70	0	70
Mathematics	6	A-N	96	0	96
Mathematics	6	A-R	78	0	78
Mathematics	6	B-E	117	0	117
Mathematics	6	C-G	71	0	71
Mathematics	6	D-S	80	0	80
Mathematics	7	A-N	51	0	51
Mathematics	7	A-R	91	0	91
Mathematics	7	B-E	99	0	99
Mathematics	7	C-G	82	0	82
Mathematics	7	D-S	74	0	74
Mathematics	8	A-N	57	0	57
Mathematics	8	B-E	121	0	121
Mathematics	8	B-F	56	0	56
Mathematics	8	C-G	79	0	79
Mathematics	8	D-S	71	0	71

Content Area	Grade/Level	Reporting Category	MC Count	EBSR Count	Total Count
Algebra I	HS	A1.1	213	0	213
Algebra I	HS	A1.1.1	73	0	73
Algebra I	HS	A1.1.2	70	0	70
Algebra I	HS	A1.1.3	70	0	70
Algebra I	HS	A1.2	210	0	210
Algebra I	HS	A1.2.1	70	0	70
Algebra I	HS	A1.2.2	70	0	70
Algebra I	HS	A1.2.3	70	0	70
ELA	3	A	148	25	173
ELA	3	B	161	23	184
ELA	4	A	153	25	178
ELA	4	B	156	23	179
ELA	5	A	151	27	178
ELA	5	B	148	29	177
ELA	6	A	163	18	181
ELA	6	B	164	19	183
ELA	7	A	156	20	176
ELA	7	B	160	18	178
ELA	8	A	156	22	178
ELA	8	B	161	21	182
Literature	HS	L.F	224	22	246
Literature	HS	L.F.1	85	4	89
Literature	HS	L.F.2	139	18	157
Literature	HS	L.N	228	18	246
Literature	HS	L.N.1	82	8	90
Literature	HS	L.N.2	146	10	156

APPENDIX B: ITEM RESIDUAL CORRELATIONS

As discussed in Chapter Seven, local item independence (LI) is an important assumption in the Rasch Model. The data presented in Chapter Seven provided an overview of the LI statistic used, item residual correlations. The data presented in this appendix supports the information previously supplied by showing the item pairs with high correlations (greater than |0.2|). As shown in the table below, all math item pairs identified assessed the same standard, and all ELA and Literature item pairs identified used the same passage.

Content Area	Grade/Level	ItemID	Item Type	Standard	Passage	Associated ItemID	Item Type	Associated Standard	Associated Passage	Residual Correlation
Mathematics	6	1190592	MC	C-G.1.1.3	NA	1190593	MC	C-G.1.1.3	NA	0.20
Algebra I	HS	1191781	MC	A1.2.1.1.3	NA	1191844	MC	A1.2.1.1.3	NA	0.24
Algebra I	HS	1191928	MC	A1.2.3.1.1	NA	1191937	MC	A1.2.3.1.1	NA	0.23
ELA	3	1192124	MC	A-K.1.1.2	150141	1192125	MC	A-V.4.1.1	150141	0.37
ELA	3	1192149	MC	B-V.4.1.1	150148	1192150	MC	B-K.1.1.1	150148	1.00
ELA	3	1192208	MC	A-K.1.1.2	150161	1192210	MC	A-V.4.1.2	150161	0.25
ELA	3	1192215	MC	B-C.3.1.2	150163	1192234	MC	B-V.4.1.2	150163	-0.27
ELA	3	1192216	MC	B-K.1.1.2	150163	1192234	MC	B-V.4.1.2	150163	-0.36
ELA	3	1192291	MC	B-C.3.1.3	150167	1192293	MC	B-V.4.1.2	150167	-0.21
ELA	3	1192346	MC	B-K.1.1.1	150187	1192348	MC	B-C.2.1.2	150187	0.20
ELA	3	1192346	MC	B-K.1.1.1	150187	1192350	MC	B-V.4.1.2	150187	0.20
ELA	3	1192346	MC	B-K.1.1.1	150187	1192352	MC	B-C.2.1.1	150187	0.21
ELA	3	1192348	MC	B-C.2.1.2	150187	1192350	MC	B-V.4.1.2	150187	1.00
ELA	3	1192348	MC	B-C.2.1.2	150187	1192352	MC	B-C.2.1.1	150187	-0.91
ELA	3	1192350	MC	B-V.4.1.2	150187	1192352	MC	B-C.2.1.1	150187	-0.91
ELA	4	1192481	MC	A-C.3.1.1	150207	1192482	MC	A-C.3.1.1	150207	-1.00
ELA	4	1192481	MC	A-C.3.1.1	150207	1192483	MC	A-C.3.1.1	150207	-1.00
ELA	4	1192482	MC	A-C.3.1.1	150207	1192483	MC	A-C.3.1.1	150207	1.00
ELA	4	1192485	MC	A-V.4.1.1	150208	1192486	MC	A-K.1.1.2	150208	-0.20
ELA	4	1192507	MC	B-V.4.1.1	150212	1192510	MC	B-C.2.1.2	150212	-0.22

Content Area	Grade/Level	ItemID	Item Type	Standard	Passage	Associated ItemID	Item Type	Associated Standard	Associated Passage	Residual Correlation
ELA	4	1192535	MC	A-K.1.1.1	150222	1192536	MC	A-K.1.1.3	150222	-0.25
ELA	4	1192535	MC	A-K.1.1.1	150222	1192538	MC	A-V.4.1.2	150222	-0.30
ELA	4	1192536	MC	A-K.1.1.3	150222	1192537	MC	A-C.2.1.1	150222	0.29
ELA	4	1192755	MC	B-C.2.1.2	150248	1192756	MC	B-V.4.1.1	150248	0.27
ELA	5	1192839	MC	A-K.1.1.2	150272	1192840	MC	A-V.4.1.2	150272	0.95
ELA	5	1192839	MC	A-K.1.1.2	150272	1192841	MC	A-C.2.1.1	150272	0.99
ELA	5	1192839	MC	A-K.1.1.2	150272	1192919	ESR	A-C.2.1.1	150272	0.53
ELA	5	1192840	MC	A-V.4.1.2	150272	1192841	MC	A-C.2.1.1	150272	0.98
ELA	5	1192840	MC	A-V.4.1.2	150272	1192919	ESR	A-C.2.1.1	150272	0.68
ELA	5	1192841	MC	A-C.2.1.1	150272	1192919	ESR	A-C.2.1.1	150272	0.61
ELA	5	1192843	MC	A-K.1.1.1	150273	1192846	MC	A-K.1.1.1	150273	0.26
ELA	5	1192859	MC	B-V.4.1.1	150277	1192861	MC	B-C.3.1.3	150277	0.29
ELA	5	1192860	MC	B-K.1.1.1	150277	1192861	MC	B-C.3.1.3	150277	0.26
ELA	5	1192867	MC	A-K.1.1.1	150279	1192869	MC	A-V.4.1.2	150279	-0.32
ELA	5	1192869	MC	A-V.4.1.2	150279	1192929	ESR	A-K.1.1.1	150279	-0.22
ELA	5	1192954	MC	A-K.1.1.2	150304	1192962	MC	A-C.3.1.1	150304	0.20
ELA	5	1193053	MC	A-K.1.1.2	150316	1193054	MC	A-K.1.1.3	150316	0.21
ELA	5	1193096	MC	B-C.2.1.1	150310	1193097	MC	B-C.3.1.1	150310	-0.74
ELA	5	1193096	MC	B-C.2.1.1	150310	1193099	MC	B-V.4.1.2	150310	-0.68
ELA	5	1193096	MC	B-C.2.1.1	150310	1193100	MC	B-K.1.1.1	150310	1.00
ELA	5	1193097	MC	B-C.3.1.1	150310	1193100	MC	B-K.1.1.1	150310	-0.74
ELA	5	1193099	MC	B-V.4.1.2	150310	1193100	MC	B-K.1.1.1	150310	-0.68
ELA	5	1193138	MC	A-C.2.1.1	150324	1193139	MC	A-K.1.1.3	150324	-0.24
ELA	6	1193189	MC	A-C.2.1.1	150334	1193190	MC	A-V.4.1.2	150334	0.35
ELA	6	1193190	MC	A-V.4.1.2	150334	1193191	MC	A-V.4.1.1	150334	-0.55

Content Area	Grade/Level	ItemID	Item Type	Standard	Passage	Associated ItemID	Item Type	Associated Standard	Associated Passage	Residual Correlation
ELA	6	1193213	MC	B-C.2.1.1	150342	1193217	MC	B-C.3.1.1	150342	-0.25
ELA	6	1193230	MC	B-K.1.1.1	150337	1193232	MC	B-C.3.1.1	150337	0.26
ELA	6	1193231	MC	B-V.4.1.1	150337	1193241	MC	B-K.1.1.1	150337	0.39
ELA	6	1193270	MC	A-K.1.1.1	150346	1193271	MC	A-V.4.1.2	150346	0.46
ELA	6	1193270	MC	A-K.1.1.1	150346	1193272	MC	A-C.2.1.1	150346	0.51
ELA	6	1193273	MC	A-K.1.1.2	150347	1193275	MC	A-C.2.1.1	150347	0.51
ELA	6	1193274	MC	A-V.4.1.2	150347	1193275	MC	A-C.2.1.1	150347	0.33
ELA	6	1193285	MC	A-V.4.1.1	150354	1193286	MC	A-K.1.1.3	150354	0.20
ELA	6	1193286	MC	A-K.1.1.3	150354	1193287	MC	A-C.2.1.2	150354	0.22
ELA	6	1193288	MC	A-K.1.1.3	150355	1193295	MC	A-C.3.1.1	150355	-0.55
ELA	6	1193301	MC	A-V.4.1.2	150356	1193303	MC	A-C.2.1.2	150356	-0.27
ELA	6	1193317	MC	A-K.1.1.3	150357	1193318	MC	A-C.2.1.2	150357	0.27
ELA	6	1193352	MC	B-C.2.1.2	150363	1193354	MC	B-C.3.1.2	150363	-0.25
ELA	6	1193352	MC	B-C.2.1.2	150363	1193355	MC	B-V.4.1.2	150363	-0.22
ELA	6	1193352	MC	B-C.2.1.2	150363	1193357	ESR	B-K.1.1.2	150363	-0.22
ELA	6	1193353	MC	B-C.2.1.3	150363	1193356	MC	B-K.1.1.1	150363	-0.28
ELA	6	1193353	MC	B-C.2.1.3	150363	1193357	ESR	B-K.1.1.2	150363	0.23
ELA	6	1193354	MC	B-C.3.1.2	150363	1193355	MC	B-V.4.1.2	150363	0.32
ELA	6	1193496	MC	A-V.4.1.1	150392	1193497	MC	A-C.2.1.1	150392	0.23
ELA	7	1193567	MC	A-V.4.1.2	150411	1193628	MC	A-C.2.1.3	150411	-0.24
ELA	7	1193582	MC	B-K.1.1.2	150413	1193587	MC	B-K.1.1.1	150413	0.23
ELA	7	1193588	MC	A-V.4.1.1	150414	1193590	MC	A-C.2.1.3	150414	-0.23
ELA	7	1193588	MC	A-V.4.1.1	150414	1193623	MC	A-C.2.1.2	150414	-0.30
ELA	7	1193610	MC	B-K.1.1.1	150409	1193619	MC	B-K.1.1.1	150409	0.31
ELA	7	1193622	MC	A-C.2.1.2	150411	1193628	MC	A-C.2.1.3	150411	0.35

Content Area	Grade/Level	ItemID	Item Type	Standard	Passage	Associated ItemID	Item Type	Associated Standard	Associated Passage	Residual Correlation
ELA	7	1193633	MC	A-K.1.1.3	150425	1193635	MC	A-C.3.1.1	150425	0.32
ELA	7	1193634	MC	A-C.2.1.1	150425	1193635	MC	A-C.3.1.1	150425	0.21
ELA	7	1193685	MC	B-V.4.1.1	150419	1193686	MC	B-V.4.1.2	150419	0.23
ELA	7	1193769	MC	B-V.4.1.1	150448	1193770	MC	B-V.4.1.2	150448	0.35
ELA	7	1193771	ESR	B-K.1.1.2	150448	1193773	MC	B-C.2.1.3	150448	-0.22
ELA	8	1193918	MC	A-V.4.1.2	150478	1193919	MC	A-C.2.1.3	150478	0.36
ELA	8	1193919	MC	A-C.2.1.3	150478	1193920	MC	A-K.1.1.2	150478	-0.44
ELA	8	1193921	MC	A-V.4.1.2	150476	1193922	MC	A-K.1.1.1	150476	0.56
ELA	8	1193925	MC	B-K.1.1.1	150481	1193972	MC	B-K.1.1.3	150481	-0.56
ELA	8	1193948	MC	B-C.2.1.1	150475	1193949	MC	B-K.1.1.2	150475	1.00
ELA	8	1193948	MC	B-C.2.1.1	150475	1193956	MC	B-V.4.1.1	150475	1.00
ELA	8	1193949	MC	B-K.1.1.2	150475	1193956	MC	B-V.4.1.1	150475	1.00
ELA	8	1193953	MC	B-V.4.1.1	150471	1193962	MC	B-K.1.1.2	150471	1.00
ELA	8	1193953	MC	B-V.4.1.1	150471	1193975	MC	B-K.1.1.3	150471	1.00
ELA	8	1193962	MC	B-K.1.1.2	150471	1193975	MC	B-K.1.1.3	150471	1.00
ELA	8	1193966	MC	A-K.1.1.1	150477	1193967	MC	A-V.4.1.2	150477	0.22
ELA	8	1194032	MC	B-K.1.1.2	150493	1194036	MC	B-V.4.1.1	150493	-0.21
ELA	8	1194062	MC	B-C.2.1.3	150497	1194063	MC	B-V.4.1.2	150497	0.25
ELA	8	1194062	MC	B-C.2.1.3	150497	1194065	ESR	B-K.1.1.2	150497	-0.25
ELA	8	1194122	MC	A-V.4.1.2	150514	1194123	MC	A-C.2.1.3	150514	0.26
ELA	8	1194147	MC	B-V.4.1.2	150506	1194150	MC	B-C.2.1.2	150506	-0.22
ELA	8	1194149	MC	B-K.1.1.3	150506	1194150	MC	B-C.2.1.2	150506	-0.35
ELA	8	1194208	MC	A-V.4.1.1	150523	1194210	MC	A-V.4.1.1	150523	0.20
Literature	HS	1194280	MC	L.F.1.2.4	150537	1194281	MC	L.F.1.3.2	150537	0.38
Literature	HS	1194281	MC	L.F.1.3.2	150537	1194282	MC	L.F.2.3.2	150537	-0.24

Content Area	Grade/Level	ItemID	Item Type	Standard	Passage	Associated ItemID	Item Type	Associated Standard	Associated Passage	Residual Correlation
Literature	HS	1194291	MC	L.N.2.3.1	150565	1194292	MC	L.N.2.3.2	150565	-0.97
Literature	HS	1194291	MC	L.N.2.3.1	150565	1194293	MC	L.N.2.3.4	150565	0.40
Literature	HS	1194292	MC	L.N.2.3.2	150565	1194293	MC	L.N.2.3.4	150565	-0.61
Literature	HS	1194297	MC	L.N.2.3.2	150564	1194299	MC	L.N.1.3.2	150564	-0.40
Literature	HS	1194299	MC	L.N.1.3.2	150564	1194480	MC	L.N.1.3.3	150564	0.41
Literature	HS	1194323	MC	L.F.1.1.2	150558	1194326	MC	L.F.2.3.1	150558	-0.27
Literature	HS	1194325	MC	L.F.2.1.1	150558	1194326	MC	L.F.2.3.1	150558	0.25
Literature	HS	1194346	MC	L.N.2.5.1	150575	1194353	MC	L.N.1.2.2	150575	0.21
Literature	HS	1194385	MC	L.N.2.5.5	150576	1194386	MC	L.N.1.1.2	150576	0.62
Literature	HS	1194399	MC	L.N.1.1.1	150577	1194401	MC	L.N.2.3.5	150577	0.40
Literature	HS	1194399	MC	L.N.1.1.1	150577	1194477	MC	L.N.1.1.4	150577	0.24
Literature	HS	1194400	MC	L.N.2.1.1	150577	1194477	MC	L.N.1.1.4	150577	0.67
Literature	HS	1194401	MC	L.N.2.3.5	150577	1194477	MC	L.N.1.1.4	150577	-0.21
Literature	HS	1194412	MC	L.N.1.1.3	150551	1194424	MC	L.N.2.5.4	150551	-0.28
Literature	HS	1194415	MC	L.N.2.1.1	150552	1194439	MC	L.N.1.1.1	150552	0.23
Literature	HS	1194415	MC	L.N.2.1.1	150552	1194440	MC	L.N.1.3.1	150552	0.20
Literature	HS	1194436	MC	L.N.2.2.3	150553	1194437	MC	L.N.2.2.3	150553	1.00
Literature	HS	1194436	MC	L.N.2.2.3	150553	1194438	MC	L.N.2.2.3	150553	-0.49
Literature	HS	1194437	MC	L.N.2.2.3	150553	1194438	MC	L.N.2.2.3	150553	-0.49
Literature	HS	1194476	MC	L.F.2.1.1	150555	1194505	MC	L.F.2.3.2	150555	0.20
Literature	HS	1194503	MC	L.F.2.2.2	150555	1194505	MC	L.F.2.3.2	150555	-0.21
Literature	HS	1194662	MC	L.N.2.3.5	150602	1194710	ESR	L.N.2.5.6	150602	-0.20

APPENDIX C: SCALE MAINTENANCE

As discussed in Chapter Eight, items were flagged due to statistical reasons in the scale maintenance process. The items identified were removed from the anchor set during calibration to allow item parameters to be re-estimated based on the most recent Firefly data. The Rasch statistics (difficulty, mean-square infit, and displacement value) for the administration and the re-estimated Rasch statistics are presented in the tables below.

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1188889	0.90	0.53	0.78	0.92	0.00	1.04
Mathematics	3	1188891	0.93	0.58	-2.54	0.99	0.00	-2.14
Mathematics	3	1188894	0.97	0.55	-2.68	1.03	0.00	-2.29
Mathematics	3	1188895	0.93	-0.56	-2.44	1.00	0.00	-3.24
Mathematics	3	1188896	0.97	0.05	-1.47	0.96	0.00	-1.72
Mathematics	3	1188899	1.01	1.04	-2.38	1.00	0.00	-1.56
Mathematics	3	1188901	0.91	0.16	-2.95	0.91	0.00	-2.98
Mathematics	3	1188902	1.04	0.39	-2.39	0.94	0.00	-2.26
Mathematics	3	1188903	0.94	-0.10	0.22	0.94	0.00	-0.15
Mathematics	3	1188904	0.93	-0.71	1.05	0.82	0.00	0.02
Mathematics	3	1188905	1.02	-0.86	1.38	0.84	0.00	0.20
Mathematics	3	1188906	1.05	-0.66	-1.93	1.02	0.00	-2.87
Mathematics	3	1188908	0.90	-0.20	-2.20	0.92	0.00	-2.67
Mathematics	3	1188909	0.97	-0.87	-0.25	0.93	0.00	-1.42
Mathematics	3	1188910	0.95	-0.79	0.13	0.94	0.00	-0.95
Mathematics	3	1188912	0.95	-0.37	-0.90	0.94	0.00	-1.59
Mathematics	3	1188913	0.97	-0.20	-0.88	0.97	0.00	-1.40
Mathematics	3	1188916	0.92	-0.26	-2.75	0.92	0.00	-3.23
Mathematics	3	1188918	0.98	-0.37	-0.19	0.97	0.00	-0.84
Mathematics	3	1188919	0.98	-0.09	-2.22	0.97	0.00	-2.59
Mathematics	3	1188920	0.90	-0.33	0.31	0.89	0.00	-0.30

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1188921	0.95	-0.67	0.57	0.89	0.00	-0.40
Mathematics	3	1188922	0.96	0.23	-0.83	0.97	0.00	-0.91
Mathematics	3	1188923	0.93	0.24	-0.58	0.93	0.00	-0.62
Mathematics	3	1188924	0.96	-0.34	-0.99	0.95	0.00	-1.65
Mathematics	3	1188925	0.96	-0.20	-1.07	0.95	0.00	-1.60
Mathematics	3	1188926	0.97	0.94	-1.77	0.93	0.00	-1.10
Mathematics	3	1188927	0.90	1.01	-1.00	0.93	0.00	-0.28
Mathematics	3	1188930	0.91	0.75	-2.93	0.92	0.00	-2.36
Mathematics	3	1188932	1.01	-0.22	-2.29	1.00	0.00	-2.77
Mathematics	3	1188933	0.90	0.63	-3.80	0.97	0.00	-3.26
Mathematics	3	1188936	0.98	0.77	-2.78	0.98	0.00	-2.20
Mathematics	3	1188939	0.87	0.81	0.63	0.90	0.00	1.18
Mathematics	3	1188940	0.85	0.82	0.06	0.83	0.00	0.62
Mathematics	3	1188941	0.91	1.37	-1.19	0.96	0.00	-0.11
Mathematics	3	1188942	0.92	1.01	-1.11	0.97	0.00	-0.39
Mathematics	3	1188945	0.91	1.28	-1.25	0.94	0.00	-0.26
Mathematics	3	1188946	0.94	0.64	-4.02	0.93	0.00	-3.46
Mathematics	3	1188947	0.96	1.09	0.06	0.92	0.00	0.92
Mathematics	3	1188948	0.98	0.88	-1.67	0.96	0.00	-1.07
Mathematics	3	1188951	1.09	-1.05	-3.37	0.95	0.00	-4.58
Mathematics	3	1188952	0.98	-0.79	-1.03	0.93	0.00	-2.16
Mathematics	3	1188953	0.91	0.06	-0.40	0.90	0.00	-0.63
Mathematics	3	1188954	0.91	-0.72	0.64	0.85	0.00	-0.38
Mathematics	3	1188955	1.17	-1.07	2.16	0.82	0.01	0.74
Mathematics	3	1188956	1.16	-1.01	1.85	0.83	0.00	0.49

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1188957	0.89	-0.80	-1.05	0.92	0.00	-2.16
Mathematics	3	1188958	0.98	-0.51	-1.78	0.99	0.00	-2.57
Mathematics	3	1188959	1.16	-0.92	-3.00	0.99	0.00	-4.11
Mathematics	3	1188960	1.43	-1.72	-2.55	1.05	0.00	-4.53
Mathematics	3	1188962	1.05	-0.65	-3.39	0.96	0.00	-4.18
Mathematics	3	1188964	0.98	0.79	-2.76	0.97	0.00	-2.17
Mathematics	3	1188965	0.89	0.57	-3.50	0.96	0.00	-3.04
Mathematics	3	1188969	0.92	0.75	-0.87	0.93	0.00	-0.40
Mathematics	3	1188971	0.90	0.77	-0.45	0.90	0.00	0.06
Mathematics	3	1188973	0.91	1.28	0.61	0.94	0.00	1.65
Mathematics	3	1188974	0.91	0.93	-0.76	0.90	0.00	-0.11
Mathematics	3	1188975	0.92	1.00	-0.53	0.92	0.00	0.21
Mathematics	3	1188976	0.92	0.73	-3.36	0.97	0.00	-2.77
Mathematics	3	1188980	1.16	-0.68	-2.85	1.04	0.00	-3.72
Mathematics	3	1188981	0.98	0.95	-2.75	0.95	0.00	-1.99
Mathematics	3	1188984	0.97	0.67	-2.38	0.98	0.00	-1.94
Mathematics	3	1188989	0.92	0.71	-3.25	0.95	0.00	-2.69
Mathematics	3	1188990	0.99	-0.43	-2.23	0.99	0.00	-2.91
Mathematics	3	1188992	1.20	-0.68	-1.95	1.06	0.00	-2.87
Mathematics	3	1188993	0.95	-0.08	-3.01	0.95	0.00	-3.27
Mathematics	3	1188994	0.93	-0.57	-0.90	0.97	0.00	-1.77
Mathematics	3	1188996	1.16	0.74	-2.24	0.97	0.00	-1.77
Mathematics	3	1188997	0.90	-1.14	-1.55	0.94	0.00	-2.97
Mathematics	3	1188998	1.12	-1.40	-0.27	0.90	0.00	-2.00
Mathematics	3	1188999	1.08	-0.56	0.17	1.00	0.00	-0.68

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1189000	0.92	-0.54	-1.44	0.95	0.00	-2.28
Mathematics	3	1189001	0.94	0.24	-1.99	0.89	0.00	-2.03
Mathematics	3	1189002	0.88	0.17	-1.87	0.86	0.00	-1.96
Mathematics	3	1189003	1.23	-0.61	0.82	1.04	0.00	-0.09
Mathematics	3	1189004	0.84	-1.03	-1.68	0.88	0.00	-2.99
Mathematics	3	1189005	1.02	-0.29	-0.43	1.01	0.00	-1.00
Mathematics	3	1189006	1.03	-0.05	-0.47	1.04	0.00	-0.81
Mathematics	3	1189007	0.93	-0.14	-0.82	0.92	0.00	-1.25
Mathematics	3	1189008	1.04	-1.06	-0.35	0.97	0.00	-1.71
Mathematics	3	1189010	1.06	-1.25	-0.13	0.92	0.00	-1.68
Mathematics	3	1189011	1.04	-0.20	-0.58	1.03	0.00	-1.05
Mathematics	3	1189012	0.93	-0.48	-1.96	0.95	0.00	-2.71
Mathematics	3	1189013	1.06	-0.35	-1.03	1.02	0.00	-1.69
Mathematics	3	1189014	1.06	-0.43	-0.08	1.01	0.00	-0.79
Mathematics	3	1189015	0.98	-0.25	-1.58	0.99	0.00	-2.13
Mathematics	3	1189016	0.93	0.96	-4.08	0.96	0.00	-3.20
Mathematics	3	1189017	0.92	0.62	-3.38	0.98	0.00	-2.88
Mathematics	3	1189020	1.10	-0.28	0.13	1.08	0.00	-0.44
Mathematics	3	1189021	1.00	-0.58	-1.51	0.96	0.00	-2.40
Mathematics	3	1189025	0.94	0.61	-2.78	0.93	0.00	-2.37
Mathematics	3	1189026	0.95	1.05	-2.87	0.96	0.00	-2.00
Mathematics	3	1189029	1.12	-0.34	-1.93	1.08	0.00	-2.52
Mathematics	3	1189031	0.87	0.85	-3.48	0.97	0.00	-2.74
Mathematics	3	1189033	0.89	0.54	-3.69	0.96	0.00	-3.24
Mathematics	3	1189037	0.91	0.66	-3.44	0.98	0.00	-2.90

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1189038	0.95	1.27	-2.87	0.97	0.00	-1.77
Mathematics	3	1189039	0.92	1.05	-3.11	0.97	0.00	-2.21
Mathematics	3	1189051	0.89	0.62	-3.84	0.95	0.00	-3.31
Mathematics	3	1189052	0.96	0.29	-2.31	0.88	0.00	-2.30
Mathematics	3	1189053	0.89	0.13	-0.96	0.88	0.00	-1.13
Mathematics	3	1189055	1.16	-1.19	-1.51	1.02	0.00	-3.03
Mathematics	3	1189056	1.07	-0.73	-1.34	1.01	0.00	-2.38
Mathematics	3	1189057	1.05	-1.16	-1.63	0.95	0.00	-3.10
Mathematics	3	1189058	1.02	-0.28	-0.84	1.01	0.00	-1.41
Mathematics	3	1189059	1.04	-0.49	-1.48	1.00	0.00	-2.27
Mathematics	3	1189060	1.02	-0.53	-1.70	0.98	0.00	-2.53
Mathematics	3	1189061	0.92	-0.42	-1.78	0.92	0.00	-2.48
Mathematics	3	1189062	1.02	-1.25	-0.63	0.95	0.00	-2.18
Mathematics	3	1189063	0.93	-0.53	-2.73	0.92	0.00	-3.44
Mathematics	3	1189064	1.06	-0.25	-0.27	1.06	0.00	-0.80
Mathematics	3	1189065	0.88	0.84	-3.48	0.97	0.00	-2.76
Mathematics	3	1189066	0.92	0.86	-3.26	0.96	0.00	-2.54
Mathematics	3	1189067	0.88	1.01	-3.29	0.93	0.00	-2.41
Mathematics	3	1189068	0.86	0.95	-3.37	0.94	0.00	-2.55
Mathematics	3	1189072	0.94	0.77	-2.90	0.95	0.00	-2.31
Mathematics	3	1189075	0.91	0.65	-3.02	0.93	0.00	-2.54
Mathematics	3	1189083	0.91	0.74	-3.81	0.97	0.00	-3.15
Mathematics	3	1189089	0.94	0.04	-1.43	0.94	0.00	-1.69
Mathematics	3	1189090	0.98	0.42	-0.36	0.97	0.00	-0.20
Mathematics	3	1189091	1.00	-0.72	-0.92	0.96	0.00	-1.96

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1189092	1.04	-0.87	-0.48	0.97	0.00	-1.66
Mathematics	3	1189093	0.95	-0.52	-1.98	0.96	0.00	-2.76
Mathematics	3	1189094	0.92	-0.37	-3.41	0.90	0.00	-3.92
Mathematics	3	1189097	1.01	-0.45	-0.03	0.97	0.00	-0.76
Mathematics	3	1189098	0.98	-0.40	-1.65	0.94	0.00	-2.35
Mathematics	3	1189099	0.93	-0.28	-2.33	0.92	0.00	-2.85
Mathematics	3	1189100	1.00	0.05	-0.91	1.01	0.00	-1.16
Mathematics	3	1189101	0.92	-0.42	-2.28	0.93	0.00	-2.94
Mathematics	3	1189103	0.98	-0.50	-0.46	0.99	0.00	-1.24
Mathematics	3	1189105	1.09	-0.99	-1.83	1.00	0.00	-3.10
Mathematics	3	1189106	0.92	-0.96	-0.86	0.95	0.00	-2.11
Mathematics	3	1189107	1.03	-0.67	-1.91	0.93	0.00	-2.83
Mathematics	3	1189108	0.98	-0.59	-2.28	0.94	0.00	-3.12
Mathematics	3	1189109	1.00	-0.68	-3.24	0.91	0.00	-4.07
Mathematics	3	1189110	1.17	-0.62	-1.91	1.07	0.00	-2.78
Mathematics	3	1189112	0.95	-1.12	-0.99	0.94	0.00	-2.43
Mathematics	3	1189113	0.97	-0.24	-0.09	0.97	0.00	-0.61
Mathematics	3	1189114	1.04	-0.25	-1.62	1.03	0.00	-2.17
Mathematics	3	1189116	0.96	-0.79	-0.17	0.96	0.00	-1.24
Mathematics	3	1189117	1.06	-0.82	-1.80	1.00	0.00	-2.90
Mathematics	3	1189119	0.99	-0.40	-3.60	0.94	0.00	-4.13
Mathematics	3	1189120	0.92	-0.67	-1.32	0.95	0.00	-2.29
Mathematics	3	1189121	1.12	-0.79	-1.70	1.07	0.00	-2.79
Mathematics	3	1189122	0.91	-0.73	-0.98	0.92	0.00	-2.03
Mathematics	3	1189125	0.95	-0.07	-1.04	0.95	0.00	-1.40

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1189126	0.99	-0.47	-0.40	0.97	0.00	-1.16
Mathematics	3	1189127	1.03	-0.50	-2.66	0.95	0.00	-3.36
Mathematics	3	1189128	1.02	-0.88	-2.84	0.94	0.00	-3.94
Mathematics	3	1189129	0.95	-0.15	-3.26	0.93	0.00	-3.57
Mathematics	3	1189132	0.94	-0.29	-0.40	0.93	0.00	-0.97
Mathematics	3	1189133	0.98	-0.61	-1.85	0.91	0.00	-2.72
Mathematics	3	1189134	0.95	-0.74	-1.09	0.94	0.00	-2.16
Mathematics	3	1189136	1.18	-0.20	-0.02	1.16	0.00	-0.50
Mathematics	3	1189137	0.91	-0.54	-3.74	0.90	0.00	-4.40
Mathematics	3	1189138	0.99	-0.56	-0.77	0.97	0.00	-1.64
Mathematics	3	1189139	1.02	-0.52	-3.14	0.95	0.00	-3.81
Mathematics	3	1189140	0.90	-0.30	-1.73	0.87	0.00	-2.32
Mathematics	3	1189141	0.94	-0.86	-0.91	0.94	0.00	-2.08
Mathematics	3	1189143	0.97	0.01	0.61	0.96	0.00	0.34
Mathematics	3	1189144	0.94	-0.89	-2.29	0.93	0.00	-3.41
Mathematics	3	1189145	1.15	-1.06	-2.74	1.02	0.00	-3.98
Mathematics	3	1189146	1.06	-1.03	-2.49	0.96	0.00	-3.74
Mathematics	3	1189147	0.96	-0.53	-2.73	0.91	0.00	-3.47
Mathematics	3	1189148	1.00	-0.88	-0.87	0.95	0.00	-2.07
Mathematics	3	1189149	1.05	-0.72	-2.88	0.95	0.00	-3.76
Mathematics	3	1189150	1.04	-0.86	-1.53	0.98	0.00	-2.71
Mathematics	3	1189151	1.36	-1.28	-1.80	1.14	0.00	-3.34
Mathematics	3	1189152	1.09	-1.64	0.59	0.94	0.00	-1.34
Mathematics	3	1189153	0.99	-0.07	-0.72	1.00	0.00	-1.08
Mathematics	3	1189154	0.90	-0.72	-1.92	0.89	0.00	-2.90

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1189155	1.03	-0.96	-3.01	0.91	0.00	-4.15
Mathematics	3	1189156	1.01	-0.61	-1.40	0.99	0.00	-2.33
Mathematics	3	1189157	0.97	-0.49	-1.41	0.94	0.00	-2.21
Mathematics	3	1189158	0.92	-1.02	-1.41	0.95	0.00	-2.77
Mathematics	3	1189159	1.01	1.00	-4.27	0.93	0.00	-3.34
Mathematics	3	1189162	1.03	0.95	-0.53	0.99	0.00	0.17
Mathematics	3	1189163	1.05	0.87	-0.88	0.98	0.00	-0.28
Mathematics	3	1189164	1.10	-0.45	-3.16	1.04	0.00	-3.78
Mathematics	3	1189173	1.05	-0.27	-2.79	1.00	0.00	-3.24
Mathematics	3	1189176	1.13	-0.22	-1.78	1.10	0.00	-2.26
Mathematics	3	1189210	1.06	1.34	-2.22	0.94	0.00	-1.10
Mathematics	3	1189212	1.01	-0.82	-1.02	1.03	0.00	-2.16
Mathematics	3	1189213	0.99	0.64	-2.29	0.96	0.00	-1.88
Mathematics	3	1189219	0.93	-0.29	-2.10	0.93	0.00	-2.66
Mathematics	3	1189220	0.93	0.65	-3.51	0.95	0.00	-2.97
Mathematics	3	1189223	0.99	-0.34	-2.32	0.96	0.00	-2.89
Mathematics	3	1189224	1.08	0.26	-2.98	1.10	0.00	-2.86
Mathematics	3	1189226	0.94	0.10	-1.77	0.92	0.00	-1.95
Mathematics	3	1189227	1.05	0.75	-0.11	1.07	0.00	0.38
Mathematics	3	1189228	0.95	0.12	-0.29	0.95	0.00	-0.45
Mathematics	3	1189230	1.00	-0.05	1.90	0.97	0.00	1.56
Mathematics	3	1189233	1.01	-0.46	-1.67	0.97	0.00	-2.44
Mathematics	3	1189234	0.88	-0.06	-1.18	0.89	0.00	-1.55
Mathematics	3	1189235	1.08	-0.91	-2.08	0.90	0.00	-3.25
Mathematics	3	1189237	0.98	-0.60	-2.30	0.88	0.00	-3.14

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1189238	1.18	-0.33	-1.79	1.12	0.00	-2.38
Mathematics	3	1189240	1.07	-0.22	-1.09	1.09	0.00	-1.62
Mathematics	3	1189241	0.97	-0.39	-3.20	0.91	0.00	-3.74
Mathematics	3	1189242	1.45	-1.22	-0.80	1.23	0.00	-2.39
Mathematics	3	1189243	0.99	-0.50	-1.19	0.97	0.00	-2.01
Mathematics	3	1189244	0.95	-0.50	-1.18	0.93	0.00	-2.01
Mathematics	3	1189245	0.87	-0.55	-2.20	0.88	0.00	-3.01
Mathematics	3	1189248	1.08	-0.15	-1.45	1.06	0.00	-1.90
Mathematics	3	1189249	0.94	0.72	-1.63	0.97	0.00	-1.17
Mathematics	3	1189250	0.92	-0.11	-1.38	0.92	0.00	-1.79
Mathematics	3	1189251	1.02	-0.44	-1.41	0.98	0.00	-2.18
Mathematics	3	1189252	1.06	-0.33	-1.86	1.03	0.00	-2.44
Mathematics	3	1189261	1.07	0.83	-0.79	1.02	0.00	-0.23
Mathematics	3	1189279	1.05	-0.27	-3.15	1.02	0.00	-3.58
Mathematics	3	1189281	1.16	-0.25	-2.49	1.10	0.00	-2.94
Mathematics	3	1189284	1.00	0.86	-1.56	0.98	0.00	-0.98
Mathematics	3	1189288	0.98	0.74	-0.77	0.98	0.00	-0.27
Mathematics	3	1189291	1.00	1.01	-0.12	0.95	0.00	0.65
Mathematics	3	1189293	1.05	-0.24	-2.25	1.03	0.00	-2.74
Mathematics	3	1189294	1.26	-0.23	-0.98	1.22	0.00	-1.49
Mathematics	3	1189295	1.02	-0.23	-2.23	1.01	0.00	-2.72
Mathematics	3	1189301	1.07	1.10	-0.34	1.08	0.00	0.50
Mathematics	3	1189303	1.08	1.14	-0.87	1.03	0.00	0.01
Mathematics	3	1189306	1.05	1.02	-0.50	1.01	0.00	0.28
Mathematics	3	1189312	0.92	0.07	-1.16	0.93	0.00	-1.40

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1189314	1.15	1.65	-0.92	0.97	0.00	0.48
Mathematics	3	1189317	0.99	-0.08	-0.70	1.00	0.00	-1.08
Mathematics	3	1189318	1.66	-0.75	0.57	1.32	0.00	-0.54
Mathematics	3	1189319	1.11	-0.77	-2.34	0.97	0.00	-3.34
Mathematics	3	1189320	1.07	-0.35	-2.71	1.02	0.00	-3.26
Mathematics	3	1189321	0.96	-0.82	-2.14	0.96	0.00	-3.23
Mathematics	3	1189324	0.97	-0.20	-1.03	0.98	0.00	-1.53
Mathematics	3	1189325	0.95	0.21	-1.54	0.94	0.00	-1.64
Mathematics	3	1189326	0.94	-0.62	-1.19	0.89	0.00	-2.14
Mathematics	3	1189327	1.23	-1.12	-1.84	1.06	0.00	-3.21
Mathematics	3	1189329	1.31	-1.08	-1.28	1.07	0.00	-2.66
Mathematics	3	1189330	0.92	0.54	-3.82	0.90	0.00	-3.38
Mathematics	3	1189331	1.02	-0.31	-1.22	1.00	0.00	-1.86
Mathematics	3	1189336	1.69	-0.53	0.94	1.39	0.00	0.07
Mathematics	3	1189337	0.92	0.76	-0.49	1.02	0.00	0.02
Mathematics	3	1189339	1.04	-0.97	-1.41	0.92	0.00	-2.72
Mathematics	3	1189340	1.15	0.03	-0.33	1.16	0.00	-0.58
Mathematics	3	1189341	1.25	-0.26	-1.58	1.16	0.00	-2.10
Mathematics	3	1189342	1.10	-0.57	-0.57	1.03	0.00	-1.46
Mathematics	3	1189343	1.06	0.15	-0.24	1.08	0.00	-0.35
Mathematics	3	1189344	1.07	-0.65	-1.72	0.95	0.00	-2.65
Mathematics	3	1189345	0.99	-0.26	-1.17	1.00	0.00	-1.74
Mathematics	3	1189347	1.13	-0.39	-1.26	1.10	0.00	-1.97
Mathematics	3	1189350	0.86	-0.12	-1.95	0.86	0.00	-2.34
Mathematics	3	1189352	1.22	-0.37	-0.96	1.12	0.00	-1.63

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	3	1189355	1.09	0.71	-1.17	1.06	0.00	-0.74
Mathematics	3	1189357	1.09	-0.31	-2.30	1.04	0.00	-2.83
Mathematics	3	1189359	0.99	-0.20	-1.69	0.97	0.00	-2.18
Mathematics	3	1189368	1.03	0.75	-2.08	0.95	0.00	-1.59
Mathematics	3	1189369	1.19	-0.21	-1.45	1.13	0.00	-1.93
Mathematics	3	1189370	1.09	-0.24	-2.23	1.05	0.00	-2.69
Mathematics	3	1189375	0.93	0.71	-1.38	0.95	0.00	-0.95
Mathematics	4	1189378	1.11	-1.07	0.50	1.01	0.00	-0.75
Mathematics	4	1189380	1.16	-0.65	-0.62	0.98	0.00	-1.41
Mathematics	4	1189381	1.07	-0.92	-1.81	0.91	0.00	-2.83
Mathematics	4	1189382	0.95	-0.34	-0.73	0.94	0.00	-1.20
Mathematics	4	1189384	1.05	0.71	-0.98	1.06	0.00	-0.37
Mathematics	4	1189386	1.12	-0.34	0.83	1.05	0.00	0.31
Mathematics	4	1189387	1.05	-0.74	-2.03	0.97	0.00	-2.85
Mathematics	4	1189388	1.02	-0.43	-2.19	0.96	0.00	-2.68
Mathematics	4	1189389	1.04	-1.42	0.23	0.91	0.00	-1.35
Mathematics	4	1189390	1.07	-0.61	-0.47	0.91	0.00	-1.22
Mathematics	4	1189392	1.04	-0.48	-0.14	1.04	0.00	-0.78
Mathematics	4	1189393	1.01	-0.52	-1.20	0.97	0.00	-1.84
Mathematics	4	1189394	0.97	-1.36	0.26	0.86	0.00	-1.26
Mathematics	4	1189395	1.04	-0.52	-1.41	0.99	0.00	-2.04
Mathematics	4	1189396	1.04	-0.88	0.46	0.95	0.00	-0.59
Mathematics	4	1189398	1.19	-0.62	1.14	1.01	0.00	0.28
Mathematics	4	1189402	1.03	-0.60	-1.27	0.99	0.00	-1.99
Mathematics	4	1189404	1.00	0.50	-1.73	1.05	0.00	-1.30

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	4	1189405	0.89	-0.45	0.14	0.90	0.00	-0.47
Mathematics	4	1189406	0.92	-0.68	-2.11	0.95	0.00	-2.85
Mathematics	4	1189407	1.00	-0.61	-1.17	0.95	0.00	-1.91
Mathematics	4	1189408	1.39	-1.81	1.23	0.97	0.01	-0.77
Mathematics	4	1189409	1.03	-1.43	0.02	0.92	0.00	-1.57
Mathematics	4	1189410	0.99	-0.75	-0.64	0.98	0.00	-1.53
Mathematics	4	1189412	0.97	-0.47	-1.57	0.94	0.00	-2.15
Mathematics	4	1189413	1.02	-0.94	-0.09	0.95	0.00	-1.19
Mathematics	4	1189417	1.02	0.25	-2.19	1.04	0.00	-1.98
Mathematics	4	1189419	0.91	1.14	-1.14	0.93	0.00	-0.07
Mathematics	4	1189420	0.93	1.80	-1.86	1.01	0.00	-0.12
Mathematics	4	1189421	1.01	1.57	-1.46	0.97	0.00	0.05
Mathematics	4	1189422	0.97	0.23	-2.57	0.99	0.00	-2.35
Mathematics	4	1189424	1.00	0.19	-2.53	1.02	0.00	-2.36
Mathematics	4	1189425	0.97	1.54	-2.02	1.01	0.00	-0.53
Mathematics	4	1189426	0.91	1.42	-1.98	0.96	0.00	-0.60
Mathematics	4	1189427	1.03	-0.63	-2.81	0.99	0.00	-3.47
Mathematics	4	1189432	0.94	0.45	-2.68	0.98	0.00	-2.24
Mathematics	4	1189434	1.02	0.09	-2.53	1.02	0.00	-2.45
Mathematics	4	1189435	0.94	0.22	-2.51	0.96	0.00	-2.32
Mathematics	4	1189437	1.05	0.66	0.12	1.09	0.00	0.68
Mathematics	4	1189438	0.99	0.17	-1.17	1.01	0.00	-1.11
Mathematics	4	1189439	0.94	0.64	-1.18	0.97	0.00	-0.62
Mathematics	4	1189440	0.97	0.92	-1.09	0.99	0.00	-0.27
Mathematics	4	1189441	0.99	0.76	-0.97	0.98	0.00	-0.31

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	4	1189442	0.92	0.73	-0.89	0.95	0.00	-0.25
Mathematics	4	1189443	1.12	0.90	-0.76	1.09	0.00	0.03
Mathematics	4	1189444	1.00	0.43	-2.13	1.04	0.00	-1.74
Mathematics	4	1189447	1.00	0.54	-1.60	1.02	0.00	-1.14
Mathematics	4	1189448	1.06	-0.85	-0.17	1.02	0.00	-1.17
Mathematics	4	1189449	0.98	-1.05	0.53	0.88	0.00	-0.70
Mathematics	4	1189450	1.08	-1.35	0.14	0.97	0.00	-1.37
Mathematics	4	1189451	0.88	-0.55	0.31	0.87	0.00	-0.42
Mathematics	4	1189453	1.06	-0.72	-1.36	1.00	0.00	-2.19
Mathematics	4	1189454	1.13	-0.74	-1.47	1.02	0.00	-2.32
Mathematics	4	1189455	1.23	-0.59	-0.18	1.14	0.00	-0.94
Mathematics	4	1189456	1.08	-0.61	-1.62	0.95	0.00	-2.33
Mathematics	4	1189457	1.38	-1.12	0.87	1.07	0.00	-0.57
Mathematics	4	1189458	1.31	-1.10	-0.41	1.07	0.00	-1.70
Mathematics	4	1189460	0.99	-0.94	-0.29	0.94	0.00	-1.38
Mathematics	4	1189462	1.03	-0.45	-1.34	1.00	0.00	-1.89
Mathematics	4	1189464	1.01	-0.62	-1.06	0.97	0.00	-1.82
Mathematics	4	1189465	0.96	-0.45	-1.80	0.92	0.00	-2.34
Mathematics	4	1189468	0.93	0.59	-2.08	0.99	0.00	-1.53
Mathematics	4	1189469	0.98	0.51	-1.37	0.99	0.00	-0.95
Mathematics	4	1189473	0.99	0.55	-0.63	0.97	0.00	-0.20
Mathematics	4	1189476	0.96	0.84	-1.22	1.00	0.00	-0.47
Mathematics	4	1189478	1.02	-1.36	-0.58	0.98	0.00	-2.05
Mathematics	4	1189480	0.97	0.60	-0.38	0.94	0.00	0.09
Mathematics	4	1189481	1.02	-0.99	-1.49	0.98	0.00	-2.58

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	4	1189485	0.94	-0.63	-1.87	0.92	0.00	-2.58
Mathematics	4	1189486	0.98	-0.90	-1.01	0.97	0.00	-2.03
Mathematics	4	1189487	1.04	-0.54	-2.24	0.95	0.00	-2.83
Mathematics	4	1189488	0.89	-0.72	-1.85	0.91	0.00	-2.66
Mathematics	4	1189490	1.00	-0.79	-1.62	0.98	0.00	-2.51
Mathematics	4	1189492	0.94	-0.48	-1.96	0.92	0.00	-2.52
Mathematics	4	1189493	1.00	-0.47	-1.21	0.99	0.00	-1.79
Mathematics	4	1189495	0.90	-0.46	-2.39	0.92	0.00	-2.90
Mathematics	4	1189496	1.13	-0.51	-0.91	1.13	0.00	-1.55
Mathematics	4	1189498	0.95	-1.35	-0.49	0.90	0.00	-1.97
Mathematics	4	1189499	1.00	0.47	-2.01	0.99	0.00	-1.60
Mathematics	4	1189500	0.95	0.63	-2.83	0.96	0.00	-2.20
Mathematics	4	1189501	1.03	-0.24	0.01	1.03	0.00	-0.39
Mathematics	4	1189503	0.97	-1.14	-0.44	0.95	0.00	-1.72
Mathematics	4	1189504	0.98	-0.59	-1.67	0.97	0.00	-2.36
Mathematics	4	1189507	0.92	0.42	-2.90	0.94	0.00	-2.48
Mathematics	4	1189508	0.98	-0.54	-1.28	0.96	0.00	-1.93
Mathematics	4	1189509	0.98	-0.52	-0.21	0.98	0.00	-0.88
Mathematics	4	1189516	0.90	0.46	-2.93	0.92	0.00	-2.47
Mathematics	4	1189517	0.99	0.47	-3.69	0.92	0.00	-3.20
Mathematics	4	1189518	0.99	0.72	0.63	1.02	0.01	1.20
Mathematics	4	1189519	1.00	-0.40	-0.12	0.95	0.00	-0.68
Mathematics	4	1189520	0.98	-0.40	-2.10	0.91	0.00	-2.55
Mathematics	4	1189522	1.18	-0.64	-1.40	1.16	0.00	-2.14
Mathematics	4	1189524	1.01	-0.19	-0.38	1.02	0.00	-0.72

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	4	1189526	0.90	1.05	-0.08	1.01	0.00	0.85
Mathematics	4	1189528	0.95	0.77	-3.30	0.93	0.00	-2.52
Mathematics	4	1189529	1.13	0.70	-4.13	0.91	0.00	-3.40
Mathematics	4	1189546	0.89	-0.43	-2.97	0.91	0.00	-3.43
Mathematics	4	1189548	0.92	0.54	-2.26	1.00	0.00	-1.75
Mathematics	4	1189551	0.97	0.26	-2.67	0.96	0.00	-2.43
Mathematics	4	1189552	1.05	-0.49	-2.66	0.98	0.00	-3.18
Mathematics	4	1189553	1.19	0.60	-4.40	0.94	0.00	-3.77
Mathematics	4	1189555	0.97	-0.45	-2.44	0.97	0.00	-2.94
Mathematics	4	1189558	1.17	-0.48	1.17	1.04	0.00	0.51
Mathematics	4	1189560	0.94	-0.48	-0.50	0.93	0.00	-1.12
Mathematics	4	1189561	0.88	-0.33	-3.63	0.92	0.00	-3.95
Mathematics	4	1189562	1.03	-0.49	-0.33	0.98	0.00	-0.99
Mathematics	4	1189564	0.98	-0.59	-1.06	0.94	0.00	-1.77
Mathematics	4	1189569	1.12	-1.30	-1.32	1.01	0.00	-2.72
Mathematics	4	1189571	0.99	-0.59	-1.12	0.94	0.00	-1.84
Mathematics	4	1189572	1.10	-0.46	-0.92	1.06	0.00	-1.51
Mathematics	4	1189574	0.94	0.68	-2.89	0.96	0.00	-2.20
Mathematics	4	1189575	0.99	0.66	0.13	1.04	0.00	0.65
Mathematics	4	1189577	0.92	0.62	-0.56	0.98	0.00	-0.05
Mathematics	4	1189579	0.90	-0.40	-1.84	0.92	0.00	-2.34
Mathematics	4	1189583	1.03	-0.50	-1.49	1.04	0.00	-2.09
Mathematics	4	1189586	0.89	0.43	-2.86	0.91	0.00	-2.44
Mathematics	4	1189588	0.91	1.40	-1.83	0.99	0.00	-0.51
Mathematics	4	1189589	1.03	0.50	-3.85	0.92	0.00	-3.31

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	4	1189592	0.92	0.55	-2.82	0.92	0.00	-2.28
Mathematics	4	1189594	1.07	-0.39	-1.41	1.05	0.00	-1.90
Mathematics	4	1189595	0.98	-0.39	-1.73	0.97	0.00	-2.21
Mathematics	4	1189600	0.97	0.56	-3.58	0.91	0.00	-3.00
Mathematics	4	1189603	0.93	0.54	-2.86	0.94	0.00	-2.31
Mathematics	4	1189605	0.95	0.40	-2.81	0.96	0.00	-2.42
Mathematics	4	1189609	1.04	0.62	-1.20	1.01	0.00	-0.68
Mathematics	4	1189611	1.03	-0.40	-0.79	1.01	0.00	-1.32
Mathematics	4	1189612	1.19	-0.71	-2.40	1.11	0.00	-3.15
Mathematics	4	1189613	1.08	-0.97	-1.47	1.01	0.00	-2.55
Mathematics	4	1189616	1.29	0.55	-5.24	0.88	0.00	-4.65
Mathematics	4	1189620	1.02	1.16	0.23	0.98	0.00	1.26
Mathematics	4	1189621	1.16	0.70	-3.97	0.96	0.00	-3.25
Mathematics	4	1189623	0.92	0.57	-1.99	0.97	0.00	-1.46
Mathematics	4	1189625	0.90	0.76	-1.07	0.94	0.00	-0.40
Mathematics	4	1189627	0.93	-0.74	0.65	0.88	0.01	-0.29
Mathematics	4	1189631	1.10	0.87	-0.41	0.99	0.00	0.33
Mathematics	4	1189632	0.98	0.93	0.57	1.00	0.01	1.37
Mathematics	4	1189634	0.99	-0.37	-2.39	0.94	0.00	-2.79
Mathematics	4	1189635	1.05	-1.19	-0.52	0.94	0.00	-1.89
Mathematics	4	1189637	0.90	0.70	-0.22	0.98	0.00	0.37
Mathematics	4	1189638	0.99	-0.27	0.21	0.97	0.00	-0.24
Mathematics	4	1189640	1.13	-0.75	-2.64	1.06	0.00	-3.39
Mathematics	4	1189647	1.06	0.68	-0.03	1.02	0.00	0.51
Mathematics	4	1189651	1.09	0.57	-0.80	1.00	0.00	-0.37

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	4	1189652	0.90	0.88	0.08	0.97	0.00	0.83
Mathematics	4	1189654	0.92	0.69	-0.35	0.97	0.00	0.22
Mathematics	4	1189657	0.95	-0.54	-1.62	0.93	0.00	-2.25
Mathematics	4	1189658	1.08	-0.27	0.31	1.05	0.00	-0.12
Mathematics	4	1189662	0.89	0.69	0.05	0.96	0.00	0.61
Mathematics	4	1189664	1.27	0.91	-0.78	1.11	0.00	0.00
Mathematics	4	1189675	0.83	0.59	0.93	0.93	0.00	1.38
Mathematics	4	1189678	1.11	1.33	-3.20	1.02	0.00	-1.86
Mathematics	4	1189679	1.06	-0.87	-2.41	0.94	0.00	-3.35
Mathematics	4	1189680	1.18	-0.47	-0.99	1.07	0.00	-1.57
Mathematics	4	1189681	0.97	-0.78	-0.95	0.93	0.00	-1.87
Mathematics	4	1189683	1.10	0.96	-0.43	0.98	0.00	0.41
Mathematics	4	1189684	1.08	-0.45	1.06	1.02	0.01	0.43
Mathematics	4	1189691	1.08	0.54	-0.44	1.06	0.00	-0.03
Mathematics	4	1189692	0.89	0.61	-2.63	0.94	0.00	-2.03
Mathematics	4	1189694	1.54	-2.48	1.43	0.95	0.01	-1.23
Mathematics	4	1189697	1.00	0.71	0.17	0.98	0.00	0.74
Mathematics	4	1189699	0.90	-0.28	0.51	0.89	0.01	0.06
Mathematics	4	1189701	1.23	-0.38	-1.37	1.18	0.00	-1.86
Mathematics	4	1189702	1.17	1.13	-0.23	1.04	0.00	0.78
Mathematics	4	1189703	0.99	-0.42	-1.59	0.96	0.00	-2.12
Mathematics	4	1189704	1.01	-0.33	-1.24	0.97	0.00	-1.68
Mathematics	4	1189707	1.00	-0.56	0.66	0.97	0.01	-0.07
Mathematics	4	1189708	0.97	-0.24	-2.09	0.94	0.00	-2.39
Mathematics	4	1189710	1.10	-0.74	-0.44	1.04	0.00	-1.33

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	4	1189713	0.98	-0.82	-0.87	0.94	0.00	-1.82
Mathematics	4	1189715	0.97	0.00	-0.82	0.96	0.00	-0.96
Mathematics	4	1189717	1.04	-0.62	-0.38	1.01	0.00	-1.15
Mathematics	4	1189718	0.99	0.02	-2.72	0.99	0.00	-2.73
Mathematics	4	1189719	0.99	-0.35	-1.77	0.96	0.00	-2.23
Mathematics	4	1189722	0.96	-0.10	-2.52	0.94	0.00	-2.65
Mathematics	4	1189725	1.00	-0.16	0.81	0.99	0.01	0.49
Mathematics	4	1189726	1.07	-0.83	-0.45	0.96	0.00	-1.44
Mathematics	4	1189728	0.99	-0.45	-3.04	0.96	0.00	-3.50
Mathematics	4	1189729	1.10	-0.74	-1.88	0.98	0.00	-2.71
Mathematics	4	1189731	0.96	-0.21	-1.11	0.96	0.00	-1.45
Mathematics	4	1189732	0.95	-0.85	-0.19	0.90	0.00	-1.19
Mathematics	4	1189746	1.23	1.02	-0.37	1.14	0.00	0.53
Mathematics	4	1189749	0.95	-0.19	-0.08	0.94	0.00	-0.43
Mathematics	4	1189754	1.10	0.80	-0.69	1.14	0.00	0.01
Mathematics	4	1189755	1.34	0.16	-0.49	1.32	0.00	-0.46
Mathematics	4	1189758	1.17	1.02	-1.33	1.14	0.00	-0.41
Mathematics	4	1189759	0.96	0.72	0.61	0.98	0.01	1.20
Mathematics	4	1189762	1.00	0.21	-2.19	1.03	0.00	-2.02
Mathematics	4	1189770	1.13	0.55	-0.68	1.04	0.00	-0.27
Mathematics	4	1189773	1.10	0.86	-0.42	0.99	0.00	0.30
Mathematics	4	1189774	1.08	0.77	-0.16	1.01	0.00	0.48
Mathematics	4	1189780	0.92	0.60	-0.67	0.93	0.00	-0.18
Mathematics	4	1189782	0.91	0.49	-2.61	0.96	0.00	-2.12
Mathematics	4	1189784	0.91	0.73	-0.97	0.96	0.00	-0.33

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	4	1189790	0.99	-0.40	0.92	0.94	0.01	0.34
Mathematics	4	1189791	0.92	0.82	-1.87	1.01	0.00	-1.11
Mathematics	4	1189798	0.97	0.72	-1.72	0.96	0.00	-1.09
Mathematics	4	1189801	1.01	0.62	-1.55	0.98	0.00	-1.03
Mathematics	4	1189803	0.99	0.70	-0.11	0.92	0.00	0.46
Mathematics	4	1189806	1.06	0.52	-1.33	1.05	0.00	-0.92
Mathematics	4	1189808	1.11	0.83	-0.61	0.97	0.00	0.09
Mathematics	4	1189809	1.08	0.74	-0.42	0.98	0.00	0.19
Mathematics	4	1189812	1.02	0.80	0.08	0.99	0.00	0.74
Mathematics	5	1189816	0.96	-0.63	-2.33	0.90	0.00	-2.97
Mathematics	5	1189817	0.92	0.65	-2.80	0.88	0.00	-2.12
Mathematics	5	1189818	0.96	-0.94	-0.30	0.92	0.00	-1.32
Mathematics	5	1189819	0.98	-1.01	-0.08	0.98	0.00	-1.18
Mathematics	5	1189820	0.89	-0.52	-0.14	0.89	0.00	-0.76
Mathematics	5	1189822	0.96	-0.46	-0.18	0.96	0.00	-0.73
Mathematics	5	1189823	0.93	-0.38	1.10	0.90	0.00	0.60
Mathematics	5	1189828	0.98	-0.31	0.13	0.98	0.00	-0.30
Mathematics	5	1189830	0.93	-0.96	0.14	0.94	0.00	-0.93
Mathematics	5	1189831	0.95	0.04	-1.99	0.95	0.00	-1.96
Mathematics	5	1189832	1.03	-0.98	-0.56	0.95	0.00	-1.63
Mathematics	5	1189833	1.01	-0.69	-1.67	0.91	0.00	-2.41
Mathematics	5	1189834	0.91	-0.69	0.82	0.88	0.00	0.00
Mathematics	5	1189835	1.15	-0.49	1.98	1.02	0.00	1.36
Mathematics	5	1189837	1.02	-1.02	-0.56	0.94	0.00	-1.67
Mathematics	5	1189839	1.01	-1.12	0.11	0.98	0.00	-1.12

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	5	1189841	0.88	-1.01	-1.38	0.93	0.00	-2.43
Mathematics	5	1189845	0.97	-0.64	-1.03	0.93	0.00	-1.74
Mathematics	5	1189846	1.05	-0.50	-0.76	0.95	0.00	-1.34
Mathematics	5	1189848	1.09	-0.64	-0.45	1.03	0.00	-1.18
Mathematics	5	1189849	0.97	0.40	-2.59	0.97	0.00	-2.18
Mathematics	5	1189854	1.08	-0.87	-0.93	0.99	0.00	-1.85
Mathematics	5	1189855	0.93	-0.52	0.66	0.92	0.00	0.01
Mathematics	5	1189856	1.01	-0.85	-1.07	0.95	0.00	-1.99
Mathematics	5	1189857	0.94	-0.56	-0.12	0.94	0.00	-0.77
Mathematics	5	1189858	0.98	-0.47	-0.42	0.96	0.00	-0.96
Mathematics	5	1189859	1.09	-0.61	-1.55	1.00	0.00	-2.21
Mathematics	5	1189860	0.96	-1.17	0.88	0.88	0.00	-0.41
Mathematics	5	1189862	0.97	-1.20	-0.05	0.95	0.00	-1.35
Mathematics	5	1189863	0.89	-0.83	0.29	0.90	0.00	-0.66
Mathematics	5	1189864	0.92	-0.33	-0.26	0.92	0.00	-0.68
Mathematics	5	1189871	0.90	0.15	-1.28	0.92	0.00	-1.17
Mathematics	5	1189872	0.99	-0.40	0.02	0.99	0.00	-0.49
Mathematics	5	1189873	0.92	0.45	-2.85	0.89	0.00	-2.37
Mathematics	5	1189874	1.01	-0.24	-0.55	1.01	0.00	-0.87
Mathematics	5	1189875	0.98	-0.72	-0.16	0.97	0.00	-0.98
Mathematics	5	1189876	0.98	-0.49	-0.52	0.95	0.00	-1.10
Mathematics	5	1189877	1.00	-0.48	-0.50	0.97	0.00	-1.06
Mathematics	5	1189878	0.98	-0.79	-0.45	0.91	0.00	-1.32
Mathematics	5	1189881	1.01	0.18	-1.10	1.01	0.00	-0.97
Mathematics	5	1189884	0.89	-0.45	-1.92	0.93	0.00	-2.40

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	5	1189885	0.99	0.06	-1.44	0.99	0.00	-1.42
Mathematics	5	1189889	0.95	-0.35	0.11	0.95	0.00	-0.35
Mathematics	5	1189891	1.02	-0.65	-1.29	0.94	0.00	-2.01
Mathematics	5	1189895	1.02	0.45	-3.24	0.95	0.00	-2.76
Mathematics	5	1189899	0.97	-0.04	-1.72	0.96	0.00	-1.79
Mathematics	5	1189901	1.03	0.53	-1.26	0.96	0.00	-0.80
Mathematics	5	1189902	1.01	-0.43	-2.29	0.97	0.00	-2.73
Mathematics	5	1189903	0.95	0.48	-1.87	0.87	0.00	-1.44
Mathematics	5	1189905	1.05	-0.43	-0.94	1.01	0.00	-1.43
Mathematics	5	1189907	1.02	-0.72	1.43	0.90	0.00	0.58
Mathematics	5	1189909	0.89	-1.19	0.17	0.91	0.00	-1.13
Mathematics	5	1189910	0.98	-0.23	0.75	0.98	0.00	0.40
Mathematics	5	1189911	0.93	-1.42	0.14	0.94	0.00	-1.41
Mathematics	5	1189912	1.01	-0.92	-1.29	0.89	0.00	-2.30
Mathematics	5	1189913	0.91	-1.52	-0.02	0.87	0.00	-1.65
Mathematics	5	1189914	0.88	-1.13	0.27	0.92	0.00	-0.97
Mathematics	5	1189919	1.01	-0.81	-2.21	0.93	0.00	-3.04
Mathematics	5	1189920	1.02	-0.54	-0.57	0.97	0.00	-1.19
Mathematics	5	1189921	1.04	-0.37	-0.97	1.02	0.00	-1.40
Mathematics	5	1189924	1.03	-0.38	-1.28	0.99	0.00	-1.72
Mathematics	5	1189935	1.01	-0.28	-2.01	0.98	0.00	-2.31
Mathematics	5	1189945	0.97	0.60	-2.85	0.93	0.00	-2.23
Mathematics	5	1189951	1.02	0.06	-1.86	1.02	0.00	-1.82
Mathematics	5	1189952	0.96	0.56	-0.09	0.93	0.00	0.38
Mathematics	5	1189953	1.14	0.97	-0.41	1.04	0.00	0.49

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	5	1189960	0.89	-0.56	0.34	0.93	0.00	-0.33
Mathematics	5	1189961	1.02	-0.53	0.13	1.03	0.00	-0.52
Mathematics	5	1189967	0.96	0.46	-1.12	0.94	0.00	-0.71
Mathematics	5	1189968	0.98	-0.74	0.32	0.98	0.00	-0.54
Mathematics	5	1189970	1.13	-0.45	-0.82	1.08	0.00	-1.33
Mathematics	5	1189971	0.94	0.47	-2.94	0.91	0.00	-2.45
Mathematics	5	1189973	1.03	0.47	-2.59	0.96	0.00	-2.15
Mathematics	5	1189974	0.98	0.47	-2.43	0.86	0.00	-2.00
Mathematics	5	1189980	0.99	-1.05	0.12	0.98	0.00	-1.04
Mathematics	5	1189981	0.94	0.47	-2.71	0.91	0.00	-2.23
Mathematics	5	1189983	1.17	0.76	0.43	1.11	0.00	1.09
Mathematics	5	1189984	0.94	0.56	-0.45	0.98	0.00	0.06
Mathematics	5	1189986	1.24	-0.36	-0.82	1.17	0.00	-1.24
Mathematics	5	1189990	1.14	-0.39	-0.90	1.07	0.00	-1.35
Mathematics	5	1189991	1.08	-0.43	-1.15	1.00	0.00	-1.64
Mathematics	5	1189992	0.88	-0.52	-0.09	0.87	0.00	-0.71
Mathematics	5	1189996	1.07	-0.79	-1.91	0.97	0.00	-2.74
Mathematics	5	1189997	1.08	0.57	0.03	1.06	0.00	0.51
Mathematics	5	1190000	1.05	-0.41	-1.61	0.98	0.00	-2.07
Mathematics	5	1190009	1.08	0.74	0.12	1.05	0.00	0.77
Mathematics	5	1190012	1.11	-0.47	-1.59	1.03	0.00	-2.10
Mathematics	5	1190017	0.89	0.82	1.61	0.97	0.00	2.34
Mathematics	5	1190019	1.06	0.64	-0.14	1.00	0.00	0.40
Mathematics	5	1190023	0.91	0.52	-2.04	0.94	0.00	-1.53
Mathematics	5	1190026	0.87	0.64	-1.84	0.95	0.00	-1.21

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	5	1190029	0.95	0.62	-1.38	0.97	0.00	-0.81
Mathematics	5	1190031	1.18	0.82	0.03	1.07	0.00	0.75
Mathematics	5	1190032	0.96	0.54	0.83	0.97	0.00	1.25
Mathematics	5	1190034	1.09	-0.66	-1.64	0.98	0.00	-2.36
Mathematics	5	1190036	0.93	0.54	-2.16	0.95	0.00	-1.62
Mathematics	5	1190038	0.92	0.74	-2.68	0.92	0.00	-1.91
Mathematics	5	1190040	1.17	0.63	-0.44	1.05	0.00	0.10
Mathematics	5	1190041	1.03	0.55	-0.38	1.04	0.00	0.09
Mathematics	5	1190047	1.16	1.02	-0.95	1.06	0.00	0.01
Mathematics	5	1190054	0.97	0.70	-2.13	0.98	0.00	-1.44
Mathematics	5	1190055	1.09	-0.34	-0.81	1.05	0.00	-1.22
Mathematics	5	1190057	1.09	0.77	-0.86	1.05	0.00	-0.16
Mathematics	5	1190058	1.11	0.47	-0.98	1.07	0.00	-0.59
Mathematics	5	1190066	1.20	0.53	-0.90	1.17	0.00	-0.43
Mathematics	5	1190067	1.09	0.76	0.23	1.01	0.00	0.88
Mathematics	5	1190069	1.02	0.55	-1.26	1.02	0.00	-0.77
Mathematics	5	1190071	1.07	0.72	-0.66	1.03	0.00	0.00
Mathematics	5	1190083	1.05	-0.28	-1.41	0.98	0.00	-1.74
Mathematics	5	1190084	1.05	-0.99	1.04	1.03	0.00	-0.11
Mathematics	5	1190086	0.94	-0.40	-1.07	0.89	0.00	-1.53
Mathematics	5	1190089	0.87	0.50	-2.30	0.90	0.00	-1.79
Mathematics	5	1190091	1.26	0.71	-0.52	1.15	0.00	0.10
Mathematics	5	1190093	0.92	0.75	-2.29	0.95	0.00	-1.53
Mathematics	5	1190094	1.06	-0.77	-0.27	0.99	0.00	-1.15
Mathematics	5	1190098	0.87	0.53	-1.81	0.93	0.00	-1.30

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	5	1190102	0.97	-0.36	-0.34	0.92	0.00	-0.78
Mathematics	5	1190103	1.06	0.81	0.02	1.02	0.00	0.73
Mathematics	5	1190104	1.03	0.72	0.21	0.96	0.00	0.83
Mathematics	5	1190109	1.03	0.70	0.09	0.97	0.00	0.69
Mathematics	5	1190110	1.04	0.61	0.03	1.01	0.00	0.54
Mathematics	5	1190112	0.89	0.61	-1.22	0.91	0.00	-0.66
Mathematics	5	1190113	0.98	-0.53	0.28	0.98	0.00	-0.37
Mathematics	5	1190121	0.92	0.61	-2.14	0.94	0.00	-1.54
Mathematics	5	1190122	1.34	-0.64	1.08	1.30	0.00	0.26
Mathematics	5	1190127	0.86	0.61	-1.96	0.93	0.00	-1.35
Mathematics	5	1190128	0.98	0.25	-2.29	0.99	0.00	-2.04
Mathematics	5	1190130	0.91	0.61	-2.11	0.94	0.00	-1.51
Mathematics	5	1190137	1.05	0.56	-1.01	1.00	0.00	-0.52
Mathematics	5	1190143	1.07	0.71	-0.23	0.95	0.00	0.37
Mathematics	5	1190145	1.02	-0.47	-1.01	0.95	0.00	-1.54
Mathematics	5	1190154	1.10	-0.79	-0.26	0.96	0.00	-1.16
Mathematics	5	1190155	0.98	0.55	-2.42	1.01	0.00	-1.85
Mathematics	5	1190158	1.03	0.70	-2.87	0.93	0.00	-2.15
Mathematics	5	1190159	1.08	-0.37	-1.31	1.01	0.00	-1.74
Mathematics	5	1190164	0.96	0.75	-2.26	0.99	0.00	-1.50
Mathematics	5	1190166	1.10	-0.36	-1.04	1.05	0.00	-1.47
Mathematics	5	1190167	0.99	-0.41	0.10	0.97	0.00	-0.42
Mathematics	5	1190168	0.97	-0.09	-0.06	0.96	0.00	-0.24
Mathematics	5	1190169	1.15	-0.64	0.82	1.16	0.00	0.02
Mathematics	5	1190171	1.03	-0.47	0.88	1.03	0.00	0.26

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	5	1190178	0.96	0.45	-2.39	0.98	0.00	-1.94
Mathematics	5	1190180	1.24	0.79	-0.48	1.13	0.00	0.23
Mathematics	5	1190183	1.13	-0.67	-0.03	1.04	0.00	-0.83
Mathematics	5	1190184	1.06	-0.38	-0.23	1.01	0.00	-0.71
Mathematics	5	1190185	1.06	0.75	0.09	0.99	0.00	0.74
Mathematics	5	1190186	1.15	0.57	-0.45	1.05	0.00	0.02
Mathematics	5	1190195	0.92	0.49	-0.20	0.94	0.00	0.22
Mathematics	5	1190197	1.01	0.53	-2.99	0.92	0.00	-2.44
Mathematics	5	1190201	1.13	0.51	0.34	1.09	0.00	0.75
Mathematics	6	1190211	1.02	-0.70	-0.23	0.95	0.00	-1.01
Mathematics	6	1190212	1.04	0.44	-2.78	0.95	0.00	-2.31
Mathematics	6	1190213	0.93	-0.63	-1.05	0.92	0.00	-1.76
Mathematics	6	1190216	1.05	-0.59	-0.27	0.98	0.00	-0.93
Mathematics	6	1190222	0.94	0.39	-2.22	0.91	0.00	-1.82
Mathematics	6	1190226	0.92	0.70	1.61	1.04	0.00	2.43
Mathematics	6	1190234	0.86	0.92	1.90	0.89	0.00	2.97
Mathematics	6	1190235	0.88	0.00	1.67	0.87	0.00	1.78
Mathematics	6	1190238	0.99	-0.36	-0.48	0.99	0.00	-0.92
Mathematics	6	1190239	0.99	-0.45	-0.66	0.98	0.00	-1.20
Mathematics	6	1190240	0.99	-0.57	-0.03	1.00	0.00	-0.67
Mathematics	6	1190241	1.05	0.40	0.29	1.07	0.00	0.71
Mathematics	6	1190244	0.97	-0.07	-1.50	0.96	0.00	-1.60
Mathematics	6	1190248	0.95	-0.46	0.20	0.91	0.00	-0.32
Mathematics	6	1190249	0.95	-0.42	-0.03	0.92	0.00	-0.54
Mathematics	6	1190252	0.96	-0.46	0.23	0.97	0.00	-0.26

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	6	1190255	0.99	-0.34	-0.97	0.97	0.00	-1.40
Mathematics	6	1190257	1.10	-0.43	-0.25	1.06	0.00	-0.76
Mathematics	6	1190259	1.02	-0.49	-1.24	0.96	0.00	-1.79
Mathematics	6	1190263	0.96	-0.39	-1.09	0.94	0.00	-1.56
Mathematics	6	1190264	0.96	-0.81	0.32	0.94	0.00	-0.54
Mathematics	6	1190267	1.05	-1.46	1.02	0.91	0.00	-0.41
Mathematics	6	1190284	1.38	-1.23	0.18	1.05	0.00	-1.17
Mathematics	6	1190285	0.92	-0.84	0.20	0.90	0.00	-0.68
Mathematics	6	1190286	0.92	-1.04	-0.04	0.88	0.00	-1.15
Mathematics	6	1190287	0.99	-0.33	-1.02	0.97	0.00	-1.43
Mathematics	6	1190288	1.06	-0.53	0.07	1.03	0.00	-0.54
Mathematics	6	1190289	0.96	-1.11	0.60	0.93	0.00	-0.50
Mathematics	6	1190290	0.96	-0.53	0.20	0.94	0.00	-0.38
Mathematics	6	1190291	0.95	-0.38	0.55	0.97	0.00	0.18
Mathematics	6	1190294	0.92	-0.50	-0.86	0.90	0.00	-1.46
Mathematics	6	1190296	1.00	-0.63	-0.34	0.94	0.00	-1.06
Mathematics	6	1190303	0.97	-0.71	-0.61	0.90	0.00	-1.44
Mathematics	6	1190304	0.99	-0.44	-1.21	0.95	0.00	-1.72
Mathematics	6	1190305	1.15	-1.01	0.33	0.91	0.00	-0.78
Mathematics	6	1190307	1.07	0.61	0.78	1.09	0.00	1.47
Mathematics	6	1190310	1.04	0.55	-0.09	1.14	0.00	0.42
Mathematics	6	1190323	1.10	0.27	-0.10	1.15	0.00	0.11
Mathematics	6	1190325	0.93	0.52	-0.04	0.93	0.00	0.42
Mathematics	6	1190326	0.89	0.63	-1.31	0.91	0.00	-0.74
Mathematics	6	1190327	0.90	0.89	-1.87	0.95	0.00	-0.98

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	6	1190329	0.93	0.43	0.75	0.93	0.00	1.24
Mathematics	6	1190330	0.96	-0.14	0.59	0.97	0.00	0.49
Mathematics	6	1190331	0.94	-0.56	0.14	0.95	0.00	-0.47
Mathematics	6	1190333	0.94	-0.33	1.36	0.94	0.00	1.12
Mathematics	6	1190334	0.93	-0.33	0.40	0.95	0.00	0.08
Mathematics	6	1190337	1.00	0.38	1.54	1.02	0.00	2.03
Mathematics	6	1190338	0.93	-0.49	0.15	0.96	0.00	-0.38
Mathematics	6	1190343	0.96	-0.73	1.84	0.90	0.00	1.20
Mathematics	6	1190344	0.94	0.52	-2.15	0.96	0.00	-1.62
Mathematics	6	1190345	0.91	0.29	0.11	0.94	0.00	0.37
Mathematics	6	1190347	0.90	0.53	-0.89	0.92	0.00	-0.43
Mathematics	6	1190349	0.96	0.36	1.15	0.97	0.00	1.59
Mathematics	6	1190352	1.00	0.47	1.56	1.01	0.00	2.14
Mathematics	6	1190353	0.94	-0.27	-1.62	0.92	0.00	-1.94
Mathematics	6	1190354	0.97	-0.45	-0.47	0.95	0.00	-1.02
Mathematics	6	1190355	1.15	-0.63	-0.62	1.04	0.00	-1.34
Mathematics	6	1190357	1.07	-0.39	-0.95	0.99	0.00	-1.41
Mathematics	6	1190359	1.06	-0.50	-0.93	0.98	0.00	-1.51
Mathematics	6	1190360	0.92	0.84	-1.12	1.00	0.00	-0.31
Mathematics	6	1190362	0.93	0.92	1.05	0.94	0.00	2.07
Mathematics	6	1190363	0.96	0.58	-0.80	1.04	0.00	-0.27
Mathematics	6	1190364	1.20	0.71	-0.06	1.28	0.00	0.62
Mathematics	6	1190372	1.12	-0.46	-0.43	1.03	0.00	-0.99
Mathematics	6	1190376	0.96	0.53	0.56	0.95	0.00	1.15
Mathematics	6	1190377	0.99	0.63	0.48	0.98	0.00	1.16

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	6	1190378	0.93	0.43	-2.08	0.94	0.00	-1.64
Mathematics	6	1190379	0.99	0.60	0.25	0.97	0.00	0.87
Mathematics	6	1190380	0.96	0.64	-2.33	0.96	0.00	-1.66
Mathematics	6	1190381	1.02	0.67	0.38	1.02	0.00	1.08
Mathematics	6	1190383	0.97	0.39	1.07	0.98	0.00	1.54
Mathematics	6	1190384	1.04	-0.85	-0.27	0.95	0.00	-1.20
Mathematics	6	1190387	0.99	-0.64	-0.31	0.94	0.00	-1.03
Mathematics	6	1190388	1.01	-0.36	0.20	0.99	0.00	-0.22
Mathematics	6	1190398	0.94	0.51	-2.46	0.93	0.00	-1.92
Mathematics	6	1190399	1.09	-1.23	0.68	0.92	0.00	-0.58
Mathematics	6	1190402	1.02	-0.89	-1.07	0.95	0.00	-2.07
Mathematics	6	1190403	1.02	-0.86	-0.30	0.93	0.00	-1.25
Mathematics	6	1190404	1.03	0.81	0.53	1.01	0.00	1.40
Mathematics	6	1190407	0.95	0.43	-2.28	0.95	0.00	-1.84
Mathematics	6	1190412	0.99	-0.50	0.47	0.98	0.00	-0.04
Mathematics	6	1190413	1.10	-0.32	0.91	1.10	0.00	0.63
Mathematics	6	1190417	0.97	0.56	1.00	1.01	0.00	1.64
Mathematics	6	1190418	0.93	0.32	1.57	0.93	0.00	2.01
Mathematics	6	1190421	1.05	-0.51	-0.54	1.01	0.00	-1.16
Mathematics	6	1190426	0.94	0.53	0.49	0.97	0.00	1.04
Mathematics	6	1190431	1.30	-1.42	-0.58	0.99	0.00	-2.13
Mathematics	6	1190432	1.05	0.44	-2.90	0.91	0.00	-2.44
Mathematics	6	1190435	0.99	-0.47	0.20	0.95	0.00	-0.33
Mathematics	6	1190437	1.32	-0.81	-0.26	1.25	0.00	-1.17
Mathematics	6	1190438	1.00	-0.83	0.40	0.92	0.00	-0.48

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	6	1190439	1.05	-0.54	-1.26	0.94	0.00	-1.85
Mathematics	6	1190441	0.94	0.44	1.24	1.00	0.00	1.77
Mathematics	6	1190442	1.01	-0.58	-0.79	0.94	0.00	-1.47
Mathematics	6	1190444	1.01	-0.75	-0.84	0.91	0.00	-1.69
Mathematics	6	1190445	0.98	-0.44	-0.88	0.92	0.00	-1.41
Mathematics	6	1190447	0.97	-0.51	1.20	0.90	0.00	0.74
Mathematics	6	1190448	1.02	-0.80	0.96	0.95	0.00	0.19
Mathematics	6	1190450	1.15	-0.84	-0.23	1.02	0.00	-1.16
Mathematics	6	1190451	1.01	0.46	-0.26	1.07	0.00	0.17
Mathematics	6	1190452	1.03	0.50	0.47	1.05	0.00	1.01
Mathematics	6	1190453	0.88	0.49	-1.78	0.90	0.00	-1.29
Mathematics	6	1190455	1.15	-0.40	-0.51	1.05	0.00	-1.00
Mathematics	6	1190458	0.93	0.68	-0.69	0.98	0.00	-0.07
Mathematics	6	1190459	1.05	-0.62	1.41	0.97	0.00	0.86
Mathematics	6	1190463	1.03	-0.50	0.44	1.02	0.00	-0.08
Mathematics	6	1190464	1.22	-0.70	0.14	1.08	0.00	-0.65
Mathematics	6	1190468	1.02	-0.80	0.15	0.93	0.00	-0.73
Mathematics	6	1190470	1.14	-0.37	1.56	1.09	0.00	1.26
Mathematics	6	1190471	1.07	0.41	-0.30	1.12	0.00	0.08
Mathematics	6	1190475	1.12	0.68	0.85	1.18	0.00	1.60
Mathematics	6	1190482	1.01	0.52	-1.23	1.04	0.00	-0.74
Mathematics	6	1190489	1.12	-0.91	-0.39	0.93	0.00	-1.42
Mathematics	6	1190490	1.02	-1.06	0.82	0.91	0.00	-0.24
Mathematics	6	1190491	1.05	-1.18	-0.52	0.96	0.00	-1.84
Mathematics	6	1190492	0.99	-0.43	-0.14	0.94	0.00	-0.66

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	6	1190493	0.90	-0.61	1.39	0.90	0.00	0.87
Mathematics	6	1190494	1.11	-0.88	1.08	1.01	0.00	0.22
Mathematics	6	1190495	1.03	-0.57	-1.22	0.95	0.00	-1.86
Mathematics	6	1190496	1.09	-0.43	0.09	1.03	0.00	-0.41
Mathematics	6	1190498	1.01	-1.41	1.18	0.92	0.00	-0.17
Mathematics	6	1190500	0.95	-0.30	1.07	0.93	0.00	0.80
Mathematics	6	1190501	1.00	-0.44	-0.61	0.92	0.00	-1.14
Mathematics	6	1190503	0.95	-0.48	0.87	0.94	0.00	0.42
Mathematics	6	1190504	1.13	-1.01	-0.31	1.00	0.00	-1.42
Mathematics	6	1190506	0.99	-0.46	-0.48	0.96	0.00	-1.04
Mathematics	6	1190507	1.01	0.49	0.78	1.03	0.00	1.33
Mathematics	6	1190510	1.02	0.58	-2.30	1.00	0.00	-1.71
Mathematics	6	1190513	0.97	0.71	-1.82	1.00	0.00	-1.13
Mathematics	6	1190514	0.89	0.92	0.89	0.96	0.00	1.89
Mathematics	6	1190515	1.07	-1.18	-1.46	0.99	0.00	-2.72
Mathematics	6	1190516	0.92	-0.19	-2.62	0.94	0.00	-2.81
Mathematics	6	1190517	0.88	0.41	1.73	0.88	0.00	2.28
Mathematics	6	1190518	0.86	0.91	0.98	0.92	0.00	1.98
Mathematics	6	1190519	1.03	-0.43	-0.95	0.97	0.00	-1.47
Mathematics	6	1190520	1.07	-0.39	-0.45	1.00	0.00	-0.94
Mathematics	6	1190523	0.95	0.40	1.22	0.97	0.00	1.72
Mathematics	6	1190526	1.19	-0.95	-0.10	1.07	0.00	-1.16
Mathematics	6	1190527	1.06	-0.71	-1.01	0.97	0.00	-1.82
Mathematics	6	1190532	0.92	0.56	0.37	0.95	0.00	0.91
Mathematics	6	1190536	0.98	0.70	1.07	0.97	0.00	1.87

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	6	1190538	0.90	0.53	0.89	0.95	0.00	1.48
Mathematics	6	1190543	0.95	0.62	0.06	0.99	0.00	0.63
Mathematics	6	1190545	0.96	0.84	1.02	0.94	0.00	1.96
Mathematics	6	1190546	1.13	0.72	-0.25	1.20	0.00	0.43
Mathematics	6	1190549	0.93	0.59	-0.63	0.96	0.00	-0.13
Mathematics	6	1190550	0.96	0.85	-0.61	0.98	0.00	0.16
Mathematics	6	1190552	0.92	0.50	-2.04	0.92	0.00	-1.54
Mathematics	6	1190555	0.92	0.64	0.53	0.92	0.00	1.20
Mathematics	6	1190556	0.96	0.61	-0.02	0.99	0.00	0.52
Mathematics	6	1190557	0.90	1.01	2.32	0.98	0.00	3.53
Mathematics	6	1190558	1.00	0.52	1.91	1.03	0.00	2.58
Mathematics	6	1190559	0.98	0.65	1.42	0.98	0.00	2.20
Mathematics	6	1190562	1.13	0.74	-0.34	1.20	0.00	0.34
Mathematics	6	1190563	0.99	0.59	0.95	1.00	0.00	1.62
Mathematics	6	1190566	0.98	0.50	-0.79	1.04	0.00	-0.36
Mathematics	6	1190568	0.95	-0.12	1.79	0.95	0.00	1.77
Mathematics	6	1190570	0.95	0.72	-1.20	1.02	0.00	-0.54
Mathematics	6	1190575	0.95	0.46	0.32	0.96	0.00	0.77
Mathematics	6	1190577	0.93	0.49	0.11	0.95	0.00	0.54
Mathematics	6	1190580	0.92	0.34	1.34	0.92	0.00	1.80
Mathematics	6	1190582	0.95	0.56	0.08	0.98	0.00	0.60
Mathematics	6	1190584	0.95	-0.67	1.17	0.89	0.00	0.57
Mathematics	6	1190588	0.97	0.41	0.59	0.97	0.00	1.04
Mathematics	6	1190591	1.03	-0.59	0.05	0.98	0.00	-0.64
Mathematics	6	1190592	1.05	-2.41	2.03	0.96	0.00	-0.24

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	6	1190593	1.03	-2.56	1.92	0.96	0.00	-0.50
Mathematics	6	1190594	0.93	-1.53	1.51	0.94	0.00	0.09
Mathematics	6	1190596	0.92	0.61	0.16	0.95	0.00	0.72
Mathematics	6	1190597	0.92	0.69	0.42	0.95	0.00	1.12
Mathematics	6	1190598	1.02	0.90	1.07	1.00	0.00	2.07
Mathematics	6	1190601	1.00	0.80	0.90	1.02	0.00	1.78
Mathematics	6	1190603	0.91	0.64	1.23	0.93	0.00	1.97
Mathematics	6	1190604	1.06	1.00	1.12	1.07	0.00	2.23
Mathematics	6	1190609	0.92	0.54	2.04	0.93	0.00	2.74
Mathematics	6	1190671	0.98	0.50	1.15	0.98	0.00	1.75
Mathematics	6	1190673	0.94	0.51	-0.64	0.98	0.00	-0.22
Mathematics	6	1190674	0.98	0.50	0.06	1.02	0.00	0.50
Mathematics	6	1190676	0.89	0.70	-0.71	0.96	0.00	-0.08
Mathematics	6	1190679	1.02	1.08	1.16	0.99	0.00	2.35
Mathematics	6	1190680	0.90	0.62	-1.61	0.94	0.00	-1.01
Mathematics	6	1190683	0.96	0.61	0.46	0.98	0.00	1.09
Mathematics	6	1190685	0.95	0.80	0.39	0.99	0.00	1.18
Mathematics	6	1190686	0.95	-0.38	0.60	0.96	0.00	0.25
Mathematics	6	1190687	1.03	-0.33	-0.10	1.01	0.00	-0.54
Mathematics	6	1190688	0.97	0.47	-1.05	1.02	0.00	-0.64
Mathematics	6	1190689	1.05	0.89	1.65	1.07	0.00	2.67
Mathematics	6	1190691	1.00	-0.15	1.30	1.01	0.00	1.24
Mathematics	6	1190692	0.97	0.68	-0.41	1.00	0.00	0.20
Mathematics	6	1190693	0.95	0.85	-0.51	0.99	0.00	0.26
Mathematics	6	1190694	0.94	0.76	-1.14	1.01	0.00	-0.43

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	6	1190695	0.95	0.95	-0.19	0.98	0.00	0.72
Mathematics	6	1190696	0.95	0.46	-0.81	1.00	0.00	-0.42
Mathematics	6	1190697	1.02	-0.33	0.09	1.00	0.00	-0.32
Mathematics	6	1190698	0.97	0.61	0.86	0.98	0.00	1.54
Mathematics	6	1190699	0.99	0.51	1.52	0.99	0.00	2.14
Mathematics	6	1190700	0.96	0.58	-0.83	1.02	0.00	-0.32
Mathematics	6	1190703	1.10	1.24	1.28	1.06	0.00	2.64
Mathematics	6	1190704	0.95	0.62	0.10	0.98	0.00	0.66
Mathematics	6	1190705	1.01	1.01	0.55	1.02	0.00	1.62
Mathematics	6	1190709	0.93	0.69	1.69	0.92	0.00	2.50
Mathematics	7	1190715	0.90	0.51	-1.18	0.95	0.00	-0.64
Mathematics	7	1190716	1.01	0.06	0.99	1.01	0.00	1.04
Mathematics	7	1190718	0.95	-0.52	-0.09	1.00	0.00	-0.70
Mathematics	7	1190719	0.97	-0.16	0.57	0.99	0.00	0.34
Mathematics	7	1190725	1.08	0.44	-2.69	0.94	0.00	-2.16
Mathematics	7	1190732	1.02	-0.45	-0.01	1.06	0.00	-0.54
Mathematics	7	1190738	0.86	-0.48	-0.07	0.92	0.00	-0.64
Mathematics	7	1190742	1.09	-0.52	0.19	1.05	0.00	-0.43
Mathematics	7	1190750	0.95	-0.36	0.26	0.92	0.00	-0.21
Mathematics	7	1190752	0.98	-0.43	0.33	0.95	0.00	-0.22
Mathematics	7	1190753	1.10	0.54	0.79	1.02	0.00	1.33
Mathematics	7	1190762	0.92	-0.33	0.69	0.97	0.00	0.31
Mathematics	7	1190764	0.95	-0.38	0.53	0.93	0.00	0.04
Mathematics	7	1190770	0.93	-0.46	0.43	0.93	0.00	-0.15
Mathematics	7	1190772	0.93	-0.38	0.61	0.94	0.00	0.15

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	7	1190775	0.94	-0.34	0.29	0.92	0.00	-0.17
Mathematics	7	1190776	0.98	-0.33	-0.68	0.91	0.00	-1.04
Mathematics	7	1190777	1.04	-0.91	0.14	0.91	0.00	-0.85
Mathematics	7	1190780	1.03	-0.52	-0.40	0.95	0.00	-0.99
Mathematics	7	1190783	0.91	-0.29	0.65	0.93	0.00	0.30
Mathematics	7	1190788	0.91	0.68	-1.37	0.94	0.00	-0.65
Mathematics	7	1190790	0.89	0.88	-0.62	0.97	0.00	0.23
Mathematics	7	1190791	0.89	1.10	-1.38	0.94	0.00	-0.24
Mathematics	7	1190803	1.02	0.67	0.97	0.95	0.00	1.64
Mathematics	7	1190806	0.88	0.77	-1.42	0.92	0.00	-0.60
Mathematics	7	1190809	1.01	-0.50	0.05	0.97	0.00	-0.54
Mathematics	7	1190813	1.03	0.86	-1.96	0.96	0.00	-1.03
Mathematics	7	1190816	0.95	-0.41	1.23	0.97	0.00	0.80
Mathematics	7	1190822	0.95	-0.56	0.60	0.95	0.00	-0.05
Mathematics	7	1190829	0.88	0.64	-0.50	0.94	0.00	0.10
Mathematics	7	1190831	0.93	0.82	2.01	0.99	0.00	2.89
Mathematics	7	1190832	1.02	0.60	0.76	0.94	0.00	1.35
Mathematics	7	1190833	0.91	0.75	-1.40	0.95	0.00	-0.61
Mathematics	7	1190839	0.94	-0.43	-0.27	0.93	0.00	-0.81
Mathematics	7	1190851	0.91	0.63	-0.61	0.99	0.00	0.01
Mathematics	7	1190852	0.99	0.81	1.01	1.04	0.00	1.81
Mathematics	7	1190855	1.00	0.48	-0.34	1.07	0.00	0.10
Mathematics	7	1190858	0.95	0.60	0.41	0.96	0.00	0.92
Mathematics	7	1190859	1.10	1.08	1.19	1.07	0.00	2.29
Mathematics	7	1190863	0.98	0.56	-0.01	1.05	0.00	0.49

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	7	1190866	1.02	0.48	1.12	1.03	0.00	1.59
Mathematics	7	1190872	0.91	0.41	-1.03	0.96	0.00	-0.60
Mathematics	7	1190873	1.08	-0.69	-0.59	0.93	0.00	-1.31
Mathematics	7	1190874	1.03	-0.80	0.82	1.03	0.00	-0.03
Mathematics	7	1190876	1.10	-0.91	-0.21	0.99	0.00	-1.22
Mathematics	7	1190877	1.29	-1.40	-0.51	0.99	0.00	-1.97
Mathematics	7	1190881	1.07	-0.45	-0.09	1.00	0.00	-0.64
Mathematics	7	1190886	1.02	-0.91	0.49	1.04	0.00	-0.50
Mathematics	7	1190887	0.92	-0.59	0.35	0.92	0.00	-0.34
Mathematics	7	1190889	1.05	0.81	1.18	1.03	0.00	2.00
Mathematics	7	1190890	1.11	-1.13	0.10	0.92	0.00	-1.10
Mathematics	7	1190892	1.02	0.93	-2.01	0.92	0.00	-1.01
Mathematics	7	1190893	1.11	-0.39	0.42	1.01	0.00	-0.08
Mathematics	7	1190896	1.10	-0.96	-0.41	0.93	0.00	-1.45
Mathematics	7	1190897	1.14	-0.98	-0.10	1.00	0.00	-1.19
Mathematics	7	1190899	0.93	0.40	-1.40	0.93	0.00	-0.96
Mathematics	7	1190900	0.95	-0.37	0.06	0.90	0.00	-0.38
Mathematics	7	1190904	1.02	-0.90	-0.13	0.90	0.00	-1.14
Mathematics	7	1190907	1.02	-1.04	-0.40	0.89	0.00	-1.51
Mathematics	7	1190912	1.50	-1.65	0.26	1.04	0.00	-1.52
Mathematics	7	1190914	1.23	-1.13	-0.38	0.96	0.00	-1.56
Mathematics	7	1190916	1.05	-1.35	0.30	0.91	0.00	-1.18
Mathematics	7	1190917	1.25	-0.70	-0.24	1.05	0.00	-1.01
Mathematics	7	1190918	1.08	-0.99	-0.43	0.89	0.00	-1.48
Mathematics	7	1190919	0.99	-0.91	0.29	0.89	0.00	-0.75

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	7	1190920	1.06	0.67	1.14	1.06	0.00	1.81
Mathematics	7	1190922	1.01	-0.38	-1.02	0.96	0.00	-1.40
Mathematics	7	1190923	1.16	0.51	0.42	1.15	0.00	0.89
Mathematics	7	1190924	1.00	0.16	-0.12	1.01	0.00	-0.01
Mathematics	7	1190925	1.11	-1.95	1.14	0.98	0.00	-0.85
Mathematics	7	1190927	1.01	0.30	0.75	1.05	0.00	1.01
Mathematics	7	1190928	1.01	-0.71	-0.28	0.92	0.00	-1.07
Mathematics	7	1190929	0.98	-0.53	0.24	0.93	0.00	-0.39
Mathematics	7	1190930	0.84	-0.86	0.41	0.87	0.00	-0.52
Mathematics	7	1190933	1.01	0.53	1.02	1.01	0.00	1.54
Mathematics	7	1190934	1.00	-0.92	1.31	0.94	0.00	0.36
Mathematics	7	1190938	0.99	-0.36	1.21	0.99	0.00	0.83
Mathematics	7	1190942	1.17	-1.11	-0.53	0.92	0.00	-1.69
Mathematics	7	1190945	1.29	-1.07	0.01	1.04	0.00	-1.15
Mathematics	7	1190947	1.17	0.84	0.78	1.03	0.00	1.62
Mathematics	7	1190948	0.96	0.29	-0.09	0.98	0.00	0.13
Mathematics	7	1190950	1.05	0.92	1.27	1.02	0.00	2.21
Mathematics	7	1190954	0.93	-0.61	1.03	0.94	0.00	0.37
Mathematics	7	1190956	1.04	-0.47	0.11	0.96	0.00	-0.44
Mathematics	7	1190968	1.00	1.04	1.60	0.97	0.00	2.69
Mathematics	7	1190978	1.08	0.65	0.96	1.00	0.00	1.62
Mathematics	7	1190980	0.97	-0.33	0.34	0.96	0.00	-0.07
Mathematics	7	1190983	0.93	-0.80	0.63	0.91	0.00	-0.25
Mathematics	7	1190986	1.01	-0.40	0.28	0.97	0.00	-0.21
Mathematics	7	1190994	0.97	0.52	-1.73	0.95	0.00	-1.15

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	7	1190995	1.86	-1.37	2.39	1.00	0.00	0.80
Mathematics	7	1191000	1.10	-0.61	-0.46	0.97	0.00	-1.11
Mathematics	7	1191001	0.97	-0.89	1.33	0.90	0.00	0.42
Mathematics	7	1191003	1.23	-1.08	1.43	1.01	0.00	0.23
Mathematics	7	1191011	1.16	-0.74	0.13	1.05	0.00	-0.69
Mathematics	7	1191018	0.97	0.64	0.82	1.01	0.00	1.42
Mathematics	7	1191019	0.96	-0.54	-0.05	0.90	0.00	-0.69
Mathematics	7	1191029	0.97	0.46	-1.72	0.96	0.00	-1.20
Mathematics	7	1191035	1.05	0.49	0.69	1.08	0.00	1.12
Mathematics	7	1191036	0.94	0.06	0.61	0.94	0.00	0.61
Mathematics	7	1191039	1.05	0.65	1.04	1.07	0.00	1.68
Mathematics	7	1191055	1.03	0.91	1.20	0.99	0.00	2.13
Mathematics	7	1191057	0.95	-0.68	0.38	0.90	0.00	-0.42
Mathematics	7	1191058	1.36	1.06	-2.60	0.97	0.00	-1.45
Mathematics	7	1191060	1.00	0.96	1.78	1.07	0.00	2.78
Mathematics	7	1191061	0.88	-0.42	0.46	0.86	0.00	-0.06
Mathematics	7	1191062	0.91	0.79	-0.05	1.02	0.00	0.68
Mathematics	7	1191065	1.12	0.46	1.44	1.18	0.00	1.92
Mathematics	7	1191066	0.88	0.47	-0.60	0.94	0.00	-0.15
Mathematics	7	1191068	1.01	-0.56	2.06	0.90	0.00	1.49
Mathematics	7	1191069	1.05	0.90	0.90	1.04	0.00	1.79
Mathematics	7	1191071	0.92	0.63	1.66	0.96	0.00	2.31
Mathematics	7	1191072	0.98	0.78	0.85	1.03	0.00	1.60
Mathematics	7	1191079	0.86	0.68	2.20	0.97	0.00	2.94
Mathematics	7	1191099	0.96	0.50	0.85	0.97	0.00	1.32

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	7	1191107	0.89	-0.41	1.52	0.89	0.00	1.10
Mathematics	7	1191108	0.93	0.51	-0.40	0.98	0.00	0.07
Mathematics	7	1191110	0.93	0.60	-0.27	0.98	0.00	0.26
Mathematics	7	1191111	0.97	0.62	1.95	1.06	0.00	2.61
Mathematics	7	1191113	0.97	0.60	0.40	1.00	0.00	0.94
Mathematics	7	1191114	1.03	-0.23	1.49	1.02	0.00	1.26
Mathematics	7	1191125	1.20	-0.50	-0.06	1.08	0.00	-0.65
Mathematics	7	1191127	0.95	0.55	-1.09	1.00	0.00	-0.52
Mathematics	7	1191131	0.97	0.60	-0.55	1.05	0.00	0.04
Mathematics	7	1191132	1.08	0.40	-0.57	1.12	0.00	-0.20
Mathematics	7	1191133	1.06	1.17	-1.73	0.98	0.00	-0.50
Mathematics	7	1191139	0.97	0.61	0.62	0.98	0.00	1.18
Mathematics	7	1191140	0.89	0.58	-0.36	0.95	0.00	0.16
Mathematics	7	1191142	0.98	0.62	0.46	0.99	0.00	1.02
Mathematics	7	1191143	0.92	0.63	-0.67	0.96	0.00	-0.05
Mathematics	7	1191144	1.02	0.72	0.51	1.02	0.00	1.18
Mathematics	8	1191160	0.90	-0.92	1.01	0.98	0.00	0.13
Mathematics	8	1191161	0.88	-0.88	1.09	0.97	0.00	0.24
Mathematics	8	1191165	1.01	-0.40	0.36	1.00	0.00	-0.07
Mathematics	8	1191167	0.91	-0.41	0.07	0.91	0.00	-0.40
Mathematics	8	1191169	0.97	-0.59	0.36	0.96	0.00	-0.26
Mathematics	8	1191171	1.12	0.63	0.69	1.06	0.00	1.34
Mathematics	8	1191175	1.08	0.76	0.56	0.97	0.00	1.34
Mathematics	8	1191177	0.98	0.23	-0.76	1.00	0.00	-0.51
Mathematics	8	1191179	0.94	0.67	-1.32	0.94	0.00	-0.61

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	8	1191180	0.93	0.46	-1.21	0.94	0.00	-0.72
Mathematics	8	1191182	0.96	-0.46	-0.70	0.89	0.00	-1.16
Mathematics	8	1191183	0.99	-0.46	-0.99	0.94	0.00	-1.44
Mathematics	8	1191185	0.97	-0.97	0.36	0.94	0.00	-0.64
Mathematics	8	1191186	0.86	-1.11	0.43	0.96	0.00	-0.70
Mathematics	8	1191188	0.95	-1.05	0.02	0.93	0.00	-1.10
Mathematics	8	1191189	0.93	-0.55	0.09	0.93	0.00	-0.53
Mathematics	8	1191190	0.95	-0.50	-0.30	0.92	0.00	-0.84
Mathematics	8	1191191	0.94	-0.41	-0.67	0.87	0.00	-1.08
Mathematics	8	1191192	1.03	-0.82	-0.12	0.91	0.00	-1.00
Mathematics	8	1191193	0.93	-0.48	0.05	0.93	0.00	-0.49
Mathematics	8	1191194	1.06	-0.73	-0.16	0.93	0.00	-0.94
Mathematics	8	1191195	0.84	-0.84	0.53	0.94	0.00	-0.31
Mathematics	8	1191196	1.02	-0.62	-0.51	0.94	0.00	-1.16
Mathematics	8	1191197	1.02	-0.63	-0.44	0.95	0.00	-1.09
Mathematics	8	1191201	0.88	-1.18	1.36	0.98	0.00	0.23
Mathematics	8	1191202	0.86	-0.57	0.38	0.88	0.00	-0.21
Mathematics	8	1191203	1.01	-0.60	1.85	0.99	0.00	1.28
Mathematics	8	1191208	0.95	0.40	-1.29	0.95	0.00	-0.85
Mathematics	8	1191210	1.10	0.97	1.20	0.94	0.00	2.24
Mathematics	8	1191215	0.94	-0.56	0.38	0.96	0.00	-0.21
Mathematics	8	1191219	1.05	-0.39	-0.19	0.98	0.00	-0.61
Mathematics	8	1191221	0.91	-0.71	1.31	0.90	0.00	0.61
Mathematics	8	1191224	1.07	0.97	0.47	1.15	0.00	1.44
Mathematics	8	1191226	1.03	-0.39	0.55	1.01	0.00	0.12

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	8	1191229	0.86	-0.92	1.31	0.90	0.00	0.41
Mathematics	8	1191231	0.99	-0.61	0.86	0.96	0.00	0.23
Mathematics	8	1191233	0.94	-1.16	1.35	0.91	0.00	0.21
Mathematics	8	1191236	1.08	-0.79	-0.13	0.95	0.00	-0.97
Mathematics	8	1191238	0.96	0.51	-0.05	1.00	0.00	0.44
Mathematics	8	1191239	0.98	-0.43	0.19	0.99	0.00	-0.27
Mathematics	8	1191240	0.97	-0.41	0.53	0.98	0.00	0.09
Mathematics	8	1191246	0.96	-0.41	0.46	1.02	0.00	0.04
Mathematics	8	1191247	0.88	-0.56	1.06	0.95	0.00	0.53
Mathematics	8	1191254	1.03	0.38	1.06	1.02	0.00	1.46
Mathematics	8	1191265	1.11	0.71	1.43	1.04	0.00	2.20
Mathematics	8	1191271	0.94	-0.46	1.19	0.94	0.00	0.74
Mathematics	8	1191274	0.98	-0.39	0.58	0.96	0.00	0.16
Mathematics	8	1191279	1.02	0.53	1.46	1.01	0.00	2.04
Mathematics	8	1191281	0.99	0.74	0.64	1.02	0.00	1.38
Mathematics	8	1191287	0.90	-0.42	1.11	0.91	0.00	0.70
Mathematics	8	1191293	0.99	0.53	0.48	1.01	0.00	1.00
Mathematics	8	1191294	1.07	0.05	-0.41	1.06	0.00	-0.36
Mathematics	8	1191295	1.12	0.44	0.20	1.12	0.00	0.62
Mathematics	8	1191296	1.04	0.53	0.66	1.06	0.00	1.19
Mathematics	8	1191297	0.99	0.20	-0.16	1.01	0.00	0.03
Mathematics	8	1191304	1.01	-0.52	0.09	0.96	0.00	-0.46
Mathematics	8	1191309	0.88	0.41	-0.40	0.92	0.00	0.01
Mathematics	8	1191310	0.91	0.42	-1.09	0.92	0.00	-0.65
Mathematics	8	1191311	0.98	-0.37	-0.10	0.93	0.00	-0.51

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	8	1191312	0.87	-0.73	0.73	0.88	0.00	-0.02
Mathematics	8	1191318	1.53	-1.06	2.23	1.15	0.00	1.06
Mathematics	8	1191321	0.96	0.27	1.53	0.96	0.00	1.85
Mathematics	8	1191327	0.93	0.67	0.73	0.95	0.00	1.40
Mathematics	8	1191329	0.99	0.60	1.41	0.91	0.00	2.06
Mathematics	8	1191331	0.92	0.50	-0.74	0.91	0.00	-0.24
Mathematics	8	1191332	0.99	0.41	0.29	1.01	0.00	0.70
Mathematics	8	1191335	0.96	0.33	-0.01	0.98	0.00	0.29
Mathematics	8	1191338	1.07	0.69	0.29	1.10	0.00	0.98
Mathematics	8	1191340	1.04	0.79	0.95	1.01	0.00	1.77
Mathematics	8	1191341	0.95	-0.56	1.12	0.97	0.00	0.56
Mathematics	8	1191342	0.84	-0.86	1.56	0.89	0.00	0.74
Mathematics	8	1191343	1.04	0.48	1.08	1.02	0.00	1.58
Mathematics	8	1191345	0.93	-0.31	1.36	0.96	0.00	1.09
Mathematics	8	1191346	1.11	-0.48	0.39	1.06	0.00	-0.13
Mathematics	8	1191351	0.86	0.61	-0.28	0.92	0.00	0.32
Mathematics	8	1191353	1.08	0.82	1.90	1.10	0.00	2.79
Mathematics	8	1191355	0.99	0.56	1.50	0.97	0.00	2.11
Mathematics	8	1191358	1.14	0.94	1.29	1.06	0.00	2.28
Mathematics	8	1191359	0.95	0.53	1.04	0.93	0.00	1.60
Mathematics	8	1191360	1.14	-0.52	1.25	1.10	0.00	0.72
Mathematics	8	1191365	1.10	-0.56	0.47	1.00	0.00	-0.12
Mathematics	8	1191366	1.14	-0.98	-0.12	0.98	0.00	-1.12
Mathematics	8	1191374	1.09	0.41	0.60	1.15	0.00	1.01
Mathematics	8	1191375	1.12	-0.91	-0.32	1.00	0.00	-1.24

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	8	1191377	0.87	0.49	-0.87	0.90	0.00	-0.36
Mathematics	8	1191382	1.05	-0.69	2.48	0.90	0.00	1.82
Mathematics	8	1191388	1.06	-0.47	0.39	1.02	0.00	-0.11
Mathematics	8	1191389	1.01	0.40	1.46	1.00	0.00	1.90
Mathematics	8	1191393	1.04	0.14	0.08	1.04	0.00	0.19
Mathematics	8	1191400	1.18	0.01	0.30	1.19	0.00	0.30
Mathematics	8	1191402	0.98	0.53	0.82	0.97	0.00	1.35
Mathematics	8	1191403	1.08	0.84	0.72	1.06	0.00	1.58
Mathematics	8	1191405	0.96	0.66	0.81	1.05	0.00	1.48
Mathematics	8	1191411	0.89	0.47	-0.97	0.93	0.00	-0.46
Mathematics	8	1191412	0.90	0.57	-0.54	0.97	0.00	0.05
Mathematics	8	1191420	1.01	0.58	1.46	0.94	0.00	2.09
Mathematics	8	1191425	1.05	0.34	0.25	1.08	0.00	0.59
Mathematics	8	1191428	0.97	0.51	0.74	0.98	0.00	1.25
Mathematics	8	1191430	1.13	0.30	0.19	1.16	0.00	0.47
Mathematics	8	1191433	1.00	0.19	1.05	1.02	0.00	1.26
Mathematics	8	1191444	1.17	-0.45	0.73	1.10	0.00	0.23
Mathematics	8	1191446	1.04	0.58	1.31	1.05	0.00	1.94
Mathematics	8	1191448	0.99	0.42	1.35	0.99	0.00	1.81
Mathematics	8	1191482	1.12	0.80	1.61	1.05	0.00	2.47
Mathematics	8	1191484	0.97	0.55	1.99	0.96	0.00	2.60
Mathematics	8	1191487	1.01	0.47	0.76	0.95	0.00	1.24
Mathematics	8	1191495	1.12	0.81	1.23	0.98	0.00	2.10
Mathematics	8	1191499	0.95	0.47	0.04	0.97	0.00	0.47
Mathematics	8	1191501	0.99	0.53	-1.55	0.96	0.00	-0.98

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Mathematics	8	1191504	0.94	0.63	-1.23	0.96	0.00	-0.57
Mathematics	8	1191505	0.91	0.47	-0.92	0.96	0.00	-0.43
Mathematics	8	1191506	0.88	0.86	-0.45	0.96	0.00	0.42
Mathematics	8	1191507	0.95	-0.38	0.48	0.96	0.00	0.05
Mathematics	8	1191509	1.02	0.36	1.80	1.00	0.00	2.21
Mathematics	8	1191513	0.87	0.50	-0.61	0.94	0.00	-0.09
Mathematics	8	1191516	0.89	-0.26	0.92	0.92	0.00	0.68
Mathematics	8	1191519	0.99	0.38	0.05	1.04	0.00	0.42
Mathematics	8	1191521	0.93	0.62	-0.06	0.94	0.00	0.52
Mathematics	8	1191522	1.14	0.78	0.70	1.07	0.00	1.50
Mathematics	8	1191523	0.99	0.56	0.71	0.91	0.00	1.29
Mathematics	8	1191525	1.22	1.01	1.04	1.04	0.00	2.11
Mathematics	8	1191526	1.04	0.53	1.24	0.95	0.00	1.82
Mathematics	8	1191530	1.21	0.94	1.03	1.06	0.00	2.02
Mathematics	8	1191532	0.94	0.51	1.31	0.99	0.00	1.87
Mathematics	8	1191533	1.09	0.44	0.56	1.13	0.00	1.00
Mathematics	8	1191534	0.93	0.42	-0.74	0.99	0.00	-0.31
Mathematics	8	1191538	0.89	0.61	-0.14	0.95	0.00	0.45
Mathematics	8	1191543	0.98	-0.23	-0.30	0.95	0.00	-0.56
Mathematics	8	1191550	0.83	-0.83	1.38	0.92	0.00	0.59
Mathematics	8	1191554	1.12	0.74	1.12	0.98	0.00	1.91
Mathematics	8	1191559	0.94	0.42	1.06	0.92	0.00	1.50
Mathematics	8	1191560	1.03	0.54	0.85	0.97	0.00	1.41
Mathematics	8	1191562	0.88	0.62	-0.95	0.95	0.00	-0.30
Algebra I	HS	1191567	0.94	0.53	-1.37	0.92	0.00	-0.83

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Algebra I	HS	1191571	0.88	0.54	-0.33	0.95	0.00	0.18
Algebra I	HS	1191573	1.01	-0.43	0.51	0.98	0.00	-0.03
Algebra I	HS	1191583	1.03	-0.35	0.29	0.96	0.00	-0.16
Algebra I	HS	1191584	0.99	-0.39	0.49	0.96	0.00	-0.03
Algebra I	HS	1191587	0.91	0.43	-1.04	0.92	0.00	-0.61
Algebra I	HS	1191593	0.96	-0.33	1.50	0.94	0.00	1.10
Algebra I	HS	1191595	1.17	-1.19	0.67	0.94	0.00	-0.63
Algebra I	HS	1191596	1.06	-1.07	1.10	0.98	0.00	-0.10
Algebra I	HS	1191597	0.99	-0.60	1.11	0.95	0.00	0.39
Algebra I	HS	1191598	1.00	-0.35	2.01	1.02	0.00	1.61
Algebra I	HS	1191599	0.96	-0.05	1.06	0.96	0.00	0.90
Algebra I	HS	1191600	1.02	0.52	1.97	0.99	0.00	2.47
Algebra I	HS	1191604	1.05	-0.61	0.69	0.96	0.00	-0.03
Algebra I	HS	1191606	1.18	-1.38	0.77	0.95	0.00	-0.76
Algebra I	HS	1191608	1.04	-0.42	-0.15	0.93	0.00	-0.62
Algebra I	HS	1191609	1.00	-0.23	0.94	0.97	0.00	0.59
Algebra I	HS	1191611	1.00	0.46	1.84	1.00	0.00	2.27
Algebra I	HS	1191614	1.08	0.59	2.06	1.06	0.00	2.63
Algebra I	HS	1191617	1.06	0.69	2.00	1.04	0.00	2.67
Algebra I	HS	1191620	1.05	1.03	1.38	1.01	0.00	2.38
Algebra I	HS	1191621	1.09	0.85	1.57	1.03	0.00	2.40
Algebra I	HS	1191623	1.13	-0.54	0.65	1.07	0.00	0.00
Algebra I	HS	1191624	0.99	-0.68	-0.09	0.94	0.00	-0.86
Algebra I	HS	1191639	0.97	0.45	1.72	0.98	0.00	2.14
Algebra I	HS	1191640	0.96	0.53	0.92	0.99	0.00	1.35

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Algebra I	HS	1191642	0.87	0.46	-0.14	0.95	0.00	0.29
Algebra I	HS	1191644	1.02	0.61	2.05	1.00	0.00	2.64
Algebra I	HS	1191650	0.91	-0.55	1.51	0.94	0.00	0.91
Algebra I	HS	1191655	1.09	0.71	1.64	1.03	0.00	2.32
Algebra I	HS	1191656	1.02	0.51	1.96	0.99	0.00	2.45
Algebra I	HS	1191660	0.96	0.53	1.19	0.99	0.00	1.65
Algebra I	HS	1191662	1.02	0.46	1.90	0.99	0.00	2.33
Algebra I	HS	1191667	1.13	-0.46	0.96	1.08	0.00	0.37
Algebra I	HS	1191674	0.89	0.51	-0.35	0.96	0.00	0.14
Algebra I	HS	1191690	0.93	-0.52	0.11	0.91	0.00	-0.52
Algebra I	HS	1191691	0.96	-0.37	-0.16	0.93	0.00	-0.60
Algebra I	HS	1191693	1.08	-0.87	0.63	0.97	0.00	-0.36
Algebra I	HS	1191697	0.98	0.69	2.15	0.98	0.00	2.82
Algebra I	HS	1191703	0.95	-0.69	1.18	0.92	0.00	0.39
Algebra I	HS	1191706	0.99	-0.54	1.32	0.97	0.00	0.69
Algebra I	HS	1191708	1.07	-0.81	0.25	0.92	0.00	-0.68
Algebra I	HS	1191711	1.12	-1.00	0.60	1.00	0.00	-0.53
Algebra I	HS	1191715	1.13	0.78	1.75	1.10	0.00	2.50
Algebra I	HS	1191718	0.96	-0.42	1.22	0.94	0.00	0.71
Algebra I	HS	1191720	1.06	-0.52	0.98	1.00	0.00	0.33
Algebra I	HS	1191725	0.99	-0.18	0.82	0.98	0.00	0.51
Algebra I	HS	1191726	1.01	-1.05	0.87	0.95	0.00	-0.33
Algebra I	HS	1191727	0.94	-0.36	1.59	0.96	0.00	1.17
Algebra I	HS	1191728	0.99	0.49	1.50	0.98	0.00	1.95
Algebra I	HS	1191731	1.01	0.46	1.98	0.99	0.00	2.42

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Algebra I	HS	1191733	1.03	-0.32	1.86	1.05	0.00	1.50
Algebra I	HS	1191738	1.12	-0.52	0.71	1.04	0.00	0.08
Algebra I	HS	1191745	0.93	-0.32	1.98	0.93	0.00	1.63
Algebra I	HS	1191748	1.10	-1.22	0.52	1.01	0.00	-0.84
Algebra I	HS	1191750	1.24	-1.49	0.67	0.97	0.00	-0.94
Algebra I	HS	1191757	1.09	-0.39	-0.11	0.99	0.00	-0.55
Algebra I	HS	1191760	1.10	-0.50	-0.55	1.00	0.00	-1.08
Algebra I	HS	1191766	0.95	0.49	-0.05	1.01	0.00	0.39
Algebra I	HS	1191767	0.91	0.42	-0.90	0.95	0.00	-0.48
Algebra I	HS	1191770	1.05	-0.79	-0.07	0.90	0.00	-0.95
Algebra I	HS	1191779	0.92	-0.73	1.46	0.94	0.00	0.67
Algebra I	HS	1191781	1.05	-0.89	2.40	0.97	0.00	1.47
Algebra I	HS	1191789	1.13	-0.41	0.11	1.05	0.00	-0.39
Algebra I	HS	1191798	0.88	0.78	-0.92	0.93	0.00	-0.14
Algebra I	HS	1191804	1.01	0.62	1.35	1.01	0.00	1.92
Algebra I	HS	1191809	1.02	0.59	1.61	1.00	0.00	2.16
Algebra I	HS	1191818	0.95	0.59	0.45	0.99	0.00	0.94
Algebra I	HS	1191832	1.06	-0.55	0.69	0.98	0.00	0.03
Algebra I	HS	1191836	1.05	0.56	1.24	1.09	0.00	1.74
Algebra I	HS	1191839	0.95	-0.69	1.45	0.96	0.00	0.70
Algebra I	HS	1191841	0.97	-0.33	1.18	0.98	0.00	0.77
Algebra I	HS	1191843	1.04	-0.44	1.01	0.99	0.00	0.44
Algebra I	HS	1191844	1.03	-0.73	2.24	0.99	0.00	1.47
Algebra I	HS	1191845	1.02	-0.80	0.50	0.96	0.00	-0.43
Algebra I	HS	1191861	0.96	0.51	1.32	0.97	0.00	1.78

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Algebra I	HS	1191877	1.04	-0.34	0.73	1.00	0.00	0.28
Algebra I	HS	1191883	1.05	0.88	1.25	1.07	0.00	2.09
Algebra I	HS	1191884	0.91	-0.93	1.94	0.93	0.00	0.97
Algebra I	HS	1191885	0.90	-0.84	1.62	0.91	0.00	0.72
Algebra I	HS	1191886	0.99	-0.56	2.28	0.96	0.00	1.68
Algebra I	HS	1191887	0.95	-0.68	1.53	0.93	0.00	0.78
Algebra I	HS	1191888	1.00	-0.40	0.28	0.96	0.00	-0.23
Algebra I	HS	1191895	1.02	-1.04	0.88	0.95	0.00	-0.31
Algebra I	HS	1191896	1.01	-0.37	0.32	0.98	0.00	-0.16
Algebra I	HS	1191897	0.95	-0.53	1.59	0.96	0.00	1.01
Algebra I	HS	1191903	0.94	-0.74	1.54	0.93	0.00	0.73
Algebra I	HS	1191912	1.00	0.41	-0.21	1.06	0.00	0.17
Algebra I	HS	1191913	1.05	-0.34	0.13	1.00	0.00	-0.30
Algebra I	HS	1191920	0.96	0.58	-0.20	1.04	0.00	0.35
Algebra I	HS	1191928	0.96	-0.24	2.04	0.96	0.00	1.77
Algebra I	HS	1191930	0.87	0.46	0.11	0.92	0.00	0.51
Algebra I	HS	1191937	1.02	-0.70	2.73	0.90	0.01	1.99
Algebra I	HS	1191938	1.12	-0.57	0.58	1.07	0.00	-0.12
Algebra I	HS	1191943	1.02	-0.34	0.23	0.96	0.00	-0.21
Algebra I	HS	1191944	1.00	-0.35	1.25	1.00	0.00	0.82
Algebra I	HS	1191946	0.90	-0.18	2.36	0.88	0.00	2.15
Algebra I	HS	1191947	1.03	-0.48	-0.50	0.92	0.00	-1.02
Algebra I	HS	1191948	0.93	-0.25	1.47	0.94	0.00	1.16
Algebra I	HS	1191949	1.00	0.82	1.41	0.98	0.00	2.20
Algebra I	HS	1191953	0.96	-0.31	0.90	0.94	0.00	0.47

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Algebra I	HS	1191993	1.01	-0.46	-0.01	0.96	0.00	-0.55
Algebra I	HS	1191994	0.99	0.49	1.88	0.96	0.00	2.35
Algebra I	HS	1191998	0.94	0.51	-1.21	0.93	0.00	-0.69
Algebra I	HS	1192006	0.99	0.59	-1.62	0.93	0.00	-1.01

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	3	1192120	0.97	0.77	-2.11	0.93	0.00	-1.35
ELA	3	1192122	1.08	-0.65	-1.73	1.02	0.00	-2.40
ELA	3	1192133	1.12	0.92	-2.56	0.93	0.00	-1.64
ELA	3	1192134	1.06	-0.53	-1.87	1.00	0.00	-2.41
ELA	3	1192135	0.96	0.45	-1.25	0.99	0.00	-0.81
ELA	3	1192137	1.25	-0.77	-0.62	0.91	0.00	-1.42
ELA	3	1192138	0.92	0.86	-1.25	0.99	0.00	-0.41
ELA	3	1192140	0.98	0.43	-1.30	1.02	0.00	-0.91
ELA	3	1192144	1.39	-0.74	-0.70	1.07	0.00	-1.47
ELA	3	1192151	0.97	1.10	-2.03	0.95	0.00	-0.90
ELA	3	1192152	2.17	-0.94	0.28	1.02	0.00	-0.64
ELA	3	1192154	1.01	-0.64	-1.19	0.91	0.00	-1.88
ELA	3	1192156	0.98	-0.13	-0.51	0.94	0.00	-0.66
ELA	3	1192157	0.96	-1.04	-1.52	0.90	0.00	-2.62
ELA	3	1192159	0.88	0.57	-0.78	1.04	0.00	-0.23
ELA	3	1192162	0.89	0.69	-1.66	0.99	0.00	-0.92
ELA	3	1192178	0.87	0.64	-0.74	0.97	0.00	-0.06
ELA	3	1192179	1.01	0.98	-1.21	0.97	0.00	-0.20
ELA	3	1192183	0.83	0.56	0.11	0.90	0.00	0.76
ELA	3	1192184	1.03	0.33	-0.72	0.96	0.00	-0.31
ELA	3	1192185	0.94	0.57	-0.03	0.99	0.00	0.63
ELA	3	1192191	1.27	-0.53	-0.44	1.15	0.00	-0.98
ELA	3	1192192	1.81	1.23	-1.42	0.88	0.00	-0.04
ELA	3	1192193	0.95	-0.30	0.36	0.92	0.00	0.19
ELA	3	1192194	1.24	0.81	-0.60	0.98	0.00	0.34

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	3	1192195	0.96	0.39	-0.02	0.94	0.00	0.50
ELA	3	1192196	0.96	0.52	-0.11	0.90	0.00	0.54
ELA	3	1192202	1.56	-0.75	-0.66	1.03	0.00	-1.39
ELA	3	1192205	1.18	1.09	-2.70	0.92	0.00	-1.57
ELA	3	1192206	0.90	0.86	-1.87	0.97	0.00	-0.97
ELA	3	1192208	0.88	1.16	-1.91	0.97	0.00	-0.76
ELA	3	1192209	1.77	-0.91	-0.51	1.01	0.00	-1.43
ELA	3	1192210	0.87	0.74	-1.77	0.97	0.00	-1.04
ELA	3	1192211	1.44	-0.51	-0.16	1.29	0.00	-0.66
ELA	3	1192212	1.54	1.69	-3.03	0.92	0.00	-1.30
ELA	3	1192215	1.81	-0.71	0.38	1.25	0.00	-0.40
ELA	3	1192216	1.67	-0.81	-0.91	1.27	0.00	-1.74
ELA	3	1192217	1.55	-0.57	-0.30	1.22	0.00	-0.92
ELA	3	1192220	1.64	-0.62	-0.16	1.11	0.00	-0.73
ELA	3	1192229	1.20	-0.43	-0.86	0.98	0.00	-1.12
ELA	3	1192233	1.46	0.24	-0.85	1.01	0.00	-0.27
ELA	3	1192236	0.96	0.38	-0.37	0.93	0.00	0.07
ELA	3	1192252	1.09	0.40	-0.24	1.03	0.00	0.24
ELA	3	1192253	1.24	-0.69	-0.50	1.09	0.00	-1.22
ELA	3	1192258	1.05	0.85	0.25	1.02	0.00	1.22
ELA	3	1192259	0.57	-0.73	-1.73	0.98	0.00	-2.35
ELA	3	1192262	0.99	0.47	0.08	0.95	0.00	0.67
ELA	3	1192263	1.34	0.25	-0.15	1.07	0.00	0.27
ELA	3	1192273	0.98	0.41	0.78	0.96	0.00	1.32
ELA	3	1192274	1.01	0.80	-1.46	0.87	0.00	-0.64

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	3	1192275	0.95	0.51	-1.85	0.96	0.00	-1.37
ELA	3	1192279	1.24	-0.67	-0.98	0.94	0.00	-1.68
ELA	3	1192281	0.71	-0.52	-0.91	0.91	0.00	-1.32
ELA	3	1192283	1.05	0.62	0.69	1.11	0.00	1.43
ELA	3	1192285	1.10	0.33	0.15	1.06	0.00	0.59
ELA	3	1192291	1.33	-0.08	-0.06	1.35	0.00	-0.09
ELA	3	1192293	0.98	0.52	-2.79	0.89	0.00	-2.25
ELA	3	1192294	0.78	0.66	0.33	0.96	0.00	1.03
ELA	3	1192296	0.94	0.23	0.48	0.96	0.00	0.78
ELA	3	1192298	1.21	-0.27	0.14	1.22	0.00	-0.10
ELA	3	1192300	1.73	-0.79	-0.32	1.02	0.00	-1.13
ELA	3	1192302	1.23	-0.47	-0.40	1.01	0.00	-0.84
ELA	3	1192310	0.95	0.36	0.60	0.99	0.00	1.10
ELA	3	1192311	0.83	0.62	-1.34	0.96	0.00	-0.73
ELA	3	1192316	1.03	0.89	0.43	1.03	0.00	1.45
ELA	3	1192318	0.68	0.03	-1.34	0.98	0.00	-0.93
ELA	3	1192322	1.42	-0.20	-0.30	1.34	0.00	-0.51
ELA	3	1192328	0.80	0.49	0.05	0.96	0.00	0.55
ELA	3	1192329	1.19	-0.61	-1.21	0.93	0.00	-1.88
ELA	3	1192336	1.27	-0.38	-0.42	1.03	0.00	-0.86
ELA	3	1192339	1.45	0.13	0.15	1.38	0.00	0.38
ELA	3	1192340	1.29	-0.42	-0.62	1.15	0.00	-1.07
ELA	3	1192343	0.91	0.36	0.28	0.98	0.00	0.70
ELA	3	1192353	0.94	0.38	-0.64	0.88	0.00	-0.21
ELA	3	1192369	0.89	0.47	0.55	0.98	0.00	1.10

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	3	1192384	0.74	-0.41	-1.39	0.91	0.00	-1.77
ELA	3	1192386	2.91	-1.38	0.12	1.03	0.00	-1.32
ELA	3	1192387	1.86	-1.01	-0.54	1.01	0.00	-1.61
ELA	3	1192390	1.71	-0.59	-0.23	1.02	0.00	-0.84
ELA	3	1192392	0.96	0.62	-1.43	0.95	0.00	-0.81
ELA	3	1192393	0.89	0.43	-0.85	0.99	0.00	-0.41
ELA	3	1192397	0.98	0.37	-2.58	0.91	0.00	-2.25
ELA	3	1192404	1.11	0.61	-0.61	0.97	0.00	0.08
ELA	3	1192405	0.96	0.40	0.26	1.00	0.00	0.75
ELA	3	1192407	0.98	0.41	0.21	1.01	0.00	0.70
ELA	3	1192408	1.20	0.28	-0.03	1.01	0.00	0.44
ELA	3	1192409	1.12	0.30	-0.71	0.98	0.00	-0.28
ELA	3	1192410	0.96	0.45	0.60	0.98	0.00	1.19
ELA	3	1192412	0.61	-0.58	-2.01	0.97	0.00	-2.49
ELA	3	1192413	1.10	0.69	0.20	1.02	0.00	1.04
ELA	3	1192414	1.38	0.26	0.31	1.15	0.00	0.61
ELA	3	1192417	0.95	0.35	-0.13	0.97	0.00	0.30
ELA	3	1192418	1.01	0.45	-0.53	0.95	0.00	-0.01
ELA	3	1192430	1.02	0.46	-0.23	0.98	0.00	0.29
ELA	3	1192433	0.71	-0.47	-1.47	0.88	0.00	-1.90
ELA	3	1192443	1.17	-0.52	-0.69	1.09	0.00	-1.22
ELA	3	1192444	0.95	0.44	-0.93	0.96	0.00	-0.48
ELA	3	1192451	0.96	0.36	-2.13	0.93	0.00	-1.78
ELA	3	1192452	1.32	-0.52	-0.72	1.05	0.00	-1.26
ELA	3	1192455	1.02	0.42	-1.09	0.92	0.00	-0.65

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	3	1192465	0.69	0.05	-1.41	0.87	0.00	-1.29
ELA	3	1192471	1.23	-0.49	-0.59	1.04	0.00	-0.94
ELA	3	1192473	1.18	-0.50	-0.81	0.95	0.00	-1.35
ELA	3	1192475	0.92	0.40	-2.11	0.89	0.00	-1.74
ELA	4	1192477	1.06	0.46	0.43	0.92	0.00	0.91
ELA	4	1192479	1.21	0.32	-1.44	0.94	0.00	-1.10
ELA	4	1192480	0.59	-0.57	-0.87	0.88	0.00	-1.41
ELA	4	1192485	1.08	0.45	-3.04	0.91	0.00	-2.59
ELA	4	1192487	1.80	-0.76	0.03	1.49	0.00	-0.83
ELA	4	1192489	0.93	0.35	-1.12	1.00	0.00	-0.75
ELA	4	1192490	1.32	-0.47	-0.03	1.00	0.00	-0.52
ELA	4	1192491	0.87	0.79	-1.35	0.96	0.00	-0.55
ELA	4	1192494	1.09	0.66	0.44	0.91	0.00	1.14
ELA	4	1192499	1.77	1.34	-3.24	0.92	0.00	-1.84
ELA	4	1192505	1.13	0.93	-1.92	0.97	0.00	-1.01
ELA	4	1192506	2.03	-1.09	0.25	1.01	0.00	-0.89
ELA	4	1192508	1.36	-0.62	-0.40	1.00	0.00	-1.12
ELA	4	1192509	1.05	-0.42	-1.34	0.99	0.00	-1.85
ELA	4	1192510	1.71	-0.79	0.18	1.02	0.00	-0.72
ELA	4	1192515	0.96	0.59	-1.34	1.00	0.00	-0.76
ELA	4	1192518	1.25	1.14	-2.36	0.93	0.00	-1.20
ELA	4	1192524	1.78	-1.01	0.01	1.02	0.00	-1.05
ELA	4	1192527	0.67	-0.43	-1.55	0.93	0.00	-1.93
ELA	4	1192530	1.05	0.25	0.32	0.99	0.00	0.63
ELA	4	1192539	0.92	0.62	-1.12	1.01	0.00	-0.54

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	4	1192542	1.91	-0.61	1.43	1.57	0.00	0.71
ELA	4	1192543	2.40	-0.96	1.12	1.45	0.00	0.09
ELA	4	1192544	0.97	-0.12	1.96	0.93	0.00	1.86
ELA	4	1192547	0.72	-0.36	-0.67	0.90	0.00	-1.02
ELA	4	1192548	1.02	-0.38	1.39	0.93	0.00	1.00
ELA	4	1192552	0.94	0.66	-2.23	0.89	0.00	-1.56
ELA	4	1192554	2.76	-1.17	0.72	1.08	0.00	-0.46
ELA	4	1192562	0.78	0.46	-0.47	0.94	0.00	-0.02
ELA	4	1192565	0.86	0.51	-0.82	0.96	0.00	-0.32
ELA	4	1192566	1.43	0.52	-0.82	0.98	0.00	-0.26
ELA	4	1192567	1.31	0.71	0.33	0.99	0.00	1.08
ELA	4	1192569	0.87	0.49	1.66	0.93	0.00	2.20
ELA	4	1192571	1.00	0.56	0.99	1.02	0.00	1.61
ELA	4	1192574	0.89	0.72	1.67	1.09	0.00	2.45
ELA	4	1192575	0.94	0.52	1.17	0.99	0.00	1.75
ELA	4	1192583	1.96	-0.83	-0.07	1.03	0.00	-0.98
ELA	4	1192587	1.52	0.46	-0.06	1.04	0.00	0.53
ELA	4	1192589	1.15	-0.44	-0.31	1.00	0.00	-0.76
ELA	4	1192602	0.92	0.44	0.18	0.96	0.00	0.62
ELA	4	1192610	1.18	-0.83	-0.53	1.01	0.00	-1.41
ELA	4	1192615	0.57	-0.07	-2.09	0.61	0.00	-2.16
ELA	4	1192619	1.34	-0.42	0.03	1.24	0.00	-0.39
ELA	4	1192630	2.09	-1.08	0.66	0.99	0.00	-0.50
ELA	4	1192636	1.01	-0.53	1.28	0.96	0.00	0.75
ELA	4	1192639	1.03	-0.23	1.03	1.04	0.00	0.80

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	4	1192641	0.97	0.56	1.04	0.95	0.00	1.65
ELA	4	1192645	0.98	0.58	0.22	0.96	0.00	0.79
ELA	4	1192649	0.93	-0.31	0.47	0.95	0.00	0.13
ELA	4	1192650	0.91	-0.29	0.28	0.94	0.00	-0.04
ELA	4	1192655	0.87	0.29	-1.29	0.90	0.00	-1.02
ELA	4	1192656	1.88	-0.85	0.37	1.28	0.00	-0.52
ELA	4	1192658	1.07	-0.14	0.63	1.09	0.00	0.47
ELA	4	1192660	0.93	0.51	1.50	1.01	0.00	2.09
ELA	4	1192661	1.24	0.73	0.46	1.03	0.00	1.26
ELA	4	1192662	1.27	0.44	-0.63	0.96	0.00	-0.13
ELA	4	1192663	0.99	0.71	1.06	0.97	0.00	1.85
ELA	4	1192664	1.10	-0.44	1.62	0.96	0.00	1.16
ELA	4	1192666	0.88	-0.17	0.43	0.90	0.00	0.24
ELA	4	1192667	1.00	0.44	0.49	0.95	0.00	0.91
ELA	4	1192668	0.88	-0.26	0.34	0.91	0.00	0.06
ELA	4	1192669	1.02	-0.18	1.94	0.96	0.00	1.75
ELA	4	1192670	1.02	-0.46	0.79	1.01	0.00	0.31
ELA	4	1192693	1.78	-0.77	0.24	1.03	0.00	-0.59
ELA	4	1192703	1.01	0.42	-0.32	0.94	0.00	0.09
ELA	4	1192707	0.86	0.55	-0.34	0.94	0.00	0.18
ELA	4	1192713	1.30	-0.38	0.41	1.20	0.00	0.00
ELA	4	1192714	1.15	0.36	-0.39	0.96	0.00	0.07
ELA	4	1192715	0.92	0.62	1.17	0.97	0.00	1.89
ELA	4	1192716	1.43	0.88	-0.10	1.02	0.00	0.88
ELA	4	1192717	0.98	0.43	0.84	0.96	0.00	1.37

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	4	1192718	1.29	0.43	-0.52	1.04	0.00	0.00
ELA	4	1192719	1.31	0.35	0.07	1.05	0.00	0.63
ELA	4	1192720	0.99	0.50	0.65	1.00	0.00	1.17
ELA	4	1192723	0.99	0.48	0.68	1.01	0.00	1.18
ELA	4	1192727	2.30	-1.06	0.54	0.99	0.00	-0.41
ELA	4	1192732	0.90	0.44	0.04	0.95	0.00	0.46
ELA	4	1192742	1.44	-0.46	0.17	1.31	0.00	-0.31
ELA	4	1192745	1.52	0.86	1.15	1.25	0.00	2.02
ELA	4	1192750	2.08	-1.09	0.37	1.02	0.00	-0.81
ELA	4	1192755	1.14	0.38	-0.42	0.90	0.00	0.01
ELA	4	1192757	1.08	0.75	0.97	1.01	0.00	1.79
ELA	4	1192758	1.02	0.74	1.11	1.00	0.00	1.92
ELA	4	1192760	1.10	0.73	0.27	0.98	0.00	1.04
ELA	4	1192761	0.89	0.44	0.89	0.95	0.00	1.36
ELA	4	1192765	2.58	-1.20	0.61	0.98	0.00	-0.67
ELA	4	1192774	1.81	-0.74	0.04	1.05	0.00	-0.65
ELA	4	1192775	1.40	-0.48	0.09	1.00	0.00	-0.46
ELA	4	1192776	1.23	-0.47	-0.62	0.97	0.00	-1.15
ELA	4	1192780	1.49	-0.39	-0.22	1.39	0.00	-0.66
ELA	4	1192782	1.09	-0.44	-0.19	1.00	0.00	-0.68
ELA	4	1192787	0.87	0.54	-0.26	1.01	0.00	0.26
ELA	4	1192816	1.09	-0.40	-0.28	0.89	0.00	-0.77
ELA	4	1192818	1.14	-0.40	-0.19	0.98	0.00	-0.65
ELA	4	1192823	0.86	0.45	0.45	1.03	0.00	0.87
ELA	4	1192832	1.34	-0.50	0.17	1.01	0.00	-0.37

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	4	1192833	0.93	0.50	-0.87	1.00	0.00	-0.40
ELA	5	1192834	1.19	0.32	-1.04	1.03	0.00	-0.63
ELA	5	1192835	1.45	0.72	0.56	1.19	0.00	1.36
ELA	5	1192836	1.22	-0.40	1.00	1.27	0.00	0.64
ELA	5	1192837	0.87	0.79	-1.50	0.77	0.00	-0.59
ELA	5	1192838	1.06	0.42	1.60	1.10	0.00	2.10
ELA	5	1192843	0.73	0.90	-0.83	0.94	0.00	0.19
ELA	5	1192844	1.03	0.59	-1.87	0.93	0.00	-1.14
ELA	5	1192845	1.35	-0.59	-0.16	0.97	0.00	-0.62
ELA	5	1192846	1.21	1.17	-2.11	0.90	0.00	-0.80
ELA	5	1192852	1.63	-0.72	1.03	1.17	0.00	0.33
ELA	5	1192854	0.78	0.47	-0.26	0.89	0.00	0.22
ELA	5	1192859	1.05	0.88	-1.82	0.97	0.00	-0.74
ELA	5	1192860	1.01	0.63	-1.77	0.97	0.00	-0.94
ELA	5	1192861	2.73	2.10	-3.20	0.91	0.00	-0.89
ELA	5	1192862	1.19	-0.44	0.10	1.01	0.00	-0.32
ELA	5	1192863	0.91	0.65	-0.55	0.95	0.00	0.13
ELA	5	1192864	1.22	-0.57	0.12	1.00	0.00	-0.44
ELA	5	1192872	1.32	0.80	0.23	1.04	0.00	1.10
ELA	5	1192875	0.92	-0.46	0.07	0.90	0.00	-0.40
ELA	5	1192876	1.08	-0.57	0.24	1.01	0.00	-0.35
ELA	5	1192885	1.29	-0.67	0.07	0.95	0.00	-0.60
ELA	5	1192887	1.02	-0.63	0.06	1.04	0.00	-0.57
ELA	5	1192891	1.36	0.89	0.38	0.97	0.00	1.34
ELA	5	1192892	0.93	0.65	-0.94	0.94	0.00	-0.26

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	5	1192895	0.87	0.39	-0.97	0.94	0.00	-0.50
ELA	5	1192896	1.14	0.94	-1.89	0.92	0.00	-0.87
ELA	5	1192900	1.23	-0.46	-0.08	0.95	0.00	-0.46
ELA	5	1192901	0.99	0.57	-0.74	1.01	0.00	-0.15
ELA	5	1192902	1.02	0.55	1.14	1.03	0.00	1.77
ELA	5	1192909	1.43	-0.50	0.40	1.18	0.00	-0.20
ELA	5	1192920	0.66	-0.53	-0.48	0.86	0.00	-0.99
ELA	5	1192921	0.98	-0.45	1.00	1.00	0.00	0.57
ELA	5	1192922	1.54	-0.50	0.56	1.01	0.00	-0.25
ELA	5	1192925	1.45	-0.63	0.02	1.06	0.00	-0.41
ELA	5	1192928	0.97	0.49	1.86	1.06	0.00	2.47
ELA	5	1192930	1.61	1.34	0.55	1.13	0.00	1.99
ELA	5	1192948	1.33	0.14	-0.05	1.07	0.00	0.30
ELA	5	1192956	0.90	0.58	-0.86	0.94	0.00	-0.25
ELA	5	1192960	2.22	-1.09	0.91	1.03	0.00	-0.19
ELA	5	1192962	1.15	0.55	-0.45	0.87	0.00	0.14
ELA	5	1192963	0.99	0.45	0.90	1.01	0.00	1.38
ELA	5	1192967	1.01	0.39	0.59	1.00	0.00	1.02
ELA	5	1192969	0.90	0.49	0.80	0.92	0.00	1.33
ELA	5	1192971	0.96	0.42	-0.22	0.91	0.00	0.21
ELA	5	1192976	0.88	0.41	-0.67	0.92	0.00	-0.24
ELA	5	1192981	0.92	0.56	-1.08	0.92	0.00	-0.49
ELA	5	1192982	1.42	-0.55	0.42	1.01	0.00	-0.11
ELA	5	1192985	1.28	-0.57	0.28	0.98	0.00	-0.23
ELA	5	1192990	1.23	0.59	0.28	1.07	0.00	0.91

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	5	1192992	0.87	0.39	1.50	0.93	0.00	1.94
ELA	5	1193003	1.46	-0.63	0.51	1.05	0.00	-0.14
ELA	5	1193008	1.08	-0.48	-0.18	1.06	0.00	-0.67
ELA	5	1193009	0.84	0.43	-0.29	0.93	0.00	0.14
ELA	5	1193019	0.74	-0.39	-0.42	0.94	0.00	-0.78
ELA	5	1193020	1.41	0.56	-0.57	0.96	0.00	0.03
ELA	5	1193021	1.02	-0.27	1.05	1.04	0.00	0.81
ELA	5	1193027	0.73	-0.47	-0.10	0.88	0.00	-0.55
ELA	5	1193042	0.99	0.49	-0.31	0.93	0.00	0.22
ELA	5	1193044	0.92	0.74	-0.01	0.92	0.00	0.77
ELA	5	1193045	1.00	0.40	-1.04	0.94	0.00	-0.58
ELA	5	1193046	2.26	-0.94	0.90	1.07	0.00	-0.15
ELA	5	1193050	0.89	0.83	-1.22	0.91	0.00	-0.31
ELA	5	1193051	0.99	0.58	-1.61	0.93	0.00	-0.93
ELA	5	1193055	0.96	0.42	0.44	0.94	0.00	0.88
ELA	5	1193057	1.37	-0.48	0.23	1.08	0.00	-0.15
ELA	5	1193058	1.42	-0.45	0.83	1.10	0.00	0.40
ELA	5	1193060	1.31	-0.48	0.43	1.02	0.00	-0.05
ELA	5	1193068	0.92	0.39	-1.58	0.93	0.00	-1.08
ELA	5	1193074	0.92	0.41	-1.60	0.94	0.00	-1.06
ELA	5	1193078	0.89	0.41	1.53	0.95	0.00	2.01
ELA	5	1193080	1.13	0.54	0.79	1.07	0.00	1.38
ELA	5	1193081	0.97	0.43	0.59	0.92	0.00	1.09
ELA	5	1193082	0.86	0.63	1.16	0.91	0.00	1.86
ELA	5	1193083	1.01	0.48	1.12	1.06	0.00	1.68

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	5	1193086	0.96	0.44	1.06	1.02	0.00	1.57
ELA	5	1193089	0.93	0.52	0.97	0.97	0.00	1.55
ELA	5	1193090	1.06	0.39	0.40	1.01	0.00	0.85
ELA	5	1193104	0.96	0.18	-0.07	0.95	0.00	0.13
ELA	5	1193107	1.02	0.38	0.71	1.02	0.00	1.12
ELA	5	1193116	0.90	0.52	-0.41	0.95	0.00	0.12
ELA	5	1193123	0.94	0.51	-0.50	0.92	0.00	0.02
ELA	5	1193125	1.13	0.56	1.08	1.11	0.00	1.71
ELA	5	1193139	1.16	-0.49	0.37	0.93	0.00	-0.15
ELA	5	1193148	0.88	0.52	-0.40	0.92	0.00	0.13
ELA	5	1193152	1.21	-0.45	-0.42	1.01	0.00	-0.83
ELA	5	1193153	1.35	-0.38	0.72	1.02	0.00	0.37
ELA	5	1193155	0.99	0.59	-1.48	0.94	0.00	-0.85
ELA	5	1193156	1.49	0.47	0.34	1.17	0.00	0.84
ELA	5	1193157	1.20	0.56	-0.24	0.96	0.00	0.40
ELA	5	1193159	0.89	0.47	0.92	0.92	0.00	1.47
ELA	5	1193163	0.96	0.44	0.44	0.94	0.00	0.93
ELA	5	1193164	0.90	0.63	0.78	0.96	0.00	1.46
ELA	5	1193170	1.06	0.62	-1.98	0.92	0.00	-1.28
ELA	5	1193179	1.32	-0.51	0.52	1.16	0.00	0.00
ELA	5	1193182	0.92	0.47	-1.51	0.94	0.00	-0.94
ELA	5	1193188	0.92	0.41	-0.77	0.95	0.00	-0.33
ELA	6	1193194	1.31	-0.53	0.20	0.98	0.00	-0.37
ELA	6	1193196	1.02	-0.41	-0.86	0.95	0.00	-1.30
ELA	6	1193199	1.64	-0.90	0.87	1.05	0.00	-0.09

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	6	1193200	1.26	1.09	-1.90	0.94	0.00	-0.76
ELA	6	1193202	1.88	-1.08	0.48	1.29	0.00	-0.56
ELA	6	1193204	0.90	-0.37	1.40	0.92	0.00	1.09
ELA	6	1193205	0.93	0.54	-1.27	0.93	0.00	-0.71
ELA	6	1193206	0.89	0.65	-1.17	0.91	0.00	-0.50
ELA	6	1193207	1.99	-0.66	0.87	1.68	0.00	0.20
ELA	6	1193208	1.09	0.74	-0.50	0.92	0.00	0.25
ELA	6	1193219	1.59	0.75	-0.12	1.03	0.00	0.73
ELA	6	1193227	0.97	0.81	-0.83	0.92	0.00	-0.02
ELA	6	1193234	1.62	-0.69	1.04	1.01	0.00	0.32
ELA	6	1193235	1.18	1.02	-1.64	0.92	0.00	-0.61
ELA	6	1193236	1.30	-0.62	-0.23	1.05	0.00	-0.85
ELA	6	1193239	3.66	-1.64	1.62	1.07	0.00	-0.08
ELA	6	1193242	1.21	0.55	-1.56	0.90	0.00	-0.90
ELA	6	1193243	1.07	0.43	1.76	1.09	0.00	2.34
ELA	6	1193245	1.12	0.32	-0.24	0.94	0.00	0.22
ELA	6	1193250	1.10	0.76	-1.81	0.93	0.00	-1.01
ELA	6	1193259	0.91	0.66	-0.93	0.97	0.00	-0.24
ELA	6	1193260	1.02	0.77	-1.30	0.97	0.00	-0.50
ELA	6	1193262	1.82	-0.93	0.60	0.99	0.00	-0.31
ELA	6	1193265	1.08	0.65	-2.16	0.91	0.00	-1.38
ELA	6	1193267	1.46	1.20	-2.49	0.93	0.00	-1.17
ELA	6	1193268	0.96	0.71	-1.64	0.94	0.00	-0.81
ELA	6	1193269	1.53	-0.71	0.03	0.98	0.00	-0.58
ELA	6	1193280	0.97	0.53	-1.09	0.93	0.00	-0.57

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	6	1193281	1.68	-0.25	2.23	1.69	0.00	2.02
ELA	6	1193282	1.80	-0.66	1.20	1.57	0.00	0.51
ELA	6	1193289	1.15	-0.47	0.29	1.02	0.00	-0.23
ELA	6	1193291	1.10	-0.43	0.39	0.98	0.00	-0.09
ELA	6	1193292	1.28	-0.61	0.22	1.07	0.00	-0.48
ELA	6	1193293	0.90	0.42	-0.97	0.91	0.00	-0.59
ELA	6	1193306	0.64	-0.50	-0.58	0.89	0.00	-1.01
ELA	6	1193307	0.98	0.61	1.71	1.00	0.00	2.43
ELA	6	1193314	1.06	0.60	0.51	1.02	0.00	1.20
ELA	6	1193315	0.95	0.55	0.76	0.96	0.00	1.39
ELA	6	1193316	0.91	-0.09	0.49	0.95	0.00	0.47
ELA	6	1193318	0.87	-0.44	1.29	0.88	0.00	0.93
ELA	6	1193320	1.17	0.44	0.24	1.03	0.00	0.80
ELA	6	1193321	1.01	0.61	0.58	0.91	0.00	1.32
ELA	6	1193323	0.91	0.41	0.88	0.90	0.00	1.43
ELA	6	1193324	1.02	0.36	1.23	1.06	0.00	1.72
ELA	6	1193325	2.23	-1.30	0.38	0.97	0.00	-0.99
ELA	6	1193327	2.02	-1.01	0.54	1.01	0.00	-0.53
ELA	6	1193328	0.92	0.57	-1.25	0.96	0.00	-0.72
ELA	6	1193330	1.15	0.53	0.89	1.09	0.00	1.54
ELA	6	1193334	1.07	0.41	1.06	1.05	0.00	1.59
ELA	6	1193335	0.96	0.54	1.38	0.96	0.00	2.03
ELA	6	1193362	0.99	0.57	0.16	0.91	0.00	0.79
ELA	6	1193375	1.02	0.34	1.04	1.01	0.00	1.47
ELA	6	1193377	1.11	0.35	0.84	1.08	0.00	1.28

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	6	1193381	0.99	0.37	1.14	1.02	0.00	1.60
ELA	6	1193394	1.35	0.35	1.39	1.10	0.00	1.78
ELA	6	1193399	1.04	0.42	0.66	0.97	0.00	1.19
ELA	6	1193401	1.19	-0.45	0.51	1.00	0.00	0.05
ELA	6	1193405	0.89	0.55	-1.21	0.86	0.00	-0.65
ELA	6	1193422	1.07	0.54	1.15	1.07	0.00	1.78
ELA	6	1193424	0.98	0.46	1.11	0.93	0.00	1.69
ELA	6	1193427	0.58	-0.45	-1.10	0.87	0.00	-1.44
ELA	6	1193428	0.92	0.35	1.32	0.91	0.00	1.78
ELA	6	1193431	1.11	0.39	0.00	0.99	0.00	0.46
ELA	6	1193433	1.04	0.53	1.33	1.07	0.00	1.96
ELA	6	1193435	0.82	0.51	-0.14	0.90	0.00	0.33
ELA	6	1193436	1.42	-0.82	0.46	1.01	0.00	-0.43
ELA	6	1193440	1.64	-0.81	0.32	1.06	0.00	-0.56
ELA	6	1193443	1.11	-0.47	0.15	1.10	0.00	-0.31
ELA	6	1193447	0.83	0.51	1.63	0.91	0.00	2.27
ELA	6	1193449	0.97	0.39	1.11	0.99	0.00	1.62
ELA	6	1193452	1.10	0.50	0.85	1.08	0.00	1.46
ELA	6	1193453	1.31	-0.49	0.36	0.90	0.00	-0.12
ELA	6	1193461	0.93	0.46	1.29	0.97	0.00	1.83
ELA	6	1193482	0.87	0.44	-0.14	0.95	0.00	0.32
ELA	6	1193491	1.29	-0.67	0.62	1.01	0.00	-0.08
ELA	6	1193507	0.97	0.41	1.67	0.93	0.00	2.21
ELA	6	1193522	0.76	0.52	0.97	0.98	0.00	1.50
ELA	6	1193523	2.24	-1.06	0.43	1.20	0.00	-0.80

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	6	1193524	1.48	-0.73	-0.25	1.11	0.00	-1.06
ELA	6	1193526	1.77	-0.82	0.54	1.13	0.00	-0.39
ELA	6	1193527	1.62	-0.75	0.40	1.09	0.00	-0.46
ELA	6	1193528	0.90	0.45	-1.54	0.89	0.00	-1.16
ELA	6	1193537	2.43	-1.02	1.43	1.08	0.00	0.38
ELA	6	1193538	1.34	-0.53	0.33	0.99	0.00	-0.23
ELA	6	1193542	1.14	0.49	0.38	1.06	0.00	0.95
ELA	6	1193543	1.05	0.63	0.86	1.05	0.00	1.58
ELA	7	1193554	0.82	0.45	-1.27	0.69	0.00	-0.82
ELA	7	1193555	0.84	-0.53	0.06	0.97	0.00	-0.45
ELA	7	1193561	1.56	-0.60	0.97	1.01	0.00	0.37
ELA	7	1193562	0.89	0.43	-0.67	0.98	0.00	-0.24
ELA	7	1193563	1.72	-0.77	0.71	1.03	0.00	-0.06
ELA	7	1193564	0.88	-0.60	0.89	0.98	0.00	0.34
ELA	7	1193566	1.01	-0.03	2.28	1.01	0.00	2.31
ELA	7	1193569	1.38	-0.87	0.21	0.96	0.00	-0.68
ELA	7	1193571	0.87	0.47	0.48	1.00	0.00	0.94
ELA	7	1193573	1.00	0.67	-1.11	0.93	0.00	-0.44
ELA	7	1193575	1.01	0.45	-1.17	0.97	0.00	-0.68
ELA	7	1193578	1.05	1.01	-1.20	0.96	0.00	-0.16
ELA	7	1193582	0.98	0.43	-0.65	0.97	0.00	-0.16
ELA	7	1193587	0.86	1.03	-0.38	0.93	0.00	0.71
ELA	7	1193588	1.24	1.51	-1.90	0.95	0.00	-0.36
ELA	7	1193589	1.34	0.16	-0.52	1.37	0.00	-0.34
ELA	7	1193590	2.20	-0.24	0.83	2.16	0.00	0.58

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	7	1193594	3.08	-1.51	1.45	1.04	0.00	-0.12
ELA	7	1193596	1.42	-0.68	1.17	1.14	0.00	0.51
ELA	7	1193598	1.52	-0.95	0.72	1.06	0.00	-0.28
ELA	7	1193602	1.51	-1.25	1.54	1.32	0.00	-0.01
ELA	7	1193607	0.74	-0.62	0.54	0.95	0.00	-0.01
ELA	7	1193608	1.33	0.54	0.52	1.08	0.00	1.12
ELA	7	1193609	1.04	-0.49	1.93	0.98	0.00	1.51
ELA	7	1193610	1.13	0.91	-1.92	0.90	0.00	-0.90
ELA	7	1193614	0.72	-0.55	0.03	0.98	0.00	-0.45
ELA	7	1193615	1.02	0.44	1.52	0.99	0.00	2.03
ELA	7	1193616	0.91	0.12	-1.71	0.82	0.00	-1.53
ELA	7	1193619	1.07	0.93	-1.71	0.92	0.00	-0.68
ELA	7	1193623	1.79	-0.74	0.63	1.51	0.00	-0.12
ELA	7	1193624	0.91	0.43	-0.75	1.01	0.00	-0.22
ELA	7	1193626	0.91	0.45	-0.14	0.90	0.00	0.33
ELA	7	1193627	1.01	0.67	-0.03	1.00	0.00	0.66
ELA	7	1193631	0.98	0.40	1.00	0.97	0.00	1.46
ELA	7	1193637	1.18	0.61	1.10	1.09	0.00	1.78
ELA	7	1193638	0.61	-0.62	-0.24	0.90	0.00	-0.79
ELA	7	1193640	0.99	0.64	0.78	0.97	0.00	1.48
ELA	7	1193647	2.19	-1.12	1.27	1.02	0.00	0.13
ELA	7	1193650	0.85	0.64	-0.16	0.96	0.00	0.48
ELA	7	1193663	0.91	0.42	-0.26	0.87	0.00	0.16
ELA	7	1193668	1.06	-0.63	0.00	1.01	0.00	-0.62
ELA	7	1193670	0.93	0.45	1.00	0.88	0.00	1.51

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	7	1193676	1.04	0.40	0.92	0.95	0.00	1.40
ELA	7	1193677	1.12	0.40	1.50	1.12	0.00	1.99
ELA	7	1193681	1.12	0.35	0.29	0.97	0.00	0.72
ELA	7	1193684	1.02	0.34	1.44	1.02	0.00	1.86
ELA	7	1193691	1.37	-0.61	0.36	0.99	0.00	-0.28
ELA	7	1193692	1.85	-0.83	1.15	1.01	0.00	0.28
ELA	7	1193697	2.08	-0.87	0.63	1.02	0.00	-0.12
ELA	7	1193699	0.92	0.69	-0.67	0.94	0.00	0.06
ELA	7	1193700	0.82	0.64	-0.14	0.96	0.00	0.54
ELA	7	1193701	1.27	-0.18	0.97	1.27	0.00	0.83
ELA	7	1193703	0.92	0.43	-0.47	0.90	0.00	-0.03
ELA	7	1193710	0.97	0.46	0.45	0.87	0.00	0.96
ELA	7	1193713	0.86	0.44	-0.77	0.89	0.00	-0.32
ELA	7	1193714	0.88	0.70	-0.86	0.93	0.00	-0.14
ELA	7	1193716	1.37	-0.49	0.67	1.06	0.00	0.19
ELA	7	1193717	1.67	-0.84	0.68	1.07	0.00	-0.16
ELA	7	1193719	1.70	-0.80	0.76	0.97	0.00	-0.04
ELA	7	1193720	0.77	0.44	0.29	0.97	0.00	0.76
ELA	7	1193721	0.85	0.39	-0.75	0.90	0.00	-0.33
ELA	7	1193726	1.09	1.05	0.82	1.04	0.00	1.95
ELA	7	1193728	0.99	0.34	1.16	1.02	0.00	1.59
ELA	7	1193731	1.10	0.77	-2.02	0.90	0.00	-1.26
ELA	7	1193732	1.56	-0.67	1.31	1.07	0.00	0.58
ELA	7	1193735	1.19	-0.45	0.03	0.99	0.00	-0.47
ELA	7	1193736	1.34	-0.57	0.43	1.12	0.00	-0.08

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	7	1193738	1.08	0.68	1.47	0.96	0.00	2.24
ELA	7	1193740	1.40	-0.39	1.13	1.29	0.00	0.77
ELA	7	1193744	1.00	0.87	1.61	0.95	0.00	2.58
ELA	7	1193761	2.12	-1.10	0.94	1.05	0.00	-0.20
ELA	7	1193772	0.81	0.48	0.75	0.96	0.00	1.25
ELA	7	1193779	1.03	0.36	0.31	1.01	0.00	0.74
ELA	7	1193781	0.91	0.52	0.74	0.93	0.00	1.34
ELA	7	1193783	0.90	1.04	1.40	0.91	0.00	2.56
ELA	7	1193796	1.48	-0.58	1.04	1.04	0.00	0.48
ELA	7	1193798	0.92	0.55	-0.47	0.98	0.00	0.08
ELA	7	1193820	0.91	0.46	0.45	0.94	0.00	0.97
ELA	7	1193846	0.94	0.50	1.67	1.01	0.00	2.24
ELA	7	1193849	1.13	0.66	0.93	1.11	0.00	1.65
ELA	7	1193854	0.97	0.36	2.07	1.03	0.00	2.51
ELA	7	1193859	1.09	0.50	0.81	1.07	0.00	1.38
ELA	7	1193862	0.99	0.45	-0.21	0.87	0.00	0.28
ELA	7	1193863	0.97	0.47	0.07	0.91	0.00	0.59
ELA	7	1193866	1.05	0.44	0.42	0.98	0.00	0.94
ELA	7	1193868	0.89	0.46	1.01	0.93	0.00	1.55
ELA	7	1193875	1.13	0.48	1.09	1.07	0.00	1.59
ELA	7	1193879	0.93	0.42	0.93	0.97	0.00	1.41
ELA	7	1193882	1.37	-0.59	0.41	1.02	0.00	-0.18
ELA	7	1193885	0.90	0.66	-0.86	0.91	0.00	-0.18
ELA	7	1193888	1.33	-0.46	0.97	1.01	0.00	0.49
ELA	7	1193899	1.37	-0.59	0.28	1.13	0.00	-0.41

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	7	1193903	1.43	-0.52	0.59	1.19	0.00	0.02
ELA	7	1193911	1.79	-0.79	0.68	1.11	0.00	-0.03
ELA	7	1193914	2.45	-1.15	0.91	1.15	0.00	-0.31
ELA	7	1193916	1.76	-0.84	0.51	1.06	0.00	-0.37
ELA	8	1193926	0.99	-0.36	1.17	1.00	0.00	0.86
ELA	8	1193931	0.91	-0.56	1.16	0.92	0.00	0.65
ELA	8	1193933	0.92	0.54	-0.39	0.97	0.00	0.17
ELA	8	1193934	1.13	0.79	-1.19	0.96	0.00	-0.38
ELA	8	1193935	1.18	-0.44	0.14	1.02	0.00	-0.28
ELA	8	1193937	1.63	-0.83	0.73	1.02	0.00	-0.13
ELA	8	1193938	0.84	0.49	-0.25	0.95	0.00	0.22
ELA	8	1193939	1.98	-0.97	1.01	1.14	0.00	-0.01
ELA	8	1193941	0.94	0.77	-0.94	0.95	0.00	-0.11
ELA	8	1193942	0.92	0.40	-0.41	1.01	0.00	0.05
ELA	8	1193943	0.94	0.48	0.31	1.10	0.00	0.84
ELA	8	1193944	1.63	-0.86	0.52	1.04	0.00	-0.39
ELA	8	1193945	1.54	-0.65	0.96	1.05	0.00	0.25
ELA	8	1193946	1.52	-0.79	0.74	0.96	0.00	-0.11
ELA	8	1193950	2.08	-1.04	1.12	0.98	0.00	0.08
ELA	8	1193951	0.92	0.63	-0.93	0.92	0.00	-0.28
ELA	8	1193952	0.94	0.54	-0.93	0.94	0.00	-0.37
ELA	8	1193954	0.93	0.56	-1.23	0.92	0.00	-0.71
ELA	8	1193955	1.02	-0.37	1.49	0.91	0.00	1.08
ELA	8	1193957	0.94	0.45	-0.59	1.01	0.00	-0.12
ELA	8	1193959	2.45	-1.14	1.50	1.03	0.00	0.32

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	8	1193960	1.44	-0.61	1.12	1.42	0.00	0.53
ELA	8	1193961	1.19	0.09	0.97	1.25	0.00	1.12
ELA	8	1193964	1.07	0.39	1.01	1.12	0.00	1.47
ELA	8	1193968	1.09	0.73	1.81	1.06	0.00	2.67
ELA	8	1193970	1.12	-0.47	0.32	1.03	0.00	-0.20
ELA	8	1193974	0.93	-0.50	0.58	0.96	0.00	0.12
ELA	8	1193977	0.97	0.07	1.14	0.98	0.00	1.25
ELA	8	1193989	1.03	0.47	1.44	1.02	0.00	2.02
ELA	8	1193992	1.03	0.59	1.28	0.96	0.00	1.98
ELA	8	1194006	1.25	-0.53	0.31	1.05	0.00	-0.25
ELA	8	1194009	0.94	0.01	2.07	0.94	0.00	2.17
ELA	8	1194010	1.05	0.22	1.16	0.99	0.00	1.48
ELA	8	1194011	0.53	-0.63	-0.77	0.86	0.00	-1.31
ELA	8	1194013	0.81	0.42	1.41	0.94	0.00	1.89
ELA	8	1194014	0.91	0.60	0.41	0.89	0.00	1.06
ELA	8	1194024	1.52	-0.44	1.28	1.25	0.00	0.83
ELA	8	1194026	1.61	-0.72	0.80	1.06	0.00	0.16
ELA	8	1194027	1.34	-0.43	0.64	1.05	0.00	0.17
ELA	8	1194032	1.32	0.29	1.53	1.32	0.00	1.92
ELA	8	1194037	0.98	-0.50	0.06	1.14	0.00	-0.43
ELA	8	1194039	1.00	-0.46	0.55	1.01	0.00	0.09
ELA	8	1194040	1.32	-0.44	1.01	1.19	0.00	0.55
ELA	8	1194043	1.00	0.35	2.12	1.00	0.00	2.58
ELA	8	1194049	1.16	0.34	1.79	1.04	0.00	2.23
ELA	8	1194050	0.95	0.54	1.98	1.00	0.00	2.68

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	8	1194055	0.87	0.40	-0.29	0.96	0.00	0.09
ELA	8	1194057	1.04	0.58	-1.45	0.91	0.00	-0.88
ELA	8	1194058	1.37	-0.77	0.42	0.99	0.00	-0.41
ELA	8	1194059	3.09	-1.50	1.39	1.11	0.00	-0.17
ELA	8	1194080	0.99	0.48	-0.61	0.96	0.00	-0.11
ELA	8	1194083	1.38	-0.35	1.59	1.34	0.00	1.25
ELA	8	1194088	1.15	-0.44	0.72	1.03	0.00	0.28
ELA	8	1194096	1.03	0.34	1.32	1.05	0.00	1.77
ELA	8	1194097	0.94	0.48	1.50	0.97	0.00	2.10
ELA	8	1194098	1.04	0.40	1.23	1.05	0.00	1.73
ELA	8	1194102	1.09	0.51	1.26	1.00	0.00	1.91
ELA	8	1194104	1.01	0.56	1.55	1.01	0.00	2.22
ELA	8	1194106	0.99	0.40	1.52	0.99	0.00	2.03
ELA	8	1194113	0.92	0.38	1.57	0.94	0.00	2.04
ELA	8	1194115	1.18	0.36	0.39	1.06	0.00	0.84
ELA	8	1194118	0.98	0.60	1.36	0.99	0.00	2.06
ELA	8	1194119	1.27	-0.47	0.45	0.99	0.00	-0.12
ELA	8	1194120	1.52	-0.69	0.55	1.07	0.00	-0.21
ELA	8	1194121	1.15	-0.41	0.04	0.99	0.00	-0.43
ELA	8	1194126	1.17	0.39	0.38	1.07	0.00	0.84
ELA	8	1194139	0.97	0.47	-0.24	0.92	0.00	0.24
ELA	8	1194145	1.00	0.43	1.11	1.00	0.00	1.62
ELA	8	1194154	0.98	0.52	-0.50	0.93	0.00	0.02
ELA	8	1194155	1.19	-0.50	0.62	1.03	0.00	0.10
ELA	8	1194159	1.29	-0.45	0.32	1.03	0.00	-0.16

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	8	1194161	1.44	-0.55	0.80	1.01	0.00	0.21
ELA	8	1194170	1.21	0.50	0.41	0.95	0.00	1.04
ELA	8	1194171	1.06	0.42	1.17	0.97	0.00	1.72
ELA	8	1194178	1.15	-0.56	0.97	0.97	0.00	0.40
ELA	8	1194185	1.02	0.54	-0.60	0.94	0.00	-0.04
ELA	8	1194204	1.04	0.64	1.44	1.11	0.00	2.16
ELA	8	1194208	0.98	0.42	-1.51	0.91	0.00	-1.11
ELA	8	1194209	1.60	-0.74	0.40	1.05	0.00	-0.37
ELA	8	1194213	1.00	0.66	-0.54	0.95	0.00	0.14
ELA	8	1194219	0.98	0.40	-1.32	0.92	0.00	-0.90
ELA	8	1194225	1.01	0.53	1.54	1.03	0.00	2.20
ELA	8	1194227	1.09	0.54	0.97	1.00	0.00	1.64
ELA	8	1194229	1.16	0.66	0.93	1.03	0.00	1.72
ELA	8	1194230	1.13	-0.42	0.22	0.98	0.00	-0.19
ELA	8	1194235	1.24	-0.41	0.84	0.98	0.00	0.42
ELA	8	1194238	1.14	-0.49	0.56	1.00	0.00	0.05
ELA	8	1194242	1.47	0.31	1.40	1.27	0.00	1.77
ELA	8	1194244	0.98	0.43	1.26	0.98	0.00	1.78
ELA	8	1194245	0.92	0.33	1.42	0.94	0.00	1.85
ELA	8	1194259	0.98	0.58	0.15	0.92	0.00	0.74
ELA	8	1194260	0.87	0.55	0.37	0.95	0.00	0.95
ELA	8	1194262	0.90	0.50	0.15	0.93	0.00	0.68
ELA	8	1194267	0.86	0.55	1.09	0.97	0.00	1.69
ELA	8	1194268	1.03	0.58	0.21	0.95	0.00	0.83
ELA	8	1194271	1.16	0.55	0.84	1.04	0.00	1.54

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
ELA	8	1194272	0.92	0.45	1.65	0.96	0.00	2.26
ELA	8	1194274	1.04	0.63	1.32	1.02	0.00	2.11
ELA	8	1194276	1.02	0.67	1.16	0.94	0.00	1.99
Literature	HS	1194286	1.17	0.42	0.41	0.95	0.00	0.90
Literature	HS	1194289	1.08	0.87	1.86	1.06	0.00	2.82
Literature	HS	1194302	1.02	0.69	0.89	0.93	0.00	1.62
Literature	HS	1194304	1.23	0.42	0.32	1.10	0.00	0.77
Literature	HS	1194306	1.05	0.00	2.02	1.06	0.00	2.06
Literature	HS	1194307	1.27	0.48	0.14	1.08	0.00	0.65
Literature	HS	1194310	1.29	0.68	0.96	1.03	0.00	1.71
Literature	HS	1194313	1.03	0.60	1.43	0.92	0.00	2.10
Literature	HS	1194315	0.99	0.48	1.69	0.96	0.00	2.23
Literature	HS	1194317	0.74	-0.53	0.71	0.93	0.00	0.22
Literature	HS	1194318	1.00	0.63	2.15	1.09	0.00	2.85
Literature	HS	1194319	0.90	-0.42	1.50	0.96	0.00	1.12
Literature	HS	1194321	1.39	0.57	-0.09	1.09	0.00	0.52
Literature	HS	1194328	1.10	0.42	0.51	1.04	0.00	0.97
Literature	HS	1194334	0.95	-0.08	0.52	1.00	0.00	0.48
Literature	HS	1194339	1.10	0.40	1.21	1.01	0.00	1.66
Literature	HS	1194341	1.49	-0.82	1.02	0.99	0.00	0.17
Literature	HS	1194346	1.19	1.23	-1.43	0.92	0.00	-0.13
Literature	HS	1194347	1.25	-0.48	0.47	0.95	0.00	0.06
Literature	HS	1194348	2.45	-1.02	1.87	1.11	0.00	0.90
Literature	HS	1194353	0.82	0.46	-0.52	0.90	0.00	0.00
Literature	HS	1194358	0.75	-0.45	0.58	0.87	0.00	0.10

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Literature	HS	1194360	1.06	-0.75	2.04	0.94	0.00	1.28
Literature	HS	1194383	1.02	0.42	0.41	0.91	0.00	0.86
Literature	HS	1194387	1.63	0.82	0.31	1.02	0.00	1.24
Literature	HS	1194388	1.11	0.50	1.61	1.04	0.00	2.23
Literature	HS	1194390	1.17	0.42	0.79	0.98	0.00	1.31
Literature	HS	1194391	1.13	0.49	1.64	1.08	0.00	2.24
Literature	HS	1194398	0.81	-0.48	0.95	0.91	0.00	0.48
Literature	HS	1194405	0.93	-0.51	1.42	0.96	0.00	0.90
Literature	HS	1194411	1.77	-0.87	0.83	0.98	0.00	0.06
Literature	HS	1194412	1.49	-0.88	0.42	1.03	0.00	-0.45
Literature	HS	1194413	0.84	0.46	0.24	0.98	0.00	0.72
Literature	HS	1194415	1.01	0.76	-1.17	0.94	0.00	-0.29
Literature	HS	1194417	0.87	0.31	-0.11	0.97	0.00	0.32
Literature	HS	1194419	1.08	0.77	0.30	1.06	0.00	1.12
Literature	HS	1194423	2.99	-1.34	2.01	1.03	0.00	0.65
Literature	HS	1194424	0.91	0.44	-0.30	0.96	0.00	0.17
Literature	HS	1194425	0.92	0.45	-0.32	0.96	0.00	0.15
Literature	HS	1194428	0.71	-0.56	0.65	0.85	0.00	0.11
Literature	HS	1194429	0.79	-0.43	0.58	0.92	0.00	0.15
Literature	HS	1194430	1.25	0.58	0.99	1.10	0.00	1.60
Literature	HS	1194433	0.82	-0.28	-0.14	0.90	0.00	-0.44
Literature	HS	1194435	1.09	-0.50	1.36	0.97	0.00	0.83
Literature	HS	1194439	1.74	1.55	-2.03	0.94	0.00	-0.36
Literature	HS	1194440	0.96	0.58	-1.00	0.95	0.00	-0.30
Literature	HS	1194442	1.03	0.36	0.63	1.07	0.00	1.05

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Literature	HS	1194444	1.16	1.24	-0.33	0.98	0.00	0.97
Literature	HS	1194448	0.98	0.86	-0.08	0.99	0.00	0.83
Literature	HS	1194449	0.91	0.51	0.33	0.90	0.00	0.88
Literature	HS	1194453	1.59	-0.69	1.29	1.12	0.00	0.58
Literature	HS	1194454	1.56	-0.76	1.04	1.02	0.00	0.27
Literature	HS	1194455	0.87	0.46	0.16	0.96	0.00	0.63
Literature	HS	1194456	0.93	0.50	-0.39	0.92	0.00	0.12
Literature	HS	1194466	1.03	1.20	-0.42	1.00	0.00	0.81
Literature	HS	1194476	1.21	1.25	-1.30	0.96	0.00	0.03
Literature	HS	1194487	1.33	-0.67	0.82	1.15	0.00	0.14
Literature	HS	1194488	0.87	0.78	-0.35	0.94	0.00	0.46
Literature	HS	1194492	1.28	0.58	0.63	1.01	0.00	1.28
Literature	HS	1194494	0.99	0.34	1.88	1.02	0.00	2.31
Literature	HS	1194496	1.14	0.86	1.51	1.07	0.00	2.46
Literature	HS	1194503	1.93	-1.00	0.89	1.00	0.00	-0.04
Literature	HS	1194505	0.97	0.82	-0.82	0.98	0.00	0.08
Literature	HS	1194506	4.86	-1.85	2.28	1.05	0.00	0.48
Literature	HS	1194510	1.10	-0.42	0.71	1.09	0.00	0.29
Literature	HS	1194512	0.98	-0.42	0.56	0.99	0.00	0.13
Literature	HS	1194513	0.93	-0.57	0.56	0.93	0.00	-0.01
Literature	HS	1194528	1.66	-0.80	1.25	1.02	0.00	0.43
Literature	HS	1194537	0.99	0.52	1.98	1.00	0.00	2.55
Literature	HS	1194542	1.05	0.46	0.47	0.91	0.00	0.94
Literature	HS	1194570	0.95	0.36	1.71	0.97	0.00	2.13
Literature	HS	1194572	1.53	-0.60	1.42	1.15	0.00	0.69

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Literature	HS	1194576	1.06	0.42	1.53	1.12	0.00	1.99
Literature	HS	1194577	1.51	-0.37	1.04	1.32	0.00	0.71
Literature	HS	1194580	1.54	-0.33	0.75	1.35	0.00	0.47
Literature	HS	1194581	0.98	-0.46	-0.57	1.04	0.00	-0.88
Literature	HS	1194582	1.24	0.20	0.53	1.12	0.00	0.84
Literature	HS	1194583	1.20	0.18	1.92	1.19	0.00	2.23
Literature	HS	1194584	0.96	0.35	2.11	0.95	0.00	2.58
Literature	HS	1194585	1.59	0.75	0.46	1.11	0.00	1.33
Literature	HS	1194586	1.79	-1.04	1.23	1.47	0.00	0.14
Literature	HS	1194587	1.00	0.70	2.17	1.03	0.00	2.91
Literature	HS	1194589	0.87	0.53	-0.61	0.91	0.00	-0.02
Literature	HS	1194596	0.98	0.52	-0.12	0.92	0.00	0.39
Literature	HS	1194597	1.28	-0.46	1.12	1.11	0.00	0.64
Literature	HS	1194601	1.10	0.61	0.39	0.88	0.00	1.06
Literature	HS	1194603	1.01	0.41	1.30	0.94	0.00	1.79
Literature	HS	1194605	1.03	0.34	1.56	1.01	0.00	1.98
Literature	HS	1194606	1.45	0.65	0.40	1.11	0.00	1.12
Literature	HS	1194608	0.83	0.35	2.86	0.96	0.00	3.29
Literature	HS	1194626	1.18	-0.44	0.28	1.02	0.00	-0.09
Literature	HS	1194627	1.31	-0.49	0.72	1.01	0.00	0.22
Literature	HS	1194629	1.89	-0.93	1.15	1.01	0.00	0.20
Literature	HS	1194630	0.82	0.81	-0.21	0.96	0.00	0.59
Literature	HS	1194633	1.01	0.46	-0.21	1.01	0.00	0.28
Literature	HS	1194634	1.00	0.57	-0.50	0.96	0.00	0.11
Literature	HS	1194662	1.36	-0.06	1.28	1.37	0.00	1.23

Content Area	Grade/Level	ItemID	Admin Infit (MnSq)	Admin Displacement	Admin Rasch Difficulty	Updated Infit (MnSq)	Updated Displacement	Updated Rasch Difficulty
Literature	HS	1194666	1.24	0.51	1.41	1.13	0.00	1.95
Literature	HS	1194681	0.97	0.45	1.72	1.02	0.00	2.23
Literature	HS	1194682	1.15	0.40	0.89	1.08	0.00	1.33
Literature	HS	1194685	1.26	0.42	0.45	1.13	0.00	0.92
Literature	HS	1194688	0.93	0.61	0.89	1.00	0.00	1.54
Literature	HS	1194692	0.93	0.49	-0.05	0.95	0.00	0.47
Literature	HS	1194697	1.09	0.62	0.69	1.07	0.00	1.33
Literature	HS	1194710	1.19	-0.61	1.06	0.80	0.00	0.47
Literature	HS	1194712	1.82	0.48	0.57	1.42	0.00	1.31
Literature	HS	1194715	1.30	-0.26	1.12	1.34	0.00	0.90
Literature	HS	1194721	1.06	0.52	-0.20	0.90	0.00	0.32
Literature	HS	1194728	0.90	0.44	-0.32	0.92	0.00	0.15
Literature	HS	1194733	1.28	-0.43	0.74	1.13	0.00	0.28
Literature	HS	1194740	1.21	-0.54	0.89	0.98	0.00	0.60
Literature	HS	1194741	1.59	-0.61	1.27	1.04	0.00	0.70
Literature	HS	1194743	1.04	0.56	-1.06	0.95	0.00	-0.46
Literature	HS	1194745	0.97	0.50	-0.67	0.96	0.00	-0.13
Literature	HS	1194750	0.75	0.63	1.36	0.94	0.00	1.99
Literature	HS	1194753	0.98	0.45	0.55	1.05	0.00	1.00
Literature	HS	1194761	1.36	-0.47	0.74	1.12	0.00	0.28
Literature	HS	1194763	1.04	-0.49	-0.15	0.97	0.00	-0.63
Literature	HS	1194766	1.26	-0.44	0.57	1.02	0.00	0.15
Literature	HS	1194767	0.95	0.71	-0.62	0.93	0.00	0.12

APPENDIX D: SUBSCORE RELIABILITY

As discussed in Chapter Sixteen, subscore reliability was calculated for each reporting category. Because subscores are based on fewer items than the total test, it is expected that subscore reliabilities are lower than total test score reliability. Furthermore, it can be noted that reporting categories with fewer items (i.e., 6 items) have lower reliabilities than reporting categories with more items.

Content Area	Grade/ Level	Reporting Category	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Mathematics	3	Numbers and Operations in Base Ten	31621	6	6	6.00	468.15	175.24	81.84	0.74
Mathematics	3	Numbers and Operations – Fractions	31621	6	6	6.00	448.44	130.00	72.44	0.65
Mathematics	3	Operations and Algebraic Thinking	31621	9	9	9.00	486.58	120.20	56.25	0.76
Mathematics	3	Geometry	31621	6	6	6.00	470.46	135.29	71.30	0.69
Mathematics	3	Measurement and Data	31621	9	9	9.00	456.59	104.16	55.04	0.70
Mathematics	4	Numbers and Operations in Base Ten	29043	7	7	7.00	518.57	134.74	67.60	0.71
Mathematics	4	Numbers and Operations – Fractions	29043	8	8	8.00	503.62	134.04	63.93	0.74
Mathematics	4	Operations and Algebraic Thinking	29043	9	9	9.00	504.40	104.80	54.20	0.72
Mathematics	4	Geometry	29043	6	6	6.00	465.38	127.35	75.27	0.59
Mathematics	4	Measurement and Data	29043	6	6	6.00	492.37	121.59	71.64	0.60
Mathematics	5	Numbers and Operations in Base Ten	31646	8	8	8.00	547.91	126.82	62.94	0.72
Mathematics	5	Numbers and Operations – Fractions	31646	10	10	10.00	547.17	110.23	52.97	0.74
Mathematics	5	Operations and Algebraic Thinking	31646	6	6	6.00	510.11	127.69	73.34	0.62
Mathematics	5	Geometry	31646	6	6	6.00	507.45	122.03	72.73	0.59
Mathematics	5	Measurement and Data	31646	6	6	6.00	529.88	129.50	73.51	0.63
Mathematics	6	The Number System	29745	7	7	7.00	580.20	121.45	66.35	0.66
Mathematics	6	Ratios and Proportional Relationships	29745	6	6	6.00	557.98	129.43	71.39	0.65
Mathematics	6	Expressions and Equations	29745	10	10	10.00	571.82	99.57	51.32	0.71
Mathematics	6	Geometry	29745	6	6	6.00	559.60	126.25	71.92	0.63
Mathematics	6	Statistics and Probability	29745	7	7	7.00	526.70	112.96	66.72	0.60

Content Area	Grade/ Level	Reporting Category	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Mathematics	7	The Number System	26581	6	6	6.00	575.36	137.56	75.73	0.64
Mathematics	7	Ratios and Proportional Relationships	26581	8	8	8.00	580.05	114.71	59.76	0.69
Mathematics	7	Expressions and Equations	26581	9	9	9.00	610.01	109.16	56.42	0.70
Mathematics	7	Geometry	26581	7	7	7.00	570.29	108.07	65.75	0.57
Mathematics	7	Statistics and Probability	26581	6	6	6.00	563.34	124.95	74.59	0.58
Mathematics	8	The Number System	24033	6	6	6.00	634.37	160.07	82.67	0.67
Mathematics	8	Expressions and Equations	24033	10	10	10.00	614.23	107.79	52.09	0.74
Mathematics	8	Functions	24033	8	8	8.00	612.40	103.22	60.21	0.62
Mathematics	8	Geometry	24033	6	6	6.00	596.03	127.72	74.85	0.60
Mathematics	8	Statistics and Probability	24033	6	6	6.00	590.69	128.41	73.46	0.62
Algebra I	HS	Operations and Linear Equations & Inequalities	26927	18	18	18.00	618.89	86.84	36.10	0.82
Algebra I	HS	Operations with Real Numbers and Expressions	26927	6	6	6.00	618.43	126.75	73.17	0.61
Algebra I	HS	Linear Equations	26927	6	6	6.00	608.24	123.94	71.97	0.61
Algebra I	HS	Linear Inequalities	26927	6	6	6.00	616.27	121.88	71.89	0.60
Algebra I	HS	Linear Functions & Data Organizations	26927	18	18	18.00	618.80	94.46	36.34	0.85
Algebra I	HS	Functions	26927	6	6	6.00	617.57	128.17	72.52	0.63
Algebra I	HS	Coordinate Geometry	26927	6	6	6.00	620.49	128.99	71.72	0.65
Algebra I	HS	Data Analysis	26927	6	6	6.00	607.54	126.38	73.09	0.61
ELA	3	Literature Text	26345	14	16	15.79	431.07	111.04	53.64	0.76
ELA	3	Informational Text	26345	14	16	15.23	429.86	106.83	55.28	0.73
ELA	4	Literature Text	25689	14	16	15.06	490.20	125.51	55.99	0.80
ELA	4	Informational Text	25689	14	16	15.68	492.87	117.74	55.54	0.77
ELA	5	Literature Text	28077	14	16	15.78	537.25	119.27	53.83	0.79
ELA	5	Informational Text	28077	14	16	15.39	535.18	120.11	54.19	0.79

Content Area	Grade/ Level	Reporting Category	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
ELA	6	Literature Text	26547	14	16	14.95	550.24	125.43	56.49	0.79
ELA	6	Informational Text	26547	14	16	15.22	546.26	120.48	55.11	0.79
ELA	7	Literature Text	24015	14	16	14.78	560.49	121.97	55.15	0.79
ELA	7	Informational Text	24015	14	16	14.90	569.07	121.96	55.17	0.79
ELA	8	Literature Text	23941	14	16	14.85	579.39	123.90	54.85	0.80
ELA	8	Informational Text	23941	14	16	15.43	586.33	122.13	54.61	0.79
Literature	HS	Fiction	23820	15	19	17.36	625.10	117.69	51.25	0.81
Literature	HS	Reading for Meaning—Fiction	23820	6	8	6.98	618.27	151.63	91.97	0.59
Literature	HS	Analyzing and Interpreting Literature—Fiction	23820	9	11	10.37	629.72	123.24	66.89	0.69
Literature	HS	Nonfiction	23820	15	19	17.50	616.77	120.48	51.33	0.82
Literature	HS	Reading for Meaning—Nonfiction	23820	6	8	6.97	615.51	147.22	87.20	0.62
Literature	HS	Analyzing and Interpreting Literature— Nonfiction	23820	9	11	10.53	617.61	129.30	68.59	0.71

APPENDIX E: RELIABILITY BY STUDENT GROUP

As discussed in Chapter Sixteen, subscore reliability was calculated by student group, including gender (female, male), race/ethnicity, economically disadvantaged, English Learners (EL), and students with individualized education plans (IEP).

Table E-1. Mathematics Grade 3

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	15631	36	36	36.00	454.85	89.94	25.27	0.92
Male	15990	36	36	36.00	472.18	99.48	25.67	0.93
AIAN	66	36	36	36.00	466.27	86.19	25.29	0.91
Asian	1079	36	36	36.00	517.55	101.76	26.49	0.93
Black	3488	36	36	36.00	410.74	86.17	25.65	0.91
Hispanic	3776	36	36	36.00	422.22	85.81	25.44	0.91
Multi-Racial	1945	36	36	36.00	446.90	92.37	25.42	0.92
NHPI	248	36	36	36.00	421.68	80.58	25.28	0.90
White	21019	36	36	36.00	479.09	92.06	25.41	0.92
Ec. Disadvantaged	10619	36	36	36.00	433.43	87.93	25.36	0.92
EL	1684	36	36	36.00	424.57	87.94	25.46	0.92
IEP	4205	36	36	36.00	409.68	92.99	25.90	0.92

Table E-2. Mathematics Grade 4

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	14130	36	36	36.00	490.03	83.62	25.16	0.91
Male	14913	36	36	36.00	506.51	93.86	25.52	0.93
AIAN	63	36	36	36.00	482.86	89.62	25.32	0.92
Asian	773	36	36	36.00	541.41	99.51	26.04	0.93
Black	3044	36	36	36.00	451.80	74.62	25.35	0.88
Hispanic	3085	36	36	36.00	466.57	83.02	25.37	0.91
Multi-Racial	1695	36	36	36.00	486.14	89.70	25.43	0.92
NHPI	188	36	36	36.00	451.56	73.24	25.47	0.88
White	20195	36	36	36.00	510.29	88.01	25.31	0.92
Ec. Disadvantaged	9435	36	36	36.00	471.32	82.61	25.30	0.91
EL	1382	36	36	36.00	457.53	79.95	25.40	0.90
IEP	4290	36	36	36.00	443.99	83.94	25.80	0.90

Table E–3. Mathematics Grade 5

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	15311	36	36	36.00	524.81	82.98	25.28	0.91
Male	16335	36	36	36.00	539.43	93.80	25.77	0.92
AIAN	43	36	36	36.00	547.58	109.06	26.98	0.93
Asian	1356	36	36	36.00	593.33	104.96	27.61	0.92
Black	2951	36	36	36.00	489.49	73.97	25.43	0.88
Hispanic	3095	36	36	36.00	493.44	76.65	25.46	0.89
Multi-Racial	1814	36	36	36.00	518.28	89.12	25.62	0.92
NHPI	198	36	36	36.00	501.91	67.51	24.93	0.86
White	22189	36	36	36.00	541.15	87.10	25.43	0.91
Ec. Disadvantaged	9841	36	36	36.00	503.00	78.88	25.36	0.90
EL	1152	36	36	36.00	500.31	83.08	25.58	0.90
IEP	4461	36	36	36.00	475.82	80.82	26.02	0.89

Table E–4. Mathematics Grade 6

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	14548	36	36	36.00	559.00	78.65	24.92	0.90
Male	15197	36	36	36.00	566.66	85.99	25.08	0.91
AIAN	410	36	36	36.00	563.40	76.30	24.82	0.89
Asian	947	36	36	36.00	601.21	93.07	25.41	0.92
Black	2945	36	36	36.00	516.46	73.64	25.47	0.88
Hispanic	3018	36	36	36.00	527.76	77.65	25.36	0.89
Multi-Racial	1539	36	36	36.00	552.76	78.95	25.01	0.90
NHPI	199	36	36	36.00	535.62	75.51	25.18	0.89
White	20687	36	36	36.00	573.91	80.19	24.87	0.90
Ec. Disadvantaged	8708	36	36	36.00	539.87	76.74	25.11	0.89
EL	953	36	36	36.00	517.03	69.90	25.42	0.87
IEP	3780	36	36	36.00	505.22	75.82	25.83	0.88

Table E–5. Mathematics Grade 7

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	12986	36	36	36.00	583.98	77.35	25.15	0.89
Male	13595	36	36	36.00	587.12	85.93	25.47	0.91
AIAN	33	36	36	36.00	586.82	85.92	25.30	0.91
Asian	901	36	36	36.00	620.51	87.41	25.49	0.91
Black	2473	36	36	36.00	543.40	73.78	25.86	0.88
Hispanic	2463	36	36	36.00	555.92	78.51	25.68	0.89
Multi-Racial	1343	36	36	36.00	566.85	79.96	25.56	0.90
NHPI	159	36	36	36.00	553.03	79.38	25.82	0.89
White	19209	36	36	36.00	594.76	79.97	25.16	0.90
Ec. Disadvantaged	7776	36	36	36.00	561.77	76.40	25.48	0.89
EL	667	36	36	36.00	529.25	67.19	26.06	0.85
IEP	3107	36	36	36.00	526.53	69.30	26.26	0.85

Table E–6. Mathematics Grade 8

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	11640	36	36	36.00	612.61	81.17	25.33	0.90
Male	12393	36	36	36.00	610.48	87.69	25.67	0.91
AIAN	63	36	36	36.00	594.44	87.49	26.03	0.91
Asian	879	36	36	36.00	661.90	91.58	26.54	0.91
Black	2150	36	36	36.00	576.39	75.58	25.58	0.88
Hispanic	2222	36	36	36.00	582.03	74.02	25.43	0.88
Multi-Racial	1199	36	36	36.00	601.82	85.61	25.66	0.91
NHPI	205	36	36	36.00	586.51	84.42	25.85	0.90
White	17315	36	36	36.00	618.13	83.81	25.44	0.91
Ec. Disadvantaged	6676	36	36	36.00	589.37	78.16	25.47	0.89
EL	597	36	36	36.00	569.15	67.55	25.50	0.86
IEP	3093	36	36	36.00	552.11	70.65	26.17	0.86

Table E–7. Algebra I

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	12634	36	36	36.00	622.62	81.08	24.91	0.90
Male	14293	36	36	36.00	616.63	86.99	25.20	0.92
AIAN	83	36	36	36.00	592.36	86.47	25.63	0.91
Asian	1372	36	36	36.00	668.90	94.78	25.38	0.93
Black	3086	36	36	36.00	583.27	76.84	25.44	0.89
Hispanic	3092	36	36	36.00	592.46	80.70	25.35	0.90
Multi-Racial	1298	36	36	36.00	608.65	83.41	25.17	0.91
NHPI	25	-	-	-	-	-	-	-
White	17971	36	36	36.00	627.39	81.72	24.92	0.91
Ec. Disadvantaged	7643	36	36	36.00	594.92	79.87	25.30	0.90
EL	1260	36	36	36.00	576.73	76.61	25.60	0.89
IEP	3545	36	36	36.00	569.31	73.71	25.76	0.88

Table E–8. ELA Grade 3

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	12913	28	32	31.01	435.71	98.82	37.39	0.86
Male	13432	28	32	31.03	427.33	98.47	37.54	0.85
AIAN	54	28	32	30.78	428.24	80.88	37.06	0.79
Asian	916	29	32	30.99	465.63	103.15	37.28	0.87
Black	3008	28	32	31.10	384.65	80.08	38.17	0.77
Hispanic	2699	29	32	31.08	392.40	85.07	38.04	0.80
Multi-Racial	1613	29	32	31.07	419.98	96.82	37.62	0.85
NHPI	229	28	32	31.00	424.30	91.43	37.39	0.83
White	17826	28	32	31.00	444.63	99.29	37.26	0.86
Ec. Disadvantaged	8801	28	32	31.06	403.57	88.72	37.78	0.82
EL	1082	28	32	31.05	388.94	82.55	38.11	0.78
IEP	3460	29	32	31.12	380.53	84.60	38.50	0.79

Table E–9. ELA Grade 4

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	12512	28	32	30.73	498.28	110.30	38.10	0.88
Male	13177	28	32	30.74	487.27	111.37	38.37	0.88
AIAN	59	30	32	30.85	464.90	107.30	38.76	0.87
Asian	642	28	32	30.77	534.90	115.67	38.16	0.89
Black	2658	28	32	30.75	446.27	95.70	38.88	0.83
Hispanic	2521	28	32	30.76	460.02	103.52	38.71	0.86
Multi-Racial	1541	28	32	30.76	482.46	116.04	38.79	0.89
NHPI	188	28	32	30.74	477.87	102.24	38.01	0.86
White	18080	28	32	30.72	503.61	110.59	38.03	0.88
Ec. Disadvantaged	8483	28	32	30.72	459.37	101.61	38.69	0.85
EL	1136	29	32	30.76	437.86	94.14	39.23	0.82
IEP	3847	28	32	30.75	423.87	93.00	39.89	0.81

Table E–10. ELA Grade 5

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	13573	28	32	31.18	543.85	108.08	36.96	0.88
Male	14504	28	32	31.16	530.50	110.33	37.21	0.88
AIAN	35	29	32	31.20	530.00	123.64	38.00	0.90
Asian	1143	29	32	31.22	595.53	113.09	37.63	0.89
Black	2744	28	32	31.14	487.31	97.01	37.58	0.85
Hispanic	2568	28	32	31.16	492.66	100.31	37.51	0.86
Multi-Racial	1636	28	32	31.16	519.56	107.01	37.23	0.88
NHPI	194	29	32	31.21	509.13	96.53	36.82	0.85
White	19757	28	32	31.17	547.94	107.96	36.93	0.88
Ec. Disadvantaged	8713	28	32	31.15	501.86	100.59	37.28	0.86
EL	956	29	32	31.16	486.93	97.28	37.58	0.85
IEP	3817	28	32	31.14	466.58	97.71	38.39	0.84

Table E–11. ELA Grade 6

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	13024	28	32	30.16	555.90	111.30	38.17	0.88
Male	13523	28	32	30.18	542.06	112.72	38.40	0.88
AIAN	574	28	32	30.18	553.57	103.38	37.71	0.87
Asian	882	28	32	30.04	599.31	108.51	38.17	0.87
Black	2648	28	32	30.28	497.11	102.75	39.03	0.85
Hispanic	2670	28	32	30.28	509.18	106.67	38.79	0.87
Multi-Racial	1441	28	32	30.20	537.44	112.02	38.44	0.88
NHPI	191	29	32	30.25	538.18	113.38	38.59	0.88
White	18141	28	32	30.14	560.66	110.81	38.11	0.88
Ec. Disadvantaged	8368	28	32	30.23	522.61	105.81	38.37	0.87
EL	763	28	32	30.27	485.99	97.33	39.24	0.84
IEP	3360	28	32	30.35	471.92	97.03	39.88	0.83

Table E–12. ELA Grade 7

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	11703	28	32	29.67	576.75	110.25	37.74	0.88
Male	12312	28	32	29.68	554.66	111.57	37.95	0.88
AIAN	22	-	-	-	-	-	-	-
Asian	1100	28	32	29.69	623.07	114.15	38.49	0.88
Black	2370	28	32	29.68	522.85	99.58	37.97	0.85
Hispanic	2164	28	32	29.67	525.78	103.33	38.09	0.86
Multi-Racial	1210	28	32	29.68	545.84	109.19	37.94	0.88
NHPI	159	28	32	29.65	551.85	113.81	38.24	0.89
White	16990	28	32	29.68	574.24	110.52	37.75	0.88
Ec. Disadvantaged	7028	28	32	29.68	533.14	102.47	37.83	0.86
EL	571	28	31	29.72	492.78	86.43	38.39	0.80
IEP	2660	28	32	29.73	496.15	97.03	38.79	0.84

Table E–13. ELA Grade 8

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	11687	28	32	30.28	595.93	111.07	37.40	0.89
Male	12254	28	32	30.29	571.13	112.64	37.67	0.89
AIAN	53	29	32	30.19	558.02	109.04	37.91	0.88
Asian	899	28	32	30.28	636.17	111.47	37.77	0.88
Black	2065	28	32	30.30	545.36	102.12	37.63	0.86
Hispanic	2174	28	32	30.28	549.44	103.67	37.63	0.87
Multi-Racial	1191	28	32	30.31	574.60	114.94	37.75	0.89
NHPI	207	28	32	30.16	585.44	117.36	37.88	0.89
White	17352	28	32	30.28	589.88	112.39	37.48	0.89
Ec. Disadvantaged	6806	28	32	30.27	553.83	105.77	37.62	0.87
EL	583	28	32	30.26	521.08	92.84	38.05	0.83
IEP	3007	28	32	30.33	508.91	97.13	38.64	0.84

Table E–14. Literature

Student Group	N Students	N Items Min	N Items Max	N Items Mean	Mean Scale Score	SD Scale Score	Mean SEM	Reliability
Female	11332	31	38	34.88	636.58	107.20	35.26	0.89
Male	12488	31	38	34.83	606.92	109.97	35.40	0.90
AIAN	66	33	38	34.80	626.83	101.92	35.03	0.88
Asian	1165	31	38	34.87	675.70	107.26	36.02	0.89
Black	2339	31	38	34.78	571.27	101.84	35.51	0.88
Hispanic	2186	31	38	34.87	585.02	107.75	35.51	0.89
Multi-Racial	1072	31	38	34.78	607.06	107.21	35.26	0.89
NHPI	13	-	-	-	-	-	-	-
White	16979	31	38	34.87	629.61	107.64	35.24	0.89
Ec. Disadvantaged	6721	31	38	34.79	588.14	105.11	35.35	0.89
EL	696	32	38	34.82	561.48	99.76	35.63	0.87
IEP	2824	31	38	34.69	548.62	95.31	35.83	0.86

APPENDIX F: DECISION CONSISTENCY – RETEST CLASSIFICATION

In Chapter Sixteen (Reliability), decision consistency is reported with two overall values: agreement rate and kappa. Yet, retest classification probability varies at different points along the scale. Thus, this appendix provides a detailed examination of retest probability with respect to each performance level. The 3 × 3 retest classification probability tables are presented for each Firefly test. As discussed in Chapter Sixteen the values in the diagonals of the contingency tables were used to calculate the overall decision consistency/agreement indices.

Table F–1. Mathematics Grade 3

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.942	0.058	0.000
Near Target (Test)	0.122	0.794	0.084
Prepared (Test)	0.000	0.165	0.835

Table F–2. Mathematics Grade 4

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.946	0.054	0.000
Near Target (Test)	0.139	0.786	0.074
Prepared (Test)	0.000	0.169	0.831

Table F–3. Mathematics Grade 5

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.947	0.053	0.000
Near Target (Test)	0.132	0.789	0.079
Prepared (Test)	0.000	0.165	0.835

Table F–4. Mathematics Grade 6

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.938	0.062	0.000
Near Target (Test)	0.155	0.783	0.062
Prepared (Test)	0.000	0.179	0.821

Table F–5. Mathematics Grade 7

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.949	0.051	0.000
Near Target (Test)	0.157	0.777	0.066
Prepared (Test)	0.000	0.203	0.797

Table F–6. Mathematics Grade 8

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.946	0.054	0.000
Near Target (Test)	0.150	0.781	0.069
Prepared (Test)	0.000	0.176	0.824

Table F–7. Algebra I

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.946	0.054	0.000
Near Target (Test)	0.148	0.782	0.070
Prepared (Test)	0.000	0.189	0.811

Table F–8. ELA Grade 3

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.921	0.079	0.000
Near Target (Test)	0.130	0.813	0.057
Prepared (Test)	0.000	0.225	0.775

Table F–9. ELA Grade 4

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.929	0.071	0.000
Near Target (Test)	0.117	0.811	0.072
Prepared (Test)	0.000	0.216	0.784

Table F–10. ELA Grade 5

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.929	0.071	0.000
Near Target (Test)	0.122	0.816	0.062
Prepared (Test)	0.000	0.235	0.765

Table F–11. ELA Grade 6

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.937	0.063	0.000
Near Target (Test)	0.130	0.811	0.059
Prepared (Test)	0.000	0.267	0.733

Table F–12. ELA Grade 7

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.940	0.060	0.000
Near Target (Test)	0.138	0.805	0.057
Prepared (Test)	0.000	0.257	0.743

Table F–13. ELA Grade 8

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.937	0.063	0.000
Near Target (Test)	0.151	0.798	0.052
Prepared (Test)	0.000	0.252	0.748

Table F–14. Literature

Performance Level	Support Needed (Retest)	Near Target (Retest)	Prepared (Retest)
Support Needed (Test)	0.930	0.070	0.000
Near Target (Test)	0.126	0.825	0.049
Prepared (Test)	0.000	0.285	0.715

REFERENCES

- Achieve, Inc. (2005). *Measuring up 2005: A report on assessment anchors and tests in reading and mathematics for Pennsylvania*. Washington, DC: Achieve, Inc.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (p. 267–281). Budapest, Hungary: Akademiai Kiado.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, NCME]. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Allman, C. (2004). *Test access: Making tests accessible for students with visual impairments – A guide for test publishers, test developers, and state assessment personnel* (2nd ed.). Louisville, KY: American Printing House for the Blind. Available from <http://www.apf.org>.
- Asparouhov, T. & Muthén, B. (2010). Simple Second Order Chi-Square Correction (Technical Note). Retrieved from: http://www.statmodel.com/download/WLSMV_new_chi21.pdf
- Bond, T. & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. (2nd Edition). Psychology Press.
- Brennan, R. L. (1998). Misconceptions at the intersection of measurement theory and practice. *Educational Measurement: Issues and Practice*, 17(1), 5–9.
- Brennan, R. (2004). BB-Class (Version 1.0). [Computer Software] Iowa City, IA: University of Iowa, Center for Advanced Studies in Measurement & Assessment. CASMA: education.uiowa.edu/casma.
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296–322.
- Buja, A. & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, 27, 509–540.
- Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265–289.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cook, L. L., & Eignor, D. R. (1991). NCME instructional module: IRT equating methods. *Educational Measurement: Issues and Practice*, 17(1), 5–9.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education. *Educational Measurement: Issues and Practice*, 10, 37–45.
- Cronbach, L., & Shavelson R. L. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64(3), 391–418.
- D’Agostino, R. B. (1998). Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 17, 2265–2281.
- Dorans, N. J., & Holland, P. W. (2000). Population invariance and equatability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, 37, 281–306.

- Dorans, N. J., Holland, P. W., Thayer, D. T., & Tateneni, K. (2003). Invariance of score linking across gender groups for three advanced placement program exams. In N. J. Dorans (Ed.), *Population invariance of score linking: Theory and applications to advanced placement program examinations* (pp. 79–118), Research Report 03-27. Princeton, NJ: Educational Testing Service.
- Dorans, N. J., & Feigenbaum, M. D. (1994). Equating issues engendered by changes to the SAT and PSAT/NMSQT®. In I. M. Lawrence, N. J. Dorans, M. D. Feigenbaum, N. J. Feryok, & N. K. Wright, Technical issues related to the introduction of the new SAT and PSAT/NMSQT (RM-94-10). Princeton, NJ: Educational Testing Service.
- Dorans, N., Schmitt, A., & Bleistein, C. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, 29, 309–319.
- Engelhard, G. Jr. (2009). Using item response theory and model-data fit to conceptualize differential item and person functioning for students with disabilities. *Educational and Psychological Measurement*, 69, 585–602.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational Measurement*, (3rd ed., pp. 105–146). New York, NY: ACE/Macmillan.
- Frisbie, D. A. (2005). Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice*, 24(3), 21–28.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley and Sons.
- Haertel, E. H. (2006). Reliability. In Brennan, R. L. (Ed.), *Educational Measurement* (4th ed., pp. 65–110). Westport, CT: Praeger.
- Hambleton, R., Swaminathan, H., and Rogers, J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hambleton, R. & Novick, M. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, 10, 159–170.
- Hambleton, R. & Rogers, H. (1986). Evaluation of the plot method for identifying potentially biased test items. In S. H. Irvine, S. Newstead, & P. Dann (Eds.), *Computer-based human assessment*, Boston, MA: Kluwer Academic Publishers.
- Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score theory models. *Journal of Educational Measurement*, 27(4), 345–359.
- Harvill, L. M. (1991). Standard error of measurement. *Educational Measurement: Issues and Practices*, 10(2), 33–41.
- Ho, D.E., Imai, K., King, G., Stuart, E.A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, 42(8), 1–28. <https://www.jstatsoft.org/v42/i08/>.
- Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 32, 179–185.
- Hu, L. & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 60(1), pp. 1–55.
- Huynh, H. (1976). On the reliability of decisions in domain referenced testing. *Journal of Educational Measurement*, 13, 253–264.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141–151.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking* (3rd ed). New York, NY: Springer.

- Koger, M. E., Thacker, A. A., & Dickinson, E. R. (2004). *Relationships among the Pennsylvania System of School Assessment (PSSA) scores, SAT scores, and self-reported high school grades for the classes of 2002 and 2003* (HumRRO Report FR-04-26). Louisville, KY: Human Resources Research Organization.
- Karkee, T., Kim, D., & Fatica K. (April, 2010). Comparability Study of Online and Paper-and-Pencil Tests Using Modified Internally and Externally Matched Criteria. Paper presented at the annual meeting of the American Educational Research Association (AERA). Denver, CO.
- Lane, S. (1999). *Validity evidence for assessments*. Paper presented at the 1999 Edward F. Reidy Interactive Lecture Series, Providence, RI.
- Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement: Issues and Practice*, 21(1), 23–30.
- Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). *Standard setting: A bookmark approach*. Symposium presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix, AZ.
- Linacre, J. M. (2019). Winsteps® Rasch measurement [computer program]. Beaverton, OR.
- Linacre, J. M. (2009). *A user's guide to WINSTEPS MINISTEP Rasch-model computer programs*. Chicago, IL: Winsteps.
- Livingston, S. & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement* 32, 179–197.
- Luecht, R. M. (2016). Applications of item response theory: Item and test information functions for designing and building mastery tests. In S. Lane, M. R. Raymond, T. M. Haladyna (Eds.), *Handbook of test development* (p. 485–506). New York, NY: Routledge.
- MacCallum, R.C. (2001). Working with imperfect models. *Multivariate Behavioral Research*, 38(1), pp. 113–139.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.
- Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, 9(3), 200–215.
- McDonald, R. P. (1979). The structural analysis of multivariate data: A sketch of general theory. *Multivariate Behavioral Research*, 14, 21–38.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 104–142). New York: Academic Press.
- Messick, S. (1989). Validity. In R. L. (Ed.), *Educational Measurement* (3rd ed., pp. 3–104). New York: American Council on Education.
- Moses, T., Deng, W., & Zhang, Y. L. (2010). The Use of Two Anchors in Nonequivalent Groups With Anchor Test (NEAT) Equating (ETS Research Report No. RR-10-23). Princeton, NJ: ETS.
- Muthén, L.K. & Muthén, B.O. (1998–2012). Mplus. Seventh Edition. Los Angeles, CA.
- National Center for Education Statistics (2020a). School District Characteristic 2018–2019 [Data file]. Retrieved from [https://data-nces.opendata.arcgis.com/datasets/No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 \(2002\)](https://data-nces.opendata.arcgis.com/datasets/No+Child+Left+Behind+Act+of+2001,+Pub.+L.+No.+107-110,+115+Stat.+1425+(2002)).
- National Center for Education Statistics (2020b). The Common Education Data Standards (CEDS). Retrieved from <http://ceds.ed.gov/>

- Pennsylvania State Board of Education. (1999, January). *Chapter 4. Academic standards and assessment*. Harrisburg, PA: Pennsylvania State Board of Education. Retrieved November 8, 2004, from <http://www.education.state.pa.us>. Also available from <http://www.pacode.com/secure/data/022/Chapter4/s4.51.html>.
- Pennsylvania Department of Education. (2021). *Assessment anchors and eligible content*. Retrieved February 9, 2021, from www.pa.gov/agencies/education/.
- Pennsylvania Department of Education. (2025). *2024 Mathematics Item and Scoring Sampler*. Retrieved March 20, 2025, from www.pa.gov/agencies/education/.
- Pennsylvania Department of Education. (2025). *2024 English Language Arts Item and Scoring Sampler*. March 20, 2025, from www.pa.gov/agencies/education/.
- Pennsylvania Department of Education. (2025). *2024 Science Item and Scoring Sampler*. Retrieved March 20, 2025, from www.pa.gov/agencies/education/.
- Pennsylvania Department of Education. (2025). *2025 Accommodations Guidelines* (PDE, revised 9/2022). Retrieved April 11, 2025, from www.pa.gov/agencies/education/.
- Pennsylvania Department of Education. (2025). *2025 Accommodations Guidelines for ELs* (PDE, revised 9/2022). Retrieved April 11, 2025, from www.pa.gov/agencies/education/.
- Pennsylvania Department of Education. (2025). *2025 Pennsylvania System of School Assessment: Handbook for Assessment Coordinators*. Retrieved April 11, 2025, from www.pa.gov/agencies/education/.
- Qualls, A. L. (1995). Estimating the reliability of a test containing multiple item formats. *Applied Measurement in Education*, 8(2), 111–120.
- R (Version 3.6) [Computer software]. Vienna, Austria: R Core Team.
- Raîche, G. (2005). Critical eigenvalue sizes in standardized residual principal components analysis. *Rasch Measurement Transactions*, 19(1), 1012.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207–230.
- Robertson, T., Wright, F. T. & Dykstra, R. L. (1988) *Order Restricted Statistical Inference*. New York: Wiley.
- Rosenbaum, P. R. (1995). *Observational studies*. New York: Springer-Verlag.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.
- Rubin, D. B. (2006). *Matched sampling for causal effects*. New York: Cambridge University Press.
- SAS (Version 7) [Computer software]. Cary, NC: SAS Institute, Inc.
- Satorra A. & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, C.C. Clogg (Eds.), *Latent Variables Analysis: Applications for Developmental Research* (pp. 399–419). Thousand Oaks, CA: Sage.
- Sinclair, A. L., & Thacker, A. A. (2005). *Relationships among Pennsylvania System of School Assessment (PSSA) scores, university proficiency exam scores, and college course grades in English and math* (HumRRO Report FR-05-55). Louisville, KY: Human Resources Research Organization.

- Sireci, S. G., and Wells, C. S. (2016). *Evaluating Test Accommodations on the Pennsylvania System of School Assessment Exams* (Research Report). East Hampton, Sireci Psychometric Services.
- Smith, R. (2004). Fit analysis in latent trait measurement models. In Smith, E. V., & Smith, R. M. (Ed.), *Introduction to Rasch Measurement: Theory, models and applications*. Maple Grove, MN: JAM Press.
- Smith, R. & Miao, C. (1994). Assessing unidimensionality for Rasch measurement. Chapter 18 in M. Wilson (Ed.) *Objective Measurement: Theory into Practice*. Vol. 2. Norwood NJ: Ablex.
- Spearman C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72–101.
- Spearman C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271–295.
- Staffa, S. & Zurakowski, D. (2018). Five steps to successfully implement and evaluate propensity score matching in clinical research studies. *Anesthesia & Analgesia*, 127(1).
- Stearns, M., & Smith R. M. (2007). *Estimation of classification consistency indices for complex assessments: Model based approaches*. Paper presented at the 2007 Annual Convention of the American Educational Research Association, Chicago, IL.
- Thacker, A. A., & Dickinson, E. R. (2004). *Item content and difficulty mapping by form and item type for the 2001–2003 Pennsylvania System of School Assessment (PSSA)*. Alexandria, VA: Human Resources Research Organization.
- Thacker, A. A., Dickinson, E. R., & Koger, M. E. (2004). *Relationships among the Pennsylvania System of School Assessment (PSSA) and other commonly administered assessments* (HumRRO Report FR-04-33). Louisville, KY: Human Resources Research Organization.
- Thompson, S., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Traub, R. E. (1994). *Reliability for the social sciences: Theory and application*. Thousand Oaks: Sage.
- Von Davier, A. A., & Wilson, C. (2008). Investigating the population sensitivity assumption of item response theory true score equating across two subgroups of examinees and two test formats. *Applied Psychological Measurement*, 32, 11–26.
- Wang, S., McCall, M., Jiao, H., Harris, G. (April, 2012). Constructing Validity and Measurement Invariance of Computer Adaptive Testing: Application to Measures of Academic Progress (MAP) using Confirmatory Factor Analysis. Paper presented at the annual meeting of the American Educational Research Association (AERA). Vancouver, British Columbia, Canada.
- Way, W. D., Lin, C., & Kong, J. (March, 2008). Maintaining Score Equivalence as Tests Transition Online: Issues, Approaches and Trends. Paper presented at the annual meeting of the National Council on Measurement in Education (NCME). New York, NY.
- Webb, N. L. (1997). *Criteria for alignment of expectations and tests in mathematics and science education* (NISE Research Monograph No. 6). Madison: University of Wisconsin–Madison, National Institute for Science Education. Washington, DC: Council of Chief State School Officers.
- Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* (NISE Research Monograph No. 18). Madison, WI: University of Wisconsin–Madison, National Institute for Science Education.
- Webb, N. L. (2002). *Alignment study in language arts, mathematics, science, and social studies of state standards and tests for four states: State collaborative on test and state standards (SCASS)*. Madison, WI: University of Wisconsin–Madison, Wisconsin Center for Education Research.

- Wright, B.D., & Linacre, M. (1994). Reasonable Mean-Square Fit Statistics. *Rasch Measurement Transactions*, 8(3), 370. Retrieved from: <https://www.rasch.org/rmt/rmt83b.htm>
- Wright, B., & Masters, G. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.
- Yen, W. M. (1993). Scaling performance assessments: strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3), 187–213.