2022-2023 LEAP 2025 Grades 3-8
Operational Technical Report
English Language Arts and Mathematics

Submitted to the
Louisiana Department of Education

December 2023

DATA RECOGNITION
DRC
CORPORATION

This online-only document was published at a cost of $33,533. This document was published for the Louisiana Department of Education, P.O. Box 94064, Baton Rouge, LA 70804-9064, by Data Recognition Corporation, 13490 Bass Lake Road, Maple Grove, MN 55311. This material was printed in accordance with the standards for printing by State Agencies established pursuant to R.S. 43:31.

Table of Contents

Executive Summary

This report is a technical summary of the 2023 administration of the Louisiana Educational Assessment Program (LEAP 2025) in English language arts (ELA) and mathematics for grades 3 through 8. The LEAP 2025 summative assessments in ELA and mathematics are administered in grades 3 through 8 and high school. These tests are designed to measure students' readiness for the next grade or course of study and proficiency in ELA and mathematics. The ELA and mathematics test forms were developed by Data Recognition Corporation (DRC) test development staff using the New Meridian item bank as well as items from the Louisiana Department of Education's own item bank. Items taken from these banks were on pre-established item response theory (IRT) scales. This section provides a summary of the 2023 operational technical report.

E.1 Overview of This Report

This technical report documents the major activities of the testing cycle and provides details that confirm that the processes and procedures applied in the LEAP 2025 assessments adhered to appropriate professional standards and practices of educational assessment. Ultimately, this report serves to document evidence that valid inferences about Louisiana student performance in ELA and mathematics can be derived from the LEAP 2025 assessments. An overview of major activities documented within this report is provided below.

The Louisiana Department of Education and Data Recognition Corporation implemented rigorous quality control procedures throughout the test development, administration, scoring, analyses, and reporting processes for the LEAP 2025 assessments. The system and procedures for monitoring, maintaining, and improving the quality of state assessment system is described in each section of the technical report as an integral part of the activities.

The Uses of Test Scores (Chapter 2)

Chapter 2 of the technical report discusses the concept of validity evidence. This technical report is composed of evidence that supports the intended uses of the LEAP 2025 test scores, and Chapter 2 discusses some of those uses.

Test Content Development (Chapter 3)

Chapter 3 of the technical report provides a summary of the test development activities that occurred in order to create the spring 2023 operational test forms. This includes quality control of Item Development, the Item Bank, and the Item Review process.

Test Administration (Chapter 4)

Chapter 4 of the technical report describes the processes implemented and the information disseminated to help ensure standardized test administration procedures and, thus, uniform test administration conditions for students. This includes quality control processes including, but not limited to, LDOE site visits, review rounds of materials, Security Checklists, and Test Security Measures (Data Forensics Analysis, Response-Change Analysis, Web Monitoring, and Plagiarism Detection).

Constructed-Response and Technology-Enhanced Scoring (Chapter 5)

Chapter 5 of the technical report describes the processes used to score constructed-response and technology-enhanced items. The quality control measures in this section include the recruitment and interview process, security protocols, and training process, including material development and qualifying procedures. This chapter discusses how scorers are trained and the measures used to ensure consistency

among scorers. Finally, this chapter presents the results of the inter-rater reliability studies. Inter-rater reliability studies along with validity and reader monitoring are additional quality control processes of scoring.

Operational Data Analyses (Chapter 6)

Chapter 6 of the technical report includes a detailed description of the operational data analyses and quality control of the 2023 LEAP 2025 assessments, which include the following major parts: the classical item analysis; calibration, scaling, and linking using IRT models; and student scoring.

Test Results (Chapter 7)

Chapter 7 of the technical report contains information on the results of the spring 2023 LEAP 2025 assessments. Detailed summary statistics based on scale scores and information about achievement levels are also provided. Finally, this chapter presents information on the score reports sent to school systems.

Performance-Level Setting (Chapter 8)

Chapter 8 of the technical report briefly discusses performance-level setting. It provides a brief overview of the quality-controlled procedures for performance-level setting and derivation of the cut scores used to classify students into achievement levels for ELA and mathematics.

Evidence of Construct-Related Reliability (Chapter 9)

Chapter 9 of the technical report provides evidence of the system and procedures for monitoring, maintaining, and improving the quality, reliability and validity of the LEAP 2025 test scores. This chapter provides detailed evidence of the reliability of the tests and information on the classification consistency of the cut scores. It also provides evidence of construct validity for the LEAP 2025 test scores.

Fairness (Chapter 10)

Chapter 10 of the technical report discusses fairness and how the LEAP 2025 assessments are constructed, with quality control procedures in place, to be fair to all Louisiana students. This chapter summarizes the results of the differential item functioning (DIF) analysis. It also discusses the results of an impact analysis designed to determine whether large differences exist with the test results of different demographic groups in Louisiana. The results of the administration mode study are also summarized.

E.2 Administration

In the spring of 2023, Louisiana administered the LEAP 2025 summative assessments in ELA and mathematics to students in grades 3–8. A paper-based test (PBT) option was administered in grade 3, and the computer-based test (CBT) was administered in grades 3–8. The CBTs were administered from April 25 to May 26, 2023. The PBTs were administered from April 26 to May 2, 2023. Test administration is discussed in Chapter 4 of this report.

A total of 99 school systems and 32 charter schools administered the ELA and mathematics LEAP 2025 tests in grades 3–8. Table E.1 shows participation rates based on census data. For the purposes of this report, participation rate is defined as the percentage of students who earned a valid scale score given the total number of students who were expected to take the test. The “Accountable” column shows the total number of students who were expected to take the test by grade and content area. The “Percentage Reportable” column shows the percentage of students who received a scale score on the LEAP 2025 by grade and content area. Further analysis of participation rates is provided in Chapter 7 of this report. The results presented in Table E.1 and Chapter 7 are presented as evidence of reliability and validity of the scores from the LEAP 2025 assessments and should not be used for state accountability purposes.

Table E.1 Participation Rates: All Students Participating in 2023 LEAP 2025 Grades 3-8

Grade	Accountable in ELA	Percentage Reportable in ELA	Accountable in Mathematics	Percentage Reportable in Mathematics*
3	≥49,370	99.65%	≥49,660	99.64%
4	≥48,770	99.81%	≥48,770	99.81%
5	≥48,280	99.79%	≥48,290	99.79%
6	≥43,000	99.61%	≥48,450	99.63%
7	≥41,920	99.46%	≥48,990	99.51%
8	≥45,050	99.35%	≥49,950	99.45%

**Students in grade 8 who were enrolled in Algebra I had the option of taking the LEAP 2025 Algebra I assessment instead of the LEAP 2025 Grade 8 Mathematics test.*

E.3 Student Performance

Tables E.2 and E.3 present the percentage of students in 2023 who were classified in each of the achievement levels for ELA and mathematics.

Table E.2 Percentage of Students Classified in Achievement Levels Using 2023 Census Data: English Language Arts

Grade	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
3	21.5	16.3	19.5	36.4	6.2
4	12.3	19.2	24.3	33.9	10.3
5	10.0	20.3	29.3	37.4	3.0
6	10.2	22.4	28.5	31.9	7.1
7	14.8	18.4	24.7	27.3	14.8
8	12.4	16.7	23.5	36.8	10.7

Table E.3 Percentage of Students Classified in Achievement Levels Using 2023 Census Data: Mathematics

Grade	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
3	13.1	22.3	28.4	30.9	5.4
4	12.4	22.6	27.9	33.1	4.0
5	13.5	25.6	28.4	28.6	4.0
6	17.2	25.6	26.1	26.8	4.2
7	10.1	31.6	31.1	23.8	3.5
8	23.2	29.9	23.3	21.5	2.3

More information on student performance may be found in Chapter 7 of this report.

E.4 Validity and Test Scores

Most sections of this technical report are designed to provide validity evidence to support the intended uses of the LEAP 2025 test scores. Chapter 2 discusses the intended uses of the LEAP 2025 test scores. Chapter 3

discusses the test development process used to create the LEAP 2025 tests, which is important to the content-related validity of the LEAP 2025 test scores. Chapter 4 presents information on test administration. Chapter 5 discusses the scoring process and the results of the inter-rater reliability studies. Chapter 6 presents the test scaling and linking procedures, student scoring methodology, and the results of other operational data analyses. Chapter 7 reviews the results of the 2023 administration and gives an overview of the score reports that were electronically delivered to the school systems for distribution to schools and parents. Chapter 8 highlights the procedures for performance-level setting implemented by Partnership for Assessment of Readiness for College and Careers (PARCC), which were used because PARCC's standards and achievement levels were used for the LEAP 2025. Chapter 9 discusses reliability and construct-related validity. Chapter 10 gives an overview of the statistical processes used to evaluate bias to ensure fairness of the LEAP 2025 for all examinees.

Chapter 1: Introduction

The LEAP 2025 assessment system is designed to measure students' knowledge of ELA, mathematics, science, and social studies. This report provides a technical overview of the LEAP 2025 ELA and mathematics assessments administered in grades 3 through 8 in the spring of 2023 and presents evidence for the validity of the 2023 LEAP 2025 ELA and mathematics assessment scores.

This chapter describes the background, purpose, and design of the LEAP 2025 assessments.

1.1 Background

In 2010, the Board of Elementary and Secondary Education (BESE) approved the Common Core State Standards (CCSS) in ELA and mathematics. After adopting the CCSS, Louisiana became a governing member of PARCC, a group of states working to develop high-quality assessments that measure the full range of the CCSS.

To prepare for the PARCC assessments and help ease the transition to the new standards, the Louisiana Department of Education (LDOE) incrementally revised the LEAP and *i*LEAP ELA and mathematics assessments in grades 3 through 8 and administered transitional tests during the 2012–2013 and 2013–2014 school years.

In the 2014–2015 school year, students in grades 3–8, except those qualifying for the LEAP Alternate Assessment, Level 1 (LAA 1), took the PARCC assessments for ELA and mathematics, which included two components: the performance-based assessment (PBA), which was administered in March, and the end-of-year assessment (EOY), which was administered in May.

As a result of a legislative agreement reached during the summer of 2015, and to maintain comparability to the 2015 assessments, the LEAP ELA and mathematics assessments in grades 3–8 for the 2015–2016 school year consisted of items taken from both the PARCC assessments (no more than 49.9%) and DRC's College and Career Readiness item bank.

In March 2016, BESE approved the Louisiana Student Standards in ELA and mathematics. In the 2016–2017, 2017–2018, 2018–2019, 2020–2021, 2021-2022, and 2022-2023 school years, students in grades 3–8, except those qualifying for an alternate assessment for students with the most significant cognitive disabilities (the LAA 1 in 2016–2017 or LEAP Connect in subsequent years), were administered forms for ELA and mathematics that consisted of New Meridian (formerly PARCC) assessment items while developing some Louisiana-owned items to enhance the New Meridian item bank. This allowed for the continued comparability to forms administered in the 2014–2015 and 2015–2016 school years. Louisiana received approval from the federal and state governments to waive the requirement to administer the spring 2020 assessment due to school facilities closing in March 2020 due to COVID-19.

The information that follows describes the technical aspects of the 2023 LEAP 2025 ELA and mathematics assessments and provides information about how to read and interpret the data.

1.2 Purpose of the LEAP 2025

The BESE and the LDOE are committed to ensuring that every student is on track to be successful in either postsecondary education or the workforce. The LEAP 2025 supports this vision by measuring the full range of student performance and providing information for educators and parents about student readiness for college and careers.

1.3 Design of the LEAP 2025

Students in grades 3–8 were administered computer-based tests (CBTs) in both ELA and mathematics; some school systems opted to administer paper-based tests (PBTs) to students in grade 3. All mathematics assessments were translated into Spanish forms. Additionally, a braille form was available for each grade and content area. The braille form was based on the PBT in grade 3 and was based on the CBT in grades 4–8. Online tools allowed students to magnify assessment items, as needed, and students with visual impairments could also take large-print versions of the PBTs. See Chapter 3, Section 3.4 for more information about the accommodations and designated supports available for students taking the LEAP 2025.

The 2023 LEAP 2025 test blueprints and test design for ELA and mathematics are based on the ELA https://resources.newmeridiancorp.org/ela-test-design/ and mathematics https://resources.newmeridiancorp.org/math-test-design/ blueprints of New Meridian's full forms. The 2023 LEAP 2025 test blueprints and test design for ELA and mathematics differ from the New Meridian blueprints and design in order to reduce testing time while maintaining full coverage and including a variety of standards.

The 2023 LEAP 2025 ELA blueprints kept a similar design as the design of New Meridian's full form, which includes both performance-based tasks and stand-alone passage sets, and a higher percentage of reading points to writing points. However, to address concerns about overtesting, only two of the three types of performance tasks—Research Simulation Task and Literary Analysis Task **or** Narrative Writing Task—are included on each of the grade-level tests. All three task types are represented across grades 3–8, which allows Louisiana flexibility in the choice of the tasks administered for each grade from year to year and encourages teachers to focus equally on all three writing types. Besides having two (instead of three) performance tasks, the 2023 LEAP 2025 Spring ELA blueprints are also different with respect to testing time and percentage of reading and writing points. Since the choice of Literary Analysis Task or Narrative Writing Task is determined during the forms construction process, alternative blueprints—one with a Literary Analysis Task and a Research Simulation Task and the other with a Research Simulation Task and a Narrative Writing Task—were created for each grade's assessment.

The passages chosen for the 2023 LEAP 2025 ELA assessments contain a variety of texts of different genres and a diverse set of authors. The assessments also contain texts that appeal to a diverse student population. Chosen passages are authentic and contain a variety of different types of text that cover a range of text complexities—Readily Accessible (RA), Moderately Complex (MOD), and Very Complex (VC).They are rich in content, engaging, high-quality, and challenging. Additionally, paired passages, which allow a mix of text complexities and sometimes types of texts–both informational and literary– are selected with careful consideration of the purpose of the standards that require the use of more than one text. This combination of criteria during passage selection allows students to demonstrate their ability to read and comprehend a range of grade-appropriate texts and topics and helps to ensure as much coverage of the standards as possible.

The LEAP 2025 ELA assessments focus on an integrated approach to reading and writing that reflects instruction in an effective ELA classroom and measures students' ability to understand what they read and express that understanding in writing. This means careful, close reading of complex grade-level literary and informational texts; a full range of texts from across the disciplines, including science, social studies, and the arts; tasks that integrate key ELA skills by asking students to read texts, answer reading and vocabulary questions about the texts, and then write using evidence from what they have read; questions worth answering, ordered in a way that builds meaning; a focus on students citing evidence from texts when answering questions about a specific passage or when writing about a set of related passages; and a focus on words that matter most in texts, are essential to understanding a particular text, and include context that allows students to determine literal and figurative meanings.

The LDOE has finalized an ELA test design that takes into account several key considerations:

- Since testing time continues to be a significant factor in test design decisions, it was determined that two of the three task types will be used in each form.
- The test must include opportunities for students to write about both literary and informational texts; therefore, each test includes a Research Simulation Task and either a Literary Analysis Task (LAT) or a Narrative Writing Task (NWT). By having both blueprints available for each grade and selecting forms based on using the best of the tasks in each form rather than following a pre-specified plan for alternating LAT/NWT forms, both of the literary task types maintain their place in the curriculum.
- The passages associated with each task and the standalone passage sets used across a form represent a range of text complexities, depending on the grade and test design.
- Although the items are dependent on the topic and complexity of the passages, the goal is to include a range of DOK levels, with more DOK 2 than DOK 3 items across a form. Item complexity is also dependent on other factors, such as item type and language complexity.
- The third session also includes a field test slot to allow for embedded field testing of one passage set per form, which provides opportunities for field testing with all students without increasing testing time. In fact, the testing time for LEAP 2025 including the field test positions is less than the testing time for New Meridian's full form. All students that are administered the ELA assessment take field test items. The field test positions contain placeholder items when field testing is not being conducted.

The 2023 LEAP 2025 mathematics blueprints kept a similar design as those of New Meridian's full form, with a few notable exceptions:

- In grades 3-5, the LEAP 2025 blueprints make use of three sessions with a total testing time of 235 minutes, instead of four sessions with a total testing time of 240 minutes.
    - In grade 3, the difference in items is a reduction of 1 Type II item worth 4 points and an increase of 2 Type I items worth 1 point with a corresponding decrease of 1 Type I item worth 2 points. Therefore, the total number of items is the same across both designs, but LEAP 2025 has 4 fewer points.
    - In grades 4 and 5, there is a bigger difference, as LEAP 2025 uses the same test design for grades 3-5, so the increase in type I 1-point items is 8 with a decrease in 4 2-point items in addition to the reduction of 1 Type II item worth 4 points.
- In grades 6-8, both assessment designs have three sessions and a total testing time of 240 minutes. However, New Meridian uses three sessions of equal testing time with 80 minutes each, while LEAP 2025 has a shorter non-calculator session 1 (60 minutes) followed by two 90-minute calculator sections. New Meridian has a split session in grade 7 mathematics for session 1 in which the non-calculator and calculator sections are split within the same session/unit. In grades 6 and 8, the entire first session/unit is designated as non-calculator. The LEAP 2025 test design has consistency across grades 6-8 in testing time per session and has either non-calculator or calculator as the designation for the entire session for ease of administration.
    - In grades 6 and 7, the LEAP 2025 design uses 8 more type I items worth 1 point, 2 fewer type I items worth 2 points, and 1 fewer type I item worth 4 points. (LEAP 2025 does not use any type I items worth 4 points.) Grades 6-8 use the same number of type II and III items in both test designs.
    - LEAP 2025 uses the same test design for grade 8, so there are 8 more type I items worth 1 point and 2 fewer type I items worth 4 points (but the same number of type I items worth 2 points).

The LEAP 2025 mathematics assessments focus on testing the Louisiana Student Standards for Mathematics (LSSM) according to the components of rigor reflected in high-quality mathematics instructional tasks that

- require students to demonstrate understanding of mathematical reasoning in mathematical and applied contexts;
- assess accurate, efficient, and flexible application of procedures and algorithms;
- rely on application of procedural skill and fluency to solve complex problems; and
- require students to demonstrate mathematical reasoning and modeling in real-world contexts.

The LSSM support students to become mathematically proficient by focusing on three components of rigor: conceptual understanding, procedural skill and fluency, and application.

- Conceptual understanding refers to understanding mathematical concepts, operations, and relations. It is more than knowing isolated facts and methods. Students should be able to make sense of why a mathematical idea is important and the kinds of contexts in which it is useful. It also allows students to connect prior knowledge to new ideas and concepts.
- Procedural skill and fluency is the ability to apply procedures accurately, efficiently, and flexibly. It requires speed and accuracy in calculation while giving students opportunities to practice basic skills. Students' ability to solve more complex application tasks is dependent on procedural skill and fluency.
- Application provides a valuable context for learning and the opportunity to solve problems in a relevant and a meaningful way. It is through real-world application that students learn to select an efficient method to find a solution, determine whether the solution(s) makes sense by reasoning, and develop critical thinking skills.

Each item on the LEAP 2025 mathematics assessment is referred to as a task and is identified by one of three types: Type I, Type II, or Type III. The tasks on the LEAP 2025 mathematics test are aligned directly to the LSSM for all reporting categories.

- Type I tasks, designed to assess conceptual understanding, fluency, and application, are aligned to the major, additional, and supporting content for each grade. Some Type I tasks may be further aligned to LEAP 2025 evidence statements for the Major Content and Additional & Supporting reporting categories and allow for the testing of more than one of the student standards on a single task.
- Type II tasks are designed to assess student reasoning ability of selected major content for the grade or the previous grade in applied contexts.
- Type III tasks are designed to assess student modeling ability of selected content for the grade or the previous grade in applied contexts. Type II and III tasks are further aligned to LEAP 2025 evidence statements for the Expressing Mathematical Reasoning and Modeling & Application reporting categories.

Each of the three task types is aligned to one of four reporting categories: Major Content, Additional & Supporting Content, Expressing Mathematical Reasoning, or Modeling & Application. Each task type is designed to align with at least one of the Louisiana Student Standards for Mathematical Practice (MP).

Additional details about the design of the ELA and mathematics assessments can be found in Chapter 3.

Chapter 2: The Uses of Test Scores

Validity is the central component of any analysis of the LEAP 2025 assessments. The following excerpt is from the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014):

> Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing system. Different components of validity evidence…include evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all test takers, as appropriate to the test interpretation in question (22).

As stated by the *Standards*, the validity of a testing program hinges on the use of the test scores. Validity evidence that supports the uses of the LEAP 2025 test scores is provided in this technical report. This chapter examines some possible uses of the LEAP 2025 test scores. However, this technical report cannot anticipate all possible interpretations and uses of the LEAP 2025 test scores.

2.1 Uses of Test Scores

To understand whether a test score is being used properly, one must understand the purpose of the test. The intended uses of the LEAP 2025 test scores include the following:

- evaluating students' overall proficiency of the Louisiana Student Standards
- identifying students' general strengths and weaknesses
- evaluating programs at the school, school system, and/or state level
- informing stakeholders, including students, teachers, school administrators, school system administrators, LDOE staff members, parents, and the public, of the status of students' progress toward meeting college and career readiness standards

This technical report refers to the uses of the test-level scores (i.e., scale scores and achievement levels), category-level scores and achievement-level classifications, and subcategory-level scores and achievement-level classifications.

2.2 Test-Level Scores

At the test level, an overall scale score that is based on student performance on the entire test is reported. In addition, an associated level of achievement is reported. These scores and achievement levels indicate, in varying ways, a student's achievement in ELA or mathematics. Test-level scores are reported at four reporting levels: the state, the school system, the school, and the student.

The LEAP 2025 high school ELA and mathematics test forms were developed by DRC's test development staff using New Meridian's item bank as well as items from the Louisiana Department of Education's own item bank. Items taken from these banks were on pre-established item response theory (IRT) scales for ELA and mathematics and were reviewed and approved for use by LDOE content experts and committees of Louisiana educators. Braille forms and Spanish translations of mathematics forms were also developed. See Chapter 3, "Test Content Development," for additional details about the processes used to develop these test forms.

The following sections discuss two types of test-level scores that are reported that indicate a student's achievement on the LEAP 2025 assessments: the scale score and its associated level of achievement.

2.3 Scale Scores

A scale score indicates a student's total performance for each content area on the LEAP 2025 assessments. The overall scale score for a content area quantifies the achievement being measured by the ELA or mathematics assessments. In other words, the scale score represents the student's level of achievement, where higher scale scores indicate higher levels of achievement on the test and lower scale scores indicate lower levels of achievement. For all LEAP 2025 test forms, the lowest obtainable scale score (LOSS) is 650 and the highest obtainable scale score (HOSS) is 850.

Scale scores are derived from raw scores (i.e., the number of items answered correctly). Raw scores depend on the items in a particular form of a test and can only be interpreted in terms of that particular set of test questions. This does not allow year-to-year or form-to-form comparison. Scale scores are more meaningful than raw scores because they maintain their meaning year-to-year, thus allowing comparisons of different test forms across the entire range of the ability scale.

2.4 Levels of Achievement

A student's performance on the ELA or mathematics LEAP 2025 assessments is reported in one of five levels of achievement: *Advanced, Mastery, Basic, Approaching Basic,* or *Unsatisfactory*. The cut scores for the ELA and mathematics achievement levels were established by PARCC using the Evidence-Based Standard Setting (EBSS) method (Beimers, Way, McClarty, & Miles, 2012) for the PARCC Performance-Level Setting (PLS) process. Details regarding the PLS process can be found in the *Performance Level Setting Technical Report* (Pearson, 2015).

Descriptions of each level of achievement in terms of what a student should know and be able to do are provided with the LEAP 2025 *Interpretive Guide* (see Chapter 7).

2.5 Use of Test-Level Scores

The LEAP 2025 scale scores and achievement levels provide summary evidence of student performance in ELA or mathematics relative to the Louisiana Student Standards. Classroom teachers may use these scores as evidence of student achievement in these content areas. At the aggregate level, school system and school administrators may use this information for activities such as curriculum planning. The results presented in this technical report provide evidence that the scale scores and achievement levels are valid and reliable indicators of what students know, understand, and are able to do relative to the Louisiana Student Standards in ELA and mathematics.

2.6 Category- and Subcategory-Level Subscores

A student's performance on the ELA categories (i.e., reading and writing) is reported by one of three ratings: *Strong*, *Moderate*, or *Weak*. Additionally, performance on the subcategories is reported at the student level for ELA and mathematics. ELA has three subcategories for reading and two subcategories for writing, as described in Table 3.1, *ELA Categories and Subcategories*. Mathematics has four reporting categories: Major Content, Additional & Supporting Content, Expressing Mathematical Reasoning, or Modeling & Application., as described in Table 3.8, *Overview of LEAP 2025 Mathematics Task Types and Reporting Categories*. Reporting categories are further broken down into subcategories, which vary by grade level. Subcategory performance is reported in one of three ratings: *Strong*, *Moderate*, or *Weak*.

Although the performance ratings are determined only by the items included within a category or subcategory, the level of knowledge and ability needed to demonstrate a performance rating is connected to the level of knowledge and ability required by the content-level assessments; a *Strong* rating requires similar knowledge and ability as the Mastery or Advanced achievement levels, a *Moderate* rating requires similar

knowledge and ability as the Basic achievement level, and a *Weak* rating requires similar knowledge and ability as the Unsatisfactory and Approaching Basic achievement levels.

2.7 Use of the Reporting Category- and Subcategory-Level Ratings

The purpose of reporting category- or subcategory-level performance ratings on LEAP 2025 assessments is to show, for each student, the relationship between the overall achievement being measured and the skills in each of the areas defined by the categories and subcategories. These ratings for individual students are best corroborated by other evidence, such as grades, teacher feedback, and scores on other tests. Chapter 3 of this technical report provides evidence of content validity that supports the use of the category- or subcategory-level performance ratings. Chapter 9 of this technical report provides evidence of construct-related validity that further supports the use of these performance ratings.

Chapter 3: Test Content Development

Content-related validity in achievement tests is evidenced by a correspondence between test content and the range of knowledge and skills that compose the construct the assessment is designed to measure, i.e., the ELA or mathematics Louisiana Student Standards. Content-related validity can be demonstrated through consistent adherence to test blueprints, through a high-quality test development process that includes review of items for accessibility to English learners and students with disabilities, and through alignment studies performed by independent groups. This chapter provides a detailed discussion of the test development process. In particular, it shows how rigorous procedures were followed to construct tests that reflect the full range of content that the 2023 LEAP 2025 assessments were expected to cover.

This chapter is particularly relevant to the following sections of the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014): Standards 4.0, 4.1, and 4.7. It also addresses Standards 3.1, 3.2, 3.9, and 4.12, which are discussed in pertinent sections of this chapter.

Standard 4.0 states the following:

> Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population (85).

Standard 4.1 states the following:

> Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s) (85).

The 2023 LEAP 2025 test specifications consisted of a test blueprint and a test design for each grade and content area. The 2023 blueprints and test designs were closely aligned to blueprints of New Meridian's full forms. The specific content area and grade-level test blueprints for the 2023 LEAP 2025 ELA assessments for grades 3–8 were designed with the goal for all students to read, understand, and express understanding of complex, grade-level texts. The specific content area and grade-level test blueprints for the 2023 LEAP 2025 mathematics assessments for grades 3–8 were designed with the goal of supporting students to become mathematically proficient by focusing on three components of rigor: conceptual understanding, procedural skill and fluency, and application. The 2023 LEAP 2025 ELA and mathematics assessments for grades 3–8 provide questions that have been reviewed by Louisiana educators to ensure their alignment to the Louisiana Student Standards and appropriateness for Louisiana students, measure the full range of student performance, and inform educators and parents about student readiness in ELA and mathematics and whether students are "on track" for college and careers. For ELA and mathematics, the 2023 LEAP 2025 assessments for grades 3–8 use the same reporting categories that were used in spring 2019, 2021, and 2022. Subcategories in mathematics were introduced for spring 2018 in response to requests from school systems. In ELA, the type and/or number of reading literary and informational passage sets changed from the 2017 LEAP 2025 assessments to the 2018 LEAP 2025 ELA assessments to reflect a similar change made in the PARCC blueprints. This change was continued for the 2021, 2022, and 2023 LEAP 2025 ELA assessments.

To construct the assessments after the test blueprints and test designs were approved, the LDOE and DRC collaborated to use items, aligned to the Louisiana Student Standards, from the New Meridian and Louisiana-

owned item banks. DRC contracted with New Meridian and was provided access to the entire bank of items and passage sets that could potentially be used on operational forms. The acquired items and passages and the Louisiana-owned items and passage sets make up the available item pool for the 2023 LEAP 2025 forms construction. The LDOE and DRC confirmed that all items selected for use on the LEAP 2025 forms were appropriate for use on Louisiana assessments by convening committees of Louisiana educators who reviewed and approved items from the item banks prior to form selection. This process is followed annually to ensure the monitoring, maintenance, and improvement of a quality item bank to use during form selection.

The ELA and mathematics LEAP 2025 assessments for grades 3–8 were developed based on the requirements of "RFP #678PUR-LEAP 2025 English Language Arts and Mathematics Assessment System" as follows:

The assessments shall be

- aligned to the ELA and mathematics Louisiana Student Standards;
- designed to be accessible for use by the widest possible range of students, including, but not limited to, students with disabilities and students with limited English proficiency [English Learners];
- constructed to yield valid and reliable test results;
- constructed to report student performance using achievement level policy definitions and reporting categories that are comparable to a significant number of other states and, for grades 3 through 8 assessments, to Louisiana's 2015–2019 and 2021-2022 assessments;
- constructed to use Louisiana's grades 3 through 8 ELA and mathematics assessments as the baseline scale[1] to report test results for grades 3 through 8 students;
- developed to limit the amount of testing time required and to be in compliance with state law regarding testing time;
- developed and reviewed with Louisiana educators;
- non-computer adaptive;
- used in assessing students' readiness to successfully transition to postsecondary education and the workplace; and
- administered, scored, and reported through a separate administration contract in both paper- and computer-based formats.

The products of the above requirements are dual-mode assessments—paper-based tests (PBTs) and computer-based tests (CBTs)—comprised of New Meridian and Louisiana-owned items aligned to the Louisiana Student Standards. Louisiana had access to the complete New Meridian item bank for forms administered in spring 2023. For grade 3, the contract with New Meridian provided for the use of enough items and passage sets, which had been approved during Item Alignment Reviews, combined with additional items and passage sets developed specifically for Louisiana, to create one complete operational test form for each content area and grade that can be administered in a dual-mode testing environment (i.e., PBT and CBT). For grades 4–8, Louisiana selected one CBT form per grade from the content that was reviewed during Item Alignment Reviews in addition to items and passage sets developed specifically for Louisiana. These items and passage sets became the available item pool used to construct the 2023 forms. DRC and LDOE content experts scrutinized each final blueprint to ensure optimal content coverage and prudent use of time and resources. In general, the blueprints represent content sampling proportions that reflect intended

[1] In the spring of 2016 and 2017, PARCC item parameters were used to place the LEAP 2025 assessments on the PARCC scale. In the spring of 2018, PARCC items that had been previously administered in Louisiana were available, so the item parameters generated from Louisiana students were used to create the LEAP 2025 scale. The LEAP 2025 scale is comparable to the PARCC scale. Future LEAP 2025 assessments will be linked to the spring 2018 LEAP 2025 scale, which is considered the baseline.

emphasis in instruction and mastery at each grade level and are comparable to New Meridian's test blueprints. The test specifications provide the numbers of items by reporting category, assessment focus, or item type, and they demonstrate the desired proportions within test delivery and available item pool constraints. These specifications can be found in the 2022-2023 *LEAP 2025 Grades 3-8 English Language Arts and Mathematics Assessment Frameworks*. All assessments were fixed forms, which means that all students who received the same form were administered the same set of items, as the forms were not adaptive.

3.1 Defining the Specific Test Blueprint

The specific content area and grade-level test blueprints were designed based on two primary factors: (1) the content requirements of the Louisiana Student Standards and (2) the reporting needs of the assessments.

3.2 English Language Arts Test Blueprints and Test Designs

The ELA test was administered during a CBT testing window (April 25-May 26, 2023) and during a PBT testing window (April 26-May 2, 2023). Only two of the three types of performance tasks—Research Simulation Task, Literary Analysis Task, and Narrative Writing Task—were included on each of the Louisiana grade-level tests; however, all three types were represented across grades 3 through 8. This allows Louisiana to rotate the tasks given for each grade from administration to administration and encourages educators to focus on all three performance task types. As the choice of Literary Analysis Task or Narrative Writing Task would be made during the forms construction process, alternative blueprints—one with a Literary Analysis Task and a Research Simulation Task and the other with a Research Simulation Task and a Narrative Writing Task—were created for each grade. During forms construction, the Narrative Writing Task was selected for grades 4 and 6 and the Literary Analysis Task was selected for grades 3, 5, 7, and 8, based on item performance and the quality of the available passage sets for each performance task.

Student performance on the LEAP 2025 ELA assessments is reported by category and subcategory as outlined in the following table.

Table 3.1 ELA Categories and Subcategories

Category	Subcategory	Subcategory Description
Reading	Reading Literary Text	Students read and demonstrate comprehension of grade-level fiction, drama, and poetry.
	Reading Informational Text	Students read and demonstrate comprehension of grade-level nonfiction, including texts about history, science, art, and music.
	Reading Vocabulary	Students use context to determine the meaning of words and phrases in grade-level texts.
Writing	Written Expression	Students use details from provided texts to compose well-developed, organized, clear writing.
	Knowledge and Use of Language Conventions	Students use the rules of Standard English (grammar, mechanics, and usage) to compose writing.

These reporting categories provide parents and educators with valuable information about

- overall student performance, including readiness to continue further study in English language arts;

- student performance broken down by subcategory which may help identify when students need additional support or more challenging work in reading and writing; and
- how well schools and school systems help students achieve expectations.

The session testing times shown in the ELA test blueprints (see Tables 3.2 through 3.6) are based on New Meridian testing times proportioned to be comparable based on the passage type being tested. The passage set that comes after the Narrative Writing Task is designed to balance the reading load between the Literary Analysis Task and the Narrative Writing Task. It is also designed to provide consistent timing in sessions 1 and 2.

Table 3.2 Grade 3 English Language Arts Test Blueprint and Test Design

Session	Content	Number of Passages	Categories/ Subcategories	Number of Two-Point SR Items	Number of Points from Two-Point SR Items	Number of PCR Items	Number of Points from PCR Items	Total Items	Total Points	Assessable ELA Student Standards (by subcategory)	Testing Time (minutes)
1	Literary Analysis Task	2	Reading: Reading Literary Text/Reading Vocabulary*	6	12	1	3	6	15	RL standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	75
			Writing: Written Expression	0	0		9	1	9	Writing standards W.1-2, 10	
			Writing: Knowledge and Use of Language Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Totals	2		6	12	1	15	7	27		
2	Research Simulation Task	2	Reading: Reading Informational Text/Reading Vocabulary*	6	12	1	3	6	15	RI Standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	75
			Writing: Written Expression	0	0		9	1	9	Writing standards W.1-2, 7-8, 10	
			Writing: Knowledge and Use of Language Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Totals	2		6	12	1	15	7	27		
3	Reading Literary Texts	2	Reading: Reading Literary Text/Reading Vocabulary*	8	16	0	0	8	16	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	60**
	Reading Informational Texts		Reading: Reading Informational Text/Reading Vocabulary*							RI standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	
	Totals	2		8	16	0	0	8	16		
Grade 3 Totals		6	Reading: Reading Literary Text/Reading Vocab*	20	40	2	3	20	46	46	210
			Reading: Reading Informational Text/Reading Vocab*				3				
			Writing: Written Expression	0	0		18	2	18	24	
			Writing: Knowledge and Use of Language Conventions	0	0		6		6		
			Total	20	40	2	30	22	70	70	

**Reading vocabulary items must constitute at least eight points on the test.*
***The time in session 3 allows for an additional passage set that is being field tested.*
As described in section 1.3, the passages associated with each task and the standalone passage sets used across a form should represent a range of text complexities as appropriate for the grade level and test design.

Table 3.3 Grade 4 English Language Arts Test Blueprint and Test Design

Session	Content	Number of Passages	Categories/ Subcategories	Number of Two-Point SR Items	Number of Points from Two-Point SR Items	Number of PCR Items	Number of Points from PCR Items	Total Items	Total Points	Assessable ELA Student Standards (by subcategory)	Testing Time (minutes)
1	Research Simulation Task	3	Reading: Reading Informational Text/Reading Vocabulary*	8	16	1	4	8	20	RI Standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	90
			Writing: Written Expression	0	0		12	1	12	Writing standards W.1-2, 4, 9, 10,	
			Writing: Knowledge and Use of Language Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Totals	3		10	20	1	19	11	39		
2	Narrative Writing Task	1	Reading: Reading Informational Text/ Reading Vocabulary*	4	8	1	0	4	8	RL standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	90
			Writing: Written Expression	0	0		9	1	9	Writing standards W.3, 4, 10	
			Writing: Knowledge and Use of Language Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Reading Literary / Informational Texts	1-2	Reading: Reading Informational Text/Reading Literature Text/Reading Vocabulary*	6	12	0	0	6	12		
	Totals	2-3		10	20	1	12	11	32		
3	Reading Literary Texts	2	Reading: Reading Literary Text/Reading Vocabulary*	8	16	0	0	8	16	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	60**
	Reading Informational Texts		Reading: Reading Informational Text/Reading Vocab*							RI standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	
	Totals	2		8	16	0	0	8	16		
Grade 4 Totals		7-8	Reading: Reading Literary Text/Reading Vocab*	26	52	2	0	26	56	56	240
			Reading: Reading Informational Text/Reading Vocab*				4				
			Writing: Written Expression	0	0		21	2	21	27	
			Writing: Knowledge and Use of Language Conventions	0	0		6		6		
			Total	26	52	2	31	28	83	83	

**Reading vocabulary items must constitute at least eight points on the test.*

***The time in session 3 allows for an additional passage set that is being field tested.*

As described in section 1.3, the passages associated with each task and the standalone passage sets used across a form should represent a range of text complexities as appropriate for the grade level and test design.

Table 3.4 Grade 5 English Language Arts Test Blueprint and Test Design

Session	Content	Number of Passages	Categories/ Subcategories	Number of Two-Point SR Items	Number of Points from Two-Point SR Items	Number of PCR Items	Number of Points from PCR Items	Total Items	Total Points	Assessable ELA Student Standards (by subcategory)	Testing Time (minutes)
1	Literary Analysis Task	2	Reading: Reading Literary Text/Reading Vocabulary*	6	12	1	4	6	16	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	90
			Writing: Written Expression	0	0		12	1	12	Writing standards W.1-2, 4, 9, 10,	
			Writing: Knowledge and Use of Language Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Reading Literary / Informational Texts	1	Reading (Reading Literary Text/Reading Informational Text/Reading Vocabulary)	4	8	0	0	4	8	RL Standards 1-3, 5-10; RI standards 1-3, 5-10; vocabulary standards RL.4, RI.4, L.4, L.5	
	Totals	3		10	20	1	19	11	39		
2	Research Simulation Task	3	Reading: Reading Informational Text/ Reading Vocabulary*	8	16	1	4	8	20	RI standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	90
			Writing: Written Expression	0	0		12	1	12	Writing standards W.1-2, 4, 7- 10,	
			Writing: Knowledge and Use of Language Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Totals	3		8	16	1	19	9	35		
3	Reading Informational Texts	1-2	Reading: Reading Informational Text/Reading Vocab*	6	12	0	0	6	12	RI standards 1-3, 5, 7-10; vocabulary standards RI.4, L.4, L.5	60**
	Totals	1-2		6	12	0	0	6	12		
Grade 5 Totals		7-8	Reading: Reading Literary Text/Reading Vocab*	10	20	2	4	10	24	56	240
			Reading: Reading Informational Text/Reading Vocab*	14	28		4	14	32		
			Writing: Written Expression	0	0		24	2	24	30	
			Writing: Knowledge and Use of Language Conventions	0	0		6		6		
			Total	24	48	2	38	26	86	86	

**Reading vocabulary items must constitute at least eight points on the test.*

***The time in session 3 allows for an additional passage set that is being field tested.*

As described in section 1.3, the passages associated with each task and the standalone passage sets used across a form should represent a range of text complexities as appropriate for the grade level and test design.

Table 3.5 Grade 6 English Language Arts Test Blueprint and Test Design

Session	Content	Number of Passages	Categories/ Subcategories	Number of Two-Point SR Items	Number of Points from Two-Point SR Items	Number of PCR Items	Number of Points from PCR Items	Total Items	Total Points	Assessable ELA Student Standards (by subcategory)	Testing Time (minutes)
1	Research Simulation Task	3	Reading: Reading Informational Text/Reading Vocabulary*	8	16	1	4	8	20	RI standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	90
			Writing: Written Expression	0	0		12	1	12	Writing standards W.1-2, 4, 7-10,	
			Writing: Knowledge and Use of Language Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Totals	3		8	16	1	19	9	35		
2	Narrative Writing Task	1	Reading: Reading Literary Text/Reading Vocabulary*	4	8	1	0	4	8	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	90
			Writing: Written Expression	0	0		12	1	12	Writing standards W.3, 4, 10	
			Writing: Knowledge and Use of Language Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Reading Literary / Informational Texts	1-2	Reading (Reading Literary Text/Reading Informational Text/Reading Vocabulary)	6	12	0	0	6	12	RL Standards 1-3, 5-10; RI standards 1-3, 5-10; vocabulary standards RL.4, RI.4, L.4, L.5	
	Totals	2-3		10	20	1	15	11	35		
3	Reading Literary Texts	2	Reading: Reading Literary Text/Reading Vocabulary*	10	20	0	0	10	20	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	80**
	Reading Informational Texts		Reading: Reading Informational Text/Reading Vocab*			0	0			RI.1-3, 5, 7-10; vocabulary standards RI.4, L.4, L.5	
	Totals	2		10	20	0	0	10	20		
Grade 6 Totals		7-8	Reading: Reading Literary Text/Reading Vocab*	28	56	2	0	28	60	60	260
			Reading: Reading Informational Text/Reading Vocab*				4				
			Writing: Written Expression	0	0		24	2	24	30	
			Writing: Knowledge and Use of Language Conventions	0	0		6		6		
			Total	28	56	2	34	30	90	90	

**Reading vocabulary items must constitute at least eight points on the test.*
***The time in session 3 allows for an additional passage set that is being field tested.*
As described in section 1.3, the passages associated with each task and the standalone passage sets used across a form should represent a range of text complexities as appropriate for the grade level and test design.

Table 3.6 Grades 7 and 8 English Language Arts Test Blueprint and Test Design

Session	Content	Number of Passages	Categories/ Subcategories	Number of Two-Point SR Items	Number of Points from Two-Point SR Items	Number of PCR Items	Number of Points from PCR Items	Total Items	Total Points	Assessable ELA Student Standards (by subcategory)	Testing Time (minutes)
1	Literary Analysis Task	2	Reading: Reading Literary Text/Reading Vocabulary*	6	12	1	4	6	16	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	90
			Writing: Written Expression	0	0		12	1	12	Writing standards W.1-2, 4, 9, 10,	
			Writing: Knowledge and Use of Language Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Reading Literary / Informational Texts	1	Reading (Reading Literary Text/Reading Informational Text/Reading Vocabulary)	4	8	0	0	4	8	RL Standards 1-3, 5-10; RI standards 1-3, 5-10; vocabulary standards RL.4, RI.4, L.4, L.5	
	Totals	3		10	20	1	19	11	39		
2	Research Simulation Task	3	Reading: Reading Informational Text/ Reading Vocabulary*	8	16	1	4	8	20	RI standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	90
			Writing: Written Expression	0	0		12	1	12	Writing standards W.1-2, 4, 7- 10,	
			Writing: Knowledge and Use of Language Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Totals	3		8	16	1	19	9	35		
3	Reading Literary Texts	2	Reading: Reading Literary Text/Reading Vocabulary*	10	20	0	0	10	20	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	80**
	Reading Informational Texts		Reading: Reading Informational Text/Reading Vocab*			0	0			RI standards 1-3, 5, 7-10; vocabulary standards RI.4, L.4, L.5	
	Totals	2		10	20	0	0	10	20		
Grades 7 and 8 Totals		8	Reading: Reading Literary Text/Reading Vocab*	28	56	2	4	28	64	64	260
			Reading: Reading Informational Text/Reading Vocab*				4				
			Writing: Written Expression	0	0		24	2	24	30	
			Writing: Knowledge and Use of Language Conventions	0	0		6		6		
			Total	28	56	2	38	30	94	94	

**Reading vocabulary items must constitute at least eight points on the test.*
***The time in session 3 allows for an additional passage set that is being field tested.*
As described in section 1.3, the passages associated with each task and the standalone passage sets used across a form should represent a range of text complexities as appropriate for the grade level and test design.

The LEAP 2025 ELA assessments consist of tasks and reading passage sets. The tasks are described below.

- Narrative Writing Task
    - This task asks students to read a literary text, answer a set of selected-response questions about the text, and create a narrative related to the text (e.g., finish the story, retell the story in another narrative form or from a different point of view).
    - This task focuses on students' ability to use narrative elements (e.g., dialogue, description) when writing.
- Literary Analysis Task
    - This task provides students with an opportunity to show their understanding of literature. It asks students to read two literary texts, answer a set of selected-response questions about the texts, and write an extended response that compares and/or explains key ideas or elements in the texts (e.g., central idea/message, contribution of illustrations, characterization).
    - This task focuses on students' ability to read complex text closely and asks them to carefully consider literature worthy of close study.
- Research Simulation Task
    - This task mirrors the research process by presenting three texts on a given topic. Students answer a set of selected-response questions about the texts and then write an extended response about some aspect of the related texts (e.g., relationship between a series of events, ideas, or concepts; comparison/contrast of key details; presentation of information).
    - This task requires students to synthesize information from related informational resources.

The following item types were included in the 2023 LEAP 2025 ELA assessments:

- Selected-Response Items:
    - Evidence-based selected response– EBSR: This item type consists of two parts; one part asks students to show their understanding of a text, and the other part asks students to identify evidence to support that understanding. The evidence supports a generalization, conclusion, or inference. This type of item is designed to provide students opportunities to make explicit the evidence that supports their close analysis of a specific text.
    - Multiple select – MS: This item type requires students to select more than one correct answer and may appear as a one-part question or as part of an EBSR item. This type of item allows for the assessment of students' ability to identify multiple pieces of evidence to support a claim.
    - Technology enhanced – TE: This item type allows measurement of learning that may not be sufficiently measured by traditional multiple-choice items: ordering of ideas within a summary; ordering of steps in a process; sorting, classifying, and categorizing ideas; matching of two themes/ideas to their unique evidence, etc. The technology offers students additional ways to show understanding that parallels the classroom instructional techniques teachers use to determine whether students are able to comprehend grade-level, complex text. TE Items may involve any of the following:
        - Highlighting text: requires a student to select text-based answer(s) from within a larger text

    - Drag and drop: requires a student to move draggable elements (e.g., words, phrases, or sentences) into one or more drop boxes (e.g., cells within a table or part[s] of a diagram)
    - Drop-down menu: requires a student to select from one or more drop-down menus to complete a phrase or sentence
    - Match interaction table: requires a student to select a checkbox in each row from two or more columns to classify statements presented in each row
- Prose constructed response – PCR: This item type appears at the end of each of the tasks and asks students to create an extended, complete written response. It elicits evidence that students have understood a text or texts they have read and can communicate that understanding well, both in terms of written expression and in terms of knowledge and use of language conventions.

A variety of item types allows for the measurement of the full range of student performance. To ensure a range of item complexity beyond the DOK level, the list below includes some of the key elements that are considered when creating items or new passage sets and selecting items for a passage set and across a form:

- The item type that best addresses the standards the item measures (e.g., standard RI.2 at some grade levels requires students to identify two main ideas, so an MS or TE item should be used when measuring this standard fully; a TEI should be using when measuring the ordering required in an RL.2 summary item.)
- A variety of items to assess more complex standards across a passage set and form (e.g., RL.6 at grades 6-8 includes point of view and purpose, which would require separate items to assess the standard fully. See the Grades 3-11 Reading and Writing Evidence Statements for more information about how each standard should be assessed.)
- The reading load and other demands of an item, which include the number of correct answers required and number of distractors for EBSR and MS items and number of interactions and distractors for TE items

All items and tasks are clearly aligned to specific standards. Most include a primary standard, as well as standard 1, which requires evidence to support the primary standard. The PCRs align to several standards since they measure reading and/or writing skills that are articulated in the RST/LAT and NWT grade-level rubrics.

The following table details the number of items and points by session and item type for each of the PBT (grade 3) and CBT (grades 3–8) forms.

Table 3.7 Distribution of ELA Items and Points by Session and Item Type

	Sub	Gr	Session	EBSR		MS		TE		PCR		Total No. of Pts.
				No. of Items	No. of Pts.	No. of Items	No. of Pts.	No. of Items	No. of Pts.	No. of Items	No. of Pts.	
Paper - Based	ELA	3	1. Literary Analysis Task	6	12					1	15	70
			2. Research Simulation Task	4	8	2	4			1	15	
			3. Reading Literary/ Informational Texts	5	10	3	6					
Computer-Based Tests (CBT)	ELA	3	1. Literary Analysis Task	5	10			1	2	1	15	70
			2. Research Simulation Task	3	6	2	4	1	2	1	15	
			3. Reading Literary/Informational Texts	4	8	2	4	2	4			
	ELA	4	1. Research Simulation Task	6	12			2	4	1	19	83
			2. Narrative Writing Task/Reading Passage	9	18			1	2	1	12	
			3. Reading Literary/ Informational Texts	5	10	1	2	2	4			
	ELA	5	1. Literary Analysis Task/Reading Passage	4	8	2	4	4	8	1	19	86
			2. Research Simulation Task	6	12			2	4	1	19	
			3. Reading Literary/Informational Texts	5	10			1	2			
	ELA	6	1. Research Simulation Task	5	10	1	2	2	4	1	19	90
			2. Narrative Writing Task/Reading Passage	5	10	4	8	1	2	1	15	
			3. Reading Literary/ Informational Texts	5	10	2	4	3	6			
	ELA	7	1. Literary Analysis Task/Reading Passage	6	12	1	2	3	6	1	19	94
			2. Research Simulation Task	5	10	1	2	2	4	1	19	
			3. Reading Literary/ Informational Texts	5	10	3	6	2	4			
	ELA	8	1. Literary Analysis Task/Reading Passage	5	10	2	4	3	6	1	19	94
			2. Research Simulation Task	6	12			2	4	1	19	
			3. Reading Literary/ Informational Texts	7	14	1	2	2	4			

3.3 Mathematics Test Blueprints and Test Designs

The mathematics assessments were administered during a CBT testing window (April 25-May 26, 2023) or during a PBT testing window (April 27-May 2, 2023). Each test session included the four mathematics reporting categories, using the three mathematics task types (see Table 3.8).

Each item on the LEAP 2025 mathematics assessment is referred to as a task and is identified by one of three types: Type I, Type II, and Type III. As shown in the following table, each task type is aligned to one or two of four reporting categories: Major Content, Additional & Supporting Content, Expressing Mathematical Reasoning, or Modeling & Application. Each task type is designed to align with at least one of the Standards for Mathematical Practice (MP).

Table 3.8 Overview of LEAP 2025 Mathematics Task Types and Reporting Categories

Task Type	Description	Reporting Categories	Mathematical Practice(s)
Type I	Conceptual understanding, fluency, and application	*Major Content:* solve problems involving the major content for the grade level. *Additional & Supporting Content:* solve problems involving the additional and supporting content for the grade level.	Can involve any or all practices
Type II	Written arguments/ justifications, critique of reasoning, or precision in mathematical statements	*Expressing Mathematical Reasoning:* express mathematical reasoning by constructing mathematical arguments and critiques.	Primarily MP.3 and MP.6 but may also involve any of the other practices
Type III	Modeling/application in a real-world context or scenario	*Modeling & Application:* solve real-world problems engaging particularly in the modeling practice.	Primarily MP.4 but may also involve any of the other practices

These reporting categories provide parents and educators with valuable information about

- overall student performance, including readiness to continue further study in mathematics;
- student performance broken down by mathematics subcategory, which may help identify when students need additional support or more challenging work; and
- how well schools and school systems help students achieve higher expectations.

Table 3.9 provides the distribution of operational points by reporting category, by grade.

Table 3.9 Distribution of Points by Reporting Category—Mathematics

Reporting Category	Grade					
	3	**4**	**5**	**6**	**7**	**8**
Major Content	30	30	30	30	30	30
Additional & Supporting Content	10	10	10	10	10	10
Expressing Mathematical Reasoning	10	10	10	14	14	14
Modeling & Application	12	12	12	12	12	12
Total	**62**	**62**	**62**	**66**	**66**	**66**

The Major Content areas for mathematics are broken into subcategories by grade as follows:

Table 3.10 Major Content Subcategories by Grade

Grade	Major Content Subcategory
3	• Products and Quotients/Solve Multiplication and Division Problems • Solve Problems with Any Operation • Fractions as Numbers and Equivalence • Solve Time, Area, Measurement, and Estimation Problems
4	• Compare and Solve Problems with Fractions • Solve Multi-step Problems • Multiplicative Comparison and Place Value
5	• Operations with Decimals/Read, Write, and Compare Decimals • Solve Fraction Problems • Interpret Fractions, Place Value, and Scaling • Recognize, Represent, and Determine Volume/Multiply and Divide Whole Numbers
6	• Rational Numbers/Multiply and Divide Fractions • Ratio and Rate • Expressions, Inequalities, and Equations
7	• Analyze Proportional Relationships and Solve Problems • Operations with Rational Numbers • Expressions, Inequalities, and Equations
8	• Radicals, Integer Exponents, and Scientific Notation • Proportional Relationships, Linear Equations, and Functions • Solving Linear Equations/Systems of Linear Equations • Congruence and Similarity/Pythagorean Theorem

The resulting 2023 LEAP 2025 mathematics test blueprints are shown in Tables 3.11–3.16.

Table 3.11 Grade 3 Mathematics Test Blueprint

Reporting Category	Task Types						Assessable Content
	Type I		Type II		Type III		
	Tasks	Points	Tasks	Points	Tasks	Points	
Major Content	27–30	30					Louisiana Student Standards for Mathematics (LSSM): 3.OA.A.1-4, 3.OA.B.6, 3.OA.C.7, 3.OA.D.8, 3.NF.A.1-3, 3.MD.A.1-2, 3.MD.C.5-7 LEAP 2025 Evidence Statements: LEAP.I.3.1-4
Additional & Supporting Content	7–10	10					LSSM: 3.NBT.A.1-3, 3.MD.B.3-4, 3.MD.D.8, 3.MD.E.9, 3.G.A.1-2 LEAP 2025 Evidence Statements: LEAP.I.3.5-6
Expressing Mathematical Reasoning			3	10			LEAP 2025 Evidence Statements: LEAP.II.3.1-8
Modeling & Application					3	12	LEAP 2025 Evidence Statements: LEAP.III.3.1-2
TOTAL	**37**	**40**	**3**	**10**	**3**	**12**	
	TOTAL TASKS		**43**	**TOTAL POINTS**		**62**	

Table 3.12 Grade 4 Mathematics Test Blueprint

Reporting Category	Task Types						Assessable Content
	Type I		Type II		Type III		
	Tasks	Points	Tasks	Points	Tasks	Points	
Major Content	27–30	30					LSSM: 4.OA.A.1-3, 4.NBT.A.1-3 4.NBT.B.4-6, 4.NF.A.1-2, 4.NF.B.3-4, 4.NF.C.5-7 LEAP 2025 Evidence Statements: LEAP.I.4.1-8
Additional & Supporting Content	7–10	10					LSSM: 4.OA.B.4, 4.OA.C.5, 4.MD.A.1-3, 4.MD.B.4, 4.MD.C.5-7, 4.MD.D.8, 4.G.A.1-3
Expressing Mathematical Reasoning			3	10			LEAP 2025 Evidence Statements: LEAP.II.4.1-7
Modeling & Application					3	12	LEAP 2025 Evidence Statements: LEAP.III.4.1-2
TOTAL	**37**	**40**	**3**	**10**	**3**	**12**	
	TOTAL TASKS		**43**	**TOTAL POINTS**		**62**	

Table 3.13 Grade 5 Mathematics Test Blueprint

Reporting Category	Task Types						Assessable Content
	Type I		Type II		Type III		
	Tasks	Points	Tasks	Points	Tasks	Points	
Major Content	27–30	30					LSSM: 5.NBT.A.1-4, 5.NBT.B.5-7 5.NF.A.1-2, 5.NF.B.3-7 5.MD.C.3-5 LEAP 2025 Evidence Statements: LEAP.I.5.1-2
Additional & Supporting Content	7–10	10					LSSM: 5.OA.A.1-2, 5.OA.B.3 5.MD.A.1, 5.MD.B.2 5.G.A.1-2, 5.G.B.3-4
Expressing Mathematical Reasoning			3	10			LEAP 2025 Evidence Statements: LEAP.II.5.1-9
Modeling & Application					3	12	LEAP 2025 Evidence Statements: LEAP.III.5.1-2
TOTAL	**37**	**40**	**3**	**10**	**3**	**12**	
	TOTAL TASKS		**43**	**TOTAL POINTS**		**62**	

Table 3.14 Grade 6 Mathematics Test Blueprint

Reporting Category	Task Types						Assessable Content
	Type I		Type II		Type III		
	Tasks	Points	Tasks	Points	Tasks	Points	
Major Content	26–30	30					LSSM: 6.RP.A.1-3, 6.NS.A.1, 6.NS.C.5-8, 6.EE.A.1-2,4, 6.EE.B.5-8, 6.EE.C.9
Additional & Supporting Content	6–10	10					LSSM: 6.NS.B.2-4, 6.G.A.1-4, 6.SP.A.1-3, 6.SP.B.4-5
Expressing Mathematical Reasoning			4	14			LEAP 2025 Evidence Statements: LEAP.II.6.1-9
Modeling & Application					3	12	LEAP 2025 Evidence Statements: LEAP.III.6.1-3
TOTAL	**36**	**40**	**4**	**14**	**3**	**12**	
	TOTAL TASKS		**43**	**TOTAL POINTS**		**66**	

Table 3.15 Grade 7 Mathematics Test Blueprint

Reporting Category	Task Types						Assessable Content
	Type I		Type II		Type III		
	Tasks	Points	Tasks	Points	Tasks	Points	
Major Content	26–30	30					LSSM: 7.RP.A.1-3, 7.NS.A.1-3, 7.EE.A.1-2, 7.EE.B.3-4
Additional & Supporting Content	6–10	10					LSSM: 7.G.A.1-3, 7.G.B.4-6, 7.SP.A.1-2, 7.SP.B.3-4, 7.SP.C.5-8
Expressing Mathematical Reasoning			4	14			LEAP 2025 Evidence Statements: LEAP.II.7.1-7
Modeling & Application					3	12	LEAP 2025 Evidence Statements: LEAP.III.7.1-4
TOTAL	**36**	**40**	**4**	**14**	**3**	**12**	
	TOTAL TASKS		**43**	**TOTAL POINTS**		**66**	

Table 3.16 Grade 8 Mathematics Test Blueprint

Reporting Category	Task Types						Assessable Content
	Type I		Type II		Type III		
	Tasks	Points	Tasks	Points	Tasks	Points	
Major Content	25-30	30					LSSM: 8.EE.A.1-4, 8.EE.B.5-6 8.EE.C.7-8, 8.F.A.1-3 8.G.A.1-4, 8.G.B.7-8
Additional & Supporting Content	5-10	10					LSSM: 8.F.B.4-5, 8.G.C.9 8.SP.A.1-4, 8.NS.A.1-2
Expressing Mathematical Reasoning			4	14			LEAP 2025 Evidence Statements: LEAP.II.8.1-5
Modeling & Application					3	12	LEAP 2025 Evidence Statements: LEAP.III.8.1-4
TOTAL	35	40	4	14	3	12	
	TOTAL TASKS		**42**	**TOTAL POINTS**		**66**	

Unlike the ELA test blueprints, which were organized by test sessions one through three, the mathematics test blueprints were organized by reporting categories, so it was necessary to define the general structure of the test forms by test session. The design goal was to have balanced test sessions with a variety of task types and equivalent testing times. For all forms in grades 3–5, students were prohibited from using calculators, except for those students with a documented calculator accommodation. For session one of the mathematics test in grades 6–8, students are prohibited from using calculators, except those students with a documented calculator accommodation. Calculators were allowed to be used by all students in grades 6–8 in sessions two and three. The general test structures (see Tables 3.17–3.22) guided test form sequencing and design. The LEAP 2025 Calculator Policy provided the basis for calculator designation of tasks and items.

Table 3.17 General Mathematics Test Structure—Grade 3

Reporting Category	Test Session						TOTAL (Operational Only)	
	Session 1 No Calculator		Session 2 No Calculator		Session 3 No Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Major Content	9–10	10	8–10	10	10	10	27–30	30
Additional & Supporting Content	3–4	4	2–4	4	2	2	7–10	10
Expressing Mathematical Reasoning	1	4	1	3	1	3	3	10
Modeling & Application	1	3	1	3	1	6	3	12
TOTAL (Operational Only)	**15**	**21**	**14**	**20**	**14**	**21**	**43**	**62**
Test Duration (minutes)*	**75**		**85**		**75**		**235**	

**The testing time includes items that are being field tested.*

Table 3.18 General Mathematics Test Structure—Grade 4

Reporting Category	Test Session						TOTAL (Operational Only)	
	Session 1 No Calculator		Session 2 No Calculator		Session 3 No Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Major Content	9–10	10	8–10	10	10	10	27–30	30
Additional & Supporting Content	3–4	4	2–4	4	2	2	7–10	10
Expressing Mathematical Reasoning	1	4	1	3	1	3	3	10
Modeling & Application	1	3	1	3	1	6	3	12
TOTAL (Operational Only)	**15**	**21**	**14**	**20**	**14**	**21**	**43**	**62**
Test Duration (minutes)*	**75**		**85**		**75**		**235**	

**The testing time includes items that are being field tested.*

Table 3.19 General Mathematics Test Structure—Grade 5

Reporting Category	Test Session						TOTAL (Operational Only)	
	Session 1 No Calculator		Session 2 No Calculator		Session 3 No Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Major Content	9–10	10	8–10	10	10	10	27–30	30
Additional & Supporting Content	3–4	4	2–4	4	2	2	7–10	10
Expressing Mathematical Reasoning	1	4	1	3	1	3	3	10
Modeling & Application	1	3	1	3	1	6	3	12
TOTAL (Operational Only)	**15**	**21**	**14**	**20**	**14**	**21**	**43**	**62**
Test Duration (minutes)*	**75**		**85**		**75**		**235**	

**The testing time includes items that are being field tested.*

Table 3.20 General Mathematics Test Structure—Grade 6

Reporting Category	Test Session						TOTAL (Operational Only)	
	Session 1 No Calculator		Session 2 Calculator		Session 3 Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Major Content	10–12	12	6–8	8	8–10	10	26–30	30
Additional & Supporting Content	6–8	8	1–2	2	0	0	6–10	10
Expressing Mathematical Reasoning	0	0	2	7	2	7	4	14
Modeling & Application	0	0	2	9	1	3	3	12
TOTAL (Operational Only)	**16–20**	**20**	**12-13**	**26**	**11–13**	**20**	**43**	**66**
Test Duration (minutes)*	**60**		**90**		**90**		**240**	

**The testing time includes items that are being field tested.*

Table 3.21 General Mathematics Test Structure—Grade 7

Reporting Category	Test Session						TOTAL (Operational Only)	
	Session 1 No Calculator		Session 2 Calculator		Session 3 Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Major Content	16–20	20	3–5	5	3–5	5	26–30	30
Additional & Supporting Content	0	0	3–5	5	3–5	5	6–10	10
Expressing Mathematical Reasoning	0	0	2	7	2	7	4	14
Modeling & Application	0	0	2	9	1	3	3	12
TOTAL (Operational Only)	**16–20**	**20**	**12-13**	**26**	**11–13**	**20**	**43**	**66**
Test Duration (minutes)*	**60**		**90**		**90**		**240**	

**The testing time includes items that are being field tested.*

Table 3.22 General Mathematics Test Structure—Grade 8

Reporting Category	Test Session						TOTAL (Operational Only)	
	Session 1 No Calculator		Session 2 Calculator		Session 3 Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Major Content	13–18	18	3–6	6	4–6	6	25–30	30
Additional & Supporting Content	2–4	4	2–3	3	2–3	3	5–10	10
Expressing Mathematical Reasoning	0	0	2	7	2	7	4	14
Modeling & Application	0	0	2	9	1	3	3	12
TOTAL (Operational Only)	**15–20**	**22**	**10–13**	**25**	**10–12**	**19**	**42**	**66**
Test Duration (minutes)*	**60**		**90**		**90**		**240**	

**The testing time includes items that are being field tested.*

The following item types were used in the 2023 LEAP 2025 mathematics assessments:

- Multiple choice: This item type requires students to select one correct answer from four answer choices. It may appear as a one-part question, as part of a two-part question, or as a part of a constructed-response item. The multiple choice items are worth one point.
- Multiple select: This item type requires students to select more than one correct answer from more than four answer choices. It may appear as a one-part question, as part of a two-part question, or as a part of a constructed-response item. The multiple select items are worth one point. Students must choose all correct answers and no incorrect answer to receive credit.
- Short answer: This item type requires students to enter a numeric response by typing from the keyboard; it allows a decimal and numbers for grades 3–8 and a negative sign for grades 6–8. It may appear as a one-part question, as part of a two-part question, or as a part of a constructed-response item. The short answer items are worth one point. Unless specified in the question, a student will earn credit for an answer that is equivalent to the correct numerical answer and proper rounding may be required.
- Keypad input: This item type requires students to enter a mathematical response using a customized pallet of numbers, operations, variables, and/or mathematical symbols; allows all rational and irrational numbers as well as expressions and equations; and scores all equivalent responses as correct unless noted otherwise. This item type may appear as a one-part question, as part of a two-part question, or as a part of a constructed-response item.
- Constructed response: This item type requires students to respond to an open-ended question which must be typed into a response box; students may use the equation builder tool (specific to the grade or grade span) to insert mathematical characters. This item type can be a single- or multi-part item. Constructed-response items ask students to write explanations or justifications, model a process, and/or solve real-world, multi-step contextual problems. A student may receive partial or full credit on constructed-response items, and maximum point values will vary by constructed-response task. Maximum values for constructed-response items are 3, 4, or 6 points.
- Technology enhanced: This item type uses technology to capture student responses. Technology-enhanced items may appear as a one-part question, as part of a two-part question, or as a part of a constructed-response item. The technology-enhanced items are worth one point. Technology-enhanced items may involve any of the following:
    - Bar graph: requires students to complete a bar graph or histogram by raising/lowering each bar to a value
    - Drag and drop: requires students to move draggable elements into one or more drop boxes
    - Dropdown menu: requires students to select from one or more dropdown menus to complete a sentence, phrase, or expression/equation/inequality
    - Hot spot: requires students to select one or more responses by choosing selectable areas on the screen

- Match interaction table: requires students to select a checkbox in each row from two or more columns
- Graph input: requires students to enter a response on a coordinate grid
- Number line input: requires a student to enter a response on a number line
- Line plot: requires students to complete a line plot with "X" as the input

A variety of item types allows for the measurement of the full range of student performance.

The following table details the number of items by point value and task type as well as the number of points per task type for each of the PBT (grades 3 and 4) and CBT (grades 3–8) forms.

Table 3.23 Distribution of Mathematics Tasks and Points by Task Type

	Content Area	Grade	Type I			Type II			Type III			Total Points
			1 pt Tasks	2 pt Tasks	Points	3 pt Tasks	4 pt Tasks	Points	3 pt Tasks	6 pt Tasks	Points	
Paper-Pencil (PBT)	Math	3	34	3	40	2	1	10	2	1	12	62
Online (CBT)	Math	3	34	3	40	2	1	10	2	1	12	62
	Math	4	34	3	40	2	1	10	2	1	12	62
	Math	5	34	3	40	2	1	10	2	1	12	62
	Math	6	32	4	40	2	2	14	2	1	12	66
	Math	7	32	4	40	2	2	14	2	1	12	66
	Math	8	30	5	40	2	2	14	2	1	12	66

3.4 Item Development and Selection

The processes of item development and selection are discussed in this section in compliance with the *Standards*.

Standard 4.7 states the following:

> The procedures used to develop, review, and try out items and to select items from the item pool should be documented (87).

The items used in the 2023 LEAP 2025 ELA and mathematics assessments came from New Meridian's and Louisiana-owned item banks.

The items selected for use on the 2023 LEAP forms were used to equate to the LEAP 2025 scale. Operational forms were selected based on LEAP 2025 test blueprint specifications, which were supported by statistical data from New Meridian operational testing.

3.5 Considerations of Test Fairness in Item Development

Standard 3.2 is particularly relevant to fairness in item development:

> Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics (64).

Bias and sensitivity guidelines used to develop the New Meridian and Louisiana-owned items help ensure the assessments are fair for all groups of test takers, despite differences in characteristics that include, but are not limited to, disability status, ethnic group, race, gender, regional background, native language, religion, sexual orientation, and socioeconomic status. DRC relied strongly on the bias and sensitivity guidelines in the development of the assessments, particularly in item selection and review. To be included in the assessments, items had to comply with the bias and sensitivity guidelines and be approved by Louisiana educators involved in the Louisiana alignment and item review meetings.

3.6 New Meridian Item Reviews

As part of New Meridian's ongoing item development practices, several educator committees had already been convened to conduct rigorous reviews of every passage and item developed for the New Meridian assessment system prior to the items becoming a part of the item bank that included items and passages available for selection on Louisiana forms. These reviews include

- text reviews of all passages (during which participants review and edit passages independently and then discuss content and bias concerns as a grade-level group),
- item reviews (during which committees review and edit items for adherence to PARCC foundational documents, basic principles of universal design, accessibility guidelines, selected metadata fields, and a style guide),
- bias and sensitivity reviews (during which educators and community members review items and tasks to confirm the absence of issues relating to bias, fairness, and sensitivity to ensure that items and tasks do not unfairly advantage or disadvantage any student subgroup over another subgroup),
- editorial reviews (during which the review committee completes a copy edit review and records member comments), and
- data reviews (during which educators evaluate item-level statistics to determine eligibility of items and tasks to move forward to the operational assessments).

Additional information on New Meridian's item review processes and procedures can be found at the New Meridian Resource Center. Only items that have been approved by expert reviewers during text reviews (ELA only), item reviews, bias and sensitivity reviews, and editorial reviews are moved forward for field testing. Of the field tested items, only those determined to have acceptable statistics, either by having acceptable item parameters according to the data review flagging criteria or by being approved by expert reviewers during data review, are eligible for review by Louisiana educators for potential use on an operational assessment. These processes follow the criteria set forth by the *Standards*.

Standard 3.1 states the following:

> Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population (63).

Standard 3.2 states the following:

> Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics (64).

Independent studies of New Meridian passages and items have found that the content being licensed assesses the skills that matter most and is rigorous, aligned to standards, and accessible to students with disabilities and English learners. For more information on the studies performed, refer to New Meridian's website: https://resources.newmeridiancorp.org/research/.

3.7 Louisiana Item Development and Item Review

3.7.1 Mathematics Item Development

To determine the mathematics item development needs for field-testing in the Spring 2023 administration, the LDOE determined the count of items needed per grade and then DRC content experts analyzed the item pool to determine the number of type I, type II, or type III items and the evidence statements/standards based on that analysis. DRC content experts reviewed standards coverage on the previous year's test by looking at the number and types of items used to cover each content standard, the difficulty range, the level of cognitive complexity covered by each content standard, and the topic/material presented in items (to ensure a variety of engaging topics are included). DRC determined gaps or holes in coverage, based on these criteria, to create an item development plan for the number and types of items to be newly developed for possible field-testing in spring 2023. DRC presented the item development plan to LDOE content experts, who then provided feedback to DRC. DRC and the LDOE collaborated to finalize the item development plan. DRC contracted with content experts to have items written. Item writers participated in item writing training with DRC and the LDOE prior to developing items. The training included:

- an overview of the assessable content and task types,
- a description of the type I, type II, and type III items,
- an explanation of how to use the standards and evidence statements when writing items,
- examples of type I, type II, and type III items,
- a discussion that covered item writing guidelines
- examples of items with issues,
- training on security and confidentiality, and
- training on universal design and bias, fairness, and sensitivity

These items were reviewed by the LDOE and revised by DRC. Once items were approved by the LDOE, they became part of the set of items that were taken to item content and bias reviews with Louisiana educators.

At the mathematics item content and bias reviews, committees met to provide feedback on the alignment and appropriateness of items. Louisiana educators reviewed items for alignment to content standards; grade appropriateness; issues of bias, fairness, and sensitivity; and difficulty and cognitive complexity (including determining whether the difficulty and cognitive complexity were appropriate for each item and whether the items available represented a range of difficulty and cognitive complexity). For a detailed description of the process followed during the item content and bias reviews, see Appendix B. Louisiana educators edited items as needed to ensure they were appropriate for use on Louisiana assessments, which allowed the items to move forward for possible field-testing. Any items deemed inappropriate were rejected if educators were not able to revise those items. Items that successfully passed through the content and bias reviews were then placed on a test form in a field test position, and data was collected on each field test item. Once field-testing was complete, the items were taken to range-finding, where committees of Louisiana educators reviewed Louisiana student responses to assign true scores to responses that would be used in training materials for the scoring of items. The field-tested constructed response items were then scored, and the data were analyzed by DRC psychometricians.

3.8 Guidelines on Bias, Fairness, and Sensitivity

Item writers and content and bias committee members were provided with guidelines on bias, fairness, and sensitivity issues as they pertain to testing. The information included definitions of bias and sensitivity, examples of different types of bias, and topics of concern, which were specific to given content areas. Writers were also provided with sample items that contained bias, fairness, and sensitivity issues and examples of how to revise items and graphics to ensure universal design is applied. The writers were also given information on accessibility and accommodations, including information on how to address language, visual elements, and design issues when considering students in special populations (e.g., students with disabilities and English Learners).

Types of Bias:

- Stereotyping
  - may result when an image is formed by relating certain characteristics to ALL members of a group and may include physical characteristics, intellectual characteristics, emotions, careers, activities, and domestic or social roles
- Gender Bias
  - may result when people of any gender are unnecessarily presented in stereotypical activities, occupations, and/or situations or are unnecessarily presented as having stereotypical emotions or characteristics
- Regionalism
  - may result from the inclusion of terms that are not commonly used nationwide or within a particular region of the state in which the test will be given
- Ethnic or Cultural Bias
  - may result from the inclusion of terms, concepts, or situations that are demeaning and/or offensive to a particular ethnic group or culture
- Socioeconomic or Class Bias
  - may result from the inclusion of activities, possessions, or ideas that may not be common to all students

- Religious Bias
  - may result from the inclusion of terms, concepts, or situations that are demeaning and/or offensive to a particular religious group
- Ageism
  - may result from the inclusion of terms, concepts, or situations that are demeaning and/or offensive to elders or to older persons (defined as people older than the reference group) and may also involve issues of bias with other age groups, including teenagers and young children, or even with the age of the reference group itself, where the grade (age) of a student is depicted negatively
- Bias against Persons with Disabilities
  - may result from the inclusion of terms, concepts, or situations that are demeaning and/or offensive to persons with disabilities

3.8.1 Louisiana Item Alignment Review

Independent of New Meridian reviews, DRC conducts the Louisiana Item Alignment Reviews, during which Louisiana educators review items and passage sets for alignment to the Louisiana Student Standards and for appropriateness of the items and tasks for students in Louisiana, including being free of issues of bias, fairness, and sensitivity.

DRC, with guidance from the LDOE, conducted the virtual Louisiana Item Alignment Review in July 2022 with committees of Louisiana educators. Grade-level committees met for three to four days to provide feedback on the alignment and appropriateness of items that made up the New Meridian item bank. To the extent possible, each committee included educators from different parts of Louisiana, who represent all Louisiana students (e.g., special education, English learners, students with disabilities, etc.). Committee members were also representative of the diverse demographics of the state.

As described in the preceding sections, items presented at these reviews went through a rigorous review process before and after the items were field-tested by New Meridian to ensure quality and appropriateness. Items were selected for inclusion in the form selection pool, imported into IDEAS (DRC's item banking system), and formatted for use on Louisiana test forms. They were placed on mock test forms to allow them to be reviewed as students would see them. Louisiana educators reviewed these items to confirm they were acceptable for use on a Louisiana assessment. Educators reviewed items individually to verify that each item aligned to the Louisiana Student Standard(s) for that item prior to discussing the items as a group. In addition, educators reviewed item keys and discussed the difficulty and cognitive complexity of each item and task. The groups came to a consensus regarding the status of each item: Accepted with Current Alignment, Accepted with Realignment, or Rejected. Items that were accepted were determined to appropriately measure the intended standard(s) and be free of issues of bias, fairness, or sensitivity that could impact student responses to the item.

3.9 Operational Test Selection

Operational item selection for the 2023 administration took place from June through September 2022 by LDOE and DRC. The New Meridian and Louisiana item pools were used to select fixed LEAP 2025 ELA and mathematics forms.

The LEAP 2025 assessments were given in two modalities: computer-based test (CBT) or paper-based test (PBT). For both ELA and mathematics, students in grades 3 through 8 took the CBTs; some school systems elected to administer the PBTs to students in grade 3. For ELA, the dual-mode grade 3 form was identical

except for a small quantity (four to five items) of technology-enhanced items (TE) in the ELA CBT form. Items used on PBT as replacements for the TE items were evidence-based selected-response items that addressed the same content standards and were of similar rigor as the TE items, when possible. For mathematics, short-answer (SA) items were reformatted as gridded-response (GR) items for use on PBTs.

3.9.1 Item and Passage Selection Process and Criteria

The item and passage selection process used for forms construction was a content-focused, collaborative process between the LDOE and DRC ELA and mathematics content specialists, and it was followed by a psychometric evaluation of each selection. The critical psychometric consideration, other than individual item performance, was the degree to which the selected items reflected the 2023 LEAP 2025 targets. Although the item pool was limited, items that were determined to be very difficult (i.e., IRT difficulty parameter $b > 2.0$) and/or not discriminating (i.e., IRT discrimination parameter $a < 0.3$) were avoided when possible.

Selection Guidelines

- Using the pool of items, content-area assessment specialists select ELA passage sets and tasks that consist of quality texts displaying diversity in topics and authors and mathematics tasks that match the blueprint. The sets and/or tasks include items that cover a range of Louisiana Student Standards and/or Evidence Statements and address the appropriate reporting categories and subcategories.
- Content-area assessment specialists and research analysts verify that each item meets psychometric guidelines for excellence as available item-performance data allows.
- Forms include adequate content coverage, as required by the detailed test blueprint.
- Each form contains an anchor set that includes passages/items from previous operational administrations. The anchor set, which is representative of the blueprint, ensures comparability to previous forms administered since 2018 (the baseline year). The remaining sets or tasks selected for a form complete the blueprint requirements.
- No item in a form should "clue" (or provide the answer to) another item on that same form.
- Clang association should be avoided. Clang is when a distractor can be associated with, or is too similar to, a stem word, or when a statement or quote is used multiple times across items in a set.
- Passage sets in ELA forms should be diverse.
- Forms should be diverse, including a variety of text types, including texts that appeal to a diverse student population.
- Forms should include a wide range of topics and a variety of questions.
- Correct answer distributions should follow best practice (no more than 3 keys of the same answer option in a row).
- Forms **must not** contain any items that have been released to the public.

3.9.2 Review of the ELA Items and Forms

DRC and LDOE ELA content specialists and members of educator committees verified that the items were in compliance with the guidelines provided by LDOE, including alignment to the content standards and appropriateness for Louisiana students. Because establishing content validity is one of the most important

aspects in the legal defensibility of a test, the alignment of the items to the content standards must be reviewed and verified at every stage of the test development process. As a result, it is essential that an item selected for a form link directly to the content standard(s) that it purports to measure. The ELA content specialists also verified all items against their designated content codes and metadata, both to evaluate the correctness of the coding and to ensure that the given item measures what it purports to measure.

In addition, the ELA content specialists reviewed each item for item quality, ensuring that the items were in compliance with industry guidelines for clarity, style, accuracy, and appropriateness for Louisiana students. While there are many published guidelines for reviewing assessment items, the following list serves to summarize the major considerations content specialists followed when reviewing items to ensure the items conformed to item quality standards for good, reliable, and fair test questions.

Guidelines for Reviewing Items Selected for Forms

A good item should

- have only one clear, correct answer and contain answer choices that are reasonably parallel in length and structure (multiple choice);
- have only the indicated number of correct answers and contain answer choices that are reasonably parallel in length and structure (multiple select);
- have a correctly assigned content code;
- measure one main idea or standard, unless the item is a complex item, such as a prose constructed-response item (PCR);
- measure the objective or content standard (s) it is designed to measure;
- be at the appropriate level of rigor;
- be simple, direct, and free of ambiguity;
- make use of vocabulary and sentence structure that is appropriate for the grade level assessed;
- be based on content that is accurate and current;
- when appropriate, contain stimulus material that is clear and concise and provides all the information needed;
- when appropriate, contain graphics that are clearly labeled;
- contain answer choices that are plausible and reasonable in terms of the requirements of the question, as well as a student's level of knowledge;
- contain distractors that relate to the question in the same way and can be supported by a rationale;
- reflect current teaching and learning practices for the content area; and
- be free of bias and sensitivity concerns.

3.9.3 Review of the Mathematics Items and Forms

DRC and LDOE mathematics content specialists also ensured the items were in compliance with the guidelines provided by LDOE, including alignment to the content standards and appropriateness for Louisiana students. Since establishing content validity is one of the most important aspects in the legal defensibility of a

test, the alignment of the items to the content standards must be reviewed and verified at every stage of the test development process. As a result, it is essential that an item selected for a form link directly to the content standard(s) that it purports to measure. The mathematics content specialists also verified all items against their designated content codes and metadata, both to evaluate the accuracy of the coding and to ensure that the given item measures what it purports to measure.

In addition, the mathematics content specialists reviewed each item for item quality, ensuring that the test items are in compliance with industry guidelines for clarity, style, accuracy, and appropriateness for Louisiana students. While there were many published guidelines for reviewing assessment items, the list below serves to summarize the major considerations mathematics content specialists followed when reviewing items to ensure they conformed to item quality standards for good, reliable, and fair test questions.

Guidelines for Reviewing Items Selected for Forms

A good item should

- contain answer choices that are reasonably parallel in length and structure;
- have the appropriate number of correct answer(s) based on item type:
  - only one clear, correct answer for a multiple-choice (MC) item
  - only the indicated number of correct answers for a multiple select (MS) item;
- have a correctly assigned content code (item map);
- measure one content standard or evidence statement;
- measure the content standard or evidence statement it is designed to measure;
- be at the appropriate level of rigor;
- be simple, direct, and free of ambiguity;
- make use of vocabulary and sentence structure that is appropriate for the grade level assessed;
- be based on content that is accurate and current;
- when appropriate, contain stimulus material that is clear and concise and provides all the necessary information;
- when appropriate, contain graphics that are clearly labeled;
- contain answer choices that are plausible and reasonable in terms of the requirements of the question and the student's level of knowledge;
- contain distractors that relate to the question in the same way and can be supported by a rationale;
- reflect current teaching and learning practices in the content area; and
- be free of gender, ethnic, racial, cultural, socioeconomic, regional, and other forms of bias.

3.9.4 Item-Selection Options for Special Cases

While every effort is made to select a test form that meets all psychometric guidelines for excellence, it may not be possible to comply with all the psychometric criteria for item/form difficulty due to item pool limitations. In these cases, critical psychometric guidelines are followed while allowing some tolerance on

less critical item-selection guidelines. The tolerance of meeting target characteristics, the relative exposure of previously used operational items, and other considerations (e.g., content coverage) may possibly be affected in such cases.

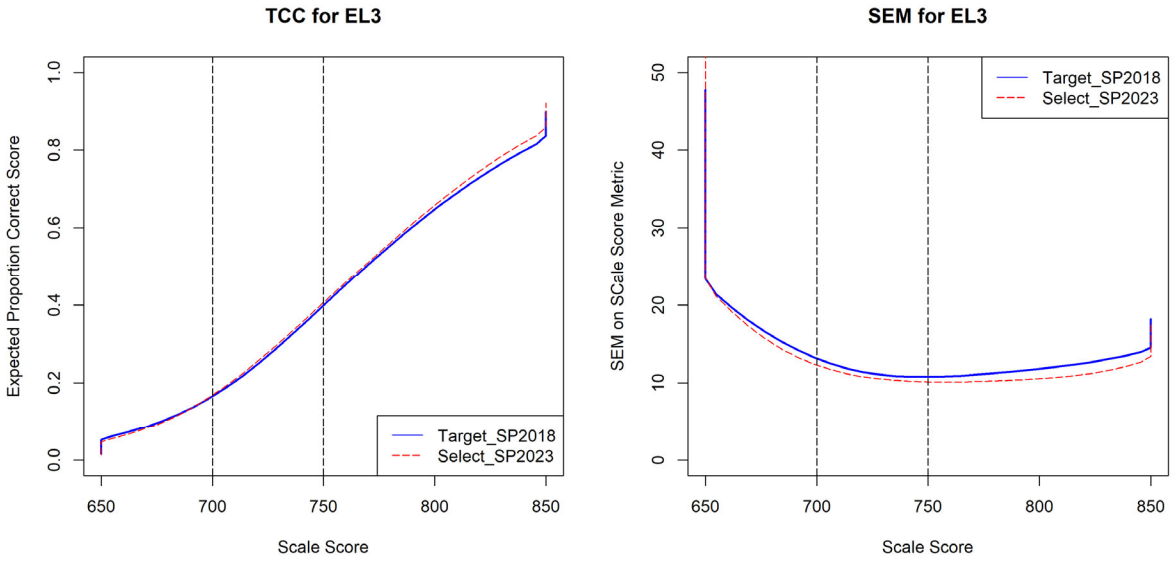
3.9.5 Psychometric Review

The psychometric evaluation of each selection was centered on reviewing the New Meridian items with operational item parameters.

Selecting Targets

The spring 2018 LEAP 2025 operational form was selected to be the target form in 2023 LEAP 2025 form construction. The rationale for the choice of the targets was that each 2018 LEAP 2025 form should be on the PARCC scale and be closely comparable to PARCC assessments. Figures 3.1 through 3.6 for ELA and Figures 3.7 through 3.12 for mathematics show the test characteristic curves (TCCs) and standard errors of measurement (SEMs) of the final forms compared to those of the target forms. The left line graph displays the TCC of the target form and the selected 2023 form, summarizing the expected proportion of the maximum raw score needed to achieve the raw score. The right line graph displays the SEM of the scale score of the target form and the selected 2023 form. This summarizes the amount of measurement error surrounding a scale score.

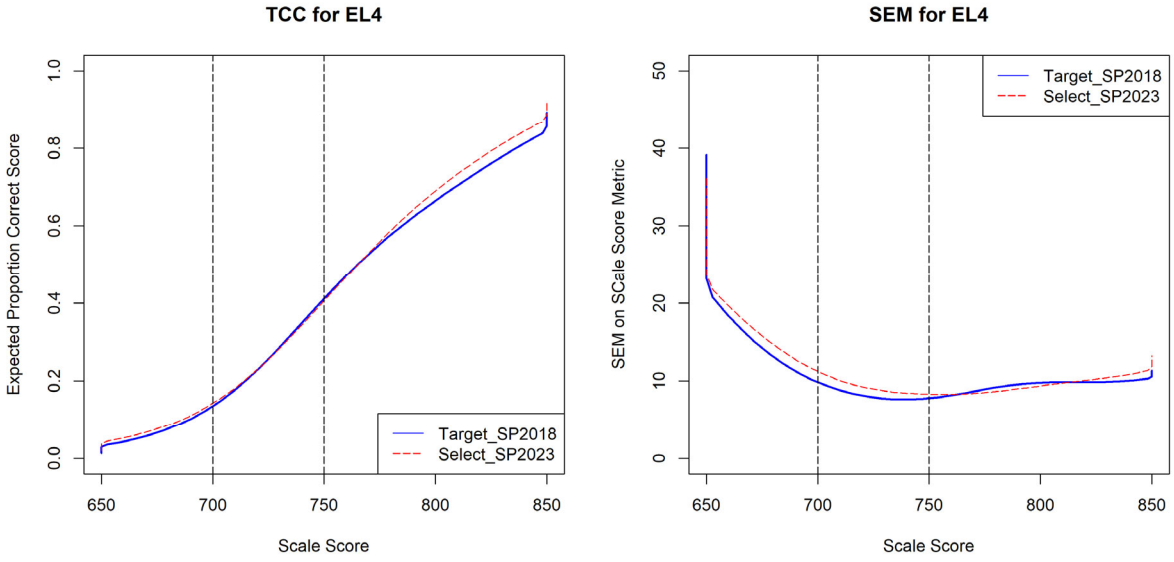
Figure 3.1 2023 ELA Form Evaluation—Grade 3



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

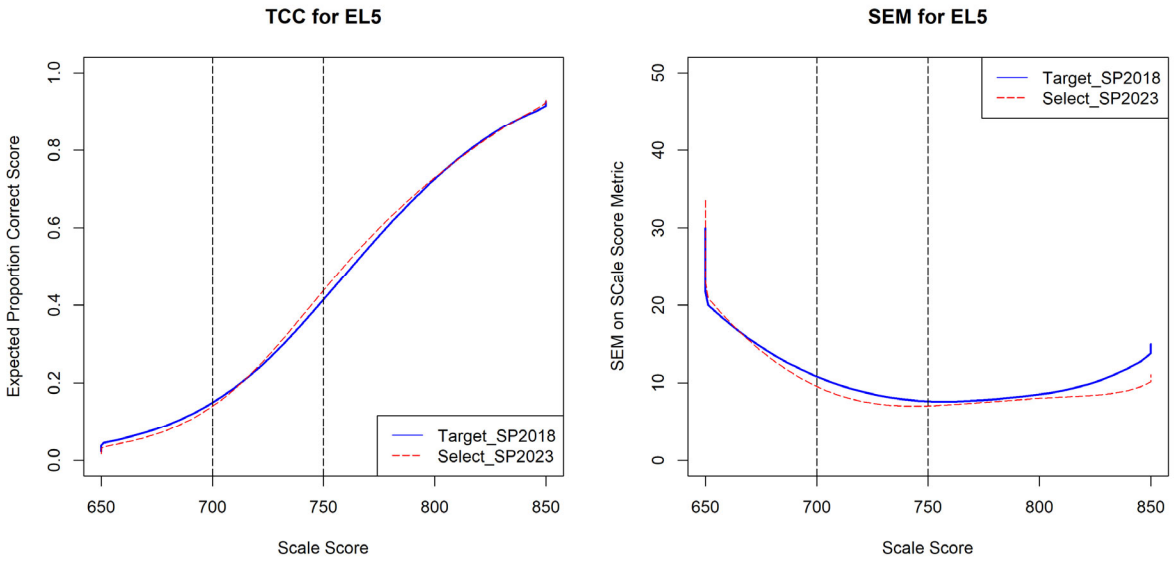
Figure 3.2 2022 ELA Form Evaluation—Grade 4



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

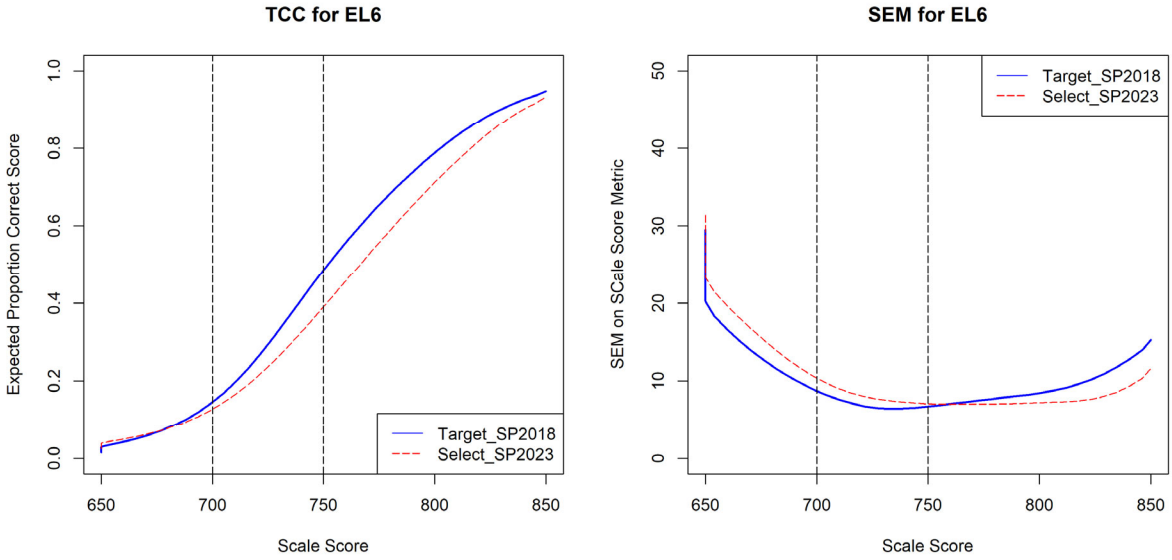
Figure 3.3 2022 ELA Form Evaluation—Grade 5



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

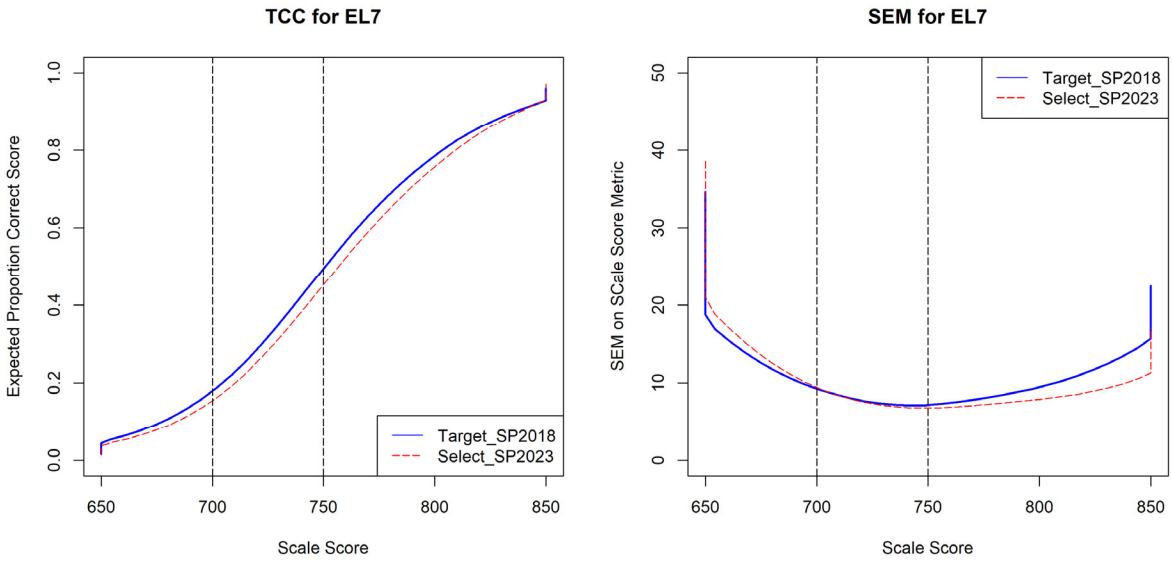
Figure 3.4 2022 ELA Form Evaluation—Grade 6



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

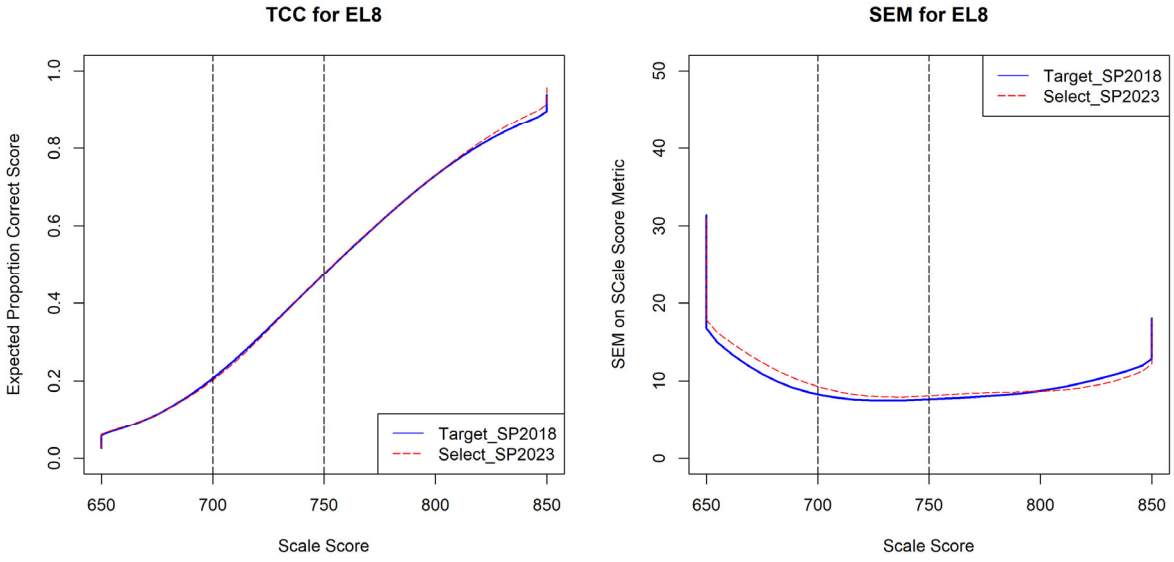
Figure 3.5 2022 ELA Form Evaluation—Grade 7



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

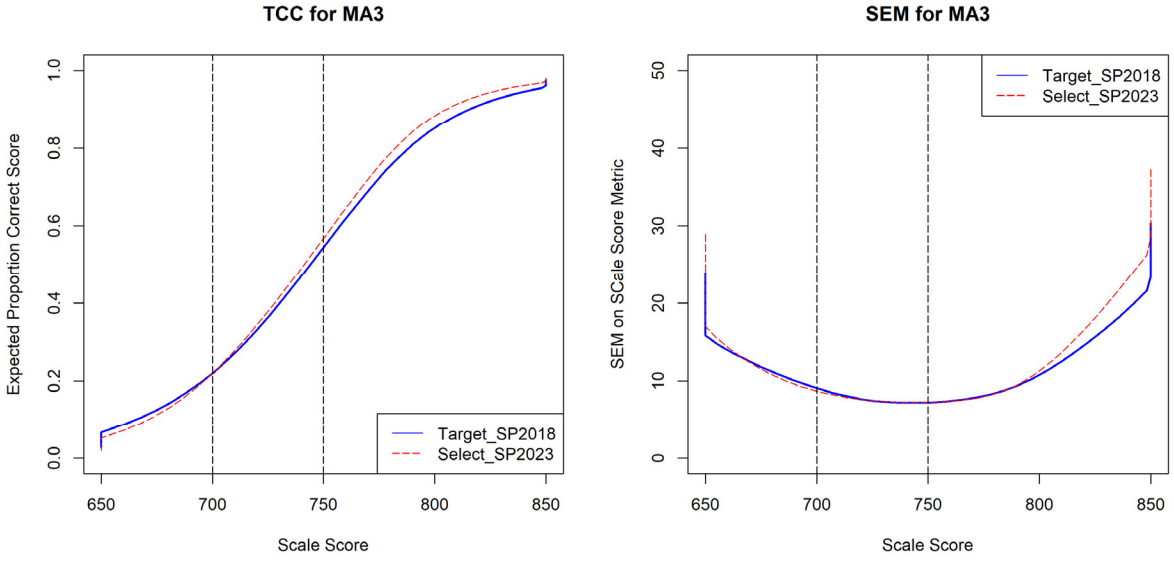
Figure 3.6 2022 ELA Form Evaluation—Grade 8



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

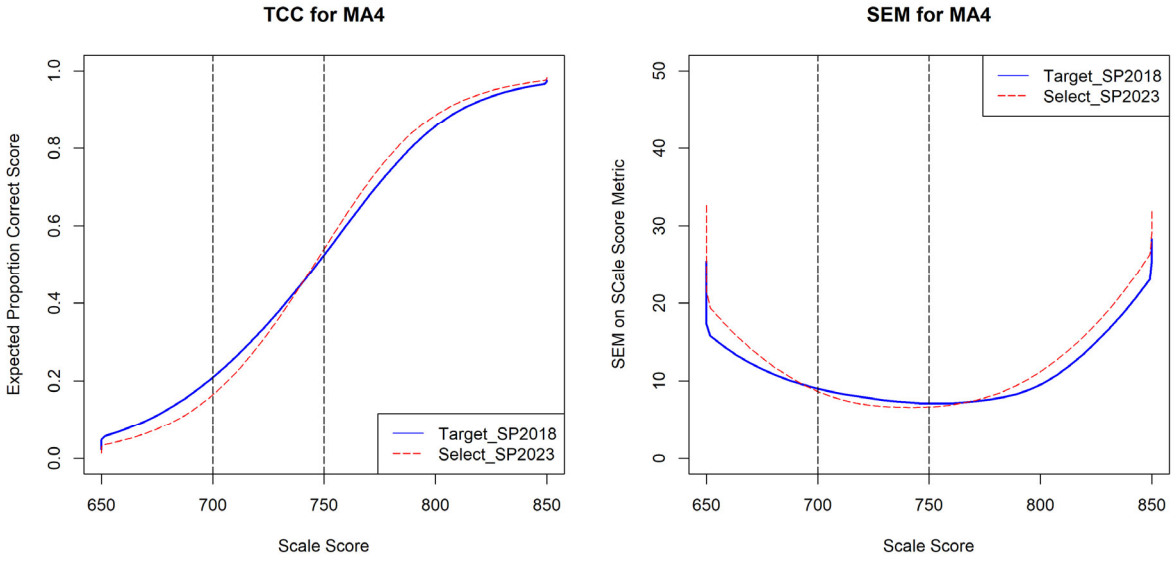
Figure 3.7 2022 Mathematics Form Evaluation—Grade 3



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

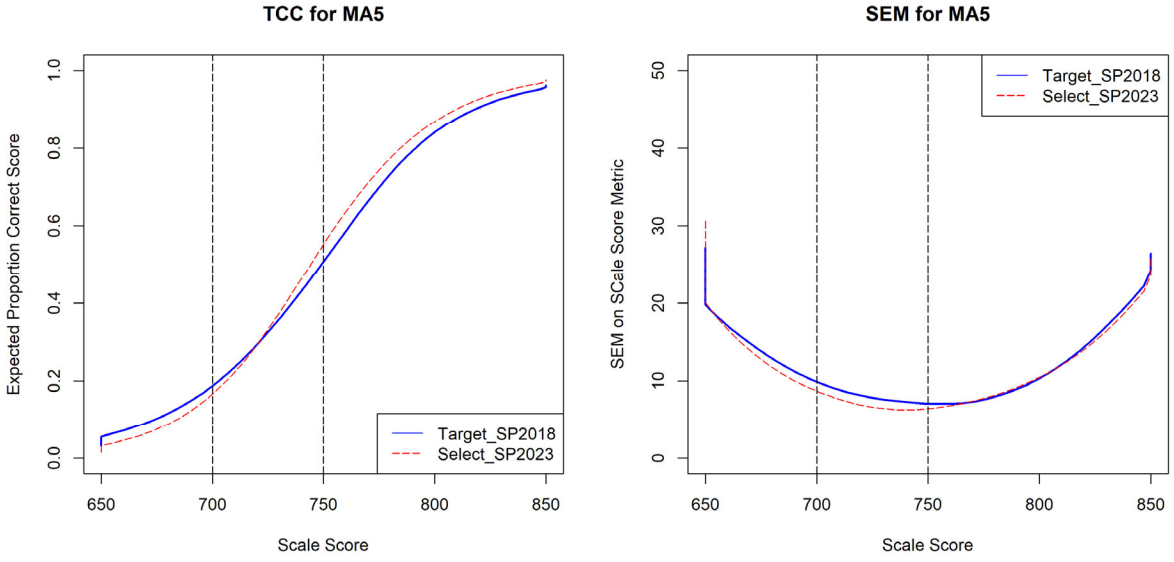
Figure 3.8 2022 Mathematics Form Evaluation—Grade 4



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

Figure 3.9 2022 Mathematics Form Evaluation—Grade 5

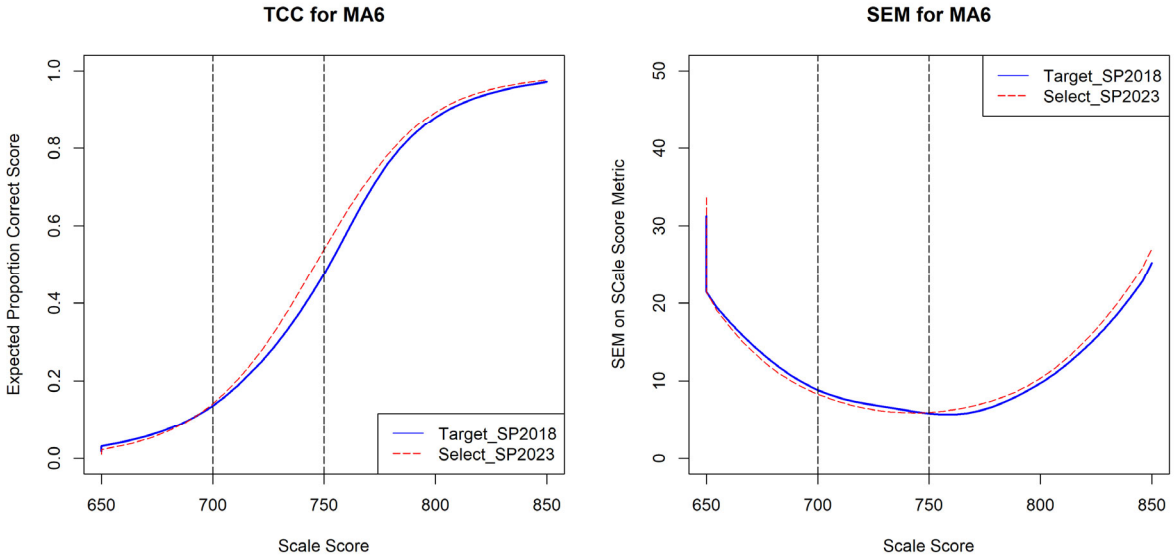


NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

.

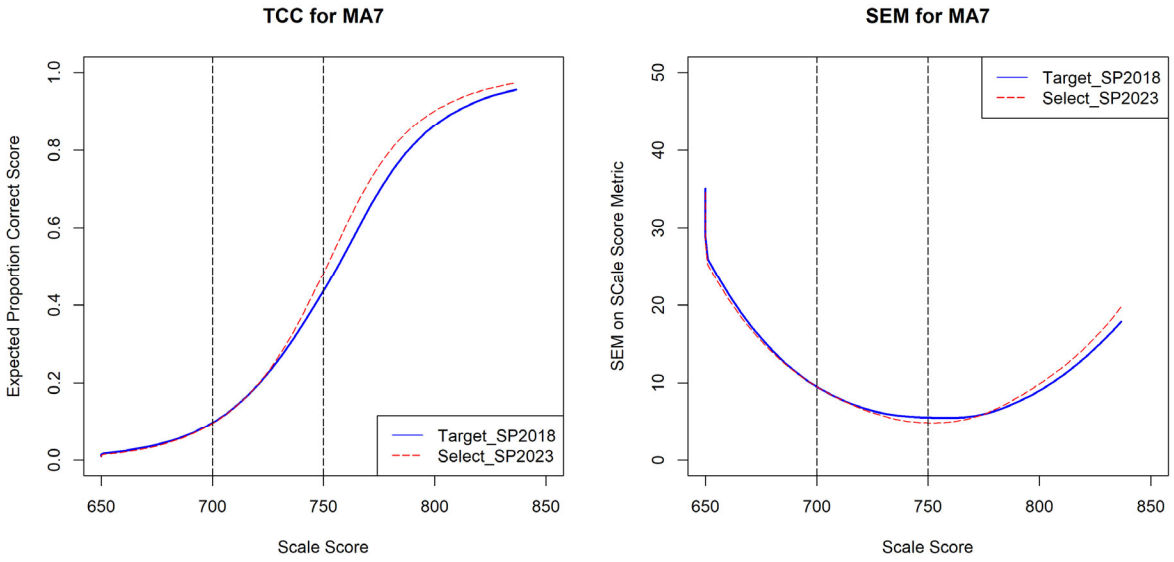
Figure 3.10 2022 Mathematics Form Evaluation—Grade 6



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

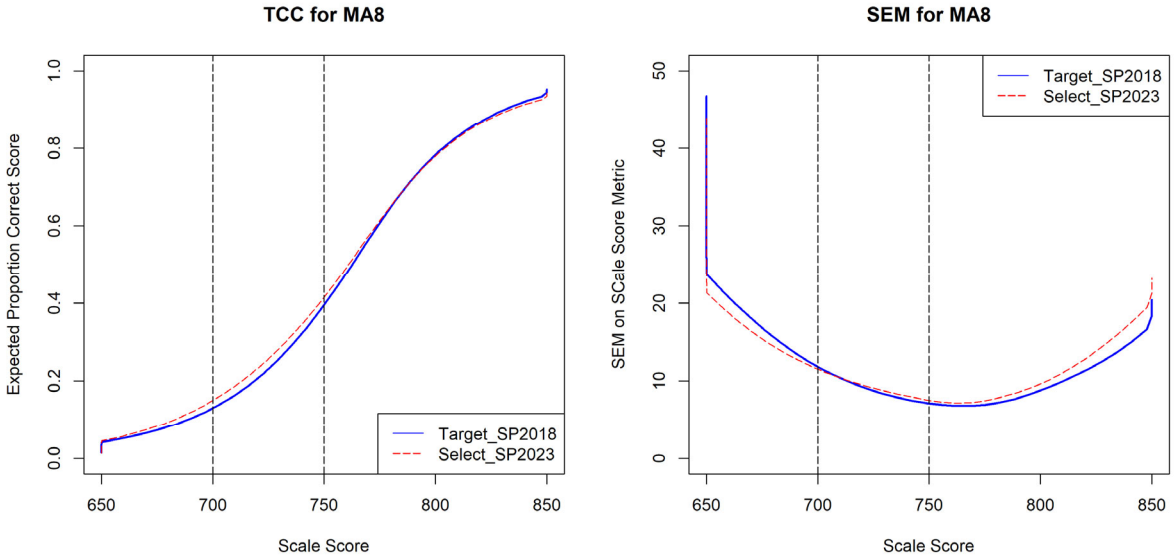
Figure 3.11 2022 Mathematics Form Evaluation—Grade 7



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

Figure 3.12 2022 Mathematics Form Evaluation—Grade 8



NOTE:

- *Target_SP2018 is the 2018 LEAP 2025 intact test form.*
- *Select_SP2023 is the selected 2023 LEAP 2025 test form.*

Selecting Anchors

Anchor sets used in the common item nonequivalent group design underwent considerable scrutiny due to the generally accepted guideline that the anchor set should mirror the total (or reference) test in terms of content and item characteristics. One of the critical psychometric considerations for an anchor set is the extent to which the TCC and SEM of the anchor set aligns to that of the total test.

3.10 Universal Design

Grade-level assessments that follow universal design guidelines allow participation of the widest possible range of students, resulting in more valid inferences about students' performances. Such assessments may reduce the need for accommodations by reducing or eliminating access barriers associated with the tests themselves. Table 3.25 presents the elements of universal design (Thompson & Thurlow, 2002). The elements of universal design are relevant to both item development and form construction. This section describes how the elements of universal design were addressed in the construction of the test forms administered in 2023 in compliance with AERA, APA, & NCME (2014) Standard 3.1, which states the following:

> Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population (63).

Universal design requires that grade-level assessments measure the performance of students with a wide range of abilities and skills, ensuring that students with diverse learning needs receive opportunities to demonstrate competence on the same content. To ensure that students can access the tests, the LEAP 2025 assessments include simple, clear, and intuitive instructions and procedures; maximum readability and comprehensibility; and maximum legibility. The online test specifications define how directions and test

items are formatted online, including the spacing between an item stem and answer choices, and other page elements (such as online tools and Help files) to ensure consistent, clean visual appearance of CBTs. Test directions at the beginning of each test session are clearly and simply stated, and the wording of such instructions is standardized as much as possible across content areas and grade levels to ensure clarity and consistency while being comparable to the requirements followed by PARCC and New Meridian.

Table 3.24 Elements of Universal Design

Element	Explanation
Inclusive Assessment Population	Tests designed for state, school system, or school accountability must include every student except those in the alternate assessment, and this is reflected in assessment design and field testing procedures.
Precisely Defined Constructs	The specific constructs tested must be clearly defined so that all construct-irrelevant cognitive, sensory, emotional, and physical barriers can be removed.
Accessible, Non-Biased Items	Accessibility is built into items from the beginning, and bias review procedures ensure that quality is retained in all items.
Amenable to Accommodations	The test design facilitates the use of needed accommodations (e.g., all items can be in braille form).
Simple, Clear, and Intuitive Instructions and Procedures	All instructions and procedures are simple, clear, and presented in understandable language.
Maximum Readability and Comprehensibility	A variety of readability and plain language guidelines are followed (e.g., sentence length and number of difficult words are kept to a minimum) to produce readable and comprehensible text.
Maximum Legibility	Characteristics that ensure easy decipherability are applied to text, tables, figures, illustrations, and response formats.

3.11 Accommodations and Designated Supports

AERA, APA, & NCME (2014) Standard 3.9 states the following:

> Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs (67).

Students with IEPs, 504 plans, and English learners (ELs) may be provided test administration accommodations as documented on their accommodation plan. More information on accommodations can be found in Section 4.3.2 of Chapter 4. Accommodation code definitions can be found in the *Paper-Based Test Administration Manual*.

Accommodated print forms were developed in grades 4–8 of ELA and mathematics for those students who were unable to participate in an online administration. For a detailed description of the process used to develop the accommodated print forms and how to modify technology-enhanced items for use in an accommodated print form, see Appendix A.

Braille and large-print test forms were constructed for each grade and content area to enable students with visual impairments to participate in the LEAP 2025 assessments. Braille and large-print forms for grade 3 of

ELA and mathematics were based on the paper-based forms. Braille forms for grades 4–8 of ELA and mathematics were based on the accommodated print forms. There are no large-print versions of the grades 4–8 accommodated print forms. Instead, students needing a large-print version in grades 4–8 use larger-sized monitors and/or the magnification features of the online testing system. All online test content has been developed to scale in relation to the available area on larger monitors while maintaining the correct aspect ratio. Specific recommendations on how to transcribe items into braille were provided by the braille publisher to produce the braille version of the LEAP 2025 assessments and the test administrator's notes that accompany the braille forms. The goal was to maximize the number of items on the braille forms that could be transcribed into braille.

The following assessment features were available to all students and do not require any documentation either prior to or during the assessment:

- blank scratch paper and graph paper
- calculators (to be used in the calculator section only)
- color overlay
- contrasting colors/reverse colors
- directions in native language
- equation builder
- bookmark
- general administration directions clarified
- general administration directions read aloud and repeated as necessary
- general masking
- headphones
- highlighters
- line guides
- magnifiers/variable zoom
- measurement tools
- redirection of student to the test
- specialized furniture or equipment
- sticky note/notepad
- strikethrough
- and writing/formatting tools (for ELA constructed response items only).

Accessibility features were available for all students with the particular need documented in their Individualized Education Programs (IEPs), Individual Accommodation Plans (IAPs), English Learner (EL) plans, or Personal Needs Profiles (PNPs). The following accessibility features were available: individual testing, small group testing, student reads assessment aloud to himself or herself, adaptive and specialized equipment or furniture, and mathematics read aloud (text-to-speech or human reader).

Accommodations were available for students who have an IEP, IAP, or EL plan, including: braille test materials, calculation device and mathematics tools for non-calculator sections of mathematics assessments, transferred answers, recorded answers, large print test materials (mathematics Spanish), mathematics Spanish read aloud, translated mathematics test, test read aloud (text-to-speech, Kurzweil, recorded audio file). For details on how these assessment and accessibility features and accommodations should be used with PBTs and CBTs, see the *LEAP 2025 Accommodations and Accessibility Features User Guide*.

For a detailed description of the process used to develop the Spanish translation forms of the mathematics tests, see Appendix B.

3.12 Item and Task Specifications

AERA, APA, & NCME (2014) Standard 4.12 states the following:

> Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications (89).

The item and task specifications are designed to ensure that the assessment items measure the assessment's claims. The purpose of the item and task specifications is to define the characteristics of the items and tasks that will provide the evidence to support one or more claims. To do this, the item and task specifications delineate the types of evidence, or targets, that should be elicited for each reporting category within a grade level. Then, the specifications provide explicit guidance on how to write items to elicit the desired evidence.

The item and task specifications provide guidance on how to measure the targets (i.e., standards) first found in the content specifications and guidelines on how to create the items that are specific to each assessment target and reporting category. In ELA and mathematics, item specifications describe the knowledge, skills, and processes being measured by each item type aligned to particular standards.

These item specifications were developed for each grade and standard to delineate the expectations of knowledge and skill to be included on test questions. In addition, the ELA and mathematics item and stimulus specifications provide guidance on determining the appropriateness of task and stimulus materials (i.e., the materials that a student must refer to when working on a test question). The stimulus specifications also provide information on the characteristics of stimuli or activities that should be avoided because they are not important to the knowledge, skill, or process being measured. This underscores DRC's efforts to select items that are accessible to the widest range of students possible; in other words, 2023 LEAP 2025 items were selected according to the elements of universal design.

3.13 Summary

In summary, the overall purpose of this chapter is to explicate the procedures used in the development of the forms administered during the spring 2023 LEAP 2025 administration. The efforts by the LDOE and DRC in developing the LEAP 2025 assessments are in alignment with multiple best practices of the test industry but, in particular, support the following AERA, APA, & NCME (2014) standards:

> ***Standard 3.1*** Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population (63).
>
> ***Standard 3.2*** Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics (64).
>
> ***Standard 3.9*** Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs (67).
>
> ***Standard 4.0*** Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide

evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population (85).

Standard 4.1 Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s) (85).

Standard 4.7 The procedures used to develop, review, and try out items and to select items from the item pool should be documented (87).

Standard 4.12 Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications (89).

Chapter 4: Test Administration

Chapter 4 of the technical report describes the processes implemented and the information disseminated to help ensure standardized test administration procedures and, thus, uniform test administration conditions for students. According to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), "The usefulness and interpretability of test scores require that a test be administered and scored according to the test developer's instructions" (111). This chapter examines how test administration procedures implemented for the 2023 Louisiana Education Assessment Program (LEAP 2025) strengthen and support the intended score interpretations and reduce construct-irrelevant variance that could threaten the validity of score interpretations.

Chapter 4 demonstrates how the LEAP 2025 assessments adhere to AERA, APA, & NCME (2014) Standards 4.15, 6.1, 6.2, 6.3, 6.4, 6.6, and 6.7. Each standard will be explicated within the relevant section of this chapter.

To ensure that the LEAP 2025 assessments are administered in accordance with the department's mandates, the LDOE takes a primary role in communicating with and training school system personnel. The development of the assessments is a collaborative effort between the LDOE and DRC. The LDOE conveys to school systems the purpose of the assessments and the importance of test administration being consistent with test industry standards. The tests and administration standards must also meet the State Board of Elementary and Secondary Education policies and the mandates of both state and federal legislation.

To accomplish these goals, the LDOE provides train-the-trainer opportunities for school system test coordinators, who, in turn, administer test-administration training to schools within their school systems. The LDOE conducts quality assurance visits during testing to ensure that school systems adhere to the standardized administration of the tests.

The district test coordinators are responsible for the schools within their school systems. They disseminate information to each school, offer assistance with test administration, and serve as liaisons between the LDOE and their school systems. The LDOE also provides assistance with and interpretation of assessment data and test results.

Ancillary materials for the LEAP 2025 test administration contribute to the body of evidence of the validity of score interpretation. This section examines how the test materials address the standards related to test administration procedures.

For the spring 2023 administration of the LEAP 2025 assessments, DRC produced the following administration manuals: *LEAP 2025 Grade 3 Paper-Based Test Administration Manual* and *LEAP 2025 Grades 3 – 8 Computer-Based Test Administration Manual* (TAMs). DRC also produced the following Test Coordinator Manuals: *LEAP 2025 Computer-Based Testing Test Coordinator Manual* and *LEAP 2025 Paper-Based Testing Test Coordinator Manual* (TCMs). LDOE assessment administration and development staff review these manuals, provide feedback, and give final approval. The TCMs include ELA, mathematics, social studies, and science in grades 3 through 8. They provide detailed instructions for district and school test coordinators' on distributing and collecting test materials and for returning them to DRC.

Paper-Based Testing *Test Coordinator Manual* Table of Contents

1. Key Dates
2. Alerts
3. Oath of Security and Confidentiality Statements

Computer-Based Testing *Test Coordinator Manual* Table of Contents

The TAMs are specific to grades, content areas, and modes of administration (i.e., online or paper). They provide detailed instructions for administering the LEAP 2025 assessments. The manuals include instructions for test security, test administrator responsibilities, test preparation, administration of tests (i.e., online or paper), and post-test procedures. Information included in the TAMs is listed below.

Paper Administration Table of Contents

Online Administration Table of Contents

18. LEAP 2025: Grades 3-8 Science (Sessions 1-2)
19. LEAP 2025: Grades 5-8 Science (Session 3 Select Schools Only
20. LEAP 2025: Grades 3-8 Social Studies (Grades 3-4 Sessions 1-2, Grades 5-8 Sessions 1-3)
21. LEAP 2025: Grades 3-4 Social Studies Session 3 Select Schools Only
22. Post-test Procedures
    22.1. Test Administrator Post-Administration Oath of Security and Confidentiality Statement
    22.2. Returning Test Materials to the School Test Coordinator
23. Index

The *Standards* contain multiple references that are relevant to test administration. Information in the TAMs addresses these standards.

The directions for test administration found in the manual address Standard 4.15, which states:

> The directions for test administration should be presented with sufficient clarity so that it is possible for others to replicate the administration conditions under which the data on reliability, validity, and (where appropriate) norms were obtained. Allowable variations in administration procedures should be clearly described. The process for reviewing requests for additional testing variations should also be documented (90).

The LEAP 2025 Test Administration Manuals provide instructions for activities conducted before, during, and after testing with sufficient detail and clarity to support reliable test administrations by qualified test administrators. To ensure uniform administration conditions throughout the state, instructions in the manuals describe the following: general rules of paper and online testing; assessment duration, timing, and sequencing information; and the materials required for testing.

Furthermore, the standardized procedures addressed in the test administration manual need to be followed, as the *Standards* state in Standard 6.1:

> Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instructions from the test user (114).

It was essential that the LEAP 2025 was administered according to the prescribed test administration manual to ensure the usefulness and interpretability of test scores and to minimize sources of construct-irrelevant variance. It should be noted that adhering to the test schedule is also a critical component. The test administration manuals include instructions for scheduling the test within the state testing window. The test administration manual also contains the schedule for timing each test session. The test timing schedule is presented in Table 4.1.

> ***Standard 6.3*** Changes or disruptions to standardized test administration procedures or scoring should be documented and reported to the test user (115).

The LDOE test administration staff reports on testing concerns that describe a wide range of improper activities that may occur during testing, including the following: copying and reviewing test questions with students; cueing students during testing, verbally or with written materials on the classroom walls; cueing students nonverbally, such as by tapping or nodding the head; using a calculator on parts of the test where it is not allowed; allowing students to correct or complete answers after tests have been submitted; splitting sessions into two parts; ignoring the standardized directions in the online assessment; reading the ELA assessment to students with the exception of those students with the read-aloud accommodation; paraphrasing parts of the test to students; changing or completing (or allowing other school personnel to change or complete) student answers; allowing accommodations that are not written in the accommodation

plan; allowing accommodations for students who do not have an accommodation plan; or defining terms on the test.

> ***Standard 6.4*** The testing environment should furnish reasonable comfort with minimal distractions to avoid construct-irrelevant variance (116).

Test administration manuals outline the steps that teachers should take to prepare classroom environment testing for administering the LEAP 2025 assessments. These steps include the following:

- Determine the layout of the classroom environment.
- Plan seating arrangements. Allow enough space between students to prevent the sharing of answers.
- Eliminate distractions such as bells or telephones.
- Use a Do Not Disturb sign on the door of the testing room.
- Make sure classroom maps, charts, and any other materials that relate to the content and processes of the test are covered, removed, or out of the students' view.

> ***Standard 6.6*** Reasonable efforts should be made to ensure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent or deceptive means (116).

The test administration manuals present instructions for post-test activities to ensure that online tests are submitted, and printed test materials are handled properly to maintain the integrity of student information and test scores. Detailed instructions guide test examiners in submitting all online test records. For students who were administered a large-print or braille test form, examiners are instructed to transcribe students' responses from the large-print test or braille test form into a consumable test booklet for grades 3 and 4, and the online testing system (INSIGHT) for grades 5 through 8, exactly as the responses appear in the original form.

> ***Standard 6.7*** Test users have the responsibility of protecting the security of test materials at all times (117).

Throughout the manuals, test coordinators and examiners are reminded of test security requirements and procedures to maintain test security. Specific actions that are direct violations of test security are so noted. Detailed information about test security procedures is presented under "Test Security" in the test administration manuals.

4.1 Return Material Forms and Guidelines

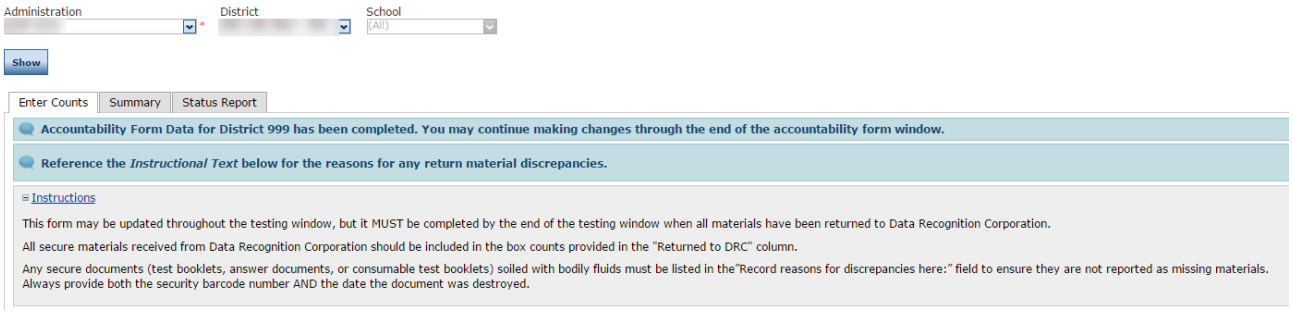
The *Test Coordinator Manual* instructs test coordinators on how to organize, pack, and return testing materials to DRC for secure inventory purposes. The LDOE assessment administration and development staff have opportunities to review these materials, provide feedback, and give final approval. The purpose of the instructions is to ensure the secure test materials are properly accounted for and organized appropriately for return shipment.

4.2 Security Checklists

As soon as printed test materials are received by a school system, the district test coordinator confirms the receipt and count of the school system materials and completes the Receipt Notice in DRC INSIGHT Portal to confirm all school system materials have been received. The district test coordinator then packages the tests to be sent to schools. Upon returning secure test materials to DRC, district test coordinators are required to complete and submit a materials accountability form that details the number of consumable test booklets or secure accommodated test materials returned. This materials accountability form also requires that school

systems document nonstandard situations, including lost, damaged, destroyed, extra, or missing test books. This form ensures all materials are accounted for. Any material not accounted for on this form is place on a missing materials list which is used by DRC and the LDOE to follow up with all districts to ensure security of all materials. A sample accountability form is shown in Figure 4.1.

Figure 4.1 Sample Accountability Form



Accountability Form for		
Science and ELA/Math Test Materials		Exact Number of Boxes Shipped to DRC
Pickup 1: UPS Ground Service (automatic pickup date)	**SCORABLE MATERIALS:**	5
	Used Science answer documents	
	Used ELA and Math consumable test booklets	
Pickup 2: UPS Ground Service (automatic pickup date)	**SCORABLE MATERIALS:**	
	Used Science makeup answer documents	
	Used ELA/Math makeup consumable test booklets	
	Used Science answer documents and ELA/Math consumable test booklets for home study program students	
	Used ELA/Math consumable test booklets for nonpublic school students	
	Accountability-coded answer documents and consumable test booklets	
	NONSCORABLE MATERIALS:	
	All unused Science answer documents	
	All unused ELA/Math consumable test booklets	
Pickup 3: Assessment Distribution Services (ADS)	**NONSCORABLE MATERIALS:**	
	All unused bar-code labels for Science and ELA/Math	
	All used and unused Science test booklets, including large print and braille	
	All ELA and Math large print and braille test booklets	

Accountability Form for		
Social Studies Test Materials		Exact Number of Boxes Shipped to DRC
Pickup 1: UPS Ground Service (automatic pickup date)	**SCORABLE AND NONSCORABLE MATERIALS:**	
	All used consumable test booklets	
	All used consumable test booklets for homestudy students	
	All unused consumable test booklets	
	All used and unused large-print and braille test booklets	

Record reasons for discrepancies here:

Enter Counts | Summary | Status Report

Instructions

Previously entered accountability form data will display. The accountability form summary information can be printed by clicking the **Print** button.

Note: The accountability form summary information is view only and cannot be edited.

Summary for District

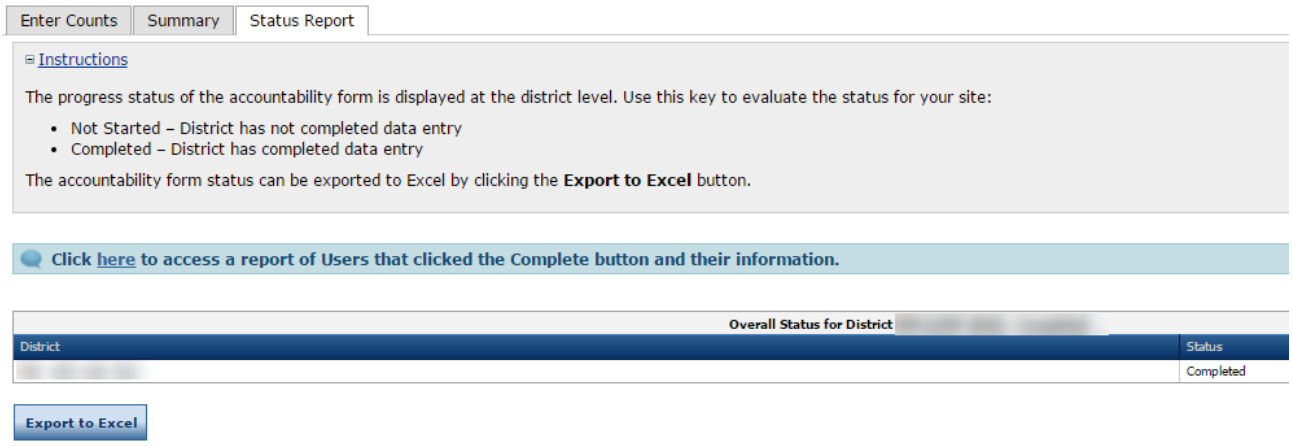
Science and ELA/Math Test Materials		Exact Number of Boxes Shipped to DRC
Pickup 1: UPS Ground Service (automatic pickup date)	**SCORABLE MATERIALS:**	5
	Used Science answer documents	
	Used ELA and Math consumable test booklets	
Pickup 2: UPS Ground Service (automatic pickup date)	**SCORABLE MATERIALS:**	
	Used Science makeup answer documents	
	Used ELA/Math makeup consumable test booklets	
	Used Science answer documents and ELA/Math consumable test booklets for home study program students	
	Used ELA/Math consumable test booklets for nonpublic school students	
	Accountability-coded answer documents and consumable test booklets	
	NONSCORABLE MATERIALS:	
	All unused Science answer documents	
	All unused ELA/Math consumable test booklets	
Pickup 3: Assessment Distribution Services (ADS)	**NONSCORABLE MATERIALS:**	
	All unused bar-code labels for Science and ELA/Math	
	All used and unused Science test booklets, including large print and braille	
	All ELA and Math large print and braille test booklets	

Summary for District

Social Studies Test Materials		Exact Number of Boxes Shipped to DRC
Pickup 1: UPS Ground Service (automatic pickup date)	**SCORABLE AND NONSCORABLE MATERIALS:**	
	All used consumable test booklets	
	All used consumable test booklets for homestudy students	
	All unused consumable test booklets	
	All used and unused large-print and braille test booklets	

Record reasons for discrepancies here:

Print



4.3 Interpretive Guides

An understanding of what test scores mean and how to interpret score reports is essential to making valid interpretations of the test scores. The *Interpretive Guide* is written for Louisiana teachers and administrators who receive the LEAP 2025 score reports. More details about the guide can be found in Chapter 7.

4.4 Test Security Measures

Maintaining the security of all test materials is crucial to preventing the possibility of random or systematic errors, such as unauthorized exposure of test items that would affect the valid interpretation of test scores. Several test security measures are implemented for the LEAP 2025 assessments. Test security procedures are discussed throughout the Test Coordinator Manuals and Test Administration Manuals.

Test coordinators and administrators are instructed to keep all test materials in locked storage, except during actual test administration, and access to secure materials must be restricted to authorized individuals only (e.g., test administrators and the school test coordinator). During testing sessions, the test administrators are directly responsible for the security of the LEAP 2025 assessments and must account for all test materials and supervise the test administration at all times.

4.5 Data Forensic Analyses

Due to the importance of the LEAP 2025 assessment, it is prudent to ensure that the results from the assessments are based on effective instruction and true student achievement. While there are many ways to achieve meaningful understanding of student knowledge via test scores, there are also ways to obtain higher test scores that are not related to actual learning. To assist ensuring that assessment results are valid, data forensic analyses are conducted to help separate meaningful gains from spurious gains. It is important to note that although the results may be used to identify potential problems within a school, the identification of a problem is not an accusation of misconduct.

Multiple methods were incorporated into the forensic analysis. The following methods were applied:

- Response Change Analysis
- Score Fluctuation Analysis
- Web Monitoring
- Plagiarism Detection

4.5.1 Response Change Analysis

Students make changes to answer choices when taking the LEAP 2025, and this is expected behavior. Unfortunately, changing student answers is also an opportunity for school personnel to improve classroom performance and, therefore, the response change analysis focuses on identifying school- and test-administrator level response-change patterns that are statistically improbable when compared to the expected pattern at the state level.

4.5.2 Score Fluctuation Analysis

It is anticipated that performance on the LEAP 2025 will improve over time from legitimate sources such as changes in the curriculum and improvement in instruction. However, large and unexpected score changes may be a sign of testing impropriety. The LDOE applied an approach where the state's level of change in performance from one year to the next is compared to a schools' and test administrators' change in performance during the same time frame. Schools and test administrators were identified when the level of change was statistically unexpected.

4.5.3 Web Monitoring

LEAP 2025 operational test content should not appear outside the boundaries of the forms administered. To protect Louisiana test content, the internet is monitored for postings which contain, or appear to contain, potentially exposed and/or copied LDOE test content. When test content is verified, steps are taken so that the infringing content is removed quickly.

4.5.4 Plagiarism Detection

The LDOE monitors for two different plagiarism situations: copying from student to student and copying from an outside source, such as Wikipedia or another internet sources. Instances of plagiarism are identified regardless if an item is scored by human scorers or artificial intelligence. Alerts are set to identify responses that may indicate the possibility of teacher interference, plagiarism, or disturbing content (e.g., possible physical or emotional abuse, suicidal ideation, threats of harm to themselves or others, etc.). Alerted responses are given additional review so the appropriate response can be taken.

4.6 Test Administration

The 2023 assessments were administered to students within the state testing window of April 25 through May 26, 2023. The paper testing window was April 26 through May 2, 2023. Each session of the assessment within each content area of the LEAP 2025 assessments was required to be administered in one block of time.

4.6.1 Time

All sessions of the ELA and mathematics LEAP 2025 assessments were timed. Only students with an extended time accommodation were permitted to exceed the established time limits of any given session. The timing schedule of the LEAP 2025 assessments is presented in Table 4.1.

Table 4.1 LEAP 2025 Administration Schedule Timing Guidelines by Session (Time in Minutes)

Grade	Session	English Language Arts	Mathematics
3	1	75	75
	2	75	85
	3	60	75
4	1	90	75
	2	90	85
	3	60	75
5	1	90	75
	2	90	85
	3	60	75
6	1	90	60
	2	90	90
	3	80	90
7	1	90	60
	2	90	90
	3	80	90
8	1	90	60
	2	90	90
	3	80	90

For the CBT administrations, data is available of how much time test takers took for each item. These time-on-items were summed and average time on test were calculated for each grade and subject and summarized in Table 4.2 (ELA) and Table 4.3 (Mathematics). The tables report the at the session level and summarize the number of students included in this analysis, the average number of items the students were administered (operational and field test), the average amount of minutes spent across all items, and the standard deviation. There are extreme test times on both ends (some are very small, and some are very large), therefore, the median is included as it is less influenced by these extremes. In this circumstance, it is a more useful description of expected values than the mean. The test times are smaller than the session-level time guidelines in Table 4.1. This indicates that test takers should have sufficient time to complete their tests.

Table 4.2 LEAP 2025 Time on Test for the Spring 2023 Administration (Time in Minutes): ELA

ELA Grade	Session	Number of Students	Number of Items	Test Mean	Test SD	Median
3	1	≥21,610	6.99	48.36	20.39	46.66
	2	≥21,750	6.99	42.00	19.42	39.79
	3	≥21,730	13.95	37.51	32.15	36.41
4	1	≥47,530	8.99	58.23	21.13	56.97
	2	≥47,580	10.99	43.44	18.25	40.64
	3	≥48,060	13.98	34.94	12.77	33.33
5	1	≥46,920	10.97	63.48	20.74	63.27
	2	≥46,950	8.99	53.16	20.75	51.80
	3	≥47,480	11.99	32.39	12.11	30.54
6	1	≥41,940	8.99	62.28	22.06	62.23
	2	≥42,240	10.99	46.90	17.79	44.79
	3	≥42,520	15.97	41.49	14.96	39.90
7	1	≥40,830	10.98	56.66	19.58	55.25
	2	≥41,390	8.99	57.85	21.67	57.68
	3	≥41,530	15.97	43.98	15.98	42.66
8	1	≥44,380	10.98	65.16	20.96	65.53
	2	≥44,890	8.99	55.72	20.99	55.06
	3	≥45,050	15.98	42.45	15.17	40.86

Table 4.3 LEAP 2025 Time on Test for the Spring 2023 Administration (Time in Minutes): Mathematics

Mathematics Grade	Session	Number of Students	Number of Items	Test Mean	Test SD	Median
3	1	≥21,860	17.97	42.01	17.53	39.61
	2	≥21,850	14.98	44.61	18.75	42.69
	3	≥21,850	15.98	39.02	16.08	37.04
4	1	≥47,710	17.96	49.96	16.98	48.72
	2	≥47,770	14.98	46.94	17.61	44.89
	3	≥47,850	15.98	42.92	20.59	40.92
5	1	≥47,330	17.97	51.07	16.12	50.15
	2	≥47,300	14.98	50.53	18.85	48.94
	3	≥47,290	15.98	47.87	21.58	46.81
6	1	≥47,600	20.93	41.65	13.11	40.79
	2	≥47,400	14.98	56.91	19.90	55.75
	3	≥47,480	12.99	42.16	17.58	40.03
7	1	≥48,130	20.93	42.08	13.45	41.56
	2	≥47,920	14.98	62.90	20.69	63.40
	3	≥48,100	12.98	45.54	17.09	43.96
8	1	≥43,660	21.95	40.78	13.15	40.11
	2	≥43,510	13.98	57.93	19.95	57.46
	3	≥43,380	11.99	51.54	19.85	50.64

4.6.2 Accommodations

Accommodations are allowed on the LEAP 2025 assessments. Accommodations may be used by a student who qualifies under the Individual with Disabilities Act (IDEA), has an IEP or a Section 504 plan of the Americans with Disabilities Act, or identifies as an English learner (EL). Accommodations must be specified in the qualifying student's individual plan and must be consistent with accommodations used during daily classroom instruction and testing. The use of any accommodation must be indicated on the student information sheet at the time of test administration. AERA, APA, & NCME Standard 6.2 states:

> When formal procedures have been established for requesting and receiving accommodations, test takers should be informed of these procedures in advance of testing (115).

In compliance with this standard, the LEAP 2025 *Test Administration Manual* contains the list of universal tools, designated supports, and accommodations permissible for the LEAP 2025 assessments. Further guidance can be found in the LEAP 2025 *Accommodations and Accessibility Features User Guide*.

Visually impaired students may be provided braille forms for any assessment and large print forms for the PBT.

Tables 4.2 through 4.5 summarize the numbers of reportable students receiving accommodations by accommodation type for the 2023 LEAP 2025. Accommodation assignment guidance is provided in the LEAP 2025 Accommodations and Accessibility User Guide. Accommodations are grouped into four sections: special education accommodation, English learner status accommodation, Section 504 status accommodation, and online accommodation. The analyses are based on census data and the number includes only those students

who were eligible to use an accommodations and received a scale score on the ELA or mathematics LEAP 2025 assessments. The percentage represents the percentage of the census population receiving that accommodation. The students who are included in the "No Accommodation" category are students who are eligible for an accommodation but have indicated that none was used.

Table 4.2 Number and Percentage of Students Receiving Special Education Accommodations by Accommodation Type, as Bubbled on the Test Booklet

Special Education Accommodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Number	Percentage	Number	Percentage
3	No Accommodation	≥1,210	4.46%	≥1,190	4.41%
3	Braille	<50	NR	<50	NR
3	Large Print	<50	NR	<50	NR
3	Answers Recorded	≥380	1.43%	≥380	1.42%
3	Extended Time	≥2,600	9.57%	≥2,630	9.70%
3	Transferred Answers	≥100	0.37%	≥100	0.37%
3	Individual/Small Group Administration	≥2,470	9.09%	≥2,460	9.07%
3	Tests Read Aloud	≥1,710	6.30%	≥1,930	7.10%
3	Calculator	-	-	≥1,180	4.35

Table 4.3 Number and Percentage of Students Receiving English Learner Accommodations by Accommodation Type, as Bubbled on the Test Booklet

English Learner Accommodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Number	Percentage	Number	Percentage
3	No Accommodation	≥80	0.32%	≥80	0.32%
3	Extended Time	≥730	2.72%	≥690	2.54%
3	Individual/Small Group Administration	≥460	1.70%	≥400	1.50%
3	English/Native Language Word-to-Word Dictionary	≥140	0.53%	≥130	0.50%
3	Test Administered by ESL Teacher	<50	NR	<50	NR
3	Directions Read Aloud/Clarified in Native Language	<50	NR	<50	NR
3	Spanish Test	-	-	<50	NR

Table 4.4 Number and Percentage of Students Receiving Section 504 Status by Accommodation Type, as Bubbled on the Test Booklet

Section 504 Status Accommodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Number	Percentage	Number	Percentage
3	No Accommodation	≥160	0.61%	≥160	0.59%
3	Large Print	<50	NR	<50	NR
3	Answers Recorded	≥60	0.23%	≥60	0.23%
3	Extended Time	≥1,720	6.32%	≥1,720	6.35%
3	Transferred Answers	<50	NR	<50	NR
3	Individual/Small Group Administration	≥1,390	5.14%	≥1,390	5.14%
3	Tests Read Aloud	≥420	1.54%	≥650	2.41%
3	Calculator	-	-	≥180	0.67%

Table 4.5 Number and Percentage of Students Receiving Online Accommodations by Accommodation Type, as valued in DRC INSIGHT Portal

Online Accommodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Number	Percentage	Number	Percentage
3	Text-to-Speech	≥2,510	11.36%	≥5,010	22.62%
3	Human Read Aloud	≥90	0.44%	≥140	0.67%
3	Native Language Word-to-Word Dictionary	≥330	1.52%	≥340	1.55%
3	Directions in Native Language	≥210	0.97%	≥210	0.97%
3	Transferred Answers	≥120	0.58%	≥130	0.59%
3	Answers Recorded	≥420	1.93%	≥430	1.94%
3	Extended Time	≥5,880	26.60%	≥5,930	26.78%
3	Individual/Small Group Administration	≥4,420	20.01%	≥4,480	20.20%
3	Accommodated Paper	<50	NR	<50	NR
3	Braille	<50	NR	<50	NR
3	Communication Assistance Scripts	<50	NR	<50	NR
3	Calculator	-	-	≥1,620	7.31%
3	Basic Calculator	-	-	≥1,680	7.58%
3	Scientific Calculator	-	-	<50	NR
3	Spanish Test	-	-	≥70	0.35%
4	Text-to-Speech	≥4,990	10.22%	≥8,440	17.27%
4	Human Read Aloud	≥290	0.60%	≥360	0.74%
4	Native Language Word-to-Word Dictionary	≥550	1.13%	≥550	1.14%
4	Directions in Native Language	≥230	0.48%	≥230	0.48%
4	Transferred Answers	≥220	0.45%	≥220	0.45%
4	Answers Recorded	≥790	1.62%	≥790	1.62%
4	Extended Time	≥11,620	23.79%	≥11,660	23.87%
4	Individual/Small Group Administration	≥9,140	18.70%	≥9,170	18.78%
4	Accommodated Paper	<50	NR	<50	NR
4	Braille	<50	NR	<50	NR
4	Communication Assistance Scripts	<50	NR	<50	NR
4	Calculator	-	-	≥3,880	7.96%
4	Basic Calculator	-	-	≥4,020	8.23%
4	Scientific Calculator	-	-	<50	NR
4	Spanish Test	-	-	≥80	0.18%

Online Accommodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Number	Percentage	Number	Percentage
5	Text-to-Speech	≥4,980	10.31%	≥8,220	17.05%
5	Human Read Aloud	≥310	0.65%	≥380	0.80%
5	Native Language Word-to-Word Dictionary	≥560	1.16%	≥560	1.16%
5	Directions in Native Language	≥180	0.38%	≥180	0.38%
5	Transferred Answers	≥240	0.51%	≥240	0.51%
5	Answers Recorded	≥690	1.44%	≥690	1.43%
5	Extended Time	≥11,470	23.74%	≥11,490	23.82%
5	Individual/Small Group Administration	≥8,980	18.59%	≥8,990	18.62%
5	Accommodated Paper	<50	NR	<50	NR
5	Braille	<50	NR	<50	NR
5	Communication Assistance Scripts	<50	NR	<50	NR
5	Calculator	-	-	≥4,580	9.51%
5	Basic Calculator	-	-	≥4,690	9.72%
5	Scientific Calculator	-	-	<50	NR
5	Spanish Test	-	-	≥90	0.20%
6	Text-to-Speech	≥4,220	9.75%	≥7,450	15.42%
6	Human Read Aloud	≥220	0.51%	≥280	0.58%
6	Native Language Word-to-Word Dictionary	≥820	1.89%	≥860	1.79%
6	Directions in Native Language	≥210	0.50%	≥220	0.46%
6	Transferred Answers	≥120	0.29%	≥150	0.32%
6	Answers Recorded	≥320	0.75%	≥360	0.74%
6	Extended Time	≥10,250	23.63%	≥11,320	23.43%
6	Individual/Small Group Administration	≥7,010	16.18%	≥7,850	16.24%
6	Accommodated Paper	<50	NR	<50	NR
6	Braille	<50	NR	<50	NR
6	Communication Assistance Scripts	<50	NR	<50	NR
6	Calculator	-	-	≥4,910	10.16%
6	Basic Calculator	-	-	≥5,070	10.50%
6	Scientific Calculator	-	-	<50	NR
6	Spanish Test	-	-	≥100	0.22%

Online Accommodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Number	Percentage	Number	Percentage
7	Text-to-Speech	≥4,070	9.59%	≥7,530	15.39%
7	Human Read Aloud	≥220	0.53%	≥280	0.59%
7	Native Language Word-to-Word Dictionary	≥960	2.28%	≥1,020	2.10%
7	Directions in Native Language	≥190	0.46%	≥200	0.43%
7	Transferred Answers	≥80	0.20%	≥90	0.20%
7	Answers Recorded	≥190	0.45%	≥220	0.47%
7	Extended Time	≥10,100	23.78%	≥11,540	23.59%
7	Individual/Small Group Administration	≥6,570	15.49%	≥7,490	15.31%
7	Accommodated Paper	<50	NR	<50	NR
7	Braille	<50	NR	<50	NR
7	Communication Assistance Scripts	<50	NR	<50	NR
7	Calculator	-	-	≥5,190	10.62%
7	Basic Calculator	-	-	≥5,280	10.79%
7	Scientific Calculator	-	-	<50	NR
7	Spanish Test	-	-	≥120	0.26%
8	Text-to-Speech	≥4,200	9.18%	≥7,370	16.67%
8	Human Read Aloud	≥170	0.38%	≥260	0.59%
8	Native Language Word-to-Word Dictionary	≥1,000	2.19%	≥1,000	2.28%
8	Directions in Native Language	≥220	0.49%	≥210	0.49%
8	Transferred Answers	≥80	0.19%	≥100	0.23%
8	Answers Recorded	≥190	0.42%	≥200	0.47%
8	Extended Time	≥10,690	23.36%	≥11,400	25.78%
8	Individual/Small Group Administration	≥6,800	14.85%	≥7,300	16.52%
8	Accommodated Paper	<50	NR	<50	NR
8	Braille	<50	NR	<50	NR
8	Communication Assistance Scripts	<50	NR	<50	NR
8	Calculator	-	-	≥5,400	12.22%
8	Basic Calculator	-	-	<50	NR
8	Scientific Calculator	-	-	≥5,390	12.19%
8	Spanish Test	-	-	≥120	0.27%

4.7 Summary

In summary, the overall purpose of each of the test administration trainings and the ancillary materials is to keep school systems informed about policies and procedures related to testing in general and the LEAP 2025 program in particular. The information imparted is clearly related to standardizing the administration of the LEAP 2025, maintaining the security of the assessment, allowing access to the assessments for special populations by clearly delineating appropriate accommodations, and maintaining integrity of the scores. These communication and training efforts by the LDOE and the ancillary information developed by DRC address multiple best practices of the testing industry but, in particular, are related to the following standards:

> ***Standard 4.15*** The directions for test administration should be presented with sufficient clarity so that it is possible for others to replicate the administration conditions under which the data on reliability, validity, and (where appropriate) norms were obtained. Allowable variations in administration procedures should be clearly described. The process for reviewing requests for additional testing variations should also be documented (90).
>
> ***Standard 6.1*** Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instructions from the test user (114).
>
> ***Standard 6.3*** Changes or disruptions to standardized test administration procedures or scoring should be documented and reported to the test user (115).
>
> ***Standard 6.4*** The testing environment should furnish reasonable comfort with minimal distractions to avoid construct-irrelevant variance (116).
>
> ***Standard 6.6*** Reasonable efforts should be made to ensure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent or deceptive means (116).
>
> ***Standard 6.7*** Test users have the responsibility of protecting the security of test materials at all times (117).

Chapter 5: Scoring of Constructed-Response and Technology-Enhanced Items

In this chapter, the scoring process used for the 2023 LEAP 2025 ELA and mathematics assessment is described, with a particular focus on the handscoring of constructed-response items and the automated scoring of technology-enhanced items. At the end of this section, the results of the inter-rater reliability for the handscoring of the LEAP 2025 constructed-response items are presented.

Chapter 5 adheres to the American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME, 2014) Standards 4.18, 4.20, 6.8, and 6.9. Each standard is presented in the pertinent section of this chapter. Standard 4.18 provides some general guidance for Chapter 5:

> Procedures for scoring and, if relevant, scoring criteria, should be presented by the test developer with sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical for extended-response items such as performance tasks, portfolios, and essays (91).

Chapter 5 explains the procedures used for scoring the LEAP 2025 ELA and mathematics constructed-response items and technology-enhanced items. The scoring criteria used for each item are not presented in this chapter to preserve the integrity of the items for future use.

5.1 Constructed-Response Item Scoring Process

Constructed-response items were scored by human raters who were trained by DRC. Handscoring and Artificial Intelligence (AI) processing rules are detailed in Appendix C. Eleven ELA items across grades 3-8 ELA (noted in the table below) were scored by an AI engine, Pearson's Intelligent Essay Assessor (IEA), using scoring models previously developed by Pearson. Second reads of 10% of these responses were completed by human scorers; handscoring supervisors also reviewed the responses that IEA was not able to score.

Table 5.1 Constructed-Response Scoring

Subject and Grade	Handscoring Only	AI Scoring	AI Vendor
ELA grade 3	Q14	Q7	Pearson
ELA grade 4	N/A	Q9, Q14	Pearson
ELA grade 5	N/A	Q7, Q20	Pearson
ELA grade 6	N/A	Q9, Q14	Pearson
ELA grade 7	N/A	Q7, Q20	Pearson
ELA grade 8	N/A	Q7, Q20	Pearson
Mathematics grades 3-8	All CRs	N/A	

5.1.1 Selection of Scoring Evaluators

Standard 4.20 states the following:

> The process for selecting, training, qualifying, and monitoring scorers should be specified by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the rubric score scale, and the procedures for training scorers should result in a degree of accuracy and agreement among scorers that allows the scores to be interpreted as originally intended by the test developer. Specifications should also describe processes for assessing scorer consistency and potential drift over time in raters' scoring (92).

The following sections explain how scorers were selected and trained for the LEAP 2025 ELA and mathematics handscoring process. Section 5.1.3 describes how the scorers were monitored throughout the handscoring process.

The Recruitment and Interview Process

DRC strives to develop a highly qualified, experienced core of evaluators to appropriately maintain the integrity of all projects.

All readers hired by DRC to score 2023 LEAP 2025 ELA and mathematics test responses had at least a four-year college degree. DRC has a human resources director dedicated solely to recruiting and retaining the handscoring staff. Applications for reader positions are screened by the handscoring project manager, the human resources director, or recruiting staff to create a large pool of potential readers. In the screening process, preference is given to candidates with previous experience scoring large-scale assessments and with degrees emphasizing the appropriate content areas. At the personal interview, reader candidates are asked to demonstrate their proficiency in writing by responding to a DRC writing topic and their proficiency in mathematics by solving word problems with correct work shown. These steps result in a highly qualified and diverse workforce. DRC personnel files for readers and team leaders include evaluations for each project completed. DRC uses these evaluations to place individuals on projects that best fit their professional backgrounds, their college degrees, and their performances on similar projects at DRC. Once placed, all readers go through rigorous training and qualifying procedures specific to the project on which they are placed. Any scorer who does not complete this training and demonstrate the ability to apply the scoring criteria by qualifying at the end of the process is not allowed to score live student responses.

5.1.2 Security

Whether training and scoring are conducted within a DRC facility or done remotely, security is essential to our handscoring process. When users log into DRC's secure, web-based scoring application, ScoreBoard, they are required to read and accept our security policy before they are allowed to access any project. For each project, scorers are also required to read and sign non-disclosure agreements, and during training emphasis is always given to what security means, the importance of maintaining security, and how this is accomplished.

Readers only have access to student responses they are qualified to score. Each scorer is assigned a unique username and password to access DRC's imaging system and must qualify before viewing any live student responses. DRC maintains full control of who may access the system and which item each scorer may score. No demographic data is available to scorers at any time.

5.1.3 Handscoring Training Process

Standard 6.9 specifies:

> Those responsible for test scoring should establish and document quality control processes and criteria. Adequate training should be provided. The quality of scoring should be monitored and documented. Any systematic source of scoring errors should be documented and corrected (118).

Training Material Development

DRC scoring supervisors trained scorers using training materials from two sources.

1. Approved training materials provided by New Meridian for ELA and math. These materials include the following:

    - Passages, prompts, and associated stimuli
    - Rubrics
    - Anchor sets
    - Practice sets
    - Qualifying sets (for prototype items only)

2. Mathematics training materials developed by DRC in conjunction with and approved by the LDOE. These materials were made for use with DRC-developed mathematics items according to processes described in DRC's response to the LDOE's "REQUEST FOR PROPOSALS For LEAP 2025 Assessment Administration (RFP #: 815200-20150723001)".

    - Prompts
    - Rubrics
    - Anchor sets
    - Practice sets
    - Qualifying sets (for all DRC-developed items)

Training and Qualifying Procedures

Handscoring involves training and qualifying team leaders and evaluators, monitoring scoring accuracy and production, and ensuring security of both the test materials and the scoring facilities. The LDOE visits the scoring centers to review training materials and oversee the training process. An explanation of the training and qualification procedures follows.

DRC used the approved mathematics and ELA training and qualifying materials to score two categories of items: "prototype" items and "abbreviated" items. Note that, like the PARCC "prototype" items for math, full sets of training and qualifying materials were also developed for all DRC-developed mathematics items. The training and qualifying procedures DRC used for these items was the same process outlined below for "prototype" mathematics items.

Prototype Items

23 items across 3-8 mathematics included in the 2023 Louisiana forms were prototype items, meaning they had a full set of associated training materials, including anchor set, practice sets, and qualifying sets. DRC started the training process with a review of the item, rubric, and anchor set, followed by the scoring and

discussion of practice sets and qualifying sets. Once this process was completed, qualified readers started scoring live student responses for that item.

Abbreviated Items

Abbreviated items required a two-step training and qualifying process. First, scorers trained and qualified as described above using approved materials for an associated prototype item that was similar to the abbreviated one they would be scoring on the Louisiana form.[2] Readers who did not qualify on the prototype item training were not allowed to continue the training.

After qualifying on the associated prototype item training, readers received additional item-specific training on the abbreviated item they were going to score. This consisted of an item-specific anchor set and two item-specific practice sets. After completing the abbreviated item training, the readers could begin scoring live student responses for the abbreviated item.

The following tables detail the composition of the training materials provided by New Meridian for mathematics and ELA.

Table 5.2 Mathematics Training Set Composition

Set Type	Prototype Item Training	Abbreviated Item Training	Annotated
Anchor Set	3 responses per score point (Composite items had 3 responses per composite score.)	3 responses per score point (Composite items had 3 responses per composite score.)	Yes
Practice Set 1	10 responses representing the range of responses	10 responses representing the range of responses	Yes
Practice Set 2	10 responses representing the range of responses	10 responses representing the range of responses	Yes
Qualifying Set 1	10 responses comparable to the anchor set responses		No
Qualifying Set 2	10 responses comparable to the anchor set responses		No
Qualifying Set 3	10 responses comparable to the anchor set responses		No
*For DRC-developed mathematics items, examples of responses at the top score points may not have been present in some anchor, training, and qualifying sets as there were few or no examples found during rangefinding or subsequent field test scoring. In such cases, DRC Scoring Directors identified examples of these scores during live scoring to supplement reader training.			

[2] Item associations were determined by PARCC/Pearson with the understanding that aspects of training are generalizable across similar items. For mathematics, the determination of prototype versus abbreviated items was made by PARCC and Pearson based on similar item types and by evidence statements. For ELA items, this determination by PARCC and Pearson was based on grade and task type.

Table 5.3 ELA Training Set Composition

Set Type	Prototype Item Training	Abbreviated Item Training	Annotated
Anchor Set*	3 responses per score point	16 responses per item: Anchor Sets for abbreviated RST and LAT item training include scores for the combined trait Reading Comprehension and Written Expression (RCWE). Anchor Sets for abbreviated NWT item training include scores for Written Expression (WE).	Yes
Practice Set 1	5 responses representing the range of responses for the Reading Comprehension and Written Expression (RCWE) trait (for LAT and RST items) the Written Expression trait (for NWT items)	10 responses representing the range of responses for the trait appropriate to the task type	Yes
Practice Set 2	5 responses representing the range of responses for the Knowledge and Use of Language Conventions trait	10 responses representing the range of responses for the conventions trait	Yes
Practice Set 3	10 responses representing the range of responses for both traits appropriate to the task type		Yes
Practice Set 4	10 responses representing the range of responses for both traits appropriate to the task type		Yes
Qualifying Set 1	10 responses comparable to the anchor set responses (included both traits appropriate to the task type)		No
Qualifying Set 2	10 responses comparable to the anchor set responses (included both traits appropriate to the task type)		No
Qualifying Set 3	10 responses comparable to the anchor set responses (included both traits appropriate to the task type)		No
Direct Copy Set**	3-5 responses composed entirely or partially of text copied from passage or passages (included both traits appropriate to the task type)	3-5 responses composed entirely or partially of text copied from passage or passages (includes both traits appropriate to the task type)	Yes

**For the ELA Knowledge and Use of Language Conventions trait, there were two mixed-prompt anchor sets per grade level (one for the narrative task and the other for the literary analysis and research simulation tasks). In addition to the mixed-prompt anchor set, depending on the task, the practice sets for prototype and abbreviated items required readers to practice scoring the Knowledge and Use of Language Conventions trait along with the Reading Comprehension and Written Expression trait (for LAT and RST items) or with the Written Expression trait (NWT). Readers were also required to qualify on the Knowledge and Use of Language Conventions trait during each prototype item qualifying session.*

***These approved sets provided additional annotated sample responses explaining the scoring rationale for responses composed entirely or partially of text copied from the source passage(s) associated with an item. DRC scoring supervisors reviewed these item-specific sets with the readers prior to scoring the associated item.*

Some items selected for use on the spring 2023 administration were previously only field tested by New Meridian. Consequently, the abbreviated training materials available for use with these items were abridged versions of typical abbreviated sets of materials. They consisted of:

- An Anchor Set (for ELA, some have annotations and some lack examples of the top scores)
- One Practice Set of 5 responses (scored but not annotated in the case of ELA)
- Approximately 10 validity responses

Since these materials were somewhat limited compared to typical abbreviated materials (the main difference being a lack of formal written annotations and fewer practice responses), DRC bolstered the training by using the field test validity responses provided by New Meridian as additional practice responses. DRC Scoring Directors then pulled additional responses from operational Louisiana student responses to use as validity responses during the scoring window. The Scoring Directors also found examples of higher-scoring responses that might be missing from the field test anchors. The validity and additional exemplar responses, along with the DRC Scoring Directors' notes for all papers used during the training of the abbreviated field-test only items, were submitted to the LDOE for approval. It is important to note that readers still had to qualify via standard qualification procedures on the prototype items for all items by first going through full training with the appropriate prototype Anchor Set, Practice Sets 1-4, and Qualifying Sets 1-3 (as well as the Conventions sets).

Qualifying Standards

DRC followed the same qualification standards that Pearson used for PARCC and New Meridian. A description of these qualifying standards follows.

Scorers demonstrated their ability to apply the scoring criteria by qualifying (i.e., scoring with acceptable agreement with true scores on qualifying sets). After each qualifying set was scored, the DRC scoring director responsible for training led the scorers in a discussion of the set.

Any scorer who did not qualify by the end of the qualifying process for an item was not allowed to score live student responses.

Table 5.4 Mathematics Qualifying Standards

	Perfect Agreement	Perfect Plus Adjacent Agreement
0, 1, 2 Rubric	80% on two of three sets	96% on two of three sets
0, 1, 2, 3 Rubric	70% on two of three sets	96% on two of three sets
0, 1, 2, 3, 4 Rubric	70% on two of three sets	95% on two of three sets

Table 5.5 Mathematics Qualifying Standards (Composite Items)*

Composite (multipart) Items	Perfect Agreement	Perfect Plus Adjacent Agreement
0, 1 Rubric	90% on two of three sets	100% on two of three sets
0, 1, 2 Rubric	80% on two of three sets	96% on two of three sets
0, 1, 2, 3 Rubric	70% on two of three sets	96% on two of three sets
0, 1, 2, 3, 4 Rubric	70% on two of three sets	95% on two of three sets

**For mathematics composite items, the appropriate qualifying standard had to be achieved on each part of the item. For example, if an item had Part A with a top score of 1, Part B with a top score of 2, and Part C with a top score of 3, a scorer/supervisor would need to achieve 90% perfect agreement on Part A, 80% perfect agreement on Part B, and 70% perfect agreement on Part C, with no more than one nonadjacent score per part across all three qualifying sets.*

Table 5.6 ELA Qualifying Standards

Perfect Agreement	Perfect Plus Adjacent Agreement
70% average for both traits on two of three qualifying sets	96% across the three qualifying sets combined on both traits
70% on each trait at least once across three qualifying sets	

ELA readers were required to meet all three of the qualifications listed in Table 5.6. Perfect plus adjacent agreement of 96% means that out of the entire pool of scores that a reader gave across the three qualifying sets for an item, no more than 4% of those scores could be nonadjacent. In other words, no more than 2 of the 60 applied scores could be nonadjacent (3 sets x 10 responses/set x 2 traits = 60 applied scores).

5.1.4 Monitoring the Scoring Process

Standard 6.8 states:

> Those responsible for test scoring should establish scoring protocols. Test scoring that involves human judgment should include rubrics, procedures, and criteria for scoring. When scoring of complex responses is done by computer, the accuracy of the algorithm and processes should be documented (118).

Section 5.1.4 explains the monitoring procedures that DRC uses to ensure that handscoring evaluators follow established scoring criteria while items are being scored. Detailed scoring rubrics, which specify the criteria for scoring, are available for handscoring evaluators for all constructed-response items.

Reader Monitoring Procedures

Throughout the handscoring process, DRC project managers, scoring directors, and team leaders reviewed the statistics that were generated on a daily basis. DRC used one team leader for every 10 to 12 readers, which was the same ratio that Pearson used for PARCC and New Meridian. If scoring concerns were apparent among individual scorers, team leaders dealt with those issues on an individual basis. If a scorer appeared to need clarification of the scoring rules, DRC supervisors typically monitored one out of five of the scorer's readings, making adjustments to that ratio as needed. If a supervisor disagreed with a reader's scores during monitoring, they provided retraining in the form of direct feedback to the reader, using rubric language and applicable training responses.

Validity Sets and Inter-Rater Reliability

In addition to the feedback that supervisors provided to readers during regular read-behinds and the continuous monitoring of inter-rater reliability and score point distributions, DRC also conducted validity scoring. Validity responses were inserted among the live student responses.

The validity responses were added to DRC's image handscoring system prior to the beginning of scoring. Validity reports compared readers' scores to pre-determined scores and were used to help detect potential room drift and individual scorer drift. This data was used to make decisions regarding the retraining and/or release of scorers, as well as the rescoring of responses.

Approximately 10% of all live student responses were scored by a second reader to establish inter-rater reliability statistics for all constructed-response items. This procedure is called a "double-blind read" because the second reader does not know the first reader's score. DRC monitored inter-rater reliability based on the responses that were scored by two readers. If a scorer fell below the expected rate of agreement, the team

leader or scoring director retrained the scorer. If a scorer failed to improve after retraining and feedback, DRC removed the scorer from the project. In this situation, DRC removed all scores assigned by the scorer in question. The responses were then reassigned and rescored.

To monitor inter-rater reliability, DRC produced scoring summary reports on a daily basis. DRC's scoring summary reports display exact, adjacent, and nonadjacent agreement rates for each reader. These rates are calculated based on responses that are scored by two readers, and their definitions are included below.

- **Percentage Exact (%EX)**—total number of responses by reader where scores are the same, divided by the number of responses that were scored twice
- **Percentage Adjacent (%AD)**—total number of responses by reader where scores are one point apart, divided by the number of responses that were scored twice
- **Percentage Nonadjacent (%NA)**—total number of responses by reader where scores are more than one score point apart, divided by the number of responses that were scored twice

The following table provided by Pearson shows the expectations for validity and inter-rater reliability:

Table 5.7 Expectations for Validity and Inter-Rater Reliability

Agreement Rate Requirements for Validity and Inter-Rater Reliability			
Content Area	**Score Point Range**	**Perfect Agreement**	**Perfect Agreement + Adjacent**
Mathematics	0–1	90%	100%
Mathematics	0–2	80%	95%
Mathematics	0–3	70%	95%
Mathematics	0–4	65%	95%
ELA	Multi-trait 0–3 or 0–4 (varies by grade and trait)	65% (each trait)	96% (each trait)

Each reader was required to maintain a level of exact agreement on validity responses and on inter-rater reliability as shown under "Perfect Agreement" in the table above. Additionally, readers were required to maintain an acceptably low rate of nonadjacent agreement. To monitor this, DRC summed each reader's exact and adjacent agreement rates and required each reader to maintain the levels shown under "Perfect Agreement + Adjacent" in the table above.

Calibration Sets

New Meridian provided DRC with approved calibration responses for all operational items that came from the leased item pool. DRC pulled calibration responses for DRC-developed mathematics items as well as additional responses for leased items. DRC used these sets to perform calibration across the entire scorer population for an item if trends were detected (e.g., low agreement between certain score points if a certain type of response was missing from initial training). These calibrations were designed to help refocus scorers on how to properly use the scoring guidelines. They were selected to help illustrate particular points and familiarize scorers with the types of responses commonly seen during operational scoring. After readers scored a calibration set, the scoring director reviewed it with the readers, using rubric language and scoring concepts exemplified by the anchor responses to explain the reasoning behind each response's score.

Reports and Reader Feedback

Reader performance and intervention information were recorded in reader feedback logs. These logs tracked information about actions taken with individual readers to ensure scoring consistency in regard to reliability, score point distribution, and validity performance. In addition to the reader feedback logs, DRC provided the LDOE with handscoring quality control reports for review throughout the scoring window. Further detail about these reports can be found in Appendix C.

5.2 Inter-Rater Reliability

A minimum of 10% of the constructed responses in ELA and mathematics were scored independently by a second reader. This was the case regardless of whether the first reader was human or AI. The statistics for inter-rater reliability were calculated for all items at all grades. To determine the reliability of scoring, the percentage of perfect agreement and adjacent agreement between the first and second scores was examined.

A total of 51 operational items were scored by human readers across all grades and both content areas. The inter-rater reliability rates and the total numbers of reads are shown in Table 5.8 for ELA items, Table 5.9 for operational mathematics items, and Table 5.10 for Spanish mathematics items.

As shown in Table 5.8, raters demonstrated at least 99% perfect and adjacent agreement for all ELA handscored items. As shown in Table 5.9 raters demonstrated at least 98% perfect and adjacent agreement for mathematics items. As shown in Table 5.10, raters demonstrated 100% perfect and adjacent agreement for Spanish mathematics items.

Table 5.8 Inter-Rater Agreement, English Language Arts Items

Grade	Task Type	Question	Trait	Total Reads	Read 2x	Inter-Rater Reliability %		
						EX	AD	EX + AD
3	Literary Analysis (PBT)	7	Reading Comprehension and Written Expression	≥32,790	≥6,390	78	22	100
			Knowledge and Use of Language Conventions	≥32,790	≥6,390	77	23	100
	Literary Analysis (CBT-AI)	7	Reading Comprehension and Written Expression	≥25,780	≥7,120	80	20	100
			Knowledge and Use of Language Conventions	≥25,780	≥7,120	81	19	100
	Research Simulation (PBT and CBT)	14	Reading Comprehension and Written Expression	≥58,130	≥12,570	79	21	100
			Knowledge and Use of Language Conventions	≥58,130	≥12,570	77	23	100
4	Research Simulation (AI)	9	Reading Comprehension and Written Expression	≥54,990	≥12,190	75	25	100
			Knowledge and Use of Language Conventions	≥54,990	≥12,190	74	26	100
	Narrative Writing (AI)	14	Written Expression	≥56,600	≥15,260	83	17	100
			Knowledge and Use of Language Conventions	≥56,600	≥15,260	78	22	100
5	Literary Analysis (AI)	7	Reading Comprehension and Written Expression	≥54,630	≥12,360	76	24	100
			Knowledge and Use of Language Conventions	≥54,630	≥12,360	75	25	100
	Research Simulation (AI)	20	Reading Comprehension and Written Expression	≥56,070	≥15,140	86	14	100
			Knowledge and Use of Language Conventions	≥56,070	≥15,140	83	17	100
6	Research Simulation (AI)	9	Reading Comprehension and Written Expression	≥49,550	≥12,060	81	19	100
			Knowledge and Use of Language Conventions	≥49,550	≥12,060	79	21	100
	Narrative Writing (AI)	14	Written Expression	≥50,370	≥13,860	85	15	100
			Knowledge and Use of Language Conventions	≥50,370	≥13,860	83	17	100
7	Literary Analysis (AI)	7	Reading Comprehension and Written Expression	≥48,270	≥11,420	82	18	100
			Knowledge and Use of Language Conventions	≥48,270	≥11,420	78	22	100
	Research Simulation (AI)	20	Reading Comprehension and Written Expression	≥48,780	≥12,470	76	24	100
			Knowledge and Use of Language Conventions	≥48,780	≥12,470	78	22	100
8	Literary Analysis (AI)	7	Reading Comprehension and Written Expression	≥51,960	≥12,150	85	14	99*
			Knowledge and Use of Language Conventions	≥51,960	≥12,150	84	16	100
		20	Reading Comprehension and Written Expression	≥52,120	≥12,350	85	15	100

Grade	Task Type	Question	Trait	Total Reads	Read 2x	Inter-Rater Reliability %		
						EX	AD	EX + AD
	Research Simulation (AI)		Knowledge and Use of Language Conventions	≥52,120	≥12,350	85	15	100

**Total Exact (EX) + Adjacent (AD) + Non-adjacent (na) does not add up to 100% due to rounding*

Table 5.9 Inter-Rater Agreement, Mathematics Items

Grade	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
3	17	Part A	≥57,800	≥12,160	95	5	100
		Part B	≥57,800	≥12,160	96	4	100
	18	Part B (PBT)	≥32,770	≥6,400	96	4	100
		Part B (CBT)	≥24,430	≥4,460	97	3	100
	32	N/A	≥57,550	≥11,730	93	7	100
	33	Part B (PBT)	≥32,720	≥6,280	97	3	100
		Part B (CBT)	≥24,340	≥4,470	98	2	100
	42	Part B (PBT)	≥32,690	≥6,220	96	4	99 (na=1)
		Part B (CBT)	≥24,550	≥4,500	96	3	99 (na =0)
		Part C (PBT)	≥32,690	≥6,220	91	8	99 (na =1)
		Part C (CBT)	≥24,550	≥4,500	89	10	99 (na =1)
	49	Part A	≥57,590	≥11,700	89	11	100
		Part B	≥57,590	≥11,700	97	3	100
4	17	Part A	≥53,740	≥10,060	92	8	100
		Part B	≥53,740	≥10,060	88	12	100
	18	N/A	≥53,350	≥11,300	96	4	100
	32	Part A	≥54,240	≥11,080	93	7	100
		Part B	≥54,240	≥11,080	92	8	100
	33	N/A	≥54,010	≥10,940	89	11	100
	48	N/A	≥53,990	≥11,180	92	8	100
	49	Part A	≥54,140	≥11,020	95	5	100
		Part B	≥54,140	≥11,020	95	5	100
		Part C	≥54,140	≥11,020	98	2	100
		Part D	≥54,140	≥11,020	95	5	100
5	17	Part A	≥53,260	≥9,690	100	0	100
		Part B	≥53,260	≥9,690	91	8	99 (na =1)
		Part C	≥53,260	≥9,690	98	2	100
	18	Part A	≥52,640	≥9,610	100	0	100
		Part B	≥52,640	≥9,610	93	6	99 (na =0)
	32	N/A	≥53,140	≥10,630	87	11	99 (na =1)
	33	N/A	≥53,140	≥10,550	88	10	98 (na =2)
	48	Part A	≥53,260	≥9,710	100	0	100
		Part B	≥53,260	≥9,710	93	7	100
	49	Part A	≥53,150	≥9,910	90	10	100 (na =1)
		Part B	≥53,150	≥9,910	96	4	100
		Part C	≥53,150	≥9,910	100	0	100

**Total Exact (EX) + Adjacent (AD) + Non-adjacent (na) does not add up to 100% due to rounding*
***N/A if an item does not have multiple parts*

Table 5.10 Inter-Rater Agreement, Mathematics Items, continued

Grade	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
6	30	N/A	≥53,140	≥10,410	89	10	99 (na =2)
	34	Part A	≥53,370	≥10,310	97	3	100
		Part B	≥53,370	≥10,310	94	5	99 (na =0)
	35	Part A	≥53,330	≥9,760	100	0	100
		Part B	≥53,330	≥9,760	95	5	100
		Part C	≥53,330	≥9,760	98	2	100
	36	N/A	≥52,920	≥11,440	93	6	99 (na =0)
	47	N/A	≥53,680	≥11,070	91	9	100
	48	Part A	≥53,370	≥9,730	100	0	100
		Part B	≥53,370	≥9,730	100	0	100
		Part C	≥53,370	≥9,730	92	7	99 (na =0)
	49	N/A	≥53,540	≥11,070	93	7	100
7	29	Part A	≥54,010	≥9,850	100	0	100
		Part B	≥54,010	≥9,850	100	0	100
		Part C	≥54,010	≥9,850	95	5	100
		Part D	≥54,010	≥9,850	96	4	100
	32	Part A	≥53,880	≥10,940	95	5	100
		Part B	≥53,880	≥10,940	99	1	100
	35	Part A	≥53,220	≥12,070	97	3	100
		Part B	≥53,220	≥12,070	95	5	100
	36	N/A	≥52,910	≥12,760	92	6	98 (na =2)
	47	Part A	≥53,870	≥9,760	100	0	100
		Part B	≥53,870	≥9,760	94	6	100
	48	Part A	≥54,000	≥10,660	92	7	99 (na =1)
		Part B	≥54,000	≥10,660	98	2	100
		Part C	≥54,000	≥10,660	95	5	100
	49	N/A	≥54,310	≥12,030	93	7	100
8	28	Part A	≥48,440	≥8,810	100	0	100
		Part B	≥48,440	≥8,810	96	4	100
		Part C	≥48,440	≥8,810	97	3	100
	34	Part A	≥48,340	≥8,800	100	0	100
		Part B	≥48,340	≥8,800	93	7	100
	35	Part A	≥48,600	≥10,470	95	5	100
		Part B	≥48,600	≥10,470	96	4	100
		Part C	≥48,600	≥10,470	96	4	100 (na =1)
	36	N/A	≥48,470	≥11,690	96	4	100

Grade	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
	44	Part A	≥48,860	≥8,890	100	0	100
		Part B	≥48,860	≥8,890	99	1	100
	46	Part A	≥48,880	≥10,740	92	7	100
		Part B	≥48,880	≥10,740	88	12	100
	48	N/A	≥48,630	≥11,650	92	7	99 (na =1)

**Total Exact (EX) + Adjacent (AD) + Non-adjacent (na) does not add up to 100% due to rounding*
***N/A if an item does not have multiple parts*

Table 5.11 Inter-Rater Agreement, Spanish Mathematics Items

Grade	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
3	17	Part A	≥100	<10	NR	NR	NR
		Part B	≥100	<10	NR	NR	NR
	18	Part B (CBT)	≥70	<10	NR	NR	NR
		Part B (PBT)	≥40	≥30	100	0	100
	32	N/A	≥130	≥50	96	4	100
	33	Part B (PBT)	≥30	<10	NR	NR	NR
		Part B (CBT)	≥80	≥10	100	0	100
	42	Part B (CBT)	≥80	≥10	100	0	100
		Part B (PBT)	≥30	<10	NR	NR	NR
		Part C (CBT)	≥80	≥10	100	0	100
		Part C (PBT)	≥30	<10	NR	NR	NR
	49	Part A	≥120	≥30	100	0	100
		Part B	≥120	≥30	100	0	100
4	17	Part A	≥90	≥10	100	0	100
		Part B	≥90	≥10	100	0	100
	18	N/A	≥90	≥10	100	0	100
	32	Part A	≥90	≥10	100	0	100
		Part B	≥90	≥10	100	0	100
	33	N/A	≥90	≥10	100	0	100
	48	N/A	≥90	≥10	100	0	100
	49	Part A	≥90	≥10	100	0	100
		Part B	≥90	≥10	100	0	100
		Part C	≥90	≥10	100	0	100
		Part D	≥90	≥10	100	0	100
5	17	Part A	≥100	≥20	100	0	100
		Part B	≥100	≥20	100	0	100
		Part C	≥100	≥20	100	0	100
	18	Part A	≥100	≥20	100	0	100
		Part B	≥100	≥20	100	0	100
	32	N/A	≥100	≥20	100	0	100
	33	N/A	≥100	≥20	100	0	100
	48	Part A	≥100	≥20	100	0	100
		Part B	≥100	≥20	100	0	100
	49	Part A	≥100	≥20	100	0	100
		Part B	≥100	≥20	100	0	100
		Part C	≥100	≥20	100	0	100

**Total Exact (EX) + Adjacent (AD) does not add up to 100% due to rounding*
***N/A if an item does not have multiple parts*

Table 5.12 Inter-Rater Agreement, Spanish Mathematics Items, continued

Grade	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
6	30	N/A	≥120	≥30	94	6	100
	34	Part A	≥120	≥30	100	0	100
		Part B	≥120	≥30	100	0	100
	35	Part A	≥110	≥20	100	0	100
		Part B	≥110	≥20	92	8	100
		Part C	≥110	≥20	100	0	100
	36	N/A	≥110	≥20	100	0	100
	47	N/A	≥120	≥30	100	0	100
	48	Part A	≥110	≥20	100	0	100
		Part B	≥110	≥20	100	0	100
		Part C	≥110	≥20	100	0	100
	49	N/A	≥110	≥30	100	0	100
7	29	Part A	≥130	≥20	100	0	100
		Part B	≥130	≥20	100	0	100
		Part C	≥130	≥20	100	0	100
		Part D	≥130	≥20	100	0	100
	32	Part A	≥120	<10	NR	NR	NR
		Part B	≥120	<10	NR	NR	NR
	35	Part A	≥130	≥30	100	0	100
		Part B	≥130	≥30	93	7	100
	36	N/A	≥140	≥40	100	0	100
	47	Part A	≥130	≥20	100	0	100
		Part B	≥130	≥20	100	0	100
	48	Part A	≥130	≥20	100	0	100
		Part B	≥130	≥20	100	0	100
		Part C	≥130	≥20	100	0	100
	49	N/A	≥130	≥30	100	0	100
8	28	Part A	≥120	≥20	100	0	100
		Part B	≥120	≥20	100	0	100
		Part C	≥120	≥20	100	0	100
	34	Part A	≥120	≥20	100	0	100
		Part B	≥120	≥20	100	0	100
	35	Part A	≥130	≥30	100	0	100

		Part B	≥130	≥30	100	0	100
		Part C	≥130	≥30	100	0	100
	36	N/A	≥130	≥30	100	0	100
	44	Part A	≥120	≥20	100	0	100
		Part B	≥120	≥20	100	0	100
	46	Part A	≥130	≥30	100	0	100
		Part B	≥130	≥30	100	0	100
	48	N/A	≥130	≥40	100	0	100

**Total Exact (EX) + Adjacent (AD) does not add up to 100% due to rounding*
***N/A if an item does not have multiple parts*

Technology-Enhanced Item Scoring Process

All technology-enhanced items, as well as EBSR, MPSR, and SA items, were processed through DRC's autoscoring engine and scored according to the assigned scoring rules as established during content creation by PARCC or DRC as applicable in conjunction with the LDOE. DRC ensured that all rubrics and scoring rules were verified for accuracy before scoring any technology-enhanced items. DRC established an adjudication process for technology-enhanced items and short-answer responses to verify that correct answers were identified. DRC's technology-enhanced scoring process included the following procedures:

- A scoring rubric was created for each technology-enhanced item. The rubric described the one and only correct answer for dichotomously scored items (i.e., items scored as either right or wrong). If partial credit was possible, the rubric described in detail the type of response that could receive credit for each score point.
- The information from the scoring rubric was entered into the scoring system within the item banking system so that the truth resided in one place along with the item image and other metadata. This scoring information included details that varied by item type. For example, for a drag-and-drop item, the information included which objects are to be placed in each drop region to receive credit.
- The information was then verified by another autoscoring expert.
- After testing started, reports were generated that showed every response, how many students gave that response, and the score the scoring system provided for that response.
- The scoring was then checked against the scoring rubric using two levels of verification.
- If any discrepancies were found, the scoring information was modified and verified again. The scoring process was then rerun. This checking and modification process continued until no other issues were found.
- As a final check, a final report was generated that showed all student responses, their frequencies, and their received scores.

In the case of braille and large-print test forms, student responses to items were transcribed into the online system by a test administrator.

5.3 Multiple-Choice and Multiple-Select Item Scoring Process

Responses to multiple-choice and multiple-select items were captured during the CBT administration and during scanning of the PBT answer documents. In the case of braille and large-print test forms, student responses to these items were transcribed into the online system by a test administrator.

5.4 Summary

The information presented in this chapter summarizes the scoring procedures for different types of items and the steps taken by DRC to ensure accuracy in the autoscoring and handscoring processes. The inter-rater reliability statistics presented in Section 5.4 demonstrate that the items were scored reliably. These efforts by DRC address multiple best practices of the testing industry but are particularly related to AERA, APA, & NCME (2014) Standards 4.18, 4.20, 6.8, and 6.9:

> ***Standard 4.18*** Procedures for scoring and, if relevant, scoring criteria, should be presented by the test developer with sufficient detail and clarity to maximize the accuracy of scoring. Instructions for

using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical for extended-response items such as performance tasks, portfolios, and essays (91).

Standard 4.20 The process for selecting, training, qualifying, and monitoring scorers should be specified by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the rubric score scale, and the procedures for training scorers should result in a degree of accuracy and agreement among scorers that allows the scores to be interpreted as originally intended by the test developer. Specifications should also describe processes for assessing scorer consistency and potential drift over time in raters' scoring (92).

Standard 6.8 Those responsible for test scoring should establish scoring protocols. Test scoring that involves human judgment should include rubrics, procedures, and criteria for scoring. When scoring of complex responses is done by computer, the accuracy of the algorithm and processes should be documented (118).

Standard 6.9 Those responsible for test scoring should establish and document quality control processes and criteria. Adequate training should be provided. The quality of scoring should be monitored and documented. Any systematic source of scoring errors should be documented and corrected (118).

Chapter 6: Operational Data Analyses

This chapter of the LEAP 2025 technical report describes the analyses that were conducted on the operational data. These include a classical item analysis and examination of the raw scores and an item response theory (IRT) analysis involving calibrating, scaling, and linking.

This section presents the classical item statistics, including aggregate raw score statistics and individual item-level statistics. Next, this section discusses the IRT models used for calibrating the data and addresses the purpose of data calibration and scaling for each content area is addressed. The calibration samples are presented next, followed by the data calibration results, including the model-data fit for the Louisiana data. If the IRT models fit the empirical item response distributions for the population about which generalizations are to be made (i.e., Louisiana students), then the claim that the scores are valid indicators of an underlying ability is strengthened. The lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS) for the LEAP 2025 tests are also presented.

Chapter 6 demonstrates how LEAP 2025 assessments adhere to American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME, 2014) Standards 1.8, 4.14, 5.2, 5.13, 5.15, and 7.2. Each standard is explicated within the appropriate section of this chapter. Standard 7.2 provides general guidance that is relevant to this chapter. It states the following:

> The population for whom a test is intended and specifications for the test should be documented (126).

For all 2023 LEAP 2025 analyses, the Louisiana student population was used. In Section 6.3, the characteristics of calibration samples, such as subgroups, are discussed. Chapter 3 presents the test specifications. Information regarding reported data is discussed in detail in Chapter 7.

In this section, summary test statistics for each form, grade, and content area of LEAP 2025 are presented. These statistics are followed by item-level statistics for each grade and content area of LEAP 2025. These statistics were produced using census data, after removing data from test takers who were administered the Spanish language and Braille versions of the test forms.

6.1 Test-Level Statistics

Table 6.1 presents the number of items, score points, mean and standard deviation of the raw scores, and average form difficulty for each test form at each grade level of the ELA and mathematics assessments, respectively. Form difficulty for an examinee was calculated by dividing the raw score of the student by total score points of the test.

As can be seen in the table, average form difficulty for ELA ranged from 0.31 to 0.46. Average form difficulty for mathematics ranged from 0.30 to 0.51. In general, the 2023 LEAP 2025 tests were relatively difficult tests across all subjects and grades. For ELA, the grade 3 computer-based test (CBT) was the most difficult, with 0.31 average form difficulty, and the grade 8 was the easiest, with 0.46 average form difficulty. For mathematics, the grade 8 test was the most difficult, with 0.30 average form difficulty, and the grade 3 paper-based test (PBT) test was the easiest, with 0.51 average form difficulty.

Table 6.1 LEAP 2025 Means and Standard Deviations for Raw Scores and Form Difficulty

Content	Grade	Mode	Total Items	Total Points	Mean Raw Score (Std. Dev.)	Average Form Difficulty (Std. Dev.)
ELA	3	CBT	26	70	21.66 (13.01)	0.31 (0.19)
	3	PBT	26	70	26.38 (12.93)	0.38 (0.18)
	4	CBT	31	83	31.98 (16.37)	0.39 (0.20)
	5	CBT	30	86	33.82 (16.32)	0.39 (0.19)
	6	CBT	33	90	34.13 (17.19)	0.38 (0.19)
	7	CBT	34	94	40.19 (20.79)	0.43 (0.22)
	8	CBT	34	94	43.40 (19.04)	0.46 (0.20)
Mathematics	3	CBT	43	62	28.27 (13.19)	0.46 (0.21)
	3	PBT	43	62	31.59 (13.74)	0.51 (0.22)
	4	CBT	43	62	27.49 (15.16)	0.44 (0.24)
	5	CBT	43	62	26.09 (14.21)	0.42 (0.23)
	6	CBT	43	66	25.51 (15.35)	0.39 (0.23)
	7	CBT	43	66	22.21 (14.98)	0.34 (0.23)
	8	CBT	41	65	19.23 (12.28)	0.30 (0.19)

Table 6.2 presents the number of items, mean and standard deviation of the item *p*-values, and item-total correlations (i.e., item discrimination values) for each test form at each grade level of the ELA and mathematics assessments, respectively.

The mean *p*-value is the average of all item *p*-values of a specific grade and content area. The mean item-total correlation (R_{it}) is the average of all item point-biserial correlations of a specific grade and content area. The *p*-value and item-total correlation are explained in the next section.

Table 6.2 LEAP 2025 Means, Standard Deviations for *p*-Values, Item-Total Correlation (R_{it})

				Item *p*-Value				Item-Total Correlation			
Content	Grade	Mode	N of Items	Mean	Std. Dev.	Min.	Max	Mean	Std. Dev.	Min.	Max
ELA	3	CBT	26	0.36	0.15	0.14	0.65	0.49	0.14	0.27	0.76
	3	PBT	26	0.42	0.17	0.21	0.79	0.48	0.13	0.25	0.70
	4	CBT	31	0.43	0.14	0.19	0.70	0.48	0.14	0.15	0.77
	5	CBT	30	0.45	0.14	0.15	0.77	0.49	0.17	0.21	0.82
	6	CBT	33	0.44	0.14	0.18	0.76	0.50	0.14	0.24	0.76
	7	CBT	34	0.46	0.12	0.27	0.78	0.53	0.15	0.37	0.85
	8	CBT	34	0.50	0.13	0.28	0.76	0.49	0.16	0.25	0.81
Mathematics	3	CBT	43	0.53	0.18	0.15	0.77	0.47	0.11	0.22	0.76
	3	PBT	43	0.58	0.17	0.19	0.83	0.47	0.11	0.19	0.76
	4	CBT	43	0.50	0.15	0.22	0.83	0.52	0.11	0.27	0.73
	5	CBT	43	0.48	0.17	0.15	0.79	0.51	0.12	0.26	0.77
	6	CBT	43	0.44	0.16	0.16	0.80	0.52	0.11	0.31	0.77
	7	CBT	43	0.37	0.14	0.16	0.76	0.53	0.13	0.27	0.77
	8	CBT	41	0.35	0.16	0.10	0.75	0.47	0.11	0.27	0.74

6.2 Item-Level Statistics

Tables 6.3–6.9 present the item statistics for each operational item included in regular test forms organized by grade for ELA. Tables 6.10–6.16 show the item statistics for each item included in regular test forms organized by grade for mathematics. The tables include administration mode, item number, *p*-value, item-total correlation (R_{it}), omit rates, total N, adjusted N (adjusted N excludes items with multiple responses [PBT only], omitted responses, responses that were not scored, or responses that received a non-score code), and the percentage at each score point, if applicable, for each item by grade and content area. The *p*-value and item-total correlations calculations used the adjusted N to determine the values. The rest of the statistics in the table are based on the total N.

p-Value

The *p*-value is a measure of item difficulty. For a multiple-choice (MC) item, the *p*-value is calculated by dividing the number of students who correctly responded to an item by the total number of students who attempted the item. The value is reported as a proportion. For a non-MC item, the *p*-value is calculated by dividing the average score for the item by the maximum points possible. This value is also reported as a proportion.

In terms of *p*-values, test scores tend to be more precise when their average *p*-values are between the mid-0.50s and the low 0.70s. However, it is important to select items on the basis of content rather than on purely statistical criteria when building a criterion-referenced test. As shown in Table 6.2, the average *p*-values associated with the ELA forms range from 0.36 in the grade 3 CBT form to 0.50 in grade 8. The average *p*-values associated with the mathematics forms range from 0.35 in grade 8 CBT to 0.58 in grade 3 PBT.

It is important that one examines the range of *p*-values, not just the average *p*-value, to determine whether a test measures well. It is desirable for a test to measure well throughout the range of skills present at a given grade. That is, it is important that the items measure the performance of students of all levels of achievement, not just students in the center of the distribution. Having a range of *p*-values also helps to

prevent floor and/or ceiling effects so that the test does not have large numbers of students at the minimum or maximum possible scores. The ELA forms have items with *p*-values ranging from 0.14 to 0.79 (see Tables 6.3–6.9) across all grade levels. The *p*-values on the mathematics forms range from 0.14 to 0.83 (see Tables 6.10–6.16). Such a broad range of *p*-values, which indicates the items measure well throughout the range of skill levels at a given grade, supports the accuracy of the LEAP 2025 test scores.

Item-Total Correlations

An item-total correlation is the correlation between an item score and the total test score, where the item score is not included in the total score. It indicates how well an item differentiates students across all levels of achievement. In general, items with correlations below 0.20 are said to be poorly discriminating. The majority of the items in the LEAP 2025 had item-total correlations above this threshold. Any item with an item-total correlation below the 0.20 threshold was further analyzed to ensure that the item was correctly keyed.

Omit Rates

The omit rate for each item indicates the percentage of students who did not answer the item. Omit rates can be used to examine possible speededness issues on tests. A test may be speeded if students do not have adequate time to answer all questions on the test. In general, an item is said to have a high omit rate if more than 5% of students failed to respond to the item. Evidence of speededness is considered a threat to validity because student test scores may not reflect their ability. Additionally, content validity may be threatened because the items that were not completed are needed to fulfill content blueprint specifications (Lu & Sireci, 2007).

This examination of omit rates complies with Standard 4.14 of the *Standards*. This standard is concerned with the speededness of a test and states the following:

> For a test that has a time limit, test development research should examine the degree to which scores include a speed component and should evaluate the appropriateness of that component, given the domain the test is designed to measure (90).

The results in this section will show that, overall, student test scores are not adversely affected by the rate at which the students complete the test. In general, students have ample time to complete all sections of the test and there is not a threat to construct or content validity.

The results presented in Tables 6.3–6.16 show that the omit rates for most of the items on the LEAP 2025 regular forms are less than 5%, suggesting that the majority of students were able to complete the test in the prescribed amount of time. There are very few items with an omit rate higher than 5%, and the omit rates for the last items in the tests rarely exceeded 3%, the largest omit rate being 3.86%. These omit rates indicate that 97% of the students completed the test. Lu & Sireci (2007) report that the Education Testing Service has used an approach where a test was considered unspeeded if at least 80% of the examinees reach the last item and all testers reach at least 75% of the items. The reported omit rates fall within these ranges.

Table 6.3 Operational Item Statistics—English Language Arts Grade 3 CBT Administration

ELA Grade 3 Computer-Based Test Administration											
Item ID	Item Type	Total N	Adj. N	*p*-Value	Pbis	Omit Rate	% at 0	% at 1	% at 2	% at 3	% Nonscore Codes
982074	ESR	≥22,120	≥22,090	0.65	0.46	0	29	12	59		0
982121	TE	≥22,120	≥21,890	0.50	0.50	1	39	21	39		0
982112	ESR	≥22,120	≥22,060	0.36	0.43	0	53	22	25		0
982109	ESR	≥22,120	≥22,060	0.54	0.58	0	35	21	43		0
982123	ESR	≥22,120	≥22,060	0.61	0.51	0	35	8	57		0
982078	ESR	≥22,120	≥22,050	0.29	0.27	0	61	18	20		0
98207902	CR	≥22,120	≥21,440	0.22	0.76	1	46	39	10	2	2
98207903	CR	≥22,120	≥21,440	0.29	0.71	1	36	40	19	2	2
1113142	TE	≥22,120	≥22,040	0.36	0.33	0	61	7	32		0
1113146	ESR	≥22,120	≥22,070	0.38	0.46	0	47	30	23		0
1113145	ESR	≥22,120	≥22,080	0.62	0.53	0	28	19	52		0
1113135	MS	≥22,120	≥22,070	0.32	0.35	0	48	39	13		0
1113144	ESR	≥22,120	≥22,070	0.56	0.48	0	38	13	49		0
1113138	MS	≥22,120	≥22,070	0.30	0.46	0	56	27	16		0
111314002	CR	≥22,120	≥21,330	0.14	0.66	1	61	31	4	0	3
111314003	CR	≥22,120	≥21,330	0.15	0.64	1	59	30	6	1	3
995132	ESR	≥22,120	≥22,080	0.40	0.45	0	54	11	35		0
995135	TE	≥22,120	≥22,050	0.37	0.52	0	46	35	19		0
995141	ESR	≥22,120	≥22,020	0.23	0.28	0	63	27	10		0
995133	ESR	≥22,120	≥22,020	0.33	0.39	0	58	17	25		0
916851	TE	≥22,120	≥21,830	0.28	0.37	1	65	11	23		0
916769	MS	≥22,120	≥21,760	0.28	0.37	2	48	44	6		0
916765	MS	≥22,120	≥21,710	0.46	0.58	2	37	33	28		0
916759	ESR	≥22,120	≥21,680	0.29	0.33	2	61	17	19		0

Table 6.4 Operational Item Statistics—English Language Arts Grade 3 PBT Administration

ELA Grade 3 Paper-Based Test Administration											
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% Nonscore Codes**
982074	ESR	≥27,200	≥27,120	0.74	0.47	0	22	9	69		0
982110	ESR	≥27,200	≥27,100	0.71	0.45	0	23	11	66		0
982112	ESR	≥27,200	≥27,090	0.41	0.42	0	48	22	30		0
982109	ESR	≥27,200	≥27,130	0.64	0.58	0	27	18	55		0
982123	ESR	≥27,200	≥27,080	0.79	0.51	0	17	7	75		0
982078	ESR	≥27,200	≥27,060	0.33	0.25	1	57	19	23		0
982079P2	CR	≥27,200	≥26,780	0.30	0.70	1	36	40	20	3	1
982079P3	CR	≥27,200	≥26,780	0.33	0.64	1	28	45	22	3	1
1113137	ESR	≥27,200	≥27,090	0.32	0.26	0	61	13	26		0
1113146	ESR	≥27,200	≥27,090	0.41	0.44	0	42	33	24		0
1113145	ESR	≥27,200	≥27,100	0.65	0.55	0	24	20	55		0
1113135	MS	≥27,200	≥27,040	0.37	0.38	1	45	36	19		0
1113144	ESR	≥27,200	≥27,000	0.61	0.45	1	32	13	54		0
1113138	MS	≥27,200	≥27,030	0.32	0.45	1	56	22	21		0
1113140P2	CR	≥27,200	≥26,750	0.21	0.63	1	43	49	7	0	1
1113140P3	CR	≥27,200	≥26,750	0.25	0.57	1	32	59	8	0	1
995132	ESR	≥27,200	≥27,030	0.48	0.46	1	46	10	43		0
995134	MS	≥27,200	≥27,010	0.47	0.50	1	34	37	29		0
995141	ESR	≥27,200	≥27,000	0.25	0.28	1	60	28	11		0
995133	ESR	≥27,200	≥27,000	0.36	0.36	1	55	16	28		0
916760	ESR	≥27,200	≥26,550	0.31	0.43	2	62	10	26		0
916769	MS	≥27,200	≥26,410	0.34	0.38	3	41	46	10		0
916765	MS	≥27,200	≥26,410	0.51	0.57	3	34	28	36		0
916759	ESR	≥27,200	≥26,200	0.33	0.34	4	55	19	22		0

Table 6.5 Operational Item Statistics—English Language Arts Grade 4 CBT Administration

ELA Grade 4 Computer-Based Test Administration												
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% Nonscore Codes**
1029266	TE	≥48,880	≥48,820	0.40	0.52	0	46	28	26			0
1029268	ESR	≥48,880	≥48,810	0.40	0.35	0	53	13	34			0
1029267	MS	≥48,880	≥48,820	0.25	0.40	0	60	29	10			0
1029269	ESR	≥48,880	≥48,810	0.42	0.43	0	52	12	36			0
1029270	ESR	≥48,880	≥48,800	0.48	0.43	0	48	9	44			0
1029272	ESR	≥48,880	≥48,800	0.55	0.53	0	32	24	43			0
1029273	ESR	≥48,880	≥48,800	0.47	0.48	0	40	25	34			0
1029275	TE	≥48,880	≥48,640	0.50	0.46	0	24	52	24			0
1029277O2	CR	≥48,880	≥48,220	0.25	0.77	0	32	37	27	2	0	1
1029277O3	CR	≥48,880	≥48,220	0.36	0.73	0	30	36	29	4		1
982190	ESR	≥48,880	≥48,860	0.70	0.54	0	24	11	64			0
982196	ESR	≥48,880	≥48,820	0.62	0.45	0	36	4	59			0
982191	ESR	≥48,880	≥48,850	0.57	0.50	0	40	6	54			0
982189	TE	≥48,880	≥48,800	0.36	0.43	0	49	29	22			0
982192O2	CR	≥48,880	≥48,260	0.25	0.72	0	49	29	19	2		1
982192O3	CR	≥48,880	≥48,260	0.30	0.70	0	33	46	17	3		1
1029306	ESR	≥48,880	≥48,820	0.39	0.31	0	58	7	35			0
1029303	ESR	≥48,880	≥48,830	0.66	0.54	0	26	15	58			0
1029304	ESR	≥48,880	≥48,830	0.33	0.36	0	60	14	26			0
1029305	ESR	≥48,880	≥48,820	0.23	0.35	0	74	7	19			0
1029307	ESR	≥48,880	≥48,820	0.54	0.54	0	38	16	46			0
1029302	ESR	≥48,880	≥48,810	0.52	0.50	0	42	13	45			0
913589	TE	≥48,880	≥48,690	0.49	0.28	0	21	60	18			0
913590	ESR	≥48,880	≥48,810	0.49	0.37	0	41	20	39			0
913588	ESR	≥48,880	≥48,820	0.53	0.42	0	38	18	44			0
1029329	TE	≥48,880	≥48,800	0.60	0.50	0	21	38	41			0
913533	ESR	≥48,880	≥48,750	0.44	0.51	0	49	14	37			0
913532	MS	≥48,880	≥48,720	0.33	0.46	0	55	25	20			0
913529	ESR	≥48,880	≥48,700	0.54	0.46	0	36	20	43			0
913531	ESR	≥48,880	≥48,680	0.19	0.15	0	77	8	14			0

Table 6.6 Operational Item Statistics—English Language Arts Grade 5 CBT Administration

ELA Grade 5 Computer-Based Test Administration												
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% Nonscore Codes**
980751	ESR	≥48,310	≥48,270	0.43	0.36	0	45	24	31			0
980793	ESR	≥48,310	≥48,280	0.77	0.49	0	20	5	74			0
980797	ESR	≥48,310	≥48,230	0.44	0.37	0	47	17	36			0
980796	MS	≥48,310	≥48,220	0.42	0.50	0	45	25	29			0
980754	TE	≥48,310	≥48,200	0.42	0.46	0	49	17	34			0
980752	TE	≥48,310	≥48,220	0.62	0.48	0	9	58	33			0
980755O2	CR	≥48,310	≥47,800	0.35	0.82	0	21	30	37	10	1	1
980755O3	CR	≥48,310	≥47,800	0.47	0.80	0	21	29	36	13		1
916777	MS	≥48,310	≥47,840	0.46	0.42	1	38	33	29			0
916772	ESR	≥48,310	≥47,770	0.74	0.45	1	16	20	63			0
916774	TE	≥48,310	≥47,660	0.52	0.40	1	41	12	45			0
916848	TE	≥48,310	≥47,470	0.43	0.46	2	39	34	25			0
980711	ESR	≥48,310	≥48,270	0.41	0.22	0	53	11	36			0
913620	ESR	≥48,310	≥48,270	0.50	0.21	0	48	5	47			0
913621	TE	≥48,310	≥48,240	0.49	0.51	0	38	26	36			0
980713	ESR	≥48,310	≥48,240	0.47	0.41	0	45	18	38			0
913623	ESR	≥48,310	≥48,260	0.54	0.48	0	37	19	44			0
913625	TE	≥48,310	≥48,260	0.48	0.39	0	22	60	18			0
913626	ESR	≥48,310	≥48,230	0.73	0.52	0	18	17	65			0
980715	ESR	≥48,310	≥48,260	0.50	0.45	0	47	7	46			0
913628O2	CR	≥48,310	≥47,760	0.15	0.74	0	52	37	10	1	0	1
913628O3	CR	≥48,310	≥47,760	0.33	0.75	0	30	40	28	1		1
868596	TE	≥48,310	≥48,250	0.45	0.54	0	46	18	36			0
868593	ESR	≥48,310	≥48,240	0.41	0.36	0	54	9	37			0
868594	ESR	≥48,310	≥48,230	0.41	0.51	0	55	10	36			0
868592	ESR	≥48,310	≥48,210	0.38	0.47	0	58	8	34			0
868591	ESR	≥48,310	≥48,220	0.55	0.43	0	37	17	47			0
868595	ESR	≥48,310	≥48,200	0.27	0.22	0	67	11	22			0

Table 6.7 Operational Item Statistics—English Language Arts Grade 6 CBT Administration

ELA Grade 6 Computer-Based Test Administration												
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% Nonscore Codes**
1030066	ESR	≥43,370	≥43,330	0.48	0.36	0	45	13	41			0
1030062	ESR	≥43,370	≥43,310	0.29	0.43	0	67	8	25			0
1030058	TE	≥43,370	≥43,250	0.53	0.24	0	7	80	13			0
1030067	ESR	≥43,370	≥43,280	0.34	0.31	0	49	33	17			0
1030060	ESR	≥43,370	≥43,290	0.34	0.42	0	57	18	25			0
1030059	MS	≥43,370	≥43,270	0.30	0.40	0	55	31	14			0
1030064	TE	≥43,370	≥43,250	0.45	0.57	0	36	36	27			0
1030061	ESR	≥43,370	≥43,240	0.44	0.44	0	50	13	37			0
1030068O2	CR	≥43,370	≥42,860	0.25	0.76	1	28	46	22	2	0	1
1030068O3	CR	≥43,370	≥42,860	0.32	0.74	1	32	41	24	3		1
980268	ESR	≥43,370	≥43,330	0.54	0.41	0	21	51	28			0
980277	MS	≥43,370	≥43,280	0.37	0.51	0	49	27	23			0
980269	ESR	≥43,370	≥43,300	0.60	0.51	0	35	11	54			0
980270	MS	≥43,370	≥43,310	0.49	0.61	0	38	25	37			0
980264O2	CR	≥43,370	≥42,590	0.18	0.75	1	54	26	13	4	1	1
980264O3	CR	≥43,370	≥42,590	0.25	0.75	1	47	32	15	4		1
1114056	MS	≥43,370	≥43,300	0.54	0.48	0	22	48	30			0
1114060	ESR	≥43,370	≥43,300	0.59	0.50	0	37	8	55			0
1114062	MS	≥43,370	≥43,290	0.37	0.48	0	42	42	16			0
1114057	ESR	≥43,370	≥43,280	0.72	0.57	0	21	14	64			0
1114061	ESR	≥43,370	≥43,270	0.32	0.42	0	64	8	28			0
1114059	TE	≥43,370	≥43,260	0.67	0.52	0	17	33	51			0
1030149	TE	≥43,370	≥43,320	0.76	0.37	0	18	11	71			0
1030153	ESR	≥43,370	≥43,250	0.38	0.39	0	59	7	34			0
1030154	MS	≥43,370	≥43,260	0.46	0.56	0	40	28	32			0
1030152	ESR	≥43,370	≥43,270	0.39	0.47	0	56	9	35			0
1030158	ESR	≥43,370	≥43,280	0.48	0.47	0	45	14	41			0
1030159	TE	≥43,370	≥43,240	0.54	0.51	0	26	40	34			0
980274	TE	≥43,370	≥43,200	0.31	0.29	0	52	34	14			0
980271	ESR	≥43,370	≥43,230	0.45	0.48	0	39	31	29			0
980273	MS	≥43,370	≥43,240	0.49	0.40	0	18	64	17			0
980276	ESR	≥43,370	≥43,240	0.58	0.59	0	37	11	52			0

Table 6.8 Operational Item Statistics—English Language Arts Grade 7 CBT Administration

ELA Grade 7 Computer-Based Test Administration												
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% Nonscore Codes**
1031079	ESR	≥42,460	≥42,430	0.61	0.40	0	34	10	56			0
1031074	ESR	≥42,460	≥42,420	0.78	0.46	0	19	5	75			0
1031072	TE	≥42,460	≥42,410	0.49	0.52	0	37	28	35			0
1031080	ESR	≥42,460	≥42,390	0.43	0.45	0	55	4	41			0
1031077	TE	≥42,460	≥42,400	0.46	0.43	0	44	19	37			0
1031073	ESR	≥42,460	≥42,410	0.71	0.38	0	25	8	67			0
1031081O2	CR	≥42,460	≥41,660	0.28	0.82	1	29	37	26	6	1	1
1031081O3	CR	≥42,460	≥41,660	0.41	0.80	1	24	35	30	9		1
1031186	MS	≥42,460	≥42,320	0.44	0.54	0	43	26	31			0
1031190	MS	≥42,460	≥42,290	0.35	0.49	0	51	27	21			0
1031187	ESR	≥42,460	≥42,270	0.44	0.58	0	47	17	35			0
1031191	TE	≥42,460	≥42,130	0.27	0.45	1	65	16	18			0
1031150	ESR	≥42,460	≥42,420	0.58	0.49	0	31	20	48			0
1031141	ESR	≥42,460	≥42,400	0.67	0.50	0	27	12	61			0
1031149	ESR	≥42,460	≥42,400	0.38	0.49	0	55	13	32			0
1031151	ESR	≥42,460	≥42,390	0.39	0.37	0	47	29	24			0
1031152	ESR	≥42,460	≥42,410	0.47	0.42	0	43	20	37			0
1031156	ESR	≥42,460	≥42,400	0.44	0.48	0	51	9	39			0
1031158	TE	≥42,460	≥42,410	0.46	0.45	0	35	39	26			0
1031154	TE	≥42,460	≥42,360	0.35	0.53	0	53	24	23			0
1031145O2	CR	≥42,460	≥41,520	0.42	0.85	1	20	21	32	21	5	1
1031145O3	CR	≥42,460	≥41,520	0.54	0.81	1	19	22	32	24		1
995797	ESR	≥42,460	≥42,390	0.52	0.50	0	38	20	42			0
995790	TE	≥42,460	≥42,350	0.46	0.42	0	34	40	26			0
995795	MS	≥42,460	≥42,370	0.47	0.52	0	32	41	27			0
995793	ESR	≥42,460	≥42,390	0.54	0.46	0	27	37	36			0
995324	ESR	≥42,460	≥42,360	0.47	0.41	0	44	20	37			0
995336	ESR	≥42,460	≥42,350	0.50	0.51	0	41	18	41			0
995328	ESR	≥42,460	≥42,320	0.44	0.45	0	51	9	40			0
995330	ESR	≥42,460	≥42,310	0.39	0.46	0	54	13	32			0
1002831	TE	≥42,460	≥42,260	0.40	0.51	0	38	43	18			0
995331	ESR	≥42,460	≥42,270	0.33	0.43	0	62	9	29			0

Table 6.9 Operational Item Statistics—English Language Arts Grade 8 CBT Administration

ELA Grade 8 Computer-Based Test Administration												
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% Nonscore Codes**
913913	MS	≥45,780	≥45,750	0.46	0.32	0	41	26	33			0
982282	MS	≥45,780	≥45,720	0.31	0.41	0	53	33	15			0
913915	ESR	≥45,780	≥45,710	0.46	0.42	0	50	8	42			0
913917	ESR	≥45,780	≥45,710	0.63	0.46	0	33	10	58			0
913916	ESR	≥45,780	≥45,710	0.47	0.25	0	49	8	43			0
913918	TE	≥45,780	≥45,670	0.43	0.48	0	27	60	13			0
913919O2	CR	≥45,780	≥44,810	0.37	0.81	1	17	28	42	9	2	1
913919O3	CR	≥45,780	≥44,810	0.53	0.79	1	14	25	46	12		1
932831	ESR	≥45,780	≥45,540	0.61	0.54	1	33	11	55			0
932830	MS	≥45,780	≥45,470	0.31	0.31	1	52	34	13			0
955625	TE	≥45,780	≥45,310	0.59	0.38	1	19	43	37			0
932819	ESR	≥45,780	≥45,270	0.60	0.45	1	35	9	55			0
1117994	ESR	≥45,780	≥45,740	0.67	0.38	0	31	5	64			0
1117998	ESR	≥45,780	≥45,700	0.46	0.34	0	48	11	41			0
1117996	ESR	≥45,780	≥45,720	0.76	0.45	0	21	6	73			0
1117999	TE	≥45,780	≥45,700	0.41	0.54	0	56	7	37			0
1118000	ESR	≥45,780	≥45,710	0.28	0.30	0	65	13	22			0
1117995	ESR	≥45,780	≥45,710	0.60	0.35	0	28	24	48			0
1118002	ESR	≥45,780	≥45,710	0.44	0.40	0	47	18	35			0
1118001	TE	≥45,780	≥45,700	0.62	0.51	0	20	36	44			0
1118006O2	CR	≥45,780	≥45,000	0.37	0.81	1	18	33	32	12	4	1
1118006O3	CR	≥45,780	≥45,000	0.48	0.80	1	18	31	37	13		1
1117890	ESR	≥45,780	≥45,700	0.47	0.38	0	41	24	34			0
1117888	ESR	≥45,780	≥45,650	0.51	0.46	0	41	17	42			0
1117892	MS	≥45,780	≥45,670	0.37	0.41	0	41	44	16			0
1117894	TE	≥45,780	≥45,660	0.70	0.51	0	12	35	53			0
1031119	ESR	≥45,780	≥45,660	0.52	0.47	0	43	9	48			0
1031118	TE	≥45,780	≥45,610	0.48	0.45	0	29	45	26			0
1031120	ESR	≥45,780	≥45,660	0.45	0.45	0	47	16	37			0
1031124	TE	≥45,780	≥45,630	0.75	0.47	0	9	31	60			0
1031121	ESR	≥45,780	≥45,630	0.65	0.56	0	29	10	60			0
1031133	MS	≥45,780	≥45,630	0.41	0.48	0	50	18	31			0

Table 6.10 Operational Item Statistics—Mathematics Grade 3 CBT Administration

Mathematics Grade 3 Computer-Based Test Administration														
Item ID	Item Type	Total N	Adj. N	*p*-Value	Pbis	Omit Rate	% at 0	% at 1	% at 2	% at 3	% at 4	% at 5	% at 6	% Nonscore Codes
896678	MC	≥22,100	≥22,080	0.68	0.38	0		15	68	2	15			0
870659	SA	≥22,100	≥21,990	0.50	0.61	0	49	50						0
896771	MC	≥22,100	≥22,070	0.65	0.39	0		3	21	64	11			0
1075004	SA	≥22,100	≥22,050	0.54	0.59	0	21	50	29					0
1114964	MC	≥22,100	≥22,040	0.70	0.33	0		70	7	7	15			0
896881	MS	≥22,100	≥22,060	0.46	0.48	0	54	46						0
1074999	MS	≥22,100	≥22,050	0.26	0.53	0	74	26						0
1075006	MC	≥22,100	≥22,050	0.67	0.44	0		17	8	8	66			0
896778	MC	≥22,100	≥22,040	0.55	0.44	0		7	55	23	15			0
1074997	MC	≥22,100	≥22,040	0.68	0.38	0		68	23	6	4			0
896876	MC	≥22,100	≥22,060	0.71	0.38	0		12	5	13	71			0
896677	SA	≥22,100	≥22,020	0.62	0.38	0	38	62						0
896759	SA	≥22,100	≥22,010	0.48	0.58	0	51	48						0
897734	CR	≥22,100	≥21,170	0.15	0.56	1	67	11	10	3	4			3
979828	CR	≥22,100	≥21,930	0.32	0.62	1	27	56	9	8				0
896896	SA	≥22,100	≥22,060	0.72	0.40	0	28	72						0
896765	MC	≥22,100	≥22,060	0.76	0.44	0		14	4	76	6			0
914028	MC	≥22,100	≥22,050	0.37	0.50	0		35	19	37	9			0
896900	SA	≥22,100	≥22,050	0.50	0.60	0	26	47	27					0
1075167	MC	≥22,100	≥22,060	0.59	0.36	0		25	58	12	5			0
896899	MC	≥22,100	≥22,050	0.20	0.22	0		9	20	21	50			0
896870	MC	≥22,100	≥22,050	0.64	0.33	0		12	5	63	20			0
896676	MC	≥22,100	≥22,050	0.53	0.52	0		16	8	24	52			0
1026036	MC	≥22,100	≥22,050	0.69	0.39	0		68	13	8	10			0
1025975	MPSR	≥22,100	≥22,020	0.44	0.44	0	35	42	23					0
800078	MC	≥22,100	≥22,020	0.55	0.56	0		36	3	55	5			0
896681	MS	≥22,100	≥22,040	0.54	0.57	0	46	54						0
897730	CR	≥22,100	≥21,270	0.31	0.56	1	33	38	23	3				3
898004	CR	≥22,100	≥21,840	0.16	0.59	1	68	20	5	5				0
1114939	SA	≥22,100	≥22,050	0.53	0.57	0	47	53						0
896775	MS	≥22,100	≥22,050	0.19	0.47	0	81	19						0
896878	MC	≥22,100	≥22,050	0.58	0.45	0		6	10	57	26			0
981776	MS	≥22,100	≥22,060	0.49	0.52	0	50	49						0
870695	SA	≥22,100	≥22,050	0.68	0.41	0	32	68						0
981765	MS	≥22,100	≥22,060	0.77	0.41	0	23	77						0
896859	SA	≥22,100	≥22,040	0.49	0.61	0	51	48						0
870678	CR	≥22,100	≥22,030	0.31	0.76	0	34	21	12	13	8	7	6	0
896893	MC	≥22,100	≥22,040	0.54	0.28	0		54	11	18	16			0
914004	SA	≥22,100	≥22,020	0.77	0.49	0	23	76						0
1075200	MC	≥22,100	≥22,020	0.51	0.37	0		14	12	51	22			0
1114945	SA	≥22,100	≥21,990	0.74	0.48	0	26	74						0
1026140	MC	≥22,100	≥22,010	0.77	0.44	0		76	12	9	2			0
981741	CR	≥22,100	≥21,300	0.36	0.47	1	35	29	21	11				3

Table 6.11 Operational Item Statistics—Mathematics Grade 3 PBT Administration

Mathematics Grade 3 Paper-Based Test Administration														
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**	**% Nonscore Codes**
896678	MC	≥27,170	≥26,520	0.73	0.38	0		13	71	1	12			2
870659	SA	≥27,170	≥26,430	0.53	0.58	3	46	51						0
896771	MC	≥27,170	≥26,750	0.71	0.41	1		3	17	70	9			0
1075004	SA	≥27,170	≥27,020	0.57	0.60	1	20	46	33					0
1114964	MC	≥27,170	≥26,940	0.75	0.32	1		74	6	5	14			0
896881	MS	≥27,170	≥26,910	0.55	0.48	1	44	55						0
1074999	MS	≥27,170	≥26,920	0.32	0.56	1	68	31						0
1075006	MC	≥27,170	≥26,590	0.72	0.42	1		15	7	6	70			1
896778	MC	≥27,170	≥26,890	0.58	0.44	1		7	58	21	13			0
1074997	MC	≥27,170	≥26,590	0.72	0.36	1		71	20	4	3			1
896876	MC	≥27,170	≥26,510	0.74	0.39	2		11	4	11	72			0
896677	SA	≥27,170	≥26,430	0.58	0.45	3	41	57						0
896759	SA	≥27,170	≥26,560	0.53	0.57	2	46	52						0
897734	CR	≥27,170	≥26,150	0.22	0.57	3	59	12	13	5	8			1
979828	CR	≥27,170	≥27,020	0.38	0.65	1	22	54	10	13				0
896896	SA	≥27,170	≥26,670	0.75	0.42	2	25	73						0
896765	MC	≥27,170	≥26,960	0.81	0.43	1		11	3	80	5			0
914028	MC	≥27,170	≥26,920	0.40	0.53	1		35	16	40	9			0
896900	SA	≥27,170	≥26,880	0.54	0.60	1	23	46	30					0
1075167	MC	≥27,170	≥26,860	0.62	0.38	1		25	61	9	4			0
896899	MC	≥27,170	≥26,640	0.19	0.19	2		10	19	19	50			0
896870	MC	≥27,170	≥26,870	0.66	0.35	1		10	4	65	20			0
896676	MC	≥27,170	≥26,730	0.59	0.51	2		13	6	22	58			0
1026036	MC	≥27,170	≥26,680	0.74	0.41	2		73	10	6	9			0
1025975	MPSR	≥27,170	≥26,910	0.47	0.50	1	32	41	26					0
800078	MC	≥27,170	≥26,700	0.61	0.58	1		32	3	60	4			0
896681	MS	≥27,170	≥26,920	0.62	0.54	1	38	62						0
897730	CR	≥27,170	≥26,410	0.40	0.55	2	24	34	37	3				0
898004	CR	≥27,170	≥26,830	0.24	0.59	1	58	22	7	11				0
1114939	SA	≥27,170	≥26,220	0.55	0.54	3	44	53						0
896775	MS	≥27,170	≥26,960	0.27	0.54	1	73	27						0
896878	MC	≥27,170	≥26,800	0.59	0.49	1		6	9	58	25			0
981776	MS	≥27,170	≥26,910	0.54	0.50	1	46	53						0
870695	SA	≥27,170	≥26,080	0.70	0.41	4	29	67						0
981765	MS	≥27,170	≥27,000	0.83	0.39	1	16	83						0
896859	SA	≥27,170	≥26,430	0.52	0.57	3	47	50						0
870678	CR	≥27,170	≥26,920	0.40	0.76	1	25	17	12	15	9	11	11	0
896893	MC	≥27,170	≥26,720	0.60	0.26	2		59	9	15	14			0
914004	SA	≥27,170	≥26,500	0.80	0.46	2	20	78						0
1075200	MC	≥27,170	≥26,880	0.54	0.41	1		15	9	54	22			0
1114945	SA	≥27,170	≥25,950	0.78	0.45	4	21	75						0
1026140	MC	≥27,170	≥26,740	0.83	0.41	1		82	9	7	2			0
981741	CR	≥27,170	≥26,760	0.54	0.47	1	19	26	26	27				0

Table 6.12 Operational Item Statistics—Mathematics Grade 4 CBT Administration

Mathematics Grade 4 Computer-Based Test Administration														
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**	**% Nonscore Codes**
1114984	MC	≥48,790	≥48,600	0.73	0.34	0		14	9	72	4			0
1075439	SA	≥48,790	≥48,670	0.37	0.60	0	49	28	23					0
896767	MC	≥48,790	≥48,730	0.59	0.55	0		59	9	13	18			0
1075317	MC	≥48,790	≥48,700	0.46	0.60	0		11	46	10	33			0
981863	MS	≥48,790	≥48,720	0.47	0.30	0	53	47						0
897440	MC	≥48,790	≥48,670	0.48	0.53	0		12	48	7	33			0
1026610	SA	≥48,790	≥48,580	0.37	0.49	0	63	37						0
1115006	SA	≥48,790	≥48,590	0.40	0.60	0	60	40						0
897465	SA	≥48,790	≥48,600	0.52	0.46	0	48	52						0
935180	MS	≥48,790	≥48,700	0.40	0.61	0	60	40						0
897469	MC	≥48,790	≥48,600	0.37	0.28	0		18	37	34	11			0
981883	MS	≥48,790	≥48,650	0.54	0.68	0	46	54						0
897288	MC	≥48,790	≥48,520	0.49	0.57	1		49	10	25	16			0
878671	CR	≥48,790	≥48,450	0.44	0.72	1	26	25	16	15	18			0
981848	CR	≥48,790	≥46,600	0.35	0.67	3	30	39	17	9				2
897442	MS	≥48,790	≥48,760	0.63	0.43	0	37	63						0
1075438	SA	≥48,790	≥48,590	0.45	0.56	0	55	44						0
914074	MC	≥48,790	≥48,750	0.55	0.47	0		10	55	28	7			0
897463	MC	≥48,790	≥48,730	0.67	0.43	0		15	11	66	8			0
897436	MC	≥48,790	≥48,730	0.60	0.27	0		18	14	60	8			0
981855	SA	≥48,790	≥48,650	0.45	0.36	0	55	45						0
981872	MC	≥48,790	≥48,720	0.68	0.43	0		68	11	12	8			0
981885	MC	≥48,790	≥48,720	0.75	0.51	0		8	6	11	75			0
1075170	MPSR	≥48,790	≥48,740	0.43	0.61	0	39	36	25					0
897445	MC	≥48,790	≥48,720	0.64	0.43	0		16	5	64	16			0
981877	MPSR	≥48,790	≥48,710	0.54	0.53	0	25	41	34					0
981875	SA	≥48,790	≥48,630	0.52	0.59	0	48	52						0
1075173	CR	≥48,790	≥47,860	0.22	0.59	1	52	31	12	3				1
1115014	CR	≥48,790	≥47,650	0.31	0.68	1	47	24	13	14				1
1075437	MC	≥48,790	≥48,730	0.79	0.37	0		10	4	79	6			0
1026760	SA	≥48,790	≥48,630	0.63	0.46	0	37	63						0
897310	MC	≥48,790	≥48,720	0.54	0.58	0		24	7	16	54			0
1114977	MS	≥48,790	≥48,720	0.42	0.55	0	58	42						0
1026698	MC	≥48,790	≥48,720	0.67	0.51	0		7	4	22	67			0
897303	MS	≥48,790	≥48,730	0.69	0.53	0	31	69						0
898013	SA	≥48,790	≥48,650	0.36	0.55	0	63	36						0
897287	MS	≥48,790	≥48,730	0.44	0.48	0	56	44						0
981893	MS	≥48,790	≥48,720	0.83	0.39	0	17	83						0
1026634	SA	≥48,790	≥48,700	0.49	0.58	0	51	49						0
914094	SA	≥48,790	≥48,610	0.29	0.57	0	70	29						0
897443	SA	≥48,790	≥48,620	0.33	0.49	0	67	33						0
981832	CR	≥48,790	≥47,360	0.28	0.63	1	54	16	15	13				2
981823	CR	≥48,790	≥47,820	0.25	0.73	1	50	13	10	7	5	5	8	1

Table 6.13 Operational Item Statistics—Mathematics Grade 5 CBT Administration

Mathematics Grade 5 Computer-Based Test Administration														
Item ID	Item Type	Total N	Adj. N	*p*-Value	Pbis	Omit Rate	% at 0	% at 1	% at 2	% at 3	% at 4	% at 5	% at 6	% Nonscore Codes
982523	MC	≥48,170	≥48,130	0.77	0.43	0		6	2	15	77			0
1075851	SA	≥48,170	≥48,010	0.46	0.31	0	54	46						0
903375	MS	≥48,170	≥48,130	0.50	0.52	0	50	50						0
914160	SA	≥48,170	≥48,090	0.76	0.35	0	24	76						0
897988	TE	≥48,170	≥47,980	0.79	0.44	0	21	79						0
800149	MC	≥48,170	≥48,110	0.63	0.59	0		17	10	10	63			0
982488	MC	≥48,170	≥47,940	0.42	0.59	0		25	14	19	42			0
982495	SA	≥48,170	≥48,050	0.63	0.46	0	37	62						0
1075322	SA	≥48,170	≥48,140	0.51	0.61	0	33	31	36					0
1119177	TE	≥48,170	≥48,090	0.38	0.48	0	62	38						0
914578	TE	≥48,170	≥47,890	0.34	0.39	1	65	34						0
898027	MC	≥48,170	≥48,060	0.77	0.37	0		76	6	5	12			0
982492	TE	≥48,170	≥48,010	0.41	0.64	0	59	41						0
902412	CR	≥48,170	≥48,020	0.27	0.61	0	39	24	25	11	1			0
902413	CR	≥48,170	≥47,480	0.22	0.66	1	58	23	11	7				0
1026898	TE	≥48,170	≥48,120	0.53	0.46	0	47	53						0
1026866	MC	≥48,170	≥48,090	0.51	0.38	0		31	9	51	9			0
898169	MC	≥48,170	≥48,110	0.42	0.38	0		34	16	8	42			0
1075859	MS	≥48,170	≥48,110	0.45	0.55	0	55	45						0
982534	MS	≥48,170	≥48,080	0.29	0.53	0	71	29						0
1027100	MC	≥48,170	≥48,110	0.29	0.26	0		20	42	29	9			0
898019	MC	≥48,170	≥48,090	0.77	0.43	0		5	10	77	7			0
1075879	SA	≥48,170	≥48,100	0.45	0.52	0	25	59	16					0
898165	SA	≥48,170	≥48,020	0.57	0.55	0	43	57						0
1119182	SA	≥48,170	≥47,930	0.46	0.47	1	54	45						0
1119175	TE	≥48,170	≥48,090	0.40	0.75	0	47	26	27					0
1026865	TE	≥48,170	≥48,050	0.29	0.58	0	71	29						0
1119213	CR	≥48,170	≥46,900	0.18	0.55	1	56	32	7	2				1
982476	CR	≥48,170	≥47,010	0.20	0.56	1	61	20	11	6				1
898024	SA	≥48,170	≥48,080	0.70	0.47	0	30	70						0
898164	SA	≥48,170	≥48,070	0.51	0.34	0	49	51						0
904183	TE	≥48,170	≥48,100	0.49	0.68	0	51	49						0
898147	MC	≥48,170	≥48,060	0.64	0.40	0		9	10	16	64			0
982544	MC	≥48,170	≥48,070	0.48	0.44	0		9	11	33	48			0
1027055	MC	≥48,170	≥48,080	0.49	0.49	0		20	15	17	49			0
904184	TE	≥48,170	≥48,060	0.57	0.67	0	43	57						0
914156	MS	≥48,170	≥48,070	0.15	0.42	0	85	15						0
1075855	MS	≥48,170	≥48,080	0.36	0.63	0	64	36						0
898158	MC	≥48,170	≥48,080	0.60	0.44	0		15	60	14	10			0
1026930	MC	≥48,170	≥48,070	0.38	0.45	0		38	6	32	23			0
914579	SA	≥48,170	≥48,070	0.71	0.52	0	29	71						0
902411	CR	≥48,170	≥48,050	0.35	0.68	0	36	30	26	8				0
982497	CR	≥48,170	≥47,840	0.33	0.77	1	38	12	11	13	12	8	6	0

Table 6.14 Item Statistics—Mathematics Grade 6 Computer-Based Test Administration

Mathematics Grade 6 Computer-Based Test Administration														
Item ID	Item Type	Total N	Adj. N	*p*-Value	Pbis	Omit Rate	% at 0	% at 1	% at 2	% at 3	% at 4	% at 5	% at 6	% Nonscore Codes
1075675	MC	≥48,240	≥48,190	0.74	0.43	0		74	12	12	3			0
981980	TE	≥48,240	≥48,100	0.54	0.58	0	46	54						0
1075866	MC	≥48,240	≥48,150	0.73	0.42	0		8	73	12	6			0
914267	SA	≥48,240	≥47,780	0.32	0.48	1	68	31						0
1116275	MC	≥48,240	≥48,170	0.52	0.35	0		5	27	52	16			0
1116259	SA	≥48,240	≥47,930	0.55	0.65	1	45	55						0
1075955	MC	≥48,240	≥47,970	0.43	0.48	1		13	23	20	43			0
1075940	SA	≥48,240	≥48,050	0.32	0.54	0	50	35	15					0
1075945	MC	≥48,240	≥48,110	0.49	0.36	0		17	48	11	23			0
981997	TE	≥48,240	≥48,120	0.66	0.53	0	34	65						0
1076105	MC	≥48,240	≥48,140	0.51	0.59	0		32	8	9	51			0
981971	SA	≥48,240	≥48,020	0.53	0.45	0	47	52						0
800191	MC	≥48,240	≥47,940	0.55	0.42	1		11	55	16	18			0
1027521	TE	≥48,240	≥47,980	0.32	0.63	1	68	31						0
1075676	SA	≥48,240	≥47,830	0.45	0.64	1	54	45						0
981977	MC	≥48,240	≥47,890	0.51	0.48	1		50	22	20	6			0
1027593	SA	≥48,240	≥47,270	0.17	0.44	2	82	16						0
981969	SA	≥48,240	≥47,260	0.54	0.48	2	45	53						0
914292	TE	≥48,240	≥47,530	0.54	0.59	1	45	54						0
901540	MC	≥48,240	≥48,180	0.67	0.40	0		7	67	7	19			0
868816	TE	≥48,240	≥48,160	0.55	0.62	0	45	55						0
900537	SA	≥48,240	≥48,080	0.56	0.40	0	44	56						0
1075880	SA	≥48,240	≥48,110	0.38	0.61	0	39	46	15					0
903093	SA	≥48,240	≥48,120	0.80	0.47	0	20	80						0
903080	SA	≥48,240	≥47,980	0.23	0.52	1	76	23						0
981964	CR	≥48,240	≥47,070	0.30	0.71	1	54	7	11	18	8			1
1116287	MC	≥48,240	≥48,140	0.38	0.46	0		38	44	13	4			0
904180	TE	≥48,240	≥48,120	0.29	0.61	0	54	34	12					0
902993	CR	≥48,240	≥47,490	0.36	0.52	1	26	41	30	2				1
1116301	CR	≥48,240	≥48,090	0.25	0.72	0	40	23	10	12	9	3	3	0
981958	CR	≥48,240	≥45,800	0.16	0.62	3	68	14	7	5				2
903094	MC	≥48,240	≥48,170	0.51	0.46	0		24	15	11	51			0
901546	MS	≥48,240	≥48,160	0.53	0.56	0	46	53						0
981982	SA	≥48,240	≥48,140	0.58	0.67	0	24	37	39					0
901552	SA	≥48,240	≥48,030	0.24	0.50	0	75	24						0
1027312	MS	≥48,240	≥48,170	0.30	0.31	0	70	30						0
982031	MC	≥48,240	≥48,160	0.39	0.43	0		11	39	19	31			0
903101	MC	≥48,240	≥48,150	0.40	0.42	0		40	37	13	10			0
901535	MS	≥48,240	≥48,150	0.25	0.48	0	74	25						0
903084	MS	≥48,240	≥48,160	0.33	0.55	0	67	33						0
914242	CR	≥48,240	≥47,010	0.23	0.60	1	53	31	3	10				2
981965	CR	≥48,240	≥48,140	0.37	0.77	0	29	30	17	12	11			0
981959	CR	≥48,240	≥46,840	0.31	0.56	1	43	32	6	15				2

Table 6.15 Item Statistics—Mathematics Grade 7 Computer-Based Test Administration

Mathematics Grade 7 Computer-Based Test Administration														
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**	**% Nonscore Codes**
1116314	MC	≥48,790	≥48,690	0.62	0.46	0		10	11	17	62			0
914341	MC	≥48,790	≥48,730	0.76	0.40	0		10	76	6	8			0
899863	MS	≥48,790	≥48,660	0.39	0.57	0	61	39						0
982970	TE	≥48,790	≥48,680	0.22	0.51	0	78	22						0
897992	MC	≥48,790	≥48,700	0.64	0.44	0		8	64	20	8			0
1027689	TE	≥48,790	≥48,670	0.24	0.52	0	76	24						0
1116319	MS	≥48,790	≥48,690	0.47	0.64	0	53	47						0
1027870	SA	≥48,790	≥48,480	0.35	0.53	1	65	35						0
1024431	TE	≥48,790	≥48,660	0.40	0.50	0	38	44	18					0
898450	SA	≥48,790	≥48,490	0.35	0.50	1	64	35						0
899319	MC	≥48,790	≥48,670	0.51	0.41	0		19	22	51	8			0
899320	SA	≥48,790	≥47,950	0.25	0.67	2	74	24						0
982983	MC	≥48,790	≥48,610	0.64	0.47	0		15	63	14	7			0
1027651	SA	≥48,790	≥47,550	0.28	0.59	3	71	27						0
983014	SA	≥48,790	≥48,160	0.33	0.39	1	66	32						0
1116315	MC	≥48,790	≥48,380	0.54	0.37	1		9	20	53	17			0
900522	MC	≥48,790	≥48,280	0.27	0.30	1		27	27	29	16			0
1116302	SA	≥48,790	≥47,440	0.23	0.61	3	75	23						0
1116320	MC	≥48,790	≥48,180	0.34	0.27	1		21	32	34	13			0
1116330	MC	≥48,790	≥48,740	0.47	0.27	0		4	11	38	47			0
896994	SA	≥48,790	≥48,210	0.36	0.66	1	64	35						0
1024536	TE	≥48,790	≥48,560	0.33	0.63	0	66	33						0
1027725	SA	≥48,790	≥48,680	0.37	0.62	0	44	36	19					0
1024433	MC	≥48,790	≥48,690	0.47	0.56	0		22	13	18	47			0
899315	MS	≥48,790	≥48,660	0.29	0.46	0	70	29						0
982919	CR	≥48,790	≥48,700	0.24	0.77	0	32	36	13	6	5	4	3	0
982951	TE	≥48,790	≥48,640	0.44	0.68	0	41	29	30					0
983507	CR	≥48,790	≥47,330	0.16	0.62	2	69	12	14	2				1
982942	TE	≥48,790	≥48,530	0.20	0.35	1	79	20						0
983508	CR	≥48,790	≥45,350	0.17	0.64	4	66	10	12	4				3
982928	CR	≥48,790	≥44,380	0.17	0.71	5	64	9	5	7	5			4
982961	MS	≥48,790	≥48,720	0.20	0.51	0	80	20						0
914313	MC	≥48,790	≥48,690	0.40	0.27	0		15	34	11	40			0
915699	TE	≥48,790	≥48,690	0.47	0.57	0	53	46						0
1024539	MC	≥48,790	≥48,680	0.50	0.55	0		50	15	13	22			0
900173	MC	≥48,790	≥48,710	0.53	0.37	0		53	14	18	14			0
900180	SA	≥48,790	≥48,630	0.26	0.65	0	73	26						0
1075674	MPSR	≥48,790	≥48,720	0.53	0.57	0	27	39	34					0
982958	SA	≥48,790	≥48,580	0.35	0.60	0	65	34						0
1024541	SA	≥48,790	≥48,570	0.28	0.59	0	72	28						0
902989	CR	≥48,790	≥48,630	0.48	0.67	0	19	44	11	25				0
982925	CR	≥48,790	≥47,800	0.29	0.61	1	40	14	35	6	3			1
982929	CR	≥48,790	≥46,680	0.24	0.66	2	55	19	14	7				3

Table 6.16 Item Statistics—Mathematics Grade 8 Computer-Based Test Administration

Mathematics Grade 8 Computer-Based Test Administration														
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**	**% Nonscore Codes**
897447	MC	≥44,120	≥44,070	0.75	0.28	0		11	75	9	5			0
900515	TE	≥44,120	≥44,010	0.30	0.27	0	70	30						0
898442	TE	≥44,120	≥44,070	0.55	0.36	0	45	55						0
870345	MPSR	≥44,120	≥44,090	0.46	0.45	0	28	52	21					0
1075642	MC	≥44,120	≥44,030	0.47	0.49	0		47	12	23	17			0
903087	SA	≥44,120	≥43,460	0.18	0.58	1	81	18						0
983065	MC	≥44,120	≥43,980	0.39	0.42	0		39	33	16	12			0
1022886	MPSR	≥44,120	≥44,070	0.36	0.33	0	42	45	13					0
914425	MC	≥44,120	≥43,990	0.60	0.44	0		22	60	7	11			0
1023523	TE	≥44,120	≥43,970	0.72	0.40	0	28	71						0
897069	MC	≥44,120	≥44,040	0.56	0.27	0		56	12	24	7			0
983028	SA	≥44,120	≥43,200	0.16	0.58	2	83	15						0
1027999	SA	≥44,120	≥43,350	0.24	0.55	2	75	23						0
1022881	MS	≥44,120	≥43,930	0.33	0.53	0	67	33						0
1022882	MC	≥44,120	≥43,920	0.41	0.39	0		13	41	35	11			0
901197	MC	≥44,120	≥43,960	0.45	0.53	0		45	15	21	19			0
1023326	TE	≥44,120	≥43,840	0.39	0.56	1	60	39						0
898148	MC	≥44,120	≥43,910	0.45	0.39	0		9	32	15	44			0
1117491	TE	≥44,120	≥43,830	0.39	0.41	1	61	39						0
1023521	MC	≥44,120	≥43,830	0.51	0.39	1		21	20	51	8			0
1117513	SA	≥44,120	≥43,850	0.32	0.54	1	67	32						0
983090	MC	≥44,120	≥44,030	0.52	0.44	0		24	52	17	7			0
878967	SA	≥44,120	≥43,860	0.22	0.57	1	77	22						0
900500	TE	≥44,120	≥44,030	0.29	0.59	0	52	36	11					0
984139	CR	≥44,120	≥43,650	0.16	0.58	1	57	11	17	9	2	1	1	0
1117482	MC	≥44,120	≥44,050	0.46	0.30	0		19	26	46	9			0
1075648	TE	≥44,120	≥44,050	0.34	0.59	0	48	36	16					0
901194	MS	≥44,120	≥44,020	0.18	0.28	0	82	18						0
1027981	CR	≥44,120	≥43,550	0.18	0.74	1	71	13	7	8				0
1023323	CR	≥44,120	≥42,030	0.16	0.49	3	63	14	11	4	3			2
984137	CR	≥44,120	≥40,640	0.12	0.44	4	70	13	9	1				4
914421	MC	≥44,120	≥44,020	0.40	0.49	0		15	28	17	40			0
897073	SA	≥44,120	≥43,840	0.12	0.51	1	88	12						0
914408	TE	≥44,120	≥43,980	0.19	0.55	0	81	19						0
904551	MC	≥44,120	≥44,020	0.51	0.43	0		14	16	18	51			0
898438	MS	≥44,120	≥44,030	0.39	0.50	0	61	39						0
878742	MC	≥44,120	≥44,020	0.49	0.35	0		49	17	23	11			0
1117519	CR	≥44,120	≥44,020	0.21	0.54	0	47	46	5	2				0
983050	SA	≥44,120	≥43,990	0.25	0.54	0	61	28	11					0
899327	CR	≥44,120	≥42,070	0.26	0.64	2	44	25	11	13	3			2
899328	CR	≥44,120	≥40,700	0.10	0.58	4	76	8	5	3				4

These item level statistics are reviewed at the beginning of the operational analyses process to ensure that items are unflawed, and a careful quality control review is given to determine that the answer key is correct.

A multiple-choice (MC) item is reviewed during the key check process if

- it has a *p*-value less than 0.25 or more than .95,
- greater number of high-performing students (top 20%) choosing a distractor than are choosing the key,
- the item-total correlation of the keyed response is less than 0.20,
- any of the incorrect answer options yields a positive distractor-total correlation, or
- the percentage of students omitting or not reaching each item is 5 or greater.

Other types of autoscored items are also flagged during the key check for review if

- they have a *p*-value less than 0.30 or more than .80,
- the percentage of students who reached any possible score point is less than 3,
- the item-total correlation is less than 0.30, or
- the flagging criteria for omit item is 15%.

6.3 Item Response Theory

Item parameters for items included in the ELA and mathematics tests were estimated using a marginal maximum-likelihood (MML) procedure and the 2-parameter logistic (2PL) model for MC items and the generalized partial credit (GPC) model (Muraki, 1992) for non-MC items. Under the 2PL model, the probability that a student with a trait or scale score of θ will respond correctly to MC item *j* is

$$P_j(\theta) = 1/[1 + exp(-1.7a_j(\theta - b_j))].$$

In the equation, a_j is the item discrimination and b_j is the item difficulty. Under the GPC model, the probability that a student with a trait or scale score of θ will respond in category *x* to partial-credit item *j* is

$$P_{jx}(\theta) = \exp\left[\sum_{k=0}^{x}(Z_{jk}(\theta))\right] / \sum_{h=0}^{m_i}\exp\left[\sum_{k=0}^{x}(Z_{jk}(\theta))\right],$$

$$\text{where } z_{jk}(\theta) = Da_j(\theta - b_j + d_{jx}),$$

where d_{jx} is the relative difficulty of score category *x* of item *j*.

The software IRTPRO (Cai, Thissen, & du Toit, 2011) was used for the IRT calibrations. IRTPRO is a multipurpose program that implements a variety of IRT models associated with mixed-item formats and associated statistics. IRTPRO has been used to calibrate large data sets, such as those of PARCC assessments. The program implements MML estimation techniques for items and MLE estimation of theta.

This section describes the calibration sample in adherence to Standard 1.8 of the AERA, APA, & NCME (2014) *Standards for Educational and Psychological Testing*. Standard 1.8 states the following:

> The composition of any sample of test takers from which validity evidence is obtained should be described in as much detail as is practical and permissible, including major relevant socio-demographic and developmental characteristics (25).

All student data available at the time of calibration was used for the grade 3 PBT and grades 3 to 8 CBT calibration, resulting in a near-census data file. Tables 6.17 and 6.18 show the representativeness of the calibration samples compared to the census data. These tables demonstrate that the calibration sample was representative of the state. Grade 3 includes both CBT and PBT students.

Table 6.17 Summary of Calibration and Census Data: English Language Arts

Calibration and Census Data: English Language Arts						
		Calibration Sample		Census Data		
Grade		N	%	N	%	Census % - Calib %
3	All Students	≥49,550	100.00%	≥49,550	100.00%	0.00%
	Gender					
	Male	≥25,240	50.95%	≥25,240	50.95%	0.00%
	Female	≥24,300	49.05%	≥24,300	49.05%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥5,280	10.66%	≥5,280	10.67%	0.00%
	American Indian or Alaska Native	≥270	0.55%	≥270	0.55%	0.00%
	Asian	≥750	1.53%	≥750	1.53%	0.00%
	Black or African American	≥20,420	41.22%	≥20,420	41.21%	-0.01%
	Native Hawaiian or Other Pacific	≥50	0.11%	≥50	0.11%	0.00%
	White	≥20,770	41.91%	≥20,770	41.92%	0.00%
	Two or More Races	≥1,960	3.96%	≥1,960	3.97%	0.01%
4	All Students	≥49,070	100.00%	≥49,020	100.00%	0.00%
	Gender					
	Male	≥25,090	51.13%	≥25,070	51.13%	0.00%
	Female	≥23,980	48.87%	≥23,950	48.87%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥5,180	10.56%	≥5,170	10.56%	0.00%
	American Indian or Alaska Native	≥280	0.57%	≥280	0.58%	0.00%
	Asian	≥830	1.70%	≥830	1.70%	0.00%
	Black or African American	≥20,210	41.20%	≥20,180	41.17%	-0.03%
	Native Hawaiian or Other Pacific	≥40	0.08%	≥40	0.08%	0.00%
	White	≥20,600	41.99%	≥20,600	42.02%	0.03%
	Two or More Races	≥1,880	3.84%	≥1,880	3.84%	0.00%
5	All Students	≥48,570	100.00%	≥48,510	100.00%	0.00%
	Gender					
	Male	≥24,850	51.17%	≥24,820	51.16%	-0.01%
	Female	≥23,720	48.83%	≥23,690	48.84%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥5,190	10.70%	≥5,190	10.71%	0.01%
	American Indian or Alaska Native	≥250	0.53%	≥250	0.53%	0.00%
	Asian	≥810	1.68%	≥810	1.68%	0.00%
	Black or African American	≥20,390	41.99%	≥20,370	42.00%	0.01%
	Native Hawaiian or Other Pacific	≥30	0.07%	≥30	0.07%	0.00%
	White	≥20,170	41.54%	≥20,130	41.51%	-0.03%
	Two or More Races	≥1,680	3.46%	≥1,680	3.47%	0.00%

Calibration and Census Data: English Language Arts						
		Calibration Sample		Census Data		
Grade		N	%	N	%	Census % - Calib %
6	All Students	≥43,640	100.00%	≥43,610	100.00%	0.00%
	Gender					
	Male	≥22,350	51.23%	≥22,330	51.22%	-0.01%
	Female	≥21,280	48.77%	≥21,270	48.78%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥4,720	10.84%	≥4,720	10.84%	0.01%
	American Indian or Alaska Native	≥200	0.47%	≥200	0.47%	0.00%
	Asian	≥700	1.62%	≥700	1.62%	0.00%
	Black or African American	≥18,440	42.25%	≥18,410	42.22%	-0.04%
	Native Hawaiian or Other Pacific	≥40	0.09%	≥40	0.09%	0.00%
	White	≥18,000	41.25%	≥18,000	41.27%	0.03%
	Two or More Races	≥1,490	3.43%	≥1,490	3.43%	0.00%
7	All Students	≥42,710	100.00%	≥42,710	100.00%	0.00%
	Gender					
	Male	≥21,940	51.38%	≥21,940	51.38%	0.00%
	Female	≥20,760	48.62%	≥20,760	48.62%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥4,880	11.42%	≥4,880	11.43%	0.01%
	American Indian or Alaska Native	≥210	0.50%	≥210	0.50%	0.00%
	Asian	≥720	1.69%	≥720	1.69%	0.00%
	Black or African American	≥17,770	41.60%	≥17,760	41.60%	0.00%
	Native Hawaiian or Other Pacific	≥30	0.07%	≥30	0.07%	0.00%
	White	≥17,610	41.23%	≥17,610	41.23%	0.00%
	Two or More Races	≥1,460	3.42%	≥1,460	3.42%	0.00%
8	All Students	≥46,130	100.00%	≥46,110	100.00%	0.00%
	Gender					
	Male	≥23,350	50.62%	≥23,340	50.62%	0.00%
	Female	≥22,780	49.38%	≥22,770	49.38%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥4,850	10.53%	≥4,850	10.53%	0.00%
	American Indian or Alaska Native	≥220	0.48%	≥220	0.48%	0.00%
	Asian	≥730	1.60%	≥730	1.60%	0.00%
	Black or African American	≥19,750	42.82%	≥19,740	42.80%	0.02%
	Native Hawaiian or Other Pacific	≥40	0.09%	≥40	0.09%	0.00%
	White	≥18,920	41.02%	≥18,920	41.03%	0.01%
	Two or More Races	≥1,590	3.46%	≥1,590	3.46%	0.00%

Table 6.18 Summary of Calibration and Census Data: Mathematics

Calibration and Census Data: Mathematics						
		Calibration Sample		Census Data		
Grade		N	%	N	%	Census % - Calib %
3	All Students	≥49,515	100.00%	≥49,500	100.00%	0.00%
	Gender					
	Male	≥25,217	50.93%	≥25,210	50.92%	-0.01%
	Female	≥24,297	49.07%	≥24,290	49.07%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥5,190	10.48%	≥5,180	10.48%	0.00%
	American Indian or Alaska Native	≥273	0.55%	≥270	0.55%	0.00%
	Asian	≥762	1.54%	≥760	1.54%	0.00%
	Black or African American	≥20,477	41.36%	≥20,470	41.35%	-0.01%
	Native Hawaiian or Other Pacific	≥54	0.11%	≥50	0.11%	0.00%
	White	≥20,770	41.95%	≥20,760	41.94%	-0.01%
	Two or More Races	≥1,963	3.96%	≥1,960	3.97%	0.00%
4	All Students	≥48,979	100.00%	≥48,930	100.00%	0.00%
	Gender					
	Male	≥25,039	51.12%	≥25,020	51.13%	0.01%
	Female	≥23,940	48.88%	≥23,910	48.87%	-0.01%
	Race Ethnicity					
	Hispanic/Latino	≥5,097	10.41%	≥5,090	10.41%	0.01%
	American Indian or Alaska Native	≥282	0.58%	≥280	0.58%	0.00%
	Asian	≥833	1.70%	≥830	1.70%	0.00%
	Black or African American	≥20,211	41.26%	≥20,170	41.22%	-0.04%
	Native Hawaiian or Other Pacific	≥40	0.08%	≥40	0.08%	0.00%
	White	≥20,609	42.08%	≥20,600	42.11%	0.03%
	Two or More Races	≥1,882	3.84%	≥1,880	3.84%	0.00%
5	All Students	≥48,387	100.00%	≥48,360	100.00%	0.00%
	Gender					
	Male	≥24,751	51.15%	≥24,740	51.15%	0.00%
	Female	≥23,636	48.85%	≥23,620	48.85%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥5,099	10.54%	≥5,090	10.54%	0.00%
	American Indian or Alaska Native	≥258	0.53%	≥250	0.53%	0.00%
	Asian	≥810	1.67%	≥810	1.67%	0.00%
	Black or African American	≥20,374	42.11%	≥20,360	42.10%	-0.01%
	Native Hawaiian or Other Pacific	≥34	0.07%	≥30	0.07%	0.00%
	White	≥20,123	41.59%	≥20,110	41.59%	0.00%
	Two or More Races	≥1,675	3.46%	≥1,670	3.46%	0.00%

Calibration and Census Data: Mathematics						
		Calibration Sample		Census Data		
Grade		N	%	N	%	Census % - Calib %
6	All Students	≥48,500	100.00%	≥48,470	100.00%	0.00%
	Gender					
	Male	≥24,800	51.13%	≥24,780	51.13%	0.00%
	Female	≥23,700	48.87%	≥23,680	48.87%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥4,920	10.15%	≥4,920	10.15%	0.01%
	American Indian or Alaska Native	≥250	0.53%	≥250	0.53%	0.00%
	Asian	≥740	1.53%	≥740	1.53%	0.00%
	Black or African American	≥20,530	42.33%	≥20,500	42.30%	-0.03%
	Native Hawaiian or Other Pacific	≥40	0.08%	≥40	0.08%	0.00%
	White	≥20,310	41.88%	≥20,310	41.91%	0.03%
	Two or More Races	≥1,670	3.46%	≥1,670	3.46%	0.00%
7	All Students	≥49,040	100.00%	≥49,040	100.00%	0.00%
	Gender					
	Male	≥25,260	51.52%	≥25,260	51.52%	0.00%
	Female	≥23,770	48.48%	≥23,770	48.48%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥5,120	10.44%	≥5,120	10.44%	0.00%
	American Indian or Alaska Native	≥270	0.56%	≥270	0.56%	0.00%
	Asian	≥790	1.62%	≥790	1.62%	0.00%
	Black or African American	≥20,600	42.01%	≥20,600	42.01%	0.00%
	Native Hawaiian or Other Pacific	≥30	0.07%	≥30	0.07%	0.00%
	White	≥20,510	41.83%	≥20,510	41.83%	0.00%
	Two or More Races	≥1,670	3.41%	≥1,670	3.41%	0.00%
8	All Students	≥44,400	100.00%	≥44,400	100.00%	0.00%
	Gender					
	Male	≥22,560	50.81%	≥22,560	50.81%	0.00%
	Female	≥21,840	49.19%	≥21,840	49.19%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥4,430	9.98%	≥4,430	9.98%	0.00%
	American Indian or Alaska Native	≥260	0.59%	≥260	0.59%	0.00%
	Asian	≥510	1.16%	≥510	1.16%	0.00%
	Black or African American	≥20,300	45.73%	≥20,300	45.73%	0.00%
	Native Hawaiian or Other Pacific	≥30	0.08%	≥30	0.08%	0.00%
	White	≥17,360	39.11%	≥17,360	39.11%	0.00%
	Two or More Races	≥1,470	3.33%	≥1,470	3.33%	0.00%

6.4 Calibration and Linking

All 2023 LEAP 2025 item calibration and linking were performed based on IRT. The calibration and linking methodology used for the Spring 2023 LEAP 2025 administration closely followed most of the PARCC methods referenced in the PARCC document *Final Technical Report for 2015 Administration*. To maintain comparability to PARCC, the 2PL/GPC IRT model was applied to item calibration using the software IRTPRO (Cai et al., 2011). To avoid local independence between traits, the writing traits written expression (WE) and written knowledge and use of language (WKL) were separately calibrated using the sparse matrix method.

The Stocking & Lord (1983) procedure was applied using the transformation and scaling software STUIRT (Kim & Kolen, 2004), which can be downloaded at https://www.education.uiowa.edu/centers/casma/computer-programs#c0748e48-f88c-6551-b2b8-ff00000648cd. PARCC scale score transformation constants for the PARCC 2016 baseline scale were used to generate final scoring tables. All IRTPRO and STUIRT command files were prepared following PARCC examples.

Descriptions of the PARCC calibration and equating approach can be found in the PARCC documents *Final Technical Report for 2015 Administration* and *Final Technical Report for 2016 Administration*.

There were two test forms, CBT and PBT, for the 2023 LEAP 2025 grade 3 ELA and mathematics assessments. Only CBT forms were administered for the grades 4 through 8 ELA and mathematics assessments. In general, a school administered the same test mode for ELA and mathematics. Table 6.19 summarizes the student count and item count by test mode for each grade and content area.

The following two steps were taken to place the 2023 LEAP 2025 tests on the LEAP 2025 scale, which are on the 2016 PARCC baseline scale:

1. Calibrate the 2023 LEAP 2025 tests.
2. Link 2023 LEAP 2025 tests, to the LEAP 2025 scale under the non-equivalent common item design.

PARCC established a new baseline scale using 2016 PARCC spring tests. The 2016 and 2017 LEAP 2025 tests were directly linked to this new PARCC 2016 baseline scale using PARCC item parameters as anchor item parameters. Therefore, LEAP 2016 and 2017 were placed on the PARCC scale. Since the 2016 and 2017 LEAP 2025 tests were calibrated with Louisiana students, the scale for these tests will be referred to as the LEAP 2025 scale, although its scale was placed on PARCC scales built with PARCC associated states' data. The 2018 LEAP 2025 tests were equated to the 2017 LEAP 2025 tests using the anchor item parameters of the 2017 LEAP 2025 tests. The 2023 LEAP 2025 forms were linked to the LEAP 2025 scale using LEAP items, which were administered in LEAP 2025 forms in 2016-2019, 2021, and 2022 as anchors by the Stocking & Lord procedure. Since the 2023 anchor items are on the PARCC scale, the 2023 LEAP 2025 forms continue to be considered on the PARCC scale.

6.4.1 Calibration of the 2023 LEAP 2025 Tests

For 2023 LEAP 2025 item calibration, the 2PL/GPC IRT model was applied to the Louisiana students' calibration samples using the software IRTPRO (Cai et al., 2011). Table 6.19 shows the number of students in the calibration samples and number of calibration items by mode. About 55% of grade 3 students took the PBT, and about 45% of grade 3 students took the CBT. More students in grades 6, 7, and 8 took the mathematics test than the ELA tests due to some schools voluntarily selecting to administer the Innovative Assessment Program (IAP) instead of the ELA LEAP 2025 test. More students in grade 8 took the ELA assessment than the mathematics assessment because high-performing students could take the LEAP 2025 HS Algebra I test instead of the mathematics grade 8 test. For ELA, reading items (RL/RI) in writing prompts

are not counted in the N-Items columns because calibration does not include reading item scores; it only includes WE item scores. A reading item score and a WE item score for the same writing prompt are the same. There were between 24 and 32 ELA items and between 41 and 43 mathematics items across grades.

Table 6.19 Summary of Student Count in Calibration Sample and Item Count by Test Mode

		N			Percentage		N-Items	
Content	**Grade**	**All**	**CBT**	**PBT**	**CBT**	**PBT**	**CBT**	**PBT**
ELA	3	≥49,550	≥22,350	≥27,190	45.12	54.88	24	24
	4	≥49,070	≥49,070	*	100.00	*	30	*
	5	≥48,570	≥48,570	*	100.00	*	28	*
	6	≥43,640	≥43,640	*	100.00	*	32	*
	7	≥42,710	≥42,710	*	100.00	*	32	*
	8	≥46,130	≥46,130	*	100.00	*	32	*
Mathematics	3	≥49,510	≥22,340	≥27,170	45.12	54.88	43	43
	4	≥48,970	≥48,970	*	100.00	*	43	*
	5	≥48,380	≥48,380	*	100.00	*	43	*
	6	≥48,500	≥48,500	*	100.00	*	43	*
	7	≥49,040	≥49,040	*	100.00	*	43	*
	8	≥44,400	≥44,400	*	100.00	*	41	*

** Grades 4–8 did not have a PBT form.*

6.4.1.1. Concurrent Calibration for PBT and CBT

For the 2023 LEAP 2025 calibration, CBT and PBT were combined and calibrated together for grade 3 based on mode effect study (section 10.4). A DIF analysis between CBT and PBT was performed for grade 3. Mantel-Haenszel (MH) DIF statistic were calculated for MC items and for dichotomously-scored constructed-response items, and the standardization DIF (Dorans & Schmitt, 1991; Zwick, Thayer & Mazzeo, 1997; Dorans, 2013) was applied to polytomously scored constructed-response items in conjunction with the Mantel chi-square statistic (Mantel, 1963; Mantel & Haenszel, 1959). Items were assigned severity classifications based on National Assessment of Educational Progress (NAEP) and Educational Testing Service (ETS) guidelines. Only |C| classifications were flagged following PARCC rules. Items with |A| or |B| classifications were considered as mode-neutral items and treated as common items across modes. Items with |C| classifications were treated as unique items across forms. DRC and LDOE content experts were asked to review the items with |C| classifications. One item in grade 3 mathematics was determined to have mode effect. Therefore, this mode-effect item was separately calibrated by test mode and other PBT and CBT items were concurrently calibrated. A separate scoring table was generated for each PBT and CBT form.

6.4.1.2. Separate Calibration for ELA Prose Constructed-Response Tasks

To address the issue of local independence for ELA prose-constructed response (PCR) tasks, the sparse matrix method was applied for grades 3 to 8. Each ELA test consisted of two PCR tasks; each task had a written expression (WE) and a written knowledge and use of the language (WKL) trait. As can be seen in Table 6.20, a single calibration was performed for grades 3 to 8 by randomly splitting the students into two groups. Almost half of the data set included responses to other items and responses to two WE traits, and the other calibration data set included the same responses to other items and responses to two WKL traits. Therefore, WE item parameters were estimated using the responses from the first group and WKL item parameters were estimated using the responses from the second group. Because these two sets of item responses were

calibrated together, there is only one unique set of item parameters for each item. PARCC took this sparse matrix approach for all grades.

Table 6.20 Calibration Data Structure for ELA WE and WKL Traits with Sparse Matrix

Group	Other Items	WE	WKL
I	XXXXXXXX	XX	
II	XXXXXXXX		XX

6.4.1.3. IRT Item Fit

The usefulness of IRT models is dependent on the extent to which they effectively reflect the data. Hambleton, Swaminathan, and Rogers (1991) explain, "The advantages of item response models can be obtained only when the fit between the model and the test data of interest is satisfactory. A poorly fitting IRT model will not yield invariant item and ability parameters" (p. 53).

It is important to note that while items may be flagged for misfit, these flags may not be of practical importance. Misfitting items that have content validity are often retained for use in one assessment and monitored over a period of usage. A large number of misfitting items in an assessment would indicate that caution should be exercised in the interpretation of the overall score.

After convergence was achieved for each IRT data set, an item characteristic curve (ICC) for each item was plotted with empirical students' performances from theta ability -4 to 4. One item in grade 8 mathematics was suppressed from calibration and scoring due to poor fit. One additional item in grade 3 mathematics was removed from the anchor set used to link the 2022 form to the LEAP 2025 scale. Additionally, four items across the mathematics grades were removed from the anchor sets used to establish comparability of the 2023 forms to the existing PARCC scale. No ELA items exhibited item misfit. The fit plots for the items removed from calibration are seen in Figure 6.1. Figure 6.2 displays the fit plots for the items removed from the LEAP 2025 anchor set.

Figure 6.1 Item Fit Plots of Items Removed from Calibration and Scoring

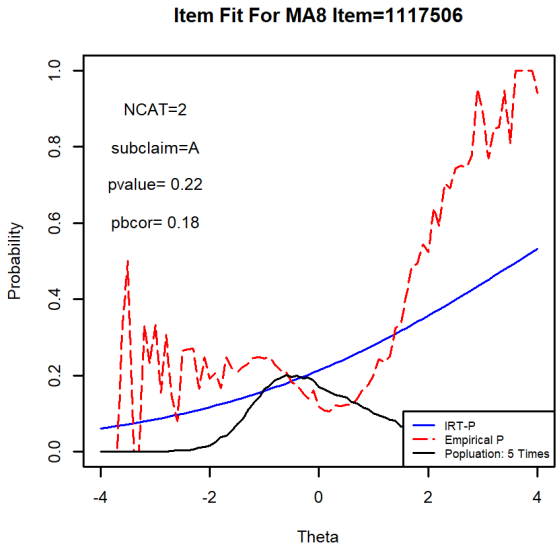
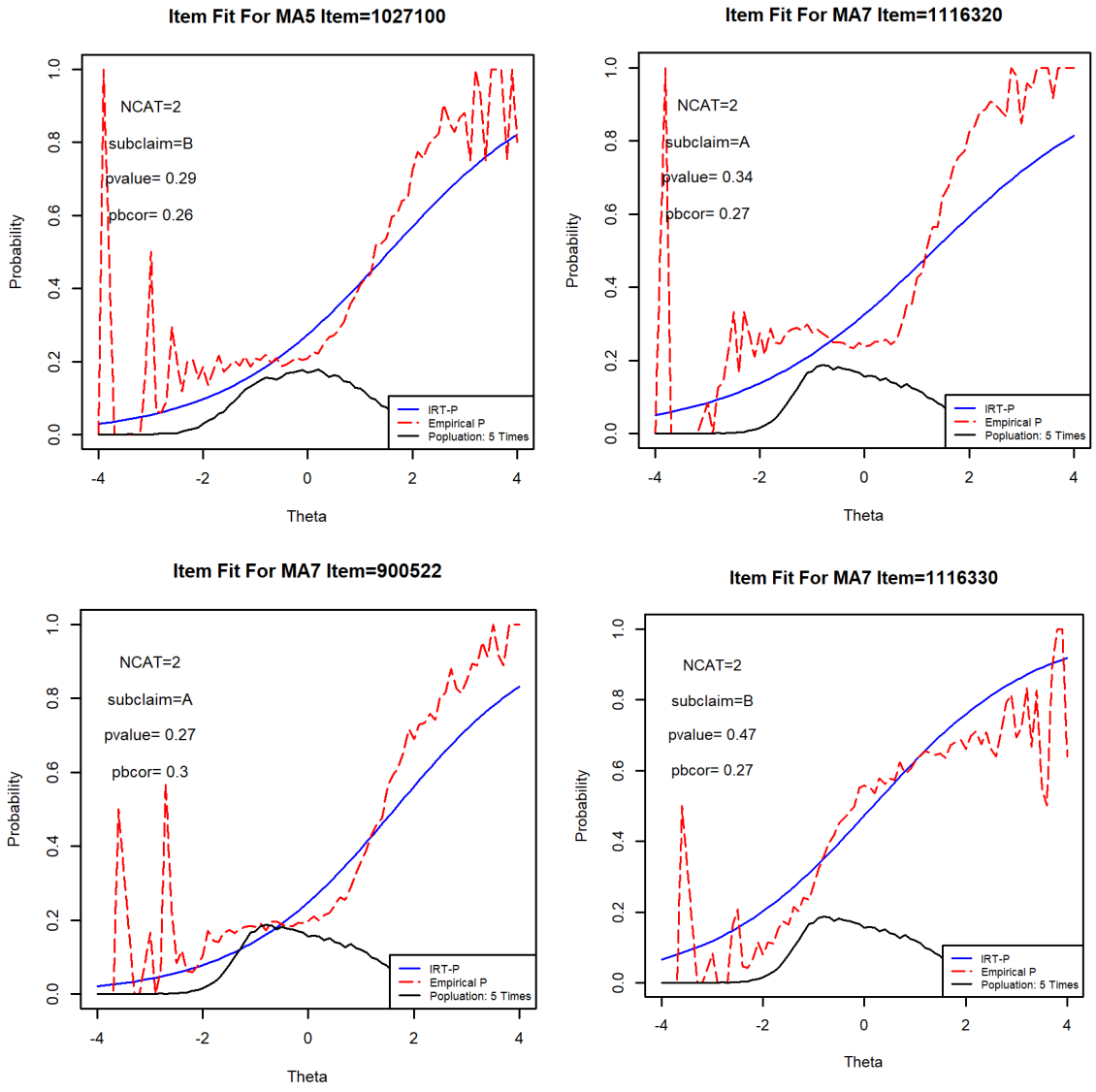


Figure 6.2 Item Fit Plots of Items Removed from Anchor Sets



After calibration, the IRT model fit was evaluated by reviewing item chi-squared statistic that were calculated using IRTPRO item parameters and item responses from students in the calibration sample. Adjusted fit values were calculated and flagged if they exceeded 0.35 (Pearson, 2018).

Since chi-square values are sensitive to sample size, these statistics are not easily compared when the number of students varies across items. As a result, adjusted fit values were calculated by dividing the chi-square fit statistic by the sample size using the following formula:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

Tables 6.21 and 6.22 show the adjusted item fit C values using the chi-square statistics and calibration sample sizes for ELA and mathematics, respectively. The average adjusted fit ranged from 0.12 to 0.14 for ELA and

0.08 to 0.09 for mathematics. No items were excluded based on model fit statistics because the adjusted item fits for all items were lower than the criterion value of 0.35, as can be seen in the maximum values for both ELA and mathematics. The largest adjusted fit value was 0.31 for ELA grades 4 and 5.

Table 6.21 Summary of Adjusted Fit for ELA

Grade	Mode	No. Items	Mean	Std. Dev.	Min.	Max.	No. Flagged Items
3	CBT/PBT	28	0.12	0.05	0.03	0.19	0
4	CBT	30	0.13	0.06	0.05	0.31	0
5	CBT	28	0.14	0.06	0.04	0.31	0
6	CBT	32	0.13	0.05	0.05	0.23	0
7	CBT	32	0.14	0.05	0.05	0.25	0
8	CBT	32	0.13	0.06	0.03	0.25	0

Table 6.22 Summary of Adjusted Fit for Mathematics

Grade	Mode	No. Items	Mean	Std. Dev.	Min.	Max.	No. Flagged Items
3	CBT/PBT	44	0.09	0.04	0.02	0.20	0
4	CBT	43	0.08	0.04	0.01	0.23	0
5	CBT	43	0.08	0.04	0.02	0.19	0
6	CBT	43	0.09	0.06	0.02	0.29	0
7	CBT	43	0.09	0.05	0.02	0.20	0
8	CBT	42	0.08	0.05	0.02	0.23	0

6.4.2 Linking 2023 LEAP 2025 Grades 3–8 to PARCC Scale

The 2016 and 2017 LEAP 2025 forms were linked to the PARCC scale using intact PARCC items embedded into the LEAP 2025 forms by using the Stocking & Lord procedure (1983). Therefore, these item parameters were placed on the PARCC scale. However, these equated Louisiana item parameters are based on only Louisiana students' responses while intact PARCC item parameters were estimated based on PARCC associated states' responses. To distinguish these two sets of item parameters, item parameters based on only Louisiana student responses will be called LEAP 2025 item parameters and its scale is referred to as the LEAP 2025 scale.

Two anchor sets were created for the 2023 Spring LEAP 2025 ELA and mathematics assessments equating process. Anchor 1 items were intact PARCC items embedded in the 2023 LEAP 2025 form. Anchor 2 items were items common to the 2023 LEAP 2025 spring forms and previous years' forms, and their item parameters were from previously operational LEAP 2025 item parameters. Anchor 2 was used in the operational analyses to link to the LEAP 2025 scale, which is the same as the PARCC scale, and Anchor 1 were used to help evaluate drift from the PARCC scale. Table 6.23 provides the Stocking & Lord transformation constants that were used to link to scale. Table 6.24 summarizes the number and score points of the initial anchor item selection before equating. Table 6.24 also summarizes the number and score points of the final anchor item selections. The difference between the initial number of anchor items and the final number of anchor items is the number of anchor items that were dropped.

Table 6.23 Stocking & Lord Transformation Constants

Content	Grade	Slope	Intercept
ELA	3	1.115429	0.133858
	4	1.049011	0.099498
	5	0.984195	0.049594
	6	1.063924	0.080348
	7	1.058298	0.025761
	8	1.043813	0.082408
Mathematics	3	0.975508	-0.25848
	4	1.069157	-0.08422
	5	1.024551	-0.27415
	6	1.072196	-0.21787
	7	1.039849	-0.13967
	8	0.974327	-0.09254

Table 6.24 Number and Score Points of Initial and Final Anchor Item Sets

			Anchor 1		Anchor 2	
Content	Grade	Anchor Set	Number of Items	Score Points	Number of Items	Score Points
ELA	3	Initial	18	40	14	30
		Final	17	38	14	30
	4	Initial	30	65	12	26
		Final	25	53	12	26
	5	Initial	24	54	14	31
		Final	19	41	14	31
	6	Initial	32	70	10	23
		Final	30	63	10	23
	7	Initial	22	50	12	27
		Final	22	50	12	27
	8	Initial	28	62	12	27
		Final	27	60	12	37
Mathematics	3	Initial	13	21	15	20
		Final	11	19	15	20
	4	Initial	21	31	15	21
		Final	20	30	15	21
	5	Initial	19	27	15	19
		Final	17	20	15	19
	6	Initial	19	30	15	22
		Final	18	29	15	22
	7	Initial	21	33	15	23
		Final	20	32	15	23
	8	Initial	15	21	15	22
		Final	14	19	14	21

**Following OP2 approach for counting Writing dimensions: Count WE and WKL only*

Figures 6.3 to 6.14 show test characteristic curves (TCCs) for anchor items, corresponding 2023 LEAP 2025 estimated anchor items (EQ_ANC), 2018 LEAP 2025 operational items (LEAP 2018), and all 2023 LEAP 2025 estimated items (EQ_ALL) for ELA and mathematics after applying the Stocking & Lord equating procedure. The blue solid line illustrates the anchor items, the red dotted line is the 2023 LEAP 2025 equated anchor items, the **black** solid line is for all the 2023 LEAP 2025 equated items, and the green dotted line is the 2018 LEAP 2025 operational items. Anchor items for each anchor set, 1 and 2, are different as mentioned above. For most ELA and mathematics grades, the TCCs for anchor items and the corresponding 2018 estimated anchor items were overlapped across most ability levels.

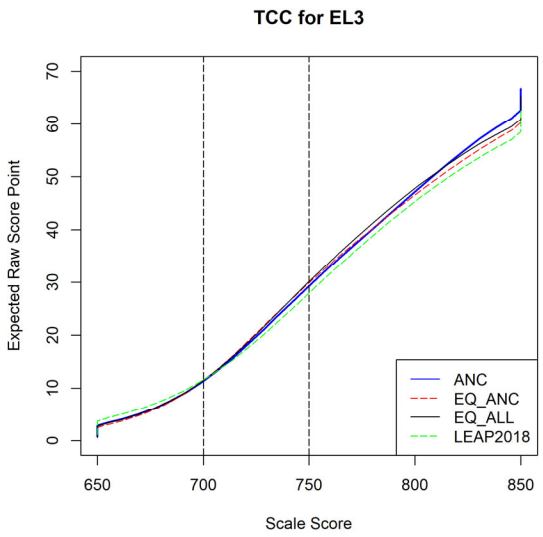
When the anchor 2, which is used for score reporting, was considered, the TCC of the anchor 2 items (ANC) and 2023 LEAP 2025 estimated anchor items (EQ_ANC) overlapped or were close to each other for all ELA grades. The same pattern was found for all mathematics grades. Anchor sets represented the overall test form in most grades. There were some differences at the extreme ranges, such as low ability or high ability. In mathematics grade 8, anchor set was a little easier than total test.

Figures 6.15 to 6.26 present scatter plots of slope item parameters and difficulty item parameters for ELA and mathematics and their correlation after linking 2023 LEAP 2025 to the PARCC 2016 scale.

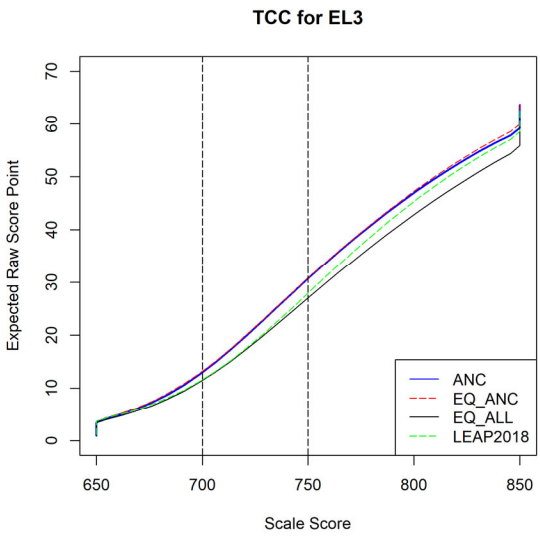
As can be seen in the ELA slope parameter plots, most parameters were around the identity line. The correlation between anchor item parameters and estimated parameters ranged from 0.94 to 0.99 with Anchor 2. For mathematics, most item slope parameters were around the identity line, and the correlations ranged from 0.97 to 0.99 with Anchor 2.

For ELA, most item difficulty parameters were around the identity line, and the correlations ranged from 0.97 to 1.00 with Anchor 2. For mathematics as well, most item difficulty parameters were around the identity line. Correlations ranged from 0.98 to 0.99 across grades with Anchor 2. It is common to find higher correlations for difficulty parameters than those for slope parameters.

Figure 6.3 ELA Grade 3 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

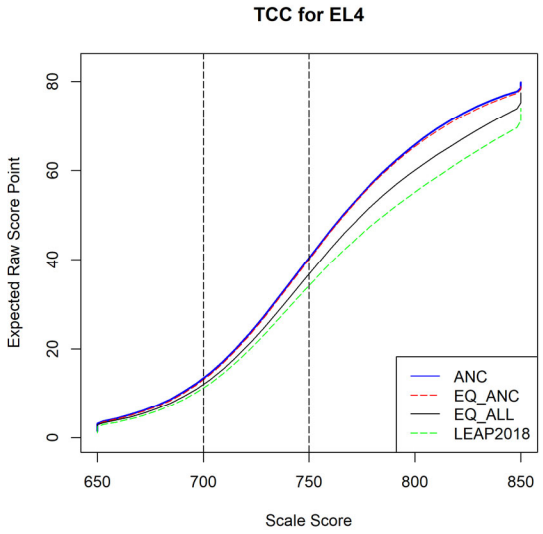


Anchor 1

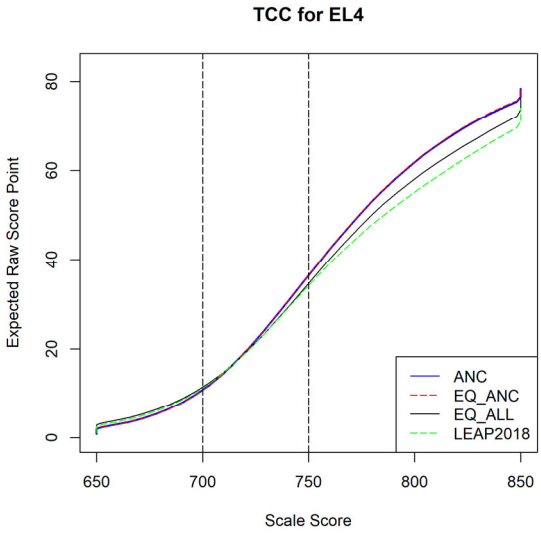


Anchor 2

Figure 6.4 ELA Grade 4 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

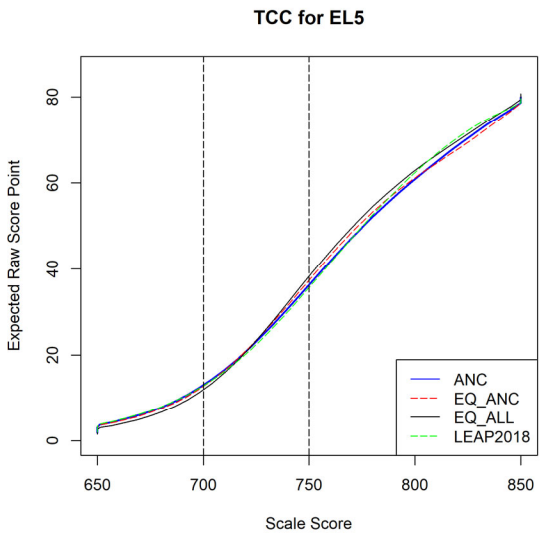


Anchor 1

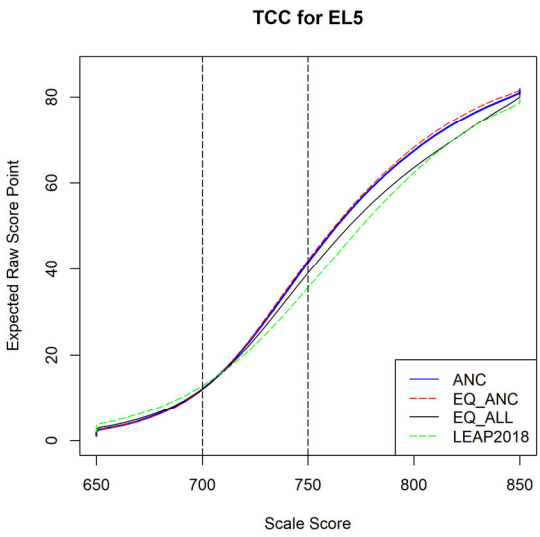


Anchor 2

Figure 6.5 ELA Grade 5 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

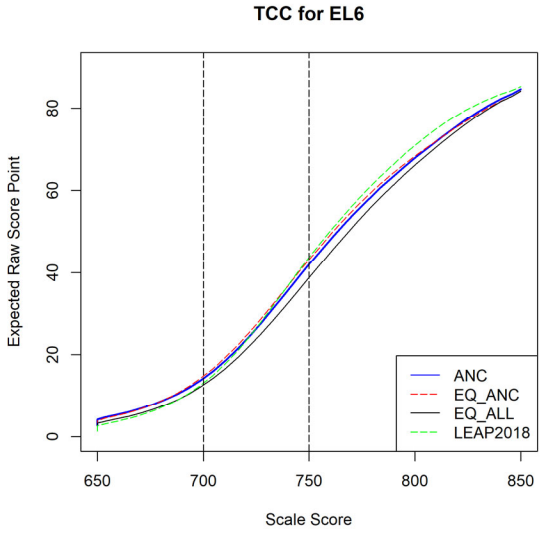


Anchor 1

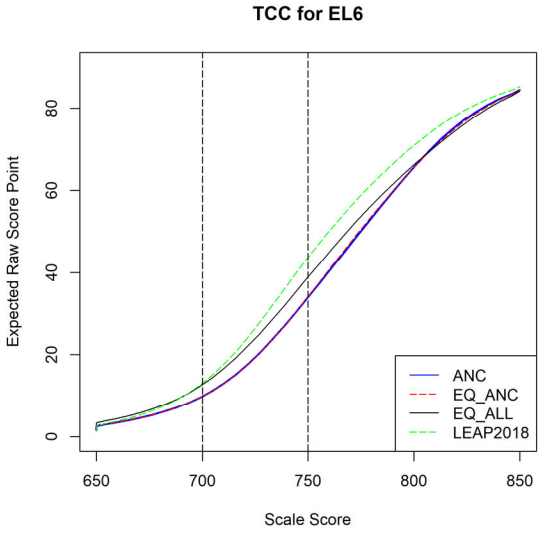


Anchor 2

Figure 6.6 ELA Grade 6 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

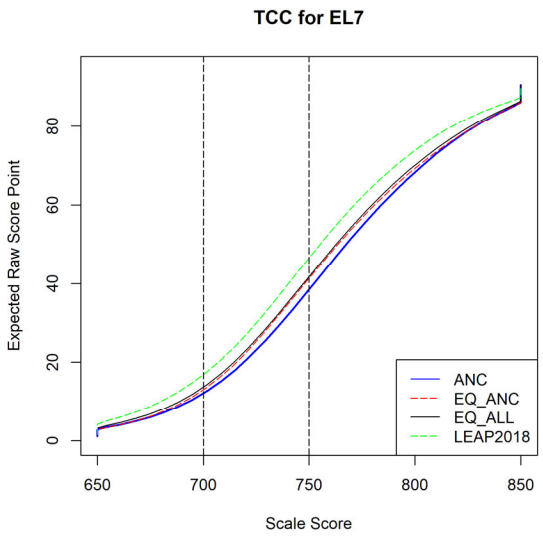


Anchor 1

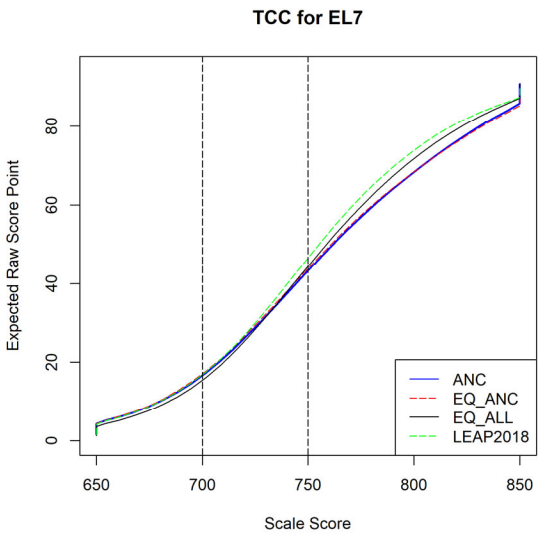


Anchor 2

Figure 6.7 ELA Grade 7 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

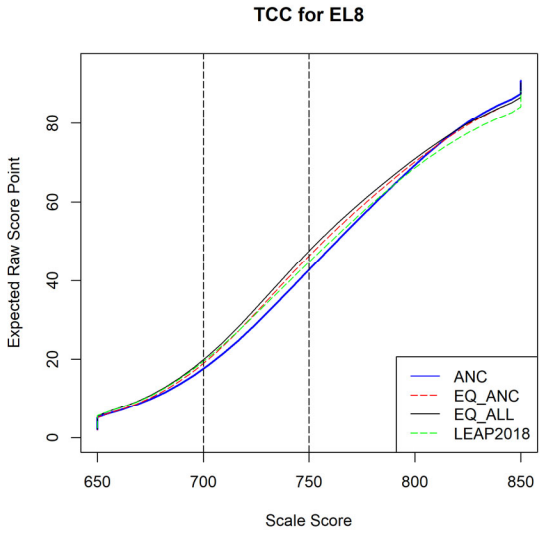


Anchor 1

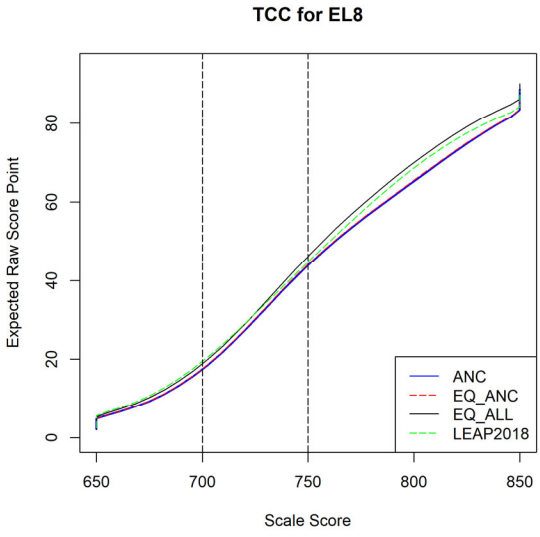


Anchor 2

Figure 6.8 ELA Grade 8 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

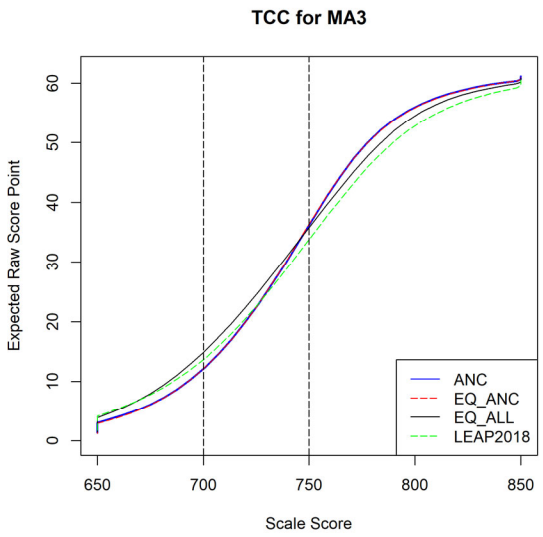


Anchor 1

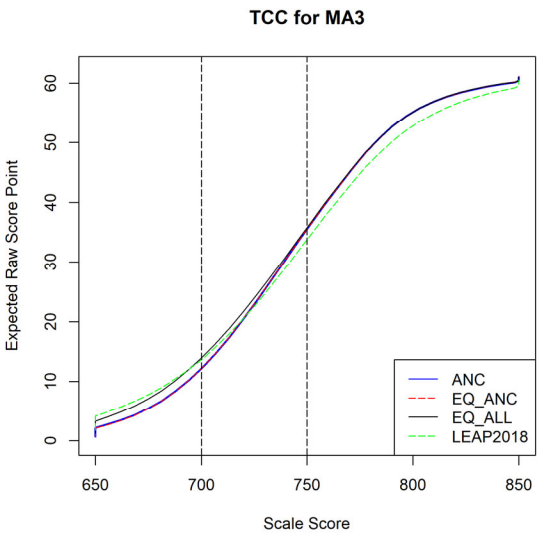


Anchor 2

Figure 6.9 Mathematics Grade 3 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

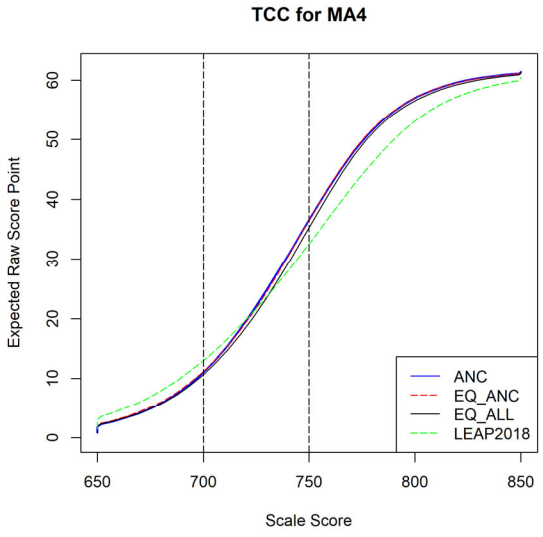


Anchor 1

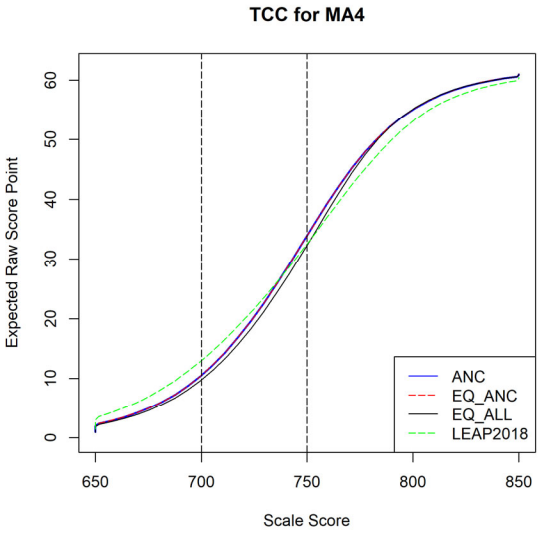


Anchor 2

Figure 6.10 Mathematics Grade 4 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

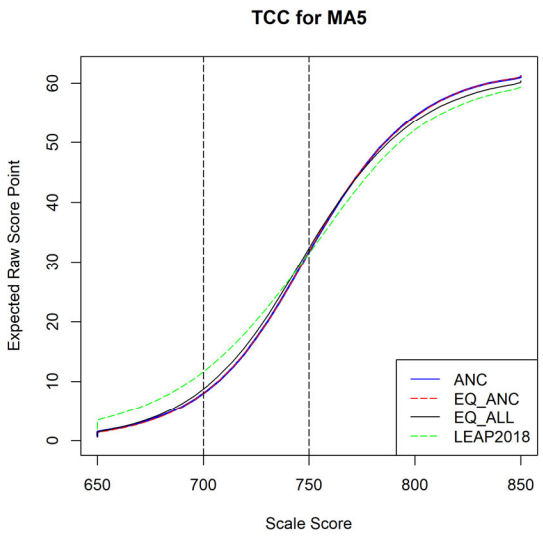


Anchor 1

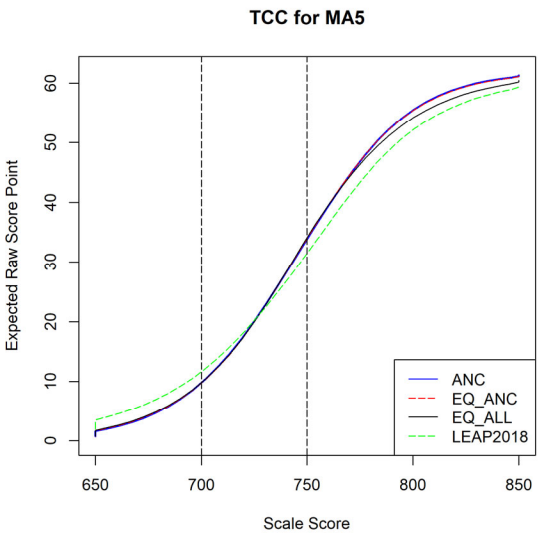


Anchor 2

Figure 6.11 Mathematics Grade 5 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

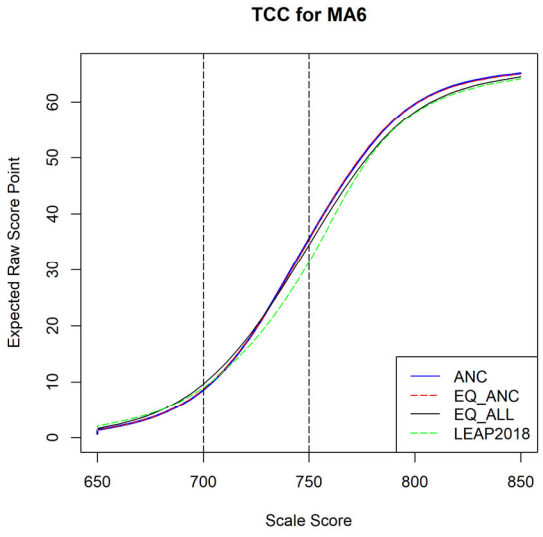


Anchor 1

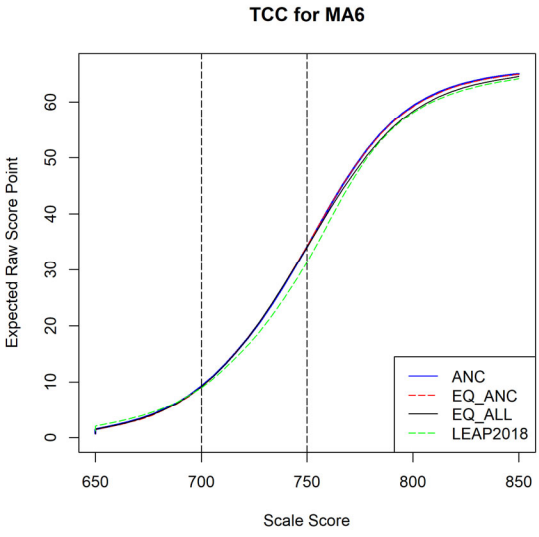


Anchor 2

Figure 6.12 Mathematics Grade 6 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

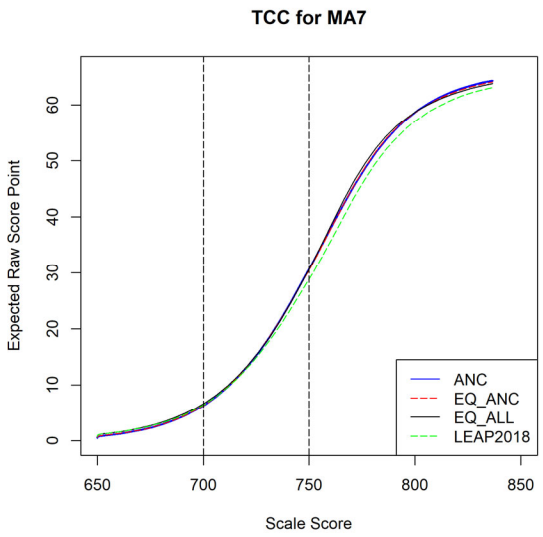


Anchor 1

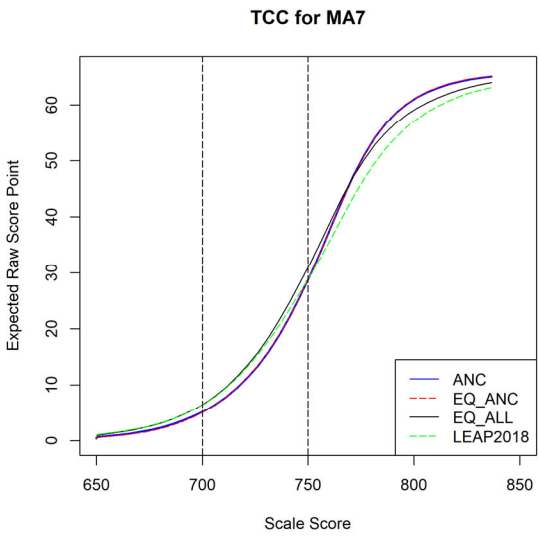


Anchor 2

Figure 6.13 Mathematics Grade 7 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items

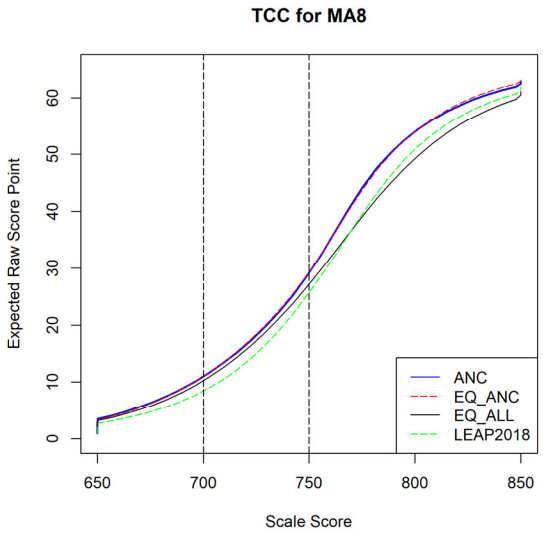


Anchor 1

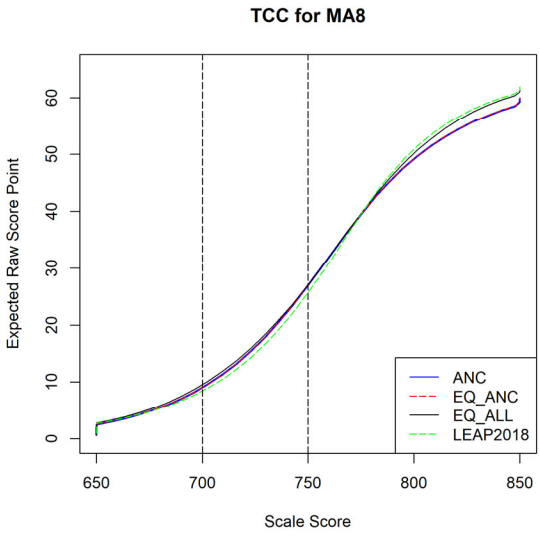


Anchor 2

Figure 6.14 Mathematics Grade 8 TCC between Pre-equated Anchor, Equated Anchor, LEAP 2018, and All LEAP 2025 Items



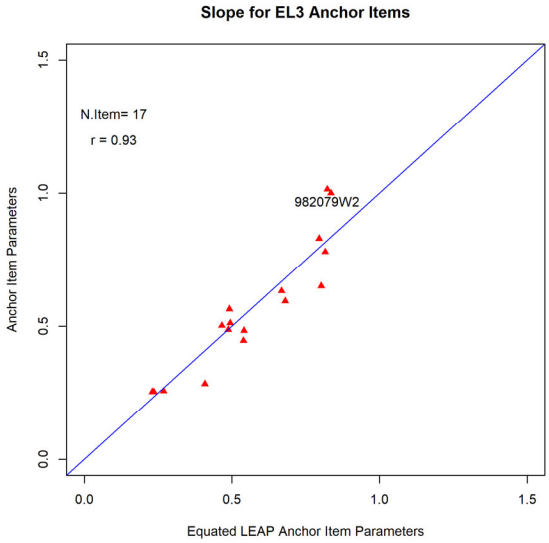
Anchor 1

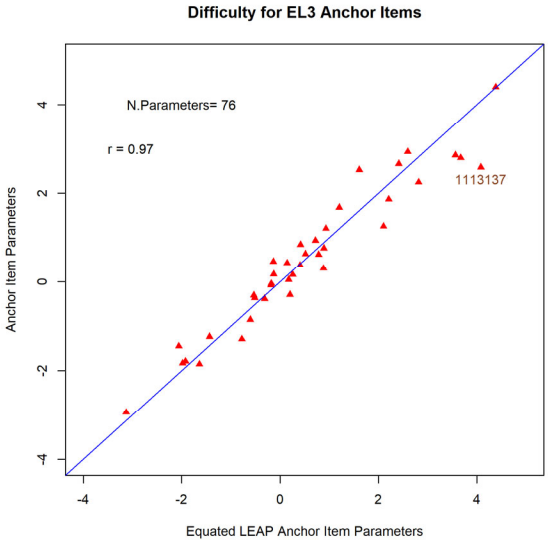


Anchor 2

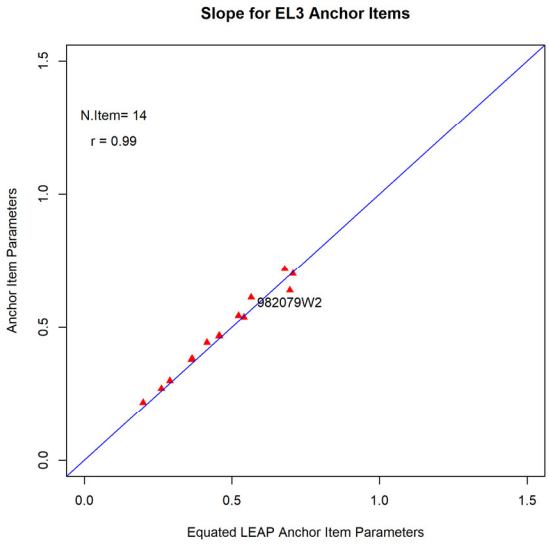
Figure 6.15 ELA Grade 3 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1





Anchor 2



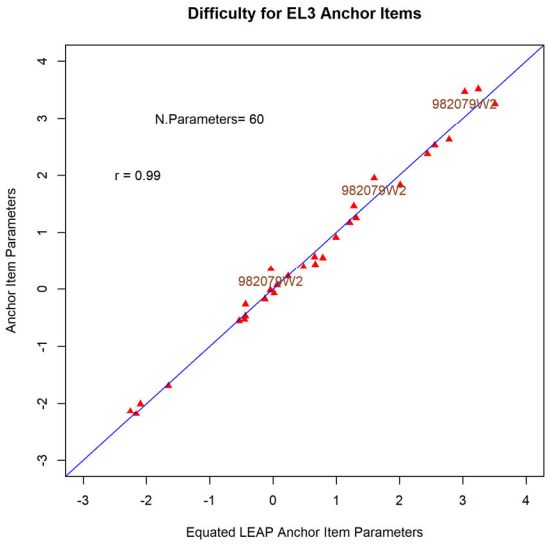
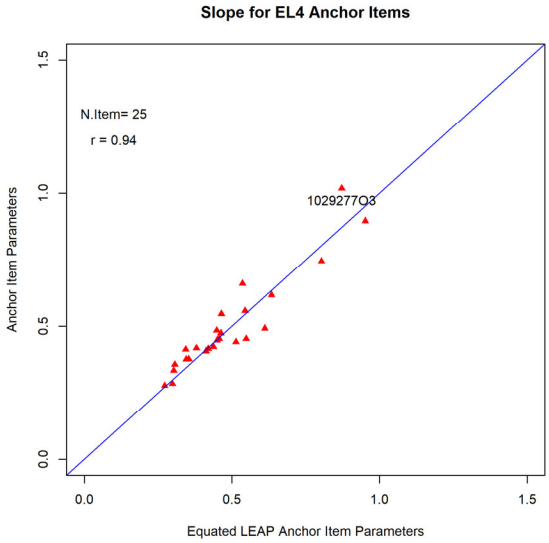
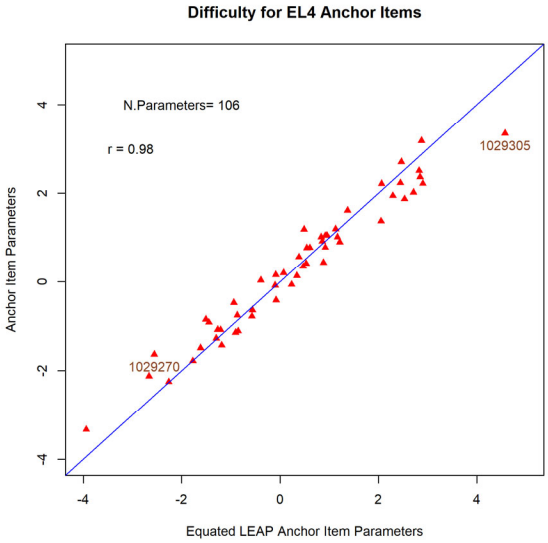


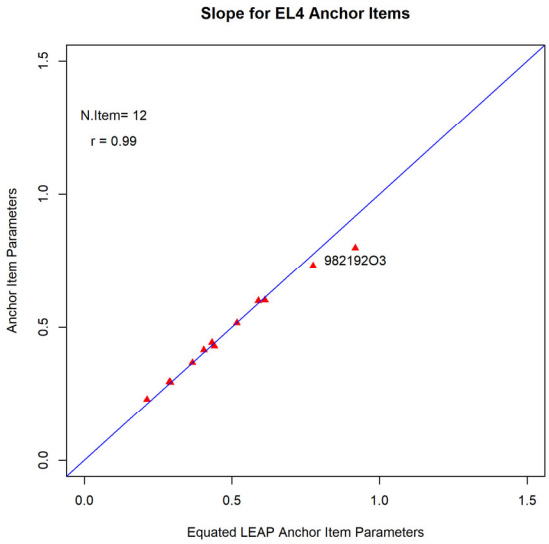
Figure 6.16 ELA Grade 4 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1





Anchor 2



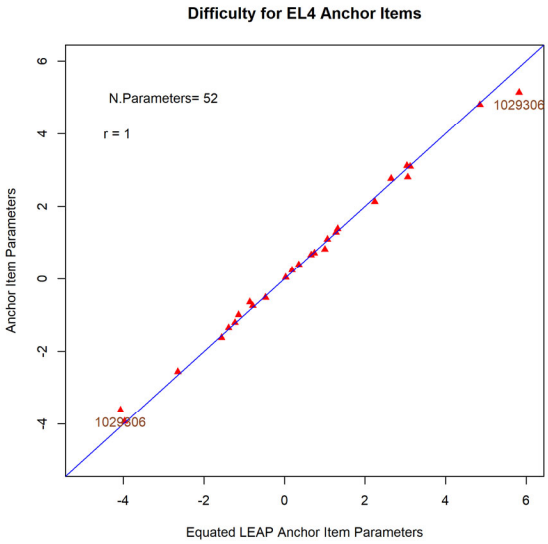
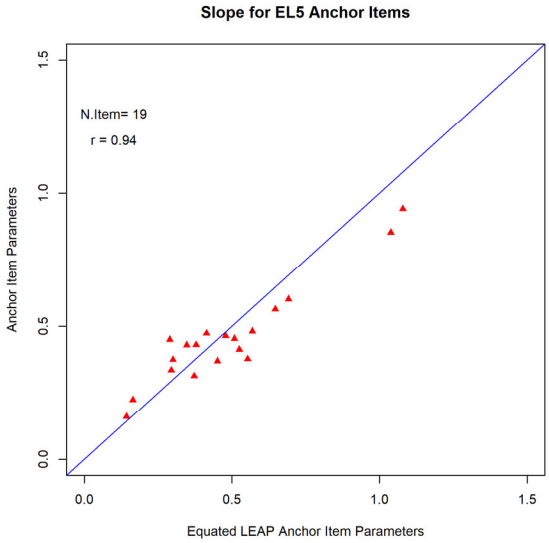
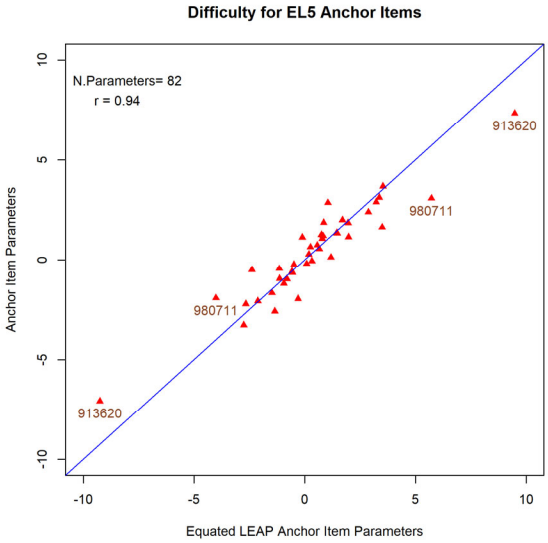


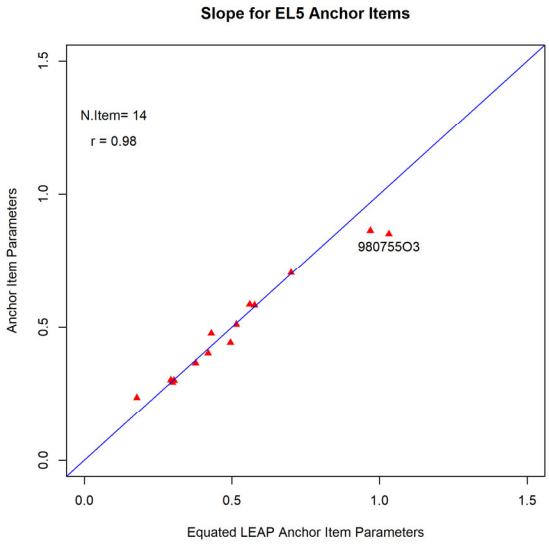
Figure 6.17 ELA Grade 5 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1





Anchor 2



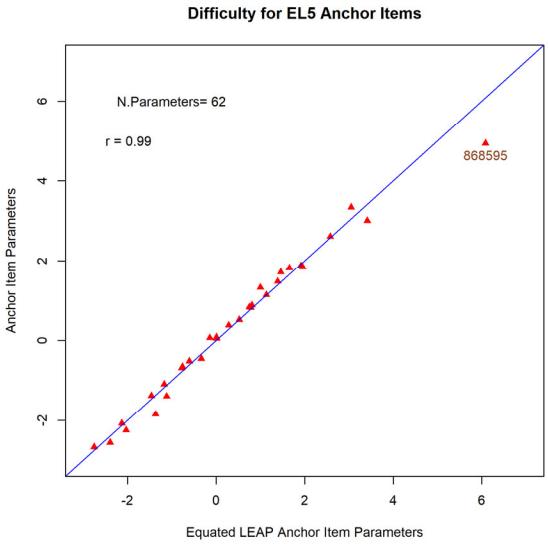
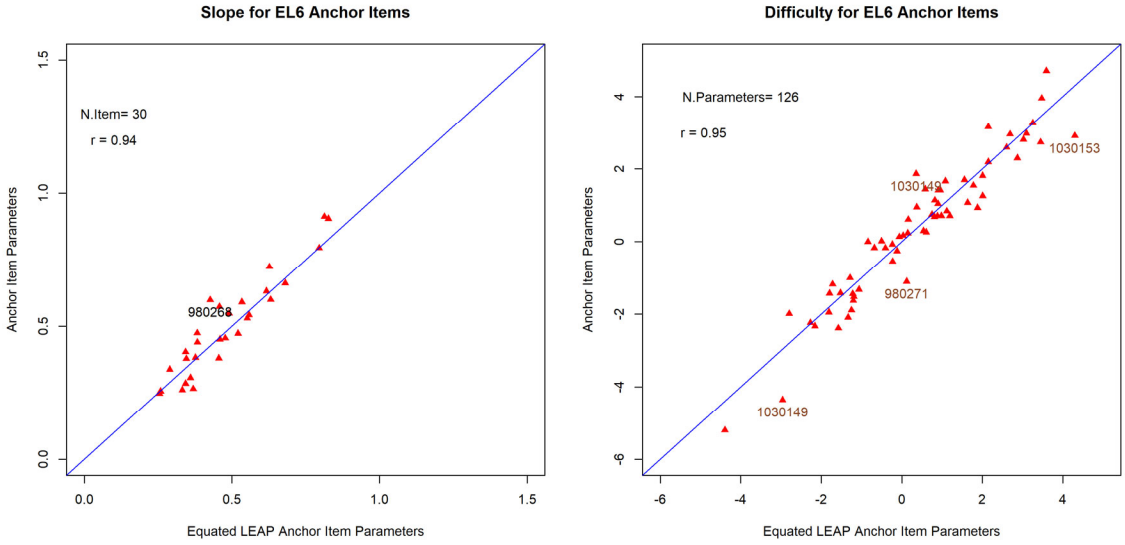


Figure 6.18 ELA Grade 6 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1



Anchor 2

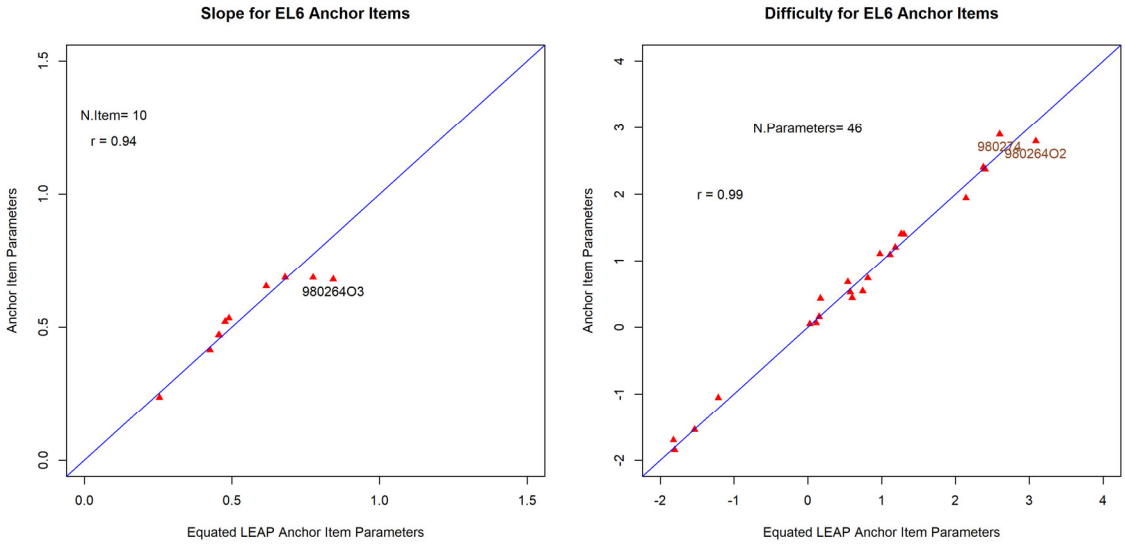
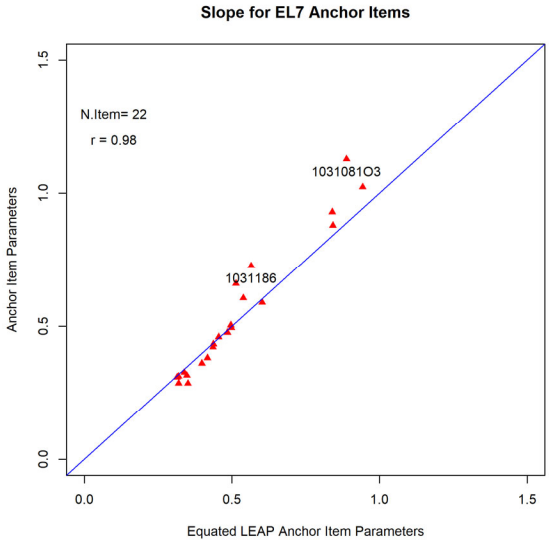
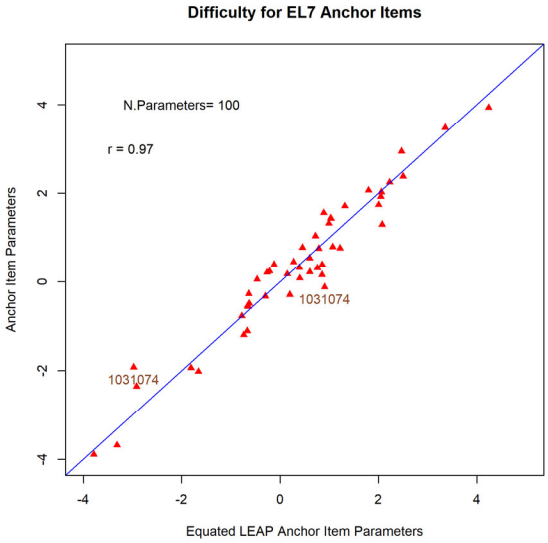


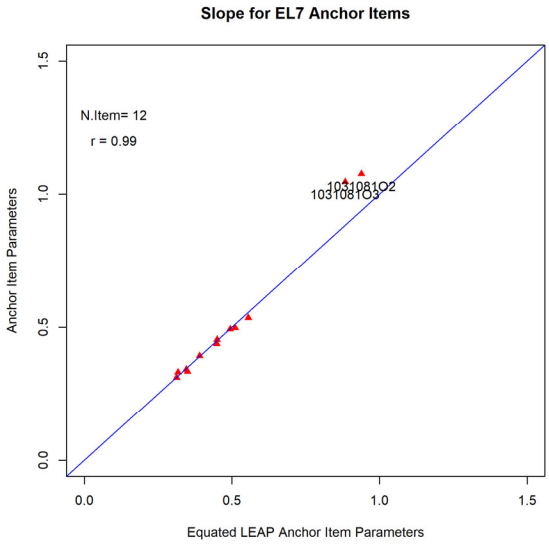
Figure 6.19 ELA Grade 7 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1





Anchor 2



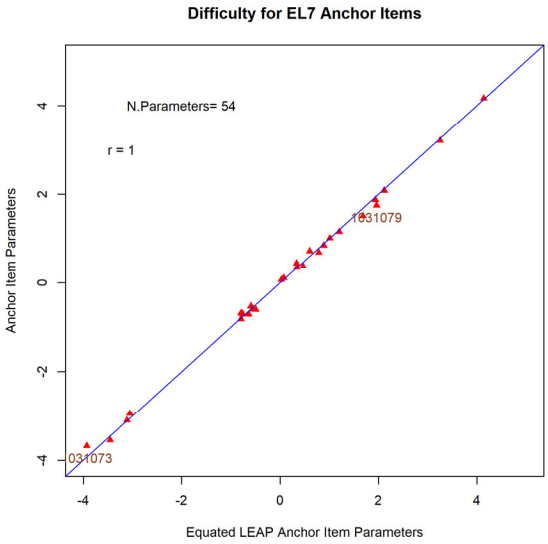
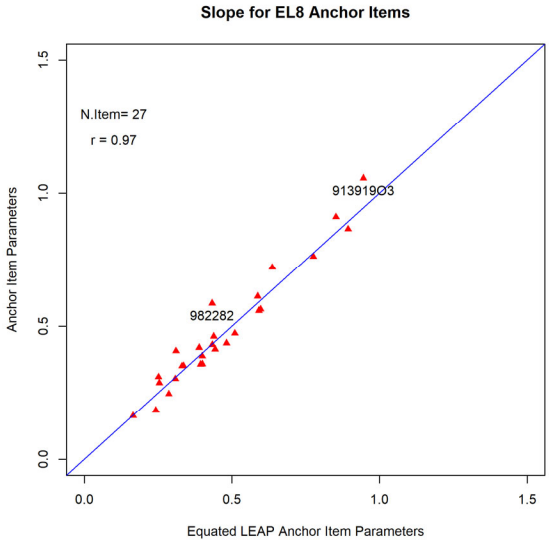
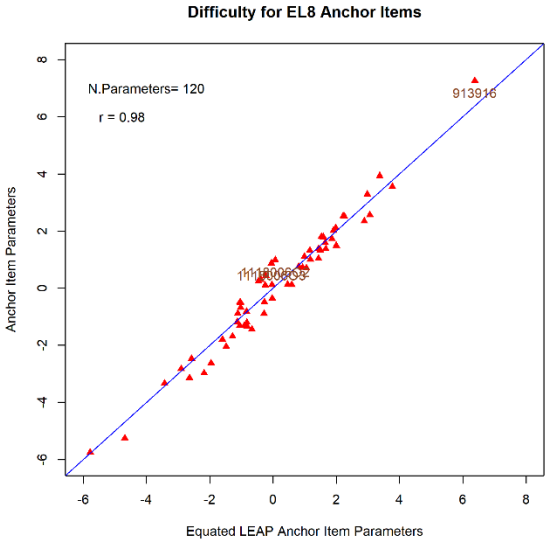


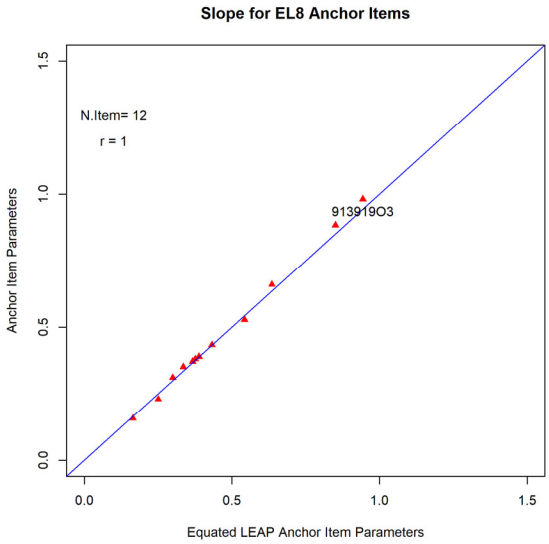
Figure 6.20 ELA Grade 8 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1





Anchor 2



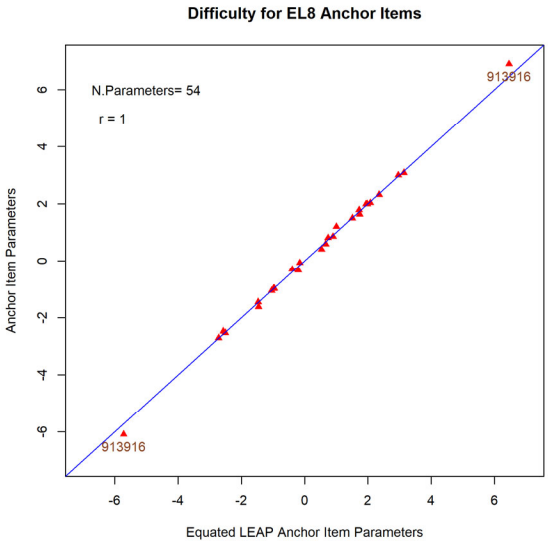
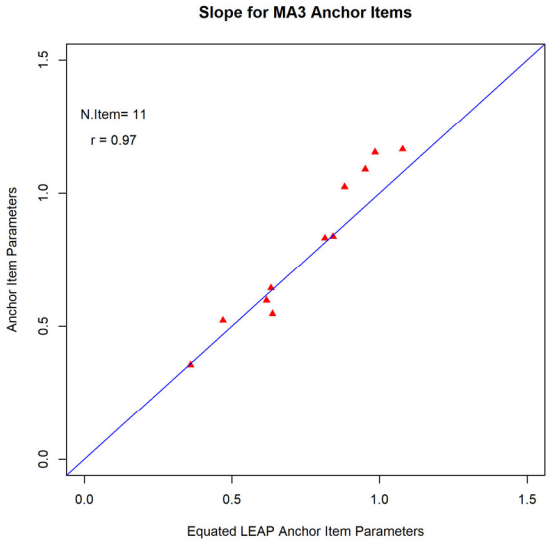
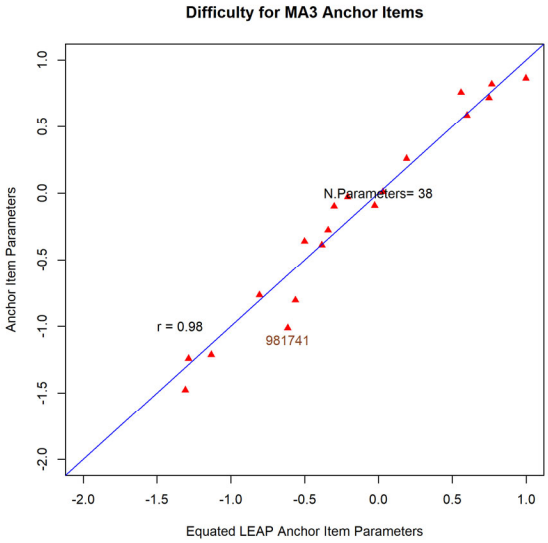


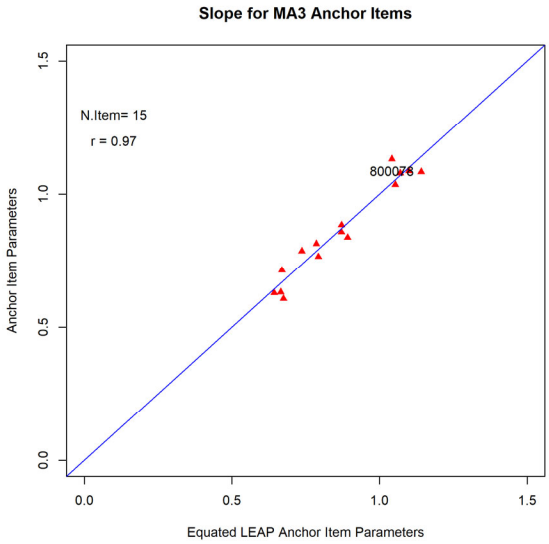
Figure 6.21 Mathematics Grade 3 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1





Anchor 2



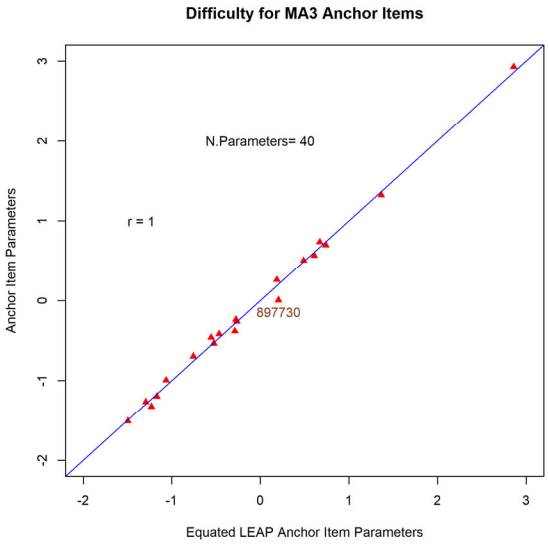
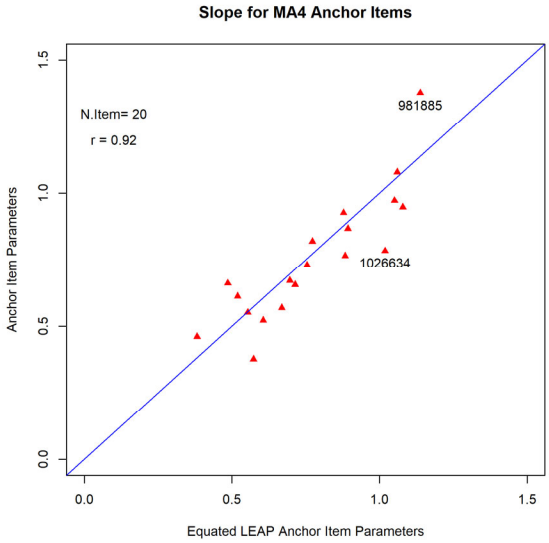
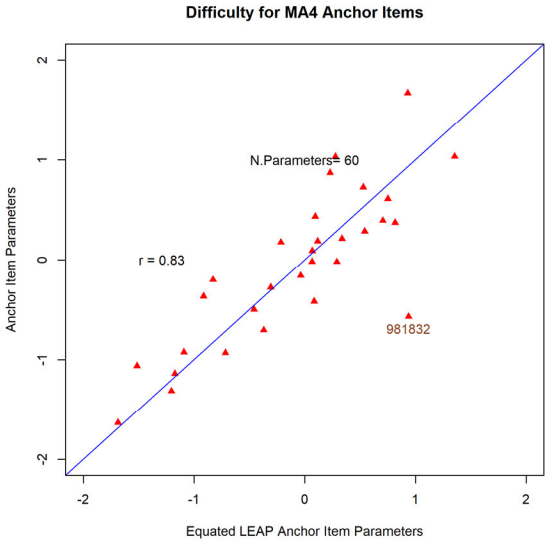


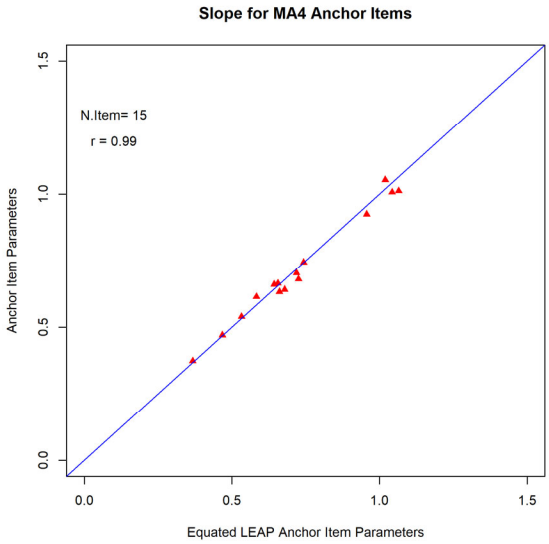
Figure 6.22 Mathematics Grade 4 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1





Anchor 2



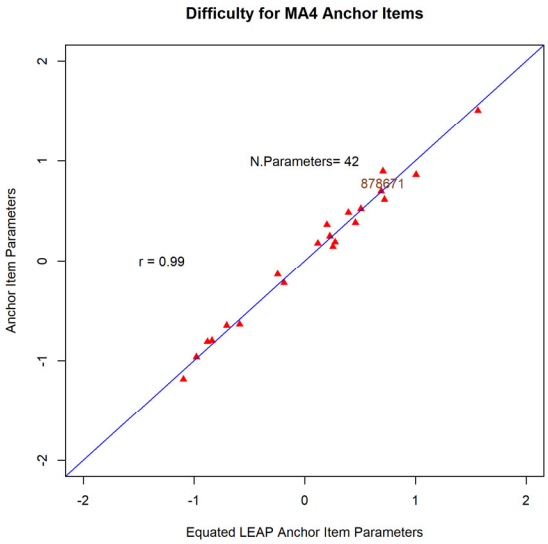
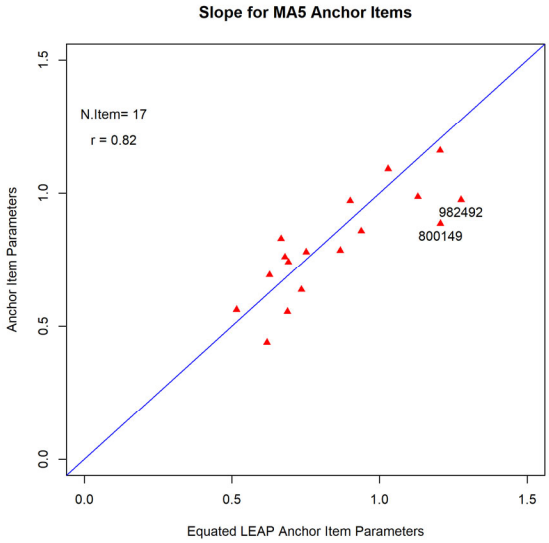
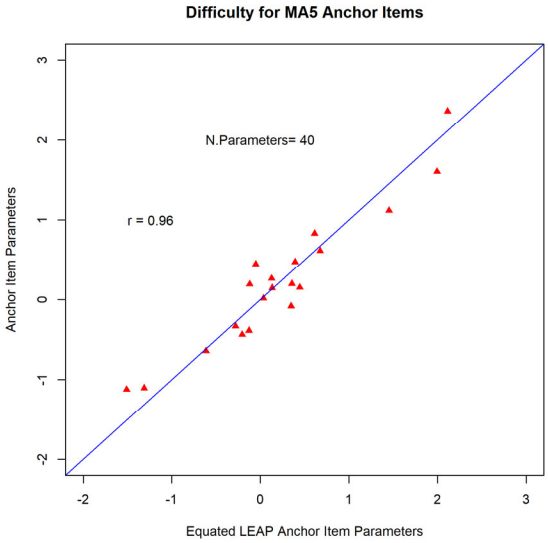


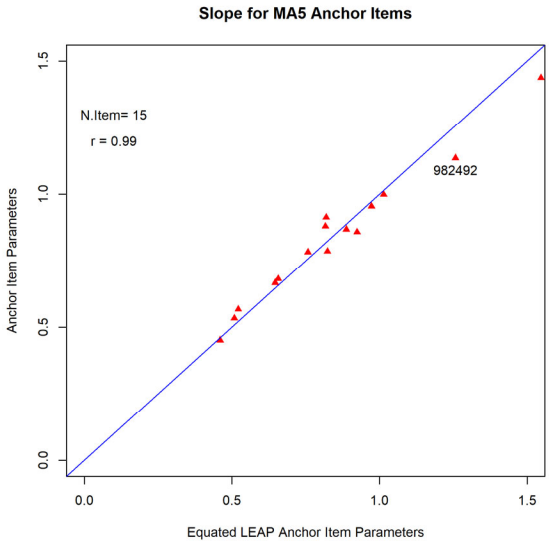
Figure 6.23 Mathematics Grade 5 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1





Anchor 2



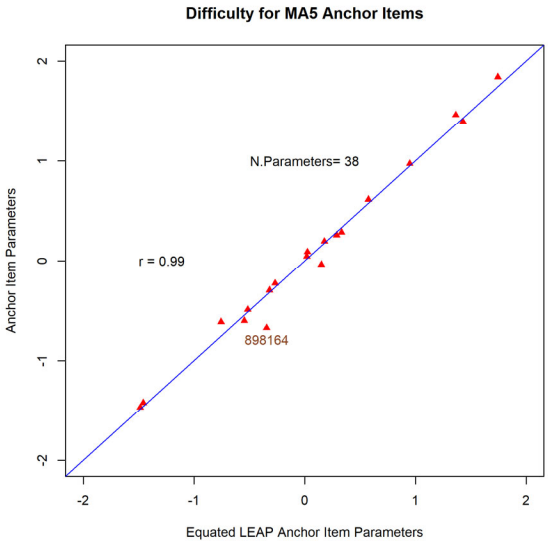
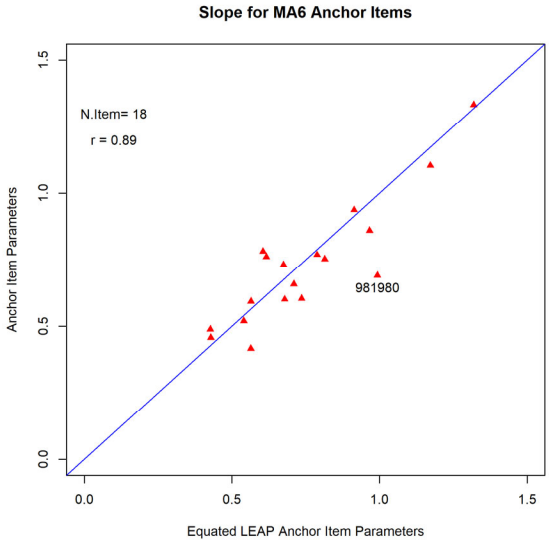
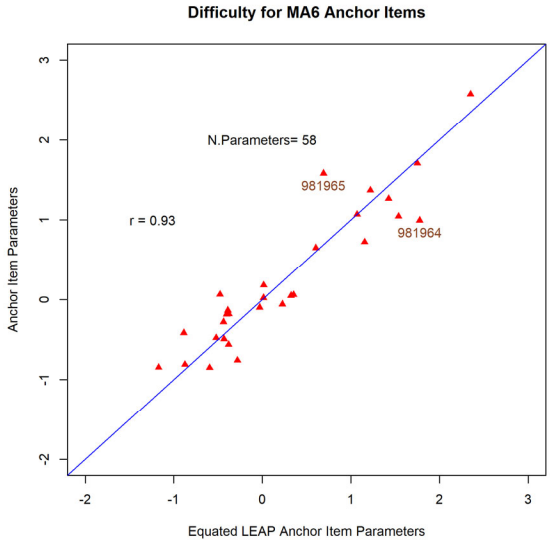


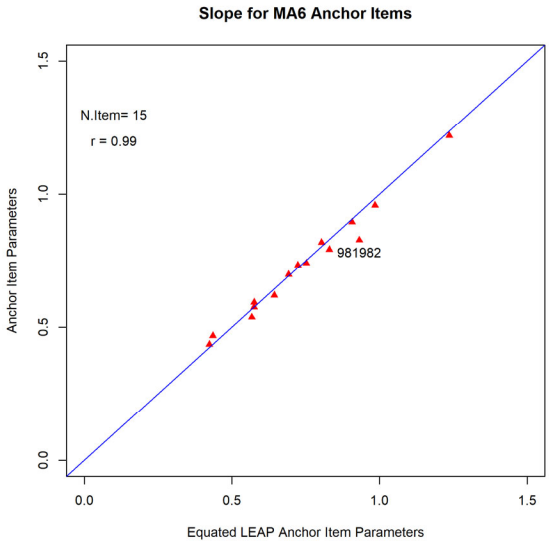
Figure 6.24 Mathematics Grade 6 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1





Anchor 2



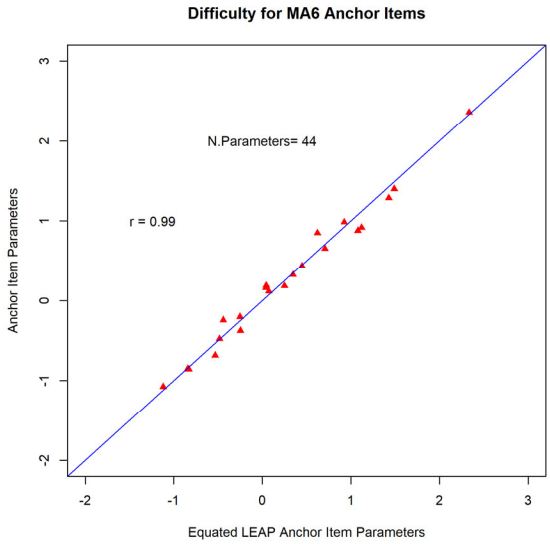
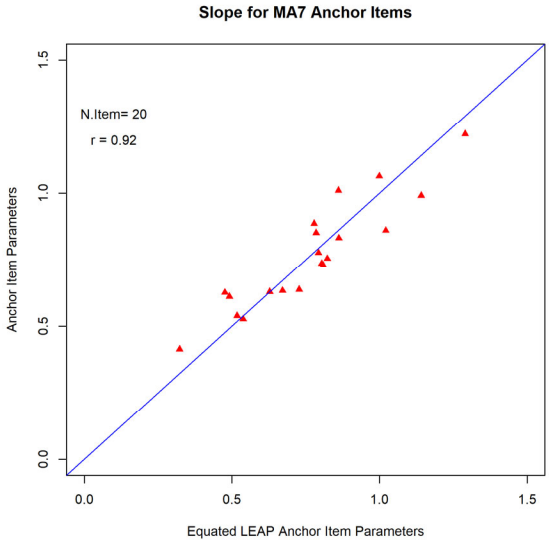
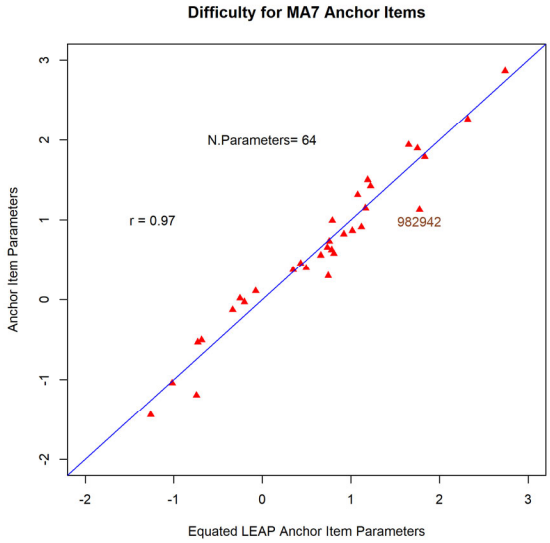


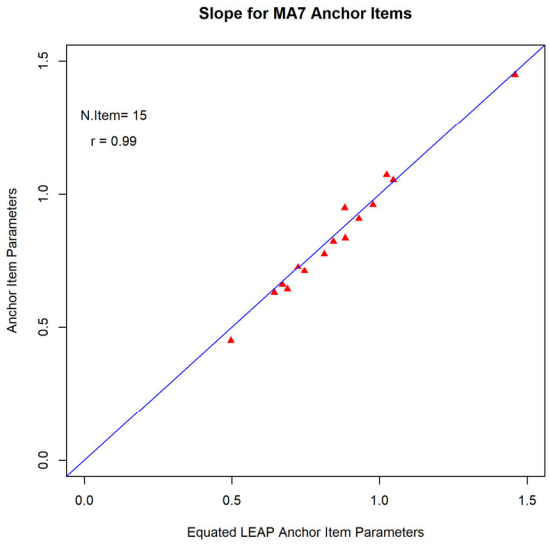
Figure 6.25 Mathematics Grade 7 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

Anchor 1





Anchor 2



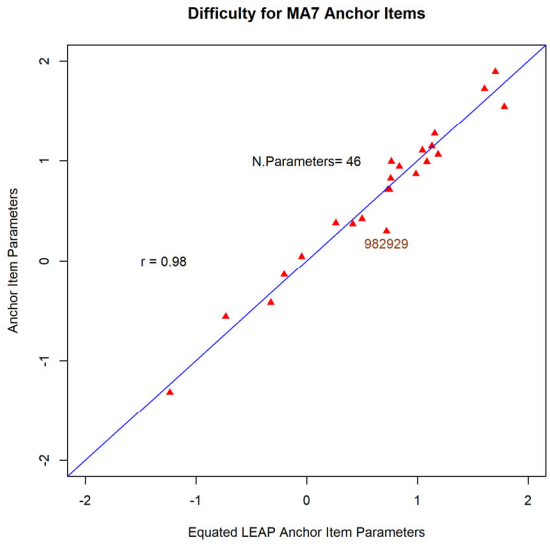
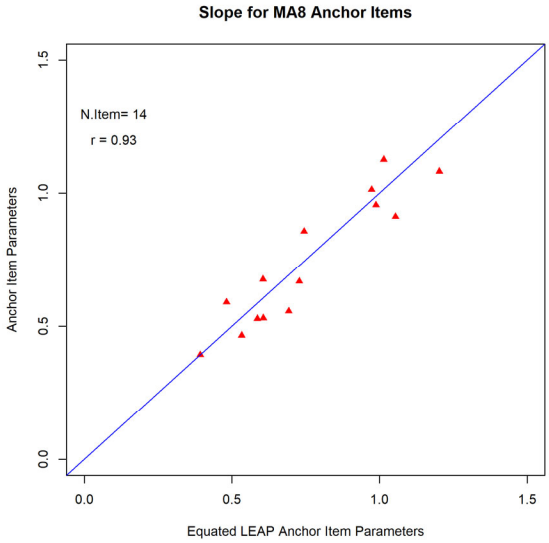
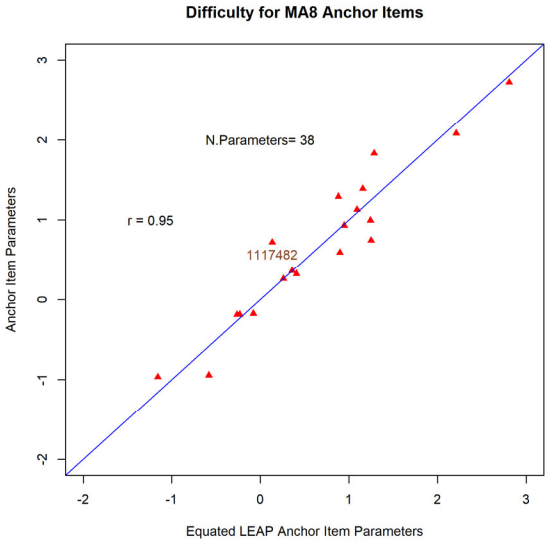


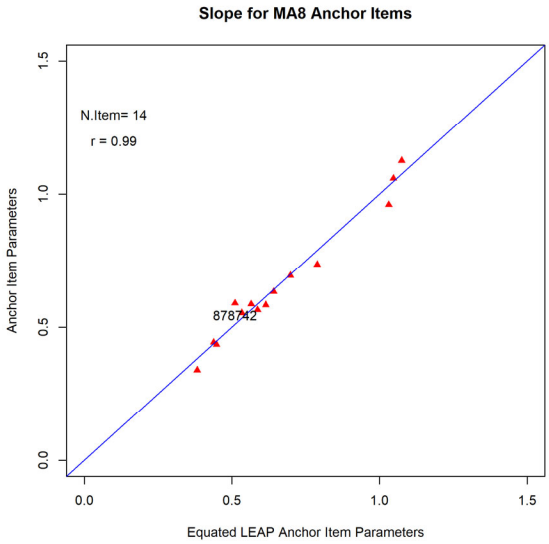
Figure 6.26 Mathematics Grade 8 Slope and Difficulty Parameters Between Pre-equated and Equated Anchor Item Parameters

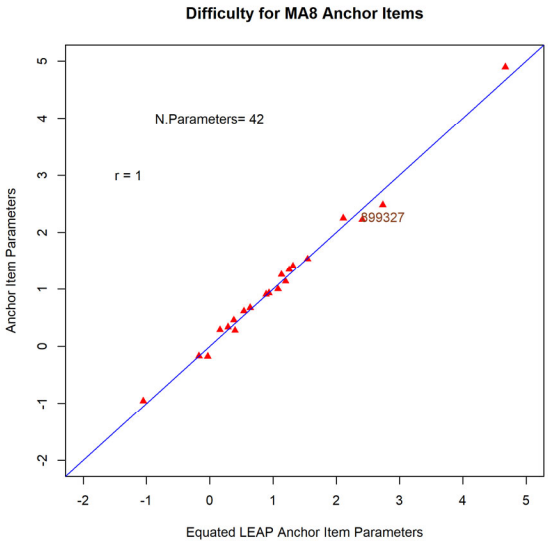
Anchor 1





Anchor 2





6.4.2.1. Evaluation of Anchor Item Stability

Standard 5.15 requires that information about the anchors be presented, stating the following:

> In equating studies that employ an anchor test design, the characteristics of the anchor test and its similarity to the forms being equated should be presented, including both content specifications and empirically determined relationships among test scores. If anchor items are used in the equating study, the representativeness and psychometric characteristics of the anchor items should be presented (105).

One of the key requirements of anchor items in deriving valid reliable linking results is that the anchor items should form a miniature version of the test in terms of content coverage or test blueprint. Dropping flagged anchor items based solely on statistical criteria may change the content coverage and impact the validity of the results. Before an anchor item may be dropped from an anchor set, the item characteristics, adequacy of the content coverage, and impact to the size of the anchor set should be evaluated.

Outliers of anchor items were reviewed with the Robust *Z* (Huynh & Meyer, 2010) and the weighted root mean square difference (WRMSD) method in addition to being verified from a content perspective, when reviewers considered aspects of the outliers, such as the number of items and score points for each category and subcategory. If approved by the LDOE, the outliers were dropped from anchor sets and considered to be non-common anchor items during equating. The following evaluation rules were applied in order to check the quality of anchor items and the anchor set.

- Exclude CR items from anchor set if categories were collapsed due to small sample size.
- Exclude items with content or parameter estimation issues.
- Run Robust Z method and remove flagged items from anchor set using the criterion value of |1.96|
- Run STUIRT and flag items if the WRMSD was greater than the values in Table 6.25.
- Remove an item from the anchor set if it is flagged by both Robust Z and WRMSD.
- Flag outliers using the plots of slope and difficulty item parameters with their correlations (Kolen & Brennan, 2014).
- Check score points and the numbers of items by reporting category and subcategory before and after dropping an anchor item.

Huynh and Meyer (2010) suggested to applying a z statistic that is robust under the presence of outliers. The robustification is established by replacing mean with median and standard deviation with interquartile range (IQR) for anchor items. A multiplicative constant (0.74) is applied to IQR to emulate the standard deviation of the normal distribution:

$$z = \frac{(D - Md)}{0.74 \times IQR},$$

where D is the difference between intact and estimated item parameters of an anchor item and *Md* is a median of differences between intact and estimate item parameters for all items. The critical value of ±1.96 is often used to evaluate estimated robust z values.

The WRMSD values were calculated to compare to the ICCs using intact and estimated anchor item parameters. WRMSD is defined as

$$SQRT\{\sum_{Q=1}^{41} W_Q[\mathrm{ICC}_Q(EST) - \mathrm{ICC}_Q\ (INTACT)]^2\},$$

where *Q* represents a quadrature point (i.e., node), *W* represents its weight given quadrature point *Q* from the standard normal distribution, *INTACT* represents intact item parameters, and *EST* represents estimated item parameters corresponding to intact item parameters. Table 6.25 summarizes WRMSD flagging criteria for inspection and possible removal of linking items.

Table 6.25 PARCC WRMSD Flagging Criteria

Categories	Points	WRMSD/Points	WRMSD
2	1	0.100	0.100
3	2	0.075	0.150
4	3	0.075	0.225
5	4	0.075	0.300
6	5	0.075	0.375
7	6	0.075	0.450
> = 8	> = 7	0.090	0.999

6.4.2.2. Lowest and Highest Obtainable Scale Scores

A maximum likelihood (MML) procedure cannot produce scale score estimates for students with perfect scores or scores below the level expected when students are guessing. In addition, although MML estimates are available for students with extreme scores other than zero or perfect, occasionally these estimates have standard errors of measurement that are very large, and differences between these extreme values have little meaning. Therefore, scores are established for these students based on a rational but necessary non-MML procedure. These values, which are set separately by grade, are called the lowest obtainable scale score (LOSS) and the highest obtainable scale score (HOSS). All grades and content areas in LEAP 2025 used the same LOSS and HOSS values. The LOSS value was 650, and the HOSS value was 850.

6.4.2.3. Reporting Category and Subcategory Subscores

A student's performance on the ELA reporting categories (i.e., Reading and Writing) and mathematics categories (i.e., Major Content, Additional & Supporting Content, Expressing Mathematical Reasoning, and Modeling & Application) is reported in one of three ratings: *Weak*, *Moderate*, or *Strong*.

Additionally, subcategory ratings are reported at the student level for ELA and mathematics. ELA has three subcategories for reading (i.e., literary text, informational text, and vocabulary) and two subcategories for writing (i.e., written expression and knowledge and use of language conventions). Mathematics has subcategories that differ by grade. Subcategory performance is reported in one of three ratings of achievement: *Strong*, *Moderate*, or *Weak*. The 2023 LEAP 2025 reporting categories are summarized in chapter 3. Please see Table 3.1 for ELA and Table 3.8 and 3.9 for mathematics.

Although the performance ratings are determined only by the items included within a category or subcategory, the level of knowledge and ability needed to achieve a performance rating is connected to the level of knowledge and ability required to reach the subject-level achievement levels in the overall tests: a *Weak* rating requires similar knowledge and ability as the *Unsatisfactory* and *Approaching Basic* achievement levels, a *Moderate* rating requires similar knowledge and ability as the *Basic* achievement level, and a *Strong* rating requires similar knowledge and ability as the *Mastery* or *Advanced* achievement levels.

Reading and writing reporting category scores were produced for ELA assessments only. The reading category score range was 10–90 and the writing category score range was 10–60. The method for scaling categories

followed the PARCC methodology (Pearson, 2017). For the reading category, two theta score points corresponding to ELA scale scores of 700 and 750 were used for scaling. Linear transformation constants mapping the two theta points to scale score points of 30 and 50 were calculated. After these transformation values were applied to item parameters belonging to the reading category, a scoring table was generated using the TCC inverse method. A similar approach was applied to scale the writing category, using two scale score points of 30 and 35. Two cut scores, 40 and 50 for reading and 30 and 35 for writing, were used to produce three performance-level ratings for each category (see Table 6.26 for cut scores for summatives, categories, and subcategories).

For reporting categories in mathematics and subcategories in ELA and mathematics, only performance-level ratings were reported. Therefore, there is no need to scale these scores. Using the item parameters belonging to a given category (mathematics) or subcategory (ELA), a raw-score-to-theta scoring table is generated by applying the TCC inverse method. PARCC estimated Θ_{L3} and Θ_{L4} corresponding to scale scores of 725 and 750 for each content/grade using PARCC 2016 operational items by the TCC inverse method, and these values are the same across years. The two raw scores corresponding to Θ_{L3} and Θ_{L4} are cut scores for the category (mathematics) and subcategory (ELA).

This is also illustrated in Table 6.26.

Table 6.26 Cut Scores for Summative, Reporting Categories, and Subcategories

Performance Level	Summative Test	Category (ELA)		Category (Mathematics)/Subcategory (Mathematics and ELA)
		Reading	Writing	
1				
2	700	30	25	
3	725	40	30	Θ_{L3}
4	750	50	35	Θ_{L4}
5	Around 800			

**Subcategory thetas are those from summative tests (i.e., 725 & 750).*
***Yellow highlight shows cut scores for category and subcategory.*

6.4.3 Item Difficulty-Student Ability Maps

LEAP 2025 item difficulties based on item response theory (IRT) were plotted to show the distribution of the item difficulties across student performance. The plots allow easy visualization of the relationship between the distributions of item difficulty and student ability. While the item difficulty parameters estimated with the Rasch model directly place item difficulty on the student performance scale (i.e., ability/theta), those estimated with the 2PL/GPMC model cannot be placed on the student performance scale because of an additional parameter, item slope. LEAP 2025 uses the 2PL/GPMC model. To resolve this issue, the concept of response probability (RP) from item mapping procedures, such as the Bookmark Standard Setting Procedure (BSSP; Lewis, Mitzel & Green, 1996), was applied to all spring 2023 LEAP 2025 operational items.

In the BSSP, an RP specifies the probability with which a student with a given ability would be able to correctly answer an item of the same difficulty. For example, if the RP criterion is 0.67 (RP67), students with a given ability would have a 67% chance of correctly answering items with difficulty at the same level. For a BSSP, it is common to use an RP67 to clearly define when students have mastery of an item (Huynh, 1988). The choice of RP criterion to use in a BSSP is a policy decision, and many states have selected different RP criteria for different purposes, and other RP criteria are often used (Cizek & Bunch, 2007, p. 162; Mitzel, et al., 2001). For the purposes of aligning item difficulty with student performance, an RP50 was selected. This

indicates that students with a given ability would have a 50% chance of correctly answering items with difficulty at the same level.

Figure 6.27 through Figure 6.32 plot the ELA distributions and Figure 6.33 through Figure 6.38 plot the Mathematics distribution. There is one RP50 value for a multiple-choice item. There is one value where it is considered that test takers of a certain achievement level will answer the MC item correctly 1/2 of the time. In a BSSP, the RP for a polytomous item is generally split by score point; however, in this study, one RP50 was estimated under the assumption that the RP50 of a polytomous item can be considered as an appropriate mastery of the item.

The upper plot presents the scale score distribution of the test takers based on census data, including those who were administered the braille forms and Spanish language version of the mathematics forms. The X-axis shows the scale score. The Y-axis is the density of the scale scores. The density is the number of students with a scale score divided by the total number of students who received a score.

The lower plot presents the RP50 values, as expressed on the scale score metric, for the spring 2023 LEAP 2025 operational items. The X-axis shows the scale score; this is the same scale as the upper plot. The Y-axis is a subcategory: RI, RL, RV, WE, & WKL for ELA and A, B, C, and D for Math. Each red dot represents the RP50 value of an item aligned to the subcategory. The four vertical lines are the cut scores. For all ELA grades and mathematics grades 6, 7, and 8, most RP50 values were located in performance levels 3 and 4, which indicates many items were difficult for lower performing students. For Mathematics grades 3-5, most RP50 values were located in performance levels 2, 3, and 4 where most students are located.

Figure 6.27 Item Difficulty-Student Ability Map: ELA 3

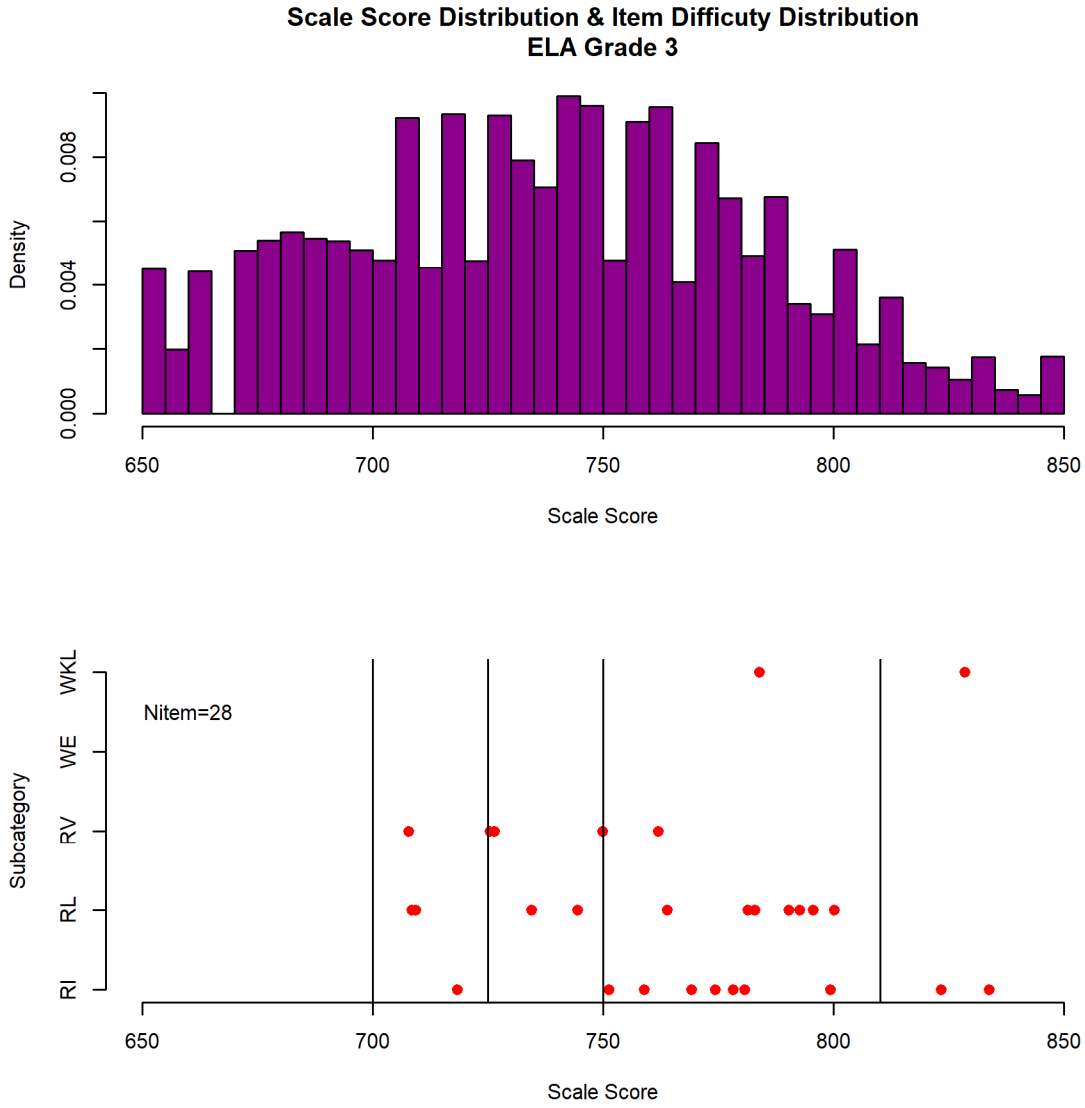


Figure 6.28 Item Difficulty-Student Ability Map: ELA 4

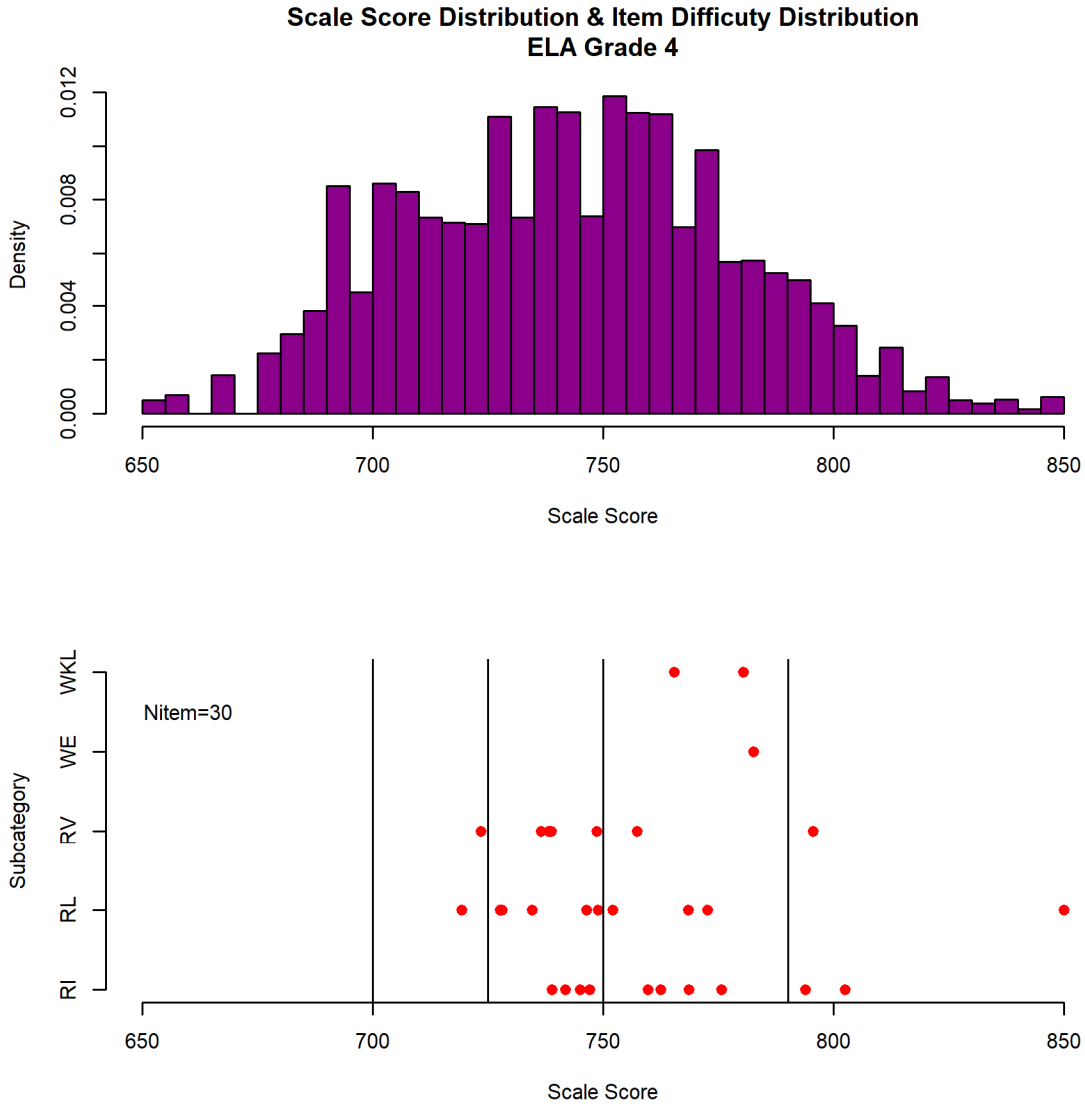


Figure 6.29 Item Difficulty-Student Ability Map: ELA 5

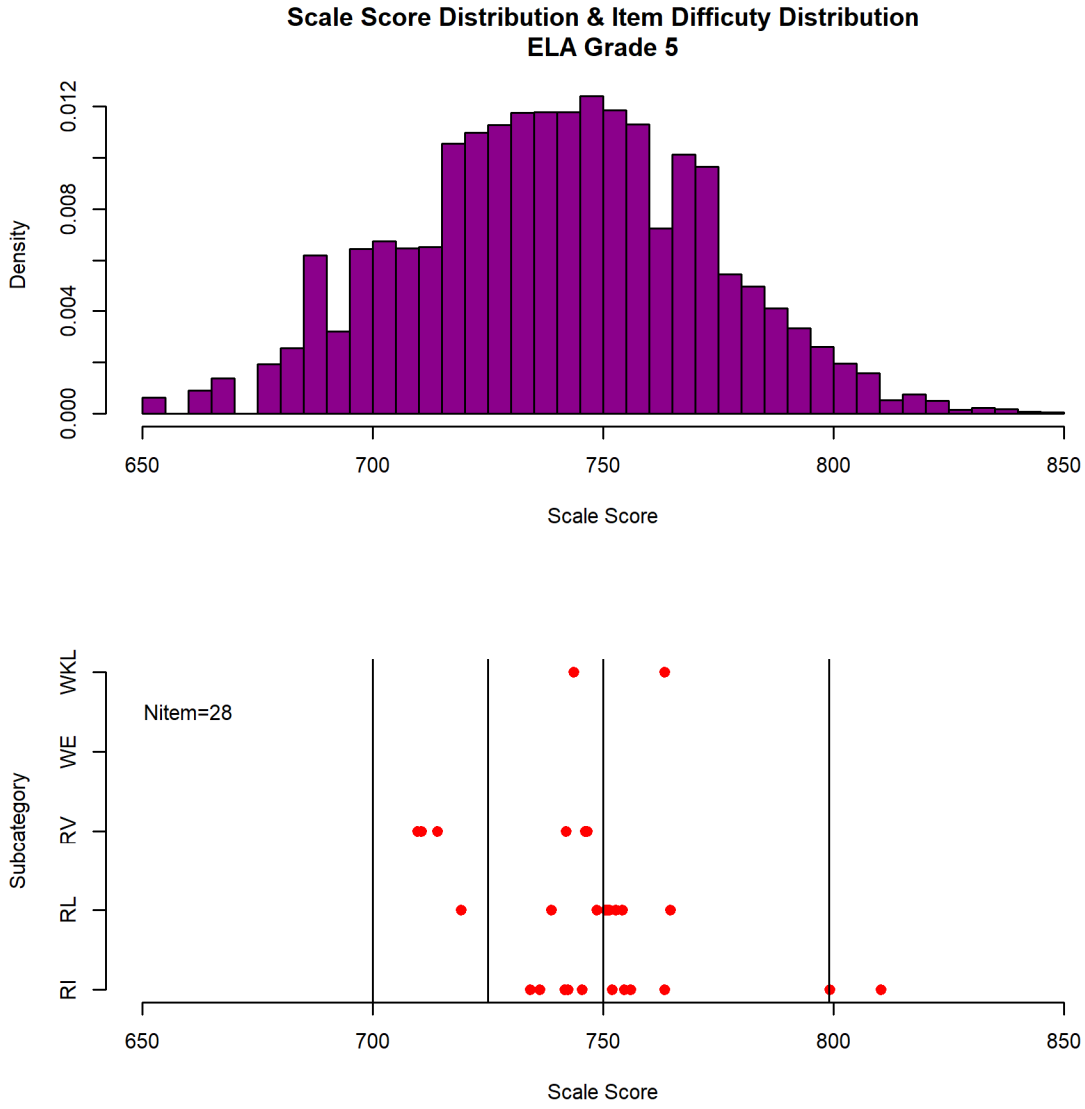


Figure 6.30 Item Difficulty-Student Ability Map: ELA 6

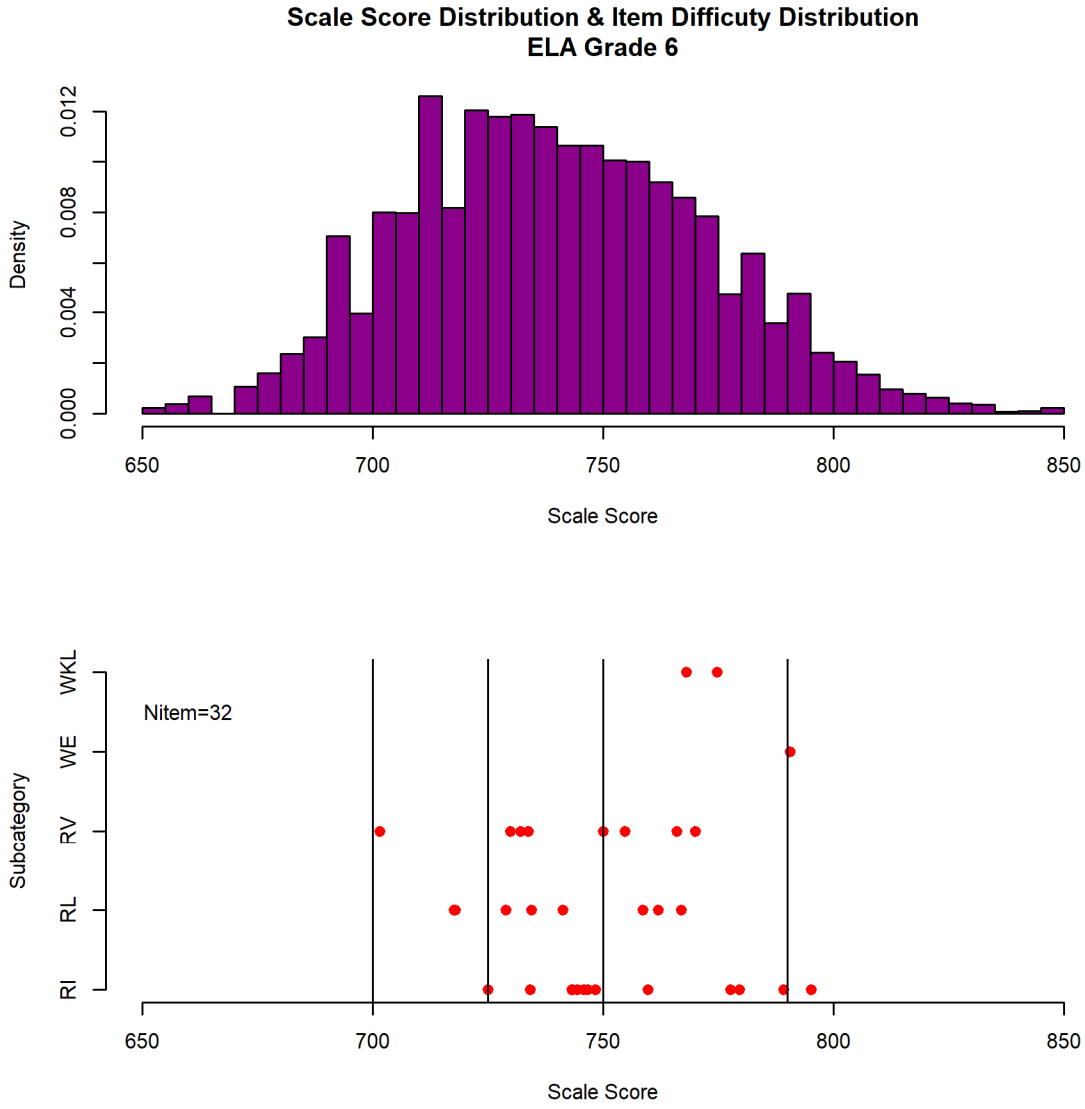


Figure 6.31 Item Difficulty-Student Ability Map: ELA 7

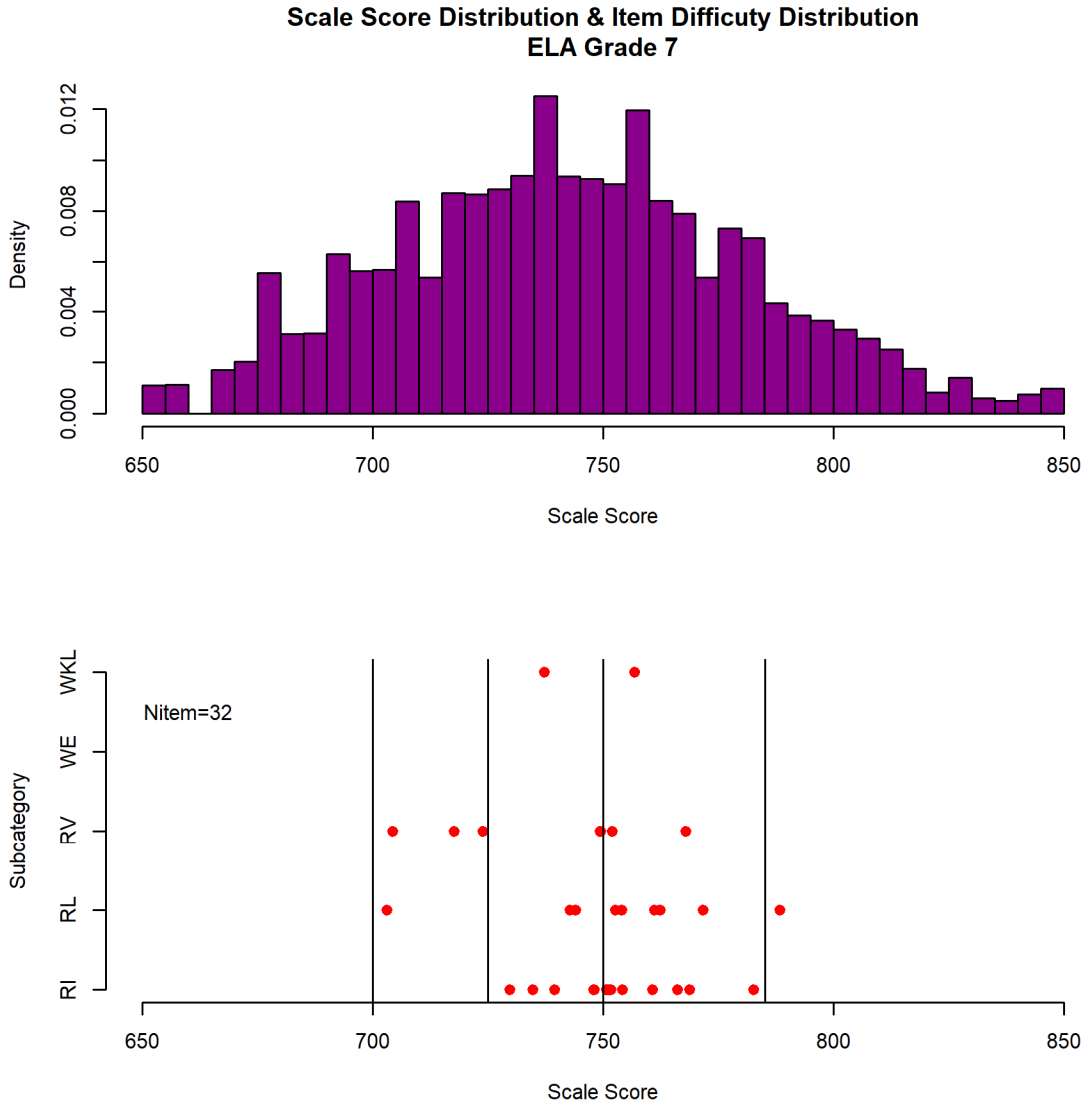


Figure 6.32 Item Difficulty-Student Ability Map: ELA 8

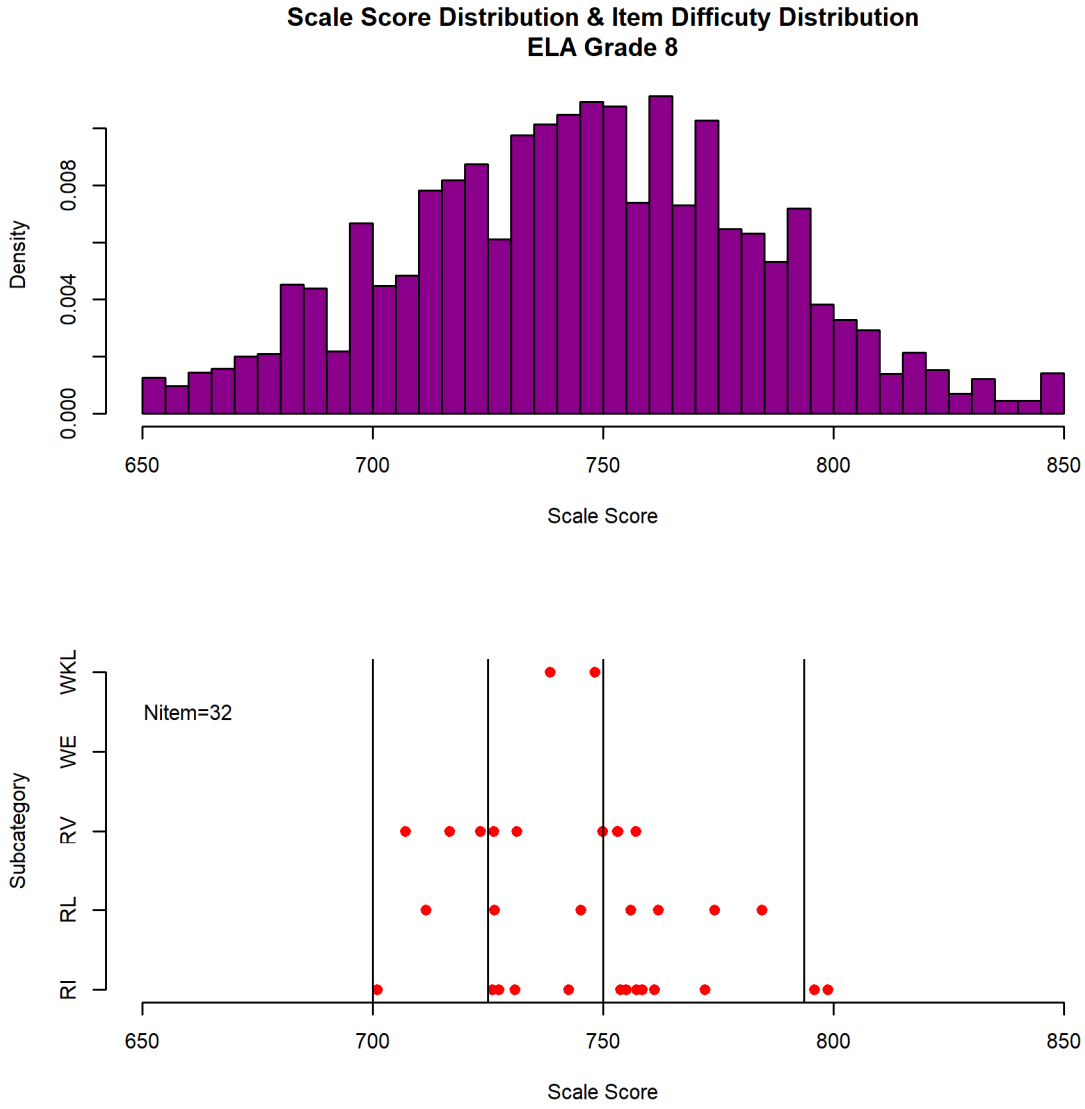


Figure 6.33 Item Difficulty-Student Ability Map: Mathematics 3

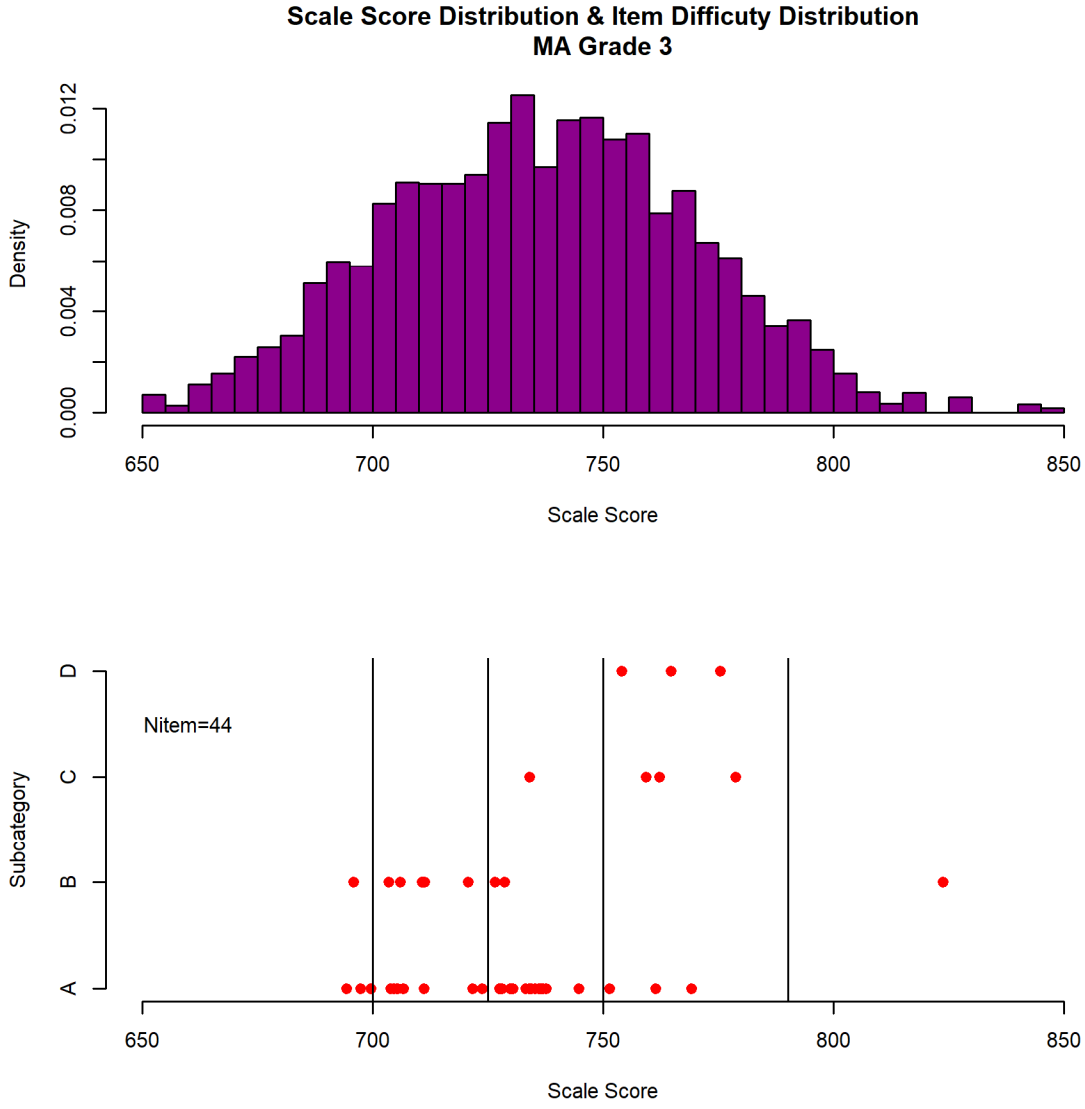


Figure 6.34 Item Difficulty-Student Ability Map: Mathematics 4

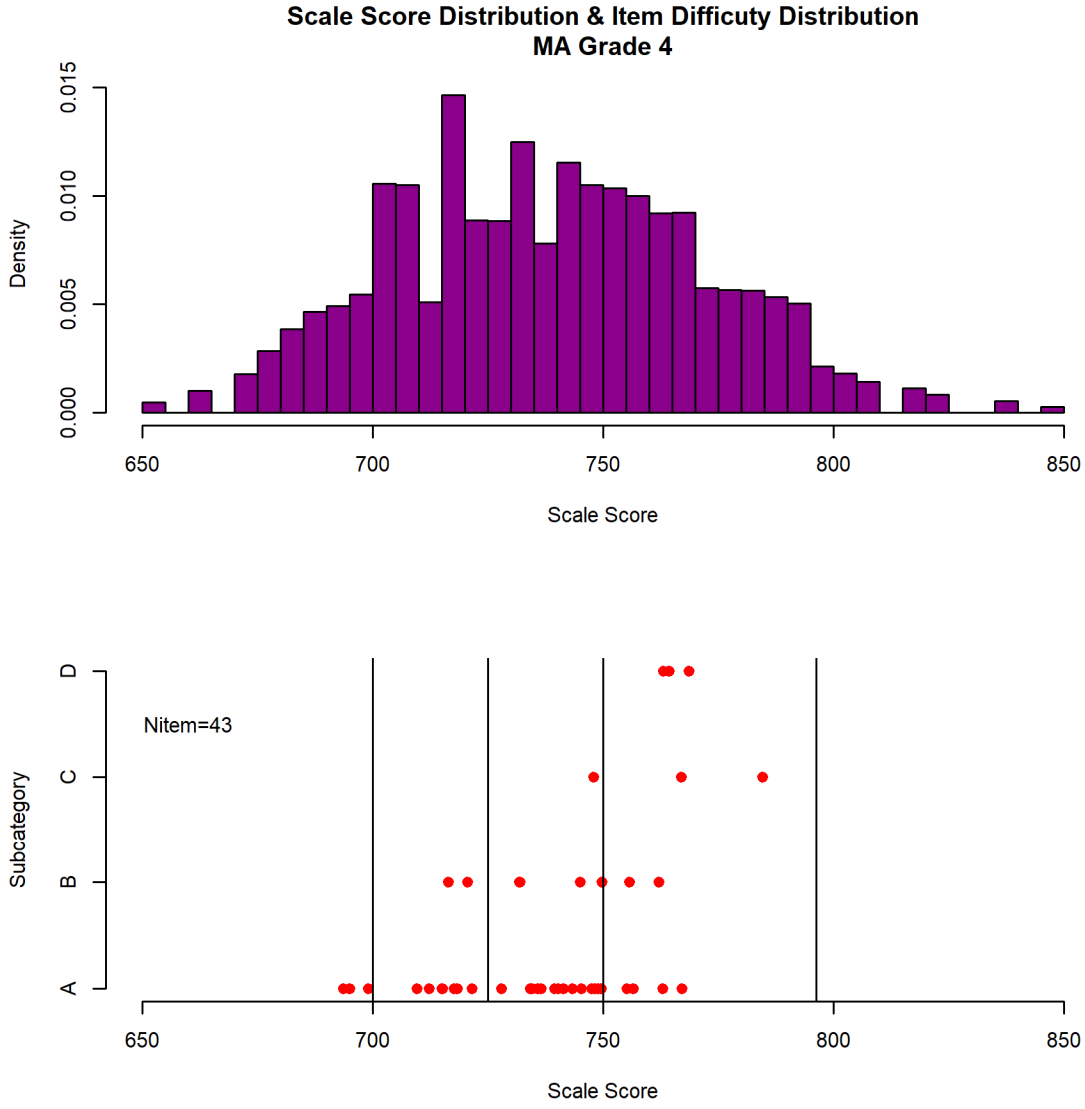


Figure 6.35 Item Difficulty-Student Ability Map: Mathematics 5

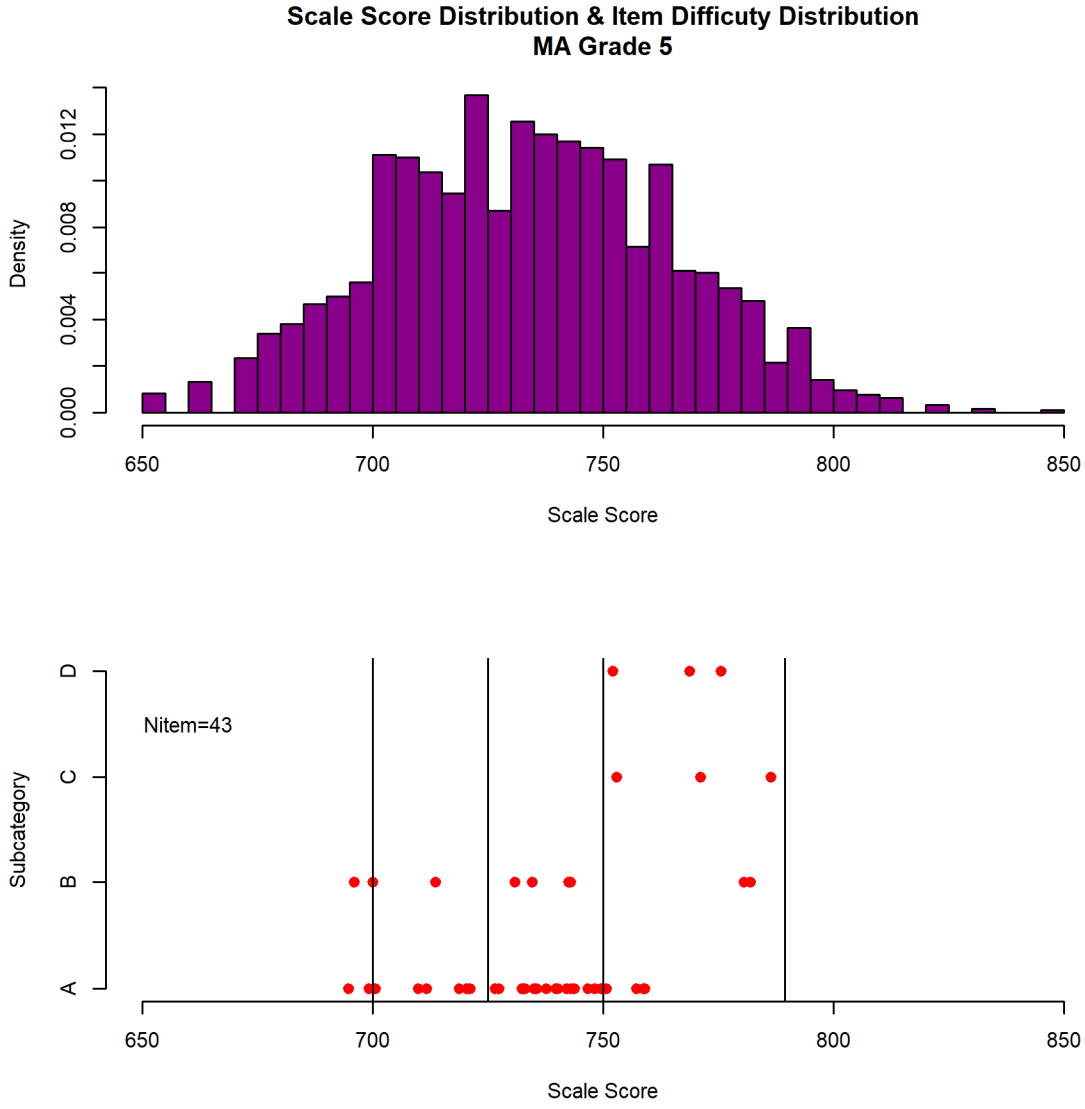


Figure 6.36 Item Difficulty-Student Ability Map: Mathematics 6

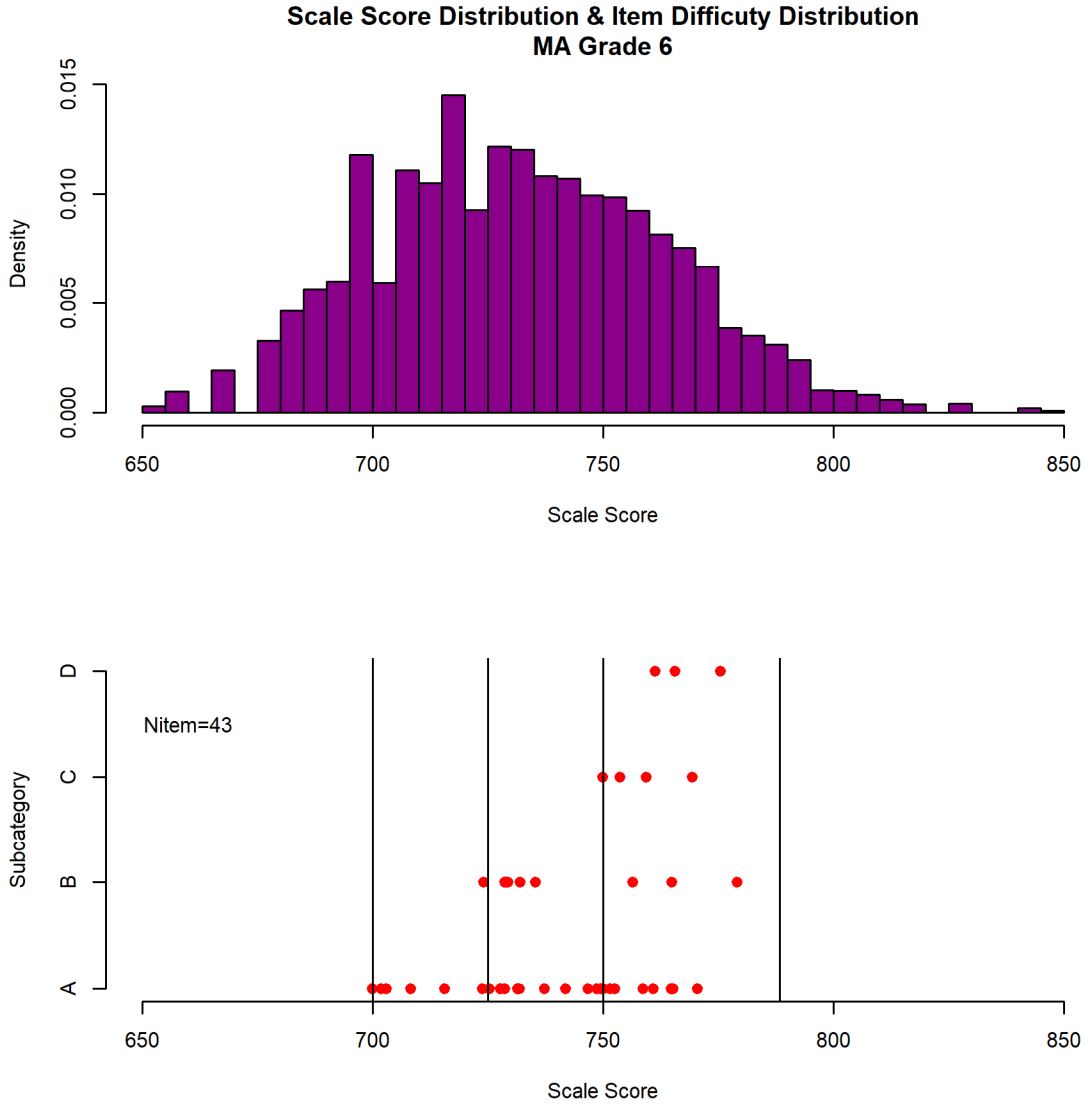


Figure 6.37 Item Difficulty-Student Ability Map: Mathematics 7

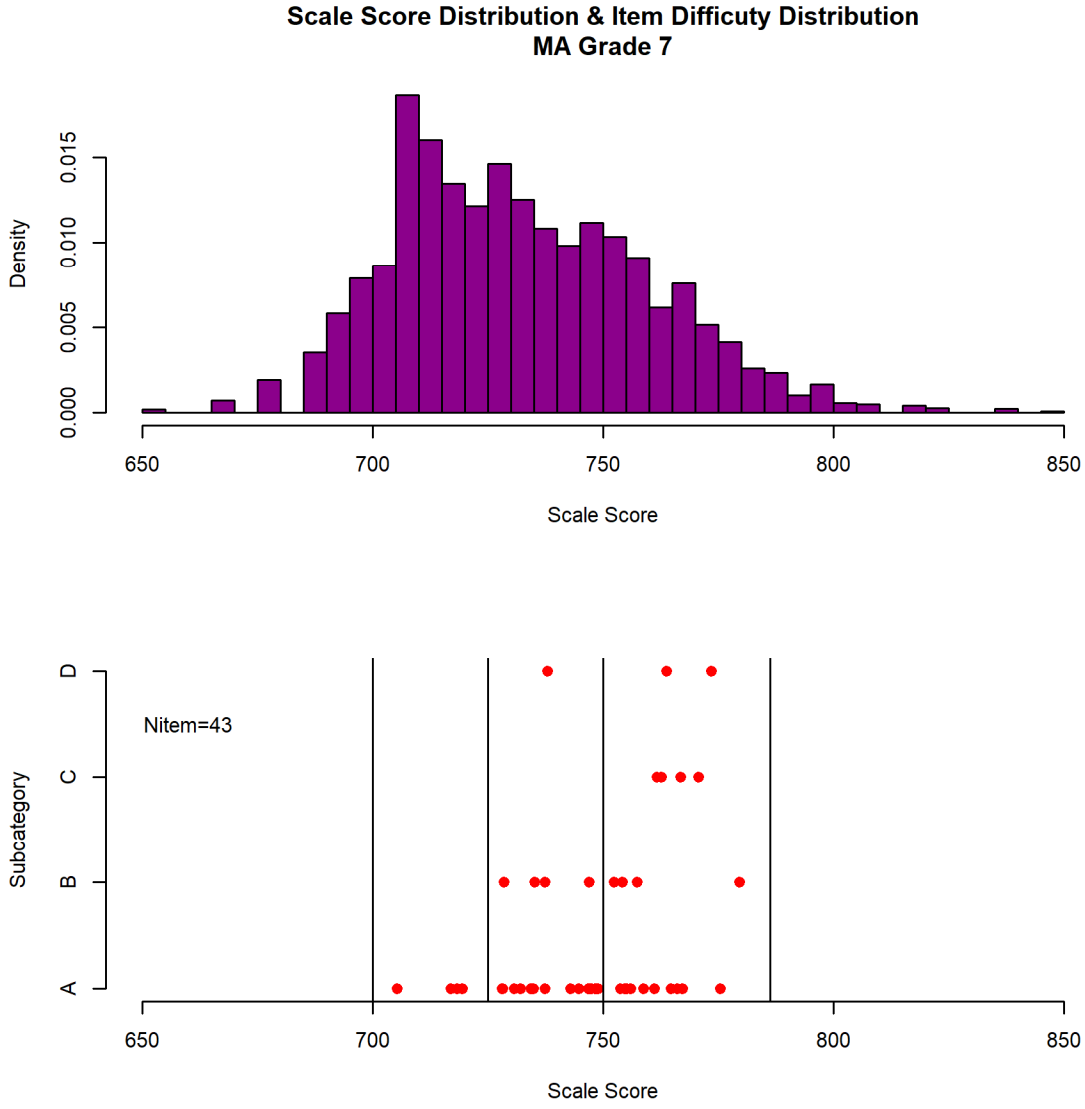
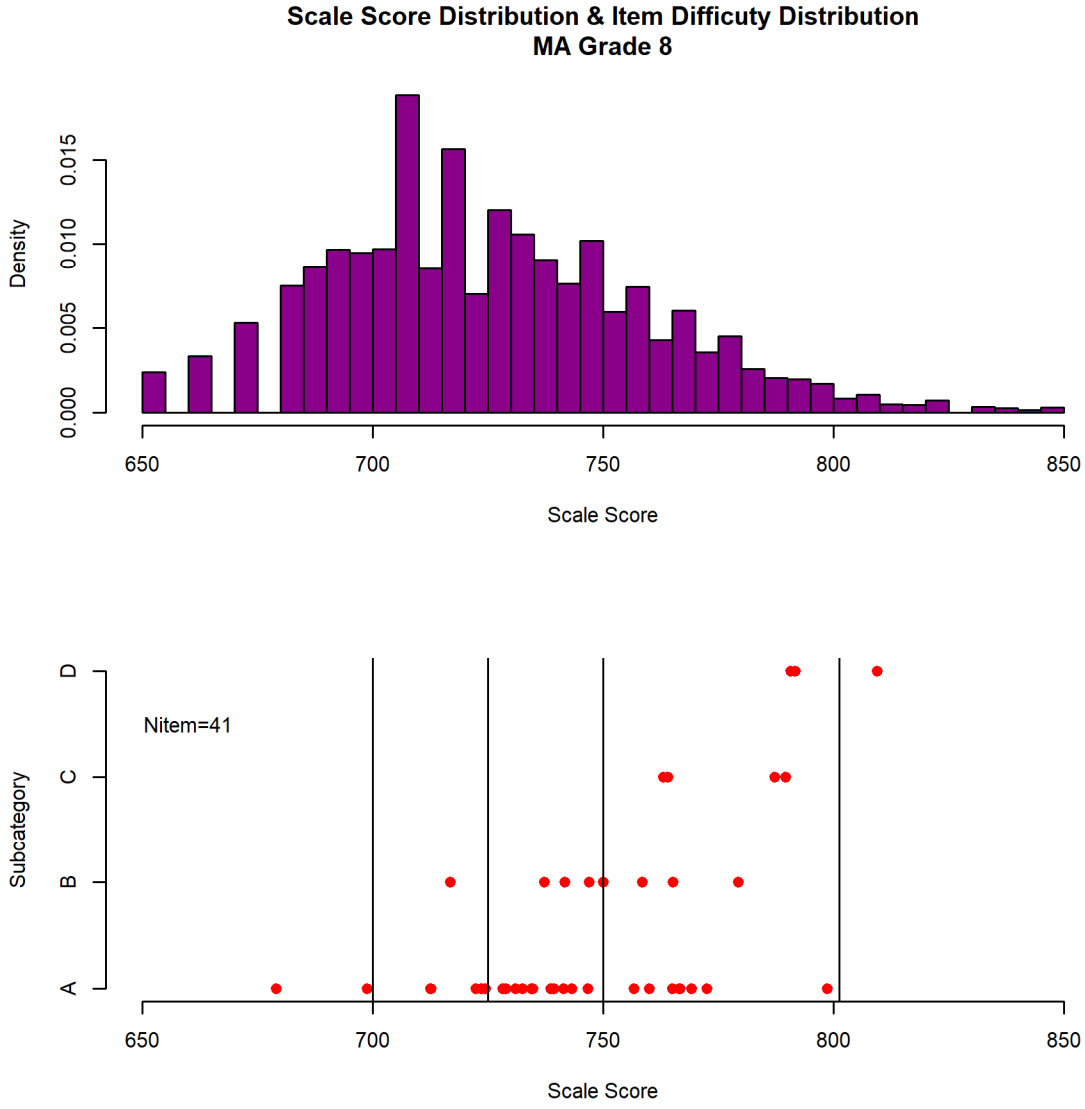


Figure 6.38 Item Difficulty-Student Ability Map: Mathematics 8



6.4.4 Across Year Form Comparability

The primary purpose of form equating is to establish score equivalency between two (or more) forms. Equivalency is established by first building the forms to be equated according to tight content specifications. Then the form scores are placed on the same scale (by equating), such that students performing on an assessment at the same level of (underlying) achievement should receive the same scale score, although they may not receive the same number-correct score (or raw score). The raw-to-scale-score relationship performs this leveling function based on form-equating studies. Theoretically, differences in the raw-to-scale-score relationship between the two forms can be partially due to differences in the samples utilized for calibration and the differences in item difficulty. The LDOE and DRC strive to maintain equivalent samples or use near-census samples over the years, minimizing the potential differences due to the samples. Differences in the raw-to-scale-score relationship, therefore, can be primarily attributed to the differences in item difficulty.

The forms used in the spring 2023 were post-equated forms. Just as in previous years, equating was conducted using the test characteristic transformation function method in the common-item non-equivalent-groups design (Stocking & Lord, 1983). Tables 6.27 through 6.38. provide scale scores at selected percentiles that can be used to compare the distributional characteristics of the Spring 2023 forms to previous administrations, based on census data. Although these scale scores are rounded values, there were differences in the scale-score values for a given percentile across the forms. These variations could arise for several reasons: (1) differences in the proficiency (i.e., achievement) of students in the samples or growth in student achievement across years; (2) unevenness in the respective distributions that combine with the number-correct-to-scale-score scoring method, leaving "gaps" in the scale; or (3) other sources of equating error. Other sources of equating error can include subtle content differences between forms, handscoring differences, or unusual student samples. Some equating errors will always be present between forms. This means that the forms will not measure identically, even under optimal testing conditions. In general, however, the test characteristic function equating techniques will "level" the equated forms through the raw-to-scale-score adjustment.

Table 6.27 Comparisons of Scale Scores at Selected Percentiles—Grade 3 ELA

	2016	2017	2018	2019	2021	2022	2023
Percentile	**Form A**	**Form B**	**Form C**	**Form D**	**Form D**	**Form E**	**Form F**
99	822	839	842	845	845	839	844
95	796	810	810	816	812	809	815
90	783	793	797	802	795	792	797
85	774	784	788	792	785	783	787
80	768	775	779	782	776	773	780
75	762	770	773	776	767	767	771
70	757	762	768	770	761	761	765
65	751	757	762	764	755	755	759
60	746	752	757	758	749	749	753
55	741	748	752	752	743	743	747
50	738	743	746	746	737	736	741
45	732	739	741	740	731	730	735
40	727	734	736	734	725	724	727
35	721	727	730	728	719	718	720
30	715	723	724	722	712	711	714
25	712	718	715	715	708	705	706
20	706	710	708	708	700	697	698
15	695	701	701	700	690	688	690
10	687	695	692	690	679	678	679
5	676	679	676	679	664	662	665
1	654	655	650	650	650	650	650

Table 6.28 Comparisons of Scale Scores at Selected Percentiles—Grade 4 ELA

	2016	2017	2018	2019	2021	2022	2023
Percentile	**Form A**	**Form B**	**Form C**	**Form D**	**Form D**	**Form E**	**Form F**
99	816	818	821	824	828	826	829
95	794	796	800	801	802	802	802
90	785	785	789	789	789	788	792
85	777	777	778	780	780	782	782
80	769	771	774	774	772	773	775
75	765	765	767	768	766	767	769
70	760	761	763	762	761	762	763
65	755	756	757	758	755	758	758
60	751	752	753	753	751	752	754
55	746	748	749	750	746	748	749
50	744	744	744	744	742	744	743
45	740	741	740	741	737	738	740
40	735	737	736	736	732	734	734
35	731	733	731	731	727	728	728
30	727	728	727	726	721	723	724
25	722	724	721	721	716	716	718
20	715	717	714	714	709	711	710
15	709	711	707	706	703	702	705
10	701	702	698	699	693	695	695
5	691	691	687	688	684	682	687
1	666	670	668	665	664	661	669

Table 6.29 Comparisons of Scale Scores at Selected Percentiles—Grade 5 ELA

	2016	2017	2018	2019	2021	2022	2023
Percentile	**Form A**	**Form B**	**Form C**	**Form D**	**Form D**	**Form E**	**Form F**
99	816	813	817	821	821	818	813
95	792	793	795	798	798	796	793
90	782	782	782	784	784	781	782
85	774	775	777	776	776	775	773
80	767	769	769	770	768	768	767
75	763	763	765	765	763	762	762
70	758	758	760	759	758	759	758
65	754	754	756	754	752	754	753
60	749	750	753	751	747	749	750
55	745	747	749	745	742	746	745
50	740	743	746	742	738	741	742
45	738	739	740	737	733	738	737
40	733	735	736	733	729	734	734
35	728	731	732	729	725	729	729
30	723	727	728	725	718	724	723
25	720	721	724	718	713	718	720
20	714	716	716	713	710	713	714
15	708	709	711	707	704	706	707
10	701	701	702	701	697	698	697
5	692	691	691	693	688	687	686
1	675	673	676	676	676	669	669

Table 6.30 Comparisons of Scale Scores at Selected Percentiles—Grade 6 ELA

	2016	2017	2018	2019	2021	2022	2023
Percentile	**Form A**	**Form B**	**Form C**	**Form D**	**Form D**	**Form E**	**Form F**
99	813	814	808	812	812	815	818
95	792	790	789	791	788	794	795
90	780	779	777	778	776	783	783
85	772	770	770	771	769	774	775
80	765	763	763	766	762	768	768
75	760	759	758	761	758	763	763
70	756	754	753	756	753	756	758
65	752	748	749	751	748	752	753
60	748	745	746	747	744	747	749
55	745	741	742	743	740	743	744
50	741	736	737	740	735	738	739
45	737	733	735	735	731	735	735
40	734	729	730	731	726	730	730
35	730	724	726	728	723	725	727
30	727	721	721	723	718	721	721
25	723	716	718	718	714	716	717
20	718	711	713	714	708	709	711
15	713	705	707	708	703	704	706
10	706	698	700	701	698	696	698
5	696	689	691	692	688	688	691
1	676	671	675	675	675	674	672

Table 6.31 Comparisons of Scale Scores at Selected Percentiles—Grade 7 ELA

	2016	2017	2018	2019	2021	2022	2023
Percentile	**Form A**	**Form B**	**Form C**	**Form D**	**Form D**	**Form E**	**Form F**
99	825	826	831	826	834	828	837
95	800	800	801	804	804	804	808
90	787	786	789	789	789	791	794
85	777	778	780	782	780	783	783
80	771	770	774	775	773	775	776
75	766	765	767	769	767	769	769
70	761	759	762	764	761	764	764
65	756	756	757	759	756	758	757
60	751	751	752	756	751	753	752
55	747	745	749	750	747	748	748
50	742	742	744	747	742	743	742
45	740	737	740	741	738	738	737
40	735	733	735	736	733	733	732
35	730	728	730	731	728	728	728
30	726	723	726	727	722	723	721
25	721	717	719	720	716	715	716
20	714	711	713	714	710	709	708
15	706	702	707	705	703	700	701
10	697	692	697	695	692	689	691
5	683	675	685	681	681	674	680
1	655	654	662	659	659	658	659

Table 6.32 Comparisons of Scale Scores at Selected Percentiles—Grade 8 ELA

	2016	2017	2018	2019	2021	2022	2023
Percentile	**Form A**	**Form B**	**Form C**	**Form D**	**Form D**	**Form E**	**Form F**
99	825	834	824	831	831	836	839
95	804	806	801	804	806	809	810
90	790	791	789	793	793	795	795
85	781	782	781	785	783	786	786
80	775	776	774	777	775	778	779
75	770	770	768	771	769	772	773
70	764	764	764	766	764	766	767
65	759	758	758	760	758	761	761
60	754	754	754	755	753	755	757
55	752	749	751	750	748	750	751
50	747	745	745	746	743	746	746
45	743	740	741	741	738	741	741
40	739	734	737	736	734	735	736
35	735	731	732	732	728	730	731
30	731	725	726	727	723	724	725
25	727	719	722	721	717	719	720
20	721	714	716	714	710	712	713
15	714	707	708	707	702	703	705
10	706	696	699	696	693	695	693
5	693	681	683	686	682	681	681
1	670	651	657	667	660	660	659

Table 6.33 Comparisons of Scale Scores at Selected Percentiles—Grade 3 Mathematics

	2016	2017	2018	2019	2021	2022	2023
Percentile	**Form A**	**Form B**	**Form C**	**Form D**	**Revised Form D**	**Form E**	**Form F**
99	824	822	817	815	816	810	811
95	802	796	793	796	790	789	791
90	789	786	783	784	778	779	779
85	781	776	775	776	768	771	772
80	775	772	771	771	764	767	766
75	770	765	764	764	758	760	760
70	765	761	759	760	752	756	755
65	760	756	755	756	748	750	751
60	756	752	750	752	742	746	746
55	751	747	746	748	738	742	742
50	746	743	742	744	734	738	737
45	741	738	740	738	727	733	733
40	738	733	735	735	723	727	729
35	733	728	731	731	719	722	723
30	728	725	726	724	711	718	719
25	722	720	719	720	706	713	713
20	716	715	713	713	700	705	708
15	710	706	708	705	694	700	701
10	703	699	698	700	686	694	694
5	692	689	686	686	677	683	684
1	672	667	664	672	658	669	664

Table 6.34 Comparisons of Scale Scores at Selected Percentiles—Grade 4 Mathematics

	2016	2017	2018	2019	2021	2022	2023
Percentile	Form A	Form B	Form C	Form D	Revised Form D	Form E	Form F
99	819	812	812	813	803	808	816
95	797	792	790	792	785	790	795
90	786	779	780	781	775	779	783
85	777	774	772	774	768	772	774
80	771	767	768	769	762	766	768
75	766	762	762	763	757	760	763
70	761	756	757	759	751	756	758
65	756	752	753	755	746	750	753
60	752	748	749	750	741	746	748
55	747	744	744	746	737	740	742
50	743	740	740	742	732	736	739
45	738	736	735	737	726	730	733
40	732	732	733	732	722	726	729
35	728	727	728	728	717	720	723
30	723	722	723	724	711	715	718
25	718	717	718	719	706	710	713
20	713	712	715	712	699	705	707
15	708	706	710	706	693	702	701
10	703	700	700	699	688	695	697
5	693	693	689	688	679	687	689
1	677	674	670	673	658	671	671

Table 6.35 Comparisons of Scale Scores at Selected Percentiles—Grade 5 Mathematics

	2016	2017	2018	2019	2021	2022	2023
Percentile	Form A	Form B	Form C	Form D	Revised Form D	Form E	Form F
99	819	808	810	809	803	807	804
95	792	784	784	788	782	787	788
90	779	774	774	778	772	774	777
85	771	767	765	769	765	768	767
80	766	760	759	763	757	761	761
75	759	755	755	757	751	755	757
70	754	751	749	753	747	749	752
65	749	747	745	748	741	745	746
60	745	742	743	744	737	739	743
55	740	740	738	740	733	735	738
50	735	735	734	737	729	731	734
45	731	730	729	733	724	727	729
40	728	728	727	728	719	722	725
35	722	723	722	724	716	717	721
30	720	720	720	719	710	712	717
25	714	715	714	714	707	709	712
20	711	709	711	711	703	703	707
15	705	706	705	705	699	700	701
10	699	699	698	699	690	692	694
5	691	691	689	690	685	688	685
1	678	675	672	674	671	670	665

Table 6.36 Comparisons of Scale Scores at Selected Percentiles—Grade 6 Mathematics

	2016	2017	2018	2019	2021	2022	2023
Percentile	Form A	Form B	Form C	Form D	Revised Form D	Form E	Form F
99	803	808	800	804	798	805	807
95	783	781	780	783	777	783	784
90	771	771	770	773	768	772	773
85	765	762	762	765	760	763	766
80	758	757	757	758	754	758	759
75	753	752	752	754	749	752	755
70	747	746	748	750	743	747	750
65	744	742	743	745	740	742	745
60	740	738	739	742	735	738	740
55	735	734	736	739	731	734	737
50	731	732	732	733	727	728	732
45	729	727	728	729	723	724	728
40	724	724	723	725	718	720	722
35	722	719	721	721	713	715	718
30	717	717	716	717	710	713	713
25	714	711	713	714	704	708	708
20	709	708	707	709	701	702	703
15	706	701	704	703	693	695	699
10	699	697	696	696	689	692	692
5	692	688	686	687	683	683	682
1	679	671	672	667	656	663	668

Table 6.37 Comparisons of Scale Scores at Selected Percentiles—Grade 7 Mathematics

	2016	2017	2018	2019	2021	2022	2023
Percentile	**Form A**	**Form B**	**Form C**	**Form D**	**Form D**	**Form E**	**Form F**
99	797	796	797	796	793	799	800
95	779	777	777	776	773	776	780
90	768	766	766	766	764	767	770
85	760	760	759	761	757	760	763
80	754	754	755	756	752	755	756
75	750	749	750	752	748	749	751
70	746	746	745	748	743	745	747
65	742	741	742	743	740	741	742
60	738	737	739	740	736	736	737
55	734	734	735	736	732	733	734
50	730	731	731	732	728	729	730
45	728	727	729	730	724	725	727
40	723	723	725	726	722	723	722
35	721	721	721	722	719	718	718
30	719	717	718	719	714	715	715
25	714	712	713	714	711	709	712
20	712	709	710	711	708	706	709
15	706	706	706	705	701	702	706
10	703	699	702	701	697	697	698
5	695	694	693	692	687	686	693
1	678	673	679	680	671	666	679

Table 6.38 Comparisons of Scale Scores at Selected Percentiles—Grade 8 Mathematics

	2016	2017	2018	2019	2021	2022	2023
Percentile	Form A	Form B	Form C	Form D	Revised Form D	Form E	Form F
99	808	809	807	812	806	814	816
95	787	784	784	788	781	783	787
90	775	771	773	775	768	771	772
85	766	763	764	766	759	760	762
80	761	757	757	758	751	754	754
75	753	751	752	752	747	748	748
70	749	746	746	746	740	741	741
65	744	741	742	742	735	737	736
60	737	736	737	737	732	732	731
55	734	730	732	732	726	729	726
50	731	727	727	730	723	723	723
45	727	724	721	724	716	720	717
40	724	718	718	721	712	713	713
35	720	714	715	715	708	709	710
30	712	710	707	711	703	705	706
25	708	706	702	707	698	700	702
20	704	698	697	699	693	695	698
15	699	693	691	694	686	690	693
10	695	687	684	689	679	683	687
5	684	674	676	677	671	666	674
1	663	656	654	659	650	650	653

Additional evidence of comparability can be found by reviewing the test characteristic curves (TCCs) for the LEAP 2025 across administrations, see figures 6.39 and 6.40. For most content areas and grades, the TCCs for the six years were similar across ability ranges. Grades 3 and 8 forms have been gradually becoming more difficult from 2017 to 2023. For ELA grade 5 and 6, the 2018 forms were slightly easier than the 2017 and 2019 forms for high-performing students. Please note that ELA 2019 and 2021 administrations used the same forms.

Except for mathematics grade 5, 2017 to 2022 mathematic forms were similar across most ability ranges. For grade 5, the 2019/2021 forms were easier than the 2017 and 2018 forms for high-performing students. Please note that most items on the mathematics 2021 forms were the same as items on the 2019 forms.

Note that this different form difficulty is adjusted by reporting different scale scores for given raw scores; a scale score of a difficult form is higher than that of an easy form given the same raw score.

Figures 6.41 and 6.42 show SEMs for the 2017- 2023 LEAP 2025 assessments. For most content areas and grades, the SEMs were similar across ability ranges, especially in the middle ability ranges.

Figure 6.39 TCCs Across Years: ELA

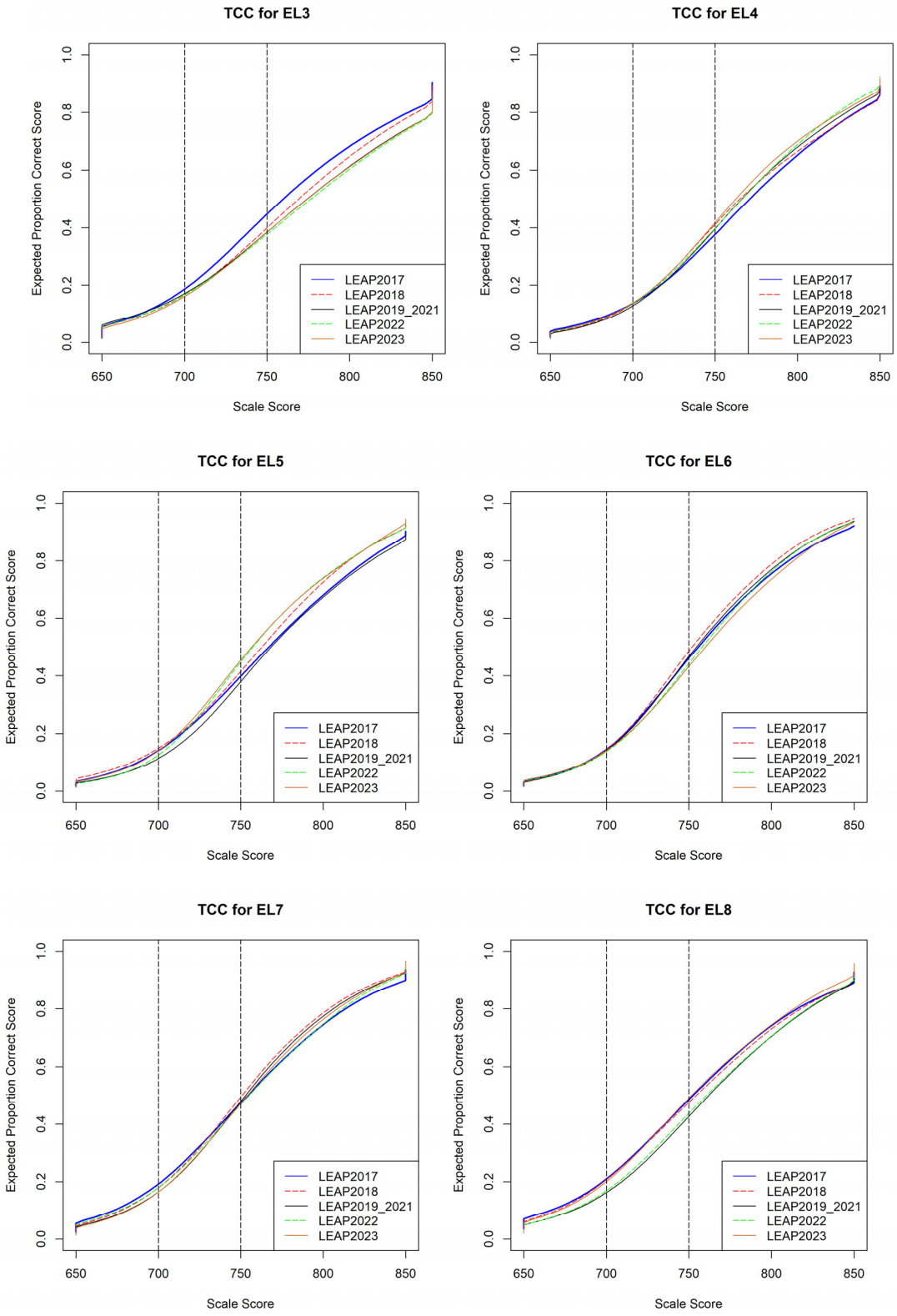


Figure 6.40 TCCs Across Years: Mathematics

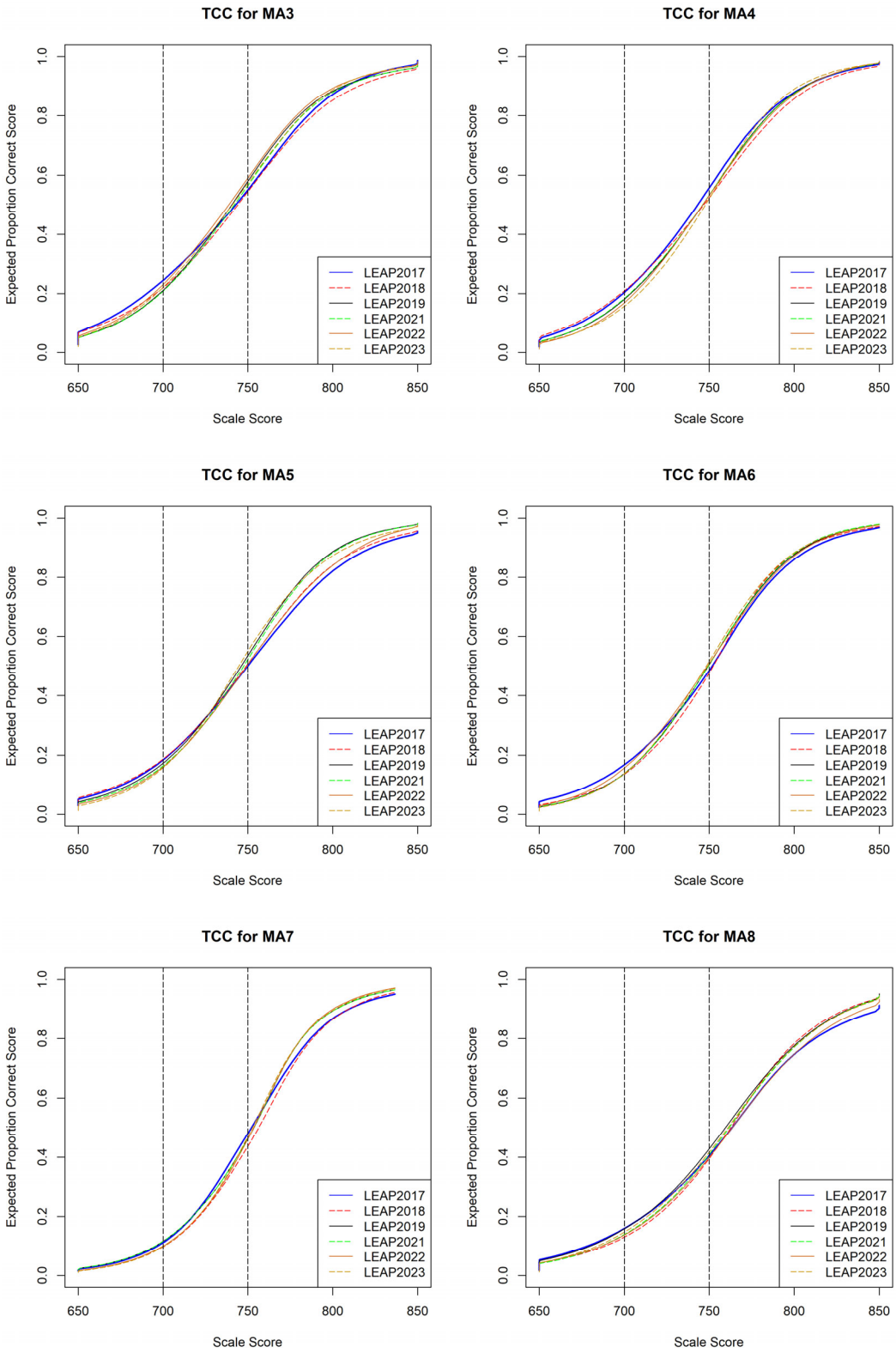


Figure 6.41 SEM Across Years: ELA

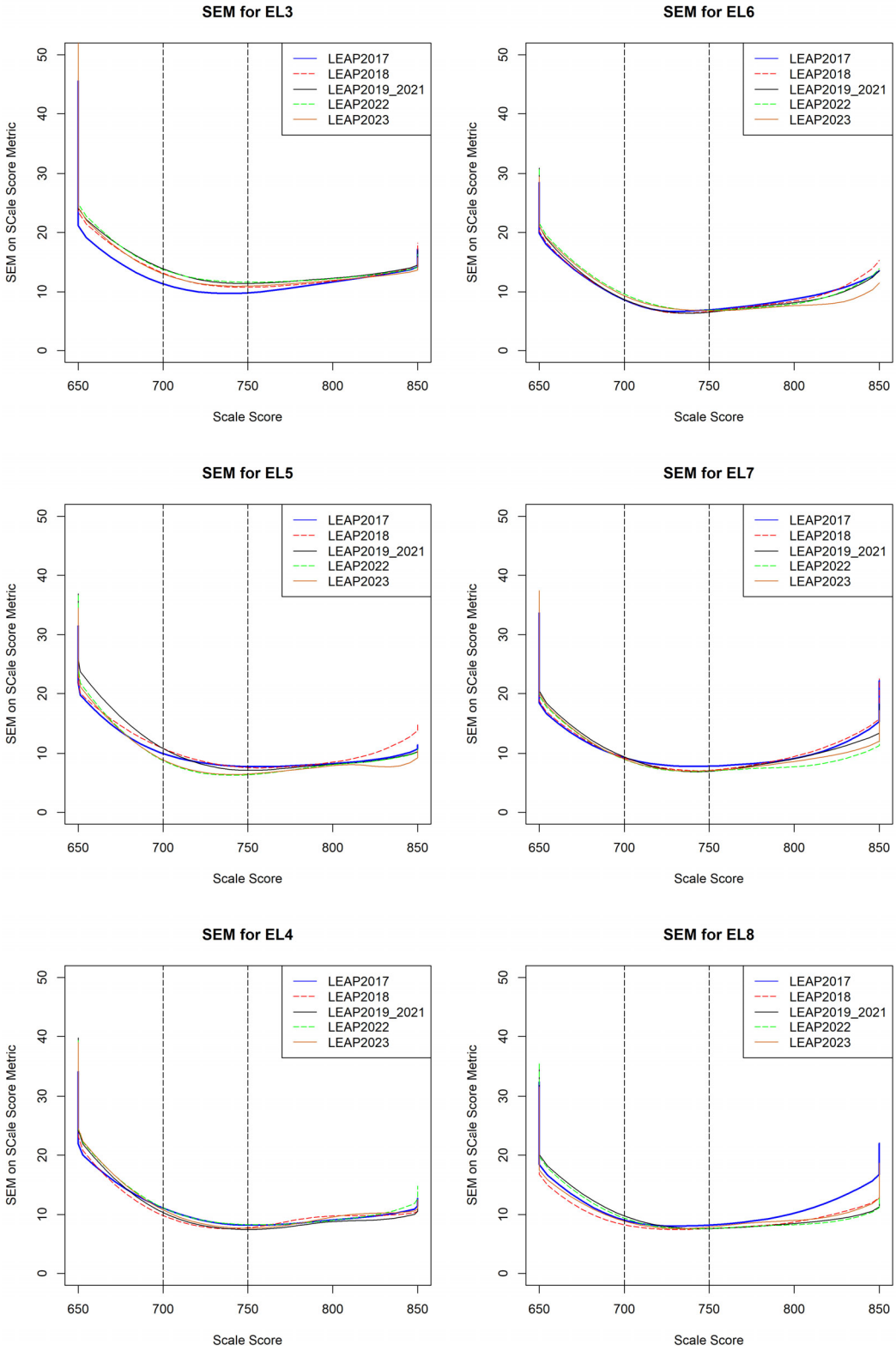
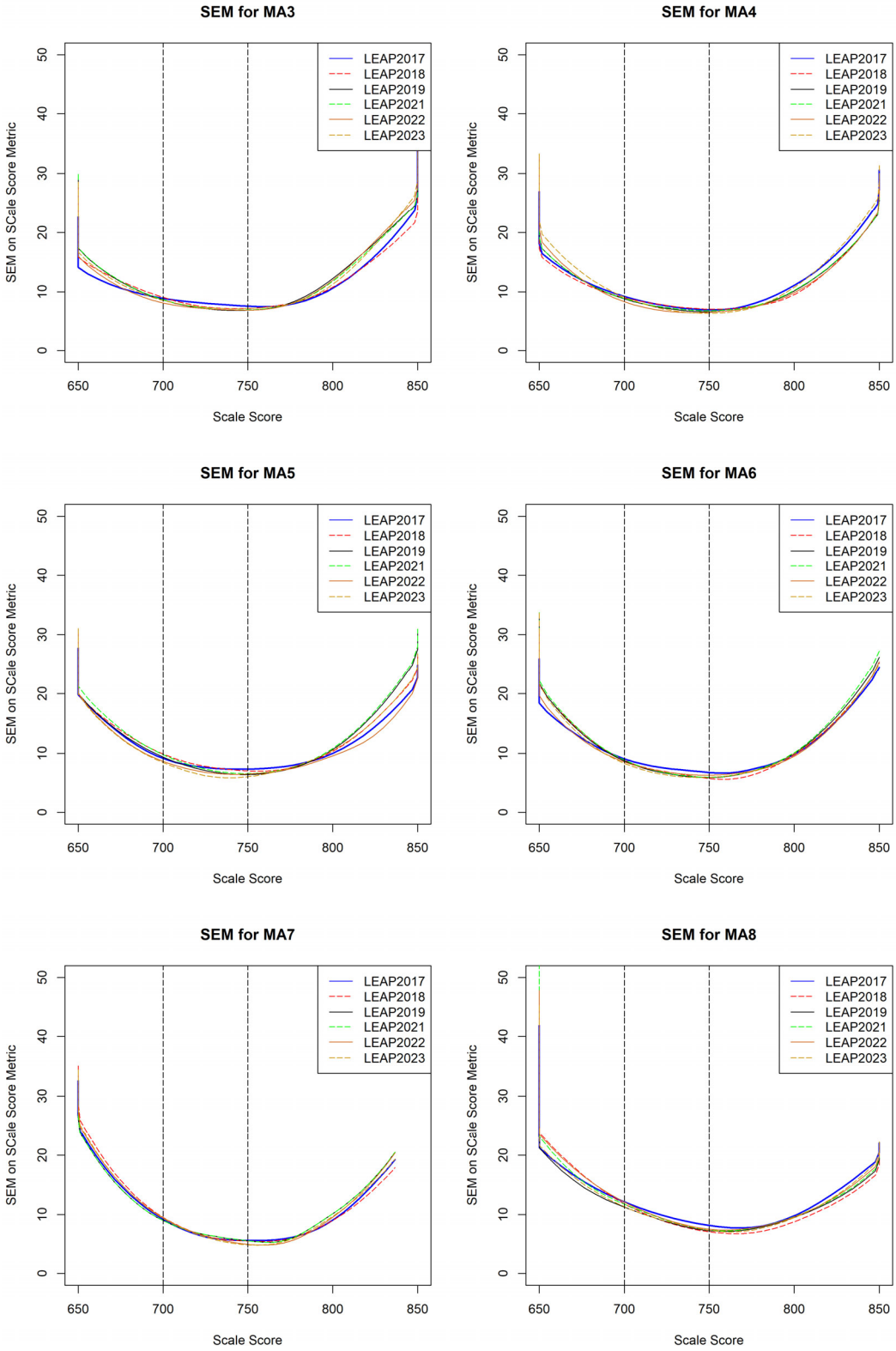


Figure 6.42 SEM Across Years: Mathematics



6.5 Summary

In summary, the overall purpose of the operational data analyses is to ensure that the test items, as well as the overall test, are functioning appropriately. Operational data analyses also help maintain the test scale so that test results may be appropriately compared across years. The data analyses undertaken by DRC address multiple best practices of the testing industry but are particularly related to the following standards:

> ***Standard 1.8*** The composition of any sample of test takers from which validity evidence is obtained should be described in as much detail as is practical and permissible, including major relevant socio-demographic and developmental characteristics (25).
>
> ***Standard 4.14*** For a test that has a time limit, test development research should examine the degree to which scores include a speed component and should evaluate the appropriateness of that component, given the domain the test is designed to measure (90).
>
> ***Standard 5.2*** The procedures for constructing scales used for reporting scores and the rationale for these procedures should be described clearly (102).
>
> ***Standard 5.13*** When claims of form-to-form score equivalence are based on equating procedures, detailed technical information should be provided on the method by which equating functions were established and on the accuracy of the equating functions (105).
>
> ***Standard 5.15*** In equating studies that employ an anchor test design, the characteristics of the anchor test and its similarity to the forms being equated should be presented, including both content specifications and empirically determined relationships among test scores. If anchor items are used in the equating study, the representativeness and psychometric characteristics of the anchor items should be presented (105).
>
> ***Standard 7.2*** The population for whom a test is intended and specifications for the test should be documented. If normative data are provided, the procedures used to gather the data should be explained; the norming population should be described in terms of relevant demographic variables; and the year(s) in which the data were collected should be reported (126).

Chapter 7: Test Results

This chapter of the technical report contains information on the results of the spring 2023 LEAP 2025 ELA and mathematics assessments. The scale score results and achievement level information are presented here. Presenting the results by achievement level translates the quantitative scale provided through scale scores into a qualitative description of student achievement. The levels are *Advanced*, *Mastery*, *Basic*, *Approaching Basic*, and *Unsatisfactory*.

While the scale score provides an essential quantitative reference for student achievement, the achievement-level information plainly outlines the meanings of the scores to parents, students, and educators. When combined, scale scores and achievement levels provide a comprehensive set of tools to assess Louisiana student achievement by content and grade level.

This chapter also provides descriptions of the score reports, data structure, and interpretive guide. The American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME, 2014) *Standards for Educational & Psychological Testing* addressed in Chapter 7 are 5.1, 6.10, 7.0, and 12.18. Each standard is presented in the pertinent section of this chapter.

The results presented in this chapter are based on census data. The results presented here may differ slightly from the official state summary report of all student populations due to ongoing resolution of test materials and student information. The results in the tables in this chapter are presented as evidence of the reliability and validity of the scores from the LEAP 2025 assessments and should not be used for state accountability purposes.

The following are subgroups reported during the administration of the LEAP 2025 tests:

- Gender: Female and Male
- Race and Ethnicity: Hispanic/Latino, American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, and Two or More Races
- Education Classification
- Economic Status
- English Learner Status (EL)
- Section 504 Status
- Migrant Status
- Homeless Status
- Foster Care Status
- Military Affiliation

For the purposes of this report, participation rate is defined as the percentage of students who received a valid scale score given the total number of students who were expected to take the online test or receive a test book. These participation rates are summarized in Table 7.1. Both the percentage of students classified as reportable and the number of students classified as accountable are reported. Reportable students include all students with a valid scale score. The “Accountable” columns show the total numbers of students who were expected to take the online test or receive a test book. These include students who should have received a LEAP 2025 scale score but who did not take the test and could not be assigned a scale score.

Table 7.1 Participation Rates

Participation Rates by Grade and Subgroup					
Grade	**Group**	**Accountable in ELA**	**Percentage Reportable in ELA**	**Accountable in Mathematics**	**Percentage Reportable in Mathematics**
3	All Students	≥49,370	99.65%	≥49,660	99.64%
	Gender				
	Female	≥24,210	99.75%	≥24,330	99.77%
	Male	≥25,130	99.63%	≥25,280	99.64%
	Ethnicity				
	Hispanic/Latino	≥4,960	99.76%	≥4,980	99.76%
	American Indian or Alaska Native	≥270	100.00%	≥270	100.00%
	Asian	≥740	100.00%	≥750	100.00%
	Black or African American	≥20,520	99.60%	≥20,690	99.64%
	Native Hawaiian or Other Pacific	≥50	100.00%	≥50	100.00%
	White	≥20,800	99.75%	≥20,850	99.73%
	Two or More Races	≥1,960	99.80%	≥1,970	99.80%
	Education Classification				
	Regular	≥43,090	99.69%	≥43,350	99.68%
	Special	≥6,270	99.39%	≥6,300	99.38%
	Economic Status				
	Not Economically Disadvantaged	≥14,820	99.64%	≥14,910	99.54%
	Economically Disadvantaged	≥34,550	99.66%	≥34,740	99.69%
	English Learner Status				
	Not English Learner	≥46,820	99.64%	≥47,090	99.63%
	English Learner	≥2,550	99.88%	≥2,570	99.88%
	Section 504 Status				
	Not Section 504	≥45,700	99.65%	≥45,970	99.64%
	Section 504	≥3,670	99.65%	≥3,680	99.67%
	Migrant Status				
	Not Migrant	≥49,270	99.65%	≥49,560	99.64%
	Migrant	≥90	100.00%	≥90	100.00%
	Homeless Status				
	Not Homeless	≥48,220	99.65%	≥48,480	99.64%
	Homeless	≥1,140	99.65%	≥1,170	99.66%
	Foster Care Status				
	Not in Foster Care	≥49,200	99.65%	≥49,480	99.64%
	In Foster Care	≥170	100.00%	≥170	100.00%
	Military Affiliation				
	Not Military Affiliated	≥48,390	99.65%	≥48,670	99.64%
	Military Affiliated	≥970	100.00%	≥980	100.00%

Participation Rates by Grade and Subgroup					
Grade	**Group**	**Accountable in ELA**	**Percentage Reportable in ELA**	**Accountable in Mathematics**	**Percentage Reportable in Mathematics**
4	All Students	≥48,770	99.81%	≥48,770	99.81%
	Gender				
	Female	≥23,830	99.82%	≥23,830	99.82%
	Male	≥24,930	99.80%	≥24,930	99.80%
	Ethnicity				
	Hispanic/Latino	≥4,880	99.86%	≥4,880	99.86%
	American Indian or Alaska Native	≥280	99.65%	≥280	99.65%
	Asian	≥820	99.88%	≥820	99.88%
	Black or African American	≥20,220	99.80%	≥20,220	99.81%
	Native Hawaiian or Other Pacific	≥40	100.00%	≥40	100.00%
	White	≥20,610	99.80%	≥20,610	99.79%
	Two or More Races	≥1,880	99.89%	≥1,880	99.89%
	Education Classification				
	Regular	≥42,720	99.84%	≥42,720	99.84%
	Special	≥6,050	99.59%	≥6,050	99.59%
	Economic Status				
	Not Economically Disadvantaged	≥14,860	99.86%	≥14,860	99.87%
	Economically Disadvantaged	≥33,900	99.78%	≥33,910	99.78%
	English Learner Status				
	Not English Learner	≥46,400	99.80%	≥46,400	99.80%
	English Learner	≥2,360	99.96%	≥2,360	99.96%
	Section 504 Status				
	Not Section 504	≥44,440	99.81%	≥44,450	99.81%
	Section 504	≥4,320	99.79%	≥4,320	99.82%
	Migrant Status				
	Not Migrant	≥48,710	99.81%	≥48,720	99.81%
	Migrant	≥50	100.00%	≥50	100.00%
	Homeless Status				
	Not Homeless	≥47,640	99.81%	≥47,650	99.81%
	Homeless	≥1,120	99.64%	≥1,120	99.65%
	Foster Care Status				
	Not in Foster Care	≥48,590	99.81%	≥48,600	99.81%
	In Foster Care	≥170	100.00%	≥170	100.00%
	Military Affiliation				
	Not Military Affiliated	≥47,850	99.81%	≥47,860	99.81%
	Military Affiliated	≥910	99.89%	≥910	99.89%

Participation Rates by Grade and Subgroup					
Grade	**Group**	**Accountable in ELA**	**Percentage Reportable in ELA**	**Accountable in Mathematics**	**Percentage Reportable in Mathematics**
5	All Students	≥48,280	99.79%	≥48,290	99.79%
	Gender				
	Female	≥23,570	99.82%	≥23,570	99.82%
	Male	≥24,710	99.76%	≥24,710	99.77%
	Ethnicity				
	Hispanic/Latino	≥4,890	99.88%	≥4,890	99.88%
	American Indian or Alaska Native	≥250	100.00%	≥250	100.00%
	Asian	≥790	100.00%	≥790	100.00%
	Black or African American	≥20,400	99.81%	≥20,400	99.82%
	Native Hawaiian or Other Pacific	≥30	100.00%	≥30	100.00%
	White	≥20,200	99.75%	≥20,200	99.74%
	Two or More Races	≥1,680	99.70%	≥1,680	99.70%
	Education Classification				
	Regular	≥42,590	99.81%	≥42,590	99.82%
	Special	≥5,690	99.61%	≥5,690	99.61%
	Economic Status				
	Not Economically Disadvantaged	≥15,030	99.79%	≥15,030	99.79%
	Economically Disadvantaged	≥33,250	99.79%	≥33,250	99.79%
	English Learner Status				
	Not English Learner	≥46,390	99.78%	≥46,400	99.79%
	English Learner	≥1,890	99.95%	≥1,890	99.95%
	Section 504 Status				
	Not Section 504	≥43,540	99.79%	≥43,540	99.80%
	Section 504	≥4,740	99.77%	≥4,740	99.77%
	Migrant Status				
	Not Migrant	≥48,220	99.79%	≥48,230	99.79%
	Migrant	≥60	100.00%	≥60	100.00%
	Homeless Status				
	Not Homeless	≥47,200	99.79%	≥47,200	99.79%
	Homeless	≥1,080	99.72%	≥1,080	99.72%
	Foster Care Status				
	Not in Foster Care	≥48,140	99.79%	≥48,140	99.79%
	In Foster Care	≥140	100.00%	≥140	100.00%
	Military Affiliation				
	Not Military Affiliated	≥47,310	99.79%	≥47,310	99.79%
	Military Affiliated	≥970	99.90%	≥970	99.90%

Participation Rates by Grade and Subgroup					
Grade	**Group**	**Accountable in ELA**	**Percentage Reportable in ELA**	**Accountable in Mathematics**	**Percentage Reportable in Mathematics**
6	All Students	≥43,000	99.61%	≥48,450	99.63%
	Gender				
	Female	≥20,970	99.69%	≥23,640	99.72%
	Male	≥22,030	99.54%	≥24,800	99.55%
	Ethnicity				
	Hispanic/Latino	≥4,400	99.77%	≥4,730	99.79%
	American Indian or Alaska Native	≥200	100.00%	≥250	100.00%
	Asian	≥700	99.86%	≥730	99.86%
	Black or African American	≥18,230	99.56%	≥20,610	99.60%
	Native Hawaiian or Other Pacific	≥30	100.00%	≥40	100.00%
	White	≥17,900	99.64%	≥20,350	99.66%
	Two or More Races	≥1,490	99.33%	≥1,690	99.35%
	Education Classification				
	Regular	≥38,210	99.66%	≥43,000	99.68%
	Special	≥4,780	99.25%	≥5,440	99.30%
	Economic Status				
	Not Economically Disadvantaged	≥13,740	99.75%	≥15,390	99.75%
	Economically Disadvantaged	≥29,250	99.55%	≥33,060	99.58%
	English Learner Status				
	Not English Learner	≥41,450	99.60%	≥46,790	99.63%
	English Learner	≥1,540	99.81%	≥1,650	99.82%
	Section 504 Status				
	Not Section 504	≥38,460	99.59%	≥43,420	99.62%
	Section 504	≥4,530	99.80%	≥5,030	99.78%
	Migrant Status				
	Not Migrant	≥42,930	99.61%	≥48,390	99.63%
	Migrant	≥60	100.00%	≥60	100.00%
	Homeless Status				
	Not Homeless	≥42,010	99.62%	≥47,420	99.65%
	Homeless	≥980	99.09%	≥1,030	99.13%
	Foster Care Status				
	Not in Foster Care	≥42,880	99.61%	≥48,320	99.64%
	In Foster Care	≥110	99.15%	≥130	99.25%
	Military Affiliation				
	Not Military Affiliated	≥42,110	99.61%	≥47,530	99.63%
	Military Affiliated	≥890	99.89%	≥920	99.89%

Participation Rates by Grade and Subgroup					
Grade	**Group**	**Accountable in ELA**	**Percentage Reportable in ELA**	**Accountable in Mathematics**	**Percentage Reportable in Mathematics**
7	All Students	≥41,920	99.46%	≥48,990	99.51%
	Gender				
	Female	≥20,390	99.52%	≥23,730	99.58%
	Male	≥21,530	99.40%	≥25,260	99.45%
	Ethnicity				
	Hispanic/Latino	≥4,480	99.69%	≥4,890	99.71%
	American Indian or Alaska Native	≥210	99.53%	≥270	99.27%
	Asian	≥690	100.00%	≥760	99.74%
	Black or African American	≥17,560	99.28%	≥20,740	99.37%
	Native Hawaiian or Other Pacific	≥20	100.00%	≥30	100.00%
	White	≥17,470	99.55%	≥20,570	99.60%
	Two or More Races	≥1,440	99.51%	≥1,670	99.58%
	Education Classification				
	Regular	≥37,540	99.50%	≥43,790	99.55%
	Special	≥4,380	99.11%	≥5,200	99.21%
	Economic Status				
	Not Economically Disadvantaged	≥13,680	99.76%	≥15,730	99.76%
	Economically Disadvantaged	≥28,230	99.31%	≥33,260	99.39%
	English Learner Status				
	Not English Learner	≥40,350	99.45%	≥47,290	99.51%
	English Learner	≥1,570	99.75%	≥1,700	99.71%
	Section 504 Status				
	Not Section 504	≥37,370	99.47%	≥43,730	99.52%
	Section 504	≥4,550	99.38%	≥5,260	99.47%
	Migrant Status				
	Not Migrant	≥41,860	99.46%	≥48,930	99.51%
	Migrant	≥50	98.28%	≥60	98.51%
	Homeless Status				
	Not Homeless	≥40,980	99.48%	≥47,990	99.53%
	Homeless	≥940	98.40%	≥1,000	98.51%
	Foster Care Status				
	Not in Foster Care	≥41,800	99.47%	≥48,850	99.52%
	In Foster Care	≥110	95.80%	≥140	97.26%
	Military Affiliation				
	Not Military Affiliated	≥41,070	99.45%	≥48,110	99.50%
	Military Affiliated	≥850	100.00%	≥870	100.00%

Participation Rates by Grade and Subgroup					
Grade	**Group**	**Accountable in ELA**	**Percentage Reportable in ELA**	**Accountable in Mathematics**	**Percentage Reportable in Mathematics**
8	All Students	≥45,050	99.35%	≥49,950	99.45%
	Gender				
	Female	≥22,200	99.44%	≥24,660	99.47%
	Male	≥22,850	99.25%	≥25,280	99.42%
	Ethnicity				
	Hispanic/Latino	≥4,430	99.48%	≥4,740	99.56%
	American Indian or Alaska Native	≥210	99.54%	≥280	99.64%
	Asian	≥720	99.72%	≥740	99.73%
	Black or African American	≥19,480	99.18%	≥21,650	99.37%
	Native Hawaiian or Other Pacific	≥40	100.00%	≥40	100.00%
	White	≥18,590	99.47%	≥20,760	99.48%
	Two or More Races	≥1,560	99.36%	≥1,710	99.42%
	Education Classification				
	Regular	≥40,450	99.40%	≥44,760	99.50%
	Special	≥4,600	98.87%	≥5,190	99.02%
	Economic Status				
	Not Economically Disadvantaged	≥14,780	99.61%	≥16,240	99.69%
	Economically Disadvantaged	≥30,270	99.21%	≥33,700	99.33%
	English Learner Status				
	Not English Learner	≥43,570	99.34%	≥48,350	99.45%
	English Learner	≥1,480	99.39%	≥1,600	99.44%
	Section 504 Status				
	Not Section 504	≥40,180	99.36%	≥44,540	99.45%
	Section 504	≥4,870	99.26%	≥5,400	99.41%
	Migrant Status				
	Not Migrant	≥45,000	99.34%	≥49,890	99.44%
	Migrant	≥50	100.00%	≥60	100.00%
	Homeless Status				
	Not Homeless	≥44,120	99.37%	≥48,960	99.46%
	Homeless	≥930	98.40%	≥990	98.48%
	Foster Care Status				
	Not in Foster Care	≥44,910	99.35%	≥49,790	99.45%
	In Foster Care	≥140	97.24%	≥150	97.48%
	Military Affiliation				
	Not Military Affiliated	≥44,190	99.34%	≥49,050	99.44%
	Military Affiliated	≥860	99.65%	≥900	99.67%

**Students in grade 8 who enrolled in Algebra I had the option of taking the Algebra LEAP 2025 HS test instead of the LEAP 2025 Mathematics grade 8 test.*

7.1 Current Administration Data

Tables 7.2 through 7.13 show the percentage of students in each achievement level based on the state population for the 2023 administration of the ELA and mathematics assessments. Results from previous years are presented as well for comparison purposes.

Table 7.2 Comparison of Percentage of Students in Achievement Levels: ELA Grade 3

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥56,800	13.4	17.8	24.7	38.9	5.1
2018	≥55,390	14.2	18.2	22.3	39.8	5.6
2019	≥52,940	13.2	17.2	23.7	39.5	6.4
2021	≥49,630	19.3	19.0	23.1	33.4	5.2
2022	≥49,380	21.9	18.9	21.2	33.6	4.4
2023	≥49,330	21.5	16.3	19.5	36.4	6.2

Table 7.3 Comparison of Percentage of Students in Achievement Levels: ELA Grade 4

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥56,230	8.8	18.3	29.3	36.2	7.3
2018	≥55,760	10.8	17.0	28.7	34.8	8.8
2019	≥54,800	10.3	18.1	26.6	36.1	8.9
2021	≥49,550	13.7	19.1	25.7	32.3	9.3
2022	≥48,980	13.6	17.9	24.5	34.1	10.0
2023	≥48,880	12.3	19.2	24.3	33.9	10.3

Table 7.4 Comparison of Percentage of Students in Achievement Levels: ELA Grade 5

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥53,300	8.7	18.8	31.1	37.9	3.4
2018	≥55,310	8.8	17.7	30.4	39.3	3.7
2019	≥54,910	8.4	21.1	30.0	36.0	4.4
2021	≥49,780	10.7	24.0	28.1	32.7	4.4
2022	≥48,980	10.2	20.0	29.9	36.2	3.6
2023	≥48,310	10.0	20.3	29.3	37.4	3.0

Table 7.5 Comparison of Percentage of Students in Achievement Levels: ELA Grade 6

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥52,370	10.4	24.9	29.8	29.4	5.5
2018	≥52,810	9.3	24.6	31.5	30.3	4.4
2019	≥54,800	9.2	23.5	29.8	32.2	5.3
2021	≥51,430	12.1	26.1	28.3	28.7	4.9
2022	≥49,450	12.1	21.6	28.5	31.3	6.5
2023	≥43,380	10.2	22.4	28.5	31.9	7.1

Table 7.6 Comparison of Percentage of Students in Achievement Levels: ELA Grade 7

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥51,930	13.2	19.2	26.5	30.3	10.8
2018	≥51,540	10.7	19.2	26.8	31.4	11.9
2019	≥52,350	11.6	16.7	25.1	33.0	13.7
2021	≥52,180	13.4	18.3	26.2	29.1	13.0
2022	≥46,360	14.7	16.6	24.0	30.6	14.0
2023	≥42,460	14.8	18.4	24.7	27.3	14.8

Table 7.7 Comparison of Percentage of Students in Achievement Levels: ELA Grade 8

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥50,450	11.4	17.4	27.0	35.1	9.0
2018	≥51,020	10.8	17.4	26.6	36.9	8.4
2019	≥50,720	11.7	16.2	25.4	37.6	9.2
2021	≥51,680	14.3	16.4	25.2	34.9	9.2
2022	≥50,820	12.4	18.2	22.9	35.7	10.8
2023	≥45,790	12.4	16.7	23.5	36.8	10.7

Table 7.8 Comparison of Percentage of Students in Achievement Levels: Mathematics Grade 3

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥56,800	11.1	18.4	27.1	36.2	7.1
2018	≥55,360	10.3	19.7	28.1	34.6	7.3
2019	≥52,820	9.7	20.6	26.4	36.5	6.7
2021	≥49,590	18.2	22.9	25.3	28.3	5.3
2022	≥49,390	13.9	21.8	27.3	32.5	4.5
2023	≥49,380	13.1	22.3	28.4	30.9	5.4

Table 7.9 Comparison of Percentage of Students in Achievement Levels: Mathematics Grade 4

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥56,230	8.2	23.2	29.7	35.0	3.8
2018	≥55,680	8.6	22.8	30.3	34.4	3.9
2019	≥54,690	11.1	20.5	27.1	38.0	3.3
2021	≥49,490	20.0	23.1	25.2	29.7	2.1
2022	≥48,960	14.8	24.6	24.3	32.6	3.7
2023	≥48,880	12.4	22.6	27.9	33.1	4.0

Table 7.10 Comparison of Percentage of Students in Achievement Levels: Mathematics Grade 5

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥53,310	11.1	24.9	32.4	27.7	3.9
2018	≥55,200	10.2	25.8	34.0	25.7	4.2
2019	≥54,730	10.3	26.8	28.3	30.5	4.1
2021	≥49,700	18.5	28.6	26.7	23.2	3.1
2022	≥48,890	13.4	28.0	29.2	24.8	4.6
2023	≥48,270	13.5	25.6	28.4	28.6	4.0

Table 7.11 Comparison of Percentage of Students in Achievement Levels: Mathematics Grade 6

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥52,350	12.6	30.8	29.2	23.7	3.7
2018	≥52,670	11.6	29.0	32.0	24.8	2.6
2019	≥54,710	11.4	26.7	31.7	26.6	3.6
2021	≥51,340	18.8	27.9	28.9	21.9	2.5
2022	≥49,390	18.0	27.4	27.4	23.9	3.3
2023	≥48,350	17.2	25.6	26.1	26.8	4.2

Table 7.12 Comparison of Percentage of Students in Achievement Levels: Mathematics Grade 7

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥51,800	11.2	28.9	35.2	22.6	2.1
2018	≥51,420	9.9	29.0	35.7	22.9	2.4
2019	≥52,090	9.1	29.5	34.7	24.5	2.3
2021	≥52,080	12.0	33.0	32.6	20.5	1.9
2022	≥51,100	13.4	29.7	32.8	21.3	2.7
2023	≥48,920	10.1	31.6	31.1	23.8	3.5

Table 7.13 Comparison of Percentage of Students in Achievement Levels: Mathematics Grade 8

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥44,710	20.3	28.2	25.0	24.7	1.8
2018	≥44,910	20.9	27.4	23.7	26.1	1.9
2019	≥44,520	20.9	25.7	25.4	25.7	2.3
2021	≥45,840	27.3	25.8	25.2	20.2	1.5
2022	≥44,990	23.5	27.7	25.2	21.5	2.1
2023	≥44,250	23.2	29.9	23.3	21.5	2.3

Score reports are the primary means of communicating test scores to appropriate school system personnel (e.g., testing coordinators or superintendents), teachers, and parents. Standard 6.10 of the *Standards* states:

> When test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the test covers, what scores represent, the precision/reliability of the scores, and how scores are intended to be used (119).

Standard 5.1 is related to Standard 6.10. It states:

> Test users should be provided with clear explanations of the characteristics, meaning, and intended interpretation of scale scores, as well as their limitations (102).

Interpretations of test scores are disseminated in two ways: the individual score report and the *LEAP 2025 Interpretive Guide* (2023).

In addition to providing interpretation of the test results, the LODE and DRC must ensure that the information is understandable for the target audience. Standard 7.0 states:

> Information relating to tests should be clearly documented so that those who use tests can make informed decisions regarding which test to use for a specific purpose, how to administer the chosen test, and how to interpret test scores (125).

The LDOE and DRC strive to create documents that will be accessible to parents, teachers, and all other stakeholders.

The Individual Student-Level Report (ISR) is the primary means for sharing student test results with parents. As such, it is a stand-alone document from which parents can glean information that is relevant to understanding their children's test scores. For more information about the test, parents are provided A Parent Guide to the LEAP 2025 Student Reports. In the 2023 administration year, student reports for each school were posted by grade, then downloaded and printed from DRC INSIGHT Portal by school systems and schools. DRC INSIGHT Portal is DRC's secure online system that provides schools and districts access to student tests and reports.

7.1.1 Description of Each Type of Report

In this section, descriptions of the School Roster Report and the ISR are provided.

In compliance with AERA, APA, & NCME (2014) Standard 12.18, the LEAP 2025 score reports provide clear information about the results of individual students and of specific groups of students. Standard 12.18 states:

> In educational settings, score reports should be accompanied by a clear presentation of information on how to interpret the scores, including the degree of measurement error associated with each score or classification level, and by supplementary information related to group summary scores. In addition, dates of test administration and relevant norming studies should be included in score reports (200).

School Roster Report

A School Roster Report, which provides summary information about student performance on the LEAP 2025 ELA and Mathematics tests, is available to school systems and schools through DRC INSIGHT Portal. Total test scores and achievement-level indicators are shown for the content area of interest. Reporting category and subcategory performance ratings are also reported for students. At the school level, the percentage of students at each achievement level and rating by category and subcategory are summarized. More details can be found in the *LEAP 2025 Interpretive Guide*.

Individual Student-Level Report

The ISR is another type of report available through the DRC INSIGHT Portal system. ISRs may be downloaded and printed by schools to be sent home to parents. At the top of the page, overall student performance is reported by scale scores and achievement level. To give context to the student score, the student's school system and state averages are presented to the right of the student information. In the middle of the page, category and subcategory performance indicators are reported. achievement-level descriptors and the percentage of students in each achievement level by school, school system, and the state, which allows comparisons of the student's overall achievement level to those of their peers, are found at the bottom of the page. When a student does not receive a scale score, their achievement level will be left blank. ISRs for students whose scores were invalidated will display a blank scale score for a given content area.

A data file referred to as Louisiana Department of Education Student File (LDESTD) was provided to the LDOE by DRC. It contains one record for every student tested; each record contains demographic information, responses for multiple-choice (MC) items, scores for items that are not MC items, raw scores, content and process standard raw scores, scale scores, and performance-level data for each content area.

The *LEAP 2025 Interpretive Guide* was written to help Louisiana school system and school administrators, teachers, parents, and the general public to better understand the LEAP 2025 ELA and mathematics tests. The *LEAP 2025 Interpretive Guide* was developed collaboratively by DRC and LDOE staff. LDOE staff had opportunities to review the guide, provide feedback, and give final approval.

The *LEAP 2025 Interpretive Guide* has three sections. The first section presents an introduction and an overview of key terms and test-related concepts. The second section discusses assessment terms and types of scores that are presented on the ISRs. Sample ISRs are included in the guide. The third section discusses information that is presented on the School Roster Report and an example of the report.

In summary, the overall purpose of reporting test results is to communicate information on student performance to stakeholders. These results are presented in the context of score reports that aid the user in understanding the meaning of the test scores. The reports and ancillary information developed by DRC address multiple best practices of the testing industry but are particularly related to the following standards:

> ***Standard 5.1*** Test users should be provided with clear explanations of the characteristics, meaning, and intended interpretation of scale scores, as well as their limitations (102).
>
> ***Standard 6.10*** When test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the test covers, what scores represent, the precision/reliability of the scores, and how scores are intended to be used (119).
>
> ***Standard 7.0*** Information relating to tests should be clearly documented so that those who use tests can make informed decisions regarding which test to use for a specific purpose, how to administer the chosen test, and how to interpret test scores (125).
>
> ***Standard 12.18*** In educational settings, score reports should be accompanied by a clear presentation of information on how to interpret the scores, including the degree of measurement error associated with each score or classification level, and by supplementary information related to group summary scores. In addition, dates of test administration and relevant norming studies should be included in score reports (200).

Chapter 8: Performance-Level Setting

This chapter briefly describes the LEAP 2025 performance-level setting and presents the cut scores and achievement-level descriptors derived from the performance-level setting. Since the LDOE uses PARCC cut scores for the LEAP 2025 ELA and mathematics tests, a brief overview of the PARCC performance-level setting procedures is included in this chapter. A more detailed discussion and the results of the PARCC performance-level setting may be found in the *Performance Level Setting Technical Report* (Pearson, 2015).

The AERA, APA, & NCME (2014) Standards addressed by the *Performance Level Setting Technical Report* (Pearson, 2015) are 5.21 and 5.22.

Starting in the spring of 2015, the ELA and mathematics assessments measured different content and constructs than did previous tests were administered to Louisiana students. The new tests were built using the PARCC item bank and were fully aligned to the Louisiana Student Standards. The new tests were reported on new scales, and students were classified by achievement levels based on their knowledge and ability to perform different tasks in relation to the new test content and standards.

In terms of the validity of the LEAP 2025 test scores, it is essential to understand that descriptors and cut scores are established in a collaborative and participatory process. The descriptors clearly establish, in plain language, the proper frame of reference for understanding how to interpret test scores, particularly cut scores.

8.1 PARCC Performance-Level Setting Process for English Language Arts and Mathematics

According to the *Performance Level Setting Technical Report* (Pearson, 2015), PARCC used the evidence-based standard setting (EBSS) method (Beimers, Way, McClarty, & Miles, 2012) for the PARCC performance-level setting (PLS) process. The EBSS method is used to combine various considerations into the process for setting performance levels, including policy considerations, content standards, research, and educator judgment about what students should know and be able to demonstrate, and to support PARCC's policy goals related to college- and career-readiness expectations. Additional details about the EBSS method can be found in the *Performance Level Setting Technical Report* (Pearson, 2015).

8.2 Cut Scores

This section presents the cut scores for each grade and content area of the LEAP 2025. Tables 8.1 and 8.2 show the ELA and mathematics cut scores for students in grades 3 through 8.

Table 8.1 English Language Arts Cut Scores

Grade	Cut Scores			
	Approaching Basic	***Basic***	***Mastery***	***Advanced***
3	700	725	750	810
4	700	725	750	790
5	700	725	750	799
6	700	725	750	790
7	700	725	750	785
8	700	725	750	794

Table 8.2 Mathematics Cut Scores

Grade	Cut Scores			
	Approaching Basic	***Basic***	***Mastery***	***Advanced***
3	700	725	750	790
4	700	725	750	796
5	700	725	750	790
6	700	725	750	788
7	700	725	750	786
8	700	725	750	801

8.2.1 Reporting Category Cut Scores

As stated in Section 6.4.2.3, student performance on ELA and mathematics reporting categories and subcategories was classified into one of three performance ratings: *Strong*, *Moderate*, and *Weak*. Detailed rules for calculating performance ratings for ELA and mathematics reporting categories and subcategories can be found in that section.

The cut scores divide the continuum of student achievement into the following five achievement levels used by the LDOE for reporting purposes:

- *Advanced*: Students performing at this level have **exceeded** college- and career-readiness expectations and are well prepared for the next level of study in this content area.
- *Mastery*: Students performing at this level have **met** college- and career-readiness expectations and are prepared for the next level of study in this content area.
- *Basic*: Students performing at this level have **nearly met** college- and career-readiness expectations and may need additional support to be fully prepared for the next level of study in this content area.
- *Approaching Basic*: Students performing at this level have **partially met** college- and career-readiness expectations and will need much support to be prepared for the next level of study in this content area.
- *Unsatisfactory*: Students performing at this level have **not yet met** the college- and career-readiness expectations and will need extensive support to be prepared for the next level of study in this content area.

Table 8.3 summarizes the LEAP 2025 ELA and mathematics scale score ranges for each level of achievement.

Table 8.3 Achievement-Level Scale Score Ranges

ELA						
Achievement Level	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Advanced	810–850	790–850	799–850	790–850	785–850	794–850
Mastery	750–809	750–789	750–798	750–789	750–784	750–793
Basic	725–749					
Approaching Basic	700–724					
Unsatisfactory	650–699					
MATHEMATICS						
Achievement Level	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Advanced	790–850	796–850	790–850	788–850	786–850	801–850
Mastery	750–789	750–795	750–789	750–787	750–785	750–800
Basic	725–749					
Approaching Basic	700–724					
Unsatisfactory	650–699					

This chapter presented a brief overview of PARCC's performance-level setting process, which set the cut scores used by the LDOE for reporting student performance on the LEAP 2025 ELA and mathematics tests. These procedures are addressed in more detail in relevant technical reports.

The performance-level setting process undertaken by PARCC addresses the following standards:

> ***Standard 5.21*** When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly (107).
>
> ***Standard 5.22*** When cut scores defining pass-fail or proficiency levels are based on direct judgments about the adequacy of item or test performances, the judgmental process should be designed so that the participants providing the judgments can bring their knowledge and experience to bear in a reasonable way (108).

Chapter 9: Evidence of Validity

Evidence for validity—the meaning of test scores and the inferences they support—is the central concept underlying the LEAP 2025 validation process. Validity evidence, from the design of the test to item development and scoring, is created throughout the entire assessment process. Therefore, evidence of validity is described throughout the LEAP 2025 technical report. Table 9.1 summarizes the sources of evidence of validity and indicates where the evidence can be found in the technical report.

Table 9.1 Summary of Evidence of Validity and the Report Chapter in Which it is Found

Source of Validity	Related Information	Related Chapter/Source
Evidence Based on Test Content	Item Development Process	Chapter 3 2022–2023 LEAP Grades 3-8 ELA and Mathematics Assessment Frameworks
	Test Blueprint and Item Alignment to Curriculum and Standards	Chapter 3 2022–2023 LEAP Grades 3-8 ELA and Mathematics Assessment Frameworks
	Item Bias, Sensitivity, and Content Appropriateness	Chapter 3
	Accommodations	Chapters 3 and 4
Evidence Based on Response Processes	Testing Time	Chapter 4
	Evaluation of the criteria used by hand scorers	Spring 2023 LEAP 2025 Handscoring Specifications Chapter 5
	Features Scored by Artificial Intelligence (AI) Engines	2023 LEAP 2025 Handscoring Specifications Chapter 5
Evidence Based on Internal Structure	Inter-rater Agreement	Chapter 5
	Reliability and Standard Errors of Measurement	Chapter 9
	Decision Accuracy	Chapter 9
	Dimensionality	Chapter 9
	Differential Item Functioning	Chapter 10
	Student Group Reliability	Chapter 10
Evidence Based on Relationships to Other Variables	Divergent Validity	Chapter 9
	Regression of LEAP 2025 from 2022 to 2023	Chapter 9
Evidence Based on the Consequences of Testing	Scale Score and Performance Level Information	Chapter 7
	Test Interpretive Guide	Chapter 4

In this chapter, DRC presents evidence of construct-related validity through studies of test reliability, convergent validity, and divergent validity. All analyses in this chapter are based on census data, after removing data from the test takers who were administered the Spanish language and Braille versions of the test forms.

Chapter 9 of this report demonstrates adherence to the American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME, 2014) Standards 1.13, 1.21, 2.0, 2.3, 2.13, 2.14, 2.16, and 2.19. Each standard is discussed in the pertinent section of this chapter.

9.1 Construct-Irrelevant Variance and Construct Underrepresentation

Minimization of construct-irrelevant variance and construct underrepresentation is addressed in the following steps of the test development process: (1) specification, (2) item writing, (3) review, (4) field testing, (5) test construction, and (6) item calibration (see Chapter 3 for more information on steps 1–5 and Chapter 6 for more information on step 6).

Construct-irrelevant variance refers to error variance that is caused by factors unrelated to the constructs measured by the test. For example, when tests are not administered under standardized conditions (e.g., one administration may be timed, but another administration is untimed), differences in student performance related to different administration conditions may result. Careful specification of content and review of the items representing that content are first steps in minimizing construct-irrelevant variance. Then, empirical evidence, especially item-level data, is used to infer construct irrelevance.

Construct underrepresentation occurs when the content of the assessment does not reflect the full range of content that the assessment is expected to cover. Specification and review, a process through which test blueprints are developed and reviewed, are primary steps in the development process designed to ensure that content is appropriately represented.

9.2 Reliability

Reliability refers to the consistency of students' test scores on parallel forms of a test. A reliable test is one that produces scores that are expected to be relatively stable if the test is administered repeatedly under similar conditions. Often, however, it is impractical to administer multiple forms of the test, and reliability is estimated on a single administration of the test. This type of reliability, known as internal consistency, provides an estimate of how consistently examinees perform across items within a test during a single test administration (Crocker & Algina, 1986). Reliability is a necessary, but not sufficient, condition of validity.

The 2014 *Standards* indicates the following:

> The term *reliability* has been used in two ways in the measurement literature. First, the term has been used to refer to the reliability coefficients of classical test theory, defined as the correlation between scores on two equivalent forms of the test, presuming that taking one form has no effect on performance on the second form. Second, the term has been used in a more general sense, to refer to the consistency of scores across replications of a testing procedure, regardless of how this consistency is estimated or reported (e.g., in terms of standard errors, reliability coefficients per se, generalizability coefficients, error/tolerance ratios, item response theory (IRT) information functions, or various indices of classification consistency) (33).

In accordance with the *Standards* in developing and maintaining tests of the highest quality, DRC has calculated the reliability of each LEAP 2025 test in a variety of ways: reliability of raw scores, overall standard

error of measurement (SEM), IRT-based conditional SEM, and decision consistency of achievement-level classifications.

There are several specific standards that this chapter addresses. These include Standards 2.0, 2.3, 2.13, and 2.19, each of which is articulated below.

> ***Standard 2.0*** Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use (42).

> ***Standard 2.3*** For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported (43).

The total score reliabilities are discussed in Section 9.2.1 of this chapter. The SEMs and subscore reliabilities are presented in Sections 9.4.2 and 9.4.3. The SEM of the total score is discussed in Section 9.2.2.

> ***Standard 2.13*** The standard error of measurement, both overall and conditional (if reported), should be provided in units of each reported score (45).

The SEM based on raw scores is discussed in Section 9.2.2 and is reported in raw score units. The conditional SEM is discussed in Section 9.2.3 and is presented in scale score units.

> ***Standard 2.19*** Each method of quantifying the reliability/precision of scores should be described clearly and expressed in terms of statistics appropriate to the method. The sampling procedures used to select test takers for reliability/precision analyses and the descriptive statistics on these samples, subject to privacy obligations where applicable, should be reported (47).

Section 9.2 discusses different ways of measuring test reliability, including reliability of raw scores and test-form SEM, IRT-based conditional SEM, and decision consistency of achievement-level classifications. These statistics were computed based on the census data.

9.2.1 Test Reliability

The reliability of raw scores by test form was evaluated using Cronbach's (1951) coefficient alpha, which is a lower-bound estimate of test reliability. The reliability coefficient is a ratio of the variance of true test scores to the variance of the total observed scores, with the values ranging from 0 to 1. The closer the value of the reliability coefficient is to 1, the more consistent the scores, where 1 refers to a perfectly consistent test. In general, reliability coefficients that are equal to or greater than 0.8 are considered acceptable for tests of moderate lengths.

Cronbach's coefficient alpha was computed using the formula

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n}\sigma_i^2}{\sigma_X^2}\right], \qquad (9.1)$$

where n is the number of items on the test, σ_i^2 is the variance of item i, and σ_x^2 is the variance of the total test score.

Total test reliability measures, such as Cronbach's coefficient alpha and SEM, consider the consistency (i.e., reliability) of performance over all test questions in a given form, the results of which imply how well the questions measure the content domain and could continue to do so over repeated administrations. The number of items in the test influences these statistics; for example, a longer test can be expected to be more reliable than a shorter test.

The reliability coefficients for the LEAP 2025 are reported in Table 9.2. These reliability coefficients were computed using the census data. The reliability statistics ranged from 0.88 to 0.93 for all ELA forms. The ELA forms have one writing component (RI or RL) that is the same score of another component (WE); the item score for the RI/RL component was excluded from the reliability computation. For mathematics, the reliabilities ranged from 0.92 to 0.94. These results indicate acceptable reliability coefficients for the LEAP 2025 tests.

Table 9.2 Reliability in English Language Arts and Mathematics

Content	Grade	Mode	Number of Items	Number of Score Points	SEM	Cronbach's Alpha	N-Count
ELA	3	CBT	24	52	4.37	0.89	≥22,120
ELA	3	PBT	24	52	4.46	0.88	≥27,200
ELA	4	CBT	30	65	5.01	0.91	≥48,880
ELA	5	CBT	28	62	5.23	0.90	≥48,310
ELA	6	CBT	32	70	4.89	0.92	≥43,370
ELA	7	CBT	32	70	5.61	0.93	≥42,460
ELA	8	CBT	32	70	5.64	0.91	≥45,780
Mathematics	3	CBT	43	62	3.63	0.92	≥22,100
Mathematics	3	PBT	43	62	3.83	0.92	≥27,170
Mathematics	4	CBT	43	62	3.83	0.94	≥48,790
Mathematics	5	CBT	43	62	3.63	0.93	≥48,170
Mathematics	6	CBT	43	66	3.78	0.94	≥48,240
Mathematics	7	CBT	43	66	3.61	0.94	≥48,790
Mathematics	8	CBT	41	65	3.43	0.92	≥44,120

The reliability statistics by subgroup are reported and discussed in Chapter 10.

9.2.2 Standard Error of Measurement

The reliability of reported test scores can be characterized by the standard errors associated with the scores. The SEM may be used to determine the range within which a student's true score is likely to fall. An observed score should be regarded not as a student's true score but as an estimate of a student's true score. It is expected that the score a student obtains from a single test administration would fall within one SEM of the student's true score 68% of the time and within approximately two SEMs of the true score 95% of the time. The SEM is an index of the random variability in test scores and is defined as follows:

$$\mathrm{SEM} = SD\sqrt{1 - R_{xx'}}\ , \tag{9.2}$$

where SD represents standard deviation of the raw score distribution, and $R_{xx'}$ is estimated by $\hat{\alpha}$ as expressed in Equation 9.1.

The SEM at the test-form level was computed in raw score metric and is also presented in Table 9.2 for ELA and mathematics.

9.2.3 Conditional Standard Error of Measurement

In contrast to SEM, conditional standard error of measurement (CSEM) expresses the degree of measurement error in scale score units and is conditioned on the ability of the student. DRC reports the CSEM in support of Standard 2.14, which states:

> When possible and appropriate, conditional standard errors of measurement should be reported at several score levels unless there is evidence that the standard error is constant across score levels. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score (46).

In further compliance with Standard 2.14, the CSEM of each cut score is reported in Table 9.3.

The CSEMs are defined as the reciprocal of the square root of the test information function and can be estimated across all points of the ability continuum (Hambleton & Swaminathan, 1985). The CSEM is defined in the following equation:

$$\text{CSEM}(\theta_i) = \frac{1}{\sqrt{I(\theta_i)}}, \tag{9.3}$$

where $I(\vartheta_i)$ is the test information function, as a sum of item information function 2, obtained as

$$I(\theta_i) = \sum_j \frac{p'_{ij}(\theta_i)^2}{p_{ij}(\theta_i) q_{ij}(\theta_i)}, \tag{9.4}$$

where $p'_{ij}(\theta_i)$ is the derivative of $p_{ij}(\theta_i)$ and $q_{ij}(\theta_i) = 1 - p_{ij}(\theta_i)$.

Note that the CSEMs vary in magnitude across the entire range of student ability estimates (i.e., scale scores) and are smaller in the middle of the score distribution and higher at the tails. This pattern is expected when IRT methods are used. Since LEAP 2025 was first administered, every effort has been made to make the TCC and CSEM values at the cut scores between the PARCC assessments and the LEAP 2025 assessments similar. Both TCC and CSEM values have been similar across the LEAP 2025 alternate forms given the same content because similar or the same statistical properties are important for alternate forms. To provide context regarding the magnitude of the CSEMs, it is important to also refer to sections 9.2.1 Test Reliability and 9.2.4 Classification Accuracy and Consistency where evidence is provided of high measures of form reliability and levels of accurate student classification at the cutpoints to support the use of the LEAP 2025 assessments. The CSEMs at the four cut scores that define the performance levels are presented in Table 9.3.

Table 9.3 Conditional Standard Errors of Measurement at the *Approaching Basic*, *Basic*, *Mastery*, and *Advanced* Cut Scores

			Approaching Basic		*Basic*		*Mastery*		*Advanced*	
Content Area	**Grade**	**Mode**	**Cut Score**	**CSEM**	**Cut Score**	**CSEM**	**Cut Score**	**CSEM**	**Cut Score**	**CSEM**
ELA	3	CBT	700	13	725	11	750	11	810	12
ELA	3	PBT	700	13	725	11	750	11	810	12
ELA	4	CBT	700	11	725	8	750	8	790	9
ELA	5	CBT	700	9	725	7	750	7	799	8
ELA	6	CBT	700	9	725	7	750	7	790	8
ELA	7	CBT	700	9	725	7	750	7	785	8
ELA	8	CBT	700	9	725	8	750	8	794	9
Mathematics	3	CBT	700	9	725	7	750	7	790	10
Mathematics	3	PBT	700	9	725	7	750	7	790	10
Mathematics	4	CBT	700	9	725	7	750	6	796	11
Mathematics	5	CBT	700	8	725	6	750	6	790	9
Mathematics	6	CBT	700	8	725	6	750	6	788	8
Mathematics	7	CBT	700	9	725	6	750	5	786	7
Mathematics	8	CBT	700	11	725	9	750	7	801	10

Figures 9.1 and 9.2 display the CSEM (conditional standard error of measurement) curves for each grade and content area by mode. Typically, with fixed-form assessments, the estimates of measurement error tend to be higher at the low and high ends of the scale-score range where few items measure those ability levels. Generally, there are few students with extreme scores, and these score levels cannot be estimated as accurately as levels toward the middle of the ability range. The middle ability range, where cut scores are located, shows lower measurement error than the low and high ends of the ability ranges. Figures 9.1 and 9.2 demonstrate that the tests are designed so that measurement error is minimized in the middle of the scale range, where most students are located.

Figure 9.1 CSEM Curves for ELA Grades 3 through 8

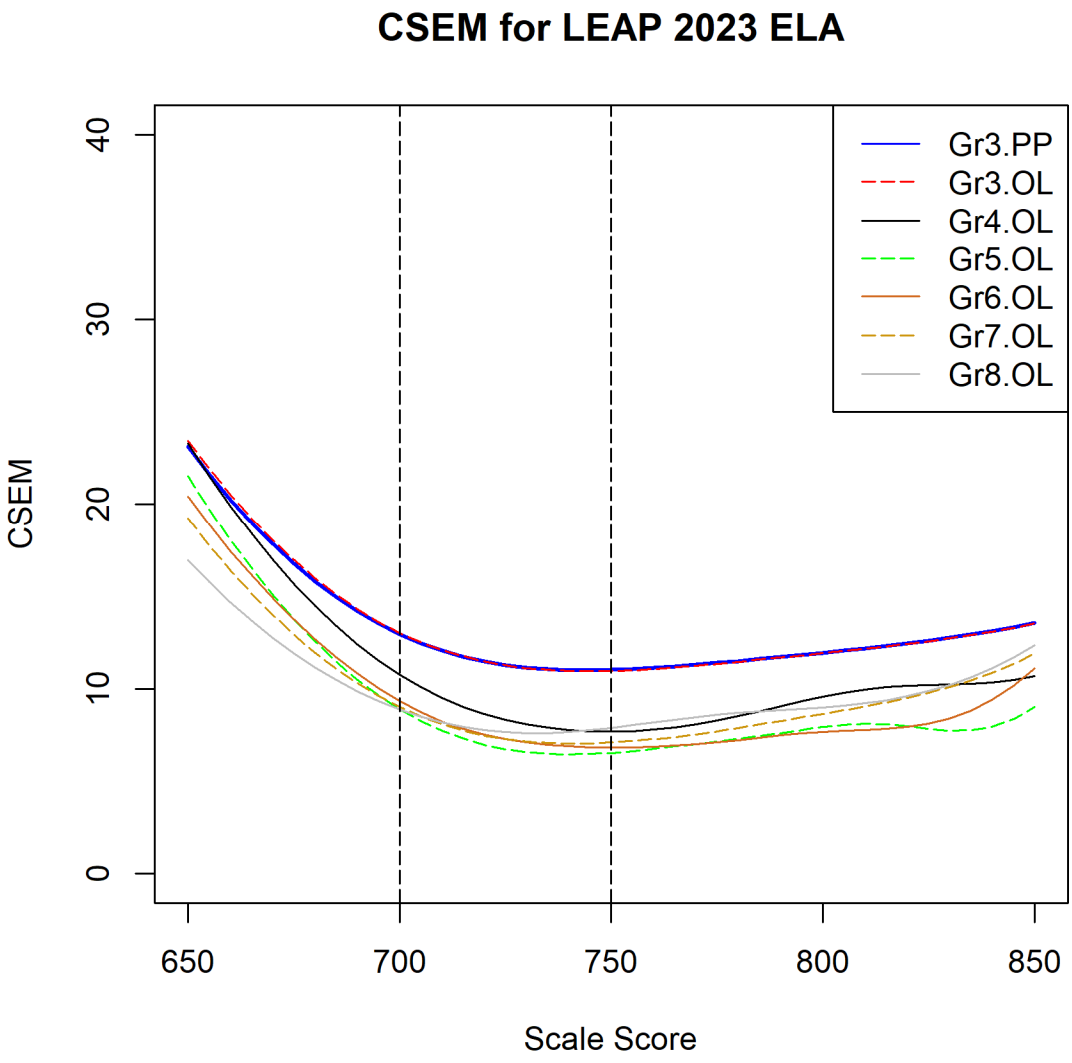
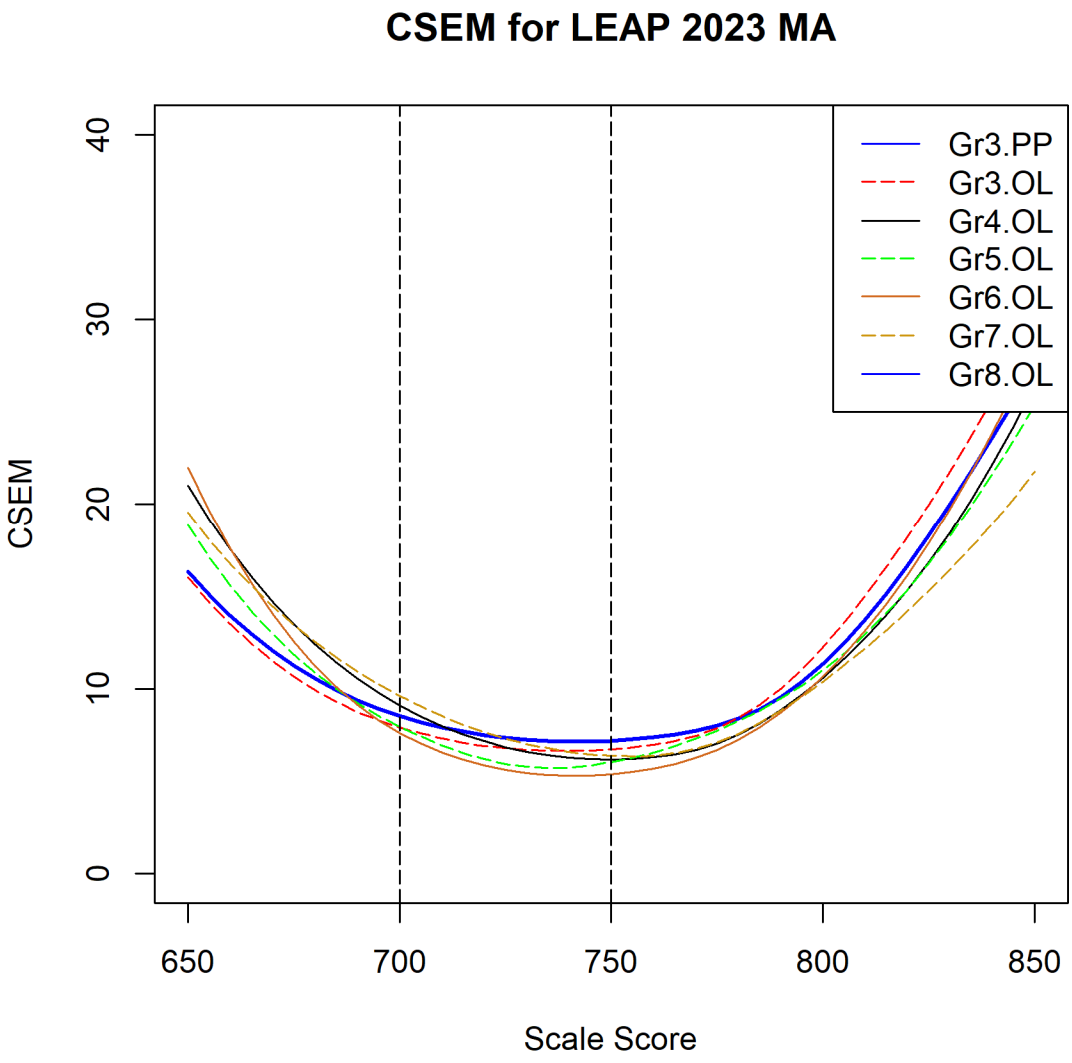


Figure 9.2 CSEM Curves for Mathematics Grades 3 through 8



9.2.4 Classification Accuracy and Consistency

Classification Accuracy

Classification accuracy is defined as the extent to which the actual classifications of test takers into various achievement levels match classifications made based on their true scores (Livingston & Lewis, 1995). Classification accuracy refers to the agreement between the observed score and the true score, whereas classification consistency refers to the agreement between two observed scores.

Classification Consistency

Classification consistency is defined as the extent to which the classifications of students in a particular achievement level match based on two independent administrations of the same test form or one administration of two parallel test forms. It is often logistically infeasible, as well as expensive, to obtain data from repeated administrations of a test, be it re-administration of the same test or administration of a parallel form. Therefore, a common practice is to estimate classification consistency from one administration of a test.

The Livingston-Lewis (1995) methodology was used to calculate classification accuracy statistics based on the spring LEAP 2025 results. The Livingston-Lewis procedure utilizes a beta-binomial model that requires two steps: (1) fitting proportion-correct true scores to a four-parameter beta distribution and (2) using the binomial distribution to estimate classification accuracy and consistency. All calculations for classification accuracy and consistency are based on census data.

Classification consistency and classification accuracy conditioned on achievement level (see Table 9.4 and Table 9.5) and on cut score (see Table 9.6 and Table 9.7) are presented for the 2023 LEAP 2025 in this section of the report. The magnitude of classification consistency and accuracy measures is influenced by several key features of the test design, including the number of items, the location and number of cut scores, the score distribution, and the reliability and associated SEM. As can be seen in Table 9.4, classification accuracy conditioned on achievement level ranged from 0.51 to 0.87 for ELA and 0.48 to 0.88 for mathematics. Classification consistency (see Table 9.5) conditioned on achievement level ranged from 0.34 to 0.78 for ELA and 0.48 to 0.81 for mathematics. Table 9.6 shows that classification accuracy at achievement cut points ranged from 0.90 to 0.97 for ELA and 0.87 to 0.99 for mathematics. Classification consistency (see Table 9.7) conditioned at achievement cut points ranged from 0.86 to 0.96 for ELA and 0.82 to 0.99 for mathematics. Classification consistency and accuracy at achievement cut points tend to be higher values than those conditioned on achievement level. For some tests, classification accuracy and consistency conditioned on the *Advanced* level were lower than 0.50. One reason for these relatively low *Advanced* level values is few highly difficult items to distinguish the *Advanced* level from other achievement levels.

Table 9.4 Classification Accuracy Conditioned on Level of Achievement

	Classification Accuracy						
Content Area	**Grade**	**Mode**	**Unsatisfactory**	**Approaching Basic**	**Basic**	**Mastery**	**Advanced**
ELA	3	CBT	0.80	0.56	0.58	0.79	0.58
ELA	3	PBT	0.74	0.56	0.60	0.83	0.51
ELA	4	CBT	0.67	0.70	0.67	0.80	0.70
ELA	5	CBT	0.69	0.71	0.72	0.87	0.29
ELA	6	CBT	0.63	0.73	0.72	0.84	0.70
ELA	7	CBT	0.77	0.70	0.71	0.77	0.80
ELA	8	CBT	0.74	0.68	0.68	0.81	0.73
Mathematics	3	CBT	0.79	0.73	0.74	0.83	0.63
Mathematics	3	PBT	0.76	0.71	0.73	0.84	0.66
Mathematics	4	CBT	0.68	0.70	0.75	0.88	0.51
Mathematics	5	CBT	0.74	0.73	0.79	0.85	0.48
Mathematics	6	CBT	0.78	0.72	0.79	0.84	0.68
Mathematics	7	CBT	0.42	0.74	0.78	0.86	0.71
Mathematics	8	CBT	0.74	0.63	0.71	0.83	0.74

Table 9.5 Classification Consistency Conditioned on Level of Achievement

	Classification Consistency						
Content Area	**Grade**	**Mode**	**Unsatisfactory**	**Approaching Basic**	**Basic**	**Mastery**	**Advanced**
ELA	3	CBT	0.76	0.42	0.43	0.70	0.57
ELA	3	PBT	0.71	0.43	0.44	0.74	0.51
ELA	4	CBT	0.65	0.53	0.57	0.70	0.63
ELA	5	CBT	0.66	0.58	0.61	0.78	0.34
ELA	6	CBT	0.62	0.58	0.63	0.74	0.67
ELA	7	CBT	0.73	0.56	0.59	0.66	0.75
ELA	8	CBT	0.72	0.55	0.55	0.71	0.67
Mathematics	3	CBT	0.72	0.61	0.64	0.77	0.58
Mathematics	3	PBT	0.70	0.58	0.62	0.77	0.60
Mathematics	4	CBT	0.61	0.56	0.66	0.81	0.52
Mathematics	5	CBT	0.68	0.62	0.68	0.78	0.48
Mathematics	6	CBT	0.72	0.60	0.66	0.79	0.62
Mathematics	7	CBT	0.41	0.61	0.70	0.81	0.66
Mathematics	8	CBT	0.65	0.50	0.57	0.79	0.69

Perhaps the most important indices for accountability systems are those for the accuracy and consistency of classification decisions made at specific cut points. To evaluate decisions at specific cut points, the joint distribution of all the achievement levels is collapsed into a dichotomized distribution around that specific cut point. As an example, for the LEAP 2025 assessments, a dichotomization at the cut point between the *Basic* and *Mastery* classifications was formed. The proportion of correct classifications below this particular cut

point is equal to the sum of all the cells at the *Unsatisfactory*, *Approaching Basic*, and *Basic* levels, and the proportion of correct classifications above that particular cut point is equal to the sum of all the cells at the *Mastery* and *Advanced* levels. Table 9.6 shows the classification accuracy and Table 9.7 shows the consistency estimates when conditioned on LEAP 2025 cut scores. The classification accuracy statistics are at or above 0.87, while the classification consistency statistics are at or above 0.82. These results suggest that consistent and accurate achievement-level classifications are being made for students in Louisiana based on the LEAP 2025.

Table 9.6 Classification Accuracy at Achievement Cut Points

			Classification Accuracy			
Content Area	**Grade**	**Mode**	***Unsatisfactory/ Approaching Basic***	***Approaching Basic/ Basic***	***Basic/ Mastery***	***Mastery/ Advanced***
ELA	3	CBT	0.91	0.91	0.91	0.97
ELA	3	PBT	0.94	0.92	0.90	0.95
ELA	4	CBT	0.94	0.92	0.92	0.95
ELA	5	CBT	0.95	0.92	0.91	0.97
ELA	6	CBT	0.94	0.92	0.92	0.97
ELA	7	CBT	0.95	0.93	0.92	0.95
ELA	8	CBT	0.95	0.93	0.91	0.95
Mathematics	3	CBT	0.94	0.92	0.93	0.98
Mathematics	3	PBT	0.95	0.93	0.92	0.96
Mathematics	4	CBT	0.93	0.93	0.94	0.97
Mathematics	5	CBT	0.94	0.93	0.94	0.97
Mathematics	6	CBT	0.93	0.93	0.94	0.98
Mathematics	7	CBT	0.90	0.92	0.95	0.98
Mathematics	8	CBT	0.87	0.91	0.95	0.99

Table 9.7 Classification Consistency at Achievement Cut Points

	Classification Consistency					
Content Area	Grade	Mode	Unsatisfactory/ Approaching Basic	Approaching Basic/ Basic	Basic/ Mastery	Mastery/ Advanced
ELA	3	CBT	0.87	0.87	0.88	0.96
ELA	3	PBT	0.91	0.88	0.86	0.93
ELA	4	CBT	0.91	0.89	0.88	0.92
ELA	5	CBT	0.93	0.89	0.87	0.96
ELA	6	CBT	0.92	0.89	0.89	0.95
ELA	7	CBT	0.92	0.90	0.89	0.93
ELA	8	CBT	0.94	0.90	0.87	0.93
Mathematics	3	CBT	0.91	0.89	0.90	0.97
Mathematics	3	PBT	0.93	0.89	0.89	0.95
Mathematics	4	CBT	0.90	0.90	0.91	0.96
Mathematics	5	CBT	0.91	0.90	0.91	0.96
Mathematics	6	CBT	0.90	0.90	0.92	0.97
Mathematics	7	CBT	0.87	0.88	0.93	0.98
Mathematics	8	CBT	0.82	0.87	0.92	0.99

9.2.5 Convergent Validity

Convergent validity is a subtype of construct validity that can be estimated by the extent to which measures of constructs that theoretically should be related to each other are, in fact, observed as related to each other. Analyses of the internal structure of a test can indicate the extent to which the relationships among test items conform to the construct the test purports to measure. For example, the LEAP 2025 mathematics test is designed to measure a single overall construct—mathematics achievement; therefore, the items comprising the LEAP 2025 mathematics test should measure only mathematics, not language or reading.

This technical report summarizes additional statistics that contribute to construct validity (Cronbach's coefficient alpha is reported previously in this section, and item fit is reported in Chapter 6). The internal consistency coefficient (i.e., Cronbach's alpha) reported is typically measured via correlations among the test items and indicates of the degree of the same general construct (Pearson, 2015, page 128). Table 9.2 shows test reliability statistics for ELA and mathematics. The reliability statistics ranged from 0.88 to 0.93 for ELA forms and from 0.92 to 0.94 for mathematics forms, indicating that items on the 2023 LEAP 2025 assessments are homogenous. For a group of items to be homogeneous, the items must measure the same construct (i.e., construct validity) or represent the same content domain (i.e., content validity). Because IRT models were used to calibrate test items and to report student scores, item fit is also relevant to construct validity. The extent to which test items function as the IRT model prescribes is relevant to the validation of test scores. As shown in Chapter 6, few items were flagged for poor model/data fit.

9.3 Principal Components Analysis

As another measure of construct validity, DRC examined the unidimensionality of each grade-level LEAP 2025 test. One of the underlying assumptions of the IRT models used to scale the LEAP 2025 tests is that the tests being calibrated are unidimensional; that is, items in each grade and content area measure a single content domain. For example, mathematics items should measure mathematics ability and not reading skills. Standard 1.13 of the *Standards* states:

> If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided (26–27).

This section examines the internal structure of the LEAP 2025 tests by evaluating the unidimensionality assumption through principal components analysis (PCA), which is one of the frequently used methods to do so (Chou and Wang, 2010). This analysis seeks evidence that there exists a single primary factor, the first principal component, which accounts for much of the relationship between items. The presence of a single or dominant factor suggests that a test is sufficiently unidimensional (i.e., that it measures one underlying construct).

A PCA was conducted for each grade, content area, and mode of the LEAP 2025 assessments. A large first principal component is evident in each analysis. It is common to have additional eigenvalues greater than 1.0, which may suggest the presence of other factors. For all grades, content areas, and modes of the LEAP 2025 assessments, the ratio of variance accounted for by the first factor to variance accounted for by the second is sufficiently large to indicate that the unidimensionality assumption holds. All the LEAP 2025 content-area tests exhibit first principal components accounting for more than 20% of the test variance for ELA (see Table 9.8) and for mathematics (see Table 9.9). Reckase (1979) proposed that the first component should account for at least 20% of the variance to claim unidimensionality.

To further investigate the unidimensionality of the ELA and mathematics assessments, the ratio of the first eigenvalue to the second eigenvalue was determined (see Tables 9.8 and 9.9). When the first eigenvalue is sufficiently larger than the second eigenvalue, that is considered evidence of unidimensionality (Lord, 1980; Lumsden, 1957, 1961). These ratios show that the first eigenvalue is at least four times as large as the second eigenvalue for all the grades, content areas, and modes. This substantial difference in magnitude indicates that one factor appears to be dominant and that the ELA and mathematics tests are essentially unidimensional.

This evidence supports the claim that there is a dominant dimension underlying the items and tasks in each test and that scores from each test represent performance primarily determined by that ability. Construct-irrelevant variance, such as factual knowledge irrelevant to doing well in a subject, does not appear to create significant nuisance factors.

Table 9.8 Principal Component Analysis for English Language Arts

Grade	Mode	Components	Eigenvalue	Percentage of Variance Explained	Cumulative Percentage of Variance Explained
3	CBT	First Component	7.03	29.30	29.30
3	CBT	Second Component	1.26	5.25	34.55
3	CBT	Ratio (First/Second)	5.59		
3	PBT	First Component	6.71	27.96	27.96
3	PBT	Second Component	1.20	5.02	32.97
3	PBT	Ratio (First/Second)	5.57		
4	CBT	First Component	8.47	28.22	28.22
4	CBT	Second Component	1.27	4.23	32.45
4	CBT	Ratio (First/Second)	6.68		
5	CBT	First Component	8.00	28.57	28.57
5	CBT	Second Component	1.36	4.85	33.42
5	CBT	Ratio (First/Second)	5.88		
6	CBT	First Component	9.46	29.55	29.55
6	CBT	Second Component	1.41	4.41	33.96
6	CBT	Ratio (First/Second)	6.70		
7	CBT	First Component	10.04	31.38	31.38
7	CBT	Second Component	1.46	4.57	35.95
7	CBT	Ratio (First/Second)	6.87		
8	CBT	First Component	9.02	28.20	28.20
8	CBT	Second Component	1.47	4.58	32.78
8	CBT	Ratio (First/Second)	6.16		

Table 9.9 Principal Component Analysis for Mathematics

Grade	Mode	Components	Eigenvalue	Percentage of Variance Explained	Cumulative Percentage of Variance Explained
3	CBT	First Component	11.55	26.86	26.86
3	CBT	Second Component	1.68	3.90	30.76
3	CBT	Ratio (First/Second)	6.88		
3	PBT	First Component	11.70	27.20	27.20
3	PBT	Second Component	1.66	3.87	31.07
3	PBT	Ratio (First/Second)	7.04		
4	CBT	First Component	13.49	31.37	31.37
4	CBT	Second Component	1.47	3.42	34.80
4	CBT	Ratio (First/Second)	9.16		
5	CBT	First Component	13.02	30.28	30.28
5	CBT	Second Component	1.56	3.62	33.90
5	CBT	Ratio (First/Second)	8.37		
6	CBT	First Component	13.60	31.63	31.63
6	CBT	Second Component	1.43	3.32	34.95
6	CBT	Ratio (First/Second)	9.53		
7	CBT	First Component	13.98	32.52	32.52
7	CBT	Second Component	1.34	3.12	35.64
7	CBT	Ratio (First/Second)	10.43		
8	CBT	First Component	11.06	26.97	26.97
8	CBT	Second Component	1.44	3.50	30.48
8	CBT	Ratio (First/Second)	7.70		

9.4 Analyses by Reporting Categories and Subcategories

Three sets of analyses were conducted at the reporting category and subcategory levels for ELA and mathematics in another attempt to assess the construct validity of the LEAP 2025 assessments. First, correlation coefficients that measure the relationship between the reporting category scores and subcategory scores in both subjects were computed. Second, the reliability of each reporting category and subcategory was computed. Finally, the SEM was computed for each reportable category and subcategory.

9.4.1 Correlations among Reporting Categories and Subcategories

This section reports the strength of the interrelationships among the categories or subcategories by computing the correlation between them. Tables 9.10–9.13 report the uncorrected Pearson product-moment (PPM) correlation coefficients, the PPM corrected for attenuation (CAPPM), and the reliability coefficients described above. The PPM among the categories and subcategories is presented below the diagonal portion of the matrix, the CAPPM is presented above the diagonal portion of the matrix, and the reliability coefficients used are shown in Tables 9.10–9.13.

The uncorrected PPM in Tables 9.10–9.13 should be interpreted in the context of the reliability coefficient. In general, lower PPM coefficients are expected between variables that are less reliable. In most cases, the PPM

coefficients show that performance on one category or subcategory is moderately to strongly related to performance on another category or subcategory within the same grade and content area. The value of the correlation coefficients will be affected by the limited number of items measuring each category or subcategory. Therefore, caution should be used when comparing the PPM coefficients that measure the relationships between categories or subcategories to those that measure the relationships between content areas. A more modest relationship (i.e., smaller correlation coefficients) is expected to be reported between the categories and subcategories as a consequence of the lower number of items measuring each of the reporting categories. The PPM between two category or subcategory scores may be artificially low because of measurement error.

Standard 1.21 states:

> When statistical adjustments, such as those for restriction of range or attenuation, are made, both adjusted and unadjusted coefficients, as well as the specific procedure used, and all statistics used in the adjustment, should be reported. Estimates of the construct-criterion relationship that remove the effects of measurement error on the test should be clearly reported as adjusted estimates (29).

The attenuation of the PPM can be corrected statistically using Spearman's formula:

$$CAPPM = \frac{r_{xy}}{\sqrt{r_{xx} r_{yy}}}, \tag{9.5}$$

where r_{xy} is the PPM between two claims or GLE strands, r_{xx} is the reliability of one of those claims or GLE strands, and r_{yy} is the reliability of the other claim or GLE strand.

ELA shows moderate relationships between the reading and writing reporting categories across all grades, indicating that these two categories measure some different traits. Across all tables, the CAPPM indicates moderate or strong relationships between subcategories. The CAPPM for reading vocabulary, written expression, and knowledge and use of language are moderate. In some cases, the CAPPM is greater than 1.0. "Disattenuated values greater than 1.00 indicate that measurement error is not randomly distributed" (Schumacker, 1996). The moderate or strong relationships suggested by the CAPPM in Tables 9.10–9.13 are further evidence of the validity of the test construct. Since the overall content area is comprised of the category or subcategories subscores and the content area is expected to be moderately or highly related.

Table 9.10 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Reporting Category: English Language Arts

Uncorrected and Corrected Correlation Coefficients: English Language Arts Reporting Category						
Grade	**Mode**	**No.**	**Category**	**N Items**	**1**	**2**
3	CBT	1	Reading	20	.	0.85
	CBT	2	Writing	4	0.73	.
	PBT	1	Reading	20	.	0.87
	PBT	2	Writing	4	0.71	
4	CBT	1	Reading	26	.	0.82
	CBT	2	Writing	4	0.73	.
5	CBT	1	Reading	24	.	0.84
	CBT	2	Writing	4	0.75	.
6	CBT	1	Reading	28	.	0.79
	CBT	2	Writing	4	0.71	.
7	CBT	1	Reading	28	.	0.83
	CBT	2	Writing	4	0.77	.
8	CBT	1	Reading	28	.	0.83
	CBT	2	Writing	4	0.75	.

Table 9.11 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Reporting Subcategories: English Language Arts

Uncorrected and Corrected Correlation Coefficients: English Language Arts Reporting Subcategory									
Grade	**Mode**	**No.**	**Subcategory**	**N Items**	**1**	**2**	**3**	**4**	**5**
3	CBT	1	Reading Literary Text	8	.	1.05	1.01	0.98	0.89
	CBT	2	Reading Information Text	7	0.70	.	1.00	0.99	0.91
	CBT	3	Reading Vocabulary	5	0.69	0.66	.	0.80	0.77
	CBT	4	Written Expression	4	0.67	0.65	0.54	.	1.14
	CBT	5	Knowledge & Use of Language	2	0.63	0.62	0.54	0.80	.
	PBT	1	Reading Literary Text	8	.	1.03	1.01	1.07	0.93
	PBT	2	Reading Information Text	7	0.68	.	0.99	1.04	0.92
	PBT	3	Reading Vocabulary	5	0.69	0.64	.	0.88	0.82
	PBT	4	Written Expression	4	0.66	0.61	0.53	.	1.30
	PBT	5	Knowledge & Use of Language	2	0.60	0.57	0.52	0.75	.
4	CBT	1	Reading Literary Text	10	.	1.01	0.99	0.84	0.76
	CBT	2	Reading Information Text	9	0.72	.	1.03	0.97	0.86
	CBT	3	Reading Vocabulary	7	0.70	0.72	.	0.84	0.76
	CBT	4	Written Expression	4	0.62	0.70	0.61	.	1.07
	CBT	5	Knowledge & Use of Language	2	0.59	0.65	0.58	0.84	.
5	CBT	1	Reading Literary Text	8	.	1.01	1.01	0.96	0.88
	CBT	2	Reading Information Text	10	0.71	.	0.99	0.87	0.81
	CBT	3	Reading Vocabulary	6	0.67	0.67	.	0.79	0.77
	CBT	4	Written Expression	4	0.71	0.66	0.56	.	1.07
	CBT	5	Knowledge & Use of Language	2	0.69	0.64	0.57	0.88	.
6	CBT	1	Reading Literary Text	8	.	0.96	0.97	0.78	0.76
	CBT	2	Reading Information Text	12	0.73	.	1.01	0.94	0.90
	CBT	3	Reading Vocabulary	8	0.73	0.74	.	0.79	0.76
	CBT	4	Written Expression	4	0.61	0.71	0.59	.	1.17
	CBT	5	Knowledge & Use of Language	2	0.61	0.70	0.58	0.92	.
7	CBT	1	Reading Literary Text	8	.	1.02	1.01	0.91	0.87
	CBT	2	Reading Information Text	13	0.81	.	1.00	0.87	0.84
	CBT	3	Reading Vocabulary	7	0.70	0.74	.	0.80	0.77
	CBT	4	Written Expression	4	0.73	0.75	0.60	.	1.08
	CBT	5	Knowledge & Use of Language	2	0.70	0.73	0.59	0.95	.
8	CBT	1	Reading Literary Text	7	.	1.05	1.04	0.97	0.95
	CBT	2	Reading Information Text	12	0.74	.	1.01	0.87	0.85
	CBT	3	Reading Vocabulary	9	0.70	0.75	.	0.77	0.78
	CBT	4	Written Expression	4	0.72	0.71	0.60	.	1.09
	CBT	5	Knowledge & Use of Language	2	0.71	0.70	0.62	0.94	.

Table 9.12 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Reporting Categories: Mathematics

Uncorrected and Corrected Correlation Coefficients: Mathematics Reporting Category								
Grade	**Mode**	**No.**	**Category**	**N Items**	**1**	**2**	**3**	**4**
3	CBT	1	Major Content	27	.	0.99	0.95	1.00
	CBT	2	Additional & Supporting Con	10	0.77	.	0.91	1.03
	CBT	3	Expressing Mathematical Rea	3	0.70	0.58	.	1.01
	CBT	4	Modeling & Application	3	0.78	0.69	0.64	.
	PBT	1	Major Content	27	.	0.99	0.99	1.01
	PBT	2	Additional & Supporting Con	10	0.78	.	0.93	1.05
	PBT	3	Expressing Mathematical Rea	3	0.71	0.59	.	1.01
	PBT	4	Modeling & Application	3	0.79	0.71	0.63	.
4	CBT	1	Major Content	29	.	0.99	1.00	0.99
	CBT	2	Additional & Supporting Con	8	0.81	.	1.00	1.00
	CBT	3	Expressing Mathematical Rea	3	0.79	0.71	.	1.07
	CBT	4	Modeling & Application	3	0.79	0.72	0.75	.
5	CBT	1	Major Content	28	.	0.98	0.99	0.99
	CBT	2	Additional & Supporting Con	9	0.79	.	0.95	0.99
	CBT	3	Expressing Mathematical Rea	3	0.77	0.66	.	1.03
	CBT	4	Modeling & Application	3	0.78	0.70	0.70	.
6	CBT	1	Major Content	26	.	0.99	1.01	0.94
	CBT	2	Additional & Supporting Con	10	0.81	.	0.99	0.95
	CBT	3	Expressing Mathematical Rea	4	0.84	0.74	.	1.02
	CBT	4	Modeling & Application	3	0.75	0.68	0.74	.
7	CBT	1	Major Content	28	.	1.00	0.96	1.02
	CBT	2	Additional & Supporting Con	8	0.80	.	0.98	1.05
	CBT	3	Expressing Mathematical Rea	4	0.80	0.72	.	1.04
	CBT	4	Modeling & Application	3	0.82	0.74	0.77	.
8	CBT	1	Major Content	26	.	1.00	0.95	0.93
	CBT	2	Additional & Supporting Con	8	0.80	.	0.99	0.99
	CBT	3	Expressing Mathematical Rea	4	0.76	0.73	.	1.02
	CBT	4	Modeling & Application	3	0.64	0.63	0.64	.

Table 9.13 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Reporting Subcategories: Mathematics

Uncorrected and Corrected Correlation Coefficients: Mathematics Reporting Subcategory								
Grade	**Mode**	**No.**	**Subcategory**	**N Items**	**1**	**2**	**3**	**4**
3	CBT	1	A1	9	.	0.89	0.89	0.97
	CBT	2	A2	4	0.65	.	0.92	0.96
	CBT	3	A3	8	0.66	0.62	.	0.95
	CBT	4	A4	6	0.67	0.60	0.60	.
	PBT	1	A1	9	.	0.90	0.90	0.97
	PBT	2	A2	4	0.66	.	0.94	0.96
	PBT	3	A3	8	0.67	0.65	.	0.96
	PBT	4	A4	6	0.67	0.62	0.63	.
4	CBT	1	A1	7	.	0.89	0.98	.
	CBT	2	A2	7	0.63	.	0.86	.
	CBT	3	A3	8	0.73	0.64	.	.
5	CBT	1	A1	7	.	0.93	0.98	0.94
	CBT	2	A2	7	0.69	.	0.99	0.92
	CBT	3	A3	7	0.74	0.71	.	0.90
	CBT	4	A4	6	0.65	0.60	0.59	.
6	CBT	1	A1	9	.	0.97	1.00	.
	CBT	2	A2	8	0.77	.	0.98	.
	CBT	3	A3	9	0.76	0.77	.	.
7	CBT	1	A1	10	.	1.00	0.99	.
	CBT	2	A2	12	0.81	.	1.00	.
	CBT	3	A3	6	0.74	0.76	.	.
8	CBT	1	A1	5	.	0.84	0.87	0.86
	CBT	2	A2	9	0.50	.	1.00	0.95
	CBT	3	A3	6	0.49	0.66	.	0.97
	CBT	4	A4	6	0.50	0.65	0.63	.

9.4.2 Reliability of Reporting Categories and Subcategories

Raw score summary statistics (i.e., mean and standard deviation), Cronbach's (1951) coefficient alpha, and SEM were computed for each of the reporting categories or subcategories by grade, content area, and mode using the census data. These statistics are presented in Tables 9.14–9.17 for ELA and mathematics. Reliability indices, such as Cronbach's coefficient alpha (and resulting SEM), are a function of the number of items on a test, the average covariance between item-pairs, and the variance of a test's total score. In general, it is expected that the coefficient alpha would be lower for a reporting category or subcategory assessed by a small number of items than for one assessed by a larger number of items.

9.4.3 Standard Error of Measurement of Reporting Categories and Subcategories

This chapter also reports the SEM associated with each of the reporting categories and subcategories in Tables 9.14–9.17 for ELA and mathematics. In these tables the RI/RL writing component was included. These SEMs are reported in the raw score metric.

Table 9.14 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of English Language Arts Reporting Categories

Grade	Mode	Category	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
3	CBT	Reading	22	46	17.22	9.30	3.55	0.85
	CBT	Writing	4	24	4.44	4.58	1.75	0.85
	PBT	Reading	22	46	20.08	9.24	3.56	0.85
	PBT	Writing	4	24	6.30	4.61	2.12	0.79
4	CBT	Reading	27	56	24.89	11.60	4.04	0.88
	CBT	Writing	4	27	7.09	5.85	1.92	0.89
5	CBT	Reading	26	56	25.54	11.22	4.17	0.86
	CBT	Writing	4	30	8.27	6.14	1.74	0.92
6	CBT	Reading	29	60	27.36	12.29	3.94	0.90
	CBT	Writing	4	30	6.77	6.11	1.81	0.91
7	CBT	Reading	30	64	29.17	14.19	4.39	0.90
	CBT	Writing	4	30	11.01	7.77	1.75	0.95
8	CBT	Reading	30	64	31.66	12.95	4.43	0.88
	CBT	Writing	4	30	11.75	7.28	1.76	0.94

Table 9.15 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of English Language Arts Reporting Subcategories

Mean, Standard Deviation, and SEM: English Language Arts								
Grade	**Mode**	**Subcategory**	**Number of Items**	**Number of Score Pts.**	**Mean Raw Score**	**Raw Score Std. Dev.**	**SEM**	**Cronbach's Alpha**
3	CBT	Reading Literary Text	9	19	6.72	4.10	2.31	0.68
	CBT	Reading Information Text	8	17	5.45	3.42	2.05	0.64
	CBT	Reading Vocabulary	5	10	5.04	2.93	1.66	0.68
	CBT	Written Expression	2	18	3.16	3.49	1.97	0.68
	CBT	Knowledge & Use of Language	2	6	1.28	1.28	0.66	0.73
	PBT	Reading Literary Text	9	19	8.24	4.12	2.30	0.69
	PBT	Reading Information Text	8	17	6.11	3.46	2.08	0.64
	PBT	Reading Vocabulary	5	10	5.73	2.85	1.65	0.67
	PBT	Written Expression	2	18	4.56	3.64	2.46	0.54
	PBT	Knowledge & Use of Language	2	6	1.73	1.20	0.76	0.60
4	CBT	Reading Literary Text	10	20	9.53	4.49	2.34	0.73
	CBT	Reading Information Text	10	22	8.57	4.76	2.58	0.71
	CBT	Reading Vocabulary	7	14	6.79	3.62	2.01	0.69
	CBT	Written Expression	2	21	5.14	4.51	2.28	0.75
	CBT	Knowledge & Use of Language	2	6	1.95	1.53	0.64	0.82
5	CBT	Reading Literary Text	9	20	8.81	4.43	2.45	0.69
	CBT	Reading Information Text	11	24	9.44	5.10	2.70	0.72
	CBT	Reading Vocabulary	6	12	7.29	3.05	1.86	0.63
	CBT	Written Expression	2	24	5.89	4.62	2.11	0.79
	CBT	Knowledge & Use of Language	2	6	2.39	1.67	0.61	0.87
6	CBT	Reading Literary Text	8	16	8.14	4.13	1.90	0.79
	CBT	Reading Information Text	13	28	11.39	5.35	2.68	0.75
	CBT	Reading Vocabulary	8	16	7.84	4.04	2.14	0.72
	CBT	Written Expression	2	24	5.06	4.66	2.26	0.76
	CBT	Knowledge & Use of Language	2	6	1.71	1.54	0.67	0.81
7	CBT	Reading Literary Text	9	20	8.59	4.80	2.43	0.74
	CBT	Reading Information Text	14	30	13.10	7.10	2.88	0.84
	CBT	Reading Vocabulary	7	14	7.48	3.53	2.09	0.65
	CBT	Written Expression	2	24	8.20	5.95	2.10	0.88
	CBT	Knowledge & Use of Language	2	6	2.81	1.90	0.65	0.88
8	CBT	Reading Literary Text	8	18	8.29	3.78	2.24	0.65
	CBT	Reading Information Text	13	28	13.09	6.12	2.93	0.77
	CBT	Reading Vocabulary	9	18	10.29	4.33	2.31	0.71
	CBT	Written Expression	2	24	8.77	5.62	2.15	0.85
	CBT	Knowledge & Use of Language	2	6	2.98	1.74	0.63	0.87

Table 9.16 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Mathematics Reporting Categories

Mean, Standard Deviation, and SEM: Mathematics								
Grade	**Mode**	**Category**	**Number of Items**	**Number of Score Points**	**Mean Raw Score**	**Raw Score Std. Dev.**	**SEM**	**Cronbach's Alpha**
3	CBT	Major Content	27	30	16.50	7.10	2.30	0.90
	CBT	Additional & Supporting Content	10	10	5.98	2.36	1.35	0.67
	CBT	Expressing Mathematical Reasoning	3	10	2.51	2.17	1.37	0.60
	CBT	Modeling & Application	3	12	3.27	2.98	1.71	0.67
	PBT	Major Content	27	30	17.50	7.11	2.28	0.90
	PBT	Additional & Supporting Content	10	10	6.25	2.33	1.32	0.68
	PBT	Expressing Mathematical Reasoning	3	10	3.61	2.41	1.56	0.58
	PBT	Modeling & Application	3	12	4.23	3.34	1.90	0.68
4	CBT	Major Content	29	30	15.99	7.66	2.28	0.91
	CBT	Additional & Supporting Content	8	10	4.90	2.71	1.40	0.73
	CBT	Expressing Mathematical Reasoning	3	10	3.21	2.70	1.51	0.69
	CBT	Modeling & Application	3	12	3.40	3.38	1.82	0.71
5	CBT	Major Content	28	30	15.23	7.78	2.32	0.91
	CBT	Additional & Supporting Content	9	10	4.97	2.37	1.27	0.71
	CBT	Expressing Mathematical Reasoning	3	10	2.68	2.15	1.24	0.66
	CBT	Modeling & Application	3	12	3.21	3.22	1.80	0.69
6	CBT	Major Content	26	30	13.92	7.58	2.25	0.91
	CBT	Additional & Supporting Content	10	10	4.36	2.60	1.33	0.74
	CBT	Expressing Mathematical Reasoning	4	14	4.38	3.55	1.74	0.76
	CBT	Modeling & Application	3	12	2.84	2.95	1.64	0.69
7	CBT	Major Content	28	30	12.04	7.59	2.23	0.91
	CBT	Additional & Supporting Content	8	10	3.92	2.52	1.39	0.70
	CBT	Expressing Mathematical Reasoning	4	14	2.92	3.20	1.53	0.77
	CBT	Modeling & Application	3	12	3.32	2.86	1.53	0.71
8	CBT	Major Content	26	29	11.82	6.19	2.27	0.87
	CBT	Additional & Supporting Content	8	10	3.18	2.53	1.28	0.75
	CBT	Expressing Mathematical Reasoning	4	14	2.37	2.92	1.52	0.73
	CBT	Modeling & Application	3	12	1.86	2.03	1.38	0.54

Table 9.17 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Mathematics Reporting Subcategories

Mean, Standard Deviation, and SEM: Mathematics								
Grade	Mode	Major Content Subcategory	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
3	CBT	A1	9	9	5.51	2.67	1.18	0.81
	CBT	A2	4	7	3.44	1.92	1.11	0.67
	CBT	A3	8	8	4.29	2.01	1.13	0.68
	CBT	A4	6	6	3.27	1.69	1.08	0.60
	PBT	A1	9	9	5.81	2.55	1.18	0.79
	PBT	A2	4	7	3.63	1.96	1.11	0.68
	PBT	A3	8	8	4.56	2.06	1.12	0.71
	PBT	A4	6	6	3.50	1.70	1.07	0.60
4	CBT	A1	7	8	4.47	2.23	1.18	0.72
	CBT	A2	7	7	3.18	2.02	1.10	0.70
	CBT	A3	8	8	4.47	2.42	1.14	0.78
5	CBT	A1	7	7	3.25	2.25	1.04	0.78
	CBT	A2	7	8	3.82	2.32	1.26	0.71
	CBT	A3	7	7	3.46	2.06	1.07	0.73
	CBT	A4	6	6	3.91	1.59	1.01	0.60
6	CBT	A1	9	10	4.50	2.68	1.28	0.77
	CBT	A2	8	10	4.53	2.88	1.24	0.81
	CBT	A3	9	10	4.89	2.69	1.35	0.75
7	CBT	A1	10	11	4.57	2.90	1.32	0.79
	CBT	A2	12	13	5.31	3.53	1.51	0.82
	CBT	A3	6	6	2.17	1.77	0.97	0.70
8	CBT	A1	5	6	2.92	1.52	1.08	0.49
	CBT	A2	9	9	3.37	2.28	1.26	0.70
	CBT	A3	6	7	2.49	1.78	1.07	0.64
	CBT	A4	6	7	3.03	1.88	1.08	0.67

9.5 Divergent (Discriminant) Validity

Measures of different constructs should not be highly correlated with each other. Divergent validity is a subtype of construct validity that can be assessed by the extent to which measures of constructs that theoretically should not be related to each other are, in fact, observed as not related to each other. Typically, correlation coefficients among measures of unrelated or distantly related constructs are examined in support of divergent validity.

To assess the divergent validity of the LEAP 2025 assessments, correlations were computed between the ELA, mathematics, social studies, and science scale scores for students who took more than one LEAP 2025 content-area test in 2023. These correlations are based on the census data, and the results are shown in Table 9.18. The correlation coefficients ranged from 0.72 (between mathematics and social studies in grades 5 and 8) to 0.85 (between ELA and social studies in grade 6). The correlation coefficients suggest that individual student scores across subjects are moderately related, indicating that these tests measure a similar knowledge base or general underlying ability while still measuring some different traits as planned.

Table 9.18 Inter-Correlation of English Language Arts and Mathematics Scale Scores

Grade	ELA/ Mathematics	ELA/ Social Studies	ELA/ Science	Mathematics/ Social Studies	Mathematics/ Science	Social Studies/ Science
3	0.76	0.78	0.76	0.73	0.74	0.76
4	0.76	0.79	0.81	0.73	0.77	0.80
5	0.75	0.78	0.79	0.72	0.78	0.80
6	0.81	0.85	0.81	0.80	0.79	0.82
7	0.78	0.81	0.81	0.77	0.81	0.82
8	0.73	0.83	0.79	0.72	0.75	0.83

9.6 Regression of LEAP 2025 from 2022 to 2023

The LEAP 2025 assessments were designed to support an integrated educational system where the scope and sequence of each grade's curriculum will support student readiness for and achievement in the next education level. Effective measurement is expected to result in assessments that produce scores that consistently measure each grade's content and produce data that provide strong evidence of preparedness for the content measured by assessments at the education level.

This study required the collection of data from adjacent grades for each content area. For this purpose, matched longitudinal LEAP 2025 test data from spring 2022 and spring 2023 were used. For example, grade 3 students were matched with grade 4 students, and only matched students were used to estimate correlation and perform linear regression from 2022 to 2023.

Table 9.19 summarizes the correlation and regression results for 2022 and 2023 LEAP 2025. For ELA, the correlation ranged from 0.79 to 0.85, and for mathematics, the correlation ranged from 0.82 to 0.87. Correlations for mathematics were slightly higher than those for ELA. Correlations for both content areas can be considered moderate, which can often be found in state assessments. R^2 indicates how much of the 2022 performance can explain the 2023 performance. For example, 0.63 for ELA 2022 grade 3 and 2023 grade 4 means that 2022's grade 3 performance can explain (predict) about 70% of 2023's grade 4 performance. This R^2 value is generally the power of 2 for the matching correlation. The R^2 values for ELA range from 0.63 to 0.72, and those for mathematics range from 0.67 to 0.76. These also show the moderate relationships between adjacent grades for both ELA and mathematics.

Table 9.19 Correlation and Regression Summary for 2022 and 2023 LEAP 2025

Content	2022 Grade	2023 Grade	N	Correlation	R^2
ELA	3	4	≥45,480	0.79	0.63
	4	5	≥45,470	0.80	0.64
	5	6	≥40,440	0.83	0.70
	6	7	≥39,220	0.84	0.71
	7	8	≥39,980	0.85	0.72
Mathematics	3	4	≥45,390	0.84	0.70
	4	5	≥45,310	0.85	0.72
	5	3	≥45,060	0.84	0.71
	6	7	≥45,210	0.87	0.76
	7	8	≥40,770	0.82	0.67

Figures 9.3 and 9.4 show regression line and scatter plots for ELA and mathematics. The linear lines in the plots are linear regression lines from 2022 to 2023. In general, the length of band given the linear regression line shows the strength of correlation. If the band is narrow, the correlation is high, and if the band is large, the correlation is low. Every plot shows some moderate linear relationships between 2022 and 2023 adjacent grades for both ELA and mathematics.

Figure 9.3 Regression Line and Scatter Plots:

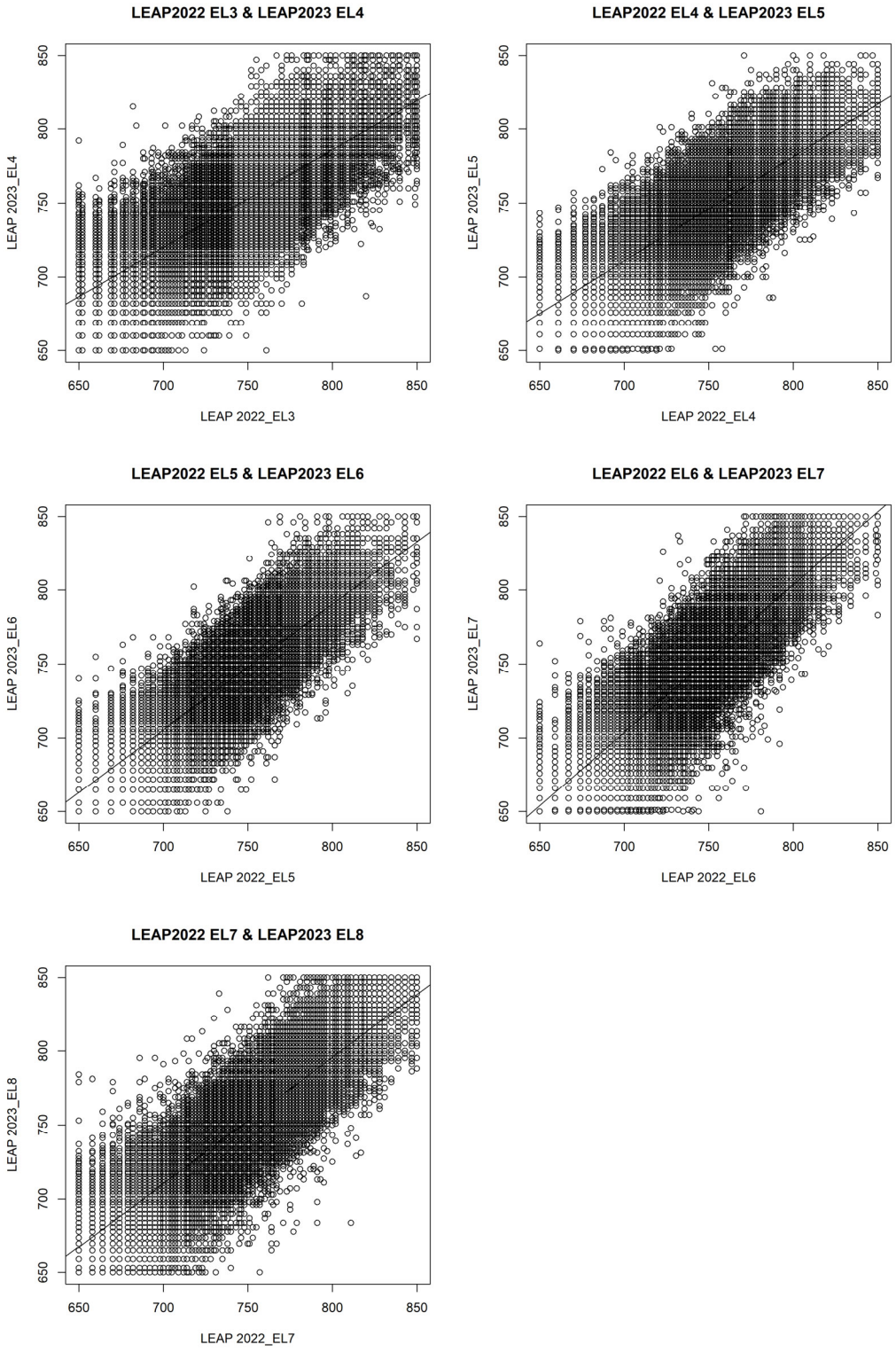
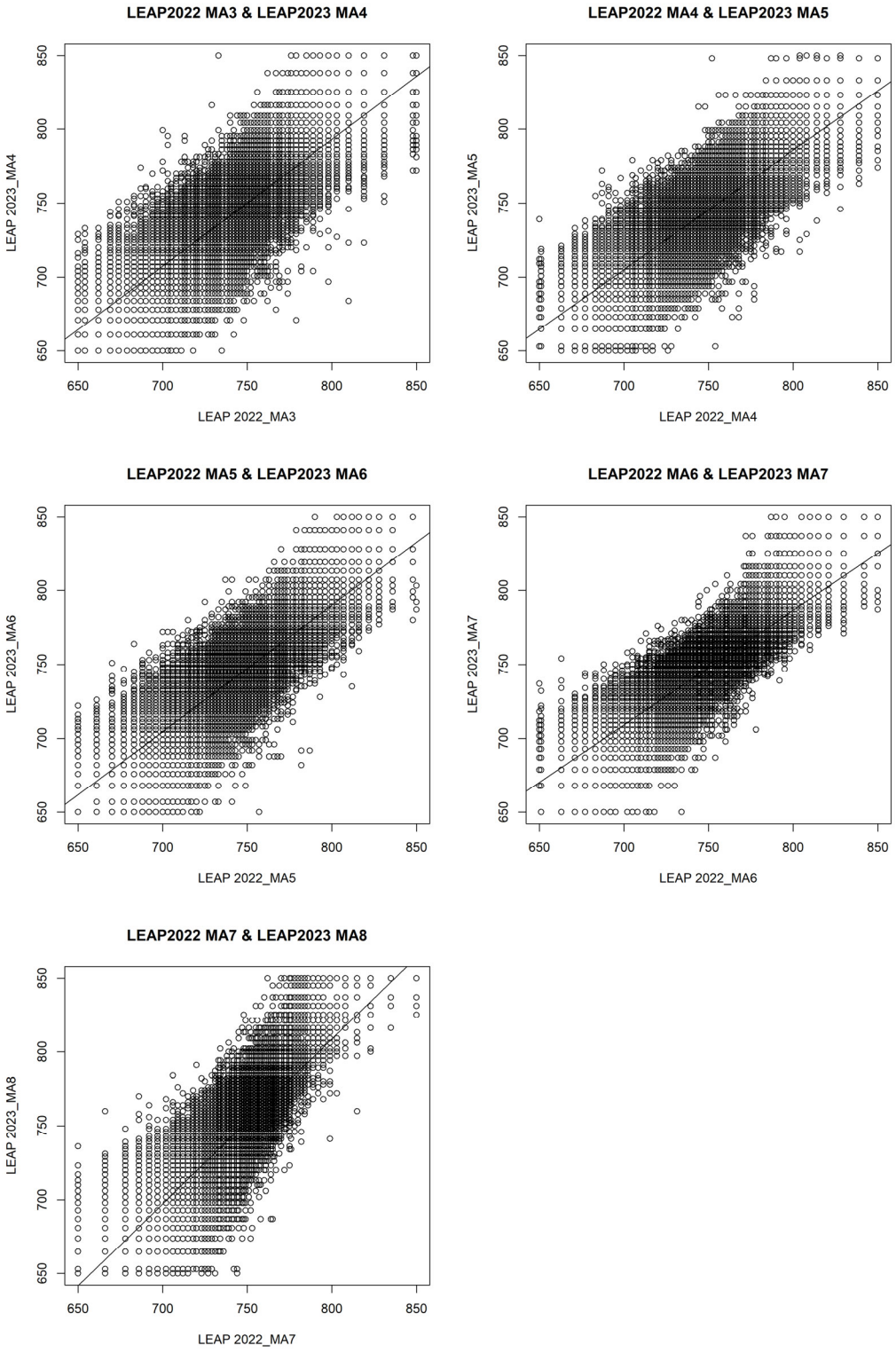


Figure 9.4 Regression Line and Scatter Plots: Mathematics



9.7 Summary

In summary, the overall purpose of establishing construct validity is to ensure that the interpretation of test scores is supported. Evidence of validity is necessary to justify the use of the LEAP 2025 test scores. This evidence addresses multiple best practices of the testing industry but particularly relates to the following standards.

> ***Standard 1.13*** If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided (26).
>
> ***Standard 1.21*** When statistical adjustments, such as those for restriction of range or attenuation, are made, both adjusted and unadjusted coefficients, as well as the specific procedure used, and all statistics used in the adjustment, should be reported. Estimates of the construct-criterion relationship that remove the effects of measurement error on the test should be clearly reported as adjusted estimates (29).
>
> ***Standard 2.0*** Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use (42).
>
> ***Standard 2.3*** For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported (43).
>
> ***Standard 2.13*** The standard error of measurement, both overall and conditional (if reported), should be provided in units of each reported score (45).
>
> ***Standard 2.14*** When possible and appropriate, conditional standard errors of measurement should be reported at several score levels unless there is evidence that the standard error is constant across score levels. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score (46).
>
> ***Standard 2.16*** When a test or combination of measures is used to make classification decisions, estimates should be provided of the percentage of test takers who would be classified in the same way on two replications of the procedure (46).
>
> ***Standard 2.19*** Each method of quantifying the reliability/precision of scores should be described clearly and expressed in terms of statistics appropriate to the method. The sampling procedures used to select test takers for reliability/precision analyses and the descriptive statistics on these samples, subject to privacy obligations where applicable, should be reported (47).

Chapter 10: Fairness

As noted in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), there are varying definitions of fairness. This chapter examines fairness as it relates to minimizing bias on a test. This chapter also discusses test performance among varying subgroups assessed by LEAP 2025 assessments. It should be noted that having differences in test performance among subgroups does not mean that a test is unfair—it simply means that groups perform differently on a test. Even when a test is carefully and properly constructed, differences may exist among subgroups as a result of differences in curriculum or learning by students in the subgroup.

This chapter demonstrates for the LEAP 2025 assessments adhere to AERA, APA, & NCME Standards 3.1–3.6. These standards are from Chapter 3 of the *Standards*, which is titled "Fairness in Testing." Each of these standards is presented in this chapter.

Standard 3.6 states:

> Where credible evidence indicates that test scores may differ in meaning for relevant subgroups in the intended examinee population, test developers and/or users are responsible for examining the evidence for validity of score interpretations for intended uses for individuals from those subgroups. What constitutes a significant difference in subgroup scores and what actions are taken in response to such differences may be defined by applicable laws (65).

Test scores of examinee subgroups that differ in meaning are an ongoing concern in any large-scale testing program. To lessen the possibility of differences in test score meaning, DRC follows several steps in the item development and item selection processes, as is explained in Section 10.1 of this chapter. In addition, the LDOE assessment research and development experts, and Louisiana educators, conduct content and bias reviews on items during the selection process, as explained in Chapter 3. These practices adhere to Standard 3.3, which states, "Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test" (64).

The PARCC consortium, as well as DRC, conducted differential item functioning (DIF) studies of their items prior to operational administrations. Items are typically evaluated for possible DIF in the field test phase of the test development process, and any items flagged for DIF are further examined to determine possible bias. During the ELA and mathematics test development process, DRC content experts tried to avoid including operational items flagged for DIF. Section 10.2 of this chapter explains the steps taken to evaluate LEAP 2025 items using DIF to adhere to Standard 3.3.

In addition, the standardized test administration practices and the extensive training process for test score interpretation for LEAP 2025 comply with Standards 3.4 and 3.5, which state:

> ***Standard 3.4*** Test takers should receive comparable treatment during the test administration and scoring process (65).
>
> ***Standard 3.5*** Test developers should specify and document provisions that have been made to test administration and scoring procedures to remove construct-irrelevant barriers for all relevant subgroups in the test-taker population (65).

Section 10.1 of this chapter is also directly relevant to Standards 3.1 and 3.2.

> ***Standard 3.1*** Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population (63).
>
> ***Standard 3.2*** Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics (64).

This chapter explains the steps taken by DRC to minimize words, phrases, and content that may be regarded as offensive by members of particular demographic subgroups. Section 3.2 of Chapter 3 discusses the content and bias review conducted for LEAP 2025. This review is also critical in fulfilling Standards 3.1 and 3.2. The New Meridian operational items used in the LEAP 2025 forms were critical to the forms construction process. Refer to the New Meridian website for the bias and sensitivity guidelines used and the processes and procedures followed by New Meridian pertaining to these items (see https://newmeridiancorp.org/).

10.1 Minimizing Bias through Careful Test Development

The construction of a test that is fair for all examinees begins in the early stages of planning and development. The item and test development processes that were used to minimize bias are summarized below.

First, careful attention was paid to content validity during the item development and item selection processes. Bias can occur only if the test is measuring different things for different groups. The possibility of bias is reduced by eliminating irrelevant skills or knowledge from the items.

Second, item writers and test developers followed PARCC Fairness and Sensitivity Guidelines for reducing or eliminating bias. DRC test development staff reviewed all items and other testing materials with these guidelines in mind. Internal editorial reviews were conducted by at least three different people: a content editor who directly supervised the item writers, a style editor, and a content supervisor. The final test was again reviewed by people in these same roles and was also subjected to an independent review by the LDOE assessment research and development specialists.

Third, careful attention was given to item statistics throughout the test development process. As part of the test assembly process, attempts were made to avoid using or reusing items with poor statistical fit or distractors with positive point biserial correlations, since this may indicate that an item is testing an ability that is irrelevant to the construct being measured. DIF statistics were also examined during test construction. Items that had exhibited significant DIF against one or more subgroups were removed from further consideration unless it was essential to include them to meet content specifications.

10.2 Evaluating Bias through Differential Item Functioning (DIF) Statistics

After administering the test, an empirical approach known as DIF was used to examine the items. The DIF statistics indicate the degree to which members of a particular focus group perform better or worse than expected on each item as compared to the reference group. The DIF procedures used, and the results of these analyses, are detailed in this section. It should be noted, however, that all items included in LEAP 2025 were thoroughly reviewed for content and bias by the LDOE and DRC content experts to ensure the items do not test knowledge or ability irrelevant to the construct the test intends to measure. Therefore, DIF flags do not necessarily indicate that an item is biased; rather, DIF flags indicate that the item functions differently for

equally able members of different groups (Camilli & Shepard, 1994). Items are not necessarily suppressed from operational scoring if they are flagged for DIF.

The position of DRC concerning test bias is based on two general propositions. First, students may differ in their background knowledge, cognitive and academic skills, languages, attitudes, and values. To the degree that these differences are large, no one curriculum and no one set of instructional materials will be equally suitable for all. Therefore, no one test will be equally appropriate for all. Furthermore, it is difficult to specify what amount of difference can be called large and to determine how these differences will affect the outcome of a particular test. Second, schools have been assigned the tasks of developing certain basic cognitive skills and supporting development of these skills equitably among all students. Therefore, there is a need for tests that measure the common skills and bodies of knowledge that are expected of all learners. The test publisher's task is to develop assessments that measure these key cognitive skills without introducing extraneous or construct-irrelevant elements into the performances on which the measurement is based. If these tests require that students have culturally specific knowledge and skills not taught in school, differences in performance among students can occur because of differences in student background and out-of-school learning. Such tests are measuring different things for different groups and can be called biased (Camilli & Shepard, 1994; Green, 1975).

To lessen this bias, DRC strives to minimize the role of extraneous elements, thereby increasing the number of students for whom the test is appropriate. As discussed above and in Chapter 3 of this report, careful attention is given during the test development and test construction processes to lessen the influence of these elements for large numbers of students. Unfortunately, in some cases these elements may continue to play a substantial role in some cases. To assess the extent to which items may be performing differently for various subgroups of interest, DIF analyses are conducted after each operational test administration.

DIF statistics are used to quantify differences in item performance between two groups after controlling for examinees' overall achievement level. Two DIF statistics that are commonly used for this purpose are the Mantel-Haenszel (MH) statistic (1959) and the standardized mean difference (SMD) between the reference and focal groups, proposed by Dorans and Schmitt (1991).

The MH statistic is computed as follows (Zwick, Donoghue, & Grima, 1993):

$$\text{Mantel } \chi^2 = \frac{\left(\sum_k F_k - \sum_k E(F_k)\right)^2}{\sum_k \text{Var}(F_k)},$$

where F_k is the sum of scores for the focal group at the kth level of the matching variable. Note that the MH statistic is sensitive to N such that larger sample sizes increase the value of chi-square.

In addition to the MH chi-square statistic, the delta statistic (MH-D DIF) was computed for all items. Educational Testing Service (ETS) first developed the MH-D DIF statistic. To compute delta, alpha (the odds ratio) is first computed as follows:

$$\alpha_{MH} = \frac{\sum_{k=1}^{K} N_{r1k} N_{f0k} / N_k}{\sum_{k=1}^{K} N_{f1k} N_{r0k} / N_k},$$

where N_{r1k} is the number of correct responses in the reference group at ability level k, N_{f0k} is the number of incorrect responses in the focal group at ability level k, N_k is the total number of responses, N_{f1k} is the

number of correct responses in the focal group at ability level *k*, and N_{r0k} is the number of incorrect responses in the reference group at ability level *k*. MH-D DIF is then computed as follows:

$$\text{MH-D DIF} = -2.35\ln(\alpha_{MH})$$

For selected-response items, the MH (χ^2_{MH}) statistic was used to evaluate potential DIF items. In the MH procedure, subgroups are matched by their raw total test score, using a contingency table with *K* ability levels. When applying the MH procedure, the log-odds ratio α is assumed to be constant across the *K* matched levels. The χ^2_{MH} , then, estimates a pooled common-odds ratio. Taking the natural logarithm of the common-odds ratio and its confidence limits and multiplying these with the constant –2.35 may then allow the resulting values to be placed on the MH delta metric (Δ_{MH}) for interpretive purposes. Items were flagged for DIF using the following criteria:

- Moderate DIF: Significant MH chi-square statistic ($p < 0.05$) and $1.0 \leq |\text{MH D-DIF}| < 1.5$
- Large DIF: Significant MH chi-square statistic ($p < 0.05$) and $|\text{MH D-DIF}| \geq 1.5$

For constructed-response items, an effect size (ES) statistic based on the MH chi-square will be used. The ES is obtained by dividing the SMD statistics by the standard deviation of the item. The SMD is an effect size index of DIF, which is relatively easy to interpret. The SMD compares the mean of the reference and focal group, adjusting for the distribution of reference and focal group members on the conditioning variable, which for these analyses is the LEAP 2025 raw score. The SMD is computed as follows (Zwick et al., 1993):

$$SMD = p_{Fk}\left(\sum_k m_{Fk} - \sum_k m_{Rk}\right)$$

where p_{Fk} = the proportion of the focal group members at the *k*th level of the matching variable, $m_{Fk} = 1/N_{F1k}$, and $m_{Rk} = 1/N_{R1k}$. Items are flagged using the same rules that are used in NAEP:

- Moderate DIF: If the MH statistic is significant ($p < .05$) and $|\text{ES}|$ is between 0.17 and 0.25
- Large DIF: If the MH statistic is significant ($p < .05$) and $|\text{ES}| \geq 0.25$

A positive DIF value indicates that the item favors the focal group, while a negative value indicates that the item disadvantages the focal group.

DIF analyses were conducted for groups defined by demographic characteristics. Data from test takers who were administered the Spanish language and Braille versions of the test were not used in the analyses. Tables 10.1 and 10.2 show the DIF results for the following subgroups:

Gender: Focal group is females; reference group is males.

Ethnicity: Focal groups are Hispanic/Latino, American Indian or Alaska Native, Asian, Black or African American, and two or more races; reference group is white.

Education Classification: Focal group is students who are classified as special education; reference group is all others.

EL Status: Focal group is students who are classified as EL; reference group is all others.

Economic Status: Focal group is students who are classified as economically disadvantaged; reference group is all others.

Section 504 Status*:* Focal group is students who are classified as Section 504; reference group is all others.

A negative SMD value implies that the focal group has a lower mean item score than the reference group, whereas a positive value implies that the focal group has a higher mean item score than the reference group, conditioned on the matching test score.

The minimum case count for the focal group was set at 200, and the minimum case count for the reference group was set at 400. The DIF analyses are not performed for subgroups of less than 200. In these cases, the statistical procedures do not have sufficient power to detect potential differences.

DIF statistics are produced and examined for all newly field-tested items and for all items from New Meridian being administered for the first time operationally in Louisiana. In the spring 2023 administration, items were field tested in all grades for mathematics. Tables 10.1 (ELA) and 10.2 (mathematics) summarize the number of DIF flags by content area, grade, and test form for each focal group that included at least 200 students. Results are not reported (NR) for groups with an insufficient number of students. The analyses were conducted by test form. The PBT form for ELA students in grade 3 (see Table 10.1) can be considered as an example. In this form, twenty-two. Of them, one items exhibited moderate negative DIF each for the Hispanic/Latino group, the Black or African American group, and for the EL group.

Table 10.1 2023 LEAP 2025 DIF Statistics: Number of Flagged Items, English Language Arts

DIF Statistics: English Language Arts					Count of Items at DIF Magnitude			
					Moderate		Large	
Grade	Mode	Number of Items	Category	Focal Group	B-	B+	C-	C+
3	CBT	8	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
3	PBT	8	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
4	CBT	17	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0

DIF Statistics: English Language Arts					Count of Items at DIF Magnitude			
					Moderate		Large	
Grade	Mode	Number of Items	Category	Focal Group	B-	B+	C-	C+
5	CBT	10	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	1	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
6	CBT	22	Gender	Female	1	1	1	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
7	CBT	14	Gender	Female	1	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0

DIF Statistics: English Language Arts					Count of Items at DIF Magnitude			
					Moderate		Large	
Grade	Mode	Number of Items	Category	Focal Group	B-	B+	C-	C+
8	CBT	20	Gender	Female	3	0	0	0
			Ethnicity	Hispanic/Latino	1	0	0	0
			Ethnicity	American Indian or Alaska Native	1	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	2	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0

Table 10.2 2023 LEAP 2025 DIF Statistics: Number of Flagged Items, Mathematics

DIF Statistics: Mathematics					Count of Items at DIF Magnitude			
					Moderate		Large	
Grade	Mode	Number of Items	Category	Focal Group	B-	B+	C-	C+
3	CBT	12	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	1	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
3	PBT	12	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	1	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0

DIF Statistics: Mathematics					Count of Items at DIF Magnitude			
					Moderate		Large	
Grade	Mode	Number of Items	Category	Focal Group	B-	B+	C-	C+
4	CBT	16	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	1	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	1	2	0	1
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
5	CBT	60	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	1	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	1	0	0	2
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
6	CBT	17	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	1	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	2
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0

DIF Statistics: Mathematics					Count of Items at DIF Magnitude			
					Moderate		Large	
Grade	Mode	Number of Items	Category	Focal Group	B-	B+	C-	C+
7	CBT	19	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	1	0	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
8	CBT	102	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	1	0	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0

10.3 Spanish and English Language Form Comparability

10.3.1 Reliability of Spanish Language Forms

Table 10.3 reports the form reliability of the Spanish language forms. Cronbach's alpha values ranged from 0.92 to 0.94. Please note that the interpretation of Cronbach's alpha values needs to be conservative due to the small case counts, especially for grade 3 PBT.

Table 10.3 Form Reliability for the Spanish Language Forms

Content	Grade	Mode	Number of Items	Number of Score Points	SEM	Cronbach's Alpha	N-Count
Mathematics	3	CBT	43	62	3.03	0.92	≥70
Mathematics	3	PBT	43	62	3.52	0.92	≥20
Mathematics	4	CBT	43	62	2.18	0.94	≥80
Mathematics	5	CBT	43	62	2.36	0.93	≥90
Mathematics	6	CBT	43	66	1.56	0.94	≥100
Mathematics	7	CBT	43	66	1.40	0.94	≥120
Mathematics	8	CBT	41	65	1.63	0.92	≥110

10.3.2 DIF Statistics for Test Language

All items on one CBT and one PBT form of the mathematics test at each grade are transadapted from English into Spanish. Transadaptation takes into consideration linguistic and cultural differences and grade-level appropriate words. By accounting for these differences, the achievement of Spanish speakers can be measured in the same way as the achievement of English speakers. Please refer to Appendix C for more information about the transadaptation of Spanish mathematics forms. To help confirm that the test items can be measured similarly regardless of the language in which the items are published, two DIF analyses were performed using the 2023 LEAP 2025 mathematics operational items, regardless of student count in the reference or focal group. Smaller counts for the groups needed to be tolerated since the overall count for those being administered the Spanish form was low.

For the first analysis, student responses for the shared operational items between 2022 and 2023 LEAP 2025 mathematics were combined. This approach increased the number of students who took the Spanish versions of the items. The Mantel-Haenszel (MH) and the Standardized Mean Difference (SMD) DIF procedures were performed on these shared items and DIF flags applied. The second analysis focused on the items that were not common between the 2022 and 2023 administrations. The MH and the SMD DIF procedures were performed on all 2023 LEAP 2025 operational items, including items that were unique to the 2023 administration in addition to those in common with the 2022 administration. However, DIF flags were applied to only the items that were not shared between 2022 and 2023.

For both analyses, DIF results were carefully reviewed whenever sample sizes were smaller than the required minimum sample size and when an item showed large (C) DIF. All items were determined by the LDOE to be suitable for scoring. Table 10.4 summarizes how many items overall exhibited moderate or large DIF in mathematics.

Table 10.4 2023 LEAP 2025 DIF Statistics: Number of Flagged Items, Mathematics

DIF Statistics: Mathematics				Count of Items at DIF Magnitude			
				Moderate		Large	
Grade	Number of Items	Category	Focal Group*	B-	B+	C-	C+
3	28	Test Language	Spanish	7	0	0	0
4	28	Test Language	Spanish	2	1	1	0
5	23	Test Language	Spanish	3	0	0	0
6	28	Test Language	Spanish	2	0	2	0
7	23	Test Language	Spanish	1	0	2	0
8	28	Test Language	Spanish	1	0	3	0

*Reference group are those that were administered the English version of the test

10.3.3 Propensity Score Matching Study

The fairness of using the transadapted form was also evaluated by examining the performance of those who took either the Spanish form or the English form. A propensity score matching study (PSM) matches groups based on similar characteristics and then compares performance. The PSM study groups were selected using covariates (matching variables), such as students' spring 2023 LEAP 2025 ELA score and their bio-demographic information, such as gender, ethnicity, economic status, and English learner status. Equivalent groups were created with the difference being which form – Spanish language or English language – was administered. The mathematics Spanish test was administered to a smaller number of students than the mathematics English test; therefore, the group who took the Spanish test was designated as the focal group for the PSM study (Rosenbaum & Rubin, 1983), and the English language test takers were considered to be the reference group. Table 10.5 shows the number of equivalent Spanish test takers and English test students matched by the PSM method using the R package, MatchIt for PSM.

Table 10.5 Number of Students Used for Propensity Score Matching

Grade	Spanish	English	
	Total	Total	Selected
3	≥90	≥48,980	≥90
4	≥80	≥48,760	≥80
5	≥90	≥48,110	≥90
6	≥100	≥43,220	≥100
7	≥110	≥42,280	≥110
8	≥100	≥39,950	≥100

*Total: Number of students who have information for all covariates

Scale scores of the Spanish language and English language administrations were estimated using the item parameters for score reporting, and their difference scores were calculated. Effect sizes (ES) of the difference scores were calculated as follows:

ES = (Spanish Mean – English Mean)/$\sqrt{(SPN\ \ VAR + ENG\ VAR)/2}$, where *VAR* = SD^2.

Table 10.6 through Table 10.11, summarize, for the flagged items, the mean, standard deviation, effect size (ES), and flag for the ES for mathematics items by Spanish and matching English tests. Two flag criteria, |0.2| and |0.5| were applied as small differences (B) and medium differences (C) flags. When |0.2| was applied, the count of items flagged ranged from seven in grade 3 to 14 in grade 5. Items with larger ES values greater

than |0.5| included one each in grade 4 and grade 8. Please note that the results of this PSM should be carefully used due to the relatively small sample size.

Table 10.6 Item Statistics and Effect Size: Grade 3

Item id	Mean		Standard Deviation		ES	Flag
	Spanish	English	Spanish	English		
870659	0.33	0.23	0.47	0.42	-0.23	B
897734	0.32	0.12	0.86	0.46	-0.28	B
1026036	0.70	0.52	0.46	0.50	-0.39	B
1025975	0.74	0.59	0.75	0.70	-0.21	B
896859	0.34	0.20	0.48	0.40	-0.33	B
870678	1.14	0.60	1.55	1.26	-0.39	B
1026140	0.73	0.58	0.45	0.50	-0.33	B

Table 10.7 Item Statistics and Effect Size: Grade 4

Item id	Mean		Standard Deviation		ES	Flag
	Spanish	English	Spanish	English		
1075317	0.17	0.33	0.38	0.47	0.35	B
897440	0.21	0.37	0.41	0.49	0.36	B
1026610	0.19	0.10	0.39	0.31	-0.23	B
897465	0.15	0.30	0.36	0.46	0.36	B
935180	0.12	0.20	0.32	0.40	0.22	B
981883	0.19	0.28	0.39	0.45	0.22	B
878671	0.59	0.80	0.93	1.14	0.20	B
897463	0.16	0.45	0.37	0.50	0.66	C
897445	0.38	0.50	0.49	0.50	0.23	B
897310	0.45	0.33	0.50	0.47	-0.26	B
1026698	0.34	0.49	0.48	0.50	0.31	B
1026634	0.06	0.22	0.24	0.42	0.48	B

Table 10.8 Item Statistics and Effect Size: Grade 5

Item id	Mean		Standard Deviation		ES	Flag
	Spanish	English	Spanish	English		
903375	0.19	0.10	0.39	0.31	-0.24	B
914160	0.57	0.69	0.50	0.47	0.24	B
897988	0.46	0.56	0.50	0.50	0.21	B
898027	0.49	0.59	0.50	0.49	0.21	B
902412	0.42	0.57	0.61	0.66	0.25	B
902413	0.21	0.10	0.50	0.31	-0.25	B
898019	0.45	0.56	0.50	0.50	0.23	B
898165	0.22	0.31	0.42	0.47	0.21	B
1119175	0.18	0.29	0.46	0.56	0.22	B
982476	0.41	0.14	0.73	0.45	-0.44	B
898024	0.29	0.42	0.46	0.50	0.26	B
904183	0.08	0.17	0.28	0.37	0.25	B
982544	0.36	0.24	0.48	0.43	-0.27	B
982497	0.91	0.48	1.67	0.94	-0.32	B

Table 10.9 Item Statistics and Effect Size: Grade 6

Item id	Mean		Standard Deviation		ES	Flag
	Spanish	English	Spanish	English		
981980	0.13	0.22	0.34	0.41	0.24	B
914267	0.10	0.04	0.30	0.20	-0.23	B
1116275	0.18	0.35	0.38	0.48	0.39	B
1116259	0.23	0.14	0.42	0.35	-0.23	B
981997	0.32	0.20	0.47	0.40	-0.27	B
800191	0.26	0.38	0.44	0.49	0.26	B
1027521	0.03	0.08	0.17	0.27	0.22	B
1075676	0.22	0.08	0.41	0.27	-0.40	B
903080	0.01	0.05	0.10	0.22	0.23	B
982031	0.15	0.24	0.36	0.43	0.23	B

Table 10.10 Item Statistics and Effect Size: Grade 7

Item id	Mean		Standard Deviation		ES	Flag
	Spanish	English	Spanish	English		
898450	0.22	0.10	0.42	0.30	-0.34	B
899319	0.24	0.38	0.43	0.49	0.31	B
983014	0.13	0.06	0.34	0.24	-0.24	B
983508	0.10	0.03	0.38	0.16	-0.24	B
915699	0.18	0.05	0.38	0.22	-0.39	B
1075674	0.75	0.57	0.60	0.66	-0.29	B
902989	0.79	0.56	0.72	0.73	-0.31	B
982929	0.06	0.15	0.28	0.48	0.22	B

Table 10.11 Item Statistics and Effect Size: Grade 8

Item id	Mean		Standard Deviation		ES	Flag
	Spanish	English	Spanish	English		
900515	0.16	0.27	0.37	0.45	0.27	B
898442	0.31	0.42	0.47	0.50	0.21	B
914425	0.34	0.48	0.48	0.50	0.28	B
1023521	0.34	0.25	0.48	0.44	-0.20	B
900500	0.42	0.26	0.60	0.48	-0.29	B
984139	0.48	0.20	0.88	0.62	-0.36	B
878742	0.50	0.31	0.50	0.46	-0.40	B
983050	0.32	0.14	0.53	0.37	-0.41	B
899327	0.46	0.10	0.74	0.41	-0.60	C

10.4 Evaluating Bias through Impact Analysis

The impact of achievement testing on subgroups can be determined and reported in the form of average scores and also in terms of test score reliability. Tables 10.12–10.18 present the number of students, test form reliability statistics (i.e., coefficient alpha; see Chapter 9), scale score means and standard deviations, and effect size (i.e., Cohen's *d*) for the various subgroups of interest by form.

10.4.1 Reliability

Tables 10.12–10.18 show the test form reliability coefficients and SEM by student subgroups. Analyses were based on census data, after removing data from the test takers who were administered the Spanish language and Braille versions of the test forms. The reliability coefficients for English language arts forms ranged from 0.82 to 0.94. For mathematics the reliability coefficients ranged from 0.86 to 0.95. These analyses show that the test reliability is of acceptable magnitude for all the subgroups. Note that the reliability coefficients are NR for subgroups with fewer than 10 students.

Table 10.12 Grade 3 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥22,120	0.89	4.37	≥22,100	0.92	3.63
Gender						
Male	≥11,400	0.89	4.28	≥11,380	0.93	3.60
Female	≥10,720	0.88	4.46	≥10,710	0.92	3.64
Ethnicity						
Hispanic/Latino	≥3,170	0.89	4.24	≥3,140	0.93	3.56
American Indian or Alaska Native	≥170	0.88	4.60	≥170	0.92	3.64
Asian	≥350	0.90	4.75	≥350	0.91	3.74
Black or African American	≥9,070	0.87	4.29	≥9,090	0.91	3.47
Native Hawaiian or Other Pacific	≥20	0.85	4.45	≥20	0.91	3.50
White	≥8,350	0.88	4.49	≥8,350	0.92	3.69
Two or More Races	≥950	0.89	4.52	≥950	0.92	3.70
Education Classification						
Regular	≥19,290	0.88	4.42	≥19,270	0.92	3.65
Special	≥2,830	0.86	3.94	≥2,830	0.92	3.37
Economic Status						
Economically Disadvantaged	≥16,520	0.87	4.32	≥16,510	0.92	3.55
Not Economically Disadvantaged	≥5,590	0.88	4.60	≥5,580	0.92	3.79
English Learner Status						
Not English Learner	≥20,150	0.88	4.41	≥20,160	0.92	3.64
English Learner	≥1,970	0.80	3.96	≥1,940	0.91	3.36
Section 504 Status						
Not Section 504	≥20,530	0.89	4.38	≥20,510	0.92	3.63
Section 504	≥1,580	0.84	4.18	≥1,580	0.91	3.45
Migrant Status						
Not Migrant	≥22,070	0.89	4.37	≥22,050	0.92	3.63
Migrant	≥40	0.88	4.60	≥40	0.90	3.48
Homeless Status						
Homeless	≥570	0.85	4.27	≥570	0.91	3.40
Not Homeless	≥21,550	0.89	4.38	≥21,530	0.92	3.63
Foster Care Status						
Foster Care	≥50	0.84	4.45	≥50	0.91	3.43
Not in Foster Care	≥22,060	0.89	4.37	≥22,040	0.92	3.63
Military Affiliation						
Military Affiliated	≥250	0.87	4.62	≥250	0.91	3.72
Not Military Affiliated	≥21,870	0.89	4.37	≥21,850	0.92	3.62

Table 10.13 Grade 3 Paper-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥27,200	0.88	4.46	≥27,170	0.92	3.83
Gender						
Male	≥13,740	0.88	4.40	≥13,720	0.93	3.80
Female	≥13,460	0.88	4.50	≥13,450	0.92	3.84
Ethnicity						
Hispanic/Latino	≥2,090	0.89	4.43	≥2,030	0.92	3.80
American Indian or Alaska Native	≥100	0.87	4.44	≥100	0.93	3.74
Asian	≥390	0.88	4.63	≥400	0.90	3.77
Black or African American	≥11,240	0.87	4.36	≥11,270	0.91	3.68
Native Hawaiian or Other Pacific	≥30	0.90	4.17	≥30	0.92	3.88
White	≥12,320	0.87	4.55	≥12,320	0.91	3.83
Two or More Races	≥1,000	0.87	4.54	≥1,000	0.91	3.87
Education Classification						
Regular	≥23,790	0.88	4.48	≥23,760	0.92	3.84
Special	≥3,410	0.87	4.21	≥3,410	0.93	3.63
Economic Status						
Economically Disadvantaged	≥18,070	0.87	4.40	≥18,050	0.92	3.74
Not Economically Disadvantaged	≥9,130	0.86	4.58	≥9,120	0.91	3.84
English Learner Status						
Not English Learner	≥26,250	0.88	4.47	≥26,290	0.92	3.83
English Learner	≥950	0.83	4.24	≥880	0.91	3.60
Section 504 Status						
Not Section 504	≥25,130	0.88	4.47	≥25,100	0.92	3.83
Section 504	≥2,070	0.85	4.36	≥2,070	0.91	3.71
Migrant Status						
Not Migrant	≥27,150	0.88	4.46	≥27,120	0.92	3.83
Migrant	≥50	0.88	4.10	≥50	0.92	3.71
Homeless Status						
Homeless	≥590	0.87	4.19	≥580	0.92	3.53
Not Homeless	≥26,610	0.88	4.47	≥26,590	0.92	3.83
Foster Care Status						
Foster Care	≥110	0.87	4.31	≥110	0.92	3.69
Not in Foster Care	≥27,090	0.88	4.46	≥27,060	0.92	3.83
Military Affiliation						
Military Affiliated	≥720	0.86	4.59	≥720	0.91	3.85
Not Military Affiliated	≥26,480	0.88	4.46	≥26,450	0.92	3.82

Table 10.14 Grade 4 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥48,880	0.91	5.01	≥48,790	0.94	3.83
Gender						
Male	≥25,010	0.91	4.93	≥24,960	0.94	3.83
Female	≥23,860	0.90	5.07	≥23,820	0.93	3.81
Ethnicity						
Hispanic/Latino	≥5,150	0.91	4.86	≥5,070	0.94	3.74
American Indian or Alaska Native	≥280	0.89	5.16	≥280	0.93	3.95
Asian	≥820	0.91	5.08	≥820	0.93	3.90
Black or African American	≥20,120	0.89	4.90	≥20,110	0.92	3.54
Native Hawaiian or Other Pacific	≥40	0.92	5.08	≥40	0.93	4.11
White	≥20,550	0.90	5.10	≥20,550	0.93	3.97
Two or More Races	≥1,880	0.89	5.14	≥1,870	0.93	3.88
Education Classification						
Regular	≥42,850	0.90	5.05	≥42,750	0.93	3.86
Special	≥6,030	0.89	4.59	≥6,030	0.93	3.41
Economic Status						
Economically Disadvantaged	≥34,060	0.89	4.96	≥33,970	0.93	3.67
Not Economically Disadvantaged	≥14,810	0.90	5.15	≥14,810	0.93	4.04
English Learner Status						
Not English Learner	≥46,190	0.90	5.04	≥46,180	0.94	3.85
English Learner	≥2,690	0.85	4.49	≥2,600	0.92	3.34
Section 504 Status						
Not Section 504	≥44,560	0.91	5.02	≥44,470	0.94	3.84
Section 504	≥4,320	0.88	4.84	≥4,320	0.93	3.60
Migrant Status						
Not Migrant	≥48,820	0.91	5.01	≥48,73	0.94	3.83
Migrant	≥60	0.92	4.84	≥50	0.93	3.86
Homeless Status						
Homeless	≥1,150	0.88	4.78	≥1,140	0.93	3.38
Not Homeless	≥47,730	0.91	5.01	≥47,640	0.94	3.83
Foster Care Status						
Foster Care	≥170	0.88	4.87	≥170	0.92	3.64
Not in Foster Care	≥48,700	0.91	5.01	≥48,610	0.94	3.83
Military Affiliation						
Military Affiliated	≥910	0.89	5.09	≥910	0.92	3.96
Not Military Affiliated	≥47,960	0.91	5.01	≥47,870	0.94	3.82

Table 10.15 Grade 5 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥48,310	0.90	5.23	≥48,170	0.93	3.63
Gender						
Male	≥24,720	0.90	5.12	≥24,640	0.94	3.61
Female	≥23,590	0.89	5.28	≥23,520	0.93	3.65
Ethnicity						
Hispanic/Latino	≥5,180	0.91	5.08	≥5,080	0.93	3.57
American Indian or Alaska Native	≥250	0.88	5.34	≥250	0.92	3.73
Asian	≥800	0.90	5.31	≥800	0.93	3.59
Black or African American	≥20,300	0.88	5.18	≥20,300	0.92	3.41
Native Hawaiian or Other Pacific	≥30	0.90	5.36	≥30	0.91	3.64
White	≥20,030	0.88	5.30	≥20,010	0.93	3.71
Two or More Races	≥1,680	0.88	5.32	≥1,670	0.93	3.66
Education Classification						
Regular	≥42,640	0.89	5.26	≥42,500	0.93	3.66
Special	≥5,670	0.88	4.65	≥5,660	0.91	3.18
Economic Status						
Economically Disadvantaged	≥33,360	0.89	5.19	≥33,280	0.93	3.53
Not Economically Disadvantaged	≥14,940	0.88	5.36	≥14,890	0.93	3.75
English Learner Status						
Not English Learner	≥46,070	0.89	5.25	≥46,030	0.93	3.65
English Learner	≥2,230	0.84	4.62	≥2,140	0.91	3.13
Section 504 Status						
Not Section 504	≥43,570	0.90	5.23	≥43,450	0.93	3.64
Section 504	≥4,730	0.88	5.10	≥4,720	0.93	3.49
Migrant Status						
Not Migrant	≥48,240	0.90	5.23	≥48,110	0.93	3.63
Migrant	≥60	0.92	5.16	≥60	0.95	3.50
Homeless Status						
Homeless	≥1,110	0.89	5.03	≥1,100	0.92	3.36
Not Homeless	≥47,190	0.90	5.23	≥47,070	0.93	3.64
Foster Care Status						
Foster Care	≥140	0.91	4.92	≥140	0.93	3.46
Not in Foster Care	≥48,160	0.90	5.23	≥48,020	0.93	3.63
Military Affiliation						
Military Affiliated	≥970	0.88	5.36	≥970	0.92	3.71
Not Military Affiliated	≥47,330	0.90	5.22	≥47,200	0.93	3.63

Table 10.16 Grade 6 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥43,370	0.92	4.89	≥48,240	0.94	3.78
Gender						
Male	≥22,220	0.92	4.72	≥24,680	0.94	3.75
Female	≥21,140	0.92	4.99	≥23,560	0.94	3.81
Ethnicity						
Hispanic/Latino	≥4,710	0.92	4.74	≥4,890	0.94	3.66
American Indian or Alaska Native	≥200	0.91	4.97	≥250	0.93	3.85
Asian	≥700	0.93	5.00	≥730	0.94	4.04
Black or African American	≥18,280	0.90	4.76	≥20,390	0.92	3.45
Native Hawaiian or Other Pacific	≥40	0.94	4.78	≥30	0.94	3.81
White	≥17,920	0.91	5.00	≥20,230	0.93	3.97
Two or More Races	≥1,490	0.91	4.97	≥1,670	0.93	3.86
Education Classification						
Regular	≥38,560	0.92	4.93	≥42,830	0.94	3.83
Special	≥4,810	0.89	4.33	≥5,400	0.92	3.15
Economic Status						
Economically Disadvantaged	≥29,580	0.91	4.80	≥32,930	0.93	3.61
Not Economically Disadvantaged	≥13,790	0.91	5.06	≥15,310	0.93	4.04
English Learner Status						
Not English Learner	≥41,480	0.92	4.91	≥46,370	0.94	3.80
English Learner	≥1,890	0.82	4.17	≥1,870	0.91	3.05
Section 504 Status						
Not Section 504	≥38,800	0.92	4.90	≥43,220	0.94	3.80
Section 504	≥4,570	0.90	4.76	≥5,020	0.93	3.56
Migrant Status						
Not Migrant	≥43,300	0.92	4.89	≥48,180	0.94	3.78
Migrant	≥60	0.92	4.63	≥60	0.92	3.59
Homeless Status						
Homeless	≥1,020	0.90	4.63	≥1,040	0.93	3.40
Not Homeless	≥42,350	0.92	4.89	≥47,200	0.94	3.79
Foster Care Status						
Foster Care	≥120	0.88	4.67	≥130	0.93	3.42
Not in Foster Care	≥43,250	0.92	4.89	≥48,110	0.94	3.79
Military Affiliation						
Military Affiliated	≥890	0.91	5.10	≥920	0.92	3.98
Not Military Affiliated	≥42,480	0.92	4.88	≥47,320	0.94	3.78

Table 10.17 Grade 7 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥42,460	0.93	5.61	≥48,790	0.94	3.61
Gender						
Male	≥21,830	0.93	5.51	≥25,150	0.95	3.58
Female	≥20,630	0.92	5.61	≥23,640	0.94	3.63
Ethnicity						
Hispanic/Latino	≥4,860	0.93	5.46	≥5,100	0.94	3.48
American Indian or Alaska Native	≥210	0.91	5.75	≥270	0.93	3.71
Asian	≥710	0.94	5.50	≥780	0.95	4.04
Black or African American	≥17,650	0.91	5.57	≥20,490	0.92	3.24
Native Hawaiian or Other Pacific	≥30	0.94	5.57	≥30	0.94	3.95
White	≥17,490	0.92	5.65	≥20,390	0.94	3.84
Two or More Races	≥1,460	0.93	5.64	≥1,670	0.94	3.67
Education Classification						
Regular	≥38,020	0.92	5.61	≥43,630	0.94	3.66
Special	≥4,430	0.90	5.05	≥5,160	0.91	2.93
Economic Status						
Economically Disadvantaged	≥28,730	0.92	5.58	≥33,170	0.93	3.40
Not Economically Disadvantaged	≥13,730	0.92	5.67	≥15,620	0.94	3.92
English Learner Status						
Not English Learner	≥40,450	0.93	5.62	≥46,790	0.94	3.63
English Learner	≥2,000	0.85	5.08	≥2,000	0.89	2.83
Section 504 Status						
Not Section 504	≥37,870	0.93	5.61	≥43,560	0.94	3.64
Section 504	≥4,580	0.91	5.50	≥5,230	0.93	3.31
Migrant Status						
Not Migrant	≥42,400	0.93	5.60	≥48,730	0.94	3.61
Migrant	≥60	0.94	5.88	≥60	0.92	3.65
Homeless Status						
Homeless	≥980	0.91	5.49	≥1,010	0.92	3.15
Not Homeless	≥41,480	0.93	5.61	≥47,780	0.94	3.62
Foster Care Status						
Foster Care	≥120	0.92	5.44	≥140	0.93	3.11
Not in Foster Care	≥42,340	0.93	5.61	≥48,650	0.94	3.61
Military Affiliation						
Military Affiliated	≥850	0.92	5.61	≥870	0.94	3.88
Not Military Affiliated	≥41,610	0.93	5.60	≥47,920	0.94	3.60

Table 10.18 Grade 8 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥45,780	0.91	5.64	≥44,120	0.92	3.43
Gender						
Male	≥23,150	0.91	5.58	≥22,410	0.93	3.38
Female	≥22,630	0.91	5.61	≥21,710	0.92	3.48
Ethnicity						
Hispanic/Latino	≥4,830	0.92	5.61	≥4,410	0.91	3.31
American Indian or Alaska Native	≥210	0.90	5.76	≥260	0.91	3.46
Asian	≥720	0.90	5.55	≥500	0.94	3.99
Black or African American	≥19,590	0.90	5.64	≥20,190	0.90	3.16
Native Hawaiian or Other Pacific	≥40	0.85	5.67	≥30	0.92	4.00
White	≥18,780	0.90	5.67	≥17,230	0.92	3.68
Two or More Races	≥1,590	0.90	5.68	≥1,480	0.92	3.49
Education Classification						
Regular	≥41,150	0.91	5.64	≥39,100	0.92	3.49
Special	≥4,630	0.88	5.27	≥5,020	0.87	2.84
Economic Status						
Economically Disadvantaged	≥30,800	0.90	5.64	≥31,160	0.91	3.29
Not Economically Disadvantaged	≥14,980	0.90	5.74	≥12,960	0.93	3.73
English Learner Status						
Not English Learner	≥43,890	0.91	5.66	≥42,310	0.92	3.45
English Learner	≥1,890	0.85	5.35	≥1,810	0.86	2.89
Section 504 Status						
Not Section 504	≥40,860	0.91	5.64	≥39,030	0.92	3.46
Section 504	≥4,910	0.90	5.61	≥5,090	0.91	3.20
Migrant Status						
Not Migrant	≥45,720	0.91	5.64	≥44,070	0.92	3.43
Migrant	≥50	0.90	5.85	≥50	0.91	3.47
Homeless Status						
Homeless	≥970	0.90	5.62	≥970	0.90	3.12
Not Homeless	≥44,800	0.91	5.64	≥43,150	0.92	3.44
Foster Care Status						
Foster Care	≥140	0.90	5.51	≥150	0.88	3.12
Not in Foster Care	≥45,630	0.91	5.64	≥43,970	0.92	3.43
Military Affiliation						
Military Affiliated	≥860	0.89	5.67	≥690	0.92	3.74
Not Military Affiliated	≥44,910	0.91	5.64	≥43,430	0.92	3.43

10.4.2 Effect Size

One way to evaluate the magnitude of the standardized mean difference (SMD) is to calculate the ES. Cohen's *d* was used on the statewide population to calculate the ES. Cohen's *d* is given by the following formula:

$$d = \frac{\overline{x_a} - \overline{x_b}}{\sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{(n_a + n_b) - 2}}},$$

where $\overline{x_a}$ is the mean score of group A, $\overline{x_b}$ is the mean score of group B, s_a^2 is the variance of group A, s_b^2 is the variance of group B, n_a is the number of students in group A, and n_b is the number of students in group B.

Cohen's *d*, then, expresses the difference in group means in terms of the standard deviation. For example, if *d* = .34 for two groups, then it may be interpreted that the SMD between the two groups is .34 of the pooled standard deviation. Cohen (1988) offered guidelines for interpreting the meaning of the *d* statistic: *d* = .20 is a small ES, *d* = .50 is a medium ES, and *d* = .80 is a large ES.

Using Cohen's (1988) guidelines, certain trends become apparent in Tables 10.19–10.25. Results are NR for subgroups with fewer than 10 students. If the effect size is negative, that means the group performs at a higher level than the group to which it's being compared. A positive effect size indicates the group performs at a lower level than the group to which it is being compared. For example, in Table 10.19 in regard to the ELA test, the effect size for the group female is -0.16 indicating that although there is less than a small difference in performance, females are scoring higher than males. On the ELA test in most grades, there are small differences in mean test scores between females and males where females outperform males. For most ELA and mathematics tests, mean scale scores and ES show that Asian and white students tend to outperform other ethnicity groups across grades. For most ELA and mathematics tests, there were clear performance differences between regular education and special education students in Education Classification, between not economically disadvantaged and economically disadvantaged in economic status, and non-EL and EL students in EL status.

Table 10.19 Impact Analysis, Grade 3 Computer-Based Test Administration

	ELA				Mathematics			
		Scale Score				Scale Score		
Group	N	Mean	Std. Dev.	Effect Size	N	Mean	Std. Dev.	Effect Size
All Students	≥22,120	731.22	44.59		≥22,180	733.37	31.93	
Gender								
Male	≥11,400	727.73	44.33		≥11,430	733.96	33.09	
Female	≥10,720	734.94	44.57	-0.16	≥10,740	732.74	30.64	0.04
Ethnicity								
White	≥8,350	745.16	43.04		≥8,350	744.57	30.71	
Hispanic/Latino	≥3,170	719.16	44.53	0.60	≥3,210	727.38	32.34	0.55
American Indian or Alaska Native	≥170	738.78	45.09	0.15	≥170	736.53	30.49	0.26
Asian	≥350	762.59	48.62	-0.40	≥350	760.33	32.28	-0.51
Black or African American	≥9,070	720.12	41.02	0.60	≥9,090	723.35	28.72	0.71
Native Hawaiian or Other Pacific	≥20	746.76	37.77	-0.04	≥20	737.33	26.57	0.24
Two or More Races	≥950	741.92	45.99	0.07	≥950	740.28	31.20	0.14
Economic Status								
Not Economically Disadvantaged	≥5,590	751.89	44.76		≥5,590	748.89	31.59	
Economically Disadvantaged	≥16,520	724.22	42.30	0.64	≥16,580	728.13	30.30	0.68
Education Classification								
Regular	≥19,290	734.99	44.21		≥19,340	735.93	31.40	
Special	≥2,830	705.52	38.17	0.68	≥2,830	715.86	29.97	0.64
English Learner Status								
Not English Learner	≥20,150	734.39	44.21		≥20,160	735.02	31.76	
English Learner	≥1,970	698.84	34.34	0.82	≥2,010	716.81	28.75	0.58
Section 504 Status								
Not Section 504	≥20,530	732.38	44.90		≥20,590	734.21	32.09	
Section 504	≥1,580	716.24	37.18	0.36	≥1,580	722.40	27.48	0.37
Migrant Status								
Not Migrant	≥22,070	731.22	44.59		≥22,130	733.37	31.94	
Migrant	≥40	733.38	45.56	-0.05	≥40	731.29	26.00	0.07
Homeless Status								
Not Homeless	≥21,550	731.64	44.64		≥21,600	733.75	31.92	
Homeless	≥570	715.42	39.54	0.36	≥570	718.88	29.00	0.47
Foster Care Status								
Not in Foster Care	≥22,060	731.25	44.60		≥22,120	733.39	31.94	
Foster Care	≥50	719.93	38.45	0.25	≥50	723.08	28.18	0.32
Military Affiliation								
Not Military Affiliated	≥21,870	730.96	44.54		≥21,930	733.19	31.93	
Military Affiliated	≥250	754.17	42.29	-0.52	≥250	749.03	28.17	-0.50

Table 10.20 Impact Analysis, Grade 3 Paper-Based Test Administration

	ELA				Mathematics			
		Scale Score				Scale Score		
Group	N	Mean	Std. Dev.	Effect Size	N	Mean	Std. Dev.	Effect Size
All Students	≥27,210	746.86	43.59		≥27,200	740.27	33.44	
Gender								
Male	≥13,740	742.84	43.71		≥13,730	740.81	34.62	
Female	≥13,460	750.96	43.08	-0.19	≥13,460	739.72	32.18	0.03
Ethnicity								
White	≥12,320	759.92	40.99		≥12,320	751.25	31.77	
Hispanic/Latino	≥2,090	738.00	45.16	0.53	≥2,050	737.31	33.36	0.44
American Indian or Alaska Native	≥100	751.02	42.14	0.22	≥100	742.61	33.34	0.27
Asian	≥390	771.71	43.80	-0.29	≥400	768.68	31.86	-0.55
Black or African American	≥11,240	732.86	41.35	0.66	≥11,280	727.58	30.54	0.76
Native Hawaiian or Other Pacific	≥30	755.55	44.89	0.11	≥30	747.70	32.64	0.11
Two or More Races	≥1,000	751.31	41.47	0.21	≥1,000	742.36	31.07	0.28
Economic Status								
Not Economically Disadvantaged	≥9,130	766.23	41.12		≥9,130	756.13	31.96	
Economically Disadvantaged	≥18,070	737.07	41.47	0.71	≥18,070	732.26	31.23	0.76
Education Classification								
Regular	≥23,790	750.52	42.69		≥23,780	742.68	32.81	
Special	≥3,410	721.34	41.18	0.69	≥3,410	723.49	32.93	0.58
English Learner Status								
Not English Learner	≥26,260	748.09	43.28		≥26,290	740.87	33.37	
English Learner	≥950	712.66	37.81	0.82	≥900	722.89	30.72	0.54
Section 504 Status								
Not Section 504	≥25,130	748.03	43.77		≥25,120	741.09	33.60	
Section 504	≥2,070	732.63	38.61	0.36	≥2,070	730.38	29.64	0.32
Migrant Status								
Not Migrant	≥27,150	746.91	43.57		≥27,140	740.30	33.43	
Migrant	≥50	719.75	42.78	0.62	≥50	728.67	33.00	0.35
Homeless Status								
Not Homeless	≥26,610	747.47	43.44		≥26,600	740.78	33.32	
Homeless	≥590	719.45	41.10	0.65	≥590	717.67	30.49	0.69
Foster Care Status								
Not in Foster Care	≥27,090	746.92	43.58		≥27,090	740.33	33.43	
Foster Care	≥110	730.49	41.65	0.38	≥110	727.23	33.03	0.39
Military Affiliation								
Not Military Affiliated	≥26,480	746.41	43.58		≥26,470	739.92	33.44	
Military Affiliated	≥720	762.96	40.80	-0.38	≥720	753.01	30.65	-0.39

Table 10.21 Impact Analysis, Grade 4 Computer-Based Test Administration

	ELA				Mathematics			
		Scale Score				Scale Score		
Group	N	Mean	Std. Dev.	Effect Size	N	Mean	Std. Dev.	Effect Size
All Students	≥48,880	743.87	35.83		≥48,880	738.73	33.53	
Gender								
Male	≥25,010	741.36	35.57		≥25,010	740.66	34.08	
Female	≥23,870	746.51	35.91	-0.14	≥23,860	736.71	32.81	0.12
Ethnicity								
White	≥20,550	756.35	34.32		≥20,550	751.70	31.52	
Hispanic/Latino	≥5,150	733.81	37.08	0.65	≥5,160	732.92	33.59	0.59
American Indian or Alaska Native	≥280	748.60	33.41	0.23	≥280	743.30	30.72	0.27
Asian	≥820	767.66	37.96	-0.33	≥820	768.52	34.35	-0.53
Black or African American	≥20,120	732.06	31.98	0.73	≥20,110	725.28	29.41	0.87
Native Hawaiian or Other Pacific	≥40	750.63	41.62	0.17	≥40	747.48	33.33	0.13
Two or More Races	≥1,880	750.37	34.22	0.17	≥1,870	742.98	32.31	0.28
Economic Status								
Not Economically Disadvantaged	≥14,820	761.56	34.89		≥14,820	756.03	32.48	
Economically Disadvantaged	≥34,060	736.18	33.43	0.75	≥34,050	731.20	31.10	0.79
Education Classification								
Regular	≥42,850	747.26	35.03		≥42,840	741.28	33.23	
Special	≥6,030	719.83	32.07	0.79	≥6,030	720.64	29.83	0.63
English Learner Status								
Not English Learner	≥46,190	745.73	35.37		≥46,190	739.95	33.38	
English Learner	≥2,690	711.98	27.73	0.96	≥2,690	717.80	28.80	0.67
Section 504 Status								
Not Section 504	≥44,560	745.11	36.01		≥44,560	739.72	33.65	
Section 504	≥4,320	731.17	31.19	0.39	≥4,320	728.54	30.36	0.34
Migrant Status								
Not Migrant	≥48,820	743.89	35.82		≥48,810	738.74	33.52	
Migrant	≥60	734.00	39.33	0.28	≥60	735.80	36.01	0.09
Homeless Status								
Not Homeless	≥47,730	744.31	35.82		≥47,720	739.18	33.48	
Homeless	≥1,150	725.78	31.20	0.52	≥1,150	720.11	29.90	0.57
Foster Care Status								
Not in Foster Care	≥48,710	743.92	35.84		≥48,700	738.77	33.54	
Foster Care	≥170	731.93	31.58	0.33	≥170	728.56	28.78	0.30
Military Affiliation								
Not Military Affiliated	≥47,970	743.56	35.81		≥47,960	738.45	33.52	
Military Affiliated	≥910	760.31	32.82	-0.47	≥910	753.50	30.24	-0.45

Table 10.22 Impact Analysis, Grade 5 Computer-Based Test Administration

	ELA				Mathematics			
		Scale Score				Scale Score		
Group	N	Mean	Std. Dev.	Effect Size	N	Mean	Std. Dev.	Effect Size
All Students	≥48,310	740.67	31.85		≥48,270	734.23	31.21	
Gender								
Male	≥24,720	736.82	31.81		≥24,700	734.29	32.10	
Female	≥23,590	744.70	31.38	-0.25	≥23,570	734.18	30.24	0.00
Ethnicity								
White	≥20,030	750.45	29.78		≥20,010	745.85	29.33	
Hispanic/Latino	≥5,180	732.35	34.26	0.59	≥5,170	728.98	31.16	0.57
American Indian or Alaska Native	≥250	746.43	30.39	0.14	≥250	739.05	28.47	0.23
Asian	≥800	765.70	34.63	-0.51	≥800	763.20	32.28	-0.59
Black or African American	≥20,300	731.57	29.65	0.64	≥20,300	722.56	27.97	0.81
Native Hawaiian or Other Pacific	≥30	751.62	32.82	-0.04	≥30	747.26	25.85	-0.05
Two or More Races	≥1,680	746.54	30.17	0.13	≥1,670	738.43	29.69	0.25
Economic Status								
Not Economically Disadvantaged	≥14,950	754.75	30.31		≥14,900	749.75	30.25	
Economically Disadvantaged	≥33,360	734.36	30.47	0.67	≥33,370	727.30	29.06	0.76
Education Classification								
Regular	≥42,640	744.12	30.61		≥42,600	736.76	31.08	
Special	≥5,670	714.77	28.91	0.96	≥5,670	715.27	25.00	0.71
English Learner Status								
Not English Learner	≥46,080	742.27	31.22		≥46,040	735.37	31.01	
English Learner	≥2,230	707.69	26.16	1.12	≥2,230	710.82	25.50	0.80
Section 504 Status								
Not Section 504	≥43,580	741.79	31.98		≥43,550	735.18	31.36	
Section 504	≥4,730	730.35	28.62	0.36	≥4,720	725.51	28.33	0.31
Migrant Status								
Not Migrant	≥48,250	740.68	31.83		≥48,210	734.24	31.20	
Migrant	≥60	729.65	38.64	0.35	≥60	730.61	34.25	0.12
Homeless Status								
Not Homeless	≥47,200	740.99	31.81		≥47,160	734.62	31.18	
Homeless	≥1,110	727.08	30.24	0.44	≥1,110	718.00	27.64	0.53
Foster Care Status								
Not in Foster Care	≥48,170	740.70	31.84		≥48,120	734.27	31.21	
Foster Care	≥140	731.54	31.64	0.29	≥150	723.91	29.33	0.33
Military Affiliation								
Not Military Affiliated	≥47,340	740.36	31.81		≥47,300	733.93	31.20	
Military Affiliated	≥970	755.84	29.98	-0.49	≥970	748.99	27.97	-0.48

Table 10.23 Impact Analysis, Grade 6 Computer-Based Test Administration

	ELA				Mathematics			
		Scale Score				Scale Score		
Group	N	Mean	Std. Dev.	Effect Size	N	Mean	Std. Dev.	Effect Size
All Students	≥43,380	740.67	32.49		≥48,350	732.52	31.58	
Gender								
Male	≥22,230	737.17	32.39		≥24,750	732.42	32.08	
Female	≥21,140	744.35	32.19	-0.22	≥23,600	732.63	31.04	-0.01
Ethnicity								
White	≥17,920	752.07	31.03		≥20,230	745.37	29.71	
Hispanic/Latino	≥4,710	732.60	33.32	0.62	≥5,000	725.97	31.60	0.64
American Indian or Alaska Native	≥200	742.43	29.97	0.31	≥250	734.97	28.73	0.35
Asian	≥700	768.78	35.78	-0.54	≥730	764.63	33.32	-0.65
Black or African American	≥18,290	730.00	28.96	0.74	≥20,390	719.81	27.19	0.90
Native Hawaiian or Other Pacific	≥40	736.18	39.03	0.51	≥40	733.03	34.73	0.42
Two or More Races	≥1,490	746.88	31.46	0.17	≥1,670	737.42	30.43	0.27
Economic Status								
Not Economically Disadvantaged	≥13,790	756.21	32.10		≥15,320	748.70	31.04	
Economically Disadvantaged	≥29,590	733.43	30.03	0.74	≥33,030	725.01	28.89	0.80
Education Classification								
Regular	≥38,560	743.82	31.66		≥42,940	735.28	31.13	
Special	≥4,820	715.48	27.69	0.91	≥5,410	710.69	26.13	0.80
English Learner Status								
Not English Learner	≥41,490	742.20	32.04		≥46,380	733.69	31.29	
English Learner	≥1,890	707.02	22.60	1.11	≥1,970	705.18	25.23	0.92
Section 504 Status								
Not Section 504	≥38,810	741.83	32.70		≥43,330	733.58	31.75	
Section 504	≥4,570	730.84	28.80	0.34	≥5,020	723.43	28.42	0.32
Migrant Status								
Not Migrant	≥43,310	740.69	32.49		≥48,290	732.54	31.58	
Migrant	≥60	729.70	32.85	0.34	≥60	720.91	29.55	0.37
Homeless Status								
Not Homeless	≥42,350	741.05	32.46		≥47,300	732.87	31.57	
Homeless	≥1,020	724.97	29.81	0.50	≥1,050	717.02	27.50	0.50
Foster Care Status								
Not in Foster Care	≥43,250	740.73	32.48		≥48,220	732.57	31.57	
Foster Care	≥120	721.10	27.43	0.60	≥130	716.04	28.22	0.52
Military Affiliation								
Not Military Affiliated	≥42,480	740.32	32.43		≥47,430	732.23	31.57	
Military Affiliated	≥890	757.36	30.76	-0.53	≥920	747.49	27.95	-0.48

Table 10.24 Impact Analysis, Grade 7 Computer-Based Test Administration

	ELA				Mathematics			
		Scale Score				Scale Score		
Group	N	Mean	Std. Dev.	Effect Size	N	Mean	Std. Dev.	Effect Size
All Students	≥42,460	742.88	38.84		≥48,920	732.60	27.42	
Gender								
Male	≥21,830	736.23	38.60		≥25,220	732.71	28.16	
Female	≥20,630	749.92	37.83	-0.36	≥23,700	732.48	26.60	0.01
Ethnicity								
White	≥17,490	754.52	37.72		≥20,400	743.26	26.89	
Hispanic/Latino	≥4,860	732.67	40.78	0.57	≥5,230	727.20	26.31	0.60
American Indian or Alaska Native	≥210	742.31	34.75	0.32	≥270	735.72	24.69	0.28
Asian	≥710	774.61	43.17	-0.53	≥790	760.92	33.30	-0.65
Black or African American	≥17,660	732.37	34.98	0.61	≥20,490	722.00	22.84	0.85
Native Hawaiian or Other Pacific	≥30	760.45	41.19	-0.16	≥30	746.36	28.49	-0.12
Two or More Races	≥1,460	749.03	38.71	0.15	≥1,670	735.32	26.66	0.30
Economic Status								
Not Economically Disadvantaged	≥13,730	759.91	38.24		≥15,640	746.56	28.09	
Economically Disadvantaged	≥28,730	734.75	36.41	0.68	≥33,280	726.04	24.49	0.80
Education Classification								
Regular	≥38,030	746.70	37.76		≥43,750	734.83	27.21	
Special	≥4,430	710.18	31.96	0.98	≥5,170	713.71	21.21	0.79
English Learner Status								
Not English Learner	≥40,460	744.91	38.17		≥46,800	733.56	27.34	
English Learner	≥2,000	701.98	28.00	1.14	≥2,120	711.46	19.44	0.82
Section 504 Status								
Not Section 504	≥37,880	744.65	39.00		≥43,690	733.69	27.60	
Section 504	≥4,580	728.27	34.15	0.43	≥5,230	723.50	24.00	0.37
Migrant Status								
Not Migrant	≥42,400	742.90	38.82		≥48,850	732.61	27.42	
Migrant	≥60	731.59	46.57	0.29	≥70	727.24	24.65	0.20
Homeless Status								
Not Homeless	≥41,480	743.31	38.80		≥47,900	732.90	27.43	
Homeless	≥980	724.65	35.75	0.48	≥1,020	718.56	22.83	0.52
Foster Care Status								
Not in Foster Care	≥42,340	742.93	38.83		≥48,780	732.64	27.41	
Foster Care	≥120	726.06	35.85	0.43	≥140	717.41	24.64	0.56
Military Affiliation								
Not Military Affiliated	≥41,610	742.48	38.77		≥48,050	732.35	27.36	
Military Affiliated	≥850	762.69	36.82	-0.52	≥870	746.24	27.37	-0.51

Table 10.25 Impact Analysis, Grade 8 Computer-Based Test Administration

	ELA				Mathematics			
		Scale Score				Scale Score		
Group	N	Mean	Std. Dev.	Effect Size	N	Mean	Std. Dev.	Effect Size
All Students	≥45,790	746.09	38.84		≥44,250	725.74	33.93	
Gender								
Male	≥23,150	739.77	38.58		≥22,470	724.41	34.45	
Female	≥22,630	752.55	38.03	-0.33	≥21,770	727.11	33.32	-0.08
Ethnicity								
White	≥18,780	758.51	37.23		≥17,240	738.86	34.06	
Hispanic/Latino	≥4,830	735.86	41.71	0.59	≥4,520	719.02	32.40	0.59
American Indian or Alaska Native	≥210	747.53	36.18	0.30	≥260	729.09	30.44	0.29
Asian	≥720	778.51	38.40	-0.54	≥500	758.45	41.50	-0.57
Black or African American	≥19,590	734.89	35.20	0.65	≥20,190	714.91	29.21	0.76
Native Hawaiian or Other Pacific	≥40	762.27	29.80	-0.10	≥30	747.43	33.05	-0.25
Two or More Races	≥1,590	753.08	37.56	0.15	≥1,480	728.84	33.66	0.29
Economic Status								
Not Economically Disadvantaged	≥14,980	762.48	37.63		≥12,970	740.52	35.64	
Economically Disadvantaged	≥30,800	738.11	36.86	0.66	≥31,270	719.61	31.20	0.64
Education Classification								
Regular	≥41,150	750.02	37.56		≥39,210	728.66	33.67	
Special	≥4,630	711.18	31.88	1.05	≥5,030	702.96	26.43	0.78
English Learner Status								
Not English Learner	≥43,890	747.91	38.11		≥42,320	726.77	33.88	
English Learner	≥1,890	703.96	30.77	1.16	≥1,920	703.11	26.04	0.70
Section 504 Status								
Not Section 504	≥40,870	747.82	38.86		≥39,150	726.95	34.14	
Section 504	≥4,910	731.73	35.51	0.42	≥5,090	716.44	30.67	0.31
Migrant Status								
Not Migrant	≥45,730	746.10	38.84		≥44,190	725.74	33.93	
Migrant	≥50	737.90	38.76	0.21	≥50	721.52	32.84	0.12
Homeless Status								
Not Homeless	≥44,810	746.50	38.76		≥43,270	726.07	33.93	
Homeless	≥970	727.12	37.54	0.50	≥980	711.10	30.27	0.44
Foster Care Status								
Not in Foster Care	≥45,640	746.15	38.84		≥44,090	725.79	33.93	
Foster Care	≥140	726.49	34.57	0.51	≥150	711.19	29.21	0.43
Military Affiliation								
Not Military Affiliated	≥44,920	745.74	38.83		≥43,560	725.45	33.86	
Military Affiliated	≥860	764.21	34.67	-0.48	≥690	744.05	33.03	-0.55

Additional data for mean scale scores are provided in Tables 10.26 and 10.27. These tables report the number of students, mean scale scores, and standard deviations for special education classification. Groups that have fewer than 50 students are NR. The analyses were based on census data.

Table 10.26 Special Education Classification Scale-Score Means and Standard Deviations: English Language Arts

Special Education Classification Scale-Score Means and Standard Deviations: English Language Arts							
Grade	Group	Yes			No		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.
3	Gifted	≥700	811.97	27.26	≥48,620	738.80	44.06
	Talented	≥630	773.39	38.83	≥48,690	739.41	44.62
	Autism	≥460	698.65	36.72	≥48,870	740.23	44.61
	Deaf-Blindness	<50	NR	NR	≥49,330	739.85	44.72
	Developmental Delay	≥780	700.69	34.52	≥48,540	740.48	44.58
	Emotional Disturbance	≥60	718.44	38.80	≥49,270	739.87	44.72
	HI—Deaf	<50	NR	NR	≥49,310	739.85	44.72
	HI—Hard-of-Hearing	<50	NR	NR	≥49,280	739.87	44.72
	Mild Mental Disability	≥330	683.49	23.03	≥49,000	740.23	44.59
	Moderate Mental Disability	<50	NR	NR	≥49,320	739.86	44.71
	Orthopedic Impairment	<50	NR	NR	≥49,290	739.86	44.72
	Other Health Impairment	≥590	705.51	35.77	≥48,740	740.26	44.65
	Specific Learning Disability	≥2,120	706.46	31.21	≥47,210	741.35	44.65
	Speech or Language Impairment	≥1,720	742.24	43.34	≥47,600	739.76	44.77
	Traumatic Brain Injury	<50	NR	NR	≥49,320	739.85	44.72
	Visual Impairment	<50	NR	NR	≥49,300	739.84	44.72
	Other	<50	NR	NR	≥49,320	739.85	44.72
	HI—Hearing Impairment	<50	NR	NR	≥49,330	739.85	44.72
	Unknown	<50	NR	NR	≥49,330	739.85	44.72

Special Education Classification Scale-Score Means and Standard Deviations: English Language Arts							
Grade	Group	Yes			No		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.
4	Gifted	≥830	802.85	25.10	≥48,040	742.85	35.12
	Talented	≥1,000	771.81	31.25	≥47,870	743.29	35.68
	Autism	≥450	714.24	34.13	≥48,430	744.15	35.73
	Deaf-Blindness	<50	NR	NR	≥48,880	743.87	35.83
	Developmental Delay	<50	NR	NR	≥48,850	743.90	35.82
	Emotional Disturbance	≥82	713.28	28.57	≥48,800	743.93	35.82
	HI—Deaf	≥17	NR	NR	≥48,870	743.88	35.83
	HI—Hard-of-Hearing	≥50	729.40	35.78	≥48,830	743.89	35.83
	Mild Mental Disability	≥460	696.24	18.33	≥48,410	744.34	35.65
	Moderate Mental Disability	<50	NR	NR	≥48,880	743.88	35.83
	Orthopedic Impairment	<50	NR	NR	≥48,840	743.89	35.83
	Other Health Impairment	≥790	715.94	29.90	≥48,090	744.33	35.74
	Specific Learning Disability	≥2,760	713.44	24.49	≥46,120	745.70	35.58
	Speech or Language Impairment	≥1,290	746.50	34.76	≥47,590	743.80	35.86
	Traumatic Brain Injury	<50	NR	NR	≥48,880	743.88	35.83
	Visual Impairment	<50	NR	NR	≥48,860	743.88	35.83
	Other	<50	NR	NR	≥48,880	743.88	35.82
	HI—Hearing Impairment	<50	NR	NR	≥48,880	743.87	35.83
	Unknown	<50	NR	NR	≥48,880	743.87	35.83
5	Gifted	≥1,000	787.43	22.14	≥47,310	739.68	31.27
	Talented	≥1,270	765.55	27.60	≥47,040	740.00	31.68
	Autism	≥390	711.46	32.54	≥47,920	740.91	31.73
	Deaf-Blindness	<50	NR	NR	≥48,310	740.67	31.85
	Developmental Delay	<50	NR	NR	≥48,310	740.67	31.85
	Emotional Disturbance	≥80	719.28	32.50	≥48,230	740.71	31.83
	HI—Deaf	<50	NR	NR	≥48,290	740.68	31.83
	HI—Hard-of-Hearing	≥50	719.11	29.16	≥48,260	740.69	31.84
	Mild Mental Disability	≥420	690.53	18.61	≥47,890	741.11	31.59
	Moderate Mental Disability	<50	NR	NR	≥48,310	740.68	31.84
	Orthopedic Impairment	≥50	730.43	31.41	≥48,260	740.68	31.85
	Other Health Impairment	≥930	711.60	27.94	≥47,380	741.24	31.65
	Specific Learning Disability	≥2,770	712.20	23.84	≥45,540	742.40	31.45
	Speech or Language Impairment	≥880	737.50	30.76	≥47,430	740.73	31.86
	Traumatic Brain Injury	<50	NR	NR	≥48,310	740.67	31.84
	Visual Impairment	<50	NR	NR	≥48,280	740.67	31.85
	Other	<50	NR	NR	≥48,310	740.67	31.85
	HI—Hearing Impairment	<50	NR	NR	≥48,310	740.67	31.85
	Unknown	<50	NR	NR	≥48,310	740.67	31.85

Special Education Classification Scale-Score Means and Standard Deviations: English Language Arts								
Grade	Group	Yes			No			
		N	Mean	Std. Dev.	N	Mean	Std. Dev.
6	Gifted	≥1,010	793.57	23.96	≥42,360	739.40	31.60
	Talented	≥1,330	767.32	29.19	≥42,040	739.82	32.23
	Autism	≥300	717.78	32.23	≥43,070	740.83	32.43
	Deaf-Blindness	<50	NR	NR	≥43,380	740.67	32.49
	Developmental Delay	<50	NR	NR	≥43,380	740.67	32.49
	Emotional Disturbance	≥110	712.18	30.34	≥43,270	740.74	32.46
	HI—Deaf	<50	NR	NR	≥43,360	740.68	32.49
	HI—Hard-of-Hearing	≥50	726.75	33.97	≥43,330	740.69	32.48
	Mild Mental Disability	≥260	693.50	16.19	≥43,110	740.96	32.35
	Moderate Mental Disability	<50	NR	NR	≥43,370	740.67	32.49
	Orthopedic Impairment	<50	NR	NR	≥43,350	740.68	32.49
	Other Health Impairment	≥910	714.17	25.95	≥42,470	741.24	32.38
	Specific Learning Disability	≥2,490	711.57	22.51	≥40,880	742.45	32.16
	Speech or Language Impairment	≥590	741.04	32.57	≥42,790	740.67	32.49
	Traumatic Brain Injury	<50	NR	NR	≥43,370	740.67	32.49
	Visual Impairment	<50	NR	NR	≥43,340	740.68	32.49
	Other	<50	NR	NR	≥43,380	740.67	32.49
	HI—Hearing Impairment	<50	NR	NR	≥43,380	740.67	32.49
	Unknown	<50	NR	NR	≥43,380	740.67	32.49
7	Gifted	≥1,000	802.53	28.11	≥41,460	741.44	37.92
	Talented	≥1,300	776.55	34.66	≥41,160	741.81	38.48
	Autism	≥250	718.10	37.66	≥42,210	743.03	38.80
	Deaf-Blindness	<50	NR	NR	≥42,460	742.88	38.84
	Developmental Delay	<50	NR	NR	≥42,460	742.88	38.84
	Emotional Disturbance	≥110	709.26	32.27	≥42,350	742.97	38.81
	HI—Deaf	<50	NR	NR	≥42,450	742.90	38.83
	HI—Hard-of-Hearing	<50	NR	NR	≥42,430	742.89	38.84
	Mild Mental Disability	≥220	686.30	18.45	≥42,240	743.18	38.70
	Moderate Mental Disability	<50	NR	NR	≥42,460	742.89	38.83
	Orthopedic Impairment	<50	NR	NR	≥42,430	742.89	38.84
	Other Health Impairment	≥910	707.92	31.75	≥41,550	743.65	38.62
	Specific Learning Disability	≥2,370	706.34	26.99	≥40,090	745.04	38.35
	Speech or Language Impairment	≥420	739.07	36.31	≥42,040	742.92	38.86
	Traumatic Brain Injury	<50	NR	NR	≥42,450	742.89	38.84
	Visual Impairment	<50	NR	NR	≥42,440	742.88	38.84
	Other	<50	NR	NR	≥42,460	742.88	38.84
	HI—Hearing Impairment	<50	NR	NR	≥42,460	742.88	38.84
	Unknown	<50	NR	NR	≥42,460	742.88	38.84

Special Education Classification Scale-Score Means and Standard Deviations: English Language Arts							
Grade	Group	Yes			No		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.
8	Gifted	≥1,030	801.86	28.98	≥44,750	744.80	38.08
	Talented	≥1,550	777.10	32.65	≥44,230	745.00	38.59
	Autism	≥240	720.62	38.43	≥45,540	746.23	38.79
	Deaf-Blindness	<50	NR	NR	≥45,790	746.09	38.84
	Developmental Delay	<50	NR	NR	≥45,790	746.09	38.84
	Emotional Disturbance	≥160	708.27	33.01	≥45,620	746.23	38.79
	HI—Deaf	<50	NR	NR	≥45,760	746.11	38.83
	HI—Hard-of-Hearing	<50	NR	NR	≥45,740	746.11	38.83
	Mild Mental Disability	≥180	687.35	20.70	≥45,600	746.33	38.71
	Moderate Mental Disability	<50	NR	NR	≥45,790	746.09	38.84
	Orthopedic Impairment	<50	NR	NR	≥45,740	746.11	38.83
	Other Health Impairment	≥1,000	711.46	33.40	≥44,780	746.87	38.60
	Specific Learning Disability	≥2,590	708.43	27.79	≥43,190	748.35	38.24
	Speech or Language Impairment	≥260	739.78	36.32	≥45,530	746.13	38.85
	Traumatic Brain Injury	<50	NR	NR	≥45,780	746.10	38.84
	Visual Impairment	<50	NR	NR	≥45,750	746.11	38.83
	Other	<50	NR	NR	≥45,790	746.09	38.84
	HI—Hearing Impairment	<50	NR	NR	≥45,790	746.09	38.84
	Unknown	<50	NR	NR	≥45,790	746.09	38.84

Table 10.27 Special Education Classification Scale-Score Means and Standard Deviations: Mathematics

Special Education Classification Scale-Score Means and Standard Deviations: Mathematics							
Grade	Group	Yes			No		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.
3	Gifted	≥700	794.79	22.50	≥48,670	736.33	32.33
	Talented	≥630	760.07	28.29	≥48,740	736.87	32.90
	Autism	≥460	710.12	34.29	≥48,920	737.43	32.83
	Deaf-Blindness	<50	NR	NR	≥49,380	737.17	32.95
	Developmental Delay	≥790	709.37	28.48	≥48,590	737.62	32.82
	Emotional Disturbance	≥60	720.82	31.57	≥49,320	737.19	32.94
	HI—Deaf	<50	NR	NR	≥49,360	737.18	32.95
	HI—Hard-of-Hearing	<50	NR	NR	≥49,330	737.19	32.95
	Mild Mental Disability	≥330	694.08	18.99	≥49,050	737.46	32.83
	Moderate Mental Disability	<50	NR	NR	≥49,370	737.18	32.94
	Orthopedic Impairment	<50	NR	NR	≥49,330	737.19	32.94
	Other Health Impairment	≥590	711.34	28.99	≥48,790	737.48	32.87
	Specific Learning Disability	≥2,120	714.39	23.60	≥47,260	738.19	32.94
	Speech or Language Impairment	≥1,730	742.46	32.13	≥47,650	736.98	32.96
	Traumatic Brain Injury	<50	NR	NR	≥49,370	737.18	32.95
	Visual Impairment	<50	NR	NR	≥49,350	737.17	32.95
	Other	<50	NR	NR	≥49,370	737.17	32.95
	HI—Hearing Impairment	<50	NR	NR	≥49,380	737.17	32.95
	Unknown	<50	NR	NR	≥49,380	737.17	32.95
4	Gifted	≥830	795.56	22.97	≥48,040	737.74	32.82
	Talented	≥1,000	760.93	30.59	≥47,870	738.27	33.43
	Autism	≥450	719.95	33.56	≥48,420	738.91	33.48
	Deaf-Blindness	<50	NR	NR	≥48,880	738.73	33.53
	Developmental Delay	<50	NR	NR	≥48,840	738.75	33.52
	Emotional Disturbance	≥80	713.40	28.42	≥48,790	738.78	33.52
	HI—Deaf	<50	NR	NR	≥48,860	738.73	33.53
	HI—Hard-of-Hearing	≥50	728.58	30.84	≥48,830	738.74	33.53
	Mild Mental Disability	≥460	697.78	16.72	≥48,410	739.13	33.40
	Moderate Mental Disability	<50	NR	NR	≥48,870	738.74	33.52
	Orthopedic Impairment	<50	NR	NR	≥48,830	738.74	33.52
	Other Health Impairment	≥790	716.92	28.24	≥48,090	739.09	33.49
	Specific Learning Disability	≥2,750	714.94	22.53	≥46,120	740.16	33.54
	Speech or Language Impairment	≥1,290	743.67	32.99	≥47,580	738.60	33.53
	Traumatic Brain Injury	<50	NR	NR	≥48,870	738.74	33.53
	Visual Impairment	<50	NR	NR	≥48,850	738.74	33.52
	Other	<50	NR	NR	≥48,870	738.74	33.52
	HI—Hearing Impairment	<50	NR	NR	≥48,880	738.73	33.53
	Unknown	<50	NR	NR	≥48,880	738.73	33.53

Special Education Classification Scale-Score Means and Standard Deviations: Mathematics							
Grade	Group	Yes			No		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.
5	Gifted	≥990	786.46	22.12	≥47,280	733.14	30.42
	Talented	≥1,250	756.26	27.45	≥47,010	733.64	31.09
	Autism	≥390	715.43	30.39	≥47,880	734.39	31.16
	Deaf-Blindness	<50	NR	NR	≥48,270	734.23	31.20
	Developmental Delay	<50	NR	NR	≥48,270	734.23	31.21
	Emotional Disturbance	≥80	718.79	29.26	≥48,190	734.26	31.20
	HI—Deaf	<50	NR	NR	≥48,250	734.24	31.20
	HI—Hard-of-Hearing	≥50	725.29	22.75	≥48,220	734.24	31.21
	Mild Mental Disability	≥420	697.40	15.77	≥47,850	734.56	31.11
	Moderate Mental Disability	<50	NR	NR	≥48,260	734.24	31.20
	Orthopedic Impairment	≥50	726.87	27.02	≥48,220	734.24	31.21
	Other Health Impairment	≥930	712.41	23.20	≥47,340	734.66	31.19
	Specific Learning Disability	≥2,770	712.55	19.92	≥45,500	735.55	31.28
	Speech or Language Impairment	≥870	733.58	30.29	≥47,390	734.25	31.22
	Traumatic Brain Injury	<50	NR	NR	≥48,260	734.24	31.21
	Visual Impairment	<50	NR	NR	≥48,240	734.24	31.21
	Other	<50	NR	NR	≥48,270	734.24	31.20
	HI—Hearing Impairment	<50	NR	NR	≥48,270	734.23	31.21
	Unknown	<50	NR	NR	≥48,270	734.23	31.21
6	Gifted	≥1,080	786.28	22.73	≥47,270	731.29	30.66
	Talented	≥1,490	755.21	28.44	≥46,860	731.80	31.40
	Autism	≥320	716.09	29.92	≥48,030	732.64	31.56
	Deaf-Blindness	<50	NR	NR	≥48,350	732.52	31.58
	Developmental Delay	<50	NR	NR	≥48,350	732.52	31.58
	Emotional Disturbance	≥110	707.82	27.22	≥48,240	732.58	31.56
	HI—Deaf	<50	NR	NR	≥48,340	732.53	31.58
	HI—Hard-of-Hearing	≥50	723.82	36.58	≥48,300	732.53	31.57
	Mild Mental Disability	≥290	692.42	14.95	≥48,060	732.77	31.49
	Moderate Mental Disability	<50	NR	NR	≥48,350	732.53	31.57
	Orthopedic Impairment	<50	NR	NR	≥48,320	732.53	31.58
	Other Health Impairment	≥1,000	709.14	24.72	≥47,350	733.02	31.52
	Specific Learning Disability	≥2,850	706.63	21.21	≥45,500	734.15	31.41
	Speech or Language Impairment	≥650	734.64	31.32	≥47,700	732.49	31.58
	Traumatic Brain Injury	<50	NR	NR	≥48,350	732.53	31.57
	Visual Impairment	<50	NR	NR	≥48,320	732.53	31.58
	Other	<50	NR	NR	≥48,350	732.53	31.58
	HI—Hearing Impairment	<50	NR	NR	≥48,350	732.52	31.58
	Unknown	<50	NR	NR	≥48,350	732.52	31.58

Special Education Classification Scale-Score Means and Standard Deviations: Mathematics							
Grade	Group	Yes			No		
		N	Mean	Std. Dev.	N	Mean	Std. Dev.
7	Gifted	≥1,110	782.71	23.36	≥47,800	731.43	26.39
	Talented	≥1,560	751.64	25.83	≥47,350	731.97	27.24
	Autism	≥280	723.84	25.69	≥48,640	732.65	27.42
	Deaf-Blindness	<50	NR	NR	≥48,920	732.60	27.42
	Developmental Delay	<50	NR	NR	≥48,920	732.60	27.42
	Emotional Disturbance	≥120	711.74	20.19	≥48,790	732.65	27.41
	HI—Deaf	<50	NR	NR	≥48,900	732.61	27.42
	HI—Hard-of-Hearing	<50	NR	NR	≥48,880	732.61	27.42
	Mild Mental Disability	≥250	698.67	12.91	≥48,670	732.77	27.36
	Moderate Mental Disability	<50	NR	NR	≥48,920	732.60	27.42
	Orthopedic Impairment	<50	NR	NR	≥48,880	732.60	27.42
	Other Health Impairment	≥1,030	713.78	21.32	≥47,890	733.00	27.39
	Specific Learning Disability	≥2,830	710.41	17.15	≥46,090	733.96	27.34
	Speech or Language Impairment	≥480	731.95	27.47	≥48,440	732.61	27.42
	Traumatic Brain Injury	<50	NR	NR	≥48,910	732.61	27.42
	Visual Impairment	<50	NR	NR	≥48,890	732.60	27.42
	Other	<50	NR	NR	≥48,920	732.60	27.42
	HI—Hearing Impairment	<50	NR	NR	≥48,920	732.60	27.42
	Unknown	<50	NR	NR	≥48,920	732.60	27.42
8	Gifted	≥440	789.09	35.17	≥43,800	725.10	33.30
	Talented	≥1,250	748.96	34.08	≥43,000	725.06	33.68
	Autism	≥250	715.67	33.39	≥43,990	725.80	33.92
	Deaf-Blindness	<50	NR	NR	≥44,250	725.74	33.93
	Developmental Delay	<50	NR	NR	≥44,250	725.74	33.93
	Emotional Disturbance	≥170	700.55	30.33	≥44,080	725.83	33.90
	HI—Deaf	<50	NR	NR	≥44,220	725.75	33.93
	HI—Hard-of-Hearing	<50	NR	NR	≥44,200	725.74	33.93
	Mild Mental Disability	≥200	687.05	16.29	≥44,040	725.92	33.88
	Moderate Mental Disability	<50	NR	NR	≥44,240	725.74	33.92
	Orthopedic Impairment	<50	NR	NR	≥44,200	725.75	33.92
	Other Health Impairment	≥1,080	703.31	26.91	≥43,160	726.30	33.89
	Specific Learning Disability	≥2,900	700.75	23.36	≥41,350	727.49	33.86
	Speech or Language Impairment	≥250	721.27	33.27	≥43,990	725.76	33.93
	Traumatic Brain Injury	<50	NR	NR	≥44,230	725.74	33.92
	Visual Impairment	<50	NR	NR	≥44,210	725.75	33.93
	Other	<50	NR	NR	≥44,240	725.74	33.93
	HI—Hearing Impairment	<50	NR	NR	≥44,250	725.74	33.93
	Unknown	<50	NR	NR	≥44,250	725.74	33.93

10.5 Mode Effect Study

It is also important to evaluate fairness in test administration in addition to evaluating fairness by examining performance among subgroups. The 2023 LEAP 2025 ELA and mathematics tests were administered as both paper-based tests (PBTs) and computer-based tests (CBTs) for grade 3. The *Standards* indicate that results across different testing modes should be comparable. The mode comparability for the 2023 LEAP 2025 CBT and PBT in grade 3 was investigated using the following steps:

- The mode effect study was performed using the CBT as the focal group and the PBT as the reference group.
- The study was based on equivalent groups design. Equivalent PBT students that match CBT students were selected using propensity score matching (PSM).
- At the item level, DIF analysis was performed using the PSM samples.
- At the test level, ESs based on difference scores of scale scores between the CBT and the PBT were used to examine the mode effect.
- Similar to PARCC's decision to not apply a mode adjustment, the LDOE also decided to not apply any mode adjustment to the LEAP 2025.

10.5.1 Mode Study by Propensity Score Matching

The CBT was administered to a smaller number of students than the PBT in grade 3; therefore, the CBT was designated as the focal group for PSM (Rosenbaum & Rubin, 1983) and the PBT was considered the reference group. That is, all CBT students and their matching PBT students were selected using covariates (matching variables), such as the 2022 LEAP 2025 ELA and mathematics scale scores and the 2023 bio-demographic information, such as gender, ethnicity, economically disadvantaged, accommodations, and ELL. Only scale scores of the grade 3 students who took the 2022 PBT were used in this study as there are no LEAP 2025 grade 2 tests. Therefore, school means from the 2022 grade 3 tests were used to match with 2023 LEAP 2025 grade 3 school means.

Table 10.28 shows the number of equivalent CBT and PBT students matched by the PSM method. Only 2023 grade 3 students who have bio-demographic information of the spring 2023 administration and spring 2022 school means were included in the matching. Also, only 2023 students whose schools administered the 2022 PBT were included in the matching. In the spring 2023 administration, about 27,000 students took the CBT form and 23,000 students took the PBT form.

For mathematics, of the 26,590 PBT students, 6,576 were selected (a number equivalent to the number of CBT students) by considering all covariates. For ELA, of the 26,652 PBT students, 6,521 were selected by considering all covariates.

Table 10.28 Number of Students Used for Propensity Score Matching

Content	Grade	CBT	PBT	
		Total*	Total*	Selected
Mathematics	3	≥6,570	≥26,590	≥6,570
ELA	3	≥6,520	≥26,650	≥6,520

**Total: Number of students who have information for all covariates*

At the item level, DIF analysis was performed using the MH statistic by Holland and Thayer (1988). There were unique items in each ELA CBT and PBT forms, and these items were dropped from analysis. Table 10.29 shows the number of mode DIF items flagged using the same rules that are used in NAEP. For mathematics, there was one item each in C+ and B+. There were two items in B+ for ELA. The negative sign indicates the CBT item was more difficult than the same PBT item.

Table 10.29 2023 LEAP 2025 Mode DIF Statistics: Number of Flagged Items

			DIF			
Content	Grade	N of Items	-C	C	-B	B
Mathematics	3	43		1		1
ELA	3	28				2

Item raw scores of matched CBT and PBTY students were used, and their difference item scores were calculated. ESs of the difference item scores were calculated as follows:

ES = (CBT Mean – PBT Mean)/$\sqrt{(CBT\ VAR + PBT\ VAR)/2}$, where *VAR* = SD^2.

Table 10.30 (mathematics) and Table 10.31 (ELA) show the mean item scores and standard deviations for the CBT and PBT administrations for the flagged items. When a flag criterion of |0.2|, which can be considered a small difference criterion, was applied, two items were flagged for mathematics and four items were flagged for ELA. One item in grade 3 mathematics was flagged for both effect size and mode DIF; therefore, it was determined to have mode effect. This item was included in the calibration two times, one with PBT responses and one with CBT responses.

Table 10.30 Mode Study Scale Score Differences and Effect Size: Mathematics Grade 3

	PBT		CBT				
Item ID	Mean	Std. Dev.	Mean	Std. Dev.	Mean Diff PBS-CBT	ES	Flag>|0.2|
870678	2.25	2.04	1.90	1.91	0.35	0.24	YES
981741	1.53	1.10	1.14	1.04	0.39	0.38	YES

Table 10.31 Mode Study Scale Score Differences and Effect Size: ELA Grade 3

	PBT		CBT				
Item ID	Mean	Std. Dev.	Mean	Std. Dev.	Mean Diff PBT-CBT	ES	Flag>|0.2|
982123	1.57	0.77	1.28	0.92	0.29	0.32	YES
982079W2	0.86	0.82	0.67	0.73	0.19	0.21	YES
1113140W2	0.60	0.60	0.42	0.60	0.18	0.23	YES
1113140W3	0.72	0.59	0.47	0.65	0.25	0.32	YES

10.6 Summary

In summary, the overall purpose of this chapter is to address fairness concerns that are relevant to the administration of LEAP 2025 assessments. The information in this chapter addresses multiple best practices of the testing industry and is particularly related to the following standards:

> ***Standard 3.1*** Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population (63).
>
> ***Standard 3.2*** Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics (64).
>
> ***Standard 3.3*** Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test (64).
>
> ***Standard 3.4*** Test takers should receive comparable treatment during the test administration and scoring process (65).
>
> ***Standard 3.5*** Test developers should specify and document provisions that have been made to test administration and scoring procedures to remove construct-irrelevant barriers for all relevant subgroups in the test-taker population (65).
>
> ***Standard 3.6*** Where credible evidence indicates that test scores may differ in meaning for relevant subgroups in the intended examinee population, test developers and/or users are responsible for examining the evidence for validity of score interpretations for intended uses for individuals from those subgroups. What constitutes a significant difference in subgroup scores and what actions are taken in response to such differences may be defined by applicable laws (65).
>
> ***Standard 3.16*** When credible research indicates that test scores for some relevant subgroups are differentially affected by construct-irrelevant characteristics of the test or of the examinees, when legally permissible, test users should use the test only for those subgroups for which there is sufficient evidence of validity to support score interpretations for the intended uses (70).

Appendix A—Accommodated Print and Braille Creation

Guidelines for Accommodated Print and Braille

Louisiana believes that all students requiring test accommodations should be presented with the same rigor as students taking tests without accommodations. To ensure this, Louisiana creates accommodated versions of the operational test form for each test administration, allowing all students to take the same items regardless of the need for an accommodated presentation. Careful consideration is given to all items that are used for Louisiana assessments for their ability to be faithfully represented in accommodated print (AP) and braille formats. Fairness for all populations, item integrity, and student-item interaction for technology-enhanced (TE) items are all factors when selecting the items that will appear on a Louisiana form. TE items are modified so that students who interact with an item on an AP or braille form will have a similar and equivalent experience to students who interact with that same item in the online environment. This maintains both the rigor and the content being assessed. Some examples of the modification process are provided below.

- Drag-and-drop items in the online environment require a student to place the answer options in an interactive table. For the AP and braille forms, the student is presented with a table with the same information as the interactive table (column or row headers, any completed cells, and blank spaces) and the answer options are listed below the table (similar to the online form in which the options are listed either below or to the right of the table). The directions are modified to ask the student to write the letter or number of the correct answer in its corresponding box. Students are also able to circle the text and draw arrows to indicate where it should be placed or add labels to the answer choices and write only the label in the box, as long as the intended response is clear to the test administrator who will transcribe the answers into the online system.
- Match interaction items in the online environment require a student to select a checkbox in one or more columns for each of multiple rows. In the AP and braille forms, the student is provided with a table and asked to mark or select the correct answer in each row.
- Highlight-text items or item parts in the online environment require a student to click on the selected text, which highlights the selected word, phrase, or sentence. In the AP and braille forms, the text is presented in the same format and the student is asked to circle the answer. Where only certain words or phrases are selectable in the online system, those options are underlined in the AP and braille forms to indicate which words and/or phrases the student should select from.
- Drop-down menu items in the online environment have answer options in a drop-down menu format, oftentimes as part of a complete sentence. The AP and braille forms display the item with a blank line in place of the drop-down menu in the sentence, with all the answer options for the drop-down menu presented vertically below the sentence and lettered or numbered. The directions are then modified to ask the student to select the letter/number of the word/phrase that belongs in the blank.
- Short answer items in the online environment require a student to type the answer in a box. In the AP and braille forms, a box is provided for the student to write the response.
- Keypad input items in the online environment require a student to enter a numeric response including all rational and irrational numbers as well as expressions and equations. In the AP forms, a box is provided for the student to write the response. In the braille forms, students are asked to answer on the paper provided.
- Graphing items, including coordinate planes, number lines, line plots, and bar graphs, in the online environment require a student to complete a graph by plotting points, adding Xs to create a line plot, or raising/lowering bars to create a bar graph or histogram. In the AP and braille forms, the student is

provided with the same coordinate plane, number line, line plot, or bar graph as in the online item, including titles, axis labels, and keys, and is asked to complete the graph.

Displaying items similarly in accommodated print and braille forms and in the online environment (and allowing students to interact with the items in a similar manner) maintains item integrity by assessing a similar construct in a similar manner regardless of how a student encounters an item. This provides students who are unable to access the assessment online with an assessment at the same level of rigor as the online test.

AP forms are thoroughly reviewed by DRC and LDOE content experts alongside the online form, and braille forms are reviewed by an outside third-party braille expert against the AP form. Throughout the braille creation process, the braille vendor relies on the AP form and consults with the content experts at LDOE for additional clarification or modifications for specific items as needed. Students' responses to the accommodated print or braille test are captured in the same online test as used by the general population, either through use of a scribe or by themselves if able. This ensures a valid and reliable assessment for students who are unable to participate in the online assessment. Louisiana's sample sizes are too small for traditional studies of comparability for both AP and braille forms.

Appendix B—Transadaptation Process for Spanish Mathematics Forms

For English Learners, the LDOE offers the mathematics assessments in Spanish for both computer-based tests (CBT) in all grades and paper-based tests (PBT) in grades 3 and 4 only to mirror the English language forms, the text-to-speech (TTS) for CBT and large print and human voice audio CDs for PBT forms. The Spanish language versions of the test were developed through transadaptation. Transadaptation takes into consideration the grade-level appropriateness of the words and sentence structures used and the linguistic and cultural differences that exist between speakers of two different languages. Accounting for these differences allows experts to ensure that a Spanish language version of an item will measure the same construct as the English-language version of the item at the same level of rigor. The item is therefore expected to measure the achievement of English learners in the same way that the English version of the item does for native speakers of English.

Once the operational form was approved in English, DRC provided item IDs for acquired items to New Meridian, who then identified which of those items had previously appeared on a Spanish transadapted form. Once New Meridian identified the items that had previously been transadapted and provided the transadaptations of those items, DRC identified the English version of all items that had not been previously transadapted (either because they were Louisiana-owned items that would appear in field-test positions or because they were acquired items that had not been previously used on a Spanish-language form by PARCC). These items were then provided to the Spanish transadaptation subcontractor for initial transadaptation. DRC's Spanish Test Development Team reviewed the previously transadapted items to ensure consistency between those items transadapted as part of the PARCC assessments and those transadapted specifically for Louisiana. The team provided guidance to the translator conducting the initial transadaptation in grade-level and culturally appropriate ways. Upon completion of the transadaptation by the subcontractor, DRC's Spanish Test Development team conducted reviews by native Spanish speakers for content and grade-level appropriateness of the transadaptation. The team also conducted an editorial review. At least two members of DRC's Spanish Test Development team compared each English item to the Spanish transadaptation to ensure that the transadaptation:

- was accurate;
- contained grade-appropriate wording;
- contained answer choices that were reasonably parallel;
- did not introduce ambiguity into the Spanish version;
- contained graphics that were clearly transadapted;
- did not alter current teaching and learning practices in the content area; and
- remained free of gender, ethnic, cultural, socioeconomic, and regional bias.

The Spanish Test Development team then reconciled any discrepancies and submitted the transadaptations to a senior Spanish Test Development team member for resolution. After approval by the senior Spanish Test Development team member, the item moved forward to be imported into DRC's item banking system.

Both previously transadapted items and newly transadapted items were imported into DRC's item banking system and formatted for online use. Each Spanish item was paired with the corresponding English item in the item bank, and the Spanish item was formatted. Graphics for the item were then finalized for review. The

finalized transadaptation was then compared to the Spanish version of the item in the DRC assessment system and the English version of the item, and all changes were verified.

DRC's Spanish Test Development team then used the final, approved communication assistance scripts in English to transadapt descriptions of graphics as necessary. These descriptions were used when preparing the TTS forms for review. Scripting the TTS forms and reviewing the finalized Spanish forms were conducted by native Spanish speakers at DRC prior to submitting the forms to the LDOE for a translation review by a third-party translation vendor. The vendor reviewed the transadapted forms and provided feedback to the LDOE and DRC. Experienced DRC Spanish Test Development team members and the translation vendor resolved any issues, and DRC made modifications as necessary. The forms were then approved by both DRC and the LDOE translation vendor.

Appendix C—LEAP 2025 Spring 2023 Handscoring/AI Documentation

Appendix D—Quality Control References

Related Information		Related Chapter/Source
Test Materials		
Item development quality procedures	Content alignment Cognitive complexity Difficult Bias, fairness, and sensitivity Technical design	Chapter 3
Form development quality procedures	Test specifications Review of statistical quality of items	Chapter 3
Test Administration		
Test administration training and procedures	Training and monitoring of test administrators Security Checklists Test Security Measurements	Chapter 4
Monitoring test administrations	LDOE site audits Data Forensics Analysis Response-Change Analysis Web Monitoring Plagiarism Detection	Chapter 4
Scoring		
Scorer recruitment, training and security procedures	Recruitment and interview process Security Training process, including material development and qualifying procedures.	Chapter 5 Appendix C
Monitoring scoring quality	Inter-rater reliability studies Validity Reader monitoring	Chapter 5 Appendix C
Psychometric Processes		
Psychometric quality procedures	Specifications document for operational analysis	Internal document between DRC and the LDOE.
Monitoring psychometric quality	Key verification Calibration Scoring table generation Psychometric quality checks on the data	Chapter 6
Performance-Level Setting	Quality-controlled procedures for performance-level setting Derivation of the cut scores	Chapter 8

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Beimers, J. N., Way, W. D., McClarty, K. L., & Miles, J. A. (2012, January). Evidence based standard setting: Establishing cut scores by integrating research evidence with expert content judgments. Austin, TX: Pearson. Retrieved from http://researchnetwork.pearson.com/wpcontent/uploads/Bulletin21_Evidence_Based_Standard_Setting.pdf

Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO for Windows [Computer software]. Lincolnwood, IL: Scientific Software International.

Camilli, G., & Shepard, A. L. (1994). *Methods for identifying biased test items.* Thousand Oaks, CA: Sage Publication.

Center for Assessment. (2017, June). *LEAP 2017: English language arts -grade 6 summary – comparability with PARCC performance standards* (Memorandum). Dove, NH.

Chou, Y., & Wang, W. (2010). Checking dimensionality in item response models with principal component analysis on standardized residuals. *Educational and Psychological Measurement, 70*, 717-731.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth.

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

Data Recognition Corporation. (2016*). Interpretive guide: Grades 3–8 ELA and math.* Maple Grove, MN.

Dorans, N. J., & Schmitt, M. P. (1991). *Constructed response and differential item functioning: A pragmatic approach* (Research Report No. RR-91-47) Princeton, NJ: Educational Testing Service.

Educational Testing Service, Pearson, & Measured Progress. (2016). *Final technical report for 2015 administration. PARCC*. Retrieved from https://eric.ed.gov/?q=source%3a%22Partnership+for+Assessment+of+Readiness+for+College+and+Careers%22&id=ED599097

Green, D. R. (1975). *Procedures for assessing bias in achievement tests*. Paper presented at the National Institute of Education Conference on Test Bias, Annapolis, MD.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Hingham, MA: Kluwer-Nijhoff Publishing.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel Procedure. In H. Wainer & H. I. Braun (Eds.), *Test Validity*, pp. 129-145. Hillsdale, NJ: Erlbaum.

Huynh, H. (1998). On score locations for binary and partial credit items and their applications to item mapping and criterion-referenced interpretation. *Journal of Educational and Behavioral Statistics, 23,* 35-56.

Huynh, H., & Meyer, P. (2010). Use of robust z in detecting unstable items in item response theory models. *Practical Assessment, Research & Evaluation*, 15, 1-5.

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

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking*. New York, NY: Springer-Verlag.

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). *Standard setting: A bookmark approach*. Paper presented at the 26th Annual CCSSO National Conference on Large Scale Assessment in Phoenix, AZ.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*(2), 179–197.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Erlbaum.

Lu, Y., & Sireci, S. G., (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice, 26*(40), 29-37.

Lumsden, J. (1957). A factorial approach to unidimensionality. *Australian Journal of Psychology*, 9, 105-111.

Lumsden, J. (1961). The construction of unidimensional tests. *Psychological Bulletin, 58*, 122-131.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4), 719–748.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249–281). Lawrence Erlbaum Associates Publishers.

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

Pearson. (2015*). Performance level setting technical report*. PARCC. Retrieved from https://eric.ed.gov/?q=source%3a%22Partnership+for+Assessment+of+Readiness+for+College+and+Careers%22&id=ED599097 /.

Pearson. (2017). *PARCC: Final technical report for 2016 administration.* PARCC. Retrieved from https://eric.ed.gov/?q=source%3a%22Partnership+for+Assessment+of+Readiness+for+College+and+Careers%22&id=ED599197.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational and Behavioral Statistics, 19,* 1012.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects*. Biometrika, 70*(1), 41–55.

Schumacker, R. E. (1996). Disattenuating correlation coefficients. *Rasch Measurement Transactions, 10*(1), 479.

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

Thompson, S., & Thurlow, M. (2002). *Universally designed assessments: Better tests for everyone!* (Policy Directions No. 14). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/NCEO/OnlinePUbs/Policy14.htm

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30*(3), 233–251.