
# 2017 LEAP 2025
# Operational Technical Report
# English Language Arts and Mathematics

**Submitted to the**
**Louisiana Department of Education**

June 2018

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 2017 administration of the Louisiana Educational Assessment Program (LEAP 2025). The LEAP 2025 is a summative assessment in English Language Arts (ELA) and Mathematics administered in grades 3 through 8. 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 Partnership for Assessment of Readiness for College and Careers (PARCC) consortium's item bank. Items taken from this bank were on pre-established item response theory (IRT) scales. This section provides a summary of the 2017 operational technical report.

## E.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 the 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 the 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 50%) and DRC's College and Career Readiness item bank.

In the 2016–2017 school year, students in grades 3–8, except those qualifying for the LAA 1, were administered forms for ELA and mathematics that consisted of PARCC assessment items. This allowed for the continued comparability to forms administered in the 2014–2015 and 2015–2016 school years.

The information that follows describes the 2017 LEAP 2025 ELA and Mathematics assessments and provides information about how to read and interpret the data on the 2017 assessment reports.

## E.2 Administration

In the spring of 2017, Louisiana administered the LEAP 2025 summative assessments in ELA and mathematics to students in grades 3–8. A paper-based test (PBT) was administered in grade 3, a PBT and a computer-based test (CBT) were administered in grade 4, and a CBT was administered in grades 5–8. The CBTs were administered from April 3 to May 5, 2017. The PBTs were administered from May 1 to May 5, 2017. Test administration is discussed in Chapter 4 of this report.

Approximately 72 school systems and 34 charter schools administered 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**

| Grade | Accountable in ELA | Percentage Reportable in ELA | Accountable in Mathematics | Percentage Reportable in Mathematics* |
|---|---|---|---|---|
| 3 | ≥ 57,110 | 99.05% | ≥ 57,350 | 99.67% |
| 4 | ≥ 56,580 | 99.05% | ≥ 56,780 | 99.60% |
| 5 | ≥ 53,310 | 99.79% | ≥ 53,340 | 99.75% |
| 6 | ≥ 52,480 | 99.66% | ≥ 52,510 | 99.63% |
| 7 | ≥ 51,960 | 99.60% | ≥ 51,980 | 99.59% |
| 8 | ≥ 50,590 | 99.46% | ≥ 50,620 | 99.43% |

**Algebra I students in grade 8 had the option of taking Algebra EOC instead of LEAP 2025 Mathematics test.*

## E.3 Student Performance

Tables E.2 and E.3 present the percentage of students who were classified in each of the 2017 achievement levels for ELA and mathematics.

**Table E.2 Percentage of Students Classified in 2017 Achievement Levels Using 2017 Census Data: English Language Arts**

| Grade | Unsatisfactory | Approaching Basic | Basic | Mastery | Advanced |
|---|---|---|---|---|---|
| 3 | 13.4 | 17.8 | 24.7 | 38.9 | 5.1 |
| 4 | 8.8 | 18.3 | 29.3 | 36.2 | 7.3 |
| 5 | 8.7 | 18.8 | 31.1 | 37.9 | 3.4 |
| 6 | 10.4 | 24.9 | 29.8 | 29.4 | 5.5 |
| 7 | 13.2 | 19.2 | 26.5 | 30.3 | 10.8 |
| 8 | 11.4 | 17.4 | 27.0 | 35.1 | 9.0 |

**Table E.3 Percentage of Students Classified in 2017 Achievement Levels Using 2017 Census Data: Mathematics**

| Grade | Unsatisfactory | Approaching Basic | Basic | Mastery | Advanced |
|---|---|---|---|---|---|
| 3 | 11.1 | 18.4 | 27.1 | 36.2 | 7.1 |
| 4 | 8.2 | 23.2 | 29.7 | 35.0 | 3.8 |
| 5 | 11.1 | 24.9 | 32.4 | 27.7 | 3.9 |
| 6 | 12.6 | 30.8 | 29.2 | 23.7 | 3.7 |
| 7 | 11.2 | 28.9 | 35.2 | 22.6 | 2.1 |
| 8 | 20.3 | 28.2 | 25.0 | 24.7 | 1.8 |

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 use of the LEAP 2025 test scores. Chapter 2 discusses the 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 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 2017 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 performance-level setting procedures implemented by PARCC since 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 and development processes used to ensure fairness of the LEAP 2025 for all examinees.

# CHAPTER 1: INTRODUCTION

The LEAP 2025 is designed to measure students' knowledge of ELA and mathematics. This report provides a technical overview of the LEAP 2025 ELA and Mathematics assessments administered in the spring of 2017 and presents evidence for the validity of the 2017 LEAP 2025 ELA and mathematics assessment scores.

This chapter describes the background, history, purpose, and design of the LEAP 2025 and provides an overview of the chapters in this technical report.

## 1.1 Purpose of the LEAP 2025

The Board of Elementary and Secondary Education (BESE) and the LDOE are committed to ensuring that every student is on track to be successful in postsecondary education and the workforce through their comprehensive plan, Louisiana Believes. The LEAP 2025 supports this vision by measuring student readiness for the next grade or course of study.

## 1.2 Design of the LEAP 2025

A paper-based test (PBT) was administered in grades 3 and 4 for both ELA and mathematics, and a computer-based test (CBT) was administered in both subjects in grades 4–8. Additionally, the mathematics form was translated to Spanish in all grades. Large-print test forms were available for PBTs and braille test forms were available for all paper-based and computer-based tests to enable students who are visually impaired to participate in LEAP 2025 testing. See Chapter 3, Section 3.5 for more information about the accommodations and designated supports available for LEAP 2025.

## 1.3 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 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. An overview of major activities documented within this report is provided below.

### *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 use of the LEAP 2025 scores, and Chapter 2 discusses some of the uses of the scores.

### *Test Content Development (Chapter 3)*

Chapter 3 of the technical report provides a summary of the test development activities that occurred to create the Spring 2017 operational test forms.

### *Test Administration (Chapter 4)*

Chapter 4 of the technical report describes the processes and activities implemented and the information disseminated to help ensure standardized test administration procedures and, thus, uniform test administration conditions for students.

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

Chapter 5 of the technical report describes the processes and activities for scoring constructed-response items. This chapter discusses how scorers are trained and the measures used for ensuring consistency among scorers. Finally, this chapter presents the results of the inter-rater reliability studies.

***Operational Data Analyses (Chapter 6)***

Chapter 6 of the technical report includes a detailed description of the operational analyses of the 2017 LEAP 2025, which include the following major parts: the classical item analysis; calibration, scaling, and linking using item response theory (IRT) models; and student scoring. This chapter also describes the demographics of the calibration samples and compares them to state census data. It reports the results of the classical item analysis and the results of the calibration, scaling, and linking processes.

***Test Results (Chapter 7)***

Chapter 7 of the technical report contains information on the results of the Spring 2017 LEAP 2025 administration. Detailed summary statistics based on scale scores and achievement-level information 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 PARCC performance-level setting procedures and derivation of 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 reliability and validity of the LEAP 2025 scores. This chapter provides detailed results of the reliability of the tests as well as information on the decision consistency of the cut scores. It also provides evidence of construct validity for the LEAP 2025 scores.

***Fairness (Chapter 10)***

Chapter 10 of the technical report discusses fairness and how the LEAP 2025 tests are constructed 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 to determine whether large differences exist between demographic groups in Louisiana. Results of the administration mode study are also summarized.

# CHAPTER 2: THE USES OF TEST SCORES

Validity is the central component of the LEAP 2025. 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 scores. It is recommended that policy and program evaluation studies, in accordance with the *Standards*, be conducted to support some of the uses of the LEAP 2025 scores that are anticipated.

## 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 scores include the following:

- evaluating students' overall proficiency of the Louisiana Student Standards
- identifying students' strengths and weaknesses
- evaluating programs at the school, school system, and/or state level
- informing stakeholders, including 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 use 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 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 ELA and mathematics test forms were developed by DRC's test development staff using the PARCC consortium's item bank. Items taken from this bank were on pre-established item response theory (IRT) scales for ELA and mathematics and were reviewed and approved for use by LDOE

content experts. Braille and large-print forms, in addition to Spanish translations of mathematics forms, were also developed.

The following sections discuss two types of test-level scores that are reported to indicate a student's achievement on the LEAP 2025: (1) the scale score and (2) its associated level of achievement.

### 2.2.1 Scale Scores

A scale score indicates a student's total performance for each content area on the LEAP 2025. The overall scale score for a content area quantifies the achievement being measured by the ELA or mathematics test. 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.

### 2.2.2 Levels of Achievement

A student's performance on the ELA or mathematics LEAP 2025 is reported in one of five levels of achievement: *Unsatisfactory*, *Approaching Basic*, *Basic*, *Mastery*, or *Advanced*. 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) (see https://parcc-assessment.org/wp-content/uploads/2017/12/PARCC_PLS_TechReport_011316_toPARCC-final.pdf).

Descriptions of each level of achievement in terms of what a student should know and be able to do are provided with the *Guide to Interpreting Results* (see Chapters 4 and 7).

### 2.2.3 Use of Test-Level Scores

The LEAP 2025 scale scores and achievement levels provide summary evidence of student achievement in ELA or mathematics. 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 student performance in ELA and mathematics.

## 2.3 Claim-Level and Subclaim-Level Subscores

A student's performance on the ELA claims (i.e., reading and writing) is reported by one of three ratings: *Weak*, *Moderate*, or *Strong*.

Additionally, subclaim subscores are reported at the student level for ELA and mathematics. ELA has three subclaims for reading and two subclaims for writing. Mathematics has four subclaims. Subclaim performance is reported in one of three ratings: *Weak, Moderate,* or *Strong*.

Although the performance ratings are determined only by the items included within a claim or subclaim, the level of knowledge and ability needed to achieve a performance rating is connected to the level of knowledge and ability required by the content-level tests: a *Weak* rating requires similar knowledge and ability as the *Unsatisfactory* and *Below 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.

### 2.3.1 Use of the Claim-Level and Subclaim-Level Subscores

The purpose of reporting claim- or subclaim-level subscores on LEAP 2025 tests 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 claims in ELA and the subclaims in ELA and mathematics. Teachers may use these subscores for individual students as indicators of strengths and weaknesses, but they 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 claim- or subclaim-level subscores. Chapter 9 of this technical report provides evidence of construct-related validity that further supports the use of these subscores.

# CHAPTER 3: TEST CONTENT DEVELOPMENT

Content-related validity in achievement tests is evidenced by a correspondence between test content and a specification of the content domain. 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 section 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 2017 LEAP 2025 was expected to cover.

This chapter is particularly relevant to the following parts 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)

## 3.1 Test Specifications

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 2017 LEAP 2025 test specifications consisted of a test blueprint and a test design for each grade and content area. To construct the assessments following the LDOE-approved test blueprints and test designs, LDOE and DRC collaborated to use items from the PARCC item bank that were aligned to the Louisiana Student Standards. The blueprints and test designs were closely aligned to the PARCC operational blueprint that had been used for the previous test administration in Louisiana and that was created by LDOE in collaboration with DRC. The ELA and Mathematics LEAP 2025 assessments for grades 3 through 8 were developed based on the requirements of RFP #678PUR-LEAP 2016 Mathematics and ELA as follows:

The assessments shall

- be aligned to ELA and mathematics Louisiana Student Standards;
- be 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;
- be constructed to yield valid and reliable test results that report student performance, using achievement levels that are comparable to the levels used by a significant number

of other states that have similarly high expectations for student learning and to the achievement levels used for Louisiana's 2015 and 2016 grades 3–8 ELA and Mathematics assessments;
- be developed and reviewed with LDOE assessment staff and educators;
- not be designed to be computer adaptive;
- be used in assessing students' readiness to successfully transition to postsecondary education and the workplace; and
- be administered through a separate administration contract in multiple modalities.

The products of the above requirements are dual-mode assessments—paper-based tests (PBTs) and computer-based tests (CBTs)—composed of PARCC test items aligned to the Louisiana Student Standards. The contract with PARCC provided for the use of enough items and related passages 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). These items and passages became the available item pool used to construct the 2017 forms. For ELA and mathematics, the 2017 LEAP 2025 test blueprints were finalized in October 2016. 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 emphasis in instruction and mastery at each grade level and are comparable to PARCC 2017 test blueprints. The test specifications provide the numbers of items by strand, assessment focus, and item type, and they demonstrate the desired proportions within test delivery and available item pool constraints. The test designs for ELA and mathematics were finalized in January 2017 by LDOE and DRC. All assessments were fixed forms.

### 3.1.1 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.

#### 3.1.1.1. English Language Arts Test Blueprints and Test Designs

The test was administered during one PBT or CBT testing window. 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, but all three types were represented across grades 3 through 8. This allows Louisiana to rotate the tasks given for each grade from year to year and encourages teachers to focus equally on all three task types. Since the choice of Literary Analysis Task or Narrative Writing Task would be made during the forms construction process, alternative blueprints—one with a Research Simulation Task and a Literary Analysis 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 grade 6 and the Literary Analysis Task was selected for grades 3, 4, 5, 7, and 8, all based on item performance and the quality of the available passage sets for each task.

The session testing times shown in the ELA test blueprints (see Tables 3.1 through 3.6) are based on PARCC testing times proportioned to be comparable based on the passage type. In the blueprints, 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 and to provide consistent timing in sessions one and two.

#### 3.1.1.2. Mathematics Test Blueprints and Test Designs

The test was administered during one PBT or CBT testing window. The 2017 mathematics test had a similar structure as the 2016 assessment; each test session included the four mathematics subclaims using the three mathematics task types (see Table 3.7). The resulting 2017 LEAP 2025 mathematics test blueprints are shown in Tables 3.8 through 3.13.

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 into test sessions. 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 through 5, students are prohibited from using calculators, except for those students with a calculator accommodation. For session one of the mathematics test in grades 6 through 8, students are prohibited from using calculators, except for those students with a calculator accommodation; calculators are allowed to be used by all students in sessions two and three. The general test structures (see Tables 3.14–3.19) guided test form sequencing and design.

PARCC's calculator guidelines, outlined in the consortium's evidence statement tables, provided the basis for calculator designation of tasks and items.

## 3.2 Item Development and Selection

The process 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 2017 LEAP 2025 ELA and mathematics assessments came from the PARCC consortium's item bank.

The PARCC items selected for use on the 2017 LEAP forms were used to equate to the PARCC scale. Refer to PARCC's website for information about processes, procedures, and timelines for item development, review, field testing, and item selection for operational use. Operational forms were selected based on LEAP 2025 test blueprint specifications, which were supported by statistical data from PARCC operational testing.

### 3.2.1 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 during the development of the PARCC 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, gender, regional background, native language, race, religion, sexual orientation, and socioeconomic status. DRC strongly relied 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 Item Alignment Review.

### 3.2.2 Item Reviews

As part of PARCC's ongoing item development practices, the consortium conducted external bias and fairness reviews that were independent of DRC's reviews. Refer to PARCC's website for information on PARCC's item review processes and procedures (see https://parcc-assessment.org/). These processes follow 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)

### 3.2.3 Louisiana Item Alignment Review

DRC, with guidance from LDOE, conducted the Louisiana Item Alignment Review in January 2017 with committees of Louisiana educators. The following four committees met for two (mathematics) or three (ELA) days to provide feedback on the alignment and appropriateness of items that made up the PARCC item bank. To the extent possible, each committee included representation for all types of students in Louisiana, including those from diverse backgrounds and learning opportunities.

During this review, educators reviewed items individually prior to discussing the items as a group to come 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 and be free of issues of bias, fairness, or sensitivity that could impact student responses to the item.

## 3.3 Operational Test Selection

Operational item selections for 2017 were performed from January through February 2017 by LDOE and DRC. The PARCC item pool was used to select fixed LEAP 2025 ELA and Mathematics forms. A small number of items developed by DRC were used to supplement the item selection when no PARCC items that met Louisiana State Standards were available.

### 3.3.1 English Language Arts 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 content specialists, and it was followed by a psychometric evaluation of each selection. Since PARCC items are the only links to the PARCC scale used for equating purposes, the critical psychometric consideration, other than individual item

performance, was the degree to which the PARCC items reflected the 2015 and 2016 targets. Although the PARCC item pool was limited, items that were determined to be very difficult (item response theory [IRT] B > 2.0) and/or not discriminating (IRT A < 0.3) were avoided when possible.

***Item Selection Guidelines***

- Using the PARCC pool of items, ELA content specialists select tasks and passage sets to match the blueprint. The tasks and sets include items that cover a range of standards and address the appropriate claims and subclaims.
- ELA content specialists verify that each item aligns to the content standards specified in the Louisiana English Language Arts Student Standards.
- ELA content specialists verify that each item meets psychometric guidelines for excellence as available item-performance data allows.
- ELA content specialists verify that each item meets technical quality requirements for well-crafted items, including
  - clear and correct answer(s) based on item specifications;
  - clear and concise language;
  - grammatical correctness;
  - appropriate range of difficulty; and
  - content that is not offensive, inappropriate, or biased.

#### 3.3.1.1. ELA Content Review and Forms Development

After ELA tasks and passage sets were selected and reviewed by DRC and LDOE ELA content specialists, they were placed in forms. In constructing the forms, DRC and LDOE ELA content specialists used the guidelines provided in the following list.

***Guidelines for Placing Tasks and Passage Sets into Forms***

- Forms should include adequate content coverage, as required by the detailed test blueprint.
- No item in a form should "clue" another item on that same form.
- Clang association should be avoided. Clang association is when a distractor can be associated with a stem word by sound (e.g., rhyming, alliteration) rather than meaning.
- Passage sets in forms should be diverse.
- Forms should include a wide range of topics and a variety of questions.
- Correct answer distributions should follow best practice (i.e., no more than three keys of the same answer option in a row).
- Forms *should not* contain any items that have been released to the public.

#### 3.3.1.2. Review of the English Language Arts Items and Forms

DRC and LDOE ELA content specialists, along with teacher committees, verified that the items are 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 item to the content standard 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 that it purports to

measure. The ELA content specialists also verify 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 review each item for item quality, making sure that the test items are 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 list below serves to summarize the major considerations ELA content specialists follow when reviewing items to make sure the items conform 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 an evidence-based selected response (EBSR) item with only four answers in each part
  - only the indicated number of correct answers for a multiple select (MS) item
- have a correctly assigned content code (i.e., item map);
- measure one main idea or standard, unless it is a complex item, such as a prose constructed-response (PCR) item;
- measure the objective or content standard 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 of the student being tested;
- be based on content that is accurate and current;
- contain stimulus material that is clear and concise and provides all the necessary information, when appropriate;
- contain graphics that are clearly labeled, when appropriate;
- contain answer choices that are plausible and reasonable in terms of the requirements of the question as well as 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, cultural, socioeconomic, and regional bias.

PBTs were developed for students in grades 3 and 4, and CBTs were developed for students in grades 4 through 8. The dual-mode forms were identical except for a small quantity (two to three items) of technology-enhanced items (TEIs) in each CBT. Items used on PBTs as replacements for the TEIs were selected-response items that addressed the same content standards and were of similar rigor as the TEIs, when possible.

### 3.3.2 Mathematics Item Selection Process and Criteria

The item selection process used for forms construction was a content-focused, collaborative process between the LDOE and DRC mathematics content specialists, and it was followed by a psychometric evaluation of each selection. For equating purposes, PARCC items are the only links to the PARCC scale. Therefore, in addition to individual PARCC item performance, a critical psychometric consideration was the degree to which the PARCC items reflected the 2015 and 2016 targets. Although the PARCC item pool was limited, items that were determined to be very difficult (IRT B > 2.0) and/or not discriminating (IRT A < 0.3) items were avoided when possible.

***Item-Selection Guidelines***

- Using the PARCC pool of items, mathematics content specialists select tasks to match the blueprint. The tasks include items that cover a range of standards and address the appropriate subclaims.
- Mathematics content specialists verify that each item aligns to the content standards specified in the Louisiana Student Standards for Mathematics.
- Mathematics content specialists verify that each item meets psychometric guidelines for excellence as available item performance data allows.
- Mathematics content specialists verify that each item meets technical quality requirements for well-crafted items, including
    - clear and correct answer(s) based on item specifications;
    - clear and concise language;
    - grammatical correctness;
    - appropriate range of difficulty; and
    - content that is not offensive, inappropriate, or biased.

#### 3.3.2.1. Mathematics Content Review and Forms Development

After items have been selected and reviewed by DRC and LDOE mathematics content specialists for both psychometric excellence and technical quality, they were placed in forms. In constructing the forms, the mathematics content specialists used the guidelines provided in the following list.

***Guidelines for Placing Items into Forms***

- Forms should include adequate content coverage, as required by the detailed test blueprint.
- No item in a form should "clue" another item on that same form. If it is necessary to use items with clueing issues due to item pool limitations, then these items should be positioned in separated testing sessions.
- Clang association should be avoided. Clang association is when a distractor can be associated with a stem word by sound (e.g., rhyming, alliteration) rather than meaning.
- Forms should be ethnically diverse, both in terms of artwork and in terms of names.
- Forms should target an equal representation of genders, both in terms of artwork and names.

- Forms should include a wide range of topics and a variety of questions.
- Correct answer distributions should follow best practice (i.e., no more than three keys of the same answer option in a row).
- Forms *should not* contain any items that have been released to the public.

#### 3.3.2.2. Review of the Mathematics Items and Forms

DRC and LDOE mathematics content specialists also ensure the items are 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 item to the content standard 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 that it purports to measure. The mathematics content specialists also verify 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 mathematics content specialists review each item for item quality, making sure that the test items are 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 list below serves to summarize the major considerations mathematics content specialists follow when reviewing items to make sure the items conform 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 (i.e., 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 of the student being tested;
- 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 as well as 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, cultural, socioeconomic, and regional bias.

PBTs were developed for students in grades 3 and 4, and CBTs were developed for students in grades 4 through 8. The dual-mode forms are identical except for a small quantity (one to two items) of one-point TEI items in each CBT. Items used on PBTs as replacements for the TEI items were selected-response items that addressed the same content standards and were are of similar rigor at the TEI items, when possible. CBT short-answer (SA) items were reformatted as gridded-response (GR) items for use on PBTs .

### 3.3.3 Item-Selection Options for Special Cases

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 (such as content coverage) may possibly be affected in such cases.

### 3.3.4 Psychometric Review

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

#### 3.3.4.1. Selecting Targets

The PARCC 2016 operational form was used to build a new PARCC 2016 scale and to select the items in the PARCC 2017 operational form. The rationale for the choice of the targets was that each 2017 LEAP 2025 form should be on the PARCC scale and 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 and the selected 2017 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 2017 form. This summarizes the amount of measurement error surrounding a scale score.

#### 3.3.4.2. 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, other than individual item performance, is the extent to which the TCC and SEM of the anchor set aligns to that of the total test.

All intact PARCC items were used as anchor items for equating. The TCCs and SEMs for all selected 2017 items and anchor items are the same or very similar because all items are PARCC items. Therefore, the TCCs and SEMs for the anchor sets and the total test were not plotted.

### 3.4 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 3.20 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 addresses how the elements of universal design were addressed in the construction of the Spring 2017 test forms 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 accommodate the largest number of students administered the LEAP 2025, the assessments include simple, clear, and intuitive instructions and procedures; maximum readability and comprehensibility; and maximum legibility. All these design components are addressed primarily through the CBTs. The page specifications define how directions and test items are placed on the pages, the location and appearance of headers and footers, spacing between an item stem and answer choices, and other page elements to ensure a consistent, legible appearance of CBTs. Written instructions 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 PARCC.

### 3.5 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 disabilities and students with limited English proficiency may be provided test administration accommodations based on their Individualized Education Plan (IEP). More information on accommodations can be found in Section 4.4.2 of Chapter 4. Accommodation code definitions can be found in the *Test Administration Manual*.

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 testing. Braille and large-print forms for ELA and mathematics were based on the standard-print forms. Specific recommendations on how to transcribe items into braille were provided by the braille publisher to produce the braille version of the LEAP 2025 and the test administrator's notes that accompany the braille forms. The goal was to maximize the number of items on the braille form, and it was possible to transcribe all items into braille.

The following additional access and accommodation features were available for PBTs and CBTs: blank scratch paper, bookmarks, calculators (to be used in the calculator section only), color overlays, contrasting colors/reverse colors, directions in native languages, general directions clarified and/or read aloud, general masking, highlighters, line guides, magnifiers/variable zoom, measurement tools, noise buffer/headphones, strike-through, write-on test, mathematics text-to-speech, mathematics human reader, and Spanish translations of the mathematics tests.

### 3.6 Standards and Content 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. To address 2017 LEAP 2025 comparability goals with PARCC 2017, PARCC claims, subclaims, and evidence statements, along with guidance provided by the *Louisiana Student Standards for ELA and Mathematics*, were used as item and task specifications.

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

These item specifications were developed for each grade level and standard to delineate the expectations of knowledge and skill to be included on test questions in each grade. In addition, the ELA and mathematics item and stimulus specifications provide guidance on determining the grade-appropriateness of task and stimulus materials (i.e., the materials that a student must refer to in working on a test question). The stimulus specifications also provide information on characteristics of stimuli or activities to avoid because they are not important to the knowledge, skill, or process being measured. This is important because it underscores DRC's efforts to select items that are accessible to the widest range of students possible; in other words, 2017 LEAP 2025 items were selected according to the elements of universal design.

Table 3.21 provides the distribution of ELA items and points on the 2017 LEAP 2025 by session and item type.

Table 3.22 provide the distribution of mathematics points on the 2017 LEAP 2025 by subclaim.

Table 3.23 provides the distribution of mathematics tasks and points on the 2017 LEAP 2025 by task type.

### 3.7 Summary

In summary, the overall purpose of this chapter is to explicate the procedures used in the development of the 2017 LEAP 2025 grade-level assessments. The efforts by LDOE and DRC in developing the LEAP 2025 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)

**Table 3.1 Grade 3 English Language Arts Test Blueprint and Test Design**

| Section | Task/ Item Set | Number of Passages | Claims/ Subclaims | Number of Two-Point EBSR Items | Number of Points from Two-Point EBSR Items | Number of PCR Items | Number of Points from PCR Items | Total Items | Total Points | Standards | Testing Time (minutes) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Literary Analysis Task | 2 | Reading: Reading Literary Text/Reading Vocabulary* | 5 | 10 | 1 | 3 | 5 | 13 | RL.1-3, 5-10, RL.4, L.4, L.5 | 75 |
| | | | Writing: Written Expression | 0 | 0 | | 9 | 1 | 9 | W.1-2, 10 | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Totals | 2 | | 5 | 10 | 1 | 15 | 6 | 25 | | |
| 2 | Research Simulation Task | 2 | Reading: Reading Informational Text/ Reading Vocabulary* | 5 | 10 | 1 | 3 | 5 | 13 | RI.1-3, 5-10, RI.4, L.4, L.5 | 75 |
| | | | Writing: Written Expression | 0 | 0 | | 9 | 1 | 9 | W.1-2, 7-8, 10 | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Totals | 2 | | 5 | 10 | 1 | 15 | 6 | 25 | | |
| 3 | Reading Literary Texts | 2 | Reading: Reading Literary Text/Reading Vocabulary* | 8 | 16 | 0 | 0 | 8 | 16 | RL.1-3, 5-10, RL.4, L.4, L.5 | 60 |
| | Reading Informational Texts | 1 | Reading: Reading Informational Text/Reading Vocab* | 6 | 12 | 0 | 0 | 6 | 12 | RI.1-3, 5-10, RI.4, L.4, L.5 | |
| | Totals | 3 | | 14 | 28 | 0 | 0 | 14 | 28 | | |
| **Grade 3 Totals** | | **7** | Reading: Reading Literary Text/Reading Vocab* | 13 | 26 | 2 | 3 | 13 | 29 | 54 | 210 |
| | | | Reading: Reading Informational Text/Reading Vocab* | 11 | 22 | | 3 | 11 | 25 | | |
| | | | Writing: Written Expression | 0 | 0 | | 18 | 2 | 18 | 24 | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 6 | | 6 | | |
| | | | **Total** | 24 | 48 | 2 | 30 | 26 | 78 | 78 | |

**There must be at least eight points of reading vocabulary items on the test.*

**Table 3.2 Grade 4 English Language Arts Test Blueprint and Test Design**

| Section | Task/ Item Set | Number of Passages | Claims/ Subclaims | Number of Two-Point EBSR Items | Number of Points from Two-Point EBSR Items | Number of PCR Items | Number of Points from PCR Items | Total Items | Total Points | Standards | Testing Time (minutes) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Literary Analysis Task | 2 | Reading: Reading Literary Text/Reading Vocabulary* | 5 | 10 | 1 | 4 | 5 | 14 | RL.1-3, 5-10, RL.4, L.4, L.5 | 90 |
| | | | Writing: Written Expression | 0 | 0 | | 12 | 1 | 12 | W.1-2, 4, 9, 10, | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Reading Literary/ Texts | 1 | Reading (Reading Literary Text/Reading Vocabulary) | 4 | 8 | 0 | 0 | 4 | 8 | RL.1-3, 5-10, RL.4, L.4, L.5 | |
| | Totals | 3 | | 9 | 18 | 1 | 19 | 10 | 37 | | |
| 2 | Research Simulation Task | 3 | Reading: Reading Informational Text/ Reading Vocabulary* | 7 | 14 | 1 | 4 | 7 | 18 | RI.1-3, 5-10, RI.4, L.4, L.5 | 90 |
| | | | Writing: Written Expression | 0 | 0 | | 12 | 1 | 12 | W.1-2, 4, 7-10, | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Totals | 3 | | 7 | 14 | 1 | 19 | 8 | 33 | | |
| 3 | Reading Literary Texts | 1 | Reading: Reading Literary Text/Reading Vocabulary* | 4 | 8 | 0 | 0 | 4 | 8 | RL.1-3, 5-10, RL.4, L.4, L.5 | 45 |
| | Reading Informational Texts | 1 | Reading: Reading Informational Text/Reading Vocab* | 6 | 12 | 0 | 0 | 6 | 12 | RI.1-3, 5-10, RI.4, L.4, L.5 | |
| | Totals | 2 | | 10 | 20 | 0 | 0 | 10 | 20 | | |
| **Grade 4 Totals** | | **8** | Reading: Reading Literary Text/Reading Vocab* | 13 | 26 | 2 | 4 | 13 | 60 | 60 | 225 |
| | | | Reading: Reading Informational Text/Reading Vocab* | 13 | 26 | | 4 | 13 | | | |
| | | | Writing: Written Expression | 0 | 0 | | 24 | 2 | 24 | 30 | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 6 | | 6 | | |
| | | | **Total** | 26 | 52 | 2 | 38 | 28 | 90 | 90 | |

**There must be at least eight points of reading vocabulary items on the test.*

**Table 3.3 Grade 5 English Language Arts Test Blueprint and Test Design**

| Section | Task/ Item Set | Number of Passages | Claims/ Subclaims | Number of Two-Point EBSR Items | Number of Points from Two-Point EBSR Items | Number of PCR Items | Number of Points from PCR Items | Total Items | Total Points | Standards | Testing Time (minutes) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Literary Analysis Task | 2 | Reading: Reading Literary Text/Reading Vocabulary* | 5 | 10 | 1 | 4 | 1 | 14 | RL.1-3, 5-10, RL.4, L.4, L.5 | 90 |
| | | | Writing: Written Expression | 0 | 0 | | 12 | | 12 | W.1-2, 4, 9, 10, | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Reading Literary Texts | 1 | Reading (Reading Literary Text/Reading Vocabulary) | 4 | 8 | 0 | 0 | 4 | 8 | RL.1-3, 5-10 RL.4, L.4, L.5 | |
| | Totals | 3 | | 9 | 18 | 1 | 19 | 10 | 37 | | |
| 2 | Research Simulation Task | 3 | Reading: Reading Informational Text/ Reading Vocabulary* | 7 | 14 | 1 | 4 | 1 | 18 | RI.1-3, 5-10, RI.4, L.4, L.5 | 90 |
| | | | Writing: Written Expression | 0 | 0 | | 12 | | 12 | W.1-2, 4, 7-10, | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Totals | 3 | | 7 | 14 | 1 | 19 | 8 | 33 | | |
| 3 | Reading Literary Texts | 1 | Reading: Reading Literary Text/Reading Vocabulary* | 4 | 8 | 0 | 0 | 4 | 8 | RL.1-3, 5-10, RL.4, L.4, L.5 | 45 |
| | Reading Informational Texts | 1 | Reading: Reading Informational Text/Reading Vocab* | 6 | 12 | 0 | 0 | 6 | 12 | RI.1-3, 5, 7-10, RI.4, L.4, L.5 | |
| | Totals | 2 | | 10 | 20 | 0 | 0 | 10 | 20 | | |
| **Grade 5 Totals** | | **8–9** | Reading: Reading Literary Text/Reading Vocab* | 13 | 26 | 2 | 4 | 2 | 30 | 60 | 225 |
| | | | Reading: Reading Informational Text/Reading Vocab* | 13 | 26 | | 4 | | 30 | | |
| | | | Writing: Written Expression | 0 | 0 | | 24 | | 24 | 30 | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 6 | | 6 | | |
| | | | **Total** | 26 | 52 | 2 | 38 | 28 | 90 | 90 | |

**There must be at least eight points of reading vocabulary items on the test.*

**Table 3.4 Grade 6 English Language Arts Test Blueprint and Test Design**

| Section | Task/ Item Set | Number of Passages | Claims/ Subclaims | Number of Two-Point EBSR Items | Number of Points from Two-Point EBSR Items | Number of PCR Items | Number of Points from PCR Items | Total Items | Total Points | Standards | Testing Time (minutes) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Research Simulation Task | 3 | Reading: Reading Informational Text/Reading Vocabulary* | 7 | 14 | 1 | 4 | 7 | 18 | RI.1-3, 5-10, RI.4, L.4, L.5 | 90 |
| | | | Writing: Written Expression | 0 | 0 | | 12 | 1 | 12 | W.1-2, 4, 7-10, | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Totals | 3 | | 7 | 14 | 1 | 19 | 8 | 33 | | |
| 2 | Narrative Writing Task | 1 | Reading: Reading Literary Text/Reading Vocabulary* | 4 | 8 | 1 | 0 | 4 | 8 | RL.1-3, 5-10, RL.4, L.4, L.5 | 90 |
| | | | Writing: Written Expression | 0 | 0 | | 12 | 1 | 12 | W.3, 4, 10 | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Reading Literary/Informational Texts | 2 | Reading (Reading Literary Text/Reading Informational Text/Reading Vocabulary) | 6 | 12 | 0 | 0 | 6 | 12 | RL.1-3, 5-10, RI.1-3, 5-10, RL.4, RI.4, L.4, L.5 | |
| | Totals | 3 | | 10 | 20 | 1 | 15 | 11 | 35 | | |
| 3 | Reading Literary Texts | 3 | Reading: Reading Literary Text/Reading Vocabulary* | 10 | 20 | 0 | 0 | 10 | 20 | RL.1-3, 5-10, RL.4, L.4, L.5 | 70 |
| | Reading Informational Texts | | Reading: Reading Informational Text/Reading Vocab* | 4 | 8 | 0 | 0 | 4 | 8 | RI.1-3, 5, 7-10, RI.4, L.4, L.5 | |
| | Totals | 3 | | 14 | 28 | 0 | 0 | 14 | 28 | | |
| **Grade 6 Totals** | | **9** | Reading: Reading Literary Text/Reading Vocab* | 16 | 32 | 2 | 0 | 16 | 32 | 66 | 250 |
| | | | Reading: Reading Informational Text/Reading Vocab* | 15 | 30 | | 4 | 15 | 34 | | |
| | | | Writing: Written Expression | 0 | 0 | | 24 | 2 | 24 | 30 | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 6 | | 6 | | |
| | | | **Total** | 31 | 62 | 2 | 34 | 33 | 96 | 96 | |

**There must be at least eight points of reading vocabulary items on the test.*

**Table 3.5 Grade 7 English Language Arts Test Blueprint and Test Design**

| Section | Task/ Item Set | Number of Passages | Claims/ Subclaims | Number of Two-Point EBSR Items | Number of Points from Two-Point EBSR Items | Number of PCR Items | Number of Points from PCR Items | Total Items | Total Points | Standards | Testing Time (minutes) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Literary Analysis Task | 2 | Reading: Reading Literary Text/Reading Vocabulary* | 5 | 10 | 1 | 4 | 1 | 14 | RL.1-3, 5-10, RL.4, L.4, L.5 | 90 |
| | | | Writing: Written Expression | 0 | 0 | | 12 | | 12 | W.1-2, 4, 9, 10, | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Reading Informational Texts | 1 | Reading (Reading Informational Text/Reading Vocabulary) | 4 | 8 | 0 | 0 | 4 | 8 | RI.1-3, 5-10, RI.4, L.4, L.5 | |
| | Totals | 3 | | 9 | 18 | 1 | 19 | 10 | 37 | | |
| 2 | Research Simulation Task | 3 | Reading: Reading Informational Text/ Reading Vocabulary* | 7 | 14 | 1 | 4 | 1 | 18 | RI.1-3, 5-10, RI.4, L.4, L.5 | 90 |
| | | | Writing: Written Expression | 0 | 0 | | 12 | | 12 | W.1-2, 4, 7- 10, | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Totals | 3 | | 7 | 14 | 1 | 19 | 8 | 33 | | |
| 3 | Reading Literary Texts | 2 | Reading: Reading Literary Text/Reading Vocabulary* | 6 | 12 | 0 | 0 | 6 | 12 | RL.1-3, 5-10, RL.4, L.4, L.5 | 70 |
| | Reading Informational Texts | 1 | Reading: Reading Informational Text/Reading Vocab* | 6 | 12 | 0 | 0 | 6 | 12 | RI.1-3, 5, 7-10, RI.4, L.4, L.5 | |
| | Totals | 3 | | 12 | 24 | 0 | 0 | 12 | 24 | | |
| **Grade 7 Totals** | | **9** | Reading: Reading Literary Text/Reading Vocab* | 11 | 22 | 2 | 4 | 2 | 26 | 64 | 250 |
| | | | Reading: Reading Informational Text/Reading Vocab* | 17 | 34 | | 4 | | 38 | | |
| | | | Writing: Written Expression | 0 | 0 | | 24 | | 24 | 30 | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 6 | | 6 | | |
| | | | **Total** | 28 | 56 | 2 | 38 | 30 | 94 | 94 | |

**There must be at least eight points of reading vocabulary items on the test.*

**Table 3.6 Grade 8 English Language Arts Test Blueprint and Test Design**

| Section | Task/ Item Set | Number of Passages | Claims/ Subclaims | Number of Two-Point EBSR Items | Number of Points from Two-Point EBSR Items | Number of PCR Items | Number of Points from PCR Items | Total Items | Total Points | Standards | Testing Time (minutes) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Literary Analysis Task | 2 | Reading: Reading Literary Text/Reading Vocabulary* | 5 | 10 | 1 | 4 | 1 | 14 | RL.1-3, 5-10, RL.4, L.4, L.5 | 90 |
| | | | Writing: Written Expression | 0 | 0 | | 12 | | 12 | W.1-2, 4, 9, 10, | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Reading Informational Texts | 1 | Reading (Reading Informational Text/Reading Vocabulary) | 4 | 8 | 0 | 0 | 4 | 8 | RI.1-3, 5-10, RI.4, L.4, L.5 | |
| | Totals | 3 | | 9 | 18 | 1 | 19 | 10 | 37 | | |
| 2 | Research Simulation Task | 3 | Reading: Reading Informational Text/ Reading Vocabulary* | 7 | 14 | 1 | 4 | 1 | 18 | RI.1-3, 5-10, RI.4, L.4, L.5 | 90 |
| | | | Writing: Written Expression | 0 | 0 | | 12 | | 12 | W.1-2, 4, 7- 10, | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 3 | | 3 | L.1, 2, plus language skills from previous grades | |
| | Totals | 3 | | 7 | 14 | 1 | 19 | 8 | 33 | | |
| 3 | Reading Literary Texts | 1 | Reading: Reading Literary Text/Reading Vocabulary* | 6 | 12 | 0 | 0 | 6 | 12 | RL.1-3, 5-10, RL.4, L.4, L.5 | 70 |
| | Reading Informational Texts | 2 | Reading: Reading Informational Text/Reading Vocab* | 6 | 12 | 0 | 0 | 6 | 12 | RI.1-3, 5, 7-10, RI.4, L.4, L.5 | |
| | Totals | 3 | | 12 | 24 | 0 | 0 | 12 | 24 | | |
| **Grade 8 Totals** | | **9** | Reading: Reading Literary Text/Reading Vocab* | 11 | 22 | 2 | 4 | 2 | 26 | 64 | 250 |
| | | | Reading: Reading Informational Text/Reading Vocab* | 17 | 34 | | 4 | | 38 | | |
| | | | Writing: Written Expression | 0 | 0 | | 24 | | 24 | 30 | |
| | | | Writing: Knowledge of Language and Conventions | 0 | 0 | | 6 | | 6 | | |
| | | | **Total** | 28 | 56 | 2 | 38 | 30 | 94 | 94 | |

**There must be at least eight points of reading vocabulary items on the test.*

**Table 3.7 Overview of LEAP Mathematics Task Types and Reporting Categories**

| Task Type | Description | Subclaim | Mathematical Practice(s) |
|---|---|---|---|
| **Type I** | conceptual understanding, fluency, and application | **Subclaim A:** solve problems involving the major content for the grade level.<br>**Subclaim B:** 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 | **Subclaim C:** 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 | **Subclaim D:** solve real-world problems engaging particularly in the modeling practice. | primarily MP.4 but may also involve any of the other practices |

**Table 3.8 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<br>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.G.A.1-2<br>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 3.9 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.4.1-11 |
| 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.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 3.10 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.5.1-3 |
| 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 LEAP 2025 Evidence Statements: LEAP.1.5.4-5 |
| 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 3.11 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 LEAP 2025 Evidence Statements: LEAP.I.6.1 |
| 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 3.12 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 3.13 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 | 28 | 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 LEAP 2025 Evidence Statement: LEAP.I.8.1 |
| Additional & Supporting Content | 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** | **38** | **40** | **4** | **14** | **3** | **12** | |
| | **TOTAL TASKS** | | **45** | **TOTAL POINTS** | | **66** | |

**Table 3.14 General Mathematics Test Structure—Grade 3**

| Reporting Category | Test Session | | | | | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | 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** | **15** | **21** | **14** | **20** | **14** | **21** | **43** | **62** |
| **Test Duration (minutes)** | **75** | | **75** | | **75** | | **225** | |

**Table 3.15 General Mathematics Test Structure—Grade 4**

| Reporting Category | Test Session | | | | | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | 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** | **15** | **21** | **14** | **20** | **14** | **21** | **43** | **62** |
| **Test Duration (minutes)** | **75** | | **75** | | **75** | | **225** | |

**Table 3.16 General Mathematics Test Structure—Grade 5**

| Reporting Category | Test Session | | | | | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | 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** | **15** | **21** | **14** | **20** | **14** | **21** | **43** | **62** |
| **Test Duration (minutes)** | **75** | | **75** | | **75** | | **225** | |

**Table 3.17 General Mathematics Test Structure—Grade 6**

| Reporting Category | Test Session | | | | | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | 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 | 6 | 1 | 6 | 3 | 12 |
| **TOTAL** | **16–20** | **20** | **12–13** | **23** | **11–13** | **23** | **43** | **66** |
| **Test Duration (minutes)** | **75** | | **75** | | **75** | | **225** | |

**Table 3.18 General Mathematics Test Structure—Grade 7**

| Reporting Category | Test Session | | | | | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | 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 | 6 | 1 | 6 | 3 | 12 |
| **TOTAL** | **16–20** | **20** | **12–14** | **23** | **11–13** | **23** | **43** | **66** |
| **Test Duration (minutes)** | **75** | | **75** | | **75** | | **225** | |

**Table 3.19 General Mathematics Test Structure—Grade 8**

| Reporting Category | Test Session | | | | | | TOTAL | |
|---|---|---|---|---|---|---|---|---|
| | 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 | 6 | 1 | 6 | 3 | 12 |
| **TOTAL** | **15–20** | **22** | **10–13** | **22** | **10–12** | **22** | **42** | **66** |
| **Test Duration (minutes)** | **75** | | **75** | | **75** | | **225** | |

**Table 3.20 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. |

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

| Claim | Subclaim | Subclaim 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 compose well-developed, organized, and clear writing, using details from provided texts. |
| | Knowledge and Use of Language Conventions | Students compose writing that correctly uses the rules of Standard English (including those for grammar, spelling, and usage). |

These reporting categories are the same as the reporting categories on the Spring 2017 ELA student reports and provide parents and educators with valuable information about

- overall student performance, including readiness to continue further studies 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 higher expectations.

**Table 3.21 Distribution of ELA Items and Points by Session and Item Type**

| | Sub | Gr | Session | EBSR | | MS | | TE | | PCR | | Total Pts |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Items | Pts | Items | Pts | Items | Pts | Items | Pts | |
| Paper -Pencil (PBT) | ELA | 3 | 1. Literary Analysis Task | 5 | 10 | | | | | 1 | 15 | 78 |
| | | | 2. Research Simulation Task | 5 | 10 | | | | | 1 | 15 | |
| | | | 3. Reading Informational and Literary Texts | 13 | 26 | 1 | 2 | | | | | |
| | ELA | 4 | 1. Literary Analysis Task/Reading Passage | 8 | 16 | 1 | 2 | | | 1 | 19 | 90 |
| | | | 2. Research Simulation Task | 5 | 10 | 2 | 4 | | | 1 | 19 | |
| | | | 3. Reading Literary/Informational Texts | 7 | 14 | 3 | 6 | | | | | |
| Online (CBT) | ELA | 4 | 1. Literary Analysis Task/Reading Passage | 7 | 14 | | | 2 | 4 | 1 | 19 | 90 |
| | | | 2. Research Simulation Task | 5 | 10 | 2 | 4 | | | 1 | 19 | |
| | | | 3. Reading Literary/Informational Texts | 6 | 12 | 2 | 4 | 2 | 4 | | | |
| | ELA | 5 | 1. Literary Analysis Task/Reading Passage | 6 | 12 | 1 | 2 | 2 | 4 | 1 | 19 | 90 |
| | | | 2. Research Simulation Task | 4 | 8 | 3 | 6 | | | 1 | 19 | |
| | | | 3. Reading Literary/Informational Texts | 7 | 14 | 2 | 4 | 1 | 2 | | | |
| | ELA | 6 | 1. Research Simulation Task | 6 | 12 | 1 | 2 | | | 1 | 19 | 96 |
| | | | 2. Narrative Writing Task/Reading Passage | 7 | 14 | 1 | 2 | 2 | 4 | 1 | 15 | |
| | | | 3. Reading Literary/Informational Texts | 11 | 22 | 1 | 2 | 2 | 4 | | | |
| | ELA | 7 | 1. Literary Analysis Task/Reading Passage | 7 | 14 | 1 | 2 | 1 | 2 | 1 | 19 | 94 |
| | | | 2. Research Simulation Task | 6 | 12 | | | 1 | 2 | 1 | 19 | |
| | | | 3. Reading Literary/Informational Texts | 8 | 16 | 3 | 6 | 1 | 2 | | | |
| | ELA | 8 | 1. Literary Analysis Task/Reading Passage | 5 | 10 | 2 | 4 | 2 | 4 | 1 | 19 | 94 |
| | | | 2. Research Simulation Task | 4 | 8 | 2 | 4 | 1 | 2 | 1 | 19 | |
| | | | 3. Reading Literary/Informational Texts | 11 | 22 | 1 | 2 | | | | | |
| | | | | | | | | | | | | |

**Table 3.22 Distribution of Points by Subclaim—Mathematics**

| | Grade | | | | | |
|---|---|---|---|---|---|---|
| **Subclaim** | **3** | **4** | **5** | **6** | **7** | **8** |
| Subclaim A | 30 | 30 | 30 | 30 | 30 | 30 |
| Subclaim B | 10 | 10 | 10 | 10 | 9 | 10 |
| Subclaim C | 10 | 10 | 10 | 14 | 14 | 14 |
| Subclaim D | 12 | 12 | 12 | 12 | 12 | 12 |
| **Total** | **62** | **62** | **62** | **66** | **65** | **66** |

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 table below, each of the three task types is aligned to one of four reporting categories (also called subclaims): Major Content (subclaim A), Additional and Supporting Content (subclaim B), Expressing Mathematical Reasoning (subclaim C), and Modeling and Application (subclaim D). Each task type is designed to align with at least one of the Standards for Mathematical Practice (MP) (see https://www.louisianabelieves.com/docs/common-core-state-standards-resources/guide---math-practices-bulleted.pdf?sfvrsn=2).

| Task Type | Description | Subclaim | Mathematical Practice(s) |
|---|---|---|---|
| **Type I** | conceptual understanding, fluency, and application | **Subclaim A:** solve problems involving the major content for the grade level.<br>**Subclaim B:** 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 | **Subclaim C:** 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 | **Subclaim D:** solve real-world problems engaging particularly in the modeling practice. | primarily MP.4 but may also involve any of the other practices |

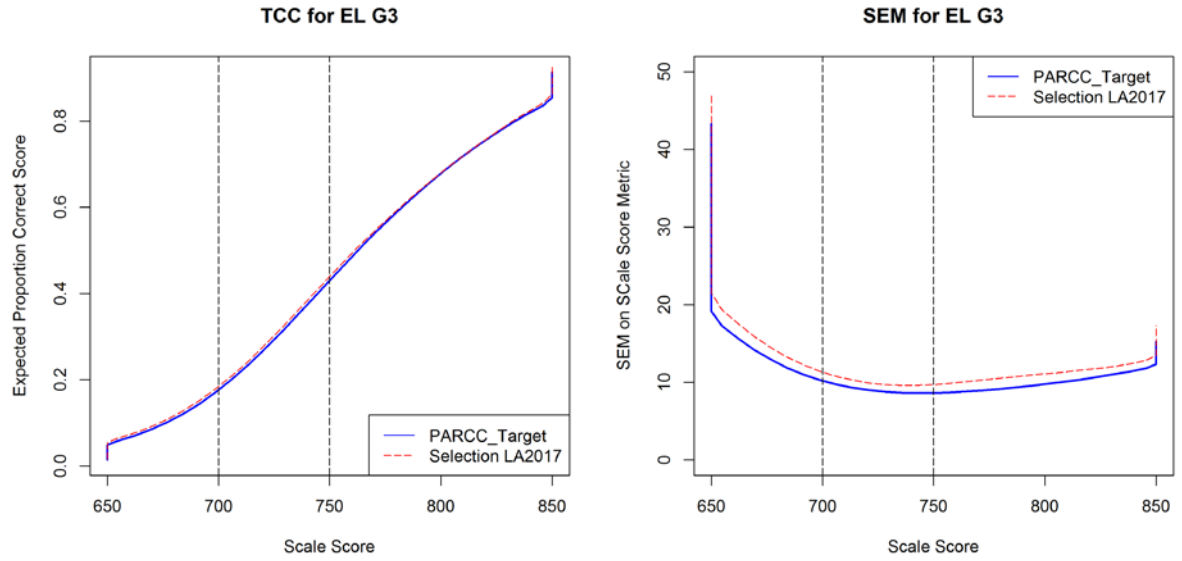
These reporting categories are the same as the reporting categories on the Spring 2016 Mathematics student reports and will provide parents and educators with valuable information about

- overall student performance, including readiness to continue further studies 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 3.23 Distribution of Mathematics Tasks and Points by Task Type**

| | Subject | Grade | Type I | | | Type II | | | Type III | | | Total Points |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | (1 pt) Tasks | (2 pts) Tasks | Points | (3 pts) Tasks | (4 pts) Tasks | Points | (3 pts) Tasks | (6 pts) Tasks | Points | |
| Paper-Pencil (PBT) | Math | 3 | 34 | 3 | 40 | 1 | 1 | 10 | 2 | 0 | 12 | 62 |
| | Math | 4 | 34 | 3 | 40 | 1 | 1 | 10 | 2 | 0 | 12 | 62 |
| Online (CBT) | 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 |

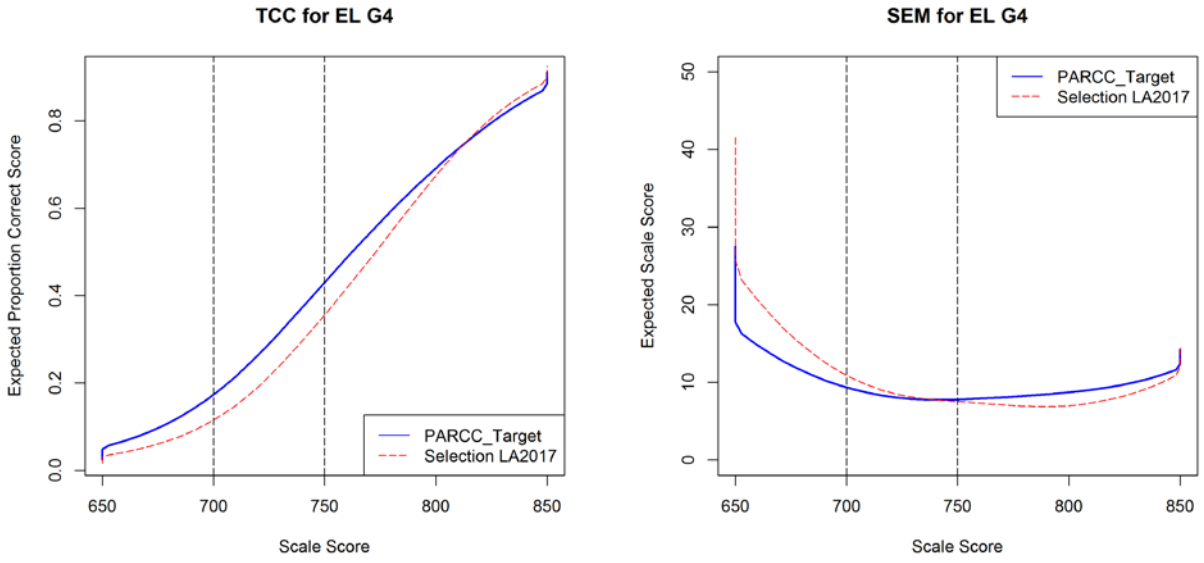
**Figure 3.1 2017 ELA Form Evaluation—Grade 3**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

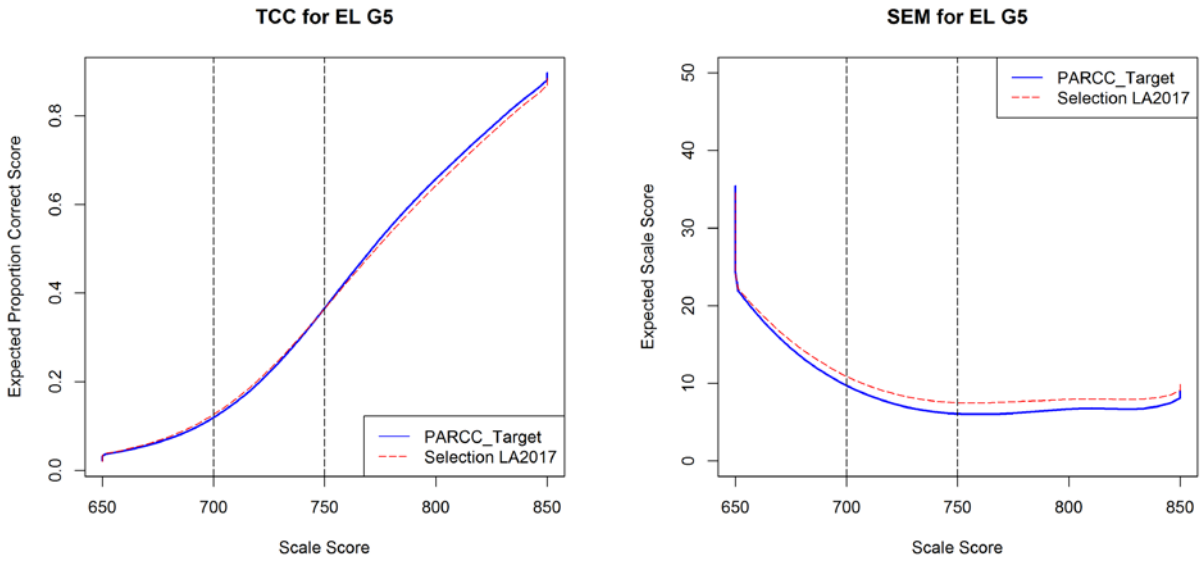
**Figure 3.2 2017 ELA Form Evaluation—Grade 4**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP test form.*

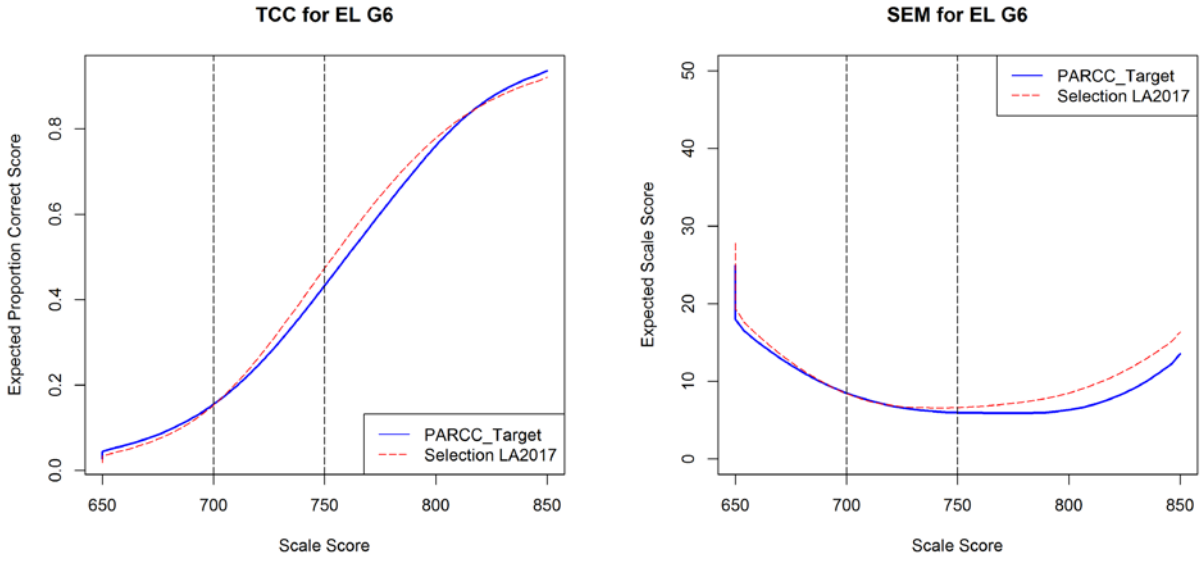
**Figure 3.3 2017 ELA Form Evaluation—Grade 5**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

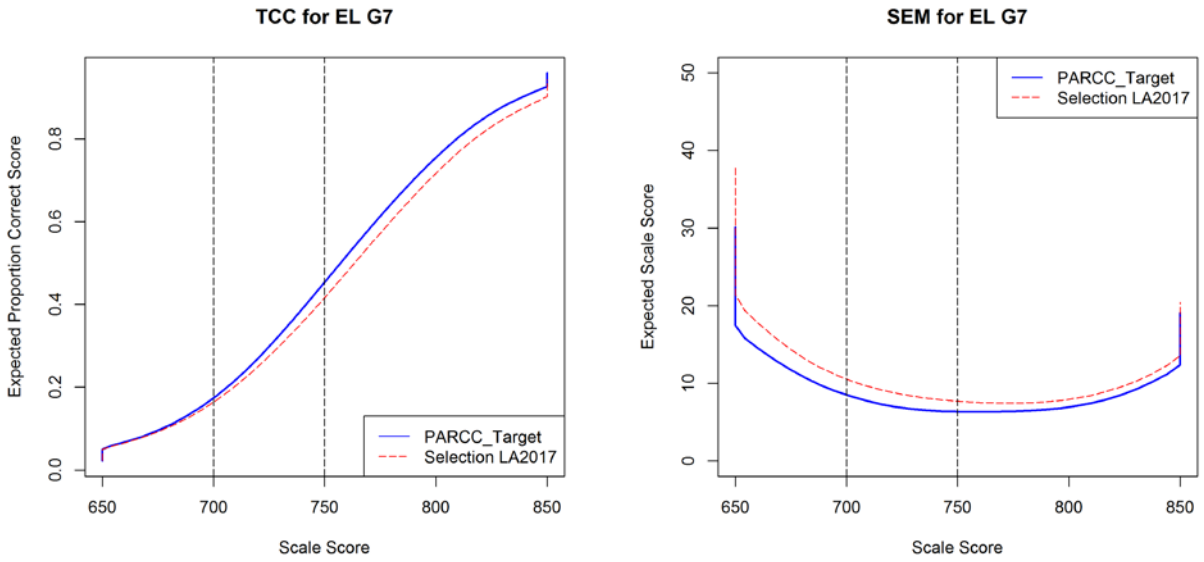
**Figure 3.4 2017 ELA Form Evaluation—Grade 6**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

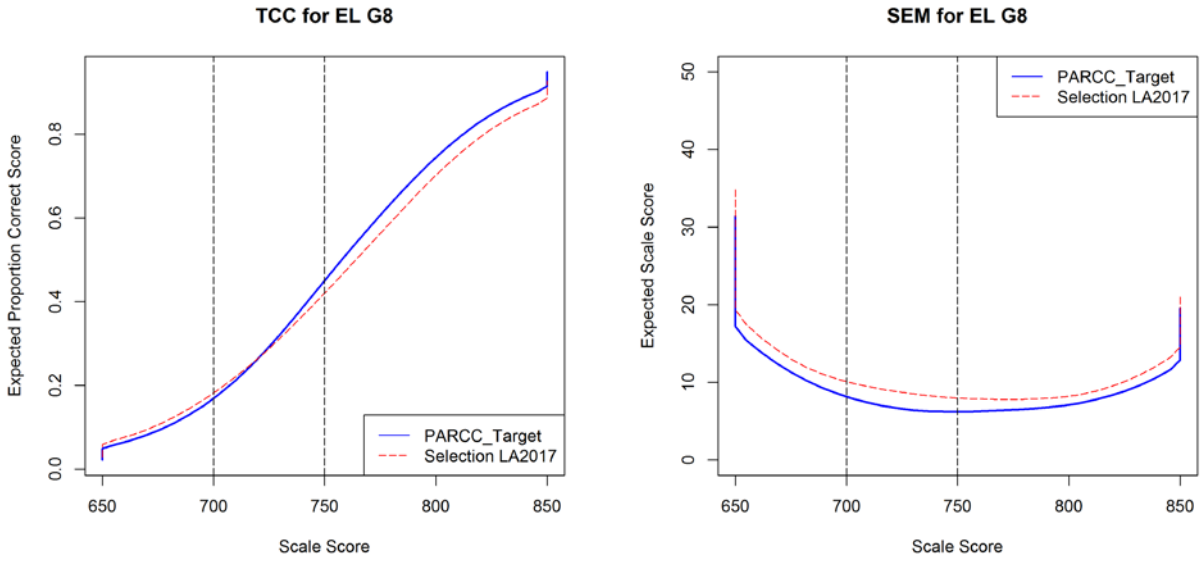
**Figure 3.5 2017 ELA Form Evaluation—Grade 7**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

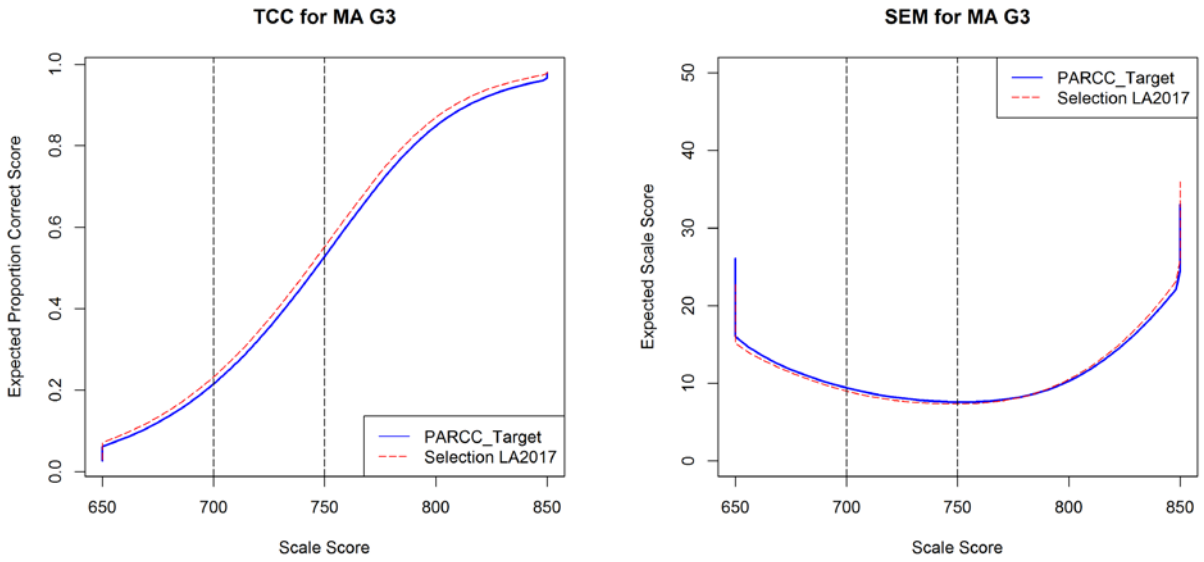
**Figure 3.6 2017 ELA Form Evaluation—Grade 8**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

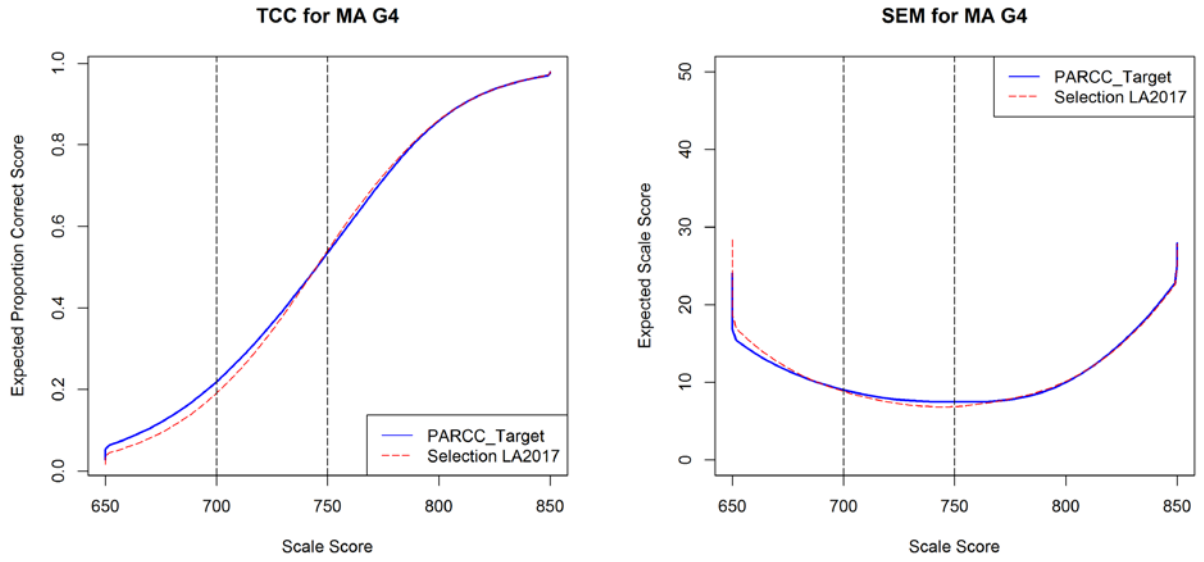
**Figure 3.7 2017 Mathematics Form Evaluation—Grade 3**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

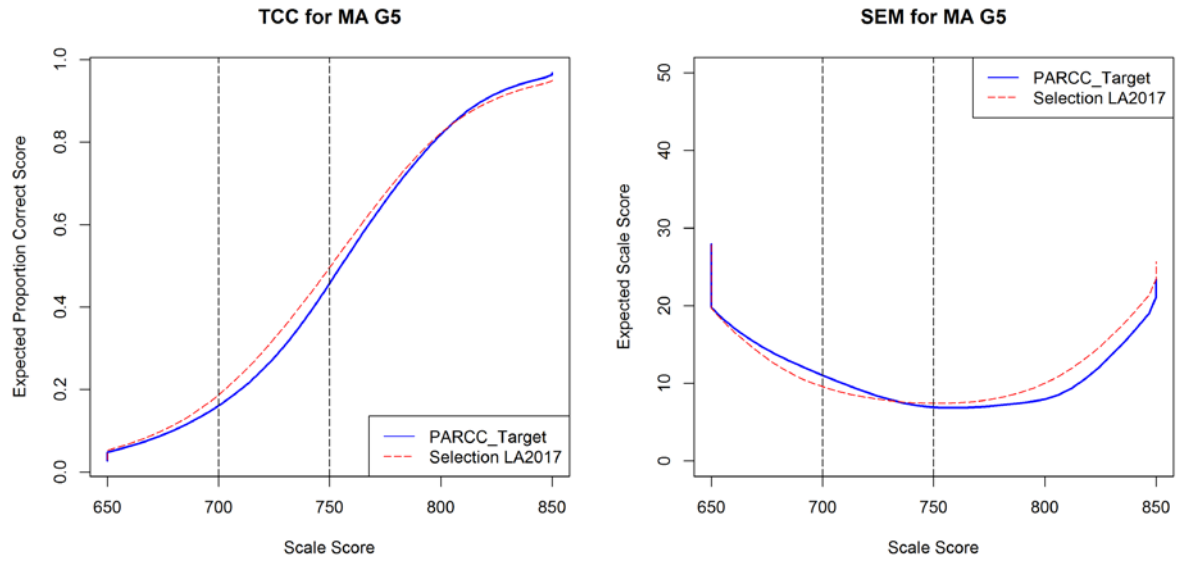
**Figure 3.8 2017 Mathematics Form Evaluation—Grade 4**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

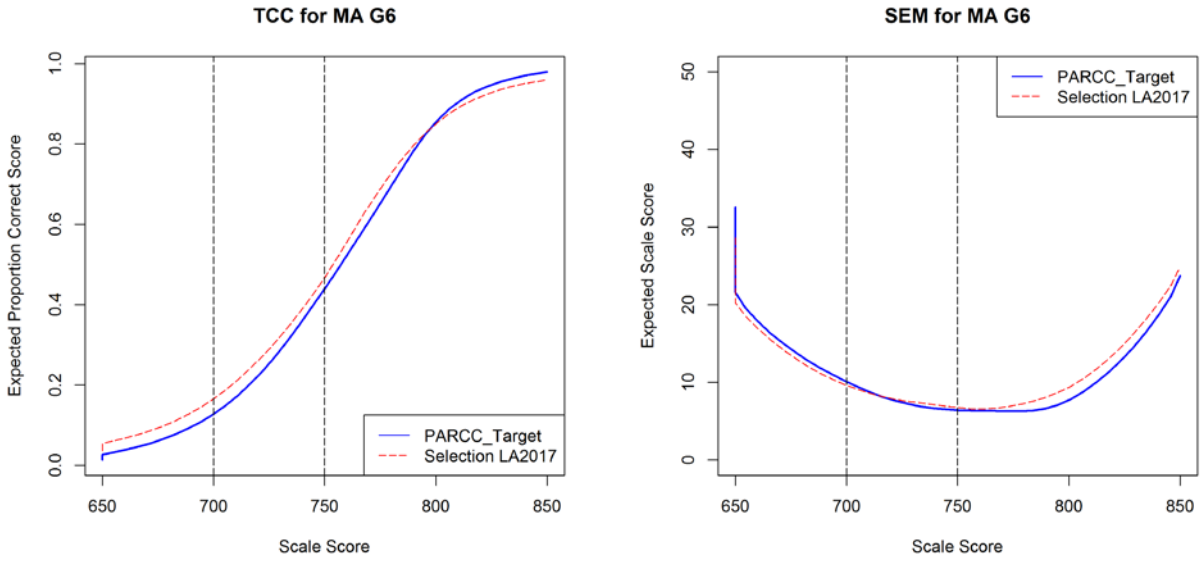
**Figure 3.9 2017 Mathematics Form Evaluation—Grade 5**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

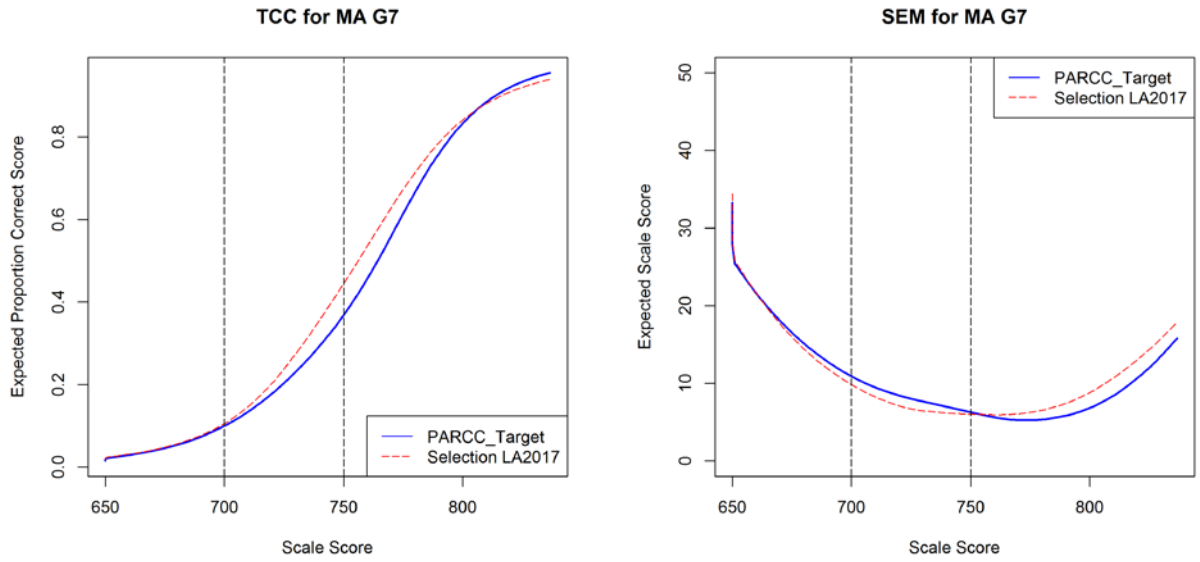
**Figure 3.10 2017 Mathematics Form Evaluation—Grade 6**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

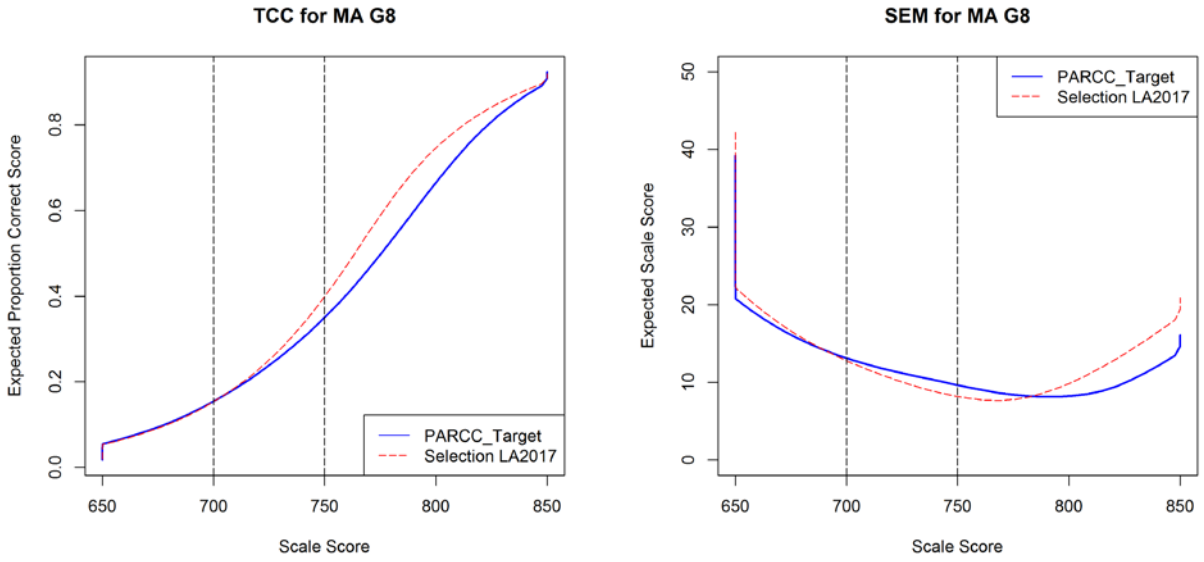
**Figure 3.11 2017 Mathematics Form Evaluation—Grade 7**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

**Figure 3.12 2017 Mathematics Form Evaluation—Grade 8**




*NOTE:*

- *PARCC_Target is the PARCC 2016 intact test form.*
- *Selection LA2017 is the selected 2017 LEAP 2025 test form.*

# CHAPTER 4: TEST ADMINISTRATION

Chapter 4 of the technical report describes the processes and activities 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 2017 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 adherence 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.

## 4.1 Training of School System Personnel

To ensure that the LEAP 2025 assessments are administered and scored in accordance with the department's mandates, LDOE takes a primary role in communicating with and training school system personnel. The development of the assessments is a collaborative effort between LDOE and DRC. 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 the administration standards must also meet the State Board of Education policies and the mandates of both state and federal legislation.

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

The school system 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 LDOE and their school systems. LDOE also provides assistance with and interpretation of assessment data and test results.

## 4.2 Ancillary Materials

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 2017 test administration of LEAP 2025, DRC produced an administration manual, the *LEAP 2025 Test Administration Manual* (TAM). DRC also produced a district *Test Coordinators Manual*. LDOE assessment administration and development staff review these

manuals, provide feedback, and give final approval. The *Test Coordinators Manual* is inclusive of ELA, mathematics, social studies, and science in grades 3 through 8. It provides detailed instructions for school system and school test coordinators' responsibilities for distributing and collecting test materials for the following programs and for returning them to DRC for scoring.

*Test Coordinators Manual* Table of Contents

1. Key Dates
2. Alerts
3. Oath of Security and Confidentiality Statement
4. General Information
5. Test Security
   - 5.1. Key Definitions
   - 5.2. Violations of Test Security
   - 5.3. Erasure Analysis
   - 5.4. Voiding Student Tests
6. Testing Guidelines
   - 6.1. Testing Eligibility
   - 6.2. Testing Conditions
   - 6.3. Testing in Class-sized Groups
   - 6.4. Test Schedule
   - 6.5. Extended Time for Testing
   - 6.6. Extended Breaks
   - 6.7. Makeup Testing
   - 6.8. Test Administration Resources
7. District Test Coordinator
   - 7.1. Conduct Training Session
   - 7.2. Receive Test Materials
   - 7.3. Large-print and Braille Test Materials
   - 7.4. Accommodated Materials
   - 7.5. Verify and Distribute Test Materials to School Test Coordinators
   - 7.6. Request Additional Test Materials and Bar-code Labels
   - 7.7. Collect Materials from Schools After Testing
   - 7.8. Used and Unused Answer Documents and Consumable Test Booklets (Defined)
   - 7.9. Unscorable Documents and Unscorable Document Labels
8. Directions for Returning Test Materials to DRC in May
   - 8.1. Pickup 1, 2, 3
   - 8.2. Final Checklist for Returning Test Materials to DRC
9. School Test Coordinator
   - 9.1. Receive and Verify Test Materials
   - 9.2. Conduct Test Administration and Security Training Session
   - 9.3. Supervise Application of Bar-code Labels and Coding of Answer Documents and Consumable Test Booklets
   - 9.4. Soiled, Damaged, and Other Unscorable Answer Documents and Consumable Test Booklets
   - 9.5. Verify and Distribute Materials to Test Administrators
   - 9.6. Supervise Test Administration
   - 9.7. Collect Test Materials
   - 9.8. Used and Unused Answer Documents and Consumable Test Booklets (Defined)

9.9. Coding Responsibilities of Principals—Before Testing
9.10. Coding Responsibilities of Principals—After Testing
10. Directions for Returning Test Materials to the DTC
10.1. Pickup 1, 2, 3
11. Void Notification
12. Index

The test administration manuals 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 test administration manuals is listed below.

Paper Administration Table of Contents

1. Spring Notes and Reminders
2. Test Administrator Oath of Security and Confidentiality Statement
3. Overview
4. Test Security
   4.1. Key Definitions
   4.2. Violations of Test Security
   4.3. Erasure Analysis
   4.4. Reporting Testing Irregularities and Security Breaches
   4.5. Voiding Student Tests
5. Test Administrator Responsibilities
6. Test Administration Checklists
   6.1. Before Testing
   6.2. During Testing
   6.3. After Testing (Daily)
   6.4. After Testing (Last Day)
7. Test Administrators' Frequently Asked Questions
8. Test Materials
   8.1. Receipt of Test Materials
9. Testing Guidelines
   9.1. Testing Eligibility
   9.2. Test Schedule
   9.3. Extended Time for Testing
   9.4. Testing Times for Grades X–X
   9.5. Makeup Testing
   9.6. Testing Conditions
10. Special Populations and Accommodations
    10.1. IDEA Special Education Students
    10.2. Students with One or More Disabilities According to Section 504
    10.3. Gifted and Talented Special Education Students
    10.4. Test Accommodations for Special Education and Section 504 Students
    10.5. Special Considerations for Deaf and Hard of Hearing Students
    10.6. Limited English Proficient (LEP) Students
11. Hand-coded Consumable Test Booklets and Answer Documents
12. Students Absent from Testing

10. Special Populations and Accommodations
    10.1. IDEA Special Education Students
    10.2. Students with One or More Disabilities According to Section 504
    10.3. Gifted and Talented Special Education Students
    10.4. Test Accommodations for Special Education and Section 504 Students
    10.5. Special Considerations for Deaf and Hard of Hearing Students
    10.6. Limited English Proficient (LEP) Students
11. Students Absent from Testing
12. Directions for Administering the Grades X–X Computer-Based LEAP 2025 Tests
13. Post-test Procedures
    13.1. Test Administrator Oath of Security and Confidentiality Statement
    13.2. Returning Test Materials to the School Test Coordinator
14. Index

The *Standards* contain multiple references that are relevant to test administration. Information in the *LEAP 2025 Test Administration Manual* addresses these standards.

Directions for test administration found in the manual addresses 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 Manual* provides 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 test administration 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 staff administer 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; 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 Individualized Education Program (IEP); allowing accommodations for students who do not have an IEP; 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 online test. 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 the online testing system (INSIGHT) 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 are presented under "Test Security" in the test administration manuals.

### 4.2.1 Return Material Forms and Guidelines

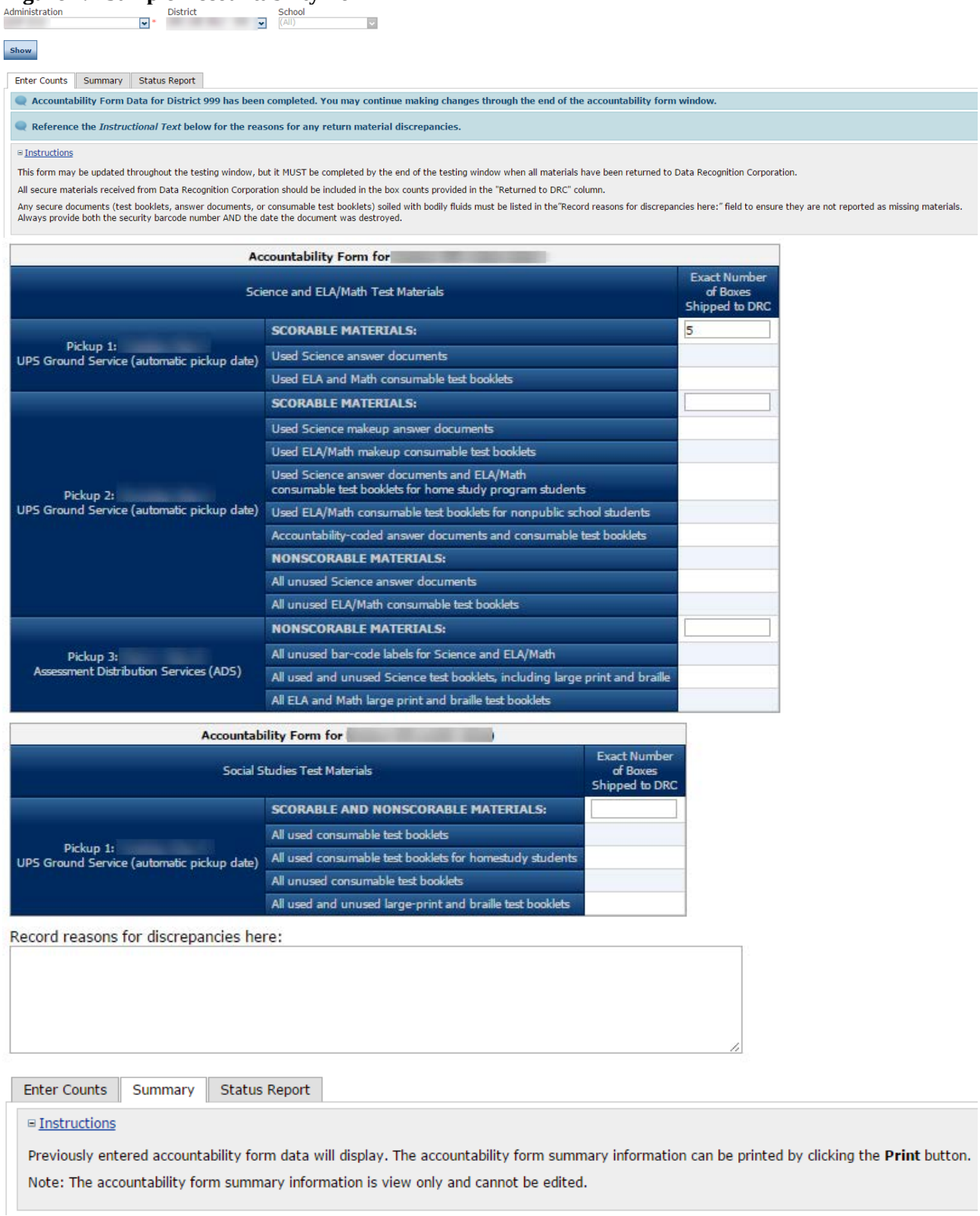
The *Test Coordinators Manual* instructs test coordinators on how to organize, pack, and return testing materials to DRC for secure inventory purposes. The LDOE assessment administration and development staff have opportunities to review these materials, provide feedback, and give final

approval. The purpose of the instructions is to ensure the secure test materials are properly accounted for and organized appropriately for return shipment.

### 4.2.2 Security Checklists

As soon as printed test materials are received by a school system, the district test coordinator ensures the first and last security barcodes on the tests match the packing list he or she received. The district test coordinator then packages the tests to be sent to schools. Upon returning test books to DRC, school and district test coordinators are required to complete and submit an accountability form that details the number of test books or printed test forms returned. This accountability form also requires that school systems and schools document nonstandard situations, including lost, damaged, destroyed, extra, or missing test books. A sample accountability form is shown in Figure 4.1.

**Figure 4.1 Sample Accountability Form**



| 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

Enter Counts | Summary | Status Report

Instructions

The progress status of the accountability form is displayed at the district level. Use this key to evaluate the status for your site:

- Not Started – District has not completed data entry
- Completed – District has completed data entry

The accountability form status can be exported to Excel by clicking the **Export to Excel** button.

**Click here to access a report of Users that clicked the Complete button and their information.**

| Overall Status for District | |
|---|---|
| District | Status |
| | Completed |

Export to Excel

### 4.2.3 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 from the LEAP 2025 administration. More details about the guide can be found in Chapter 7.

## 4.3 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. Test security procedures are discussed throughout the *Test Coordinators Manual* 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 and must account for all test materials and supervise the test administrators at all times.

## 4.4 Test Administration

The 2017 test was administered to students within the state testing window of April 3 through May 5, 2017. The paper testing window was May 1 through 5, 2017. Each session of the assessment within each content area of the LEAP 2025 was required to be administered in one block of time.

### 4.4.1 Time

Each session of the ELA and Mathematics LEAP 2025 tests was 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 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 | 75 |
| | 3 | 60 | 75 |
| 4 | 1 | 90 | 75 |
| | 2 | 90 | 75 |
| | 3 | 45 | 75 |
| 5 | 1 | 90 | 75 |
| | 2 | 90 | 75 |
| | 3 | 45 | 75 |
| 6 | 1 | 90 | 75 |
| | 2 | 90 | 75 |
| | 3 | 70 | 75 |
| 7 | 1 | 90 | 75 |
| | 2 | 90 | 75 |
| | 3 | 70 | 75 |
| 8 | 1 | 90 | 75 |
| | 2 | 90 | 75 |
| | 3 | 70 | 75 |

### 4.4.2 Accommodations

Accommodations are allowed on the LEAP 2025. Accommodations have been split into three areas: universal tools, designated supports, and accommodations.

- Universal tools are available to all students taking an assessment.
- Designated supports are available to students when deemed appropriate by a team of educators.
- Accommodations must appear in a student's IEP or 504 plan.

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.

Braille forms are provided for students who are visually challenged. Large-print forms are also available for visually challenged students using a paper-based test administration.

Tables 4.2 through 4.5 summarize the numbers of reportable students receiving accommodations by accommodation type for the 2017 LEAP 2025. Accommodations are grouped into four sections: special education accommodation, Limited English Proficient (LEP) status accommodation, Section 504 status accommodation, and online accommodation. The analyses are based on census data and include only those students who received accommodations and received a scale score on the ELA or Mathematics LEAP 2025.

**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 | ≥ 2,260 | 3.98% | ≥ 2,260 | 3.98% |
| 3 | Braille | < 50 | NR | < 50 | NR |
| 3 | Large Print | < 50 | NR | < 50 | NR |
| 3 | Answers Recorded | ≥ 540 | 0.96% | ≥ 550 | 0.97% |
| 3 | Extended Time | ≥ 4,770 | 8.41% | ≥ 4,740 | 8.36% |
| 3 | Transferred Answers | ≥ 240 | 0.43% | ≥ 230 | 0.42% |
| 3 | Individual/Small Group Administration | ≥ 4,740 | 8.35% | ≥ 4,710 | 8.3% |
| 3 | Tests Read Aloud | ≥ 3,600 | 6.35% | ≥ 4,050 | 7.15% |
| 4 | No Accommodation | ≥ 2,420 | 4.46% | ≥ 2,430 | 4.48% |
| 4 | Braille | < 50 | NR | < 50 | NR |
| 4 | Large Print | < 50 | NR | < 50 | NR |
| 4 | Answers Recorded | ≥ 480 | 0.9% | ≥ 470 | 0.87% |
| 4 | Extended Time | ≥ 4,850 | 8.95% | ≥ 4,870 | 8.97% |
| 4 | Transferred Answers | ≥ 250 | 0.47% | ≥ 250 | 0.46% |
| 4 | Individual/Small Group Administration | ≥ 4,760 | 8.77% | ≥ 4,770 | 8.79% |
| 4 | Tests Read Aloud | ≥ 3,690 | 6.8% | ≥ 4,220 | 7.77% |

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

| LEP Accommodation Type | | | | | |
|---|---|---|---|---|---|
| | | English Language Arts | | Mathematics | |
| Grade | Accommodation | Number | Percentage | Number | Percentage |
| 3 | No Accommodation | ≥ 490 | 0.88% | ≥ 480 | 0.86% |
| 3 | Extended Time | ≥ 1,940 | 3.43% | ≥ 1,940 | 3.43% |
| 3 | Individual/Small Group Administration | ≥ 1,710 | 3.01% | ≥ 1,710 | 3.02% |
| 3 | English/Native Language Word-to-Word Dictionary | ≥ 390 | 0.69% | ≥ 370 | 0.65% |
| 3 | Directions Read Aloud/Clarified in Native Language | ≥ 260 | 0.47% | ≥ 190 | 0.35% |
| 4 | No Accommodation | ≥ 300 | 0.55% | ≥ 290 | 0.55% |
| 4 | Extended Time | ≥ 1,680 | 3.1% | ≥ 1,670 | 3.08% |
| 4 | Individual/Small Group Administration | ≥ 1,470 | 2.71% | ≥ 1,450 | 2.68% |
| 4 | English/Native Language Word-to-Word Dictionary | ≥ 380 | 0.72% | ≥ 390 | 0.73% |
| 4 | Directions Read Aloud/Clarified in Native Language | ≥ 210 | 0.39% | ≥ 170 | 0.32% |

**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 | ≥ 320 | 0.57% | ≥ 300 | 0.53% |
| 3 | Large Print | < 50 | NR | < 50 | NR |
| 3 | Answers Recorded | < 50 | NR | < 50 | NR |
| 3 | Extended Time | ≥ 120 | 0.23% | ≥ 120 | 0.21% |
| 3 | Transferred Answers | ≥ 3,950 | 6.96% | ≥ 3,900 | 6.88% |
| 3 | Individual/Small Group Administration | ≥ 50 | 0.1% | ≥ 60 | 0.11% |
| 3 | Tests Read Aloud | ≥ 3,470 | 6.12% | ≥ 3,430 | 6.04% |
| 4 | No Accommodation | ≥ 1,750 | 3.08% | ≥ 2,310 | 4.07% |
| 4 | Large Print | ≥ 370 | 0.69% | ≥ 360 | 0.66% |
| 4 | Answers Recorded | < 50 | NR | < 50 | NR |
| 4 | Extended Time | < 50 | NR | < 50 | NR |
| 4 | Transferred Answers | ≥ 100 | 0.2% | ≥ 100 | 0.19% |
| 4 | Individual/Small Group Administration | ≥ 4,510 | 8.31% | ≥ 4,500 | 8.3% |
| 4 | Tests Read Aloud | ≥ 60 | 0.13% | ≥ 60 | 0.12% |

**Table 4.5 Number and Percentage of Students Receiving Online Accommodations by Accommodation Type, as valued in eDIRECT**

| Online Accommodation Type | | | | | |
|---|---|---|---|---|---|
| | | English Language Arts | | Mathematics | |
| Grade | Accommodation | Number | Percentage | Number | Percentage |
| 4 | Text-to-Speech | ≥ 130 | 7.17% | ≥ 310 | 16.1% |
| 4 | Human Read Aloud | ≥ 50 | 2.99% | ≥ 90 | 5.11% |
| 4 | Native Language Word-to-Word Dictionary | < 50 | NR | < 50 | NR |
| 4 | Directions in Native Language | < 50 | NR | < 50 | NR |
| 4 | Transferred Answers | < 50 | NR | < 50 | NR |
| 4 | Answers Recorded | < 50 | NR | < 50 | NR |
| 4 | Extended Time | ≥ 380 | 20.01% | ≥ 380 | 20.02% |
| 4 | Individual/Small Group Administration | ≥ 370 | 19.55% | ≥ 380 | 19.66% |
| 5 | Text-to-Speech | ≥ 5,370 | 10.08% | ≥ 8,500 | 15.95% |
| 5 | Human Read Aloud | ≥ 2,900 | 5.44% | ≥ 3,730 | 7% |
| 5 | Native Language Word-to-Word Dictionary | ≥ 460 | 0.87% | ≥ 460 | 0.86% |
| 5 | Directions in Native Language | ≥ 140 | 0.27% | ≥ 140 | 0.28% |
| 5 | Transferred Answers | ≥ 210 | 0.41% | ≥ 210 | 0.41% |
| 5 | Answers Recorded | ≥ 500 | 0.94% | ≥ 500 | 0.95% |
| 5 | Extended Time | ≥ 11,240 | 21.09% | ≥ 11,320 | 21.25% |
| 5 | Individual/Small Group Administration | ≥ 9,430 | 17.71% | ≥ 9,540 | 17.9% |
| 6 | Text-to-Speech | ≥ 5,430 | 10.38% | ≥ 7,660 | 14.64% |
| 6 | Human Read Aloud | ≥ 2,560 | 4.89% | ≥ 3,070 | 5.87% |
| 6 | Native Language Word-to-Word Dictionary | ≥ 620 | 1.2% | ≥ 620 | 1.2% |
| 6 | Directions in Native Language | ≥ 120 | 0.24% | ≥ 120 | 0.24% |
| 6 | Transferred Answers | ≥ 180 | 0.34% | ≥ 170 | 0.34% |
| 6 | Answers Recorded | ≥ 280 | 0.54% | ≥ 280 | 0.54% |
| 6 | Extended Time | ≥ 10,940 | 20.89% | ≥ 11,010 | 21.03% |
| 6 | Individual/Small Group Administration | ≥ 8,300 | 15.86% | ≥ 8,370 | 15.99% |
| 7 | Text-to-Speech | ≥ 5,220 | 10.06% | ≥ 7,370 | 14.23% |
| 7 | Human Read Aloud | ≥ 2,500 | 4.83% | ≥ 2,990 | 5.79% |
| 7 | Native Language Word-to-Word Dictionary | ≥ 720 | 1.4% | ≥ 720 | 1.4% |
| 7 | Directions in Native Language | ≥ 140 | 0.27% | ≥ 140 | 0.28% |
| 7 | Transferred Answers | ≥ 160 | 0.32% | ≥ 160 | 0.32% |
| 7 | Answers Recorded | ≥ 200 | 0.4% | ≥ 210 | 0.41% |
| 7 | Extended Time | ≥ 10,500 | 20.22% | ≥ 10,580 | 20.44% |
| 7 | Individual/Small Group Administration | ≥ 7,620 | 14.69% | ≥ 7,690 | 14.85% |
| 8 | Text-to-Speech | ≥ 4,450 | 8.84% | ≥ 6,310 | 14.13% |
| 8 | Human Read Aloud | ≥ 2,180 | 4.32% | ≥ 2,600 | 5.83% |
| 8 | Native Language Word-to-Word Dictionary | ≥ 710 | 1.41% | ≥ 700 | 1.58% |
| 8 | Directions in Native Language | ≥ 110 | 0.23% | ≥ 110 | 0.26% |
| 8 | Transferred Answers | ≥ 120 | 0.25% | ≥ 110 | 0.25% |

## 4.5 Summary

In summary, the overall purpose of each of the test administration workshops 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 providing guidance on appropriate interpretations of the test results. These communication and training efforts by 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: CONSTRUCTED-RESPONSE AND TECHNOLOGY-ENHANCED SCORING

In this chapter, the scoring process used for the 2017 LEAP 2025 ELA and Mathematics assessment is described, with a particular focus on the handscoring process 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.

### 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)

Sections 5.1.1 and 5.1.2 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.

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

### *The Recruitment and Interview Process*

All readers hired by DRC to score 2017 LEAP 2025 ELA and Mathematics responses had at least a four-year college degree in an appropriate field, such as a bachelor's degree in a STEM field or in English language arts.

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 also demonstrate his or her ability to apply the scoring criteria by qualifying at the end of the process is not allowed to score live student responses.

ELA and Mathematics scorers hired by DRC meet the degree requirements as specified by PARCC and conveyed by Pearson described in the following two paragraphs:

> Mathematics scorers had degrees in mathematics, science, engineering, education, or a related field; and/or teacher certification in a mathematics or science related field; or appropriate work experience that demonstrated proficiency in the subject area. Work experience outside of scoring was considered, as was college coursework. Three appropriate college-level courses (e.g., calculus, statistics, finance) were considered sufficient. Scorers of mathematics performance task responses had the mathematics knowledge needed to effectively score responses to PARCC Mathematics items.
>
> ELA scorers had degrees in reading, education, history, psychology, journalism, or a related area and/or teacher certification or other work experience that enabled them to succeed in scoring the Literary Analysis Task (LAT), Research Simulation Task (RST), or Narrative Writing Task (NWT). Work experience outside scoring was considered, as was college coursework. Following Pearson's practices for PARCC, three appropriate college-level courses with an emphasis on writing or literature were considered sufficient college-level experience for scoring ELA responses. Alternatively, successful completion of 250 hours of scoring was considered sufficient work experience for a scorer with insufficient college-level coursework to be hired to score Louisiana PARCC ELA responses.

### 5.1.2 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 2016 PARCC-approved training materials provided by PARCC, including the following:

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

***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. An explanation of the training and qualification procedures follows.

DRC used the PARCC-approved mathematics and ELA training and qualifying materials to score two categories of items: "prototype" items and "abbreviated" items.

***Prototype Items***

Some items included in the Louisiana forms were prototype items, meaning they had full sets of associated training materials, including anchor sets, practice sets, and qualifying sets. DRC started the training process with a review of passages and items, rubrics, and anchor sets, followed by the scoring and discussion of practice sets and qualifying sets. Once this process was completed for a prototype item included on the Louisiana form, 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 PARCC-approved materials for an associated prototype item that was similar to the abbreviated one they would be scoring on the Louisiana form.[1] Readers who did not qualify on the prototype item training were not allowed to continue the training.

---

[1] *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.*

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

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

**Table 5.1 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 |

**Table 5.2 ELA Training Set Composition**

| Set Type | Prototype Item Training | Abbreviated Item Training | Annotated |
|---|---|---|---|
| Anchor Set* | 3 responses per score point | 16 responses per item:<br>• Anchor Sets for abbreviated RST and LAT item training included scores for the combined trait Reading Comprehension and Written Expression (RCWE).<br>• Anchor Sets for abbreviated NWT item training included scores for Written Expression (WE). | Yes |
| Practice Set 1 | 5 responses representing the range of responses for<br>• the Reading Comprehension and Written Expression (RCWE) trait (for LAT and RST items)<br>• the Written Expression trait (for NWT items) | 10 responses representing the range of responses for both traits appropriate to the task type | Yes |
| Practice Set 2 | 5 responses representing the range of responses for the Knowledge of Language and Conventions trait | 10 responses representing the range of responses for both traits appropriate to the task type | 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 (included both traits appropriate to the task type) | Yes |

**For the ELA Knowledge of Language and Conventions trait, there were two mixed-prompt anchor sets per grade level (one for the narrative task and the other for 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 of Language and 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 of Language and Conventions trait during each prototype item qualifying session.*

***These PARCC-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.*

### *Qualifying Standards*

DRC followed the same qualification standards that Pearson used for PARCC. A description of these PARCC 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 5.3 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 5.4 Mathematics Qualifying Standards (Composite Items)***

| **Composite (multi-part) Items** | **Perfect Agreement** | **Perfect Plus Adjacent Agreement** |
|---|---|---|
| 0, 1 Rubric | 90% on two of three sets | 96% 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 5.5 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 5.5. 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).

### 5.1.3 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 5.1.3 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 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. If scoring patterns 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, he or she 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 using PARCC-approved validity responses supplied by PARCC. Validity responses were inserted among the live student responses. DRC used the same validity responses that Pearson used, which included both paper-based test (PBT) and computer-based test (CBT) 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 as well as 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.

- **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:

| Agreement Rate Requirements for Validity and Inter-Rater Reliability | | | |
|---|---|---|---|
| **Subject** | **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***

PARCC provided DRC with PARCC-approved calibration sets. 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 a reader scored a calibration set, the scoring director reviewed it from the front of the room, using rubric language and 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.

### 5.1.4 Security

Each DRC scoring center is a secure facility. All employees are issued photo identification badges and are required to wear them in plain view at all times. Access to scoring centers is limited to badge-wearing staff and to visitors accompanied by authorized staff. All readers are made aware that no scoring materials may leave the scoring center and must sign legally binding confidentiality agreements before work begins. DRC retains these agreements for the duration of

the contract. To prevent the unauthorized duplication of secure materials, cell phone and camera use within the scoring rooms is strictly forbidden. Readers only have access to the student responses they are qualified to score. Each scorer is assigned a unique username and password to access the DRC 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.

### 5.2 Technology-Enhanced Item Scoring Process

All technology-enhanced items were processed through DRC's autoscoring engine and scored according to the assigned scoring rules. 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 any 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 designated specific information 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 that are equivalent to technology-enhanced 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. 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 Inter-Rater Reliability

Approximately 10% of the responses in ELA and mathematics were scored independently by a second reader. The statistics for the 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 two readers was examined.

A total of 51 items were scored by human readers across all grades and content areas. The inter-rater reliability rates and the total numbers of reads are shown in Table 5.6 for ELA items, Table 5.7 for mathematics items, and Table 5.8 for Spanish mathematics items.

As shown in Table 5.6, raters demonstrated at least 97% perfect and adjacent agreement for all ELA handscored items. As shown in Table 5.7, raters demonstrated at least 97% perfect and adjacent agreement for mathematics items. As shown in Table 5.8, raters demonstrated 100% perfect and adjacent agreement for Spanish mathematics items.

**Table 5.6 Inter-Rater Agreement, English Language Arts Items**

| Grade | Task Type | Question | Trait | Total Reads | Read 2x | Inter-Rater Reliability % | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | EX | Adj | EX + Adj |
| 3 | Literary Analysis | 6 | Reading Comprehension and Written Expression | 63,255 | 12,314 | 81 | 18 | 99 |
| | | | Knowledge of Language and Conventions | 63,255 | 12,314 | 73 | 25 | 98 |
| 3 | Research Simulation | 12 | Reading Comprehension and Written Expression | 63,168 | 12,140 | 75 | 24 | 99 |
| | | | Knowledge of Language and Conventions | 63,168 | 12,140 | 76 | 24 | 100 |
| 4 | Literary Analysis | 6 | Reading Comprehension and Written Expression | 65,224 | 17,375 | 83 | 17 | 100 |
| | | | Knowledge of Language and Conventions | 65,224 | 17,375 | 82 | 18 | 100 |
| 4 | Research Simulation | 18 | Reading Comprehension and Written Expression | 64,741 | 16,414 | 78 | 22 | 100 |
| | | | Knowledge of Language and Conventions | 64,741 | 16,414 | 78 | 22 | 100 |
| 5 | Literary Analysis | 6 | Reading Comprehension and Written Expression | 60,223 | 13,679 | 76 | 23 | 99 |
| | | | Knowledge of Language and Conventions | 60,223 | 13,679 | 73 | 25 | 98 |
| 5 | Research Simulation | 18 | Reading Comprehension and Written Expression | 59,568 | 12,210 | 73 | 26 | 99 |
| | | | Knowledge of Language and Conventions | 59,568 | 12,210 | 72 | 27 | 99 |
| 6 | Research Simulation | 8 | Reading Comprehension and Written Expression | 57,897 | 10,928 | 72 | 27 | 99 |
| | | | Knowledge of Language and Conventions | 57,897 | 10,928 | 73 | 26 | 99 |
| 6 | Narrative Writing | 13 | Written Expression | 58,224 | 11,834 | 75 | 22 | 97 |
| | | | Knowledge of Language and Conventions | 58,224 | 11,834 | 72 | 27 | 99 |
| 7 | Literary Analysis | 6 | Reading Comprehension and Written Expression | 57,248 | 10,966 | 73 | 25 | 98 |
| | | | Knowledge of Language and Conventions | 57,248 | 10,966 | 74 | 24 | 98 |
| 7 | Research Simulation | 18 | Reading Comprehension and Written Expression | 57,320 | 10,916 | 72 | 27 | 99 |
| | | | Knowledge of Language and Conventions | 57,320 | 10,916 | 73 | 26 | 99 |
| 8 | Literary Analysis | 6 | Reading Comprehension and Written Expression | 55,800 | 10,996 | 70 | 28 | 98 |
| | | | Knowledge of Language and Conventions | 55,800 | 10,996 | 72 | 27 | 99 |
| 8 | Research Simulation | 18 | Reading Comprehension and Written Expression | 55,777 | 11,002 | 72 | 27 | 99 |
| | | | Knowledge of Language and Conventions | 55,777 | 11,002 | 78 | 21 | 99 |

**Table 5.7 Inter-Rater Agreement, Mathematics Items**

| Grade | Question | Part(s) | Total Reads | Read 2x | Inter-Rater Reliability % | | |
|---|---|---|---|---|---|---|---|
| | | | | | EX | Adj | EX + Adj |
| 3 | 14 | N/A | 63,053 | 12,008 | 86 | 12 | 98 |
| 3 | 15 | N/A | 62,797 | 11,495 | 86 | 12 | 98 |
| 3 | 28 | Part C | 62,873 | 11,648 | 87 | 13 | 100 |
| 3 | 29 | N/A | 62,907 | 11,716 | 93 | 6 | 99 |
| 3 | 42 | N/A | 62,828 | 11,558 | 79 | 19 | 98 |
| 3 | 43 | Part B | 62,802 | 11,506 | 95 | 4 | 99 |
| | | Part C | 62,802 | 11,506 | 89 | 10 | 99 |
| 4 | 14 | Part A | 62,234 | 11,456 | 94 | 6 | 100 |
| | | Part B | 62,234 | 11,456 | 95 | 5 | 100 |
| 4 | 15 | N/A | 62,221 | 11,442 | 88 | 11 | 99 |
| 4 | 28 | N/A | 62,319 | 11,638 | 85 | 13 | 98 |
| 4 | 29 | N/A | 62,230 | 11,474 | 85 | 14 | 9 |
| 4 | 42 | Part B (PBT) | 60,244 | 11,387 | 95 | 5 | 100 |
| 4 | 42 | Part B (CBT) | 2,198 | 472 | 98 | 2 | 100 |
| 4 | 43 | Part B | 60,145 | 11,190 | 97 | 2 | 99 |
| | | Part C | 60,145 | 11,190 | 94 | 5 | 99 |
| 4 | 43 | Part B | 2,177 | 436 | 98 | 2 | 100 |
| | | Part C | 2,177 | 436 | 89 | 11 | 100 |
| 5 | 14 | Part A | 58,729 | 11,112 | 84 | 15 | 99 |
| | | Part B | 58,729 | 11,112 | 90 | 9 | 99 |
| 5 | 15 | N/A | 57,946 | 11,056 | 84 | 15 | 99 |
| 5 | 28 | Part A | 58,837 | 11,104 | 79 | 21 | 100 |
| | | Part B | 58,837 | 11,104 | 93 | 7 | 100 |
| 5 | 29 | N/A | 58,655 | 11,338 | 84 | 15 | 99 |
| 5 | 42 | Part A | 58,776 | 11,374 | 95 | 5 | 100 |
| | | Part B | 58,776 | 11,374 | 98 | 2 | 100 |
| 5 | 43 | Part A | 58,738 | 11,233 | 89 | 10 | 99 |
| | | Part B | 58,738 | 11,233 | 86 | 13 | 99 |

**Table 5.7 Inter-Rater Agreement, Mathematics Items, continued**

| Grade | Question | Part(s) | Total Reads | Read 2x | Inter-Rater Reliability % | | |
|---|---|---|---|---|---|---|---|
| | | | | | EX | Adj | EX + Adj |
| 6 | 29 | N/A | 57,169 | 11,080 | 90 | 9 | 99 |
| 6 | 30 | N/A | 57,416 | 11,174 | 82 | 16 | 98 |
| 6 | 31 | N/A | 57,017 | 10,796 | 95 | 5 | 100 |
| 6 | 32 | N/A | 56,241 | 11,165 | 93 | 7 | 100 |
| 6 | 41 | Part B | 57,656 | 10,810 | 94 | 6 | 100 |
| 6 | 42 | Part B | 57,496 | 10,520 | 89 | 10 | 99 |
| 6 | 43 | Part A | 57,292 | 10,800 | 91 | 7 | 98 |
| | | Part B | 57,292 | 10,800 | 93 | 5 | 98 |
| 7 | 29 | N/A | 56,422 | 10,988 | 97 | 3 | 100 |
| 7 | 30 | N/A | 55,725 | 11,600 | 91 | 8 | 99 |
| 7 | 31 | N/A | 55,047 | 10,720 | 96 | 3 | 99 |
| 7 | 32 | Part A | 55,053 | 10,316 | 94 | 6 | 100 |
| | | Part B | 55,053 | 10,316 | 98 | 2 | 100 |
| | | Part C | 55,053 | 10,316 | 95 | 5 | 100 |
| 7 | 41 | N/A | 56,769 | 10,842 | 89 | 11 | 100 |
| 7 | 42 | Part B | 56,854 | 10,374 | 96 | 4 | 100 |
| | | Part C | 56,854 | 10,374 | 97 | 3 | 100 |
| 7 | 43 | Part C | 56,392 | 10,348 | 96 | 3 | 99 |
| | | Part D | 56,392 | 10,348 | 91 | 9 | 100 |
| 8 | 29 | N/A | 48,564 | 9,878 | 90 | 10 | 100 |
| 8 | 30 | N/A | 47,961 | 10,027 | 92 | 8 | 100 |
| 8 | 31 | N/A | 48,871 | 9,724 | 91 | 8 | 99 |
| 8 | 32 | N/A | 47,922 | 9,714 | 93 | 7 | 100 |
| 8 | 40 | Part B | 49,261 | 9,066 | 94 | 5 | 99 |
| 8 | 41 | Part B | 48,811 | 8,978 | 97 | 2 | 99 |
| 8 | 42 | Part A | 48,826 | 9,566 | 87 | 12 | 99 |
| | | Part B | 48,826 | 9,566 | 92 | 8 | 100 |

**Table 5.8 Inter-Rater Agreement, Spanish Mathematics Items**

| Grade | Question | Part(s) | Total Reads | Read 2x | Inter-Rater Reliability % | | |
|---|---|---|---|---|---|---|---|
| | | | | | EX | Adj | EX + Adj |
| 3 | 14 | N/A | 51 | 14 | 100 | 0 | 100 |
| 3 | 15 | N/A | 51 | 14 | 100 | 0 | 100 |
| 3 | 28 | Part C | 51 | 14 | 100 | 0 | 100 |
| 3 | 29 | N/A | 49 | 10 | 100 | 0 | 100 |
| 3 | 42 | N/A | 50 | 12 | 100 | 0 | 100 |
| 3 | 43 | Part B | 51 | 14 | 100 | 0 | 100 |
| | | Part C | 51 | 14 | 100 | 0 | 100 |
| 4 | 14 | Part A | 45 | 12 | 100 | 0 | 100 |
| | | Part B | 45 | 12 | 100 | 0 | 100 |
| 4 | 15 | N/A | 45 | 12 | 100 | 0 | 100 |
| 4 | 28 | N/A | 46 | 14 | 86 | 14 | 100 |
| 4 | 29 | N/A | 45 | 12 | 100 | 0 | 100 |
| 4 | 42 | Part B | 44 | 12 | 100 | 0 | 100 |
| 4 | 42 | Part B | 2 | 2 | 100 | 0 | 100 |
| 4 | 43 | Part B | 44 | 12 | 100 | 0 | 100 |
| | | Part C | 44 | 12 | 100 | 0 | 100 |
| 4 | 43 | Part B | 2 | 2 | 100 | 0 | 100 |
| | | Part C | 2 | 2 | 100 | 0 | 100 |
| 5 | 14 | Part A | 93 | 22 | 100 | 0 | 100 |
| | | Part B | 93 | 22 | 100 | 0 | 100 |
| 5 | 15 | N/A | 94 | 24 | 100 | 0 | 100 |
| 5 | 28 | Part A | 93 | 26 | 100 | 0 | 100 |
| | | Part B | 93 | 26 | 100 | 0 | 100 |
| 5 | 29 | N/A | 92 | 22 | 100 | 0 | 100 |
| 5 | 42 | Part A | 91 | 21 | 100 | 0 | 100 |
| | | Part B | 91 | 21 | 100 | 0 | 100 |
| 5 | 43 | Part A | 95 | 28 | 100 | 0 | 100 |
| | | Part B | 95 | 28 | 100 | 0 | 100 |

**Table 5.8 Inter-Rater Agreement, Spanish Mathematics Items, continued**

| Grade | Question | Part(s) | Total Reads | Read 2x | Inter-Rater Reliability % | | |
|---|---|---|---|---|---|---|---|
| | | | | | EX | Adj | EX + Adj |
| 6 | 29 | N/A | 141 | 38 | 95 | 5 | 100 |
| 6 | 30 | N/A | 140 | 38 | 100 | 0 | 100 |
| 6 | 31 | N/A | 139 | 34 | 100 | 0 | 100 |
| 6 | 32 | N/A | 139 | 38 | 100 | 0 | 100 |
| 6 | 41 | Part B | 139 | 34 | 100 | 0 | 100 |
| 6 | 42 | Part B | 137 | 30 | 100 | 0 | 100 |
| 6 | 43 | Part A | 143 | 44 | 100 | 0 | 100 |
| | | Part B | 143 | 44 | 100 | 0 | 100 |
| 7 | 29 | N/A | 160 | 36 | 100 | 0 | 100 |
| 7 | 30 | N/A | 153 | 40 | 100 | 0 | 100 |
| 7 | 31 | N/A | 160 | 48 | 100 | 0 | 100 |
| 7 | 32 | Part A | 158 | 44 | 95 | 5 | 100 |
| | | Part B | 158 | 44 | 100 | 0 | 100 |
| | | Part C | 158 | 44 | 100 | 0 | 100 |
| 7 | 41 | N/A | 161 | 42 | 100 | 0 | 100 |
| 7 | 42 | Part B | 163 | 36 | 100 | 0 | 100 |
| | | Part C | 163 | 36 | 100 | 0 | 100 |
| 7 | 43 | Part C | 164 | 42 | 95 | 5 | 100 |
| | | Part D | 164 | 42 | 100 | 0 | 100 |
| 8 | 29 | N/A | 133 | 34 | 100 | 0 | 100 |
| 8 | 30 | N/A | 135 | 44 | 100 | 0 | 100 |
| 8 | 31 | N/A | 138 | 44 | 100 | 0 | 100 |
| 8 | 32 | N/A | 131 | 30 | 100 | 0 | 100 |
| 8 | 40 | Part B | 139 | 34 | 100 | 0 | 100 |
| 8 | 41 | Part B | 133 | 24 | 100 | 0 | 100 |
| 8 | 42 | Part A | 132 | 32 | 100 | 0 | 100 |
| | | Part B | 132 | 32 | 100 | 0 | 100 |

## 5.5 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 are 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, the IRT models used for calibrating the data are discussed and 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 presented.

Chapter 6 demonstrates adherence in the LEAP 2025 program 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 2017 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.

## 6.1 Classical Item Statistics

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.

### 6.1.1 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 0.36 to 0.46. Average form difficulty for mathematics ranged from 0.31 to 0.52. In general, 2017 LEAP 2025 tests were relatively difficult tests across all subjects and grades. For ELA, the grade 4 computer-based test (CBT) was the most difficult, with 0.36 average form difficulty, and the grade 8 test was the easiest, with 0.46 average form difficulty. For mathematics, the grade 8 test was the most

difficult, with 0.31 average form difficulty, and the grade 3 test was the easiest, with 0.52 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 | PBT | 30 | 72 | 31.96 (14.38) | 0.43 (0.16) |
| | 4 | CBT | 32 | 82 | 31.18 (15.19) | 0.36 (0.12) |
| | 4 | PBT | 32 | 82 | 31.94 (14.50) | 0.37 (0.12) |
| | 5 | CBT | 32 | 82 | 32.61 (14.69) | 0.38 (0.15) |
| | 6 | CBT | 36 | 92 | 36.61 (18.08) | 0.39 (0.13) |
| | 7 | CBT | 34 | 86 | 39.87 (18.20) | 0.43 (0.11) |
| | 8 | CBT | 34 | 86 | 42.52 (17.84) | 0.46 (0.14) |
| Mathematics | 3 | PBT | 43 | 62 | 31.61 (12.74) | 0.52 (0.24) |
| | 4 | CBT | 42 | 59 | 29.55 (12.03) | 0.46 (0.21) |
| | 4 | PBT | 42 | 59 | 31.07 (12.39) | 0.49 (0.20) |
| | 5 | CBT | 42 | 61 | 25.95 (11.20) | 0.41 (0.22) |
| | 6 | CBT | 41 | 64 | 25.49 (12.55) | 0.38 (0.22) |
| | 7 | CBT | 43 | 66 | 22.50 (13.18) | 0.34 (0.18) |
| | 8 | CBT | 42 | 66 | 20.46 (11.37) | 0.31 (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 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 Raw Scores, *p*-Values, Item-Total Correlation ($R_{it}$)**

| | | | | Item *p*-Value | | | | Average Total Correlation | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Content | Grade | Mode | N of Items | Mean | Std. Dev. | Min. | Max | Mean | Std. Dev. | Min. | Max |
| ELA | 3 | PBT | 30 | 0.47 | 0.16 | 0.17 | 0.78 | 0.44 | 0.11 | 0.13 | 0.58 |
| | 4 | CBT | 32 | 0.40 | 0.12 | 0.13 | 0.68 | 0.42 | 0.12 | 0.21 | 0.63 |
| | 4 | PBT | 32 | 0.40 | 0.12 | 0.17 | 0.70 | 0.40 | 0.12 | 0.14 | 0.57 |
| | 5 | CBT | 32 | 0.42 | 0.15 | 0.21 | 0.85 | 0.41 | 0.10 | 0.28 | 0.61 |
| | 6 | CBT | 36 | 0.42 | 0.14 | 0.21 | 0.75 | 0.45 | 0.11 | 0.22 | 0.66 |
| | 7 | CBT | 34 | 0.44 | 0.12 | 0.19 | 0.75 | 0.41 | 0.14 | 0.14 | 0.69 |
| | 8 | CBT | 34 | 0.46 | 0.16 | 0.27 | 0.82 | 0.39 | 0.15 | 0.12 | 0.70 |
| Mathematics | 3 | PBT | 43 | 0.60 | 0.24 | 0.19 | 0.98 | 0.45 | 0.12 | 0.17 | 0.75 |
| | 4 | CBT | 42 | 0.50 | 0.21 | 0.16 | 0.91 | 0.46 | 0.12 | 0.21 | 0.68 |
| | 4 | PBT | 42 | 0.52 | 0.21 | 0.17 | 0.91 | 0.45 | 0.12 | 0.21 | 0.68 |
| | 5 | CBT | 42 | 0.47 | 0.22 | 0.14 | 0.88 | 0.40 | 0.14 | 0.03 | 0.67 |
| | 6 | CBT | 41 | 0.44 | 0.23 | 0.10 | 0.93 | 0.44 | 0.13 | 0.16 | 0.68 |
| | 7 | CBT | 43 | 0.37 | 0.19 | 0.05 | 0.78 | 0.45 | 0.14 | 0.09 | 0.74 |
| | 8 | CBT | 42 | 0.35 | 0.19 | 0.10 | 0.75 | 0.40 | 0.12 | 0.19 | 0.64 |

### 6.1.2 Item-Level Statistics

Tables 6.3–6.9 present the item statistics for each item included in regular test forms by grade for ELA. Tables 6.10–6.16 show the item statistics for each item included in regular test forms 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 omits), and the percentage at each score point, if applicable, for each item by grade and content area.

**p-*Value***

The *p*-value is a measure of item difficulty. For a multiple-choice (MC) item, the *p*-value is calculated from the number of students who correctly responded to an item divided 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 from the average score for the item divided 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 in the mid-0.50s to 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 0.40 in grade 4 to 0.47 in grade 3. The average *p*-values associated with the mathematics forms range from 0.35 in grade 8 to 0.60 in grade 3.

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 the 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

both low-scoring and high-scoring students, as well as 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 the 0.13 to the 0.85 (see Tables 6.3–6.9) across all grade levels. The *p*-values on the mathematics forms range from the 0.05 to 0.98 (see Tables 6.10–6.16). Such a broad range of *p*-values, which indicates the items measure well throughout the range of skills 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 and the total test score, where the item score is not included in the total score. It indicates how well an item differentiates between low-scoring and high-scoring students. 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.

This examination of omit rates complies with Standard 4.14 of the *Standards*. This standard is concerned with 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.

The results presented in Tables 6.3–6.16 show that the omit rates for 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.

**Table 6.3 Item Statistics—English Language Arts Grade 3 Paper-Based Test Administration**

| ELA Grade 3 Paper-Based Test Administration | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Item | Item Type | Total N | Adj. N | *p*-Value | Pbis | Omit Rate | % at 0 | % at 1 | % at 2 | % at 3 |
| 1 | ESR | ≥ 56,800 | ≥ 56,540 | 0.48 | 0.50 | 0.45 | 47.96 | 8.27 | 43.33 | |
| 2 | ESR | ≥ 56,800 | ≥ 56,500 | 0.73 | 0.52 | 0.52 | 23.27 | 6.25 | 69.96 | |
| 3 | ESR | ≥ 56,800 | ≥ 56,480 | 0.58 | 0.50 | 0.56 | 35.99 | 11.66 | 51.79 | |
| 4 | ESR | ≥ 56,800 | ≥ 56,370 | 0.56 | 0.32 | 0.76 | 37.24 | 12.47 | 49.53 | |
| 5 | ESR | ≥ 56,800 | ≥ 56,360 | 0.75 | 0.38 | 0.77 | 20.19 | 8.29 | 70.74 | |
| 6 | CR | ≥ 56,800 | ≥ 55,870 | 0.30 | 0.47 | 1.08 | 28.65 | 53.81 | 14.28 | 1.62 |
| 6 | CR | ≥ 56,800 | ≥ 55,870 | 0.38 | 0.54 | 1.08 | 17.65 | 51.17 | 26.89 | 2.64 |
| 7 | ESR | ≥ 56,800 | ≥ 56,550 | 0.60 | 0.50 | 0.43 | 24.20 | 31.91 | 43.46 | |
| 8 | ESR | ≥ 56,800 | ≥ 56,500 | 0.42 | 0.39 | 0.52 | 48.12 | 19.07 | 32.29 | |
| 9 | ESR | ≥ 56,800 | ≥ 56,470 | 0.17 | 0.29 | 0.58 | 75.39 | 14.57 | 9.46 | |
| 10 | ESR | ≥ 56,800 | ≥ 56,450 | 0.40 | 0.43 | 0.61 | 43.17 | 32.99 | 23.23 | |
| 11 | ESR | ≥ 56,800 | ≥ 56,330 | 0.43 | 0.44 | 0.82 | 51.95 | 9.60 | 37.63 | |
| 12 | CR | ≥ 56,800 | ≥ 56,000 | 0.24 | 0.57 | 0.84 | 41.60 | 42.95 | 12.92 | 1.13 |
| 12 | CR | ≥ 56,800 | ≥ 56,000 | 0.34 | 0.58 | 0.84 | 25.73 | 46.49 | 23.69 | 2.68 |
| 13 | ESR | ≥ 56,800 | ≥ 56,490 | 0.56 | 0.47 | 0.54 | 26.42 | 35.02 | 38.03 | |
| 14 | ESR | ≥ 56,800 | ≥ 56,260 | 0.38 | 0.32 | 0.95 | 54.41 | 14.95 | 29.68 | |
| 15 | MS | ≥ 56,800 | ≥ 56,360 | 0.36 | 0.20 | 0.77 | 31.71 | 64.45 | 3.08 | |
| 16 | ESR | ≥ 56,800 | ≥ 56,170 | 0.63 | 0.51 | 1.11 | 28.98 | 15.87 | 54.04 | |
| 17 | ESR | ≥ 56,800 | ≥ 56,180 | 0.50 | 0.45 | 1.08 | 45.64 | 8.56 | 44.71 | |
| 18 | ESR | ≥ 56,800 | ≥ 56,130 | 0.55 | 0.52 | 1.17 | 39.39 | 9.43 | 50.02 | |
| 19 | ESR | ≥ 56,800 | ≥ 55,920 | 0.78 | 0.47 | 1.54 | 17.18 | 8.40 | 72.88 | |
| 20 | ESR | ≥ 56,800 | ≥ 55,950 | 0.33 | 0.29 | 1.48 | 55.11 | 21.24 | 22.17 | |
| 21 | ESR | ≥ 56,800 | ≥ 55,550 | 0.57 | 0.53 | 2.20 | 38.18 | 6.82 | 52.80 | |
| 22 | ESR | ≥ 56,800 | ≥ 55,810 | 0.41 | 0.43 | 1.73 | 47.29 | 22.08 | 28.90 | |
| 23 | ESR | ≥ 56,800 | ≥ 55,520 | 0.50 | 0.45 | 2.24 | 43.60 | 11.10 | 43.06 | |
| 24 | ESR | ≥ 56,800 | ≥ 55,620 | 0.48 | 0.44 | 2.07 | 43.26 | 15.95 | 38.72 | |
| 25 | ESR | ≥ 56,800 | ≥ 55,100 | 0.17 | 0.13 | 2.99 | 77.06 | 7.88 | 12.08 | |
| 26 | ESR | ≥ 56,800 | ≥ 54,930 | 0.54 | 0.52 | 3.29 | 35.79 | 18.12 | 42.79 | |

**Table 6.4 Item Statistics—English Language Arts Grade 4 Computer-Based Test Administration**

| ELA Grade 4 Computer-Based Test Administration | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | Item Type | Total N | Adj. N | *p*-Value | Pbis | Omit Rate | % at 0 | % at 1 | % at 2 | % at 3 | % at 4 |
| 1 | ESR | ≥ 1,930 | ≥ 1,930 | 0.43 | 0.43 | 0.05 | 53.95 | 6.45 | 39.56 | | |
| 2 | ESR | ≥ 1,930 | ≥ 1,930 | 0.59 | 0.48 | 0.26 | 27.23 | 26.92 | 45.59 | | |
| 3 | TE | ≥ 1,930 | ≥ 1,930 | 0.58 | 0.55 | 0.31 | 33.57 | 17.02 | 49.10 | | |
| 4 | ESR | ≥ 1,930 | ≥ 1,930 | 0.46 | 0.24 | 0.36 | 49.30 | 9.49 | 40.85 | | |
| 5 | ESR | ≥ 1,930 | ≥ 1,930 | 0.41 | 0.37 | 0.10 | 56.47 | 4.64 | 38.78 | | |
| 6 | CR | ≥ 1,930 | ≥ 1,920 | 0.21 | 0.61 | 0.26 | 34.61 | 47.65 | 14.96 | 2.27 | |
| 6 | CR | ≥ 1,930 | ≥ 1,920 | 0.29 | 0.61 | 0.26 | 34.50 | 46.47 | 16.35 | 2.17 | |
| 7 | ESR | ≥ 1,930 | ≥ 1,920 | 0.42 | 0.35 | 0.62 | 51.32 | 13.41 | 34.66 | | |
| 8 | TE | ≥ 1,930 | ≥ 1,920 | 0.53 | 0.31 | 0.93 | 16.30 | 60.08 | 22.69 | | |
| 9 | ESR | ≥ 1,930 | ≥ 1,920 | 0.29 | 0.27 | 0.77 | 58.23 | 24.65 | 16.35 | | |
| 10 | ESR | ≥ 1,930 | ≥ 1,920 | 0.53 | 0.50 | 0.98 | 39.87 | 13.92 | 45.23 | | |
| 11 | ESR | ≥ 1,930 | ≥ 1,930 | 0.38 | 0.47 | 0.00 | 44.40 | 35.12 | 20.47 | | |
| 12 | MS | ≥ 1,930 | ≥ 1,930 | 0.49 | 0.53 | 0.15 | 42.55 | 16.81 | 40.48 | | |
| 13 | MS | ≥ 1,930 | ≥ 1,930 | 0.13 | 0.25 | 0.15 | 75.97 | 21.71 | 2.17 | | |
| 14 | ESR | ≥ 1,930 | ≥ 1,930 | 0.44 | 0.34 | 0.00 | 47.14 | 17.38 | 35.48 | | |
| 15 | ESR | ≥ 1,930 | ≥ 1,930 | 0.43 | 0.44 | 0.10 | 48.58 | 16.81 | 34.50 | | |
| 16 | ESR | ≥ 1,930 | ≥ 1,930 | 0.37 | 0.31 | 0.26 | 48.01 | 29.04 | 22.69 | | |
| 17 | ESR | ≥ 1,930 | ≥ 1,930 | 0.25 | 0.35 | 0.15 | 67.87 | 14.18 | 17.79 | | |
| 18 | CR | ≥ 1,930 | ≥ 1,930 | 0.27 | 0.63 | 0.26 | 26.61 | 45.59 | 21.25 | 4.59 | 1.50 |
| 18 | CR | ≥ 1,930 | ≥ 1,930 | 0.37 | 0.62 | 0.26 | 22.38 | 49.46 | 20.99 | 6.70 | |
| 19 | ESR | ≥ 1,930 | ≥ 1,930 | 0.68 | 0.27 | 0.00 | 17.17 | 28.67 | 54.15 | | |
| 20 | MS | ≥ 1,930 | ≥ 1,930 | 0.39 | 0.52 | 0.21 | 43.27 | 34.55 | 21.97 | | |
| 21 | ESR | ≥ 1,930 | ≥ 1,930 | 0.49 | 0.49 | 0.10 | 46.47 | 8.92 | 44.51 | | |
| 22 | MS/TE | ≥ 1,930 | ≥ 1,930 | 0.52 | 0.42 | 0.00 | 13.56 | 68.90 | 17.53 | | |
| 23 | ESR | ≥ 1,930 | ≥ 1,930 | 0.25 | 0.44 | 0.26 | 65.81 | 17.59 | 16.35 | | |
| 24 | ESR/TE | ≥ 1,930 | ≥ 1,930 | 0.36 | 0.48 | 0.21 | 44.77 | 38.73 | 16.30 | | |
| 25 | ESR | ≥ 1,930 | ≥ 1,930 | 0.36 | 0.41 | 0.21 | 48.79 | 29.60 | 21.40 | | |
| 26 | ESR | ≥ 1,930 | ≥ 1,920 | 0.36 | 0.47 | 0.57 | 53.95 | 19.29 | 26.20 | | |
| 27 | MS | ≥ 1,930 | ≥ 1,920 | 0.36 | 0.30 | 0.72 | 32.59 | 62.15 | 4.54 | | |
| 28 | ESR | ≥ 1,930 | ≥ 1,920 | 0.29 | 0.21 | 0.98 | 61.47 | 17.74 | 19.80 | | |

**Table 6.5 Item Statistics—English Language Arts Grade 4 Paper-Based Test Administration**

| ELA Grade 4 Paper-Based Test Administration | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | Item Type | Total N | Adj. N | *p*-Value | Pbis | Omit Rate | % at 0 | % at 1 | % at 2 | % at 3 | % at 4 |
| 1 | ESR | ≥ 54,300 | ≥ 54,120 | 0.42 | 0.43 | 0.33 | 54.76 | 5.46 | 39.45 | | |
| 2 | ESR | ≥ 54,300 | ≥ 54,050 | 0.61 | 0.49 | 0.45 | 25.29 | 26.77 | 47.49 | | |
| 3 | MS | ≥ 54,300 | ≥ 54,080 | 0.50 | 0.28 | 0.39 | 37.18 | 25.77 | 36.66 | | |
| 4 | ESR | ≥ 54,300 | ≥ 54,070 | 0.47 | 0.27 | 0.42 | 48.83 | 7.46 | 43.29 | | |
| 5 | ESR | ≥ 54,300 | ≥ 53,980 | 0.40 | 0.36 | 0.58 | 58.16 | 3.43 | 37.83 | | |
| 6 | CR | ≥ 54,300 | ≥ 53,840 | 0.29 | 0.56 | 0.78 | 19.20 | 51.22 | 24.31 | 4.19 | 0.23 |
| 6 | CR | ≥ 54,300 | ≥ 53,840 | 0.39 | 0.55 | 0.78 | 18.25 | 51.17 | 25.36 | 4.38 | |
| 7 | ESR | ≥ 54,300 | ≥ 52,010 | 0.40 | 0.33 | 4.21 | 51.31 | 11.43 | 33.06 | | |
| 8 | ESR | ≥ 54,300 | ≥ 51,660 | 0.63 | 0.53 | 4.85 | 32.30 | 5.79 | 57.06 | | |
| 9 | ESR | ≥ 54,300 | ≥ 51,410 | 0.33 | 0.28 | 5.31 | 51.04 | 25.73 | 17.92 | | |
| 10 | ESR | ≥ 54,300 | ≥ 51,140 | 0.63 | 0.53 | 5.82 | 29.92 | 10.68 | 53.58 | | |
| 11 | ESR | ≥ 54,300 | ≥ 54,150 | 0.39 | 0.44 | 0.28 | 41.49 | 38.38 | 19.85 | | |
| 12 | MS | ≥ 54,300 | ≥ 54,060 | 0.50 | 0.53 | 0.44 | 42.48 | 14.74 | 42.34 | | |
| 13 | MS | ≥ 54,300 | ≥ 54,040 | 0.17 | 0.21 | 0.48 | 69.67 | 25.86 | 3.99 | | |
| 14 | ESR | ≥ 54,300 | ≥ 53,970 | 0.44 | 0.37 | 0.60 | 47.86 | 15.76 | 35.78 | | |
| 15 | ESR | ≥ 54,300 | ≥ 53,970 | 0.44 | 0.41 | 0.59 | 46.64 | 17.90 | 34.87 | | |
| 16 | ESR | ≥ 54,300 | ≥ 54,000 | 0.37 | 0.30 | 0.55 | 47.77 | 29.80 | 21.88 | | |
| 17 | ESR | ≥ 54,300 | ≥ 53,810 | 0.23 | 0.32 | 0.89 | 69.23 | 15.00 | 14.88 | | |
| 18 | CR | ≥ 54,300 | ≥ 53,860 | 0.26 | 0.57 | 0.77 | 24.88 | 50.34 | 20.73 | 2.94 | 0.30 |
| 18 | CR | ≥ 54,300 | ≥ 53,860 | 0.35 | 0.56 | 0.77 | 22.66 | 52.10 | 21.30 | 3.12 | |
| 19 | ESR | ≥ 54,300 | ≥ 54,110 | 0.70 | 0.25 | 0.34 | 17.09 | 25.95 | 56.62 | | |
| 20 | MS | ≥ 54,300 | ≥ 54,060 | 0.42 | 0.49 | 0.43 | 40.11 | 34.89 | 24.57 | | |
| 21 | ESR | ≥ 54,300 | ≥ 53,900 | 0.43 | 0.48 | 0.73 | 52.37 | 8.81 | 38.08 | | |
| 22 | MS | ≥ 54,300 | ≥ 54,030 | 0.35 | 0.35 | 0.48 | 47.75 | 33.95 | 17.82 | | |
| 23 | ESR | ≥ 54,300 | ≥ 53,780 | 0.28 | 0.44 | 0.94 | 63.69 | 16.25 | 19.12 | | |
| 24 | ESR | ≥ 54,300 | ≥ 53,660 | 0.25 | 0.29 | 1.17 | 60.75 | 26.26 | 11.83 | | |
| 25 | ESR | ≥ 54,300 | ≥ 53,480 | 0.34 | 0.41 | 1.50 | 51.95 | 25.43 | 21.13 | | |
| 26 | ESR | ≥ 54,300 | ≥ 53,680 | 0.37 | 0.49 | 1.13 | 53.58 | 16.95 | 28.34 | | |
| 27 | MS | ≥ 54,300 | ≥ 53,410 | 0.38 | 0.29 | 1.64 | 29.87 | 62.73 | 5.76 | | |
| 28 | ESR | ≥ 54,300 | ≥ 52,920 | 0.29 | 0.14 | 2.53 | 61.07 | 16.02 | 20.38 | | |

**Table 6.6 Item Statistics—English Language Arts Grade 5 Computer-Based Test Administration**

| ELA Grade 5 Computer-Based Test Administration | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | Item Type | Total N | Adj. N | *p*-Value | Pbis | Omit Rate | % at 0 | % at 1 | % at 2 | % at 3 | % at 4 |
| 1 | ESR | ≥ 53,300 | ≥ 53,290 | 0.85 | 0.40 | 0.03 | 11.66 | 6.96 | 81.35 | | |
| 2 | ESR | ≥ 53,300 | ≥ 53,270 | 0.48 | 0.50 | 0.07 | 45.10 | 14.72 | 40.12 | | |
| 3 | MS | ≥ 53,300 | ≥ 53,260 | 0.28 | 0.29 | 0.08 | 52.30 | 38.90 | 8.72 | | |
| 4 | ESR | ≥ 53,300 | ≥ 53,260 | 0.62 | 0.39 | 0.08 | 35.86 | 3.31 | 60.75 | | |
| 5 | ESR/TE | ≥ 53,300 | ≥ 53,280 | 0.31 | 0.32 | 0.05 | 52.39 | 33.25 | 14.30 | | |
| 6 | CR | ≥ 53,300 | ≥ 53,010 | 0.21 | 0.54 | 0.31 | 34.52 | 48.47 | 14.27 | 1.97 | |
| 6 | CR | ≥ 53,300 | ≥ 53,010 | 0.37 | 0.56 | 0.31 | 24.51 | 45.22 | 24.89 | 4.84 | |
| 7 | ESR | ≥ 53,300 | ≥ 53,160 | 0.44 | 0.35 | 0.26 | 44.66 | 21.60 | 33.47 | | 0.65 |
| 8 | ESR | ≥ 53,300 | ≥ 53,140 | 0.37 | 0.28 | 0.31 | 44.04 | 37.15 | 18.51 | | |
| 9 | ESR | ≥ 53,300 | ≥ 53,120 | 0.29 | 0.33 | 0.34 | 61.98 | 18.30 | 19.38 | | |
| 10 | ESR/TE | ≥ 53,300 | ≥ 53,100 | 0.32 | 0.34 | 0.38 | 54.35 | 26.71 | 18.56 | | |
| 11 | ESR | ≥ 53,300 | ≥ 53,270 | 0.78 | 0.40 | 0.07 | 17.52 | 8.04 | 74.37 | | |
| 12 | MS | ≥ 53,300 | ≥ 53,240 | 0.46 | 0.31 | 0.11 | 13.45 | 81.96 | 4.48 | | |
| 13 | ESR | ≥ 53,300 | ≥ 53,200 | 0.40 | 0.33 | 0.19 | 48.95 | 22.74 | 28.12 | | |
| 14 | ESR | ≥ 53,300 | ≥ 53,190 | 0.44 | 0.43 | 0.22 | 39.53 | 32.39 | 27.86 | | |
| 15 | ESR | ≥ 53,300 | ≥ 53,240 | 0.46 | 0.52 | 0.11 | 40.93 | 25.16 | 33.79 | | |
| 16 | MS | ≥ 53,300 | ≥ 53,240 | 0.44 | 0.39 | 0.12 | 26.16 | 59.07 | 14.65 | | |
| 17 | MS | ≥ 53,300 | ≥ 53,230 | 0.26 | 0.32 | 0.15 | 63.48 | 19.90 | 16.47 | | |
| 18 | CR | ≥ 53,300 | ≥ 53,050 | 0.26 | 0.61 | 0.24 | 27.57 | 46.73 | 20.35 | 4.41 | |
| 18 | CR | ≥ 53,300 | ≥ 53,050 | 0.38 | 0.61 | 0.24 | 23.32 | 44.39 | 25.23 | 6.57 | |
| 19 | TE/ESR | ≥ 53,300 | ≥ 53,230 | 0.47 | 0.54 | 0.15 | 44.83 | 16.08 | 38.93 | | |
| 20 | ESR | ≥ 53,300 | ≥ 53,240 | 0.48 | 0.36 | 0.11 | 47.22 | 9.82 | 42.85 | | |
| 21 | ESR | ≥ 53,300 | ≥ 53,230 | 0.47 | 0.47 | 0.14 | 47.23 | 10.95 | 41.69 | | |
| 22 | ESR | ≥ 53,300 | ≥ 53,230 | 0.37 | 0.49 | 0.15 | 59.50 | 7.15 | 33.20 | | |
| 23 | ESR | ≥ 53,300 | ≥ 53,220 | 0.58 | 0.40 | 0.17 | 35.05 | 13.77 | 51.02 | | |
| 24 | ESR | ≥ 53,300 | ≥ 53,210 | 0.31 | 0.28 | 0.18 | 64.18 | 9.17 | 26.47 | | |
| 25 | ESR | ≥ 53,300 | ≥ 53,190 | 0.66 | 0.45 | 0.21 | 18.39 | 30.13 | 51.27 | | |
| 26 | MS | ≥ 53,300 | ≥ 53,130 | 0.22 | 0.30 | 0.33 | 60.99 | 32.73 | 5.95 | | |
| 27 | MS | ≥ 53,300 | ≥ 53,080 | 0.37 | 0.51 | 0.42 | 46.31 | 32.54 | 20.73 | | |
| 28 | ESR | ≥ 53,300 | ≥ 52,980 | 0.33 | 0.33 | 0.60 | 62.12 | 9.66 | 27.62 | | |

**Table 6.7 Item Statistics—English Language Arts Grade 6 Computer-Based Administration**

| ELA Grade 6 Computer-Based Test Administration | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | Item Type | Total N | Adj. N | *p*-Value | Pbis | Omit Rate | % at 0 | % at 1 | % at 2 | % at 3 | % at 4 |
| 1 | MS | ≥ 52,370 | ≥ 52,340 | 0.70 | 0.47 | 0.06 | 11.80 | 37.12 | 51.01 | | |
| 2 | ESR | ≥ 52,370 | ≥ 52,270 | 0.42 | 0.50 | 0.19 | 48.65 | 18.04 | 33.12 | | |
| 3 | ESR | ≥ 52,370 | ≥ 52,260 | 0.21 | 0.22 | 0.21 | 68.75 | 19.69 | 11.34 | | |
| 4 | ESR | ≥ 52,370 | ≥ 52,250 | 0.53 | 0.41 | 0.23 | 39.62 | 13.81 | 46.34 | | |
| 5 | ESR | ≥ 52,370 | ≥ 52,290 | 0.42 | 0.38 | 0.17 | 43.57 | 28.71 | 27.55 | | |
| 6 | ESR | ≥ 52,370 | ≥ 52,270 | 0.30 | 0.29 | 0.20 | 66.42 | 6.08 | 27.29 | | |
| 7 | ESR | ≥ 52,370 | ≥ 52,260 | 0.48 | 0.52 | 0.23 | 46.37 | 10.16 | 43.24 | | |
| 8 | CR | ≥ 52,370 | ≥ 52,040 | 0.30 | 0.63 | 0.39 | 24.30 | 39.69 | 28.70 | 6.00 | 0.66 |
| 8 | CR | ≥ 52,370 | ≥ 52,040 | 0.41 | 0.65 | 0.39 | 22.36 | 37.88 | 33.43 | 5.69 | |
| 9 | ESR | ≥ 52,370 | ≥ 52,320 | 0.43 | 0.36 | 0.11 | 55.00 | 4.67 | 40.23 | | |
| 10 | ESR | ≥ 52,370 | ≥ 52,300 | 0.62 | 0.56 | 0.15 | 22.13 | 32.31 | 45.42 | | |
| 11 | TE/TE | ≥ 52,370 | ≥ 52,290 | 0.48 | 0.42 | 0.16 | 17.34 | 69.82 | 12.68 | | |
| 12 | ESR | ≥ 52,370 | ≥ 52,280 | 0.27 | 0.32 | 0.19 | 55.86 | 33.01 | 10.94 | | |
| 13 | CR | ≥ 52,370 | ≥ 51,620 | 0.26 | 0.63 | 0.71 | 45.21 | 19.67 | 21.60 | 9.12 | 2.97 |
| 13 | CR | ≥ 52,370 | ≥ 51,620 | 0.37 | 0.66 | 0.71 | 29.90 | 34.63 | 26.30 | 7.73 | |
| 14 | ESR | ≥ 52,370 | ≥ 52,270 | 0.38 | 0.49 | 0.21 | 44.47 | 34.47 | 20.86 | | |
| 15 | TE | ≥ 52,370 | ≥ 52,210 | 0.34 | 0.59 | 0.32 | 48.35 | 34.77 | 16.56 | | |
| 16 | ESR | ≥ 52,370 | ≥ 52,240 | 0.37 | 0.44 | 0.26 | 45.92 | 34.19 | 19.64 | | |
| 17 | ESR | ≥ 52,370 | ≥ 52,190 | 0.34 | 0.31 | 0.34 | 55.93 | 19.23 | 24.50 | | |
| 18 | MS | ≥ 52,370 | ≥ 52,150 | 0.32 | 0.45 | 0.42 | 55.27 | 25.29 | 19.02 | | |
| 19 | ESR | ≥ 52,370 | ≥ 52,110 | 0.35 | 0.31 | 0.50 | 53.15 | 22.35 | 24.00 | | |
| 20 | ESR | ≥ 52,370 | ≥ 52,310 | 0.75 | 0.46 | 0.13 | 18.41 | 12.97 | 68.50 | | |
| 21 | ESR | ≥ 52,370 | ≥ 52,260 | 0.61 | 0.57 | 0.21 | 33.52 | 10.03 | 56.25 | | |
| 22 | ESR | ≥ 52,370 | ≥ 52,280 | 0.51 | 0.46 | 0.19 | 45.52 | 6.57 | 47.72 | | |
| 23 | TE | ≥ 52,370 | ≥ 52,260 | 0.39 | 0.45 | 0.21 | 32.25 | 57.57 | 9.97 | | |
| 24 | ESR | ≥ 52,370 | ≥ 52,250 | 0.67 | 0.60 | 0.23 | 29.51 | 6.74 | 63.52 | | |
| 25 | ESR | ≥ 52,370 | ≥ 52,270 | 0.69 | 0.57 | 0.19 | 25.12 | 10.93 | 63.75 | | |
| 26 | ESR | ≥ 52,370 | ≥ 52,230 | 0.44 | 0.49 | 0.28 | 42.07 | 28.11 | 29.53 | | |
| 27 | ESR | ≥ 52,370 | ≥ 52,220 | 0.31 | 0.44 | 0.29 | 66.33 | 5.32 | 28.06 | | |
| 28 | ESR | ≥ 52,370 | ≥ 52,230 | 0.31 | 0.33 | 0.27 | 60.09 | 18.38 | 21.26 | | |
| 29 | MS | ≥ 52,370 | ≥ 52,260 | 0.27 | 0.39 | 0.21 | 50.83 | 44.11 | 4.84 | | |
| 30 | TE | ≥ 52,370 | ≥ 52,190 | 0.42 | 0.43 | 0.36 | 39.82 | 36.32 | 23.51 | | |
| 31 | ESR | ≥ 52,370 | ≥ 52,200 | 0.35 | 0.32 | 0.34 | 53.27 | 23.45 | 22.94 | | |
| 32 | ESR | ≥ 52,370 | ≥ 52,180 | 0.34 | 0.40 | 0.38 | 51.83 | 28.20 | 19.59 | | |
| 33 | ESR | ≥ 52,370 | ≥ 52,170 | 0.26 | 0.38 | 0.39 | 63.55 | 19.65 | 16.41 | | |

**Table 6.8 Item Statistics—English Language Arts Grade 7 Computer-Based Test Administration**

| ELA Grade 7 Computer-Based Test Administration | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | Item Type | Total N | Adj. N | *p*-Value | Pbis | Omit Rate | % at 0 | % at 1 | % at 2 | % at 3 | % at 4 |
| 1 | ESR | ≥ 51,930 | ≥ 51,880 | 0.30 | 0.32 | 0.10 | 62.77 | 14.74 | 22.39 | | |
| 2 | MS | ≥ 51,930 | ≥ 51,850 | 0.37 | 0.35 | 0.15 | 48.47 | 29.60 | 21.78 | | |
| 3 | ESR | ≥ 51,930 | ≥ 51,850 | 0.47 | 0.48 | 0.16 | 42.75 | 19.35 | 37.74 | | |
| 4 | ESR | ≥ 51,930 | ≥ 51,840 | 0.52 | 0.36 | 0.17 | 36.56 | 22.31 | 40.96 | | |
| 5 | ESR/TE | ≥ 51,930 | ≥ 51,860 | 0.39 | 0.28 | 0.13 | 49.54 | 22.40 | 27.93 | | |
| 6 | CR | ≥ 51,930 | ≥ 51,260 | 0.38 | 0.68 | 0.76 | 18.73 | 31.74 | 31.75 | 12.28 | 4.21 |
| 6 | CR | ≥ 51,930 | ≥ 51,260 | 0.49 | 0.69 | 0.76 | 18.96 | 30.16 | 32.43 | 17.16 | |
| 7 | ESR | ≥ 51,930 | ≥ 51,680 | 0.59 | 0.53 | 0.49 | 35.02 | 11.95 | 52.54 | | |
| 8 | ESR | ≥ 51,930 | ≥ 51,640 | 0.43 | 0.43 | 0.56 | 51.84 | 9.66 | 37.94 | | |
| 9 | ESR | ≥ 51,930 | ≥ 51,590 | 0.57 | 0.43 | 0.65 | 36.20 | 13.22 | 49.93 | | |
| 10 | ESR | ≥ 51,930 | ≥ 51,510 | 0.30 | 0.14 | 0.80 | 62.39 | 14.36 | 22.45 | | |
| 11 | ESR | ≥ 51,930 | ≥ 51,870 | 0.75 | 0.50 | 0.13 | 20.73 | 8.14 | 71.00 | | |
| 12 | ESR | ≥ 51,930 | ≥ 51,830 | 0.55 | 0.43 | 0.20 | 34.57 | 20.91 | 44.31 | | |
| 13 | ESR | ≥ 51,930 | ≥ 51,850 | 0.63 | 0.51 | 0.16 | 30.38 | 12.29 | 57.17 | | |
| 14 | ESR | ≥ 51,930 | ≥ 51,830 | 0.53 | 0.47 | 0.20 | 27.11 | 39.52 | 33.17 | | |
| 15 | ESR | ≥ 51,930 | ≥ 51,860 | 0.54 | 0.39 | 0.14 | 41.82 | 8.36 | 49.68 | | |
| 16 | ESR/TE | ≥ 51,930 | ≥ 51,830 | 0.38 | 0.38 | 0.20 | 48.47 | 26.18 | 25.14 | | |
| 17 | ESR | ≥ 51,930 | ≥ 51,830 | 0.23 | 0.14 | 0.20 | 73.70 | 6.21 | 19.89 | | |
| 18 | CR | ≥ 51,930 | ≥ 51,430 | 0.39 | 0.68 | 0.55 | 15.94 | 32.89 | 32.19 | 14.90 | 3.10 |
| 18 | CR | ≥ 51,930 | ≥ 51,430 | 0.51 | 0.69 | 0.55 | 16.70 | 31.41 | 32.92 | 17.99 | |
| 19 | ESR | ≥ 51,930 | ≥ 51,840 | 0.41 | 0.33 | 0.18 | 36.01 | 46.26 | 17.55 | | |
| 20 | ESR | ≥ 51,930 | ≥ 51,810 | 0.41 | 0.39 | 0.23 | 56.57 | 3.80 | 39.39 | | |
| 21 | MS | ≥ 51,930 | ≥ 51,780 | 0.53 | 0.53 | 0.29 | 32.41 | 29.85 | 37.45 | | |
| 22 | MS | ≥ 51,930 | ≥ 51,780 | 0.43 | 0.47 | 0.29 | 39.08 | 35.43 | 25.19 | | |
| 23 | ESR | ≥ 51,930 | ≥ 51,770 | 0.54 | 0.31 | 0.31 | 41.67 | 9.12 | 48.90 | | |
| 24 | MS | ≥ 51,930 | ≥ 51,790 | 0.33 | 0.29 | 0.28 | 55.20 | 23.39 | 21.13 | | |
| 25 | ESR | ≥ 51,930 | ≥ 51,770 | 0.47 | 0.45 | 0.30 | 37.94 | 29.26 | 32.49 | | |
| 26 | ESR | ≥ 51,930 | ≥ 51,790 | 0.20 | 0.23 | 0.28 | 66.63 | 25.47 | 7.62 | | |
| 27 | ESR | ≥ 51,930 | ≥ 51,780 | 0.48 | 0.42 | 0.29 | 43.40 | 16.97 | 39.33 | | |
| 28 | ESR | ≥ 51,930 | ≥ 51,780 | 0.46 | 0.30 | 0.29 | 47.86 | 12.75 | 39.10 | | |
| 29 | ESR | ≥ 51,930 | ≥ 51,760 | 0.41 | 0.40 | 0.32 | 46.79 | 23.17 | 29.72 | | |
| 30 | TE | ≥ 51,930 | ≥ 51,680 | 0.19 | 0.24 | 0.48 | 69.97 | 21.27 | 8.28 | | |

**Table 6.9 Item Statistics—English Language Arts Grade 8 Computer-Based Test Administration**

| ELA Grade 8 Computer-Based Test Administration | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | Item Type | Total N | Adj. N | *p*-Value | Pbis | Omit Rate | % at 0 | % at 1 | % at 2 | % at 3 | % at 4 |
| 1 | ESR | ≥ 50,455 | ≥ 50,420 | 0.44 | 0.53 | 0.06 | 50.79 | 10.95 | 38.20 | | |
| 2 | ESR/TE | ≥ 50,455 | ≥ 50,410 | 0.64 | 0.39 | 0.09 | 28.61 | 15.23 | 56.07 | | |
| 3 | ESR | ≥ 50,455 | ≥ 50,370 | 0.43 | 0.15 | 0.16 | 51.14 | 11.09 | 37.62 | | |
| 4 | ESR | ≥ 50,455 | ≥ 50,390 | 0.45 | 0.13 | 0.11 | 43.35 | 23.92 | 32.63 | | |
| 5 | ESR | ≥ 50,455 | ≥ 50,370 | 0.33 | 0.34 | 0.16 | 49.33 | 34.59 | 15.93 | | |
| 6 | CR | ≥ 50,455 | ≥ 49,700 | 0.37 | 0.65 | 0.92 | 20.87 | 33.44 | 26.60 | 12.42 | 5.18 |
| 6 | CR | ≥ 50,455 | ≥ 49,700 | 0.48 | 0.66 | 0.92 | 20.83 | 32.04 | 27.43 | 18.22 | |
| 7 | MS | ≥ 50,455 | ≥ 50,220 | 0.31 | 0.32 | 0.46 | 51.11 | 36.09 | 12.35 | | |
| 8 | MS | ≥ 50,455 | ≥ 50,170 | 0.38 | 0.41 | 0.56 | 37.80 | 46.78 | 14.87 | | |
| 9 | ESR | ≥ 50,455 | ≥ 50,120 | 0.70 | 0.43 | 0.66 | 21.09 | 16.50 | 61.74 | | |
| 10 | TE | ≥ 50,455 | ≥ 49,940 | 0.38 | 0.38 | 1.02 | 48.14 | 26.36 | 24.49 | | |
| 11 | MS | ≥ 50,455 | ≥ 50,370 | 0.62 | 0.37 | 0.16 | 21.71 | 33.43 | 44.69 | | |
| 12 | MS | ≥ 50,455 | ≥ 50,340 | 0.53 | 0.38 | 0.22 | 16.27 | 61.42 | 22.09 | | |
| 13 | ESR | ≥ 50,455 | ≥ 50,310 | 0.29 | 0.21 | 0.28 | 58.88 | 23.86 | 16.98 | | |
| 14 | ESR | ≥ 50,455 | ≥ 50,320 | 0.31 | 0.37 | 0.25 | 60.04 | 16.84 | 22.87 | | |
| 15 | ESR | ≥ 50,455 | ≥ 50,320 | 0.33 | 0.42 | 0.26 | 57.54 | 18.61 | 23.59 | | |
| 16 | ESR | ≥ 50,455 | ≥ 50,330 | 0.52 | 0.19 | 0.23 | 43.41 | 8.74 | 47.61 | | |
| 17 | ESR/TE | ≥ 50,455 | ≥ 50,330 | 0.27 | 0.45 | 0.25 | 46.85 | 51.63 | 1.27 | | |
| 18 | CR | ≥ 50,455 | ≥ 49,680 | 0.52 | 0.70 | 0.99 | 11.51 | 20.77 | 28.61 | 25.08 | 12.51 |
| 18 | CR | ≥ 50,455 | ≥ 49,680 | 0.64 | 0.69 | 0.99 | 12.09 | 20.48 | 28.41 | 37.51 | |
| 19 | ESR | ≥ 50,455 | ≥ 50,340 | 0.46 | 0.38 | 0.22 | 44.24 | 18.79 | 36.75 | | |
| 20 | ESR | ≥ 50,455 | ≥ 50,300 | 0.30 | 0.31 | 0.29 | 63.34 | 12.77 | 23.60 | | |
| 21 | ESR | ≥ 50,455 | ≥ 50,290 | 0.35 | 0.33 | 0.31 | 52.41 | 25.15 | 22.13 | | |
| 22 | ESR | ≥ 50,455 | ≥ 50,300 | 0.38 | 0.25 | 0.30 | 57.26 | 8.63 | 33.82 | | |
| 23 | ESR | ≥ 50,455 | ≥ 50,290 | 0.45 | 0.41 | 0.32 | 51.64 | 7.09 | 40.95 | | |
| 24 | ESR | ≥ 50,455 | ≥ 50,290 | 0.36 | 0.31 | 0.32 | 51.74 | 24.36 | 23.59 | | |
| 25 | ESR | ≥ 50,455 | ≥ 50,320 | 0.82 | 0.49 | 0.27 | 12.41 | 11.15 | 76.17 | | |
| 26 | ESR | ≥ 50,455 | ≥ 50,320 | 0.35 | 0.31 | 0.25 | 57.86 | 14.56 | 27.32 | | |
| 27 | ESR | ≥ 50,455 | ≥ 50,300 | 0.73 | 0.54 | 0.30 | 18.75 | 15.70 | 65.25 | | |
| 28 | ESR | ≥ 50,455 | ≥ 50,290 | 0.81 | 0.48 | 0.32 | 16.12 | 4.92 | 78.64 | | |
| 29 | ESR | ≥ 50,455 | ≥ 50,290 | 0.43 | 0.12 | 0.33 | 45.79 | 21.98 | 31.91 | | |
| 30 | MS | ≥ 50,455 | ≥ 50,260 | 0.31 | 0.32 | 0.37 | 51.07 | 35.87 | 12.69 | | |

**Table 6.10 Item Statistics—Mathematics Grade 3 Paper-Based Test Administration**

| Mathematics Grade 3 Paper-Based Test Administration | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | 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 |
| 1 | MC | ≥ 56,800 | ≥ 56,520 | 0.78 | 0.32 | 0.38 | | | | | | | |
| 2 | MC | ≥ 56,800 | ≥ 56,270 | 0.86 | 0.35 | 0.42 | | | | | | | |
| 3 | SA | ≥ 56,800 | ≥ 55,880 | 0.73 | 0.48 | 1.62 | 26.65 | 71.74 | | | | | |
| 4 | MS | ≥ 56,800 | ≥ 56,090 | 0.41 | 0.43 | 1.24 | 57.80 | 40.95 | | | | | |
| 5 | SA | ≥ 56,800 | ≥ 55,870 | 0.38 | 0.31 | 1.63 | 60.88 | 37.49 | | | | | |
| 6 | MC | ≥ 56,800 | ≥ 56,290 | 0.90 | 0.34 | 0.55 | | | | | | | |
| 7 | SA | ≥ 56,800 | ≥ 56,150 | 0.35 | 0.69 | 1.14 | 52.21 | 24.13 | 22.52 | | | | |
| 8 | MC | ≥ 56,800 | ≥ 56,050 | 0.92 | 0.28 | 0.65 | | | | | | | |
| 9 | SA | ≥ 56,800 | ≥ 55,430 | 0.75 | 0.51 | 2.42 | 24.84 | 72.74 | | | | | |
| 10 | MC | ≥ 56,800 | ≥ 56,380 | 0.84 | 0.38 | 0.67 | | | | | | | |
| 11 | MC | ≥ 56,800 | ≥ 56,470 | 0.98 | 0.17 | 0.52 | | | | | | | |
| 12 | MS | ≥ 56,800 | ≥ 55,750 | 0.53 | 0.56 | 1.85 | 45.79 | 52.36 | | | | | |
| 13 | SA | ≥ 56,800 | ≥ 56,290 | 0.77 | 0.27 | 0.90 | 23.22 | 75.88 | | | | | |
| 14 | CR | ≥ 56,800 | ≥ 55,770 | 0.19 | 0.58 | 1.49 | 63.31 | 9.84 | 14.55 | 6.33 | 4.17 | | |
| 15 | CR | ≥ 56,800 | ≥ 54,550 | 0.29 | 0.61 | 3.87 | 50.33 | 18.73 | 16.95 | 10.02 | | | |
| 16 | SA | ≥ 56,800 | ≥ 56,260 | 0.79 | 0.32 | 0.95 | 20.96 | 78.09 | | | | | |
| 17 | MC | ≥ 56,800 | ≥ 56,510 | 0.85 | 0.42 | 0.42 | | | | | | | |
| 18 | SA | ≥ 56,800 | ≥ 55,930 | 0.70 | 0.50 | 1.52 | 29.33 | 69.15 | | | | | |
| 19 | MC | ≥ 56,800 | ≥ 56,340 | 0.75 | 0.47 | 0.69 | | | | | | | |
| 20 | ESR | ≥ 56,800 | ≥ 56,190 | 0.53 | 0.52 | 1.08 | 28.02 | 37.54 | 33.36 | | | | |
| 21 | SA | ≥ 56,800 | ≥ 56,380 | 0.29 | 0.61 | 0.74 | 54.04 | 32.67 | 12.55 | | | | |
| 22 | MS | ≥ 56,800 | ≥ 55,810 | 0.26 | 0.36 | 1.73 | 72.75 | 25.51 | | | | | |
| 23 | MC | ≥ 56,800 | ≥ 54,040 | 0.55 | 0.46 | 2.61 | | | | | | | |
| 24 | SA | ≥ 56,800 | ≥ 55,710 | 0.77 | 0.40 | 1.92 | 22.65 | 75.43 | | | | | |
| 25 | MC | ≥ 56,800 | ≥ 56,250 | 0.40 | 0.39 | 0.79 | | | | | | | |
| 26 | SA | ≥ 56,800 | ≥ 55,780 | 0.78 | 0.35 | 1.78 | 21.53 | 76.69 | | | | | |
| 27 | MC | ≥ 56,800 | ≥ 56,440 | 0.94 | 0.28 | 0.55 | | | | | | | |
| 28 | CR | ≥ 56,800 | ≥ 56,550 | 0.47 | 0.67 | 0.44 | 24.95 | 23.90 | 36.30 | 14.41 | | | |
| 29 | CR | ≥ 56,800 | ≥ 52,400 | 0.22 | 0.57 | 7.52 | 64.77 | 8.80 | 5.07 | 13.61 | | | |
| 30 | MC | ≥ 56,800 | ≥ 56,440 | 0.55 | 0.41 | 0.43 | | | | | | | |
| 31 | MC | ≥ 56,800 | ≥ 55,100 | 0.70 | 0.47 | 1.28 | | | | | | | |
| 32 | SA | ≥ 56,800 | ≥ 55,800 | 0.67 | 0.46 | 1.75 | 32.17 | 66.08 | | | | | |
| 33 | MC | ≥ 56,800 | ≥ 56,390 | 0.79 | 0.36 | 0.66 | | | | | | | |
| 34 | SA | ≥ 56,800 | ≥ 55,920 | 0.75 | 0.43 | 1.55 | 24.51 | 73.94 | | | | | |
| 35 | MS | ≥ 56,800 | ≥ 55,710 | 0.30 | 0.49 | 1.92 | 68.30 | 29.78 | | | | | |
| 36 | SA | ≥ 56,800 | ≥ 55,500 | 0.22 | 0.50 | 2.29 | 76.39 | 21.32 | | | | | |
| 37 | MC | ≥ 56,800 | ≥ 56,180 | 0.62 | 0.36 | 0.90 | | | | | | | |
| 38 | MC | ≥ 56,800 | ≥ 55,980 | 0.75 | 0.41 | 1.36 | | | | | | | |
| 39 | SA | ≥ 56,800 | ≥ 55,410 | 0.25 | 0.58 | 2.45 | 72.83 | 24.72 | | | | | |

| Mathematics Grade 3 Paper-Based Test Administration (continued) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | 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 |
| 40 | MC | ≥ 56,800 | ≥ 56,290 | 0.92 | 0.35 | 0.71 | | | | | | | |
| 41 | MS | ≥ 56,800 | ≥ 56,210 | 0.72 | 0.43 | 1.04 | 27.25 | 71.71 | | | | | |
| 42 | CR | ≥ 56,800 | ≥ 55,560 | 0.26 | 0.54 | 2.03 | 46.34 | 30.30 | 16.46 | 4.71 | | | |
| 43 | CR | ≥ 56,800 | ≥ 56,090 | 0.40 | 0.81 | 1.24 | 25.22 | 17.00 | 13.06 | 13.78 | 8.70 | 9.67 | 11.33 |

**Table 6.11 Item Statistics—Mathematics Grade 4 Computer-Based Test Administration**

| Mathematics Grade 4 Computer-Based Test Administration | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item* | 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 |
| 1 | MC | ≥ 1,930 | ≥ 1,930 | 0.91 | 0.29 | 0.10 | | | | | | | |
| 2 | MC | ≥ 1,930 | ≥ 1,930 | 0.76 | 0.35 | 0.10 | | | | | | | |
| 3 | MS | ≥ 1,930 | ≥ 1,930 | 0.88 | 0.32 | 0.00 | 12.07 | 87.93 | | | | | |
| 4 | MC | ≥ 1,930 | ≥ 1,930 | 0.25 | 0.34 | 0.15 | | | | | | | |
| 5 | MS | ≥ 1,930 | ≥ 1,930 | 0.48 | 0.61 | 0.15 | 52.06 | 47.78 | | | | | |
| 6 | SA | ≥ 1,930 | ≥ 1,930 | 0.59 | 0.40 | 0.31 | 40.76 | 58.93 | | | | | |
| 7 | SA | ≥ 1,930 | ≥ 1,930 | 0.47 | 0.55 | 0.26 | 52.53 | 47.21 | | | | | |
| 8 | MC | ≥ 1,930 | ≥ 1,930 | 0.50 | 0.62 | 0.05 | | | | | | | |
| 9 | ESR | ≥ 1,930 | ≥ 1,930 | 0.72 | 0.43 | 0.05 | 12.54 | 29.98 | 57.43 | | | | |
| 10 | MC | ≥ 1,930 | ≥ 1,930 | 0.72 | 0.47 | 0.36 | | | | | | | |
| 11 | SA | ≥ 1,930 | ≥ 1,930 | 0.63 | 0.35 | 0.15 | 36.74 | 63.11 | | | | | |
| 12 | MC | ≥ 1,930 | ≥ 1,930 | 0.57 | 0.44 | 0.10 | | | | | | | |
| 13 | MC | ≥ 1,930 | ≥ 1,930 | 0.33 | 0.41 | 0.36 | | | | | | | |
| 14 | CR | ≥ 1,930 | ≥ 1,910 | 0.30 | 0.72 | 0.62 | 33.85 | 27.92 | 25.54 | 7.28 | 4.44 | | |
| 16 | MC | ≥ 1,930 | ≥ 1,930 | 0.63 | 0.36 | 0.10 | | | | | | | |
| 17 | MC | ≥ 1,930 | ≥ 1,930 | 0.88 | 0.31 | 0.36 | | | | | | | |
| 18 | MC | ≥ 1,930 | ≥ 1,930 | 0.61 | 0.42 | 0.00 | | | | | | | |
| 19 | MC | ≥ 1,930 | ≥ 1,930 | 0.50 | 0.44 | 0.10 | | | | | | | |
| 20 | SA | ≥ 1,930 | ≥ 1,920 | 0.26 | 0.57 | 0.46 | 73.63 | 25.90 | | | | | |
| 21 | MC | ≥ 1,930 | ≥ 1,930 | 0.29 | 0.21 | 0.21 | | | | | | | |
| 22 | ESR | ≥ 1,930 | ≥ 1,930 | 0.64 | 0.49 | 0.00 | 13.00 | 45.77 | 41.23 | | | | |
| 23 | MC | ≥ 1,930 | ≥ 1,930 | 0.33 | 0.31 | 0.26 | | | | | | | |
| 24 | MS | ≥ 1,930 | ≥ 1,930 | 0.72 | 0.45 | 0.05 | 28.17 | 71.78 | | | | | |
| 25 | SA | ≥ 1,930 | ≥ 1,920 | 0.27 | 0.56 | 0.46 | 72.76 | 26.78 | | | | | |
| 26 | SA | ≥ 1,930 | ≥ 1,920 | 0.29 | 0.57 | 0.62 | 70.64 | 28.74 | | | | | |
| 27 | ESR | ≥ 1,930 | ≥ 1,930 | 0.61 | 0.57 | 0.10 | 15.53 | 46.23 | 38.13 | | | | |
| 28 | CR | ≥ 1,930 | ≥ 1,900 | 0.24 | 0.66 | 0.98 | 64.65 | 10.68 | 8.57 | 14.40 | | | |
| 29 | CR | ≥ 1,930 | ≥ 1,890 | 0.28 | 0.64 | 1.60 | 49.79 | 20.69 | 20.49 | 6.81 | | | |
| 30 | MC | ≥ 1,930 | ≥ 1,930 | 0.64 | 0.39 | 0.15 | | | | | | | |
| 31 | MC | ≥ 1,930 | ≥ 1,930 | 0.63 | 0.53 | 0.05 | | | | | | | |
| 32 | MC | ≥ 1,930 | ≥ 1,930 | 0.70 | 0.47 | 0.26 | | | | | | | |
| 33 | MS | ≥ 1,930 | ≥ 1,930 | 0.18 | 0.32 | 0.15 | 82.09 | 17.75 | | | | | |
| 34 | MC | ≥ 1,930 | ≥ 1,930 | 0.79 | 0.37 | 0.15 | | | | | | | |
| 35 | SA | ≥ 1,930 | ≥ 1,930 | 0.37 | 0.35 | 0.21 | 63.11 | 36.69 | | | | | |
| 36 | MC | ≥ 1,930 | ≥ 1,930 | 0.27 | 0.27 | 0.05 | | | | | | | |
| 37 | SA | ≥ 1,930 | ≥ 1,920 | 0.16 | 0.46 | 0.52 | 83.23 | 16.25 | | | | | |
| 38 | MS | ≥ 1,930 | ≥ 1,930 | 0.41 | 0.44 | 0.15 | 59.13 | 40.71 | | | | | |
| 39 | MC | ≥ 1,930 | ≥ 1,930 | 0.52 | 0.43 | 0.21 | | | | | | | |

| Mathematics Grade 4 Computer-Based Test Administration (continued) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Item** | **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** |
| 40 | SA | ≥ 1,930 | ≥ 1,920 | 0.55 | 0.49 | 0.46 | 44.79 | 54.75 | | | | | |
| 41 | MC | ≥ 1,930 | ≥ 1,930 | 0.64 | 0.39 | 0.10 | | | | | | | |
| 42 | CR | ≥ 1,930 | ≥ 1,910 | 0.19 | 0.60 | 0.98 | 57.53 | 30.39 | 7.64 | 3.46 | | | |
| 43 | CR | ≥ 1,930 | ≥ 1,930 | 0.37 | 0.74 | 0.36 | 15.22 | 29.46 | 15.69 | 11.09 | 17.91 | 3.87 | 6.40 |

***Item 15 was removed from test scoring.*

**Table 6.12 Item Statistics—Mathematics Grade 4 Paper-Based Test Administration**

| Mathematics Grade 4 Paper-Based Test Administration | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item* | 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 |
| 1 | MC | ≥ 54,300 | ≥ 54,080 | 0.91 | 0.28 | 0.34 | | | | | | | |
| 2 | MC | ≥ 54,300 | ≥ 53,800 | 0.75 | 0.33 | 0.44 | | | | | | | |
| 3 | MS | ≥ 54,300 | ≥ 54,030 | 0.91 | 0.28 | 0.48 | 8.67 | 90.84 | | | | | |
| 4 | MC | ≥ 54,300 | ≥ 53,940 | 0.27 | 0.36 | 0.54 | | | | | | | |
| 5 | MS | ≥ 54,300 | ≥ 53,910 | 0.48 | 0.59 | 0.72 | 51.35 | 47.93 | | | | | |
| 6 | SA | ≥ 54,300 | ≥ 53,500 | 0.61 | 0.38 | 1.47 | 38.02 | 60.52 | | | | | |
| 7 | SA | ≥ 54,300 | ≥ 52,220 | 0.45 | 0.52 | 3.82 | 53.36 | 42.83 | | | | | |
| 8 | MC | ≥ 54,300 | ≥ 53,480 | 0.51 | 0.62 | 1.39 | | | | | | | |
| 9 | ESR | ≥ 54,300 | ≥ 53,930 | 0.75 | 0.48 | 0.68 | 11.07 | 26.78 | 61.48 | | | | |
| 10 | MC | ≥ 54,300 | ≥ 53,750 | 0.75 | 0.48 | 0.98 | | | | | | | |
| 11 | SA | ≥ 54,300 | ≥ 53,410 | 0.64 | 0.37 | 1.63 | 35.11 | 63.26 | | | | | |
| 12 | MC | ≥ 54,300 | ≥ 53,930 | 0.55 | 0.45 | 0.64 | | | | | | | |
| 13 | MC | ≥ 54,300 | ≥ 53,810 | 0.33 | 0.37 | 0.84 | | | | | | | |
| 14 | CR | ≥ 54,300 | ≥ 53,940 | 0.43 | 0.72 | 0.58 | 26.26 | 20.48 | 24.45 | 10.29 | 17.86 | | |
| 16 | MC | ≥ 54,300 | ≥ 49,360 | 0.65 | 0.32 | 0.91 | | | | | | | |
| 17 | MC | ≥ 54,300 | ≥ 54,030 | 0.87 | 0.31 | 0.44 | | | | | | | |
| 18 | MC | ≥ 54,300 | ≥ 53,710 | 0.64 | 0.41 | 1.03 | | | | | | | |
| 19 | MC | ≥ 54,300 | ≥ 53,820 | 0.47 | 0.40 | 0.83 | | | | | | | |
| 20 | SA | ≥ 54,300 | ≥ 52,920 | 0.23 | 0.52 | 2.53 | 74.73 | 22.74 | | | | | |
| 21 | MC | ≥ 54,300 | ≥ 53,660 | 0.27 | 0.21 | 0.86 | | | | | | | |
| 22 | ESR | ≥ 54,300 | ≥ 53,840 | 0.64 | 0.51 | 0.84 | 13.42 | 43.95 | 41.78 | | | | |
| 23 | MC | ≥ 54,300 | ≥ 53,970 | 0.37 | 0.34 | 0.53 | | | | | | | |
| 24 | MS | ≥ 54,300 | ≥ 53,810 | 0.74 | 0.40 | 0.90 | 25.72 | 73.38 | | | | | |
| 25 | SA | ≥ 54,300 | ≥ 52,970 | 0.32 | 0.58 | 2.45 | 66.30 | 31.25 | | | | | |
| 26 | SA | ≥ 54,300 | ≥ 52,800 | 0.34 | 0.56 | 2.76 | 64.57 | 32.66 | | | | | |
| 27 | ESR | ≥ 54,300 | ≥ 54,000 | 0.65 | 0.58 | 0.55 | 14.76 | 40.14 | 44.55 | | | | |
| 28 | CR | ≥ 54,300 | ≥ 53,630 | 0.29 | 0.67 | 1.12 | 58.83 | 12.59 | 8.75 | 18.60 | | | |
| 29 | CR | ≥ 54,300 | ≥ 51,970 | 0.35 | 0.64 | 4.21 | 43.14 | 19.91 | 18.36 | 14.31 | | | |
| 30 | MC | ≥ 54,300 | ≥ 54,040 | 0.69 | 0.37 | 0.42 | | | | | | | |
| 31 | MC | ≥ 54,300 | ≥ 53,860 | 0.58 | 0.53 | 0.57 | | | | | | | |
| 32 | MC | ≥ 54,300 | ≥ 53,650 | 0.73 | 0.47 | 1.16 | | | | | | | |
| 33 | MS | ≥ 54,300 | ≥ 53,680 | 0.18 | 0.34 | 1.13 | 81.08 | 17.79 | | | | | |
| 34 | MC | ≥ 54,300 | ≥ 53,830 | 0.78 | 0.36 | 0.61 | | | | | | | |
| 35 | SA | ≥ 54,300 | ≥ 53,300 | 0.42 | 0.35 | 1.84 | 57.23 | 40.93 | | | | | |
| 36 | MC | ≥ 54,300 | ≥ 53,680 | 0.29 | 0.29 | 1.09 | | | | | | | |
| 37 | SA | ≥ 54,300 | ≥ 52,970 | 0.17 | 0.48 | 2.44 | 81.11 | 16.45 | | | | | |
| 38 | MS | ≥ 54,300 | ≥ 53,790 | 0.41 | 0.41 | 0.93 | 58.82 | 40.26 | | | | | |

| Mathematics Grade 4 Paper-Based Test Administration (continued) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | 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 |
| 39 | MC | ≥ 54,300 | ≥ 53,470 | 0.50 | 0.40 | 1.02 | | | | | | | |
| 40 | SA | ≥ 54,300 | ≥ 52,760 | 0.58 | 0.50 | 2.82 | 40.65 | 56.52 | | | | | |
| 41 | MC | ≥ 54,300 | ≥ 53,640 | 0.67 | 0.39 | 1.10 | | | | | | | |
| 42 | CR | ≥ 54,300 | ≥ 53,610 | 0.17 | 0.56 | 1.20 | 59.39 | 30.62 | 6.75 | 1.96 | | | |
| 43 | CR | ≥ 54,300 | ≥ 53,840 | 0.46 | 0.75 | 0.70 | 13.18 | 20.62 | 10.05 | 11.15 | 29.66 | 4.38 | 10.13 |

**Item 15 was removed from scoring.*

**Table 6.13 Item Statistics—Mathematics Grade 5 Computer-Based Test Administration**

| Mathematics Grade 5 Computer-Based Test Administration | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item* | 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 |
| 1 | MC | ≥ 53,310 | ≥ 53,260 | 0.84 | 0.42 | 0.07 | | | | | | | |
| 2 | ESR | ≥ 53,310 | ≥ 53,240 | 0.53 | 0.34 | 0.13 | 30.55 | 32.82 | 36.51 | | | | |
| 3 | MC | ≥ 53,310 | ≥ 53,220 | 0.41 | 0.24 | 0.14 | | | | | | | |
| 4 | MS | ≥ 53,310 | ≥ 53,200 | 0.21 | 0.45 | 0.20 | 78.38 | 21.43 | | | | | |
| 5 | SA | ≥ 53,310 | ≥ 53,280 | 0.84 | 0.26 | 0.06 | 16.45 | 83.49 | | | | | |
| 6 | MC | ≥ 53,310 | ≥ 53,260 | 0.77 | 0.36 | 0.06 | | | | | | | |
| 8 | SA | ≥ 53,310 | ≥ 52,870 | 0.16 | 0.52 | 0.83 | 83.55 | 15.62 | | | | | |
| 9 | MC | ≥ 53,310 | ≥ 53,260 | 0.82 | 0.27 | 0.07 | | | | | | | |
| 10 | SA | ≥ 53,310 | ≥ 53,120 | 0.38 | 0.40 | 0.35 | 61.66 | 37.99 | | | | | |
| 11 | MC | ≥ 53,310 | ≥ 53,250 | 0.64 | 0.42 | 0.08 | | | | | | | |
| 12 | MC | ≥ 53,310 | ≥ 53,160 | 0.27 | 0.33 | 0.26 | | | | | | | |
| 13 | SA | ≥ 53,310 | ≥ 53,040 | 0.59 | 0.45 | 0.51 | 40.54 | 58.95 | | | | | |
| 14 | CR | ≥ 53,310 | ≥ 52,810 | 0.22 | 0.71 | 0.48 | 54.21 | 18.15 | 15.34 | 7.36 | 4.00 | | |
| 15 | CR | ≥ 53,310 | ≥ 52,040 | 0.31 | 0.64 | 1.89 | 46.43 | 22.09 | 18.42 | 10.68 | | | |
| 16 | MC | ≥ 53,310 | ≥ 53,260 | 0.88 | 0.22 | 0.04 | | | | | | | |
| 17 | MC | ≥ 53,310 | ≥ 53,250 | 0.60 | 0.23 | 0.06 | | | | | | | |
| 18 | SA | ≥ 53,310 | ≥ 53,260 | 0.70 | 0.48 | 0.10 | 30.01 | 69.89 | | | | | |
| 19 | MC | ≥ 53,310 | ≥ 53,250 | 0.72 | 0.46 | 0.06 | | | | | | | |
| 20 | ESR | ≥ 53,310 | ≥ 53,270 | 0.74 | 0.39 | 0.07 | 20.05 | 11.94 | 67.95 | | | | |
| 21 | SA | ≥ 53,310 | ≥ 53,190 | 0.59 | 0.51 | 0.23 | 41.19 | 58.58 | | | | | |
| 22 | MC | ≥ 53,310 | ≥ 53,260 | 0.59 | 0.33 | 0.05 | | | | | | | |
| 23 | MC | ≥ 53,310 | ≥ 53,240 | 0.32 | 0.25 | 0.08 | | | | | | | |
| 24 | ESR | ≥ 53,310 | ≥ 53,280 | 0.46 | 0.66 | 0.06 | 40.16 | 26.88 | 32.90 | | | | |
| 25 | MC | ≥ 53,310 | ≥ 53,250 | 0.25 | 0.25 | 0.07 | | | | | | | |
| 26 | SA | ≥ 53,310 | ≥ 53,230 | 0.49 | 0.29 | 0.14 | 50.81 | 49.04 | | | | | |
| 27 | MC | ≥ 53,310 | ≥ 53,240 | 0.54 | 0.53 | 0.09 | | | | | | | |
| 28 | CR | ≥ 53,310 | ≥ 52,960 | 0.28 | 0.52 | 0.27 | 35.98 | 47.69 | 11.34 | 4.35 | | | |
| 29 | CR | ≥ 53,310 | ≥ 52,500 | 0.18 | 0.55 | 0.83 | 61.37 | 26.26 | 6.32 | 4.52 | | | |
| 30 | SA | ≥ 53,310 | ≥ 53,230 | 0.64 | 0.54 | 0.15 | 35.54 | 64.31 | | | | | |
| 31 | MC | ≥ 53,310 | ≥ 53,220 | 0.51 | 0.46 | 0.08 | | | | | | | |
| 32 | SA | ≥ 53,310 | ≥ 53,220 | 0.60 | 0.35 | 0.17 | 40.22 | 59.61 | | | | | |
| 33 | MS | ≥ 53,310 | ≥ 53,240 | 0.64 | 0.35 | 0.14 | 35.61 | 64.26 | | | | | |
| 34 | MC | ≥ 53,310 | ≥ 53,240 | 0.18 | 0.03 | 0.05 | | | | | | | |
| 35 | MC | ≥ 53,310 | ≥ 53,220 | 0.22 | 0.23 | 0.08 | | | | | | | |
| 36 | MC | ≥ 53,310 | ≥ 53,200 | 0.22 | 0.32 | 0.12 | | | | | | | |
| 37 | MC | ≥ 53,310 | ≥ 53,230 | 0.50 | 0.46 | 0.07 | | | | | | | |
| 38 | MS | ≥ 53,310 | ≥ 53,200 | 0.23 | 0.48 | 0.21 | 77.16 | 22.63 | | | | | |

| Mathematics Grade 5 Computer-Based Test Administration (continued) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | 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 |
| 39 | SA | ≥ 53,310 | ≥ 53,100 | 0.39 | 0.24 | 0.39 | 60.84 | 38.77 | | | | | |
| 40 | MC | ≥ 53,310 | ≥ 53,220 | 0.31 | 0.35 | 0.08 | | | | | | | |
| 41 | MC | ≥ 53,310 | ≥ 53,210 | 0.64 | 0.33 | 0.10 | | | | | | | |
| 42 | CR | ≥ 53,310 | ≥ 52,600 | 0.14 | 0.57 | 0.62 | 66.51 | 25.38 | 5.64 | 1.13 | | | |
| 43 | CR | ≥ 53,310 | ≥ 52,690 | 0.24 | 0.72 | 0.57 | 42.98 | 12.90 | 15.79 | 16.94 | 5.14 | 2.82 | 2.25 |

**Item 7 was removed from scoring.*

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

| Mathematics Grade 6 Computer-Based Test Administration | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item* | 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 |
| 1 | MC | ≥ 52,350 | ≥ 52,320 | 0.93 | 0.19 | 0.01 | | | | | | | |
| 2 | SA | ≥ 52,350 | ≥ 52,260 | 0.71 | 0.47 | 0.16 | 28.90 | 70.94 | | | | | |
| 3 | MC | ≥ 52,350 | ≥ 52,290 | 0.33 | 0.40 | 0.06 | | | | | | | |
| 4 | MS | ≥ 52,350 | ≥ 52,280 | 0.14 | 0.28 | 0.12 | 85.62 | 14.26 | | | | | |
| 5 | SA | ≥ 52,350 | ≥ 52,250 | 0.52 | 0.40 | 0.19 | 48.01 | 51.79 | | | | | |
| 6 | MC | ≥ 52,350 | ≥ 52,280 | 0.58 | 0.49 | 0.08 | | | | | | | |
| 7 | MS | ≥ 52,350 | ≥ 52,280 | 0.21 | 0.46 | 0.12 | 78.58 | 21.30 | | | | | |
| 8 | MC | ≥ 52,350 | ≥ 52,260 | 0.63 | 0.43 | 0.12 | | | | | | | |
| 9 | MS | ≥ 52,350 | ≥ 52,280 | 0.54 | 0.45 | 0.13 | 45.59 | 54.28 | | | | | |
| 10 | MC | ≥ 52,350 | ≥ 52,280 | 0.55 | 0.50 | 0.08 | | | | | | | |
| 11 | SA | ≥ 52,350 | ≥ 52,070 | 0.52 | 0.35 | 0.53 | 48.05 | 51.42 | | | | | |
| 12 | MC | ≥ 52,350 | ≥ 52,290 | 0.73 | 0.33 | 0.07 | | | | | | | |
| 13 | MS | ≥ 52,350 | ≥ 52,290 | 0.22 | 0.30 | 0.11 | 77.86 | 22.03 | | | | | |
| 15 | MS | ≥ 52,350 | ≥ 52,290 | 0.63 | 0.43 | 0.11 | 37.37 | 62.52 | | | | | |
| 16 | MC | ≥ 52,350 | ≥ 52,280 | 0.73 | 0.40 | 0.09 | | | | | | | |
| 17 | SA | ≥ 52,350 | ≥ 51,950 | 0.29 | 0.28 | 0.75 | 70.25 | 29.00 | | | | | |
| 18 | MC | ≥ 52,350 | ≥ 52,260 | 0.31 | 0.19 | 0.13 | | | | | | | |
| 19 | MS | ≥ 52,350 | ≥ 52,250 | 0.13 | 0.34 | 0.18 | 87.12 | 12.70 | | | | | |
| 20 | MC | ≥ 52,350 | ≥ 52,240 | 0.80 | 0.39 | 0.16 | | | | | | | |
| 21 | MC | ≥ 52,350 | ≥ 52,270 | 0.75 | 0.35 | 0.07 | | | | | | | |
| 22 | SA | ≥ 52,350 | ≥ 52,040 | 0.20 | 0.50 | 0.59 | 79.84 | 19.56 | | | | | |
| 23 | MS | ≥ 52,350 | ≥ 52,280 | 0.24 | 0.51 | 0.13 | 75.71 | 24.16 | | | | | |
| 24 | SA | ≥ 52,350 | ≥ 52,220 | 0.27 | 0.55 | 0.23 | 54.92 | 35.01 | 9.84 | | | | |
| 25 | MC | ≥ 52,350 | ≥ 52,230 | 0.43 | 0.51 | 0.14 | | | | | | | |
| 26 | MS | ≥ 52,350 | ≥ 52,250 | 0.48 | 0.58 | 0.18 | 51.74 | 48.08 | | | | | |
| 27 | SA | ≥ 52,350 | ≥ 51,700 | 0.17 | 0.54 | 1.24 | 75.35 | 12.85 | 10.57 | | | | |
| 29 | CR | ≥ 52,350 | ≥ 51,140 | 0.20 | 0.62 | 1.63 | 67.38 | 11.20 | 9.06 | 10.05 | | | |
| 30 | CR | ≥ 52,350 | ≥ 51,350 | 0.40 | 0.62 | 1.21 | 36.86 | 10.48 | 13.56 | 30.48 | 6.70 | | |
| 31 | CR | ≥ 52,350 | ≥ 51,230 | 0.10 | 0.46 | 1.60 | 81.81 | 6.17 | 6.32 | 3.57 | | | |
| 32 | CR | ≥ 52,350 | ≥ 50,050 | 0.19 | 0.65 | 3.47 | 51.76 | 35.18 | 6.13 | 2.53 | | | |
| 33 | MC | ≥ 52,350 | ≥ 52,260 | 0.83 | 0.24 | 0.05 | | | | | | | |
| 34 | SA | ≥ 52,350 | ≥ 52,220 | 0.62 | 0.45 | 0.25 | 38.01 | 61.74 | | | | | |
| 35 | SA | ≥ 52,350 | ≥ 52,230 | 0.65 | 0.51 | 0.22 | 17.79 | 33.51 | 48.48 | | | | |
| 36 | MS | ≥ 52,350 | ≥ 52,250 | 0.27 | 0.49 | 0.19 | 73.07 | 26.75 | | | | | |
| 37 | MC | ≥ 52,350 | ≥ 52,190 | 0.57 | 0.16 | 0.18 | | | | | | | |
| 38 | MC | ≥ 52,350 | ≥ 52,230 | 0.29 | 0.40 | 0.11 | | | | | | | |

| Mathematics Grade 6 Computer-Based Test Administration (continued) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Item** | **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** |
| 39 | ESR | ≥ 52,350 | ≥ 52,240 | 0.65 | 0.48 | 0.20 | 13.18 | 43.97 | 42.64 | | | | |
| 40 | MS | ≥ 52,350 | ≥ 52,260 | 0.21 | 0.38 | 0.17 | 78.92 | 20.91 | | | | | |
| 41 | CR | ≥ 52,350 | ≥ 52,140 | 0.50 | 0.63 | 0.39 | 8.58 | 55.45 | 14.07 | 21.50 | | | |
| 42 | CR | ≥ 52,350 | ≥ 52,130 | 0.11 | 0.54 | 0.42 | 77.49 | 10.95 | 3.93 | 3.76 | 3.45 | | |
| 43 | CR | ≥ 52,350 | ≥ 51,500 | 0.20 | 0.76 | 1.11 | 68.09 | 5.41 | 3.28 | 3.75 | 3.00 | 5.44 | 9.41 |

**Items 14 and 28 were removed from scoring.*

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

| Mathematics Grade 7 Computer-Based Test Administration | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | 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 |
| 1 | MC | ≥ 51,800 | ≥ 51,730 | 0.78 | 0.36 | 0.07 | | | | | | | |
| 2 | MC | ≥ 51,800 | ≥ 51,730 | 0.64 | 0.42 | 0.08 | | | | | | | |
| 3 | SA | ≥ 51,800 | ≥ 51,640 | 0.44 | 0.61 | 0.29 | 55.48 | 44.23 | | | | | |
| 4 | MC | ≥ 51,800 | ≥ 51,700 | 0.55 | 0.49 | 0.13 | | | | | | | |
| 5 | MS | ≥ 51,800 | ≥ 51,690 | 0.18 | 0.43 | 0.19 | 81.48 | 18.32 | | | | | |
| 6 | MC | ≥ 51,800 | ≥ 51,720 | 0.62 | 0.48 | 0.10 | | | | | | | |
| 7 | MS | ≥ 51,800 | ≥ 51,730 | 0.22 | 0.57 | 0.13 | 77.70 | 22.17 | | | | | |
| 8 | SA | ≥ 51,800 | ≥ 51,650 | 0.33 | 0.45 | 0.28 | 66.72 | 33.00 | | | | | |
| 9 | MC | ≥ 51,800 | ≥ 51,720 | 0.30 | 0.09 | 0.10 | | | | | | | |
| 10 | SA | ≥ 51,800 | ≥ 51,610 | 0.57 | 0.53 | 0.35 | 42.63 | 57.02 | | | | | |
| 11 | MC | ≥ 51,800 | ≥ 51,700 | 0.38 | 0.21 | 0.13 | | | | | | | |
| 12 | MC | ≥ 51,800 | ≥ 51,720 | 0.49 | 0.47 | 0.09 | | | | | | | |
| 13 | SA | ≥ 51,800 | ≥ 51,640 | 0.40 | 0.47 | 0.29 | 59.65 | 40.05 | | | | | |
| 14 | MC | ≥ 51,800 | ≥ 51,720 | 0.58 | 0.44 | 0.09 | | | | | | | |
| 15 | MS | ≥ 51,800 | ≥ 51,630 | 0.12 | 0.45 | 0.31 | 87.71 | 11.98 | | | | | |
| 16 | MC | ≥ 51,800 | ≥ 51,670 | 0.68 | 0.42 | 0.19 | | | | | | | |
| 17 | MS | ≥ 51,800 | ≥ 51,690 | 0.17 | 0.54 | 0.20 | 82.79 | 17.01 | | | | | |
| 18 | MC | ≥ 51,800 | ≥ 51,660 | 0.52 | 0.29 | 0.21 | | | | | | | |
| 19 | MS | ≥ 51,800 | ≥ 51,640 | 0.12 | 0.34 | 0.30 | 87.72 | 11.98 | | | | | |
| 20 | MC | ≥ 51,800 | ≥ 51,640 | 0.49 | 0.45 | 0.24 | | | | | | | |
| 21 | SA | ≥ 51,800 | ≥ 51,570 | 0.56 | 0.55 | 0.44 | 44.02 | 55.53 | | | | | |
| 22 | MC | ≥ 51,800 | ≥ 51,570 | 0.46 | 0.27 | 0.31 | | | | | | | |
| 23 | SA | ≥ 51,800 | ≥ 51,470 | 0.05 | 0.42 | 0.63 | 90.97 | 6.21 | 2.19 | | | | |
| 24 | MC | ≥ 51,800 | ≥ 51,650 | 0.42 | 0.46 | 0.16 | | | | | | | |
| 25 | MS | ≥ 51,800 | ≥ 51,680 | 0.20 | 0.21 | 0.23 | 79.35 | 20.42 | | | | | |
| 26 | SA | ≥ 51,800 | ≥ 51,510 | 0.23 | 0.41 | 0.56 | 61.37 | 30.67 | 7.40 | | | | |
| 27 | MS | ≥ 51,800 | ≥ 51,660 | 0.16 | 0.26 | 0.26 | 83.86 | 15.88 | | | | | |
| 28 | MC | ≥ 51,800 | ≥ 51,610 | 0.36 | 0.28 | 0.24 | | | | | | | |
| 29 | CR | ≥ 51,800 | ≥ 50,470 | 0.19 | 0.61 | 1.84 | 63.89 | 19.69 | 5.67 | 8.19 | | | |
| 30 | CR | ≥ 51,800 | ≥ 49,050 | 0.23 | 0.63 | 3.84 | 44.89 | 27.83 | 11.45 | 7.12 | 3.40 | | |
| 31 | CR | ≥ 51,800 | ≥ 49,170 | 0.07 | 0.42 | 4.25 | 80.96 | 8.76 | 4.38 | 0.82 | | | |
| 32 | CR | ≥ 51,800 | ≥ 49,680 | 0.31 | 0.73 | 3.78 | 47.44 | 20.09 | 15.80 | 12.60 | | | |
| 33 | MC | ≥ 51,800 | ≥ 51,610 | 0.71 | 0.32 | 0.19 | | | | | | | |
| 34 | ESR | ≥ 51,800 | ≥ 51,610 | 0.40 | 0.43 | 0.36 | 39.79 | 39.82 | 20.03 | | | | |
| 35 | MC | ≥ 51,800 | ≥ 51,640 | 0.65 | 0.45 | 0.12 | | | | | | | |
| 37 | SA | ≥ 51,800 | ≥ 51,490 | 0.23 | 0.47 | 0.58 | 76.63 | 22.78 | | | | | |
| 38 | SA | ≥ 51,800 | ≥ 50,930 | 0.16 | 0.60 | 1.66 | 74.63 | 15.87 | 7.83 | | | | |
| 39 | SA | ≥ 51,800 | ≥ 51,160 | 0.21 | 0.40 | 1.22 | 78.13 | 20.66 | | | | | |
| 40 | MC | ≥ 51,800 | ≥ 51,520 | 0.18 | 0.14 | 0.34 | | | | | | | |

| Mathematics Grade 7 Computer-Based Test Administration (continued) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Item** | **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** |
| 41 | CR | ≥ 51,800 | ≥ 50,990 | 0.46 | 0.67 | 0.99 | 31.82 | 21.91 | 19.00 | 25.71 | | | |
| 42 | CR | ≥ 51,800 | ≥ 51,620 | 0.24 | 0.63 | 0.35 | 38.83 | 42.90 | 6.52 | 5.40 | 6.00 | | |
| 43 | CR | ≥ 51,800 | ≥ 51,160 | 0.46 | 0.80 | 1.22 | 22.39 | 13.84 | 10.16 | 9.66 | 13.76 | 18.79 | 10.18 |

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

| Mathematics Grade 8 Computer-Based Test Administration | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | 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 |
| 1 | MC | ≥ 44,710 | ≥ 44,670 | 0.75 | 0.37 | 0.03 | | | | | | | |
| 2 | MC | ≥ 44,710 | ≥ 44,660 | 0.59 | 0.35 | 0.06 | | | | | | | |
| 3 | MS | ≥ 44,710 | ≥ 44,660 | 0.36 | 0.54 | 0.13 | 63.50 | 36.38 | | | | | |
| 4 | MC | ≥ 44,710 | ≥ 44,630 | 0.63 | 0.34 | 0.12 | | | | | | | |
| 5 | SA | ≥ 44,710 | ≥ 44,420 | 0.53 | 0.32 | 0.65 | 46.73 | 52.61 | | | | | |
| 6 | MC | ≥ 44,710 | ≥ 44,650 | 0.46 | 0.32 | 0.08 | | | | | | | |
| 7 | MS | ≥ 44,710 | ≥ 44,660 | 0.34 | 0.51 | 0.13 | 66.18 | 33.69 | | | | | |
| 8 | MC | ≥ 44,710 | ≥ 44,590 | 0.71 | 0.36 | 0.20 | | | | | | | |
| 9 | SA | ≥ 44,710 | ≥ 44,330 | 0.24 | 0.44 | 0.84 | 75.02 | 24.14 | | | | | |
| 10 | MS | ≥ 44,710 | ≥ 44,650 | 0.37 | 0.52 | 0.15 | 62.49 | 37.36 | | | | | |
| 11 | ESR | ≥ 44,710 | ≥ 44,670 | 0.56 | 0.44 | 0.10 | 23.68 | 40.32 | 35.90 | | | | |
| 12 | MC | ≥ 44,710 | ≥ 44,650 | 0.27 | 0.26 | 0.08 | | | | | | | |
| 13 | SA | ≥ 44,710 | ≥ 44,550 | 0.65 | 0.29 | 0.35 | 34.86 | 64.78 | | | | | |
| 14 | MC | ≥ 44,710 | ≥ 44,620 | 0.32 | 0.25 | 0.15 | | | | | | | |
| 15 | MS | ≥ 44,710 | ≥ 44,480 | 0.10 | 0.35 | 0.52 | 89.12 | 10.36 | | | | | |
| 16 | MC | ≥ 44,710 | ≥ 44,630 | 0.62 | 0.22 | 0.12 | | | | | | | |
| 17 | MS | ≥ 44,710 | ≥ 44,640 | 0.23 | 0.36 | 0.16 | 76.72 | 23.12 | | | | | |
| 18 | MC | ≥ 44,710 | ≥ 44,630 | 0.44 | 0.33 | 0.12 | | | | | | | |
| 19 | SA | ≥ 44,710 | ≥ 44,140 | 0.30 | 0.51 | 1.27 | 68.69 | 30.04 | | | | | |
| 20 | MC | ≥ 44,710 | ≥ 44,620 | 0.42 | 0.26 | 0.14 | | | | | | | |
| 21 | MC | ≥ 44,710 | ≥ 44,620 | 0.51 | 0.36 | 0.13 | | | | | | | |
| 22 | SA | ≥ 44,710 | ≥ 44,640 | 0.67 | 0.45 | 0.17 | 16.94 | 31.98 | 50.92 | | | | |
| 23 | SA | ≥ 44,710 | ≥ 44,460 | 0.21 | 0.51 | 0.56 | 78.45 | 20.99 | | | | | |
| 24 | MC | ≥ 44,710 | ≥ 44,630 | 0.29 | 0.23 | 0.05 | | | | | | | |
| 25 | MC | ≥ 44,710 | ≥ 44,580 | 0.52 | 0.32 | 0.15 | | | | | | | |
| 26 | MS | ≥ 44,710 | ≥ 44,630 | 0.18 | 0.35 | 0.19 | 81.96 | 17.85 | | | | | |
| 27 | SA | ≥ 44,710 | ≥ 43,870 | 0.12 | 0.48 | 1.88 | 81.25 | 9.37 | 7.50 | | | | |
| 28 | MC | ≥ 44,710 | ≥ 44,590 | 0.50 | 0.28 | 0.13 | | | | | | | |
| 29 | CR | ≥ 44,710 | ≥ 42,830 | 0.18 | 0.60 | 2.83 | 55.47 | 25.26 | 6.56 | 2.83 | 5.66 | | |
| 30 | CR | ≥ 44,710 | ≥ 41,970 | 0.15 | 0.63 | 4.33 | 66.44 | 16.34 | 7.32 | 3.76 | | | |
| 31 | CR | ≥ 44,710 | ≥ 43,400 | 0.54 | 0.58 | 1.94 | 27.88 | 16.22 | 18.25 | 34.71 | | | |
| 32 | CR | ≥ 44,710 | ≥ 42,310 | 0.12 | 0.45 | 4.03 | 74.42 | 8.36 | 10.27 | 1.56 | | | |
| 33 | MC | ≥ 44,710 | ≥ 44,590 | 0.19 | 0.19 | 0.04 | | | | | | | |
| 34 | SA | ≥ 44,710 | ≥ 44,500 | 0.20 | 0.53 | 0.47 | 79.26 | 20.27 | | | | | |
| 35 | MS | ≥ 44,710 | ≥ 44,550 | 0.16 | 0.43 | 0.35 | 83.52 | 16.13 | | | | | |
| 36 | SA | ≥ 44,710 | ≥ 44,560 | 0.37 | 0.51 | 0.35 | 37.60 | 50.01 | 12.04 | | | | |
| 37 | MS | ≥ 44,710 | ≥ 44,570 | 0.15 | 0.49 | 0.31 | 84.83 | 14.86 | | | | | |
| 38 | SA | ≥ 44,710 | ≥ 44,530 | 0.14 | 0.25 | 0.41 | 78.14 | 15.41 | 6.04 | | | | |

| Mathematics Grade 8 Computer-Based Test Administration (continued) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | 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 |
| 39 | SA | ≥ 44,710 | ≥ 44,430 | 0.22 | 0.51 | 0.62 | 77.53 | 21.85 | | | | | |
| 40 | CR | ≥ 44,710 | ≥ 44,570 | 0.17 | 0.48 | 0.31 | 56.38 | 36.74 | 4.39 | 2.17 | | | |
| 41 | CR | ≥ 44,710 | ≥ 44,160 | 0.12 | 0.26 | 1.21 | 54.78 | 40.62 | 2.42 | 0.57 | 0.39 | | |
| 42 | CR | ≥ 44,710 | ≥ 43,470 | 0.21 | 0.69 | 1.84 | 57.86 | 9.97 | 5.91 | 8.41 | 4.26 | 3.97 | 6.83 |

## 6.2 Item Response Theory

Item parameters for items included in 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 $\theta$ 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 $\theta$ 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 PARSCALE (Muraki & Bock, 2003) was used for the IRT calibrations. PARSCALE is a multipurpose program that implements a variety of IRT models associated with mixed-item formats and associated statistics. PARSCALE has been used to calibrate large data sets, such as those of PARCC and Smarter Balanced assessments. The program implements MML estimation techniques for items and MLE estimation of theta.

## 6.3 Calibration Sample

This section describes the calibration sample in adherence to Standard 1.8 of the AERA, APA, & NCME (2014) Standards. 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, 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.

**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 | % | (Calib % - Census %) |
| 3 | All Students | ≥ 49,840 | 100.00% | ≥ 56,960 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 25,820 | 51.80% | ≥ 29,000 | 50.92% | (0.88%) |
| | Female | ≥ 23,950 | 48.07% | ≥ 27,920 | 49.03% | 0.97% |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 2,750 | 5.53% | ≥ 4,250 | 7.47% | 1.95% |
| | American Indian or Alaska Native | ≥ 370 | 0.74% | ≥ 340 | 0.60% | (0.15%) |
| | Asian | ≥ 650 | 1.31% | ≥ 890 | 1.57% | 0.25% |
| | Black or African American | ≥ 22,630 | 45.41% | ≥ 25,220 | 44.28% | (1.13%) |
| | Native Hawaiian or Other Pacific | ≥ 30 | 0.07% | ≥ 70 | 0.13% | 0.05% |
| | White | ≥ 22,190 | 44.54% | ≥ 24,490 | 43.00% | (1.54%) |
| | Two or More Races | ≥ 1,090 | 2.20% | ≥ 1,630 | 2.87% | 0.67% |
| 4 | All Students | ≥ 55,340 | 100.00% | ≥ 56,230 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 28,500 | 51.51% | ≥ 28,650 | 50.95% | (0.56%) |
| | Female | ≥ 26,750 | 48.33% | ≥ 27,560 | 49.01% | 0.68% |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 3,740 | 6.76% | ≥ 3,900 | 6.94% | 0.18% |
| | American Indian or Alaska Native | ≥ 400 | 0.74% | ≥ 350 | 0.65% | (0.08%) |
| | Asian | ≥ 820 | 1.50% | ≥ 810 | 1.45% | (0.05%) |
| | Black or African American | ≥ 24,760 | 44.74% | ≥ 24,690 | 43.92% | (0.82%) |
| | Native Hawaiian or Other Pacific | ≥ 60 | 0.11% | ≥ 50 | 0.10% | (0.01%) |
| | White | ≥ 24,130 | 43.60% | ≥ 24,850 | 44.16% | 0.56% |
| | Two or More Races | ≥ 1,290 | 2.35% | ≥ 1,530 | 2.71% | 0.37% |
| 5 | All Students | ≥ 53,040 | 100.00% | ≥ 53,300 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 27,170 | 51.22% | ≥ 27,410 | 51.40% | 0.17% |
| | Female | ≥ 25,800 | 48.64% | ≥ 25,890 | 48.60% | (0.03%) |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 3,570 | 6.73% | ≥ 3,530 | 6.63% | (0.10%) |
| | American Indian or Alaska Native | ≥ 380 | 0.72% | ≥ 320 | 0.62% | (0.10%) |
| | Asian | ≥ 770 | 1.47% | ≥ 820 | 1.54% | 0.07% |
| | Black or African American | ≥ 23,650 | 44.60% | ≥ 23,670 | 44.45% | (0.15%) |
| | Native Hawaiian or Other Pacific | ≥ 50 | 0.11% | ≥ 50 | 0.10% | (0.01%) |
| | White | ≥ 23,230 | 43.80% | ≥ 23,580 | 44.19% | 0.39% |
| | Two or More Races | ≥ 1,250 | 2.37% | ≥ 1,320 | 2.47% | 0.10% |

| Calibration and Census Data: English Language Arts (continued) | | | | | | |
|---|---|---|---|---|---|---|
| | | Calibration Sample | | Census Data | | |
| Grade | | N | % | N | % | (Calib % - Census %) |
| 6 | All Students | ≥ 42,200 | 100.00% | ≥ 52,580 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 21,800 | 51.66% | ≥ 27,100 | 51.55% | (0.12%) |
| | Female | ≥ 20,310 | 48.14% | ≥ 25,470 | 48.45% | 0.31% |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 2,400 | 5.70% | ≥ 3,320 | 6.32% | 0.62% |
| | American Indian or Alaska Native | ≥ 300 | 0.73% | ≥ 360 | 0.70% | (0.02%) |
| | Asian | ≥ 550 | 1.31% | ≥ 800 | 1.53% | 0.23% |
| | Black or African American | ≥ 19,080 | 45.22% | ≥ 23,230 | 44.20% | (1.02%) |
| | Native Hawaiian or Other Pacific | ≥ 40 | 0.10% | ≥ 40 | 0.08% | (0.03%) |
| | White | ≥ 18,720 | 44.37% | ≥ 23,670 | 45.02% | 0.65% |
| | Two or More Races | ≥ 980 | 2.32% | ≥ 1,130 | 2.15% | (0.17%) |
| 7 | All Students | ≥ 51,600 | 100.00% | ≥ 51,930 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 26,800 | 51.93% | ≥ 26,480 | 50.99% | (0.95%) |
| | Female | ≥ 24,730 | 47.93% | ≥ 25,450 | 49.01% | 1.08% |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 3,540 | 6.86% | ≥ 3,150 | 6.08% | (0.78%) |
| | American Indian or Alaska Native | ≥ 350 | 0.69% | ≥ 390 | 0.78% | 0.09% |
| | Asian | ≥ 780 | 1.52% | ≥ 850 | 1.64% | 0.13% |
| | Black or African American | ≥ 23,290 | 45.13% | ≥ 23,040 | 44.41% | (0.72%) |
| | Native Hawaiian or Other Pacific | ≥ 40 | 0.09% | ≥ 40 | 0.09% | 0.00% |
| | White | ≥ 22,250 | 43.12% | ≥ 23,460 | 45.12% | 2.00% |
| | Two or More Races | ≥ 1,230 | 2.39% | ≥ 970 | 1.87% | (0.53%) |
| 8 | All Students | ≥ 50,050 | 100.00% | ≥ 50,450 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 25,480 | 50.93% | ≥ 25,830 | 51.23% | 0.30% |
| | Female | ≥ 23,950 | 47.86% | ≥ 24,610 | 48.77% | 0.91% |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 3,290 | 6.59% | ≥ 2,960 | 5.89% | (0.70%) |
| | American Indian or Alaska Native | ≥ 320 | 0.65% | ≥ 350 | 0.72% | 0.08% |
| | Asian | ≥ 760 | 1.54% | ≥ 820 | 1.62% | 0.09% |
| | Black or African American | ≥ 22,350 | 44.67% | ≥ 22,270 | 44.12% | (0.55%) |
| | Native Hawaiian or Other Pacific | ≥ 50 | 0.11% | ≥ 40 | 0.08% | (0.03%) |
| | White | ≥ 21,440 | 42.85% | ≥ 23,130 | 45.85% | 3.00% |
| | Two or More Races | ≥ 1,170 | 2.35% | ≥ 860 | 1.72% | (0.63%) |

**Table 6.18 Summary of Calibration and Census Data: Mathematics**

| Calibration and Census Data: Mathematics | | | | | | |
|---|---|---|---|---|---|---|
| | | Calibration Sample | | Census Data | | |
| Grade | | N | % | N | % | (Calib % - Census %) |
| 3 | All Students | ≥ 56,820 | 100.00% | ≥ 56,800 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 28,780 | 50.65% | ≥ 28,920 | 50.91% | 0.25% |
| | Female | ≥ 27,380 | 48.20% | ≥ 27,840 | 49.03% | 0.84% |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 3,790 | 6.68% | ≥ 4,220 | 7.46% | 0.77% |
| | American Indian or Alaska Native | ≥ 400 | 0.71% | ≥ 320 | 0.60% | (0.11%) |
| | Asian | ≥ 820 | 1.44% | ≥ 890 | 1.57% | 0.12% |
| | Black or African American | ≥ 25,010 | 44.02% | ≥ 25,150 | 44.27% | 0.25% |
| | Native Hawaiian or Other Pacific | ≥ 60 | 0.11% | ≥ 70 | 0.13% | 0.01% |
| | White | ≥ 24,760 | 43.58% | ≥ 24,430 | 42.99% | (0.60%) |
| | Two or More Races | ≥ 1,260 | 2.23% | ≥ 1,630 | 2.88% | 0.65% |
| 4 | All Students | ≥ 55,450 | 100.00% | ≥ 56,230 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 28,550 | 51.49% | ≥ 28,640 | 50.92% | (0.57%) |
| | Female | ≥ 26,810 | 48.35% | ≥ 27,560 | 49.02% | 0.67% |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 3,740 | 6.76% | ≥ 3,890 | 6.93% | 0.17% |
| | American Indian or Alaska Native | ≥ 410 | 0.74% | ≥ 350 | 0.65% | (0.09%) |
| | Asian | ≥ 820 | 1.49% | ≥ 810 | 1.45% | (0.04%) |
| | Black or African American | ≥ 24,810 | 44.75% | ≥ 24,690 | 43.91% | (0.83%) |
| | Native Hawaiian or Other Pacific | ≥ 60 | 0.11% | ≥ 50 | 0.10% | (0.01%) |
| | White | ≥ 24,170 | 43.60% | ≥ 24,860 | 44.16% | 0.57% |
| | Two or More Races | ≥ 1,300 | 2.35% | ≥ 1,530 | 2.72% | 0.37% |
| 5 | All Students | ≥ 22,940 | 100.00% | ≥ 53,310 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 11,770 | 51.31% | ≥ 27,410 | 51.40% | 0.09% |
| | Female | ≥ 11,110 | 48.44% | ≥ 25,890 | 48.60% | 0.16% |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 1,420 | 6.21% | ≥ 3,530 | 6.63% | 0.42% |
| | American Indian or Alaska Native | ≥ 180 | 0.80% | ≥ 320 | 0.62% | (0.18%) |
| | Asian | ≥ 360 | 1.58% | ≥ 820 | 1.54% | (0.04%) |
| | Black or African American | ≥ 9,530 | 41.55% | ≥ 23,670 | 44.46% | 2.91% |
| | Native Hawaiian or Other Pacific | ≥ 20 | 0.10% | ≥ 50 | 0.10% | (0.00%) |
| | White | ≥ 10,780 | 47.03% | ≥ 23,570 | 44.18% | (2.85%) |
| | Two or More Races | ≥ 540 | 2.37% | ≥ 1,320 | 2.47% | 0.10% |

| Calibration and Census Data: Mathematics (continued) | | | | | | |
|---|---|---|---|---|---|---|
| | | Calibration Sample | | Census Data | | |
| Grade | | N | % | N | % | (Calib % - Census %) |
| 6 | All Students | ≥ 28,810 | 100.00% | ≥ 52,350 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 14,870 | 51.61% | ≥ 26,960 | 51.53% | (0.08%) |
| | Female | ≥ 13,880 | 48.17% | ≥ 25,380 | 48.47% | 0.30% |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 1,590 | 5.54% | ≥ 3,300 | 6.32% | 0.77% |
| | American Indian or Alaska Native | ≥ 210 | 0.75% | ≥ 350 | 0.70% | (0.04%) |
| | Asian | ≥ 380 | 1.32% | ≥ 800 | 1.53% | 0.21% |
| | Black or African American | ≥ 12,660 | 43.96% | ≥ 23,110 | 44.18% | 0.22% |
| | Native Hawaiian or Other Pacific | ≥ 30 | 0.11% | ≥ 30 | 0.08% | (0.03%) |
| | White | ≥ 13,160 | 45.70% | ≥ 23,600 | 45.03% | (0.66%) |
| | Two or More Races | ≥ 670 | 2.33% | ≥ 1,130 | 2.15% | (0.17%) |
| 7 | All Students | ≥ 51,500 | 100.00% | ≥ 51,800 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 26,570 | 51.59% | ≥ 26,420 | 51.00% | (0.59%) |
| | Female | ≥ 24,840 | 48.24% | ≥ 25,380 | 49.00% | 0.76% |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 2,940 | 5.72% | ≥ 3,150 | 6.09% | 0.37% |
| | American Indian or Alaska Native | ≥ 370 | 0.72% | ≥ 390 | 0.78% | 0.06% |
| | Asian | ≥ 690 | 1.36% | ≥ 840 | 1.64% | 0.28% |
| | Black or African American | ≥ 23,110 | 44.88% | ≥ 23,000 | 44.46% | (0.41%) |
| | Native Hawaiian or Other Pacific | ≥ 30 | 0.07% | ≥ 40 | 0.09% | 0.02% |
| | White | ≥ 23,020 | 44.70% | ≥ 23,370 | 45.07% | 0.37% |
| | Two or More Races | ≥ 1,200 | 2.34% | ≥ 960 | 1.86% | (0.47%) |
| 8 | All Students | ≥ 44,400 | 100.00% | ≥ 44,710 | 100.00% | 0.00% |
| | Gender | | | | | |
| | Male | ≥ 22,800 | 51.36% | ≥ 23,090 | 51.68% | 0.32% |
| | Female | ≥ 21,510 | 48.46% | ≥ 21,610 | 48.32% | (0.14%) |
| | Race Ethnicity | | | | | |
| | Hispanic/Latino | ≥ 2,600 | 5.86% | ≥ 2,680 | 6.00% | 0.14% |
| | American Indian or Alaska Native | ≥ 310 | 0.72% | ≥ 320 | 0.75% | 0.03% |
| | Asian | ≥ 580 | 1.32% | ≥ 570 | 1.27% | (0.05%) |
| | Black or African American | ≥ 20,040 | 45.14% | ≥ 21,080 | 47.10% | 1.96% |
| | Native Hawaiian or Other Pacific | ≥ 30 | 0.08% | ≥ 30 | 0.08% | 0.01% |
| | White | ≥ 19,720 | 44.42% | ≥ 19,250 | 43.09% | (1.33%) |
| | Two or More Races | ≥ 990 | 2.24% | ≥ 760 | 1.70% | (0.54%) |

### 6.4 Calibration and Linking

All 2017 LEAP 2025 item calibration and linking were performed based on IRT.

Calibration and linking methodology used for the Spring 2017 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 PARSCALE (Muraki & Bock, 2003). 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 http://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 applied to generate final scoring tables. All PARSCALE 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* (see https://parcc-assessment.org/wp-content/uploads/2018/02/PARCC-2015-Tech-Report.pdf) and *Final Technial Report for 2016 Administration* (see https://parcc-assessment.org/wp-content/uploads/2018/02/PARCC-2016-Tech-Report.pdf).

There were only paper-based tests (PBTs) for the grade 3 ELA and mathematics assessments and only CBTs for the grades 5 through 8 ELA and mathematics assessments. There were two test forms, CBT and PBT, for the 2017 LEAP 2025 grade 4 ELA and mathematics assessments. In general, a school administered the same test mode for ELA and mathematics. For 2017 LEAP 2025 calibration, CBT and PBT were combined and calibrated together for grade 4 due to small sample sizes. 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 2017 LEAP 2025 tests on the PARCC 2016 baseline scale:

1. Calibrate 2017 LEAP 2025 tests.
2. Link 2017 LEAP 2025 tests, except for the grade 6 ELA test, to PARCC 2016 under the non-equivalent common item design.

The 2017 LEAP 2025 forms were linked to the PARCC scale using all intact PARCC items embedded in the 2017 LEAP 2025 tests as anchors by the Stocking & Lord (1983) procedure.

PARCC established a new baseline scale using Spring 2016 tests. LEAP 2015 and 2016 were linked to this new PARCC 2016 baseline scale and 2017 LEAP 2025 also was scaled to the

PARCC 2016 baseline scale. Therefore, all LEAP 2015, LEAP 2016, and 2017 LEAP 2025 were placed on the same PARCC 2016 baseline scale.

### 6.4.1 Calibration of 2017 LEAP 2025 Tests

For 2017 LEAP 2025 item calibration, the 2PL/GPC IRT model was applied to the Louisiana students' calibration samples using the software PARSCALE (Muraki & Bock, 2003). Table 6.19 shows the number of students in the calibration samples and number of calibration items by mode. About 97% of grade 4 students took the PBT, and about 3% of grade 4 students took the CBT. More students in grade 8 took the ELA assessment than the mathematics assessment because high-performing students were allowed to take the EOC Algebra I test instead of the mathematics grade 8 test. For ELA, reading items (RL/RI) in writing prompts are not counted 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 28~35 ELA items and 42~43 mathematics items across grades.

**Table 6.19 Summary of Student Count and Item Count by Test Mode**

| | | N | | | Percentage | | N-Items | |
|---|---|---|---|---|---|---|---|---|
| **Content** | **Grade** | **All** | **CBT** | **PBT** | **CBT** | **PBT** | **CBT** | **PBT** |
| ELA | 3 | ≥ 56,800 | * | ≥ 56,800 | * | 100.00 | * | 30 |
| | 4 | ≥ 56,230 | ≥ 1,930 | ≥ 54,300 | 3.45 | 96.55 | 30 | 30 |
| | 5 | ≥ 53,300 | ≥ 53,300 | * | 100.00 | * | 30 | * |
| | 6 | ≥ 52,370 | ≥ 52,370 | * | 100.00 | * | 36 | * |
| | 7 | ≥ 51,930 | ≥ 51,930 | * | 100.00 | * | 34 | * |
| | 8 | ≥ 50,450 | ≥ 50,450 | * | 100.00 | * | 34 | * |
| Mathematics | 3 | ≥ 56,800 | * | ≥ 56,800 | * | 100.00 | * | 43 |
| | 4 | ≥ 56,230 | ≥ 1,930 | ≥ 54,300 | 3.45 | 96.55 | 43 | 43 |
| | 5 | ≥ 53,310 | ≥ 53,310 | * | 100.00 | * | 43 | * |
| | 6 | ≥ 52,350 | ≥ 52,350 | * | 100.00 | * | 43 | * |
| | 7 | ≥ 51,800 | ≥ 51,800 | * | 100.00 | * | 43 | * |
| | 8 | ≥ 44,710 | ≥ 44,710 | * | 100.00 | * | 42 | * |

**Grade 3 did not have a CBT form. Grades 5–8 did not have a PBT form.*

#### 6.4.1.1. 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. Each ELA test consisted of two PCR tasks; eash task had a WE and a WKL trait. As can be seen in Table 6.20, a single calibration was performed for all grades 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 |

#### 6.4.1.2. 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 together with empirical students’ performances from theta ability -4 to 4. Figure 6.1 shows ICCs for four items suppressed based on IRT item fit. The first plot is for mathematics grade 4 item number 868637. The blue line indicates the ICC, and the red dotted line denotes empirical students’ performances for this item. The students’ ability distribution, similar to the normal distribution, was plotted at the bottom of the figure. The figure clearly shows that ICC did not fit to students’ performances across all theta ranges. The performances in the middle ability range were slightly poorer compared to those in the low-ability range. IRT is based on the assumption that students’ performances should monotonically increase as ability increases. This plot clearly shows this IRT assumption cannot be achieved with this item. A similar pattern was found for mathematics grade 5 MC item number 870788 and mathematics grade 6 MC item number 870820. The fourth plot is for mathematics grade 6 MC item number 870813. Students’ performances were almost flat across ability range. That is, students’ performances were close to guessing value. This implies that students may not have been taught the content of this item. After content experts reviewed the items, these four items were dropped from all analyses and scoring. Item calibration was performed again after dropping these items.

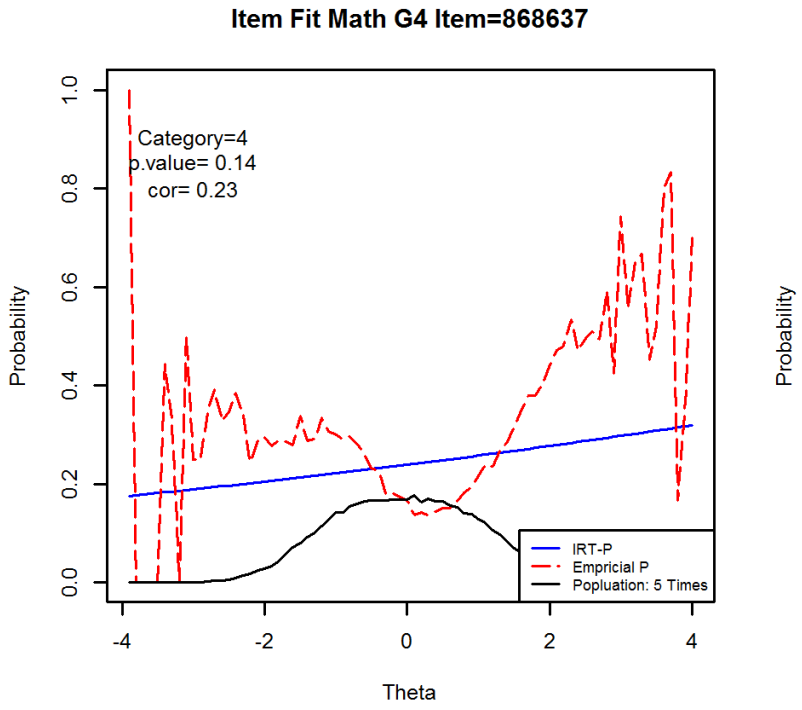
After calibration was redone for mathematics grades 4, 5, and 6, the IRT model fit for all data sets was evaluated by reviewing item chi-square values from PARSCALE (Muraki & Bock, 2003), calculating adjusted fit values and flagging them if > 0.45 (Pearson, 2015).

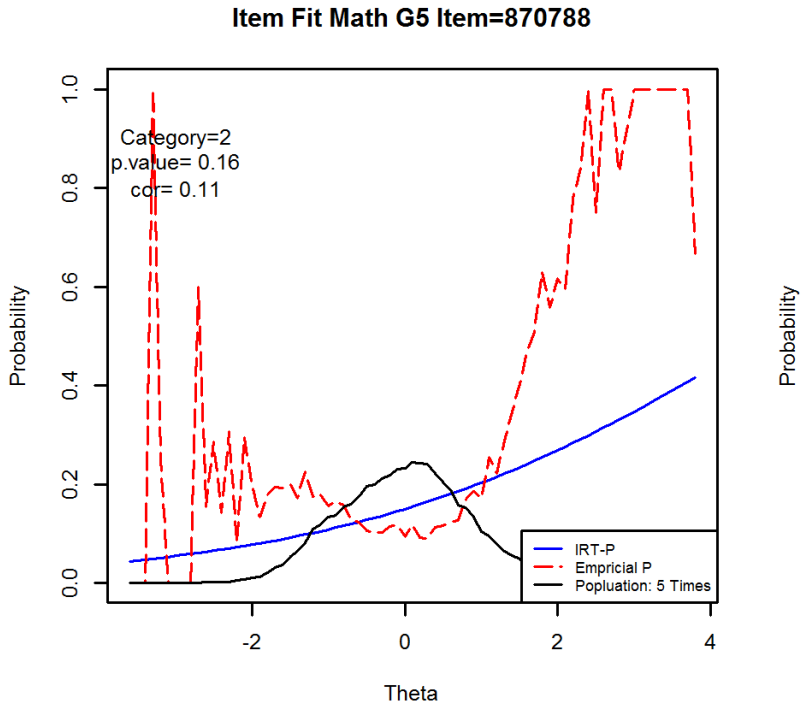
Since chi-square values are sensitive to sample size, these statistics are not easily compared when the number of students varies across items. As a result, adjusted fit values were calculated by dividing the chi-square fit statistic by the sample size using the following formula:

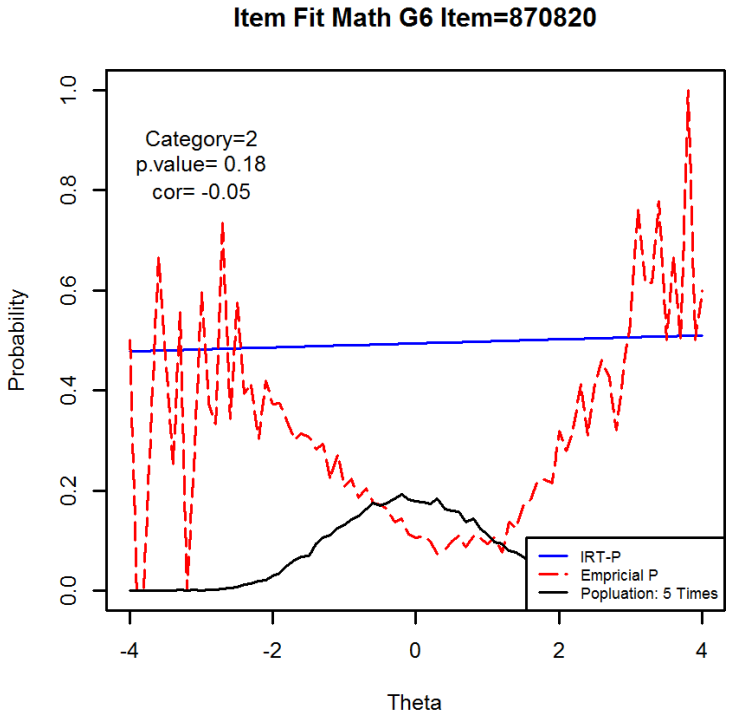
$$C = \sqrt{\frac{\chi^2}{\chi^2+N}}$$

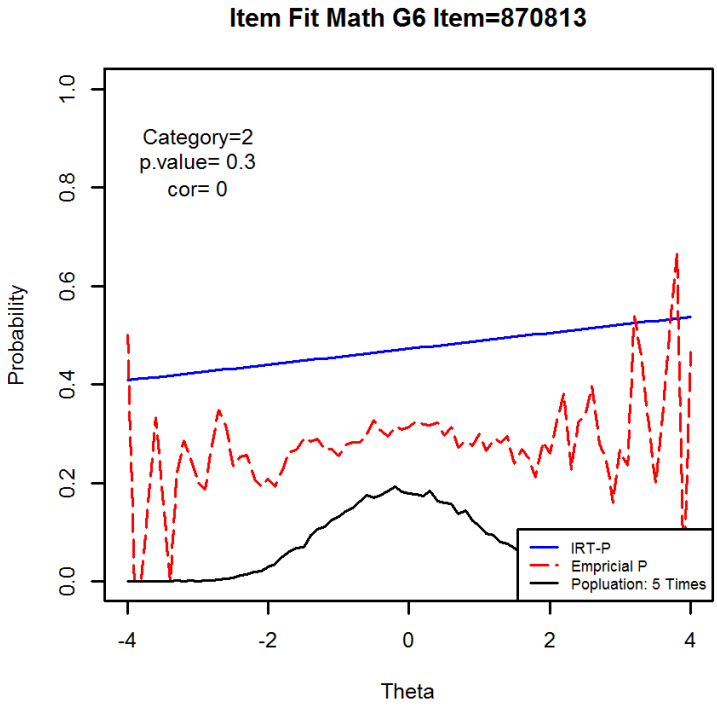
One limitation of the PARSCALE (Muraki & Bock, 2003) output is that when a chi-square value is greater than 9,999.99, PARSCALE does not print the value in the phase two output. Instead, it prints asterisks (*****). When that happens, the chi-square value or adjusted fit is not available.

**Figure 6.1 Item Characteristic Curves for Suppressed Items**







Tables 6.21 and 6.22 show adjusted item fit using PARSCALE chi-square values and calibration sample sizes for ELA and mathematics, respectively (Muraki & Bock, 2003). The average adjusted fit ranged from 0.13 to 0.15 for ELA and 0.08 to 0.10 for mathematics. Items were not excluded based on model fit statistics because the adjusted item fits for all items were lower than the criterion value of 0.45, as can be seen in the maximum values for both ELA and mathematics. The largest adjusted fit value was 0.27 for mathematics grade 7.

**Table 6.21 Summary of Adjusted Fit for ELA**

| Grade | Mode | No. Items | Mean | Std. Dev. | Min. | Max. | No. Flagged Items | No. Not Available |
|---|---|---|---|---|---|---|---|---|
| 3 | PBT | 28 | 0.15 | 0.04 | 0.07 | 0.21 | 0 | 0 |
| 4 | CBT/PBT | 34 | 0.14 | 0.05 | 0.05 | 0.23 | 0 | 0 |
| 5 | CBT | 30 | 0.14 | 0.05 | 0.06 | 0.24 | 0 | 0 |
| 6 | CBT | 35 | 0.13 | 0.04 | 0.06 | 0.21 | 0 | 0 |
| 7 | CBT | 35 | 0.13 | 0.04 | 0.06 | 0.22 | 0 | 0 |
| 8 | CBT | 32 | 0.13 | 0.05 | 0.06 | 0.23 | 0 | 0 |

**Table 6.22 Summary of Adjusted Fit for Mathematics**

| Grade | Mode | No. Items | Mean | Std. Dev. | Min. | Max. | No. Flagged Items | No. Not Available |
|---|---|---|---|---|---|---|---|---|
| 3 | PBT | 43 | 0.08 | 0.03 | 0.04 | 0.18 | 0 | 0 |
| 4 | CBT/PBT | 42 | 0.08 | 0.04 | 0.03 | 0.20 | 0 | 0 |
| 5 | CBT | 42 | 0.09 | 0.03 | 0.04 | 0.18 | 0 | 0 |
| 6 | CBT | 41 | 0.10 | 0.04 | 0.05 | 0.22 | 0 | 0 |
| 7 | CBT | 43 | 0.09 | 0.05 | 0.04 | 0.27 | 0 | 0 |
| 8 | CBT | 42 | 0.09 | 0.04 | 0.04 | 0.21 | 0 | 0 |

### 6.4.2 Linking 2017 LEAP 2025 Grades 3–8 to PARCC Scale

The 2017 LEAP 2025 forms were linked to the PARCC scale using intact PARCC items embedded in the 2017 LEAP 2025 forms as anchors by the Stocking & Lord procedure (1983). There were two sets of anchor item parameters. The first (i.e., long) anchor set included all intact PARCC items, and the second (i.e., short) anchor items were selected by maximizing the same mode item parameters as the test mode. For example, the second anchor of ELA grade 3 consists of most PBT item parameters because the ELA grade 3 test mode was PBT. The first anchor set, which includes both CBT and PBT item parameters, was planned to be used as an operational anchor set. The second anchor set was introduced to check the mode effect of PARCC item parameters by comparing equating results from these two different anchor sets. Table 6.23 summarizes the number of items in the first anchor set and the Stocking & Lord transformation constants for ELA and mathematics grades 3 through 8. The difference between the initial number of anchor items and the final number of anchor items is the number of anchor items dropped.

**Table 6.23 Stocking & Lord Transformation Constants for Linking 2017 LEAP 2025 to PARCC 2016**

| Content | Grade | Initial No. of Anchor Items | Final No. of Anchor Items | Slope | Intercept |
|---|---|---|---|---|---|
| ELA | 3 | 28 | 23 | 0.952 | 0.226 |
| | 4 | 28 | 21 | 0.892 | 0.126 |
| | 5 | 30 | 22 | 0.936 | 0.122 |
| | 6 | N/A | N/A | N/A | N/A |
| | 7 | 32 | 26 | 0.972 | 0.030 |
| | 8 | 30 | 28 | 0.955 | 0.040 |
| Mathematics | 3 | 43 | 37 | 1.007 | -0.040 |
| | 4 | 41 | 38 | 0.976 | 0.001 |
| | 5 | 38 | 33 | 0.918 | -0.161 |
| | 6 | 38 | 35 | 0.984 | -0.206 |
| | 7 | 38 | 34 | 0.988 | -0.120 |
| | 8 | 37 | 35 | 0.947 | 0.022 |

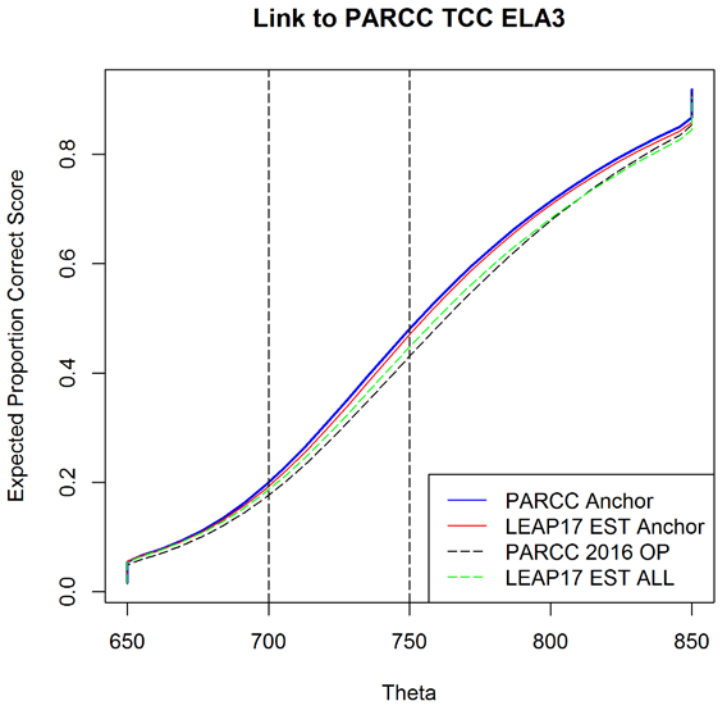
ELA grade 6 results were not included in the equating results and summaries because conventional equating results were not used for scoring. After reviewing the equating results, the Center for Assessment recommended taking a different approach, the calibration approach (Center for Assessment, 2017). Therefore, this report does not include conventional equating results for ELA grade 6.

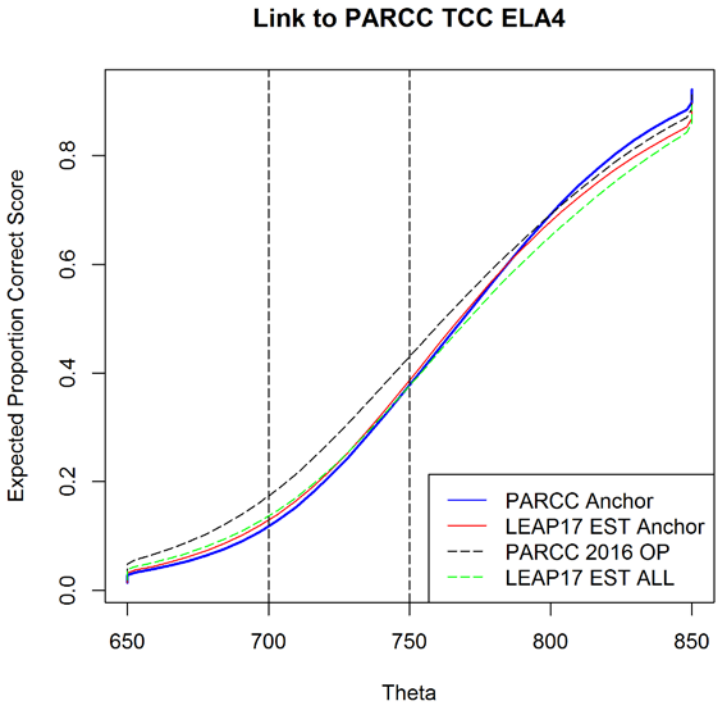
Figures 6.2 and 6.3 show test characteristic curves (TCCs) for PARCC anchor items (PARCC Anchor), corresponding 2017 LEAP 2025 estimated anchor items (LEAP17 EST Anchor), PARCC 2016 operational items (PARCC 2016 OP), and all 2017 LEAP 2025 estimated items (LEAP17 EST ALL) for ELA and mathematics after the Stocking & Lord (1983) equating procedure. For ELA, the four TCCs overlapped for most ability levels across all five grades. For mathematics, the three TCCs also overlapped for most ability levels across all six grades. There were some differences at the extreme ranges, such as low ability or high ability.

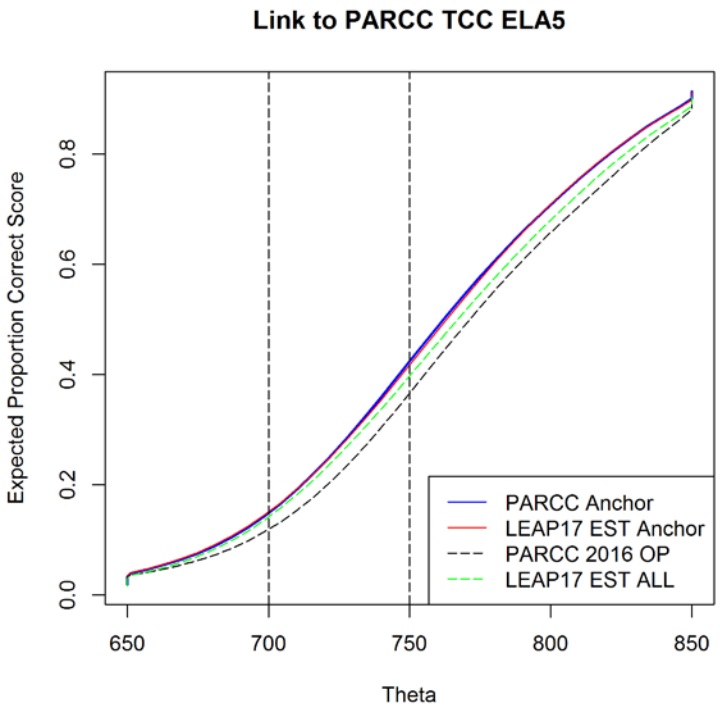
Figures 6.4 and 6.5 present scatterplots of slope item parameters for ELA and mathematics and their correlations after linking the 2017 LEAP 2025 to the PARCC 2016 scale. As can be seen in the ELA plots, most item slope parameters were around the identity line except for two items in ELA grade 5. The correlation between PARCC 2016 and 2017 LEAP 2025 anchor items was 0.766 for ELA grade 5. For the other grades, correlations ranged from 0.905 to 0.985. For mathematics, most item slope parameters were around the identity line and the correlations ranged from 0.795 to 0.933 across grades.

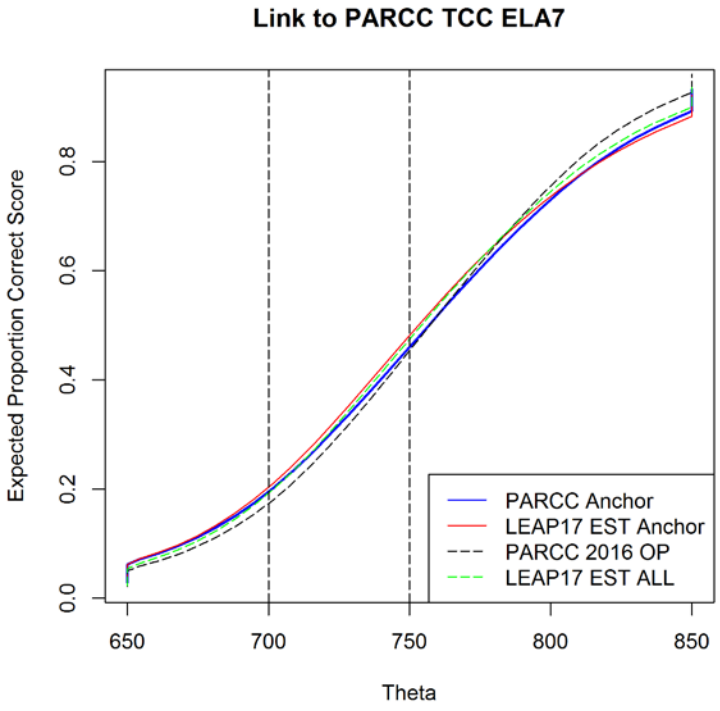
Figures 6.6 and 6.7 present scatterplots of the difficulty item parameters for ELA and mathematics and their correlations after linking the 2017 LEAP 2025 to the PARCC 2016 scale. For ELA, most item difficulty parameters were around the identity line and the correlations ranged from 0.966 to 0.991 across grades. Compared to ELA item difficulty parameters, mathematics item difficulty parameters were farther from the identity line for most grades. Correlations ranged from 0.906 to 0.977 across grades.

**Figure 6.2 ELA TCC between PARCC 2016 Anchor, 2017 LEAP 2025 Anchor, and All LEAP 2025 Items**





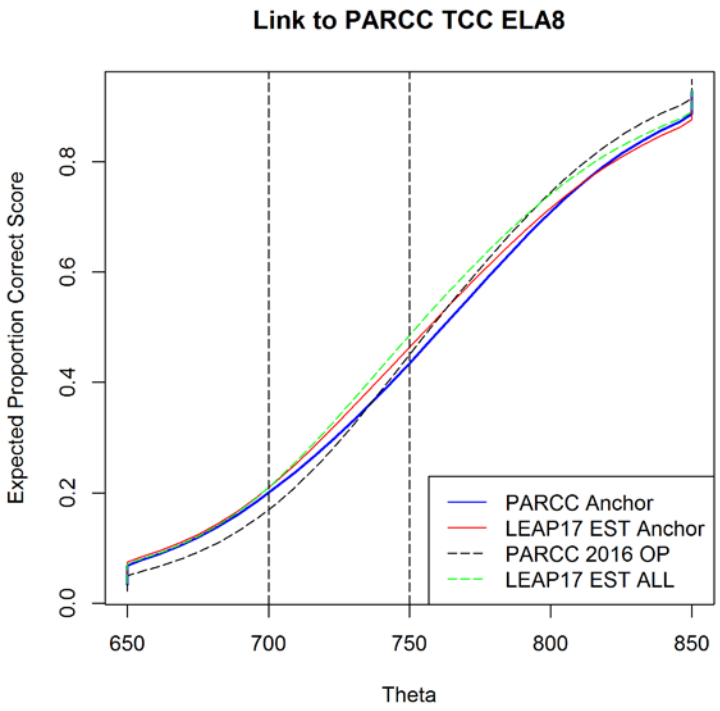
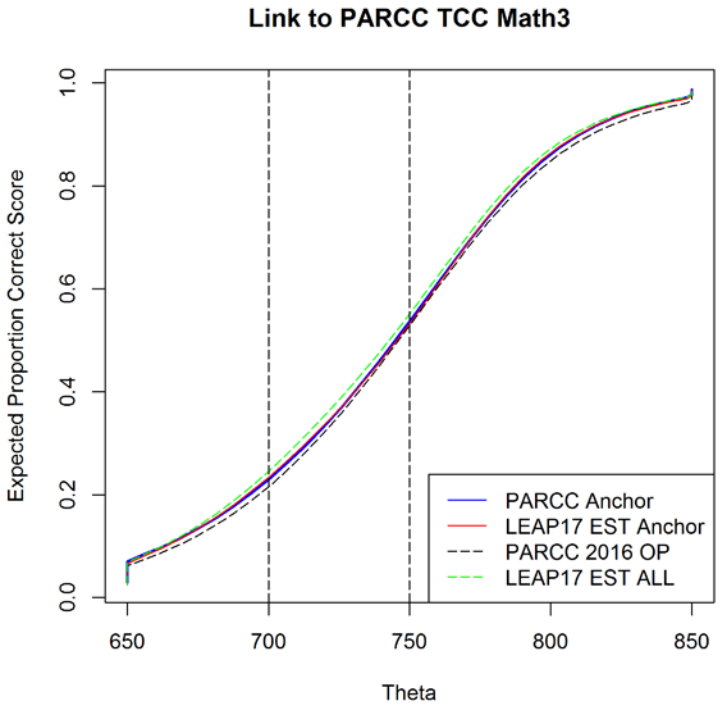
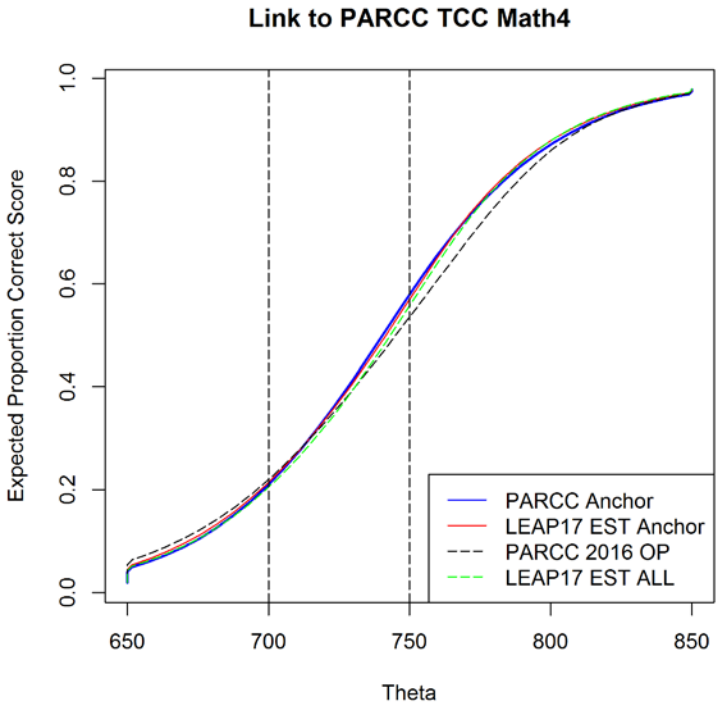
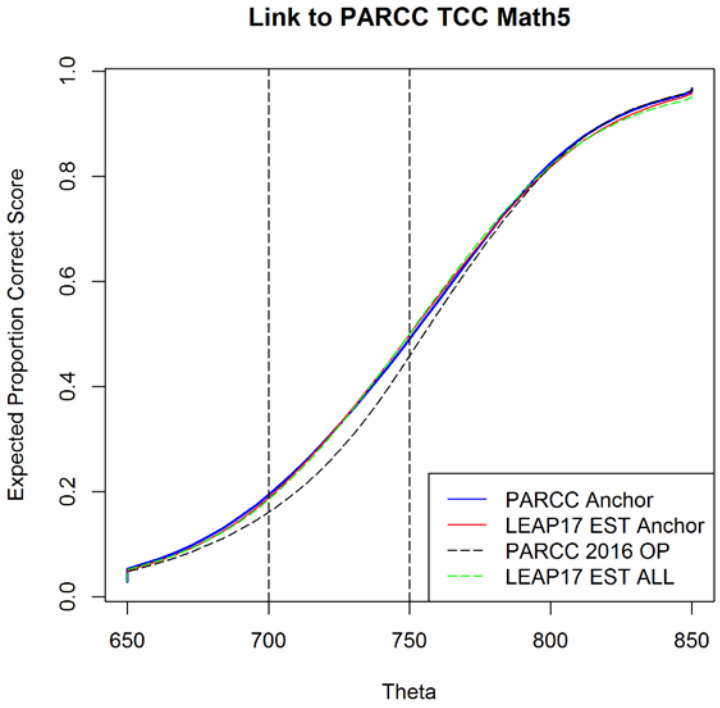
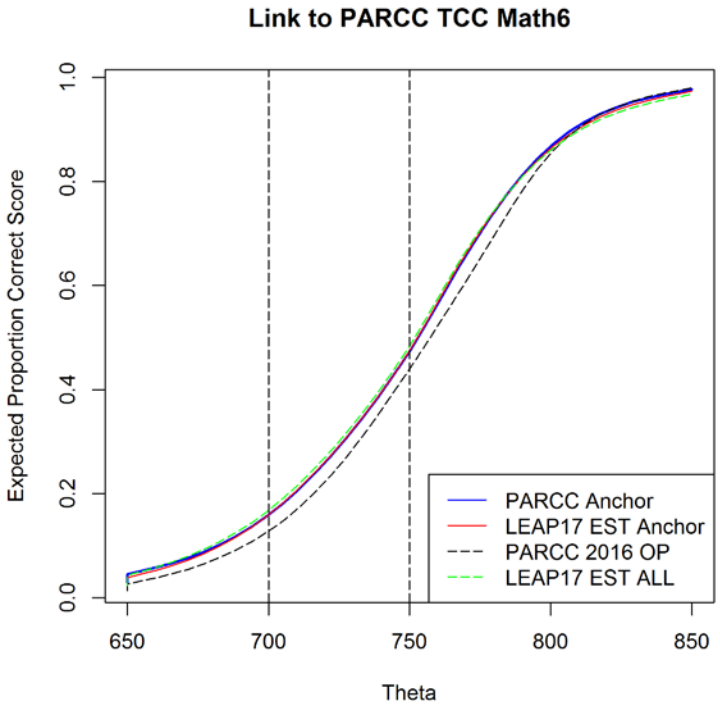


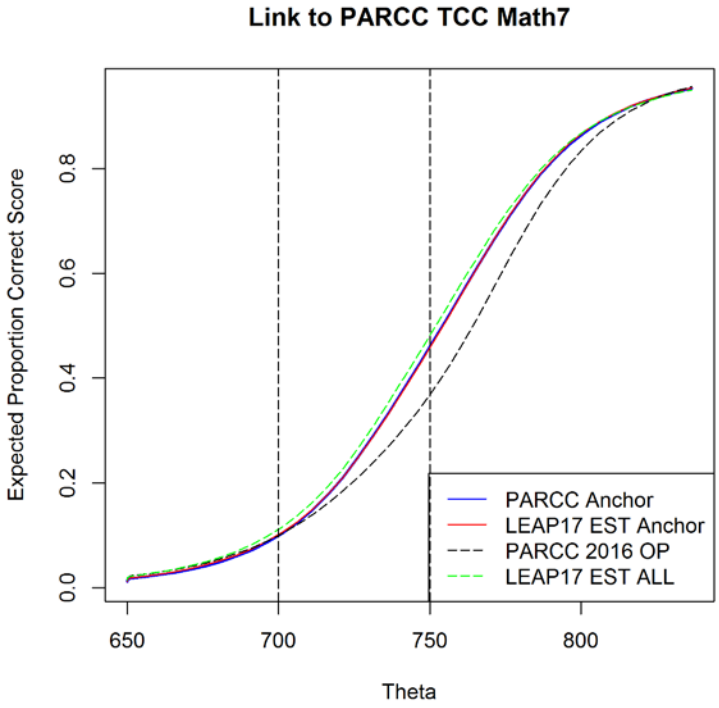
**Figure 6.3 Mathematics TCC between PARCC 2016 Anchor, 2017 LEAP 2025 Anchor, and All LEAP 2025 Items**







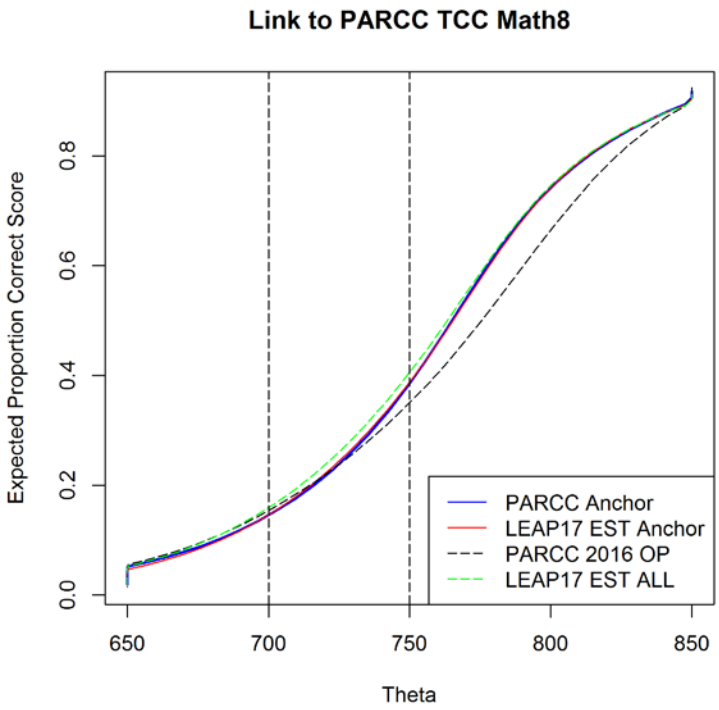
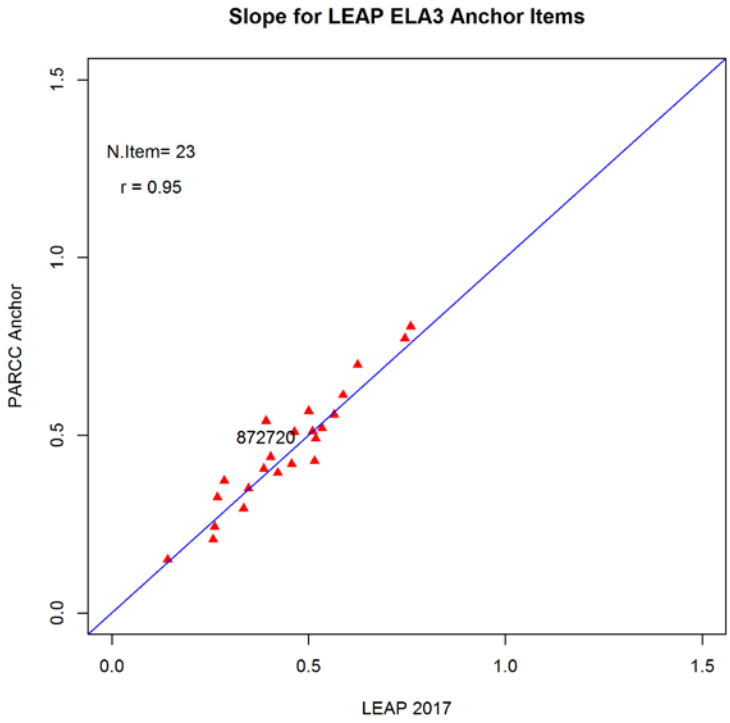
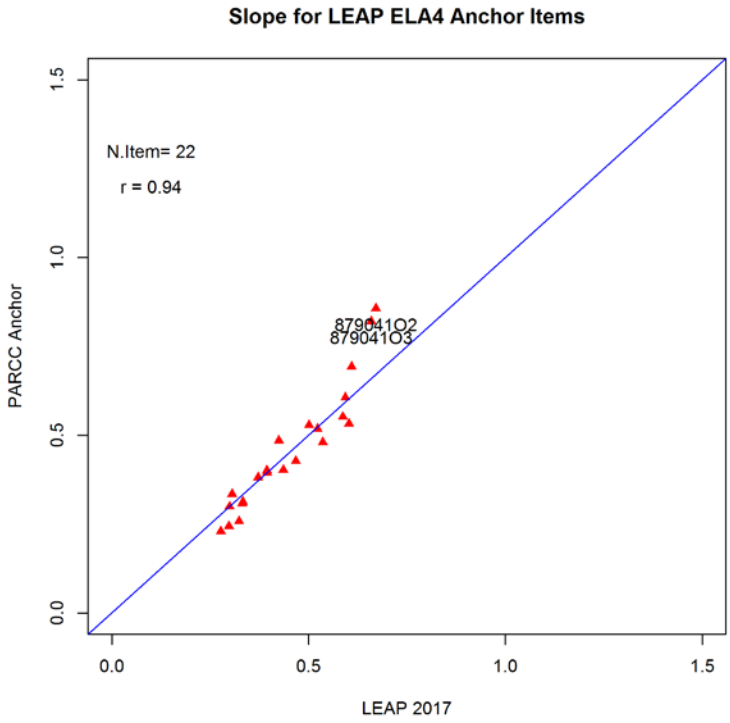
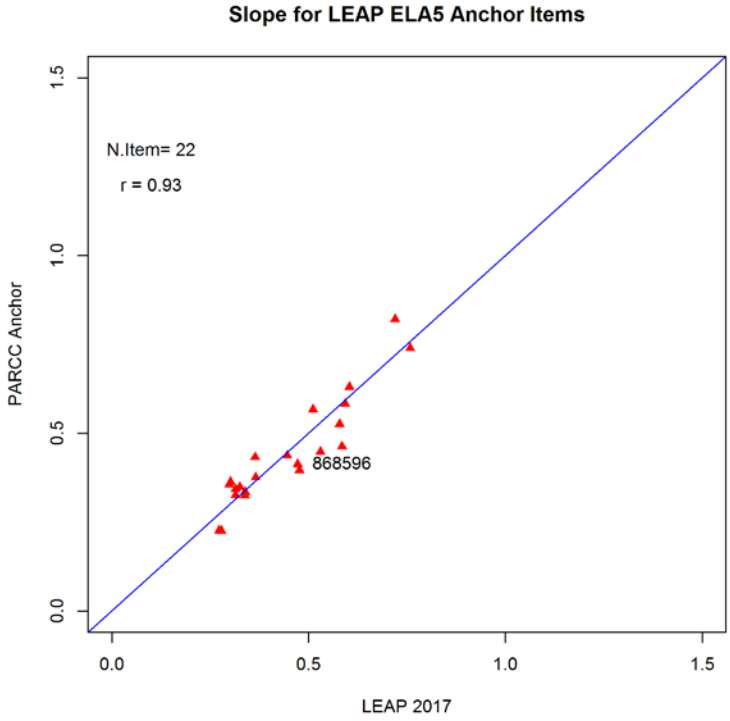
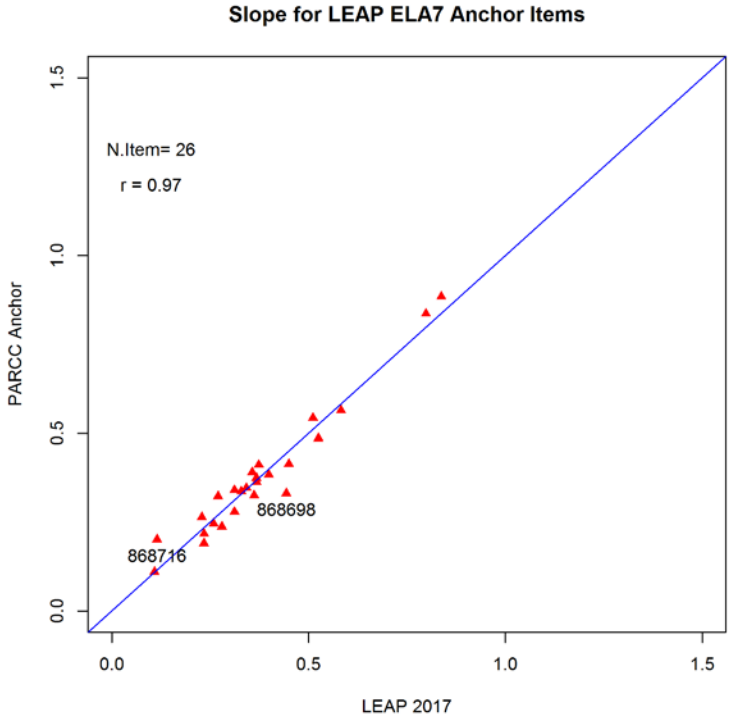



**Figure 6.4 ELA Slope Parameters after Linking 2017 LEAP 2025 to PARCC Scale**





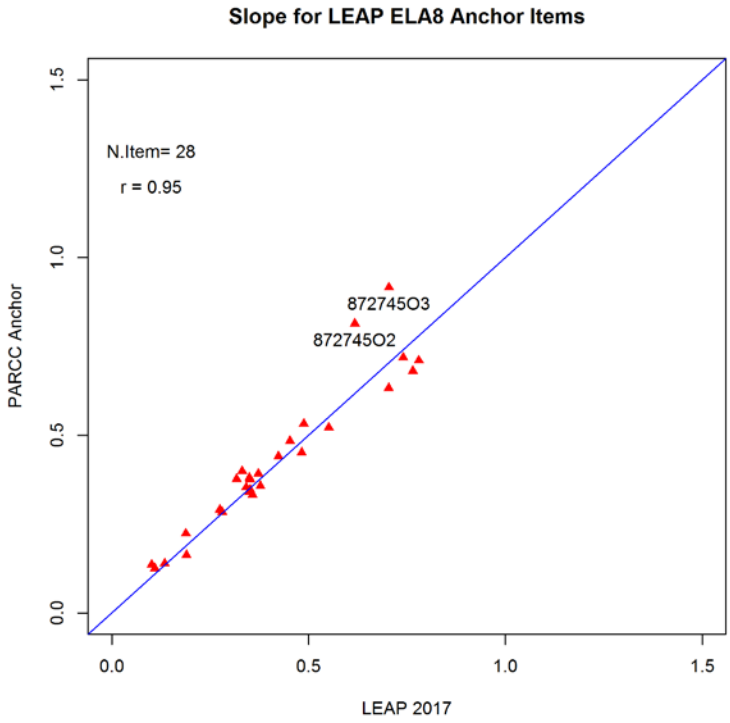
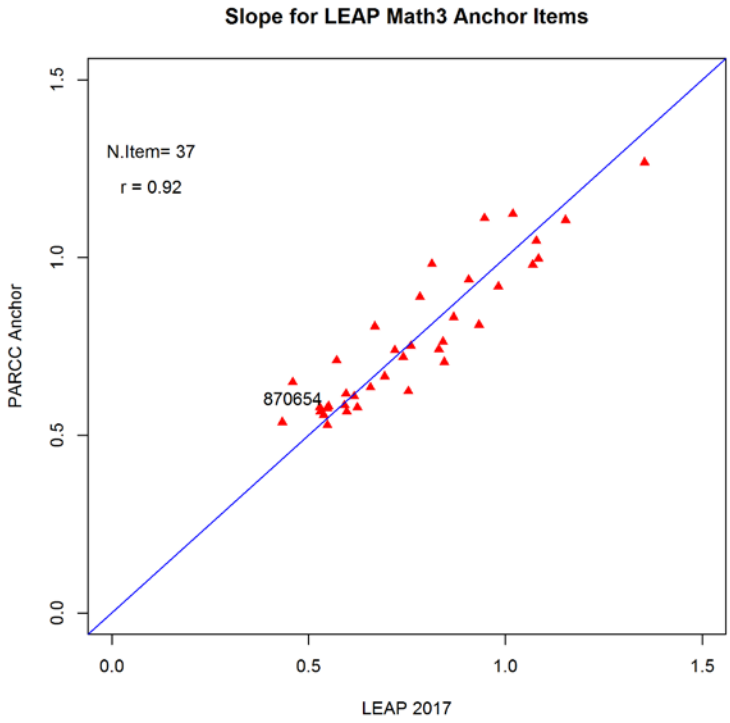
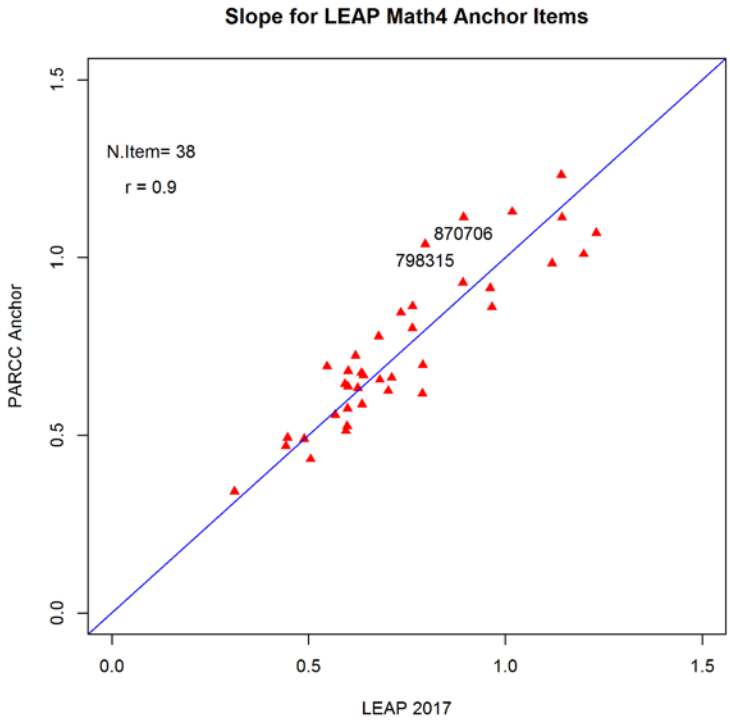
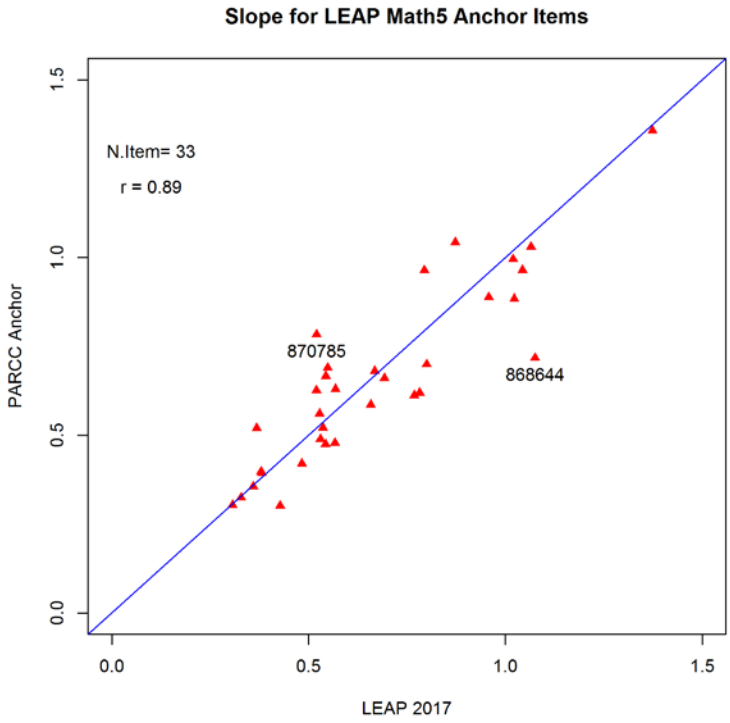
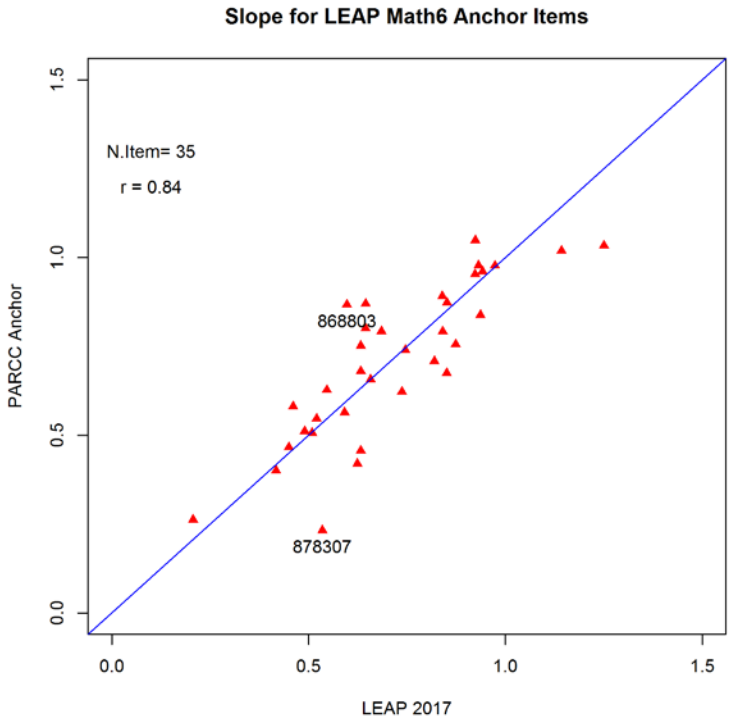


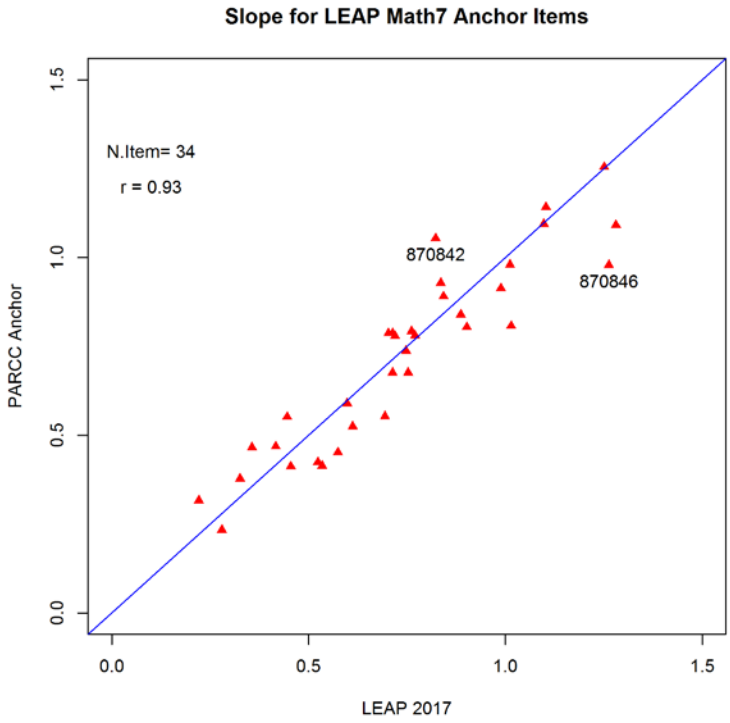
**Figure 6.5 Mathematics Slope Parameters after Linking 2017 LEAP 2025 to PARCC Scale**







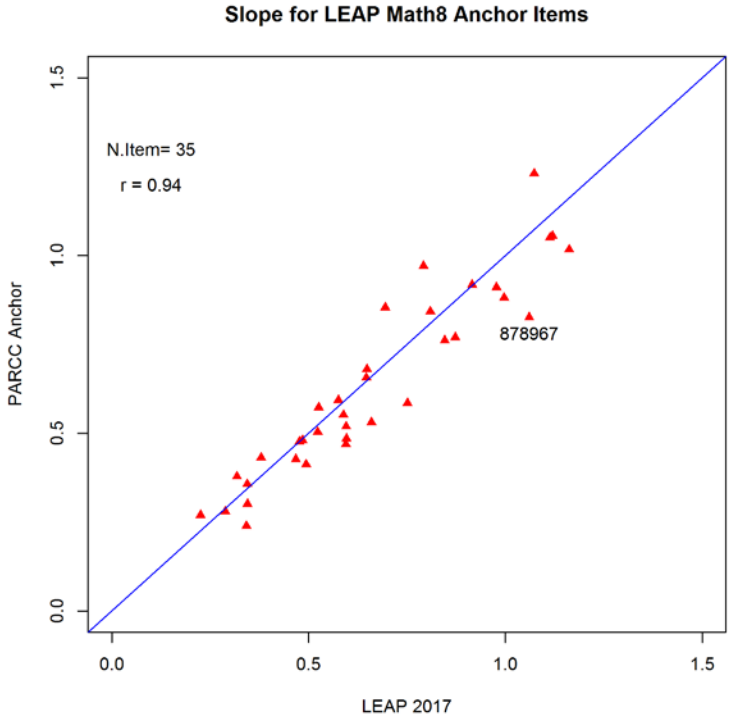
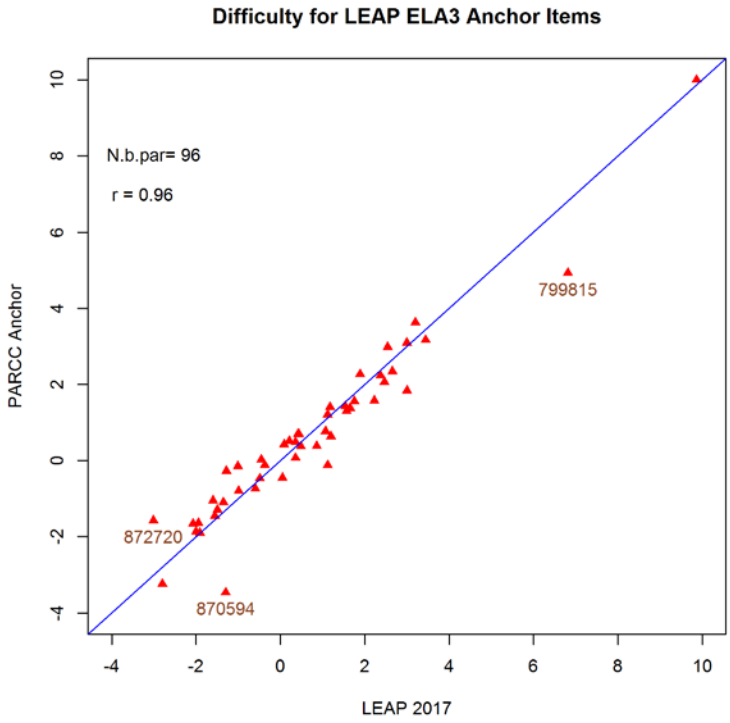
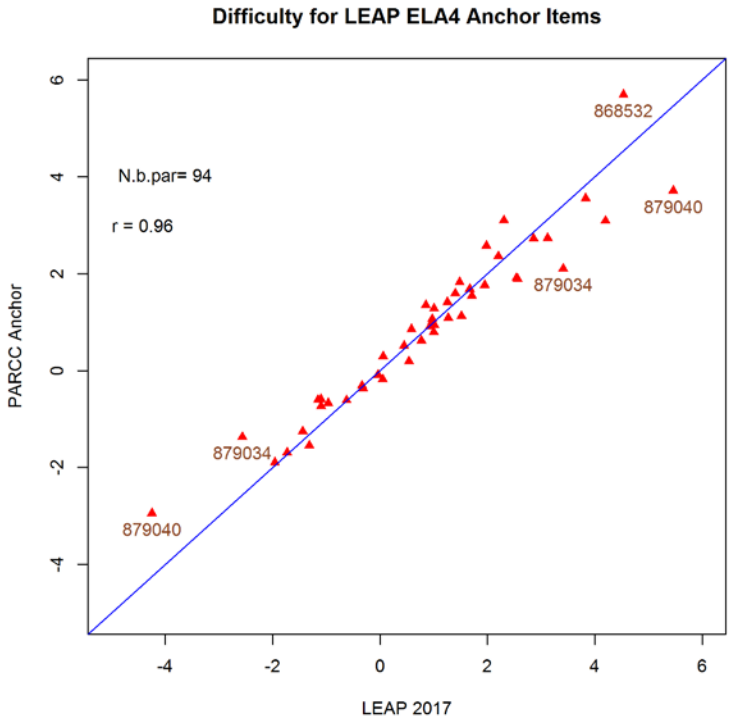
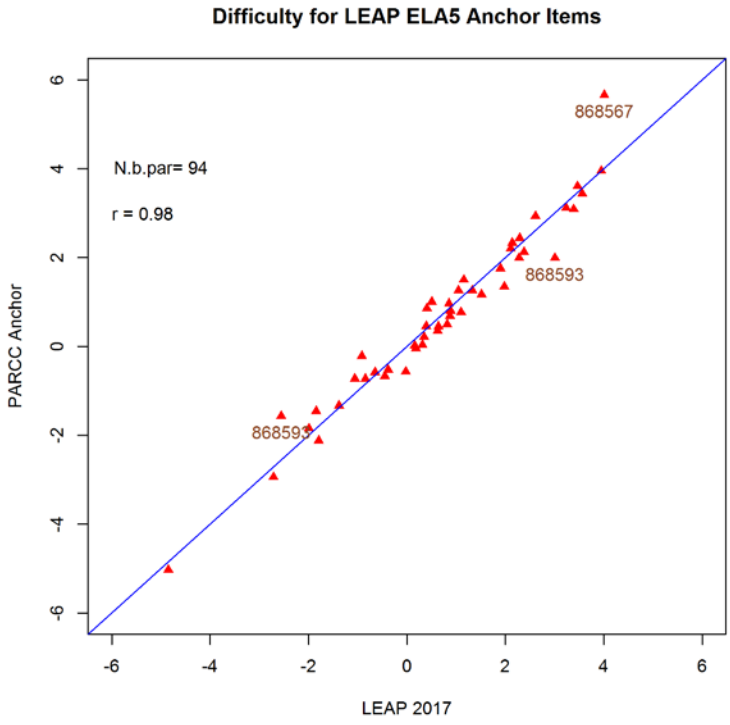
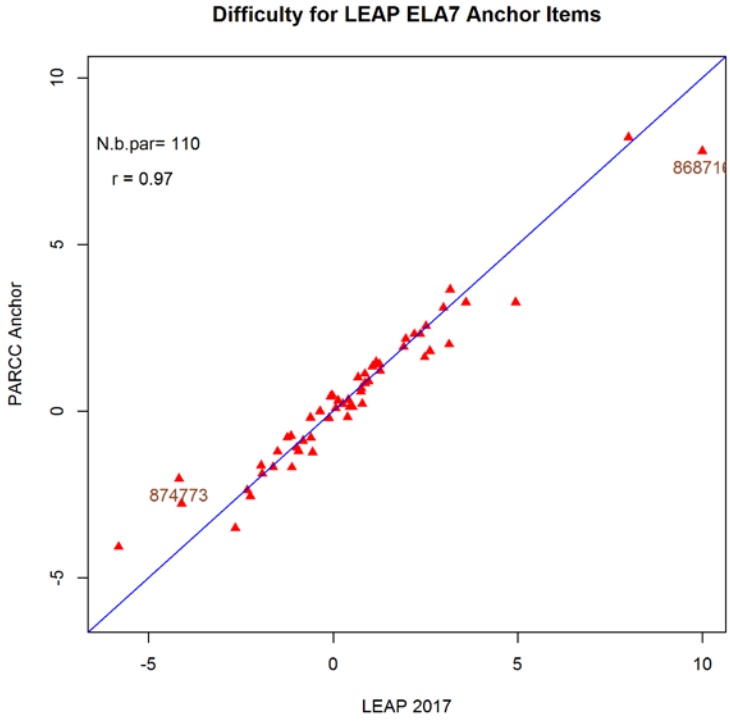



**Figure 6.6 ELA Difficulty Parameters after Linking 2017 LEAP 2025 to PARCC Scale**





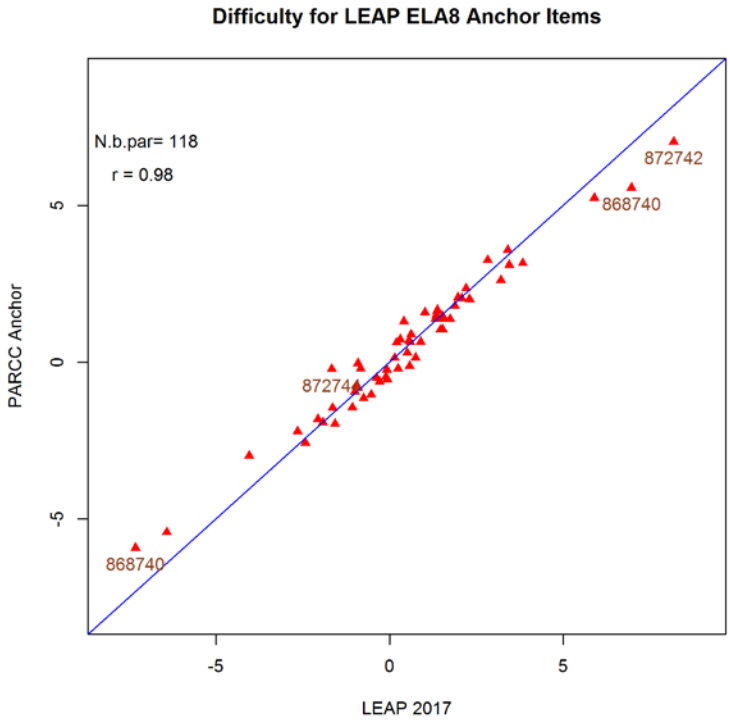
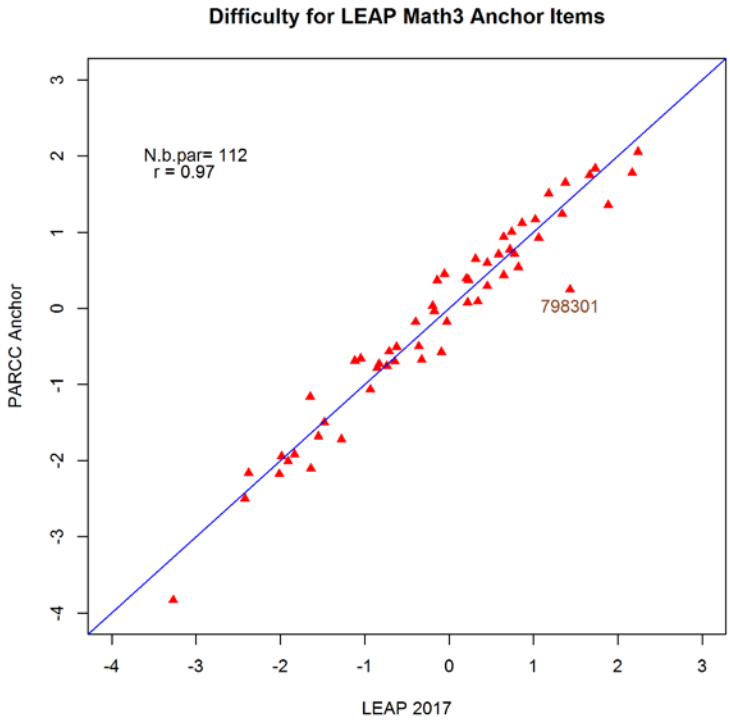
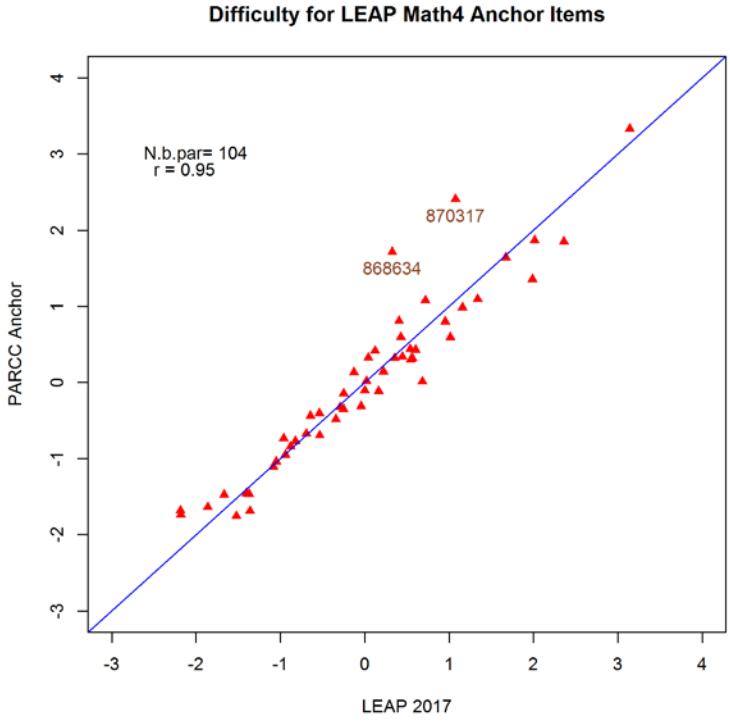
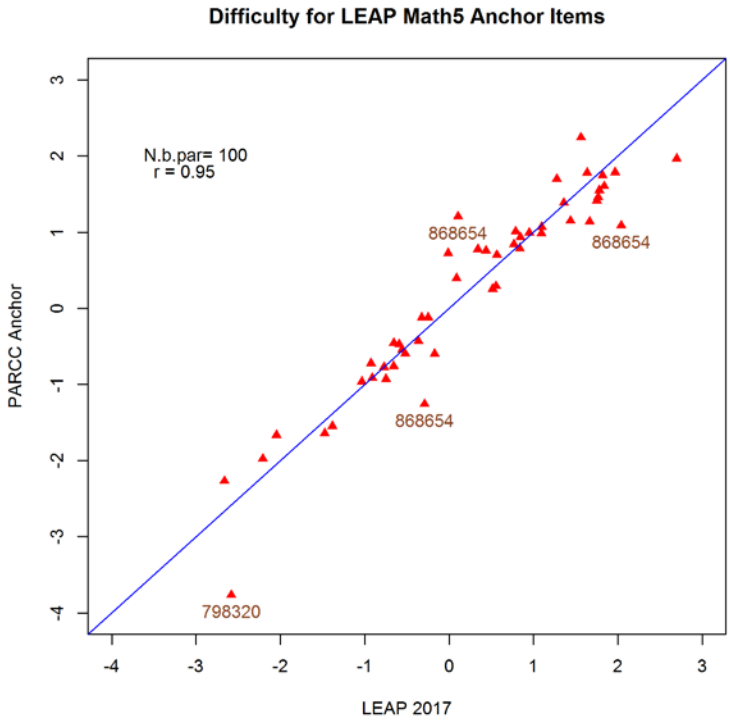
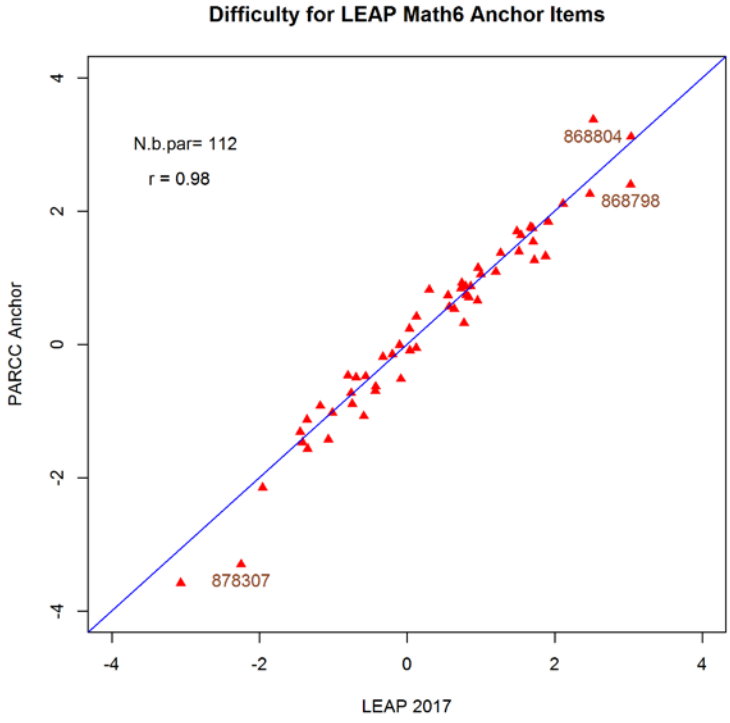


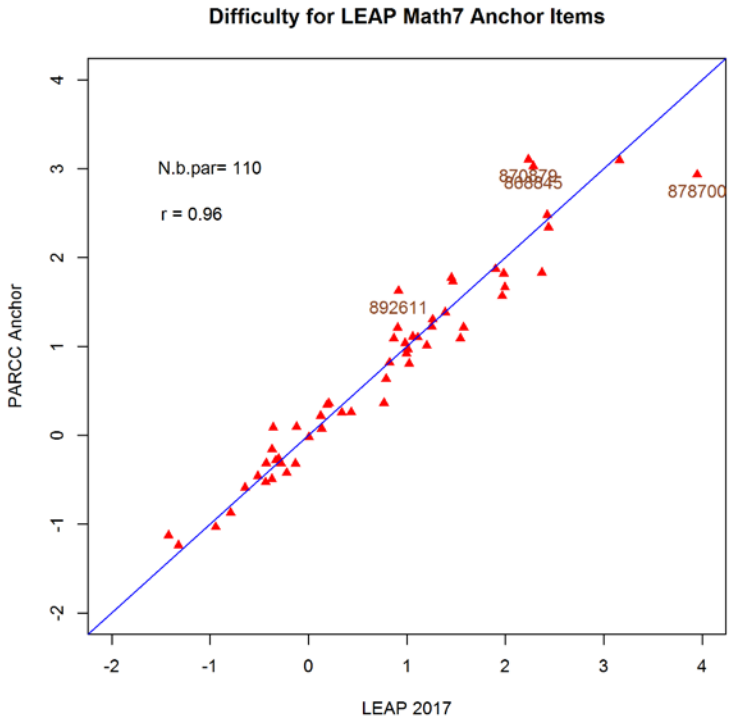
**Figure 6.7 Mathematics Difficulty Parameters after Linking 2017 LEAP 2025 to PARCC Scale**







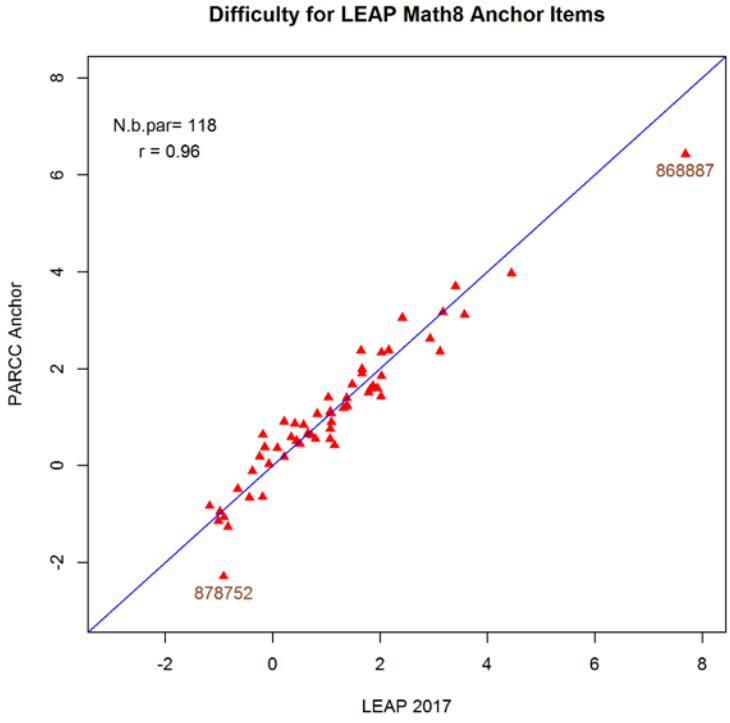



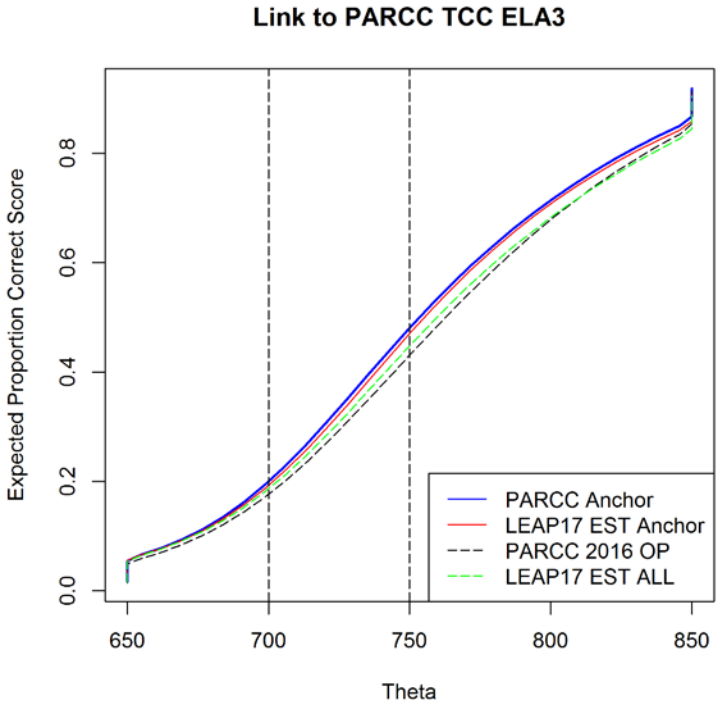
Table 6.24 presents the number of short anchor items by mode. Short anchor items were selected by maximizing the same mode item parameters as the test mode in addition to meeting test blueprint. Therefore, the number of anchor items vary across content areas and grades.

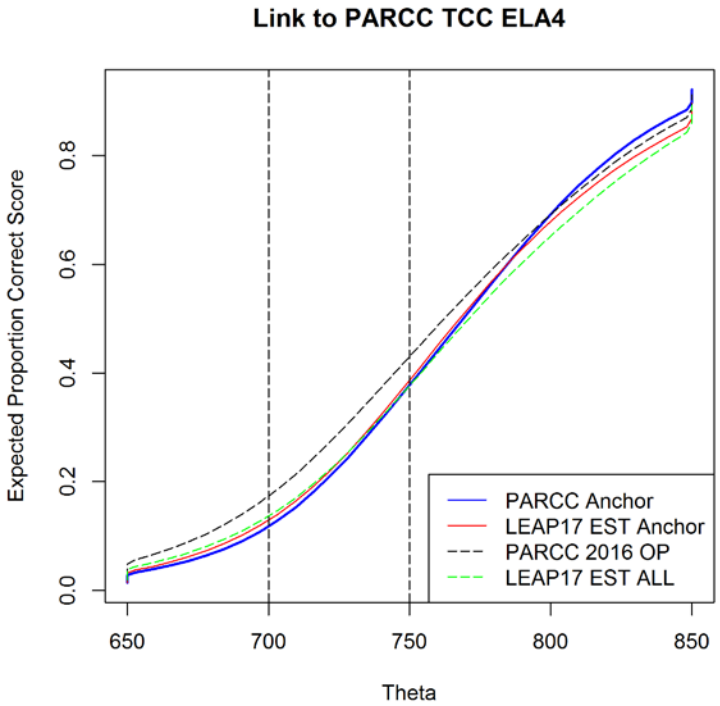
**Table 6.24 Number of Short Anchor Items by Mode**

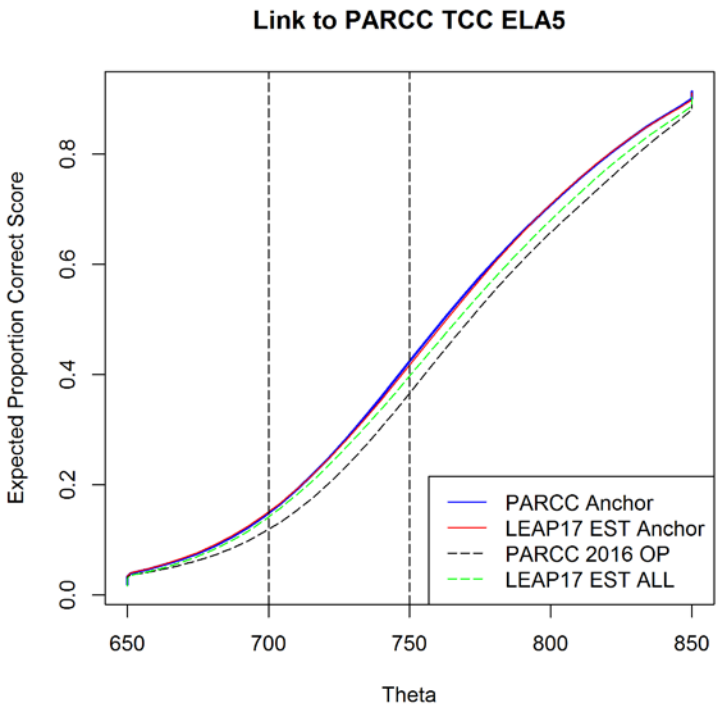
| | ELA | | | Mathematics | | |
|---|---|---|---|---|---|---|
| **Grade** | **Total** | **CBT** | **PBT** | **Total** | **CBT** | **PBT** |
| 3 | 15 | 2 | 13 | 15 | 1 | 14 |
| 4 | 17 | 9 | 8 | 14 | | 14 |
| 5 | 17 | 16 | 1 | 24 | 24 | |
| 6 | N/A | N/A | N/A | 22 | 21 | 1 |
| 7 | 17 | 17 | | 26 | 25 | 1 |
| 8 | 17 | 17 | | 24 | 23 | 1 |

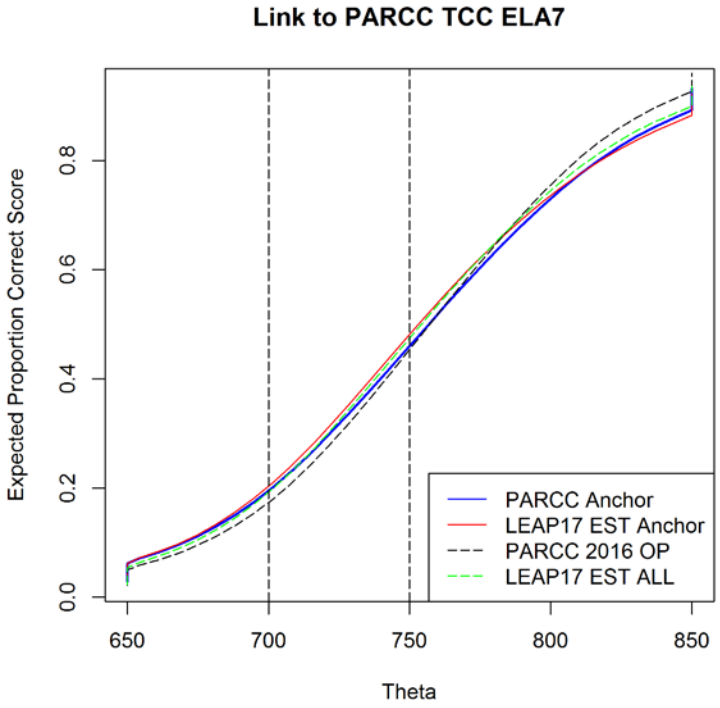
After equating with long and short anchors was performed, TCC results were compared. Figures 6.7 and 6.8 present TCCs for ELA and mathematics. Due to the calibration approach for ELA grade 6, TCC plots are unavailable. The TCCs show that the equating results with long and short anchors are very similar across most ability ranges, content areas, and grades. Equated item parameters using the long anchor sets were applied to generate scoring tables.

**Figure 6.8 TCC for 2017 LEAP 2025 ELA Grades 3 through 8 with Long and Short Anchors**







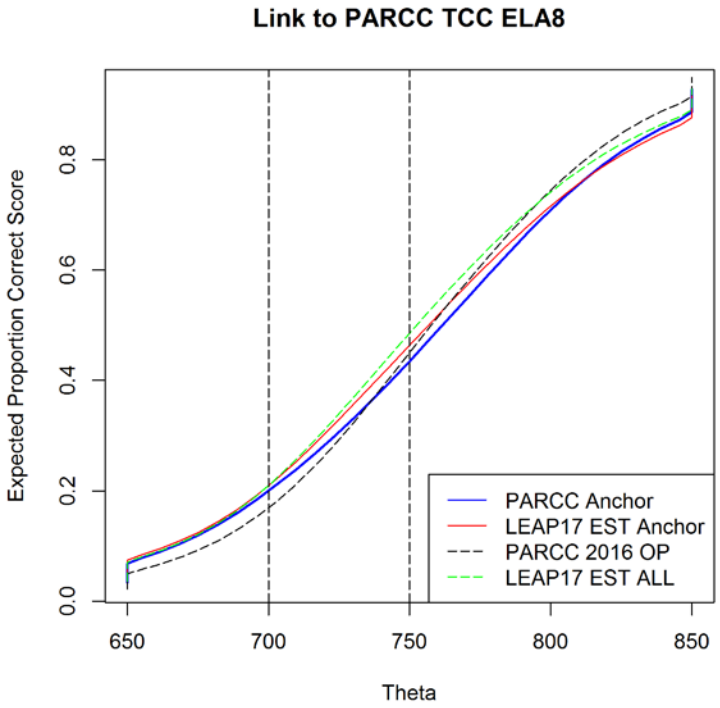
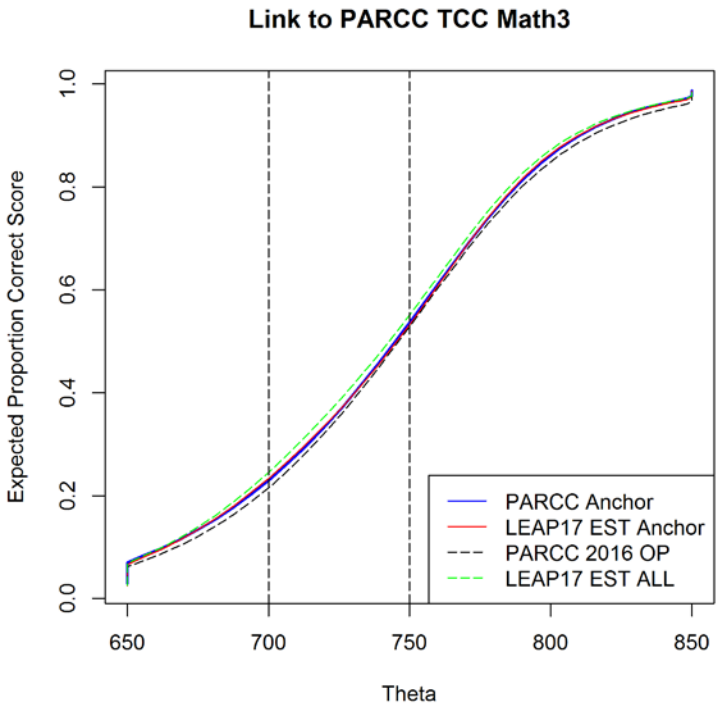
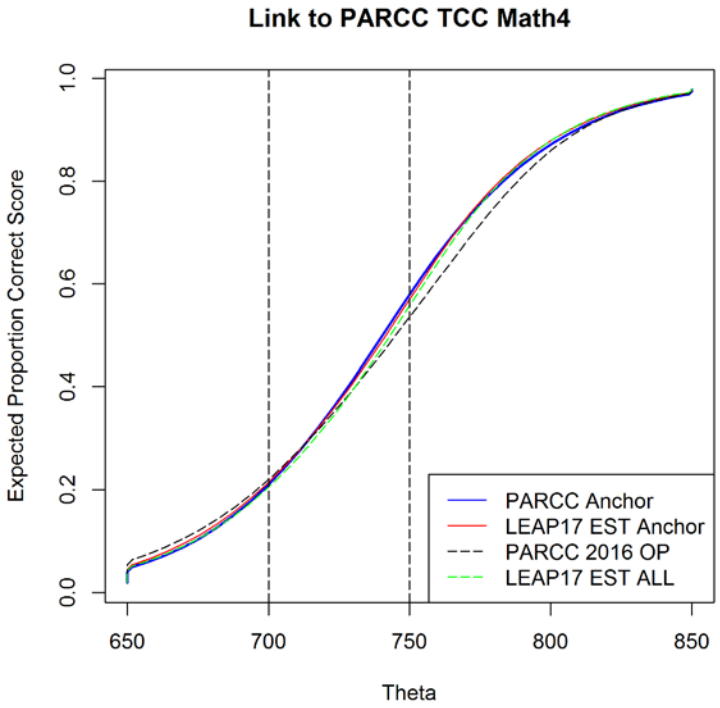
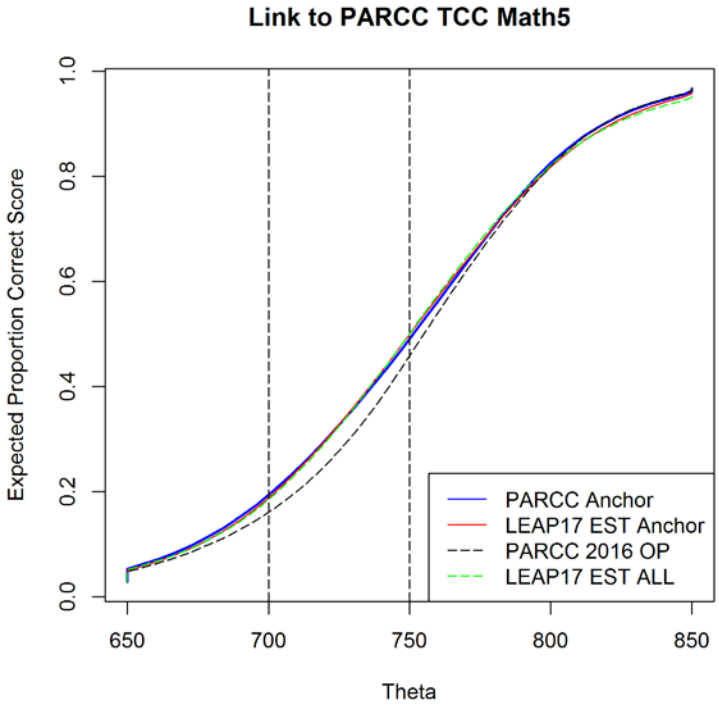
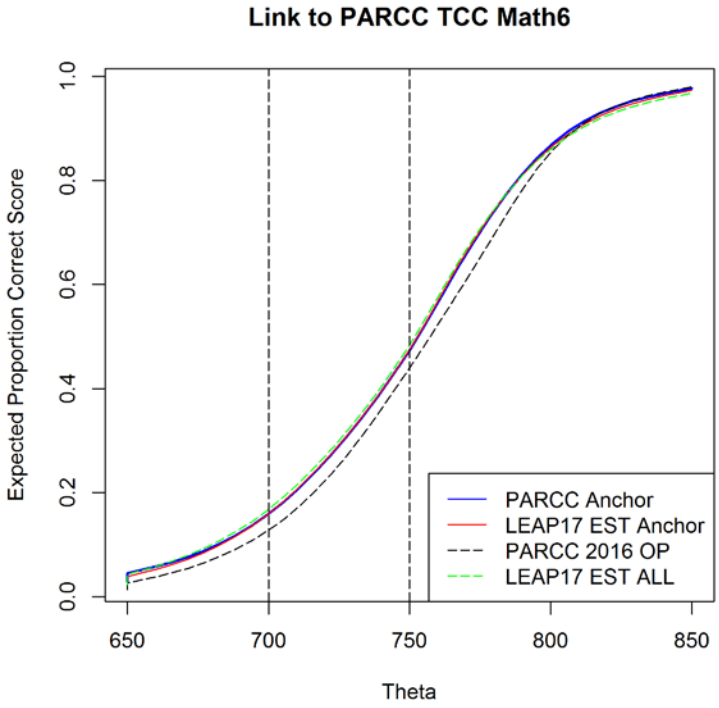



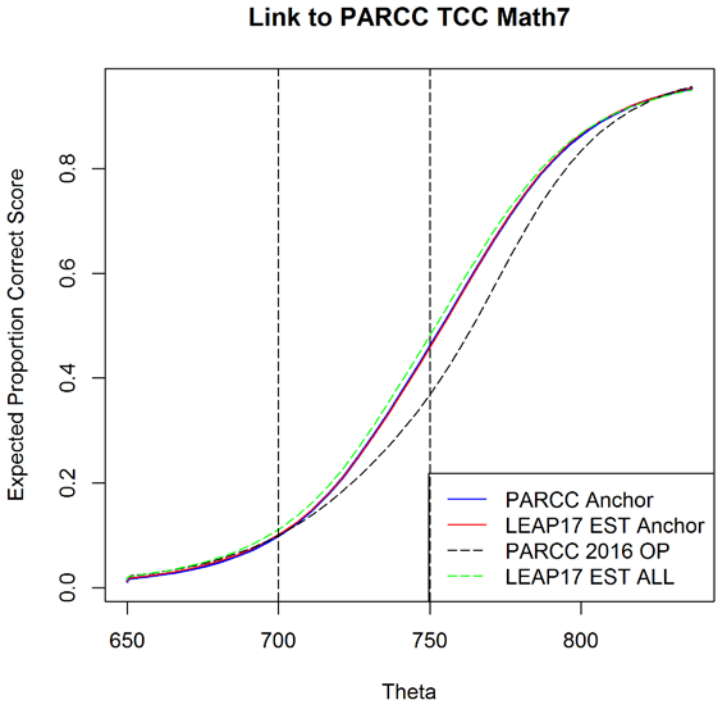
**Figure 6.9 TCC for 2017 LEAP 2025 Mathematics Grades 3 through 8 with Long and Short Anchors**

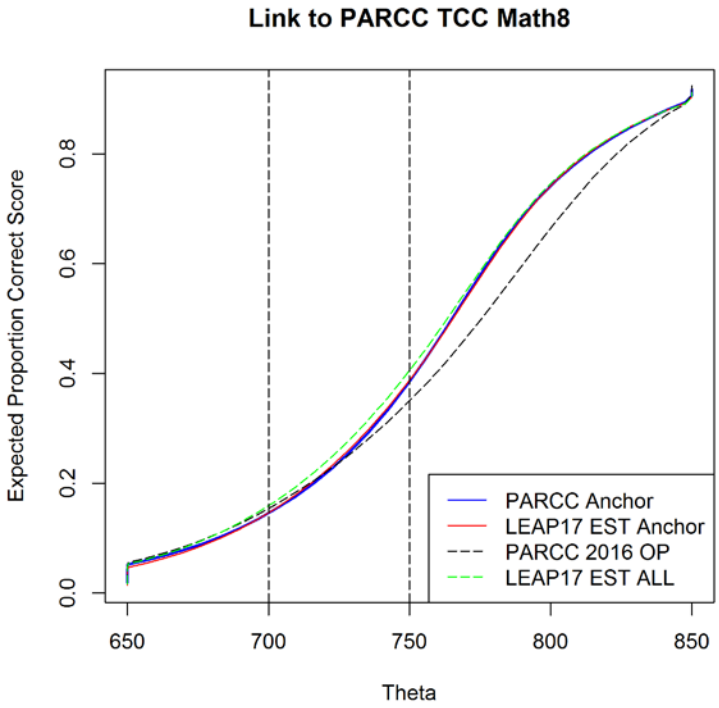






#### 6.4.2.1. Evaluation of Anchor Item Stability

Standard 5.15 requires information about the anchors, 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 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 invalidate 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 weighted root mean square difference (WRMSD) method in addition to content perspective, such as the number of items and score points for each claim and subclaim. If approved by LDOE, the outliers were dropped from anchor sets and considered to be non-common anchor items during equating. All intact 2016 PARCC items were initially used as anchor items to link the 2017 LEAP 2025 to PARCC 2016, and the quality of the anchor items was checked. The following evaluation rules were applied 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 STUIRT and flag items for further inspection if the WRMSD was greater than the values in Table 6.25. If the items were flagged, then they were removed from the anchor set and the ICC was reviewed with the WRMSD (Kim & Kolen, 2004).
- 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 claim and subclaim before and after dropping an anchor item.

The WRMSD values were calculated to compare to the ICCs using intact and estimated item parameters. PARCC 2016 item parameters were intact when linking to PARCC 2016, and their corresponding 2017 LEAP 2025 calibrated item parameters were estimated 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 PARSCALE PH2 output, *INTACT* represents intact PARCC item parameters, and *EST*

represents estimated item parameters corresponding to intact PARCC 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 |

Content representation was considered when items with large WRMSD values (and possible exclusion) from the linking sets were inspected to avoid removing large numbers of items from the same subclaim. After calculating WRMSD and excluding items with large WRMSD values from the first linking sets, STUIRT was rerun to produce new WRMSD values and the anchor items were reviewed again. This process was repeated until no items were flagged.

#### 6.4.2.2. Lowest and Highest Obtainable Scale Scores

An MML procedure cannot produce scale score estimates for students with perfect scores or scores below the level expected by 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 2017 LEAP 2025 used the same LOSS and HOSS values. The LOSS value was 650, and the HOSS value was 850.

#### 6.4.2.3. Claim-Level and Subclaim-Level Subscores

A student's performance on the ELA claims (i.e., reading and writing) is reported in one of three ratings: *Weak*, *Moderate*, or *Strong*.

Additionally, subclaim subscores are reported at the student level for ELA and mathematics. ELA has three subclaims for reading (i.e., literary text, informational text, and vocabulary) and two subclaims for writing (i.e., written expression and knowledge and use of language conventions). Mathematics has four subclaims. Subclaim performance is reported in one of three ratings of achievement: *Weak*, *Moderate*, or *Strong*.

Although the performance ratings are determined only by the items included within a claim or subclaim, the level of knowledge and ability needed to achieve a performance rating is connected to the level of knowledge and ability required by the subject-level achievement tests: a *Weak* rating requires similar knowledge and ability as the *Unsatisfactory* and *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.

For all scoring tables, the TCC inverse method was applied to generate raw-to-scale-score tables following the PARCC approach.

The 2017 LEAP 2025 reporting categories included the following:

***English Language Arts***

Claim

1. Reading
2. Writing

Subclaim

1. Reading Informational Texts—RI
2. Reading Literature—RL
3. Reading Vocabulary—RV
4. Written Expression—WE
5. Written Knowledge of Language—WKL

***Mathematics***

Subclaim

1. Mathematics Subclaim A – Major Content
2. Mathematics Subclaim B – Additional and Supporting Content
3. Mathematics Subclaim C – Expressing Mathematical Reasoning
4. Mathematics Subclaim D – Modeling and Application

Reading and writing claim scores were produced for ELA assessments only. The reading claim score range is 10–90 and the writing claim score range is 10–60. The method for scaling claims followed the PARCC methodology (Pearson, 2017). For the reading claim, 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 claim, a scoring table was generated using the TCC inverse method. A similar approach was applied to scale the writing claim, 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 claim (see Table 6.26 for cut scores for summatives, claims, and subclaims).

For subclaims, only performance-level ratings were reported. Therefore, there is no need to scale subclaims. Using the item parameters belonging to a given subclaim, a raw-score-to-theta scoring table is generated by applying the TCC inverse method. The two raw scores corresponding to $\Theta_{L3}$ and $\Theta_{L4}$ are cut scores for the subclaim.

**Table 6.26 ELA Cut Scores for Summative, Claim, and Subclaim**

| Performance Level | Summative Test | Claim | | Subclaim |
|---|---|---|---|---|
| | | Reading | Writing | |
| 1 | | | | |
| 2 | 700 | 30 | 25 | |
| 3 | 725 | 40 | 30 | $\Theta_{L3}$ |
| 4 | 750 | 50 | 35 | $\Theta_{L4}$ |
| 5 | Around 800 | | | |

**Subclaim thetas are those from summative tests (i.e., 725 & 750).*

***Yellow highlight shows cut scores for claim and subclaim.*

## 6.5 Comparability: Form Equating

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. 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.

In the spring of 2017, the forms used 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). Table 6.27 through Table 6.38 provide scale scores at selected percentiles that can be used to compare the distributional characteristics of the Spring 2017 forms to previous administrations. Althoughs these 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 proficiency (i.e., achievement) of 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; and (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 error 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 “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 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 822 | 839 |
| 95 | 796 | 810 |
| 90 | 783 | 793 |
| 85 | 774 | 784 |
| 80 | 768 | 775 |
| 75 | 762 | 770 |
| 70 | 757 | 762 |
| 65 | 751 | 757 |
| 60 | 746 | 752 |
| 55 | 741 | 748 |
| 50 | 738 | 743 |
| 45 | 732 | 739 |
| 40 | 727 | 734 |
| 35 | 721 | 727 |
| 30 | 715 | 723 |
| 25 | 712 | 718 |
| 20 | 706 | 710 |
| 15 | 695 | 701 |
| 10 | 687 | 695 |
| 5 | 676 | 679 |
| 1 | 654 | 655 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 816 | 818 |
| 95 | 794 | 796 |
| 90 | 785 | 785 |
| 85 | 777 | 777 |
| 80 | 769 | 771 |
| 75 | 765 | 765 |
| 70 | 760 | 761 |
| 65 | 755 | 756 |
| 60 | 751 | 752 |
| 55 | 746 | 748 |
| 50 | 744 | 744 |
| 45 | 740 | 741 |
| 40 | 735 | 737 |
| 35 | 731 | 733 |
| 30 | 727 | 728 |
| 25 | 722 | 724 |
| 20 | 715 | 717 |
| 15 | 709 | 711 |
| 10 | 701 | 702 |
| 5 | 691 | 691 |
| 1 | 666 | 670 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 816 | 813 |
| 95 | 792 | 793 |
| 90 | 782 | 782 |
| 85 | 774 | 775 |
| 80 | 767 | 769 |
| 75 | 763 | 763 |
| 70 | 758 | 758 |
| 65 | 754 | 754 |
| 60 | 749 | 750 |
| 55 | 745 | 747 |
| 50 | 740 | 743 |
| 45 | 738 | 739 |
| 40 | 733 | 735 |
| 35 | 728 | 731 |
| 30 | 723 | 727 |
| 25 | 720 | 721 |
| 20 | 714 | 716 |
| 15 | 708 | 709 |
| 10 | 701 | 701 |
| 5 | 692 | 691 |
| 1 | 675 | 673 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 813 | 814 |
| 95 | 792 | 790 |
| 90 | 780 | 779 |
| 85 | 772 | 770 |
| 80 | 765 | 763 |
| 75 | 760 | 759 |
| 70 | 756 | 754 |
| 65 | 752 | 748 |
| 60 | 748 | 745 |
| 55 | 745 | 741 |
| 50 | 741 | 736 |
| 45 | 737 | 733 |
| 40 | 734 | 729 |
| 35 | 730 | 724 |
| 30 | 727 | 721 |
| 25 | 723 | 716 |
| 20 | 718 | 711 |
| 15 | 713 | 705 |
| 10 | 706 | 698 |
| 5 | 696 | 689 |
| 1 | 676 | 671 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 825 | 826 |
| 95 | 800 | 800 |
| 90 | 787 | 786 |
| 85 | 777 | 778 |
| 80 | 771 | 770 |
| 75 | 766 | 765 |
| 70 | 761 | 759 |
| 65 | 756 | 756 |
| 60 | 751 | 751 |
| 55 | 747 | 745 |
| 50 | 742 | 742 |
| 45 | 740 | 737 |
| 40 | 735 | 733 |
| 35 | 730 | 728 |
| 30 | 726 | 723 |
| 25 | 721 | 717 |
| 20 | 714 | 711 |
| 15 | 706 | 702 |
| 10 | 697 | 692 |
| 5 | 683 | 675 |
| 1 | 655 | 654 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 825 | 834 |
| 95 | 804 | 806 |
| 90 | 790 | 791 |
| 85 | 781 | 782 |
| 80 | 775 | 776 |
| 75 | 770 | 770 |
| 70 | 764 | 764 |
| 65 | 759 | 758 |
| 60 | 754 | 754 |
| 55 | 752 | 749 |
| 50 | 747 | 745 |
| 45 | 743 | 740 |
| 40 | 739 | 734 |
| 35 | 735 | 731 |
| 30 | 731 | 725 |
| 25 | 727 | 719 |
| 20 | 721 | 714 |
| 15 | 714 | 707 |
| 10 | 706 | 696 |
| 5 | 693 | 681 |
| 1 | 670 | 651 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 824 | 822 |
| 95 | 802 | 796 |
| 90 | 789 | 786 |
| 85 | 781 | 776 |
| 80 | 775 | 772 |
| 75 | 770 | 765 |
| 70 | 765 | 761 |
| 65 | 760 | 756 |
| 60 | 756 | 752 |
| 55 | 751 | 747 |
| 50 | 746 | 743 |
| 45 | 741 | 738 |
| 40 | 738 | 733 |
| 35 | 733 | 728 |
| 30 | 728 | 725 |
| 25 | 722 | 720 |
| 20 | 716 | 715 |
| 15 | 710 | 706 |
| 10 | 703 | 699 |
| 5 | 692 | 689 |
| 1 | 672 | 667 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 819 | 812 |
| 95 | 797 | 792 |
| 90 | 786 | 779 |
| 85 | 777 | 774 |
| 80 | 771 | 767 |
| 75 | 766 | 762 |
| 70 | 761 | 756 |
| 65 | 756 | 752 |
| 60 | 752 | 748 |
| 55 | 747 | 744 |
| 50 | 743 | 740 |
| 45 | 738 | 736 |
| 40 | 732 | 732 |
| 35 | 728 | 727 |
| 30 | 723 | 722 |
| 25 | 718 | 717 |
| 20 | 713 | 712 |
| 15 | 708 | 706 |
| 10 | 703 | 700 |
| 5 | 693 | 693 |
| 1 | 677 | 674 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 819 | 808 |
| 95 | 792 | 784 |
| 90 | 779 | 774 |
| 85 | 771 | 767 |
| 80 | 766 | 760 |
| 75 | 759 | 755 |
| 70 | 754 | 751 |
| 65 | 749 | 747 |
| 60 | 745 | 742 |
| 55 | 740 | 740 |
| 50 | 735 | 735 |
| 45 | 731 | 730 |
| 40 | 728 | 728 |
| 35 | 722 | 723 |
| 30 | 720 | 720 |
| 25 | 714 | 715 |
| 20 | 711 | 709 |
| 15 | 705 | 706 |
| 10 | 699 | 699 |
| 5 | 691 | 691 |
| 1 | 678 | 675 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 803 | 808 |
| 95 | 783 | 781 |
| 90 | 771 | 771 |
| 85 | 765 | 762 |
| 80 | 758 | 757 |
| 75 | 753 | 752 |
| 70 | 747 | 746 |
| 65 | 744 | 742 |
| 60 | 740 | 738 |
| 55 | 735 | 734 |
| 50 | 731 | 732 |
| 45 | 729 | 727 |
| 40 | 724 | 724 |
| 35 | 722 | 719 |
| 30 | 717 | 717 |
| 25 | 714 | 711 |
| 20 | 709 | 708 |
| 15 | 706 | 701 |
| 10 | 699 | 697 |
| 5 | 692 | 688 |
| 1 | 679 | 671 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 797 | 796 |
| 95 | 779 | 777 |
| 90 | 768 | 766 |
| 85 | 760 | 760 |
| 80 | 754 | 754 |
| 75 | 750 | 749 |
| 70 | 746 | 746 |
| 65 | 742 | 741 |
| 60 | 738 | 737 |
| 55 | 734 | 734 |
| 50 | 730 | 731 |
| 45 | 728 | 727 |
| 40 | 723 | 723 |
| 35 | 721 | 721 |
| 30 | 719 | 717 |
| 25 | 714 | 712 |
| 20 | 712 | 709 |
| 15 | 706 | 706 |
| 10 | 703 | 699 |
| 5 | 695 | 694 |
| 1 | 678 | 673 |

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

| | 2016 | 2017 |
|---|---|---|
| Percentile | Form A | Form B |
| 99 | 808 | 809 |
| 95 | 787 | 784 |
| 90 | 775 | 771 |
| 85 | 766 | 763 |
| 80 | 761 | 757 |
| 75 | 753 | 751 |
| 70 | 749 | 746 |
| 65 | 744 | 741 |
| 60 | 737 | 736 |
| 55 | 734 | 730 |
| 50 | 731 | 727 |
| 45 | 727 | 724 |
| 40 | 724 | 718 |
| 35 | 720 | 714 |
| 30 | 712 | 710 |
| 25 | 708 | 706 |
| 20 | 704 | 698 |
| 15 | 699 | 693 |
| 10 | 695 | 687 |
| 5 | 684 | 674 |
| 1 | 663 | 656 |

## 6.6 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. It also helps 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 2017 LEAP 2025 administration of ELA and mathematics. The scale score results are presented here. Achievement-level information is also provided. Presenting the results by achievement level translates the quantitative scale provided through scale scores into a qualitative description of student achievement: *Advanced*, *Mastery*, *Basic*, *Approaching Basic*, and *Unsatisfactory*.

While the scale score provides an essential quantitative reference to 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 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.

Results presented below 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 reliability and validity of the scores from the LEAP 2025 assessments and should not be used for state accountability purposes.

## 7.1 Student Participation

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
- Limited English Proficient (LEP) Status
- Migrant Status

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 shows 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** | ≥ 57,110 | 99.05% | ≥ 57,350 | 99.67% |
| | **Gender** | | | | |
| | Female | ≥ 27,980 | 99.15% | ≥ 28,100 | 99.68% |
| | Male | ≥ 29,060 | 99.07% | ≥ 29,180 | 99.70% |
| | **Ethnicity** | | | | |
| | Hispanic/Latino | ≥ 4,120 | 98.52% | ≥ 4,160 | 98.87% |
| | American Indian or Alaska Native | ≥ 340 | 98.83% | ≥ 340 | 99.71% |
| | Asian | ≥ 880 | 99.55% | ≥ 880 | 99.77% |
| | Black or African American | ≥ 25,400 | 99.02% | ≥ 25,510 | 99.77% |
| | Native Hawaiian or Other Pacific | ≥ 70 | 94.44% | ≥ 70 | 95.95% |
| | White | ≥ 24,550 | 99.35% | ≥ 24,620 | 99.76% |
| | Two or More Races | ≥ 1,630 | 99.76% | ≥ 1,640 | 99.94% |
| | **Education Classification** | | | | |
| | Regular | ≥ 50,730 | 99.05% | ≥ 50,930 | 99.68% |
| | Special | ≥ 6,380 | 98.98% | ≥ 6,420 | 99.60% |
| | **Economic Status** | | | | |
| | Economically Disadvantaged | ≥ 41,120 | 99.82% | ≥ 41,300 | 99.87% |
| | Not Economically Disadvantaged | ≥ 14,750 | 99.70% | ≥ 14,770 | 99.81% |
| | **LEP Status** | | | | |
| | Non-LEP | ≥ 54,570 | 99.09% | ≥ 54,780 | 99.73% |
| | LEP | ≥ 2,540 | 98.07% | ≥ 2,570 | 98.37% |
| | **Migrant Status** | | | | |
| | Nonmigrant | ≥ 56,970 | 99.05% | ≥ 57,210 | 99.67% |
| | Migrant | ≥ 130 | 97.81% | ≥ 130 | 98.55% |
| | **Section 504 Status** | | | | |
| | Non-Section 504 | ≥ 52,180 | 99.03% | ≥ 52,400 | 99.65% |
| | Section 504 | ≥ 4,930 | 99.25% | ≥ 4,950 | 99.88% |

| Participation Rates by Grade and Subgroup (continued) | | | | | |
|---|---|---|---|---|---|
| Grade | Group | Accountable in ELA | Percentage Reportable in ELA | Accountable in Mathematics | Percentage Reportable in Mathematics |
| 4 | **All Students** | ≥ 56,580 | 99.05% | ≥ 56,780 | 99.60% |
| | **Gender** | | | | |
| | Female | ≥ 27,680 | 99.23% | ≥ 27,770 | 99.67% |
| | Male | ≥ 28,830 | 99.02% | ≥ 28,950 | 99.58% |
| | **Ethnicity** | | | | |
| | Hispanic/Latino | ≥ 3,830 | 98.41% | ≥ 3,860 | 98.73% |
| | American Indian or Alaska Native | ≥ 360 | 98.92% | ≥ 370 | 99.73% |
| | Asian | ≥ 800 | 99.13% | ≥ 810 | 99.38% |
| | Black or African American | ≥ 24,940 | 98.90% | ≥ 25,050 | 99.62% |
| | Native Hawaiian or Other Pacific | ≥ 50 | 98.04% | ≥ 50 | 96.15% |
| | White | ≥ 24,960 | 99.46% | ≥ 25,010 | 99.76% |
| | Two or More Races | ≥ 1,520 | 99.74% | ≥ 1,530 | 100.00% |
| | **Education Classification** | | | | |
| | Regular | ≥ 50,360 | 99.08% | ≥ 50,530 | 99.63% |
| | Special | ≥ 6,210 | 98.86% | ≥ 6,250 | 99.33% |
| | **Economic Status** | | | | |
| | Economically Disadvantaged | ≥ 40,000 | 99.77% | ≥ 40,140 | 99.85% |
| | Not Economically Disadvantaged | ≥ 15,400 | 99.73% | ≥ 15,420 | 99.82% |
| | **LEP Status** | | | | |
| | Non-LEP | ≥ 54,470 | 99.11% | ≥ 54,650 | 99.66% |
| | LEP | ≥ 2,100 | 97.58% | ≥ 2,130 | 97.94% |
| | **Migrant Status** | | | | |
| | Nonmigrant | ≥ 56,460 | 99.05% | ≥ 56,670 | 99.60% |
| | Migrant | ≥ 110 | 100.00% | ≥ 110 | 100.00% |
| | **Section 504 Status** | | | | |
| | Non-Section 504 | ≥ 50,770 | 99.03% | ≥ 50,960 | 99.57% |
| | Section 504 | ≥ 5,800 | 99.28% | ≥ 5,820 | 99.83% |

| Participation Rates by Grade and Subgroup (continued) | | | | | |
|---|---|---|---|---|---|
| Grade | Group | Accountable in ELA | Percentage Reportable in ELA | Accountable in Mathematics | Percentage Reportable in Mathematics |
| 5 | **All Students** | ≥ 53,310 | 99.79% | ≥ 53,340 | 99.75% |
| | **Gender** | | | | |
| | Female | ≥ 25,910 | 99.80% | ≥ 25,930 | 99.76% |
| | Male | ≥ 27,390 | 99.79% | ≥ 27,410 | 99.74% |
| | **Ethnicity** | | | | |
| | Hispanic/Latino | ≥ 3,430 | 99.53% | ≥ 3,430 | 99.56% |
| | American Indian or Alaska Native | ≥ 330 | 100.00% | ≥ 330 | 100.00% |
| | Asian | ≥ 810 | 100.00% | ≥ 810 | 100.00% |
| | Black or African American | ≥ 23,760 | 99.77% | ≥ 23,780 | 99.71% |
| | Native Hawaiian or Other Pacific | ≥ 50 | 100.00% | ≥ 50 | 100.00% |
| | White | ≥ 23,590 | 99.83% | ≥ 23,610 | 99.81% |
| | Two or More Races | ≥ 1,320 | 99.92% | ≥ 1,320 | 99.85% |
| | **Education Classification** | | | | |
| | Regular | ≥ 47,420 | 99.81% | ≥ 47,450 | 99.78% |
| | Special | ≥ 5,880 | 99.69% | ≥ 5,890 | 99.54% |
| | **Economic Status** | | | | |
| | Economically Disadvantaged | ≥ 38,000 | 99.84% | ≥ 38,020 | 99.81% |
| | Not Economically Disadvantaged | ≥ 14,910 | 99.70% | ≥ 14,910 | 99.70% |
| | **LEP Status** | | | | |
| | Non-LEP | ≥ 51,780 | 99.80% | ≥ 51,820 | 99.75% |
| | LEP | ≥ 1,520 | 99.67% | ≥ 1,520 | 99.74% |
| | **Migrant Status** | | | | |
| | Nonmigrant | ≥ 53,210 | 99.79% | ≥ 53,240 | 99.75% |
| | Migrant | ≥ 90 | 100.00% | ≥ 90 | 100.00% |
| | **Section 504 Status** | | | | |
| | Non-Section 504 | ≥ 47,650 | 99.78% | ≥ 47,690 | 99.74% |
| | Section 504 | ≥ 5,650 | 99.93% | ≥ 5,650 | 99.89% |

| Participation Rates by Grade and Subgroup (continued) | | | | | |
|---|---|---|---|---|---|
| Grade | Group | Accountable in ELA | Percentage Reportable in ELA | Accountable in Mathematics | Percentage Reportable in Mathematics |
| 6 | **All Students** | ≥ 52,480 | 99.66% | ≥ 52,510 | 99.63% |
| | **Gender** | | | | |
| | Female | ≥ 25,430 | 99.71% | ≥ 25,450 | 99.68% |
| | Male | ≥ 27,050 | 99.60% | ≥ 27,060 | 99.59% |
| | **Ethnicity** | | | | |
| | Hispanic/Latino | ≥ 3,200 | 99.31% | ≥ 3,200 | 99.38% |
| | American Indian or Alaska Native | ≥ 370 | 99.46% | ≥ 370 | 99.46% |
| | Asian | ≥ 780 | 100.00% | ≥ 780 | 100.00% |
| | Black or African American | ≥ 23,280 | 99.60% | ≥ 23,300 | 99.55% |
| | Native Hawaiian or Other Pacific | ≥ 30 | 97.37% | ≥ 30 | 97.37% |
| | White | ≥ 23,660 | 99.74% | ≥ 23,670 | 99.74% |
| | Two or More Races | ≥ 1,130 | 99.73% | ≥ 1,130 | 99.73% |
| | **Education Classification** | | | | |
| | Regular | ≥ 46,900 | 99.69% | ≥ 46,920 | 99.68% |
| | Special | ≥ 5,570 | 99.39% | ≥ 5,580 | 99.28% |
| | **Economic Status** | | | | |
| | Economically Disadvantaged | ≥ 36,980 | 99.68% | ≥ 37,010 | 99.65% |
| | Not Economically Disadvantaged | ≥ 15,060 | 99.69% | ≥ 15,060 | 99.71% |
| | **LEP Status** | | | | |
| | Non-LEP | ≥ 51,330 | 99.66% | ≥ 51,360 | 99.64% |
| | LEP | ≥ 1,150 | 99.22% | ≥ 1,150 | 99.39% |
| | **Migrant Status** | | | | |
| | Nonmigrant | ≥ 52,400 | 99.65% | ≥ 52,430 | 99.63% |
| | Migrant | ≥ 80 | 100.00% | ≥ 80 | 100.00% |
| | **Section 504 Status** | | | | |
| | Non-Section 504 | ≥ 46,640 | 99.63% | ≥ 46,670 | 99.61% |
| | Section 504 | ≥ 5,840 | 99.88% | ≥ 5,840 | 99.86% |

| Participation Rates by Grade and Subgroup (continued) | | | | | |
|---|---|---|---|---|---|
| **Grade** | **Group** | **Accountable in ELA** | **Percentage Reportable in ELA** | **Accountable in Mathematics** | **Percentage Reportable in Mathematics** |
| 7 | **All Students** | ≥ 51,960 | 99.60% | ≥ 51,980 | 99.59% |
| | **Gender** | | | | |
| | Female | ≥ 25,440 | 99.69% | ≥ 25,440 | 99.69% |
| | Male | ≥ 26,510 | 99.50% | ≥ 26,530 | 99.48% |
| | **Ethnicity** | | | | |
| | Hispanic/Latino | ≥ 3,030 | 99.70% | ≥ 3,030 | 99.70% |
| | American Indian or Alaska Native | ≥ 400 | 99.26% | ≥ 400 | 99.26% |
| | Asian | ≥ 840 | 99.88% | ≥ 840 | 99.88% |
| | Black or African American | ≥ 23,140 | 99.55% | ≥ 23,150 | 99.53% |
| | Native Hawaiian or Other Pacific | ≥ 40 | 100.00% | ≥ 40 | 100.00% |
| | White | ≥ 23,520 | 99.63% | ≥ 23,520 | 99.62% |
| | Two or More Races | ≥ 970 | 99.59% | ≥ 970 | 99.59% |
| | **Education Classification** | | | | |
| | Regular | ≥ 46,730 | 99.65% | ≥ 46,750 | 99.64% |
| | Special | ≥ 5,220 | 99.10% | ≥ 5,230 | 99.12% |
| | **Economic Status** | | | | |
| | Economically Disadvantaged | ≥ 36,160 | 99.59% | ≥ 36,180 | 99.56% |
| | Not Economically Disadvantaged | ≥ 15,360 | 99.75% | ≥ 15,360 | 99.75% |
| | **LEP Status** | | | | |
| | Non-LEP | ≥ 50,830 | 99.59% | ≥ 50,850 | 99.58% |
| | LEP | ≥ 1,120 | 99.73% | ≥ 1,120 | 99.73% |
| | **Migrant Status** | | | | |
| | Nonmigrant | ≥ 51,870 | 99.60% | ≥ 51,900 | 99.59% |
| | Migrant | ≥ 80 | 100.00% | ≥ 80 | 100.00% |
| | **Section 504 Status** | | | | |
| | Non-Section 504 | ≥ 46,390 | 99.58% | ≥ 46,410 | 99.57% |
| | Section 504 | ≥ 5,560 | 99.73% | ≥ 5,570 | 99.70% |

| Participation Rates by Grade and Subgroup (continued) | | | | | |
|---|---|---|---|---|---|
| Grade | Group | Accountable in ELA | Percentage Reportable in ELA | Accountable in Mathematics | Percentage Reportable in Mathematics |
| 8 | **All Students** | ≥ 50,590 | 99.46% | ≥ 50,620 | 99.43% |
| | **Gender** | | | | |
| | Female | ≥ 24,670 | 99.55% | ≥ 24,680 | 99.53% |
| | Male | ≥ 25,910 | 99.38% | ≥ 25,940 | 99.34% |
| | **Ethnicity** | | | | |
| | Hispanic/Latino | ≥ 2,810 | 99.36% | ≥ 2,810 | 99.33% |
| | American Indian or Alaska Native | ≥ 360 | 99.46% | ≥ 360 | 99.46% |
| | Asian | ≥ 810 | 100.00% | ≥ 810 | 100.00% |
| | Black or African American | ≥ 22,410 | 99.35% | ≥ 22,430 | 99.28% |
| | Native Hawaiian or Other Pacific | ≥ 40 | 100.00% | ≥ 40 | 100.00% |
| | White | ≥ 23,270 | 99.57% | ≥ 23,270 | 99.58% |
| | Two or More Races | ≥ 870 | 99.20% | ≥ 870 | 99.20% |
| | **Education Classification** | | | | |
| | Regular | ≥ 45,930 | 99.52% | ≥ 45,960 | 99.48% |
| | Special | ≥ 4,650 | 98.93% | ≥ 4,660 | 98.97% |
| | **Economic Status** | | | | |
| | Economically Disadvantaged | ≥ 34,780 | 99.40% | ≥ 34,810 | 99.36% |
| | Not Economically Disadvantaged | ≥ 15,370 | 99.75% | ≥ 15,370 | 99.75% |
| | **LEP Status** | | | | |
| | Non-LEP | ≥ 49,580 | 99.47% | ≥ 49,610 | 99.44% |
| | LEP | ≥ 1,010 | 99.01% | ≥ 1,010 | 99.01% |
| | **Migrant Status** | | | | |
| | Nonmigrant | ≥ 50,520 | 99.46% | ≥ 50,550 | 99.43% |
| | Migrant | ≥ 70 | 100.00% | ≥ 70 | 100.00% |
| | **Section 504 Status** | | | | |
| | Non-Section 504 | ≥ 45,500 | 99.44% | ≥ 45,530 | 99.41% |
| | Section 504 | ≥ 5,090 | 99.67% | ≥ 5,090 | 99.61% |

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

## 7.2 Current Administration Data

The LEAP 2025 ELA and Mathematics assessments were administered to students in grades 3–8. Tables 7.2 and 7.3 provide a summary of the scale scores based on the state population for the 2017 administration of the ELA and mathematics assessments, respectively.

**Table 7.2 State-Level Scale Score Statistics: English Language Arts**

| Grade | N | Mean SS | SD SS | Percentile | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 10th | 25th | 50th | 75th | 90th |
| **3** | ≥ 56,800 | 743.41 | 38.70 | 695 | 718 | 743 | 770 | 793 |
| **4** | ≥ 56,230 | 744.18 | 31.94 | 702 | 724 | 744 | 765 | 785 |
| **5** | ≥ 53,300 | 742.37 | 30.77 | 701 | 721 | 743 | 763 | 782 |
| **6** | ≥ 52,370 | 737.88 | 31.02 | 698 | 716 | 736 | 759 | 779 |
| **7** | ≥ 51,930 | 740.64 | 36.84 | 692 | 717 | 742 | 765 | 786 |
| **8** | ≥ 50,450 | 744.26 | 37.52 | 696 | 719 | 745 | 770 | 791 |

**Table 7.3 State-Level Scale Score Statistics: Mathematics**

| Grade | N | Mean SS | SD SS | Percentile | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 10th | 25th | 50th | 75th | 90th |
| **3** | ≥ 56,800 | 742.52 | 33.28 | 699 | 720 | 743 | 765 | 786 |
| **4** | ≥ 56,230 | 740.18 | 30.67 | 700 | 717 | 740 | 762 | 779 |
| **5** | ≥ 53,310 | 736.04 | 29.17 | 699 | 715 | 735 | 755 | 774 |
| **6** | ≥ 52,350 | 731.91 | 29.28 | 697 | 711 | 732 | 752 | 771 |
| **7** | ≥ 51,800 | 731.92 | 26.03 | 699 | 712 | 731 | 749 | 766 |
| **8*** | ≥ 44,710 | 728.40 | 33.09 | 687 | 706 | 727 | 751 | 771 |

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

Tables 7.4 and 7.5 show the percentage of students in each achievement level based on the state population for the 2017 administration of the ELA and mathematics assessments.

**Table 7.4 Comparison of Percentage of Students in Each Achievement Level, ELA 2017 Census Data**

| Content | Grade | N | Unsatisfactory | Approaching Basic | Basic | Mastery | Advanced |
|---|---|---|---|---|---|---|---|
| ELA | 3 | ≥ 56,800 | 13.4 | 17.8 | 24.7 | 38.9 | 5.1 |
| ELA | 4 | ≥ 56,230 | 8.8 | 18.3 | 29.3 | 36.2 | 7.3 |
| ELA | 5 | ≥ 53,300 | 8.7 | 18.8 | 31.1 | 37.9 | 3.4 |
| ELA | 6 | ≥ 52,370 | 10.4 | 24.9 | 29.8 | 29.4 | 5.5 |
| ELA | 7 | ≥ 51,930 | 13.2 | 19.2 | 26.5 | 30.3 | 10.8 |
| ELA | 8 | ≥ 50,450 | 11.4 | 17.4 | 27.0 | 35.1 | 9.0 |

**Table 7.5 Comparison of Percentage of Students in Each Achievement Level, Mathematics 2017 Census Data**

| Content | Grade | N | Unsatisfactory | Approaching Basic | Basic | Mastery | Advanced |
|---|---|---|---|---|---|---|---|
| Mathematics | 3 | ≥ 56,800 | 11.1 | 18.4 | 27.1 | 36.2 | 7.1 |
| Mathematics | 4 | ≥ 56,230 | 8.2 | 23.2 | 29.7 | 35.0 | 3.8 |
| Mathematics | 5 | ≥ 53,310 | 11.1 | 24.9 | 32.4 | 27.7 | 3.9 |
| Mathematics | 6 | ≥ 52,350 | 12.6 | 30.8 | 29.2 | 23.7 | 3.7 |
| Mathematics | 7 | ≥ 51,800 | 11.2 | 28.9 | 35.2 | 22.6 | 2.1 |
| Mathematics | 8 | ≥ 44,710 | 20.3 | 28.2 | 25.0 | 24.7 | 1.8 |

## 7.3 Reports

Score reports are the primary means of communicating test scores to relevant 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 related to the test scores are disseminated in two ways: (1) the individual score report and (2) the *LEAP Interpretive Guide* (2016).

In addition to providing interpretation, it is important that the information is understandable by 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)

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 relevant information so they can understand their child's test score. In the 2016–2017 administration year, student reports for each school were posted by grade, downloaded, and printed from eDIRECT by the school system and school.

### 7.3.1 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 individual student achievement and 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 eDIRECT. Total test scores and achievement-level indicators are shown for the content area of interest. Claim and subclaim performance ratings are also reported for students. At the school level, the percentage of students at each achievement level and rating by claim and subclaim are summarized. More details can be found in the *LEAP Interpretive Guide* on the LDOE website (see https://www.louisianabelieves.com/docs/default-source/assessment/leap-2025-grades-3-8-ela-and-math-interpretive-guide.pdf?sfvrsn=3).

***Individual Student-Level Report***

The ISR is another type of report available through the eDIRECT system. ISRs may be downloaded and printed by schools to be sent home to the 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, claim and subclaim 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 his or her peers, are found at the bottom of the page. When a student does not receive a scale score, his or her achievement level will be left blank. ISRs for students whose scores were invalidated display a blank scale score for a given content area.

## 7.4 Data Structures

A data file referred to as Louisiana Department of Education Student File (LDESTD) was provided to 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.

## 7.5 Interpreting Test Results

The *LEAP Interpretative Guide* (see https://www.louisianabelieves.com/docs/default-source/assessment/leap-2025-grades-3-8-ela-and-math-interpretive-guide.pdf?sfvrsn=3) was written for Louisiana school system and school administrators, teachers, parents, and the general public to better understand the LEAP 2025 ELA and Mathematics tests. The *LEAP Interpretative 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 Interpretative 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.

## 7.6 Summary

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 LEAP 2025 performance-level setting and presents the cut scores established and the achievement-level descriptors derived from the performance-level setting. Since the LDOE used 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, ELA and mathematics tests that measured different content and constructs than previous tests measured 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 the new scales, and the 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 to which the LEAP 2025 ELA and Mathematics assessments were aligned.

In terms of the validity of the LEAP 2025 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, educator judgment about what students should know and be able to demonstrate, and research 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 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 Claim Cut Scores

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

## 8.3 Achievement-Level Descriptors

The cut scores divide the continuum of student achievement into the following five achievement levels used by 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 studies 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 studies 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 studies 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 studies 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 studies 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 | | | | | |

## 8.4 Summary

This chapter presented a brief overview of PARCC's performance-level setting process, which set the cut scores used by 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 CONSTRUCT-RELATED VALIDITY

Evidence for construct-related validity—the meaning of test scores and the inferences they support—is the central concept underlying the LEAP 2025 validation process. In this section, DRC presents evidence of construct-related validity through studies of test reliability, convergent validity, and divergent validity. All analyses in this section are based on census data.

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 *Standards* indicates:

> 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 subscore reliabilities and SEMs are presented in Section 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, $\sigma_i^2$ is the variance of item $i$, and $\sigma_x^2$ 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; 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 9.1. These reliability coefficients were computed using the census data. The reliability statistics ranged from 0.87 to 0.91 for all ELA forms. For mathematics, the reliabilities ranged from 0.89 to 0.92. These results indicate acceptable reliability coefficients for the LEAP 2025 tests.

**Table 9.1 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 | PBT | 30 | 72 | 4.83 | 0.89 | ≥ 56,800 |
| ELA | 4 | CBT | 32 | 82 | 5.19 | 0.88 | ≥ 1,930 |
| ELA | 4 | PBT | 32 | 82 | 5.13 | 0.87 | ≥ 54,300 |
| ELA | 5 | CBT | 32 | 82 | 5.12 | 0.88 | ≥ 53,300 |
| ELA | 6 | CBT | 36 | 92 | 5.37 | 0.91 | ≥ 52,370 |
| ELA | 7 | CBT | 34 | 86 | 6.06 | 0.89 | ≥ 51,930 |
| ELA | 8 | CBT | 34 | 86 | 6.37 | 0.87 | ≥ 50,450 |
| Mathematics | 3 | PBT | 43 | 62 | 3.73 | 0.91 | ≥ 56,800 |
| Mathematics | 4 | CBT | 42 | 59 | 3.47 | 0.92 | ≥ 1,930 |
| Mathematics | 4 | PBT | 42 | 59 | 3.67 | 0.91 | ≥ 54,300 |
| Mathematics | 5 | CBT | 42 | 61 | 3.58 | 0.90 | ≥ 53,310 |
| Mathematics | 6 | CBT | 41 | 64 | 3.92 | 0.90 | ≥ 52,350 |
| Mathematics | 7 | CBT | 43 | 66 | 3.88 | 0.91 | ≥ 51,800 |
| Mathematics | 8 | CBT | 42 | 66 | 3.74 | 0.89 | ≥ 44,710 |

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:

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

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 9.1 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 9.2.

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):

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

where $I(\theta_i)$ 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. The CSEMs at the four cut scores that define the performance levels are presented in Table 9.2.

**Table 9.2 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 | PBT | 700 | 12 | 725 | 10 | 750 | 10 | 810 | 12 |
| ELA | 4 | CBT | 700 | 11 | 725 | 9 | 750 | 8 | 790 | 9 |
| ELA | 4 | PBT | 700 | 11 | 725 | 9 | 750 | 8 | 790 | 9 |
| ELA | 5 | CBT | 700 | 10 | 725 | 8 | 750 | 8 | 799 | 8 |
| ELA | 6 | CBT | 700 | 9 | 725 | 7 | 750 | 7 | 790 | 8 |
| ELA | 7 | CBT | 700 | 9 | 725 | 8 | 750 | 8 | 785 | 8 |
| ELA | 8 | CBT | 700 | 9 | 725 | 8 | 750 | 8 | 794 | 10 |
| Mathematics | 3 | PBT | 700 | 9 | 725 | 8 | 750 | 8 | 790 | 10 |
| Mathematics | 4 | CBT | 700 | 9 | 725 | 8 | 750 | 7 | 796 | 10 |
| Mathematics | 4 | PBT | 700 | 9 | 725 | 8 | 750 | 7 | 796 | 10 |
| Mathematics | 5 | CBT | 700 | 9 | 725 | 8 | 750 | 7 | 790 | 9 |
| Mathematics | 6 | CBT | 700 | 9 | 725 | 8 | 750 | 7 | 788 | 8 |
| Mathematics | 7 | CBT | 700 | 10 | 725 | 6 | 750 | 6 | 786 | 7 |
| Mathematics | 8 | CBT | 700 | 12 | 725 | 10 | 750 | 8 | 801 | 10 |

Figures 9.1 and 9.2 display the CSEM curves for each grade and content area by mode. The estimates of measurement error tend to be higher at the low and high ends of the scale score range. The measurement error increases when there are few observations at a particular ability level. 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. 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 the majority of students are located.

**Figure 9.1 CSEM Curves for ELA Grades 3 through 8**

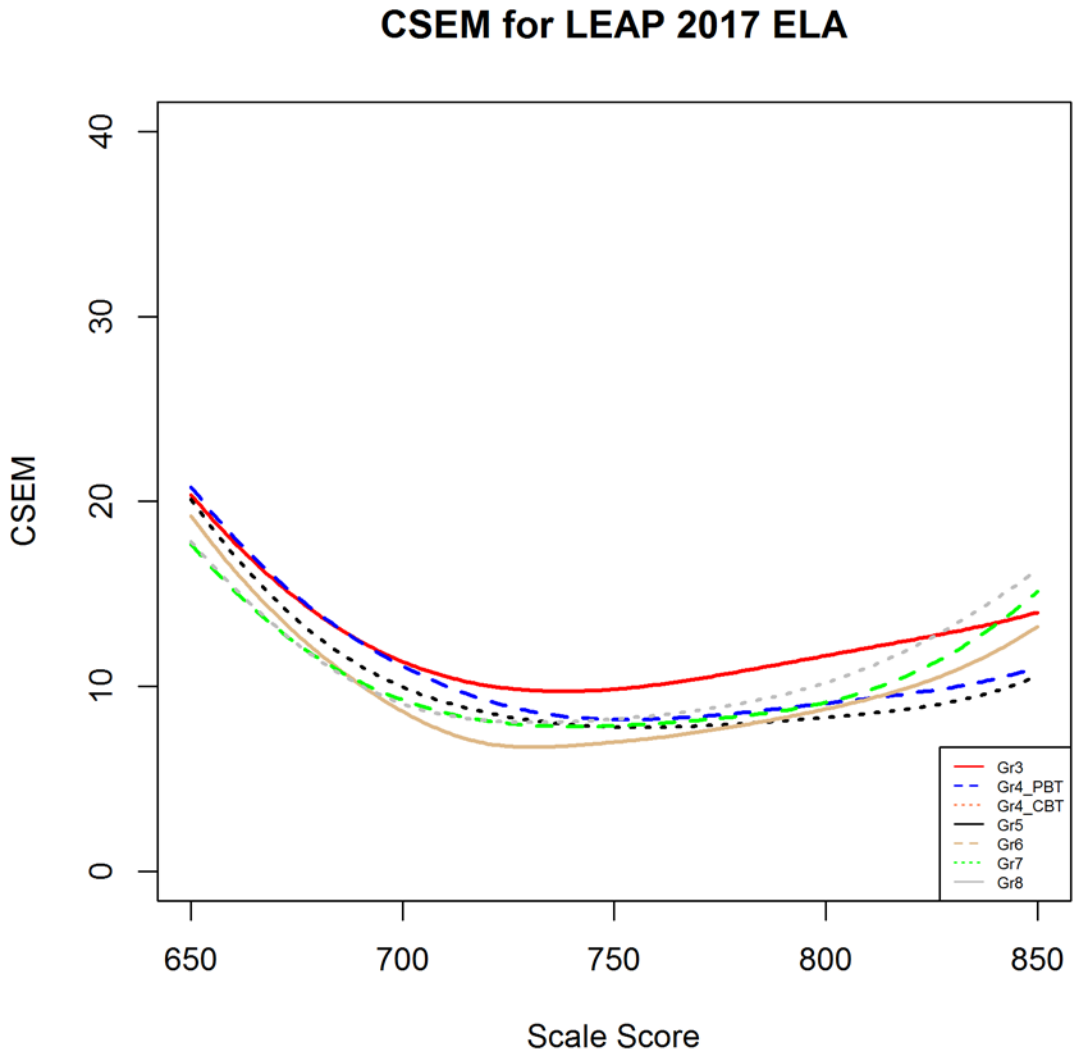
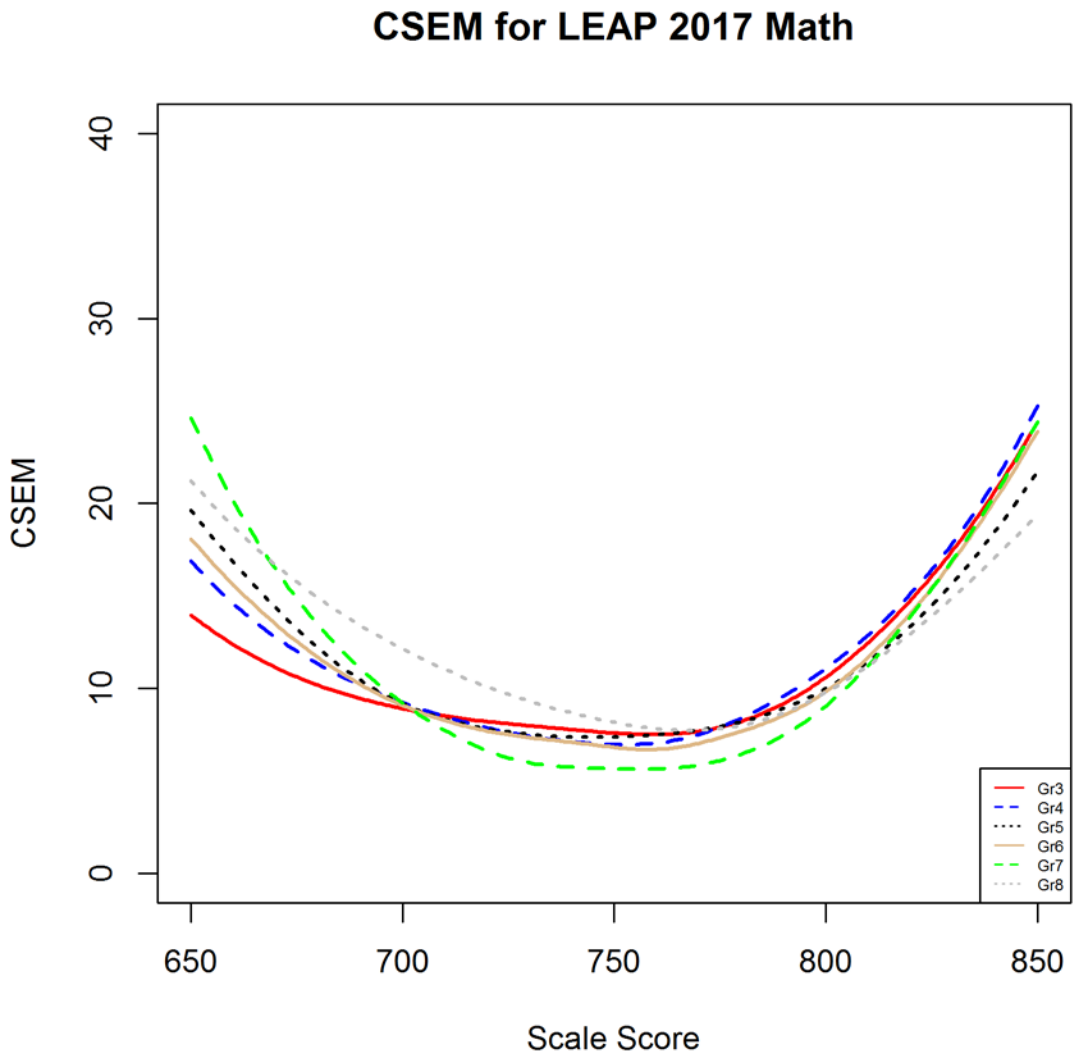


**Figure 9.2 CSEM Curves for Mathematics Grades 3 through 8**




### 9.2.4 Classification Accuracy and Consistency

***Decision Accuracy***

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

### *Decision Consistency*

Decision consistency, or classification consistency, is defined as the extent to which the classifications of students in a particular achievement level agree on the basis of two independent administrations of the test 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 decision consistency from one administration of a test.

The Livingston-Lewis (1995) methodology was used to calculate decision accuracy statistics based on the 2017 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 decision accuracy and consistency are based on census data.

Classification consistency and classification accuracy conditioned on performance level (see Table 9.3 and Table 9.4) and on cut score (see Table 9.5 and Table 9.6) are presented for the 2017 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 9.3, classification accuracy conditioned on achievement level ranged from 0.50 to 0.84 for ELA and 0.48 to 0.86 for mathematics. Classification consistency (see Table 9.4) conditioned on achievement level ranged from 0.48 to 0.76 for ELA and 0.48 to 0.80 for mathematics. Table 9.5 shows that classification accuracy at achievement cut points ranged from 0.89 to 0.98 for ELA and 0.88 to 0.99 for mathematics. Classification consistency (see Table 9.6) conditioned at achievement cut points ranged from 0.85 to 0.97 for ELA and 0.88 to 0.99 for mathematics. Classification consistency and accuracy at achievement cut points tend to be higher values than those conditioned on performance level. For some ELA 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 performance levels.

**Table 9.3 Decision Accuracy Conditioned on Level of Achievement**

| | Decision Accuracy | | | | | | |
|---|---|---|---|---|---|---|---|
| Content Area | Grade | Mode | Unsatisfactory | Approaching Basic | Basic | Mastery | Advanced |
| ELA | 3 | PBT | 0.64 | 0.66 | 0.83 | 0.51 | 0.64 |
| ELA | 4 | CBT | 0.64 | 0.68 | 0.80 | 0.65 | 0.64 |
| ELA | 4 | PBT | 0.62 | 0.71 | 0.80 | 0.63 | 0.62 |
| ELA | 5 | CBT | 0.69 | 0.70 | 0.84 | 0.50 | 0.69 |
| ELA | 6 | CBT | 0.73 | 0.75 | 0.81 | 0.65 | 0.73 |
| ELA | 7 | CBT | 0.67 | 0.67 | 0.75 | 0.68 | 0.67 |
| ELA | 8 | CBT | 0.62 | 0.66 | 0.79 | 0.62 | 0.62 |
| Mathematics | 3 | PBT | 0.65 | 0.70 | 0.85 | 0.68 | 0.65 |
| Mathematics | 4 | CBT | 0.73 | 0.74 | 0.86 | 0.50 | 0.73 |
| Mathematics | 4 | PBT | 0.72 | 0.74 | 0.86 | 0.48 | 0.72 |
| Mathematics | 5 | CBT | 0.70 | 0.75 | 0.79 | 0.65 | 0.70 |
| Mathematics | 6 | CBT | 0.69 | 0.71 | 0.82 | 0.71 | 0.69 |
| Mathematics | 7 | CBT | 0.74 | 0.77 | 0.84 | 0.58 | 0.74 |
| Mathematics | 8 | CBT | 0.60 | 0.66 | 0.82 | 0.59 | 0.60 |

**Table 9.4 Decision Consistency Conditioned on Level of Achievement**

| | Decision Consistency | | | | | | |
|---|---|---|---|---|---|---|---|
| Content Area | Grade | Mode | Unsatisfactory | Approaching Basic | Basic | Mastery | Advanced |
| ELA | 3 | PBT | 0.69 | 0.49 | 0.52 | 0.74 | 0.52 |
| ELA | 4 | CBT | 0.63 | 0.50 | 0.56 | 0.71 | 0.61 |
| ELA | 4 | PBT | 0.59 | 0.49 | 0.55 | 0.71 | 0.60 |
| ELA | 5 | CBT | 0.61 | 0.53 | 0.58 | 0.76 | 0.51 |
| ELA | 6 | CBT | 0.64 | 0.63 | 0.62 | 0.73 | 0.63 |
| ELA | 7 | CBT | 0.70 | 0.53 | 0.53 | 0.64 | 0.64 |
| ELA | 8 | CBT | 0.66 | 0.48 | 0.51 | 0.68 | 0.58 |
| Mathematics | 3 | PBT | 0.67 | 0.51 | 0.57 | 0.78 | 0.66 |
| Mathematics | 4 | CBT | 0.53 | 0.58 | 0.64 | 0.80 | 0.49 |
| Mathematics | 4 | PBT | 0.52 | 0.57 | 0.63 | 0.79 | 0.48 |
| Mathematics | 5 | CBT | 0.60 | 0.56 | 0.63 | 0.73 | 0.61 |
| Mathematics | 6 | CBT | 0.49 | 0.58 | 0.58 | 0.76 | 0.67 |
| Mathematics | 7 | CBT | 0.54 | 0.60 | 0.69 | 0.76 | 0.56 |
| Mathematics | 8 | CBT | 0.63 | 0.49 | 0.52 | 0.76 | 0.59 |

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 performance levels is collapsed into a dichotomized distribution around that specific cut point. As an example, the 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 9.5 shows the classification accuracy and Table 9.6 shows the consistency estimates when conditioned on LEAP 2025 cut scores. The classification accuracy statistics are at or above 0.88, while the classification consistency statistics are at or above 0.84. These results suggest that consistent and accurate performance-level classifications are being made for students in Louisiana based on the LEAP 2025.

**Table 9.5 Decision Accuracy at Achievement Cut Points**

| | | | Decision Accuracy | | | |
|---|---|---|---|---|---|---|
| **Content Area** | **Grade** | **Mode** | ***Unsatisfactory/ Approaching Basic*** | ***Approaching Basic/ Basic*** | ***Basic/ Mastery*** | ***Mastery/ Advanced*** |
| ELA | 3 | PBT | 0.94 | 0.91 | 0.90 | 0.96 |
| ELA | 4 | CBT | 0.95 | 0.91 | 0.90 | 0.96 |
| ELA | 4 | PBT | 0.95 | 0.91 | 0.89 | 0.96 |
| ELA | 5 | CBT | 0.95 | 0.91 | 0.89 | 0.98 |
| ELA | 6 | CBT | 0.95 | 0.91 | 0.92 | 0.97 |
| ELA | 7 | CBT | 0.95 | 0.91 | 0.90 | 0.94 |
| ELA | 8 | CBT | 0.95 | 0.91 | 0.89 | 0.95 |
| Mathematics | 3 | PBT | 0.95 | 0.92 | 0.91 | 0.97 |
| Mathematics | 4 | CBT | 0.94 | 0.92 | 0.93 | 0.97 |
| Mathematics | 4 | PBT | 0.94 | 0.92 | 0.92 | 0.97 |
| Mathematics | 5 | CBT | 0.93 | 0.90 | 0.92 | 0.98 |
| Mathematics | 6 | CBT | 0.90 | 0.90 | 0.93 | 0.98 |
| Mathematics | 7 | CBT | 0.92 | 0.91 | 0.94 | 0.99 |
| Mathematics | 8 | CBT | 0.88 | 0.89 | 0.93 | 0.99 |

**Table 9.6 Decision Consistency at Achievement Cut Points**

| | | | Decision Consistency | | | |
|---|---|---|---|---|---|---|
| Content Area | Grade | Mode | Unsatisfactory/ Approaching Basic | Approaching Basic/ Basic | Basic/ Mastery | Mastery/ Advanced |
| ELA | 3 | PBT | 0.92 | 0.88 | 0.86 | 0.95 |
| ELA | 4 | CBT | 0.92 | 0.87 | 0.86 | 0.94 |
| ELA | 4 | PBT | 0.93 | 0.87 | 0.85 | 0.94 |
| ELA | 5 | CBT | 0.93 | 0.87 | 0.85 | 0.97 |
| ELA | 6 | CBT | 0.93 | 0.88 | 0.89 | 0.96 |
| ELA | 7 | CBT | 0.92 | 0.87 | 0.86 | 0.92 |
| ELA | 8 | CBT | 0.92 | 0.87 | 0.85 | 0.93 |
| Mathematics | 3 | PBT | 0.92 | 0.88 | 0.88 | 0.96 |
| Mathematics | 4 | CBT | 0.92 | 0.89 | 0.90 | 0.96 |
| Mathematics | 4 | PBT | 0.91 | 0.88 | 0.89 | 0.96 |
| Mathematics | 5 | CBT | 0.91 | 0.86 | 0.89 | 0.97 |
| Mathematics | 6 | CBT | 0.87 | 0.86 | 0.90 | 0.98 |
| Mathematics | 7 | CBT | 0.89 | 0.87 | 0.91 | 0.98 |
| Mathematics | 8 | CBT | 0.84 | 0.85 | 0.90 | 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 9.1 shows test reliability for ELA and mathematics. The reliability statistics ranged from 0.87 to 0.91 for ELA forms and from 0.89 to 0.92 for mathematics forms, indicating items on the 2017 LEAP 2025 assessments measure the same construct or the same content domain. In order 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, no 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 LEAP 2025 is that the tests being calibrated are unidimensional; that is, items comprising LEAP 2025 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 LEAP 2025 by evaluating the unidimensionality assumption through principal components analysis (PCA). 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., measures one underlying construct).

A PCA was conducted for each grade, content area, and mode of LEAP 2025. 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 LEAP 2025, the ratio of the variance accounted for by the first factor to the second is sufficiently large, indicating 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 9.7) and for mathematics (see Table 9.8). To further investigate the unidimensionality of the ELA and mathematics assessments, the ratio of the first eigenvalue to the second eigenvalue was explored (see Tables 9.7 and 9.8). 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 9.7 Principal Component Analysis for English Language Arts**

| Grade | Mode | Components | Eigenvalue | Percentage of Variance Explained | Cumulative Percentage of Variance Explained |
|---|---|---|---|---|---|
| 3 | PBT | First Component | 7.20 | 25.72 | 25.72 |
| 3 | PBT | Second Component | 1.23 | 4.39 | 30.10 |
| 3 | PBT | Ratio (First/Second) | 5.86 | | |
| 4 | CBT | First Component | 7.29 | 24.30 | 24.30 |
| 4 | CBT | Second Component | 1.36 | 4.53 | 28.83 |
| 4 | CBT | Ratio (First/Second) | 5.36 | | |
| 4 | PBT | First Component | 6.75 | 22.49 | 22.49 |
| 4 | PBT | Second Component | 1.61 | 5.37 | 27.86 |
| 4 | PBT | Ratio (First/Second) | 4.19 | | |
| 5 | CBT | First Component | 6.93 | 23.10 | 23.10 |
| 5 | CBT | Second Component | 1.41 | 4.72 | 27.82 |
| 5 | CBT | Ratio (First/Second) | 4.90 | | |
| 6 | CBT | First Component | 9.14 | 26.12 | 26.12 |
| 6 | CBT | Second Component | 1.43 | 4.08 | 30.20 |
| 6 | CBT | Ratio (First/Second) | 6.40 | | |
| 7 | CBT | First Component | 7.66 | 23.93 | 23.93 |
| 7 | CBT | Second Component | 1.41 | 4.42 | 28.35 |
| 7 | CBT | Ratio (First/Second) | 5.42 | | |
| 8 | CBT | First Component | 7.13 | 22.28 | 22.28 |
| 8 | CBT | Second Component | 1.44 | 4.49 | 26.77 |
| 8 | CBT | Ratio (First/Second) | 4.97 | | |

**Table 9.8 Principal Component Analysis for Mathematics**

| Grade | Mode | Components | Eigenvalue | Percentage of Variance Explained | Cumulative Percentage of Variance Explained |
|---|---|---|---|---|---|
| 3 | PBT | First Component | 10.27 | 23.88 | 23.88 |
| 3 | PBT | Second Component | 2.00 | 4.66 | 28.53 |
| 3 | PBT | Ratio (First/Second) | 5.12 | | |
| 4 | CBT | First Component | 10.74 | 24.98 | 24.98 |
| 4 | CBT | Second Component | 1.86 | 4.33 | 29.31 |
| 4 | CBT | Ratio (First/Second) | 5.76 | | |
| 4 | PBT | First Component | 10.54 | 24.51 | 24.51 |
| 4 | PBT | Second Component | 1.76 | 4.10 | 28.62 |
| 4 | PBT | Ratio (First/Second) | 5.97 | | |
| 5 | CBT | First Component | 9.08 | 21.11 | 21.11 |
| 5 | CBT | Second Component | 1.75 | 4.08 | 25.19 |
| 5 | CBT | Ratio (First/Second) | 5.17 | | |
| 6 | CBT | First Component | 10.00 | 23.25 | 23.25 |
| 6 | CBT | Second Component | 1.99 | 4.63 | 27.88 |
| 6 | CBT | Ratio (First/Second) | 5.02 | | |
| 7 | CBT | First Component | 10.85 | 25.24 | 25.24 |
| 7 | CBT | Second Component | 1.79 | 4.16 | 29.40 |
| 7 | CBT | Ratio (First/Second) | 6.07 | | |
| 8 | CBT | First Component | 9.03 | 21.49 | 21.49 |
| 8 | CBT | Second Component | 1.42 | 3.39 | 24.88 |
| 8 | CBT | Ratio (First/Second) | 6.34 | | |

## 9.4 Analyses by Claims and Subclaims

Three sets of analyses were conducted at the claim and subclaim level for ELA and mathematics in another attempt to assess the construct validity of LEAP 2025. First, correlation coefficients that measure the relationship between the claim scores and subclaim scores were computed. Second, the reliability of each claim and subclaim was computed. Finally, the SEM was computed for each reportable claim and subclaim.

### 9.4.1 Correlations among Claims and Subclaims

This section reports the strength of the interrelationships among the claims or subclaims by computing the correlation between them. Tables 9.9–9.11 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 claim or subclaim subscores 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 9.9–9.11.

The uncorrected PPM in Tables 9.9–9.11 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 claim or subclaim is moderately to strongly related to performance on another claim or subclaim within the same grade and content area. The value of the correlation coefficients will be affected by the limited number of items measuring each claim or subclaim. Therefore, caution should be used when comparing the PPM coefficients measuring the relationships between claims or subclaims to those measuring the relationships between content areas (see Table 9.11). A more modest relationship (i.e., smaller correlation coefficients) is expected to be reported between the claims or subclaims as a consequence of the lower number of items measuring each of the reporting categories. The PPM between two claim or subclaim subscores 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 for the other claim or GLE strand.

ELA shows moderate relationships between reading and writing claims across all grades, indicating these two claims measure some different traits. Across all tables, the CAPPM indicates moderate or strong relationships between subclaims. The CAPPM for reading vocabulary, written expression, and knowledge and use of language was 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 9.9–9.11 are further evidence of the validity of the test construct. Since the overall content area is comprised of the claim or subclaim subscores and the content area is expected to measure a single dimension, these subscores are expected to be moderately or highly related.

**Table 9.9 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Claims: English Language Arts**

| Grade | Mode | No. | Claim | N Items | 1 | 2 |
|---|---|---|---|---|---|---|
| 3 | PBT | 1 | Reading | 24 | . | 0.84 |
| | PBT | 2 | Writing | 4 | 0.69 | . |
| 4 | CBT | 1 | Reading | 26 | . | 0.84 |
| | CBT | 2 | Writing | 4 | 0.72 | . |
| | PBT | 1 | Reading | 26 | . | 0.76 |
| | PBT | 2 | Writing | 4 | 0.65 | . |
| 5 | CBT | 1 | Reading | 26 | . | 0.81 |
| | CBT | 2 | Writing | 4 | 0.69 | . |
| 6 | CBT | 1 | Reading | 31 | . | 0.81 |
| | CBT | 2 | Writing | 4 | 0.71 | . |
| 7 | CBT | 1 | Reading | 28 | . | 0.85 |
| | CBT | 2 | Writing | 4 | 0.75 | . |
| 8 | CBT | 1 | Reading | 28 | . | 0.87 |
| | CBT | 2 | Writing | 4 | 0.75 | . |

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

| Subclaim Uncorrected and Corrected Correlation Coefficients: English Language Arts | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Grade | Mode | No. | Subclaim | N Items | 1 | 2 | 3 | 4 | 5 |
| 3 | PBT | 1 | Reading Literary Text | 11 | . | 1.03 | 1.00 | 0.97 | 0.81 |
| | PBT | 2 | Reading Information Text | 8 | 0.69 | . | 1.04 | 1.06 | 0.88 |
| | PBT | 3 | Reading Vocabulary | 7 | 0.70 | 0.68 | . | 0.86 | 0.78 |
| | PBT | 4 | Written Expression | 2 | 0.61 | 0.62 | 0.52 | . | 1.13 |
| | PBT | 5 | Knowledge & Use of Language | 2 | 0.57 | 0.58 | 0.54 | 0.70 | . |
| 4 | CBT | 1 | Reading Literary Text | 11 | . | 1.01 | 0.94 | 0.99 | 0.95 |
| | CBT | 2 | Reading Information Text | 9 | 0.67 | . | 0.99 | 1.06 | 1.03 |
| | CBT | 3 | Reading Vocabulary | 8 | 0.64 | 0.64 | . | 0.81 | 0.79 |
| | CBT | 4 | Written Expression | 2 | 0.67 | 0.68 | 0.54 | . | 1.43 |
| | CBT | 5 | Knowledge & Use of Language | 2 | 0.64 | 0.67 | 0.53 | 0.94 | . |
| | PBT | 1 | Reading Literary Text | 11 | . | 0.99 | 0.93 | 0.89 | 0.86 |
| | PBT | 2 | Reading Information Text | 9 | 0.63 | . | 1.04 | 1.04 | 1.01 |
| | PBT | 3 | Reading Vocabulary | 8 | 0.64 | 0.64 | . | 0.73 | 0.72 |
| | PBT | 4 | Written Expression | 2 | 0.60 | 0.62 | 0.48 | . | 1.51 |
| | PBT | 5 | Knowledge & Use of Language | 2 | 0.58 | 0.61 | 0.47 | 0.96 | . |

| Subclaim Uncorrected and Corrected Correlation Coefficients: English Language Arts (continued) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Grade** | **Mode** | **No.** | **Subclaim** | **N Items** | **1** | **2** | **3** | **4** | **5** |
| 5 | CBT | 1 | Reading Literary Text | 10 | . | 1.05 | 1.00 | 1.00 | 0.93 |
| | CBT | 2 | Reading Information Text | 11 | 0.69 | . | 0.98 | 0.92 | 0.87 |
| | CBT | 3 | Reading Vocabulary | 7 | 0.62 | 0.65 | . | 0.81 | 0.81 |
| | CBT | 4 | Written Expression | 2 | 0.63 | 0.62 | 0.52 | . | 1.30 |
| | CBT | 5 | Knowledge & Use of Language | 2 | 0.60 | 0.61 | 0.53 | 0.87 | . |
| 6 | CBT | 1 | Reading Literary Text | 14 | . | 0.97 | 0.97 | 0.84 | 0.79 |
| | CBT | 2 | Reading Information Text | 13 | 0.75 | . | 0.97 | 0.95 | 0.88 |
| | CBT | 3 | Reading Vocabulary | 5 | 0.64 | 0.62 | . | 0.84 | 0.80 |
| | CBT | 4 | Written Expression | 2 | 0.62 | 0.68 | 0.51 | . | 1.21 |
| | CBT | 5 | Knowledge & Use of Language | 2 | 0.62 | 0.67 | 0.51 | 0.86 | . |
| 7 | CBT | 1 | Reading Literary Text | 10 | . | 1.06 | 1.01 | 0.99 | 0.97 |
| | CBT | 2 | Reading Information Text | 13 | 0.71 | . | 1.04 | 0.91 | 0.91 |
| | CBT | 3 | Reading Vocabulary | 7 | 0.63 | 0.67 | . | 0.80 | 0.81 |
| | CBT | 4 | Written Expression | 2 | 0.71 | 0.68 | 0.55 | . | 1.18 |
| | CBT | 5 | Knowledge & Use of Language | 2 | 0.70 | 0.68 | 0.56 | 0.95 | . |
| 8 | CBT | 1 | Reading Literary Text | 9 | . | 1.16 | 1.10 | 1.15 | 1.13 |
| | CBT | 2 | Reading Information Text | 13 | 0.64 | . | 1.02 | 0.97 | 0.95 |
| | CBT | 3 | Reading Vocabulary | 8 | 0.59 | 0.67 | . | 0.77 | 0.78 |
| | CBT | 4 | Written Expression | 2 | 0.68 | 0.71 | 0.55 | . | 1.24 |
| | CBT | 5 | Knowledge & Use of Language | 2 | 0.67 | 0.69 | 0.55 | 0.96 | . |

**Table 9.11 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Subclaims: Mathematics**

| **Grade** | **Mode** | **No.** | **Subclaim** | **N Items** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|---|---|---|---|
| 3 | PBT | 1 | Major Content | 28 | . | 1.00 | 0.96 | 0.97 |
| | PBT | 2 | Additional & Supporting Con | 9 | 0.78 | . | 0.99 | 1.02 |
| | PBT | 3 | Expressing Mathematical Rea | 3 | 0.72 | 0.66 | . | 1.07 |
| | PBT | 4 | Modeling & Application | 3 | 0.75 | 0.70 | 0.71 | . |
| 4 | CBT | 1 | Major Content | 27 | . | 0.97 | 0.95 | 1.15 |
| | CBT | 2 | Additional & Supporting Con | 10 | 0.74 | . | 0.99 | 1.17 |
| | CBT | 3 | Expressing Mathematical Rea | 3 | 0.75 | 0.68 | . | 1.23 |
| | CBT | 4 | Modeling & Application | 2 | 0.72 | 0.63 | 0.69 | . |
| | PBT | 1 | Major Content | 27 | . | 0.96 | 1.00 | 1.12 |
| | PBT | 2 | Additional & Supporting Con | 10 | 0.74 | . | 1.01 | 1.12 |
| | PBT | 3 | Expressing Mathematical Rea | 3 | 0.76 | 0.68 | . | 1.20 |
| | PBT | 4 | Modeling & Application | 2 | 0.72 | 0.63 | 0.68 | . |

| Subclaim Uncorrected and Corrected Correlation Coefficients: Mathematics (continued) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Grade** | **Mode** | **No.** | **Subclaim** | **N Items** | **1** | **2** | **3** | **4** |
| 5 | CBT | 1 | Major Content | 27 | . | 0.92 | 0.96 | 0.92 |
| | CBT | 2 | Additional & Supporting Con | 9 | 0.63 | . | 0.89 | 0.86 |
| | CBT | 3 | Expressing Mathematical Rea | 3 | 0.69 | 0.52 | . | 1.01 |
| | CBT | 4 | Modeling & Application | 3 | 0.70 | 0.52 | 0.65 | . |
| 6 | CBT | 1 | Major Content | 26 | . | 0.92 | 1.00 | 1.01 |
| | CBT | 2 | Additional & Supporting Con | 8 | 0.65 | . | 0.96 | 1.08 |
| | CBT | 3 | Expressing Mathematical Rea | 4 | 0.77 | 0.60 | . | 1.11 |
| | CBT | 4 | Modeling & Application | 3 | 0.70 | 0.60 | 0.68 | . |
| 7 | CBT | 1 | Major Content | 28 | . | 0.98 | 0.98 | 1.09 |
| | CBT | 2 | Additional & Supporting Con | 8 | 0.70 | . | 1.01 | 1.06 |
| | CBT | 3 | Expressing Mathematical Rea | 4 | 0.79 | 0.67 | . | 1.09 |
| | CBT | 4 | Modeling & Application | 3 | 0.78 | 0.63 | 0.72 | . |
| 8 | CBT | 1 | Major Content | 27 | . | 1.00 | 1.00 | 0.98 |
| | CBT | 2 | Additional & Supporting Con | 8 | 0.75 | . | 0.97 | 1.01 |
| | CBT | 3 | Expressing Mathematical Rea | 4 | 0.69 | 0.61 | . | 1.05 |
| | CBT | 4 | Modeling & Application | 3 | 0.67 | 0.63 | 0.60 | . |

### 9.4.2 Reliability of Claims or Subclaims

Raw score summary statistics (i.e., mean and standard deviation), Cronbach's (1951) coefficient alpha, and SEM were computed for each of the claims or subclaims by grade, content area, and mode using the census data. These statistics are presented in Tables 9.12–9.14 for ELA and mathematics. Reliability indices, such as Cronbach's coefficient alpha (and resulting SEM), are a function of the number of test items, the average covariance between item-pairs, and the variance of the total score. In general, it is expected that the coefficient alpha would be lower for a claim or subclaim assessed by a small number of items compared to a claim or subclaim assessed by a larger number of items.

### 9.4.3 Standard Error of Measurement of Claims or Subclaims

This chapter also reports the SEM associated with each of the claims and subclaims in Tables 9.12–9.14 for ELA and mathematics. These SEMs are reported in the raw score metric.

**Table 9.12 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of English Language Arts Claims**

| Grade | Mode | Claim | Number of Items | Number of Score Points | Mean Raw Score | Raw Score Std. Dev. | SEM | Cronbach's Alpha |
|---|---|---|---|---|---|---|---|---|
| 3 | PBT | Reading | 24 | 48 | 25.03 | 10.81 | 4.01 | 0.86 |
| | PBT | Writing | 4 | 24 | 6.93 | 4.60 | 2.07 | 0.80 |
| 4 | CBT | Reading | 26 | 52 | 23.44 | 10.65 | 4.15 | 0.85 |
| | CBT | Writing | 4 | 30 | 7.74 | 5.59 | 2.00 | 0.87 |
| | PBT | Reading | 26 | 52 | 23.30 | 10.50 | 4.14 | 0.84 |
| | PBT | Writing | 4 | 30 | 8.65 | 5.28 | 1.92 | 0.87 |
| 5 | CBT | Reading | 26 | 52 | 24.78 | 10.41 | 4.09 | 0.85 |
| | CBT | Writing | 4 | 30 | 7.83 | 5.43 | 2.05 | 0.86 |
| 6 | CBT | Reading | 31 | 62 | 27.69 | 12.61 | 4.20 | 0.89 |
| | CBT | Writing | 4 | 30 | 8.92 | 6.80 | 2.45 | 0.87 |
| 7 | CBT | Reading | 28 | 56 | 27.80 | 11.86 | 4.64 | 0.85 |
| | CBT | Writing | 4 | 30 | 12.08 | 7.52 | 2.06 | 0.92 |
| 8 | CBT | Reading | 28 | 56 | 28.77 | 10.86 | 4.63 | 0.82 |
| | CBT | Writing | 4 | 30 | 13.75 | 8.18 | 2.35 | 0.92 |

**Table 9.13 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of English Language Arts Subclaims**

| Mean, Standard Deviation, and SEM: English Language Arts | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Grade | Mode | Subclaim | Number of Items | Number of Score Points | Mean Raw Score | Raw Score Std. Dev. | SEM | Cronbach's Alpha |
| 3 | PBT | Reading Literary Text | 10 | 20 | 11.07 | 4.69 | 2.51 | 0.71 |
| | PBT | Reading Information Text | 7 | 14 | 6.86 | 3.70 | 2.25 | 0.63 |
| | PBT | Reading Vocabulary | 7 | 14 | 7.09 | 3.73 | 2.09 | 0.69 |
| | PBT | Written Expression | 4 | 18 | 4.78 | 3.56 | 2.40 | 0.55 |
| | PBT | Knowledge & Use of Language | 2 | 6 | 2.15 | 1.34 | 0.73 | 0.70 |
| 4 | CBT | Reading Literary Text | 10 | 20 | 10.78 | 4.63 | 2.57 | 0.69 |
| | CBT | Reading Information Text | 8 | 16 | 6.39 | 3.77 | 2.29 | 0.63 |
| | CBT | Reading Vocabulary | 8 | 16 | 6.28 | 3.76 | 2.18 | 0.67 |
| | CBT | Written Expression | 4 | 24 | 5.77 | 4.27 | 2.51 | 0.65 |
| | CBT | Knowledge & Use of Language | 2 | 6 | 1.97 | 1.38 | 0.80 | 0.66 |
| | PBT | Reading Literary Text | 10 | 20 | 11.00 | 4.84 | 2.60 | 0.71 |
| | PBT | Reading Information Text | 8 | 16 | 6.22 | 3.45 | 2.28 | 0.56 |
| | PBT | Reading Vocabulary | 8 | 16 | 6.08 | 3.76 | 2.18 | 0.66 |
| | PBT | Written Expression | 4 | 24 | 6.46 | 4.01 | 2.43 | 0.63 |
| | PBT | Knowledge & Use of Language | 2 | 6 | 2.19 | 1.31 | 0.79 | 0.64 |
| 5 | CBT | Reading Literary Text | 9 | 18 | 6.76 | 3.80 | 2.37 | 0.61 |
| | CBT | Reading Information Text | 10 | 20 | 9.57 | 4.74 | 2.60 | 0.70 |
| | CBT | Reading Vocabulary | 7 | 14 | 8.45 | 3.27 | 1.98 | 0.63 |
| | CBT | Written Expression | 4 | 24 | 5.59 | 4.10 | 2.44 | 0.65 |
| | CBT | Knowledge & Use of Language | 2 | 6 | 2.24 | 1.47 | 0.82 | 0.69 |
| 6 | CBT | Reading Literary Text | 14 | 28 | 12.52 | 5.94 | 2.62 | 0.81 |
| | CBT | Reading Information Text | 12 | 24 | 10.06 | 5.63 | 2.79 | 0.75 |
| | CBT | Reading Vocabulary | 5 | 10 | 5.11 | 2.45 | 1.66 | 0.54 |
| | CBT | Written Expression | 3 | 24 | 6.60 | 5.36 | 3.07 | 0.67 |
| | CBT | Knowledge & Use of Language | 2 | 6 | 2.32 | 1.61 | 0.79 | 0.76 |
| 7 | CBT | Reading Literary Text | 9 | 18 | 9.22 | 4.58 | 2.71 | 0.65 |
| | CBT | Reading Information Text | 12 | 24 | 11.57 | 5.49 | 3.04 | 0.69 |
| | CBT | Reading Vocabulary | 7 | 14 | 7.00 | 3.27 | 2.05 | 0.61 |
| | CBT | Written Expression | 4 | 24 | 9.10 | 5.78 | 2.57 | 0.80 |
| | CBT | Knowledge & Use of Language | 2 | 6 | 2.98 | 1.81 | 0.80 | 0.81 |
| 8 | CBT | Reading Literary Text | 8 | 16 | 8.76 | 3.67 | 2.71 | 0.46 |
| | CBT | Reading Information Text | 12 | 24 | 10.98 | 5.19 | 2.96 | 0.67 |
| | CBT | Reading Vocabulary | 8 | 16 | 9.02 | 3.57 | 2.12 | 0.65 |
| | CBT | Written Expression | 4 | 24 | 10.44 | 6.34 | 2.99 | 0.78 |
| | CBT | Knowledge & Use of Language | 2 | 6 | 3.31 | 1.88 | 0.89 | 0.78 |

**Table 9.14 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Mathematics Subclaims**

| Mean, Standard Deviation, and SEM: Mathematics | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Grade | Mode | Subclaim | Number of Items | Number of Score Points | Mean Raw Score | Raw Score Std. Dev. | SEM | Cronbach's Alpha |
| 3 | PBT | Major Content | 28 | 30 | 18.84 | 6.18 | 2.18 | 0.88 |
| | PBT | Additional & Supporting Content | 9 | 10 | 6.07 | 2.16 | 1.21 | 0.69 |
| | PBT | Expressing Mathematical Reasoning | 3 | 9 | 2.92 | 2.36 | 1.41 | 0.64 |
| | PBT | Modeling & Application | 3 | 7 | 3.78 | 3.44 | 1.96 | 0.68 |
| 4 | CBT | Major Content | 27 | 30 | 17.24 | 6.53 | 2.23 | 0.88 |
| | CBT | Additional & Supporting Content | 10 | 10 | 4.39 | 2.28 | 1.32 | 0.66 |
| | CBT | Expressing Mathematical Reasoning | 3 | 6 | 2.46 | 2.43 | 1.31 | 0.71 |
| | CBT | Modeling & Application | 2 | 4 | 6.06 | 2.37 | 1.77 | 0.44 |
| | PBT | Major Content | 27 | 30 | 17.47 | 6.45 | 2.24 | 0.88 |
| | PBT | Additional & Supporting Content | 10 | 10 | 4.38 | 2.31 | 1.33 | 0.67 |
| | PBT | Expressing Mathematical Reasoning | 3 | 6 | 3.08 | 2.66 | 1.54 | 0.67 |
| | PBT | Modeling & Application | 2 | 4 | 6.75 | 2.61 | 1.89 | 0.47 |
| 5 | CBT | Major Content | 27 | 29 | 16.08 | 6.02 | 2.36 | 0.85 |
| | CBT | Additional & Supporting Content | 9 | 10 | 5.06 | 2.05 | 1.38 | 0.55 |
| | CBT | Expressing Mathematical Reasoning | 3 | 5 | 2.10 | 2.01 | 1.25 | 0.62 |
| | CBT | Modeling & Application | 3 | 9 | 2.87 | 2.81 | 1.59 | 0.68 |
| 6 | CBT | Major Content | 26 | 29 | 15.78 | 6.07 | 2.25 | 0.86 |
| | CBT | Additional & Supporting Content | 8 | 9 | 4.08 | 1.97 | 1.29 | 0.57 |
| | CBT | Expressing Mathematical Reasoning | 4 | 9 | 4.07 | 3.18 | 1.77 | 0.69 |
| | CBT | Modeling & Application | 3 | 9 | 2.03 | 2.92 | 1.96 | 0.55 |
| 7 | CBT | Major Content | 28 | 30 | 12.72 | 6.32 | 2.28 | 0.87 |
| | CBT | Additional & Supporting Content | 8 | 10 | 2.20 | 1.84 | 1.17 | 0.59 |
| | CBT | Expressing Mathematical Reasoning | 4 | 11 | 3.74 | 3.27 | 1.63 | 0.75 |
| | CBT | Modeling & Application | 3 | 5 | 3.83 | 3.14 | 2.01 | 0.59 |
| 8 | CBT | Major Content | 27 | 30 | 10.98 | 5.44 | 2.29 | 0.82 |
| | CBT | Additional & Supporting Content | 8 | 10 | 4.20 | 2.33 | 1.32 | 0.68 |
| | CBT | Expressing Mathematical Reasoning | 4 | 9 | 2.13 | 2.16 | 1.41 | 0.58 |
| | CBT | Modeling & Application | 3 | 9 | 3.15 | 3.04 | 2.00 | 0.57 |

## 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 tests, correlations were computed between the ELA and mathematics scale scores for students who took more than one LEAP 2025 content-area test in 2017. These correlations are based on the census data, and the results are shown in Table 9.15. The correlation coefficients ranged from 0.70 (between ELA and mathematics in grade 4) to 0.76 (between ELA and mathematics in grades 6 and 8). The correlation coefficients suggest that individual student scores for ELA and mathematics are moderately related, indicating that these two tests measure a similar knowledge base or general underlying ability but still measure some different traits as planned.

**Table 9.15 Inter-Correlation of English Language Arts and Mathematics Scale Scores**

| Grade | ELA/Mathematics |
|---|---|
| 3 | 0.74 |
| 4 | 0.70 |
| 5 | 0.71 |
| 6 | 0.76 |
| 7 | 0.76 |
| 8 | 0.70 |

## 9.6 Summary

In summary, the overall purpose of establishing construct validity is to ensure that the meaning of test scores is supported. Evidence of validity is necessary to justify the use of the LEAP 2025. 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. It should be noted that differences in test performance among subgroups does not mean that a test is unfair—it simply means that groups perform differently on the 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 is particularly relevant 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)

There is no particular research on LEAP 2025 showing that the test scores of examinee subgroups differ in meaning; however, this is 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, LDOE assessment research and development experts 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 conducted differential item functioning (DIF) studies of their items prior to PARCC operational administrations. Items are typically evaluated for possible DIF in the field test phase of the test development, and any items flagged for DIF are further examined to determine possible bias. During the ELA and mathematics test development, DRC content experts tried to avoid including PARCC operational items flagged for DIF. Section 10.2 of this chapter explains the steps taken to evaluate LEAP 2025 items through the use of DIF to adhere to this standard.

In addition, standardized test administration and extensive training of 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 section 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 PARCC operational items used in the 2017 LEAP 2025 forms were critical to the forms construction process. Refer to the PARCC website for the bias and sensitivity guidelines used and the processes and procedures followed by PARCC pertaining to these items (see http://parcconline.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 several published 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 at least these same people and was also subjected to an independent review by 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 (see Tables 10.1 and 10.2) indicate the degree to which members of a particular subgroup 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 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. 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 *k*th 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 ($\chi^2_{MH}$) 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 $\chi^2_{MH}$, 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 ($\Delta_{MH}$) 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):

$$SMD = p_{Fk}(\sum_{k} m_{Fk} - \sum_{k} m_{Rk}),$$

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

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

***LEP Status:*** Focal group is students who are classified as LEP; reference group is all others.

***Economic Status:*** Focal group is students who are classified as economically disadvantaged; 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 differences should they exist.

Tables 10.1 and 10.2 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, one item was flagged for DIF for the female subgroup: it exhibited moderate positive DIF. One item each was flagged for the Hispanic/Latino and American Indian or Alaska Native subgroups; these items showed moderate negative DIF. Two items were flagged for the Asian subgroup: one showed moderate negative DIF and one showed moderate positive DIF. Two items were flagged for the special education subgroup: one showed moderate negative DIF and one showed large negative DIF. Lastly, one item was flagged for the 504 subgroup: it showed moderate negative DIF.

**Table 10.1 2017 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 | Group | B- | B+ | C- | C+ |
| 3 | PBT | 29 | Gender | Female | 0 | 1 | 0 | 0 |
| | | | Ethnicity | Hispanic/Latino | 1 | 0 | 0 | 0 |
| | | | Ethnicity | American Indian or Alaska Native | 1 | 0 | 0 | 0 |
| | | | Ethnicity | Asian | 1 | 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 | 0 | 1 | 0 |
| | | | LEP Status | LEP | 0 | 0 | 0 | 0 |
| | | | Economic Status | Economically Disadvantaged | 0 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 1 | 0 | 0 | 0 |

| DIF Statistics: English Language Arts (continued) | | | | | Count of Items at DIF Magnitude | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Moderate | | Large | |
| Grade | Mode | Number of Items | Category | Group | B- | B+ | C- | C+ |
| 4 | CBT | 36 | Gender | Female | 0 | 2 | 0 | 0 |
| | | | Ethnicity | Hispanic/Latino | NR | NR | NR | NR |
| | | | Ethnicity | American Indian or Alaska Native | NR | NR | NR | NR |
| | | | Ethnicity | Asian | NR | NR | NR | NR |
| | | | Ethnicity | Black or African American | 2 | 0 | 0 | 0 |
| | | | Ethnicity | Two or More Races | NR | NR | NR | NR |
| | | | Education Classification | Special | 1 | 0 | 0 | 0 |
| | | | LEP Status | LEP | | | | |
| | | | Economic Status | Economically Disadvantaged | 2 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | NR | NR | NR | NR |
| | PBT | 36 | Gender | Female | 0 | 2 | 0 | 0 |
| | | | Ethnicity | Hispanic/Latino | 1 | 0 | 0 | 0 |
| | | | Ethnicity | American Indian or Alaska Native | 1 | 0 | 0 | 0 |
| | | | Ethnicity | Asian | 2 | 0 | 0 | 0 |
| | | | Ethnicity | Black or African American | 2 | 0 | 0 | 0 |
| | | | Ethnicity | Two or More Races | 0 | 0 | 0 | 0 |
| | | | Education Classification | Special | 5 | 1 | 0 | 0 |
| | | | LEP Status | LEP | 1 | 0 | 0 | 0 |
| | | | Economic Status | Economically Disadvantaged | 0 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 0 | 0 | 0 | 0 |
| 5 | CBT | 31 | Gender | Female | 0 | 4 | 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 | 5 | 0 | 0 | 0 |
| | | | LEP Status | LEP | 1 | 0 | 1 | 0 |
| | | | Economic Status | Economically Disadvantaged | 0 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 0 | 0 | 0 | 0 |

| DIF Statistics: English Language Arts (continued) | | | | | Count of Items at DIF Magnitude | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Moderate | | Large | |
| Grade | Mode | Number of Items | Category | Group | B- | B+ | C- | C+ |
| 6 | CBT | 31 | Gender | Female | 1 | 2 | 0 | 2 |
| | | | Ethnicity | Hispanic/Latino | 0 | 0 | 0 | 0 |
| | | | Ethnicity | American Indian or Alaska Native | 0 | 0 | 0 | 0 |
| | | | Ethnicity | Asian | 2 | 2 | 0 | 0 |
| | | | Ethnicity | Black or African American | 0 | 0 | 0 | 0 |
| | | | Ethnicity | Two or More Races | 0 | 0 | 0 | 0 |
| | | | Education Classification | Special | 2 | 0 | 3 | 0 |
| | | | LEP Status | LEP | 2 | 0 | 0 | 0 |
| | | | Economic Status | Economically Disadvantaged | 0 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 0 | 0 | 0 | 0 |
| 7 | CBT | 30 | Gender | Female | 1 | 0 | 0 | 4 |
| | | | Ethnicity | Hispanic/Latino | 0 | 0 | 1 | 0 |
| | | | Ethnicity | American Indian or Alaska Native | 0 | 0 | 0 | 0 |
| | | | Ethnicity | Asian | 0 | 3 | 0 | 0 |
| | | | Ethnicity | Black or African American | 0 | 0 | 1 | 0 |
| | | | Ethnicity | Two or More Races | 0 | 0 | 0 | 0 |
| | | | Education Classification | Special | 2 | 0 | 2 | 0 |
| | | | LEP Status | LEP | 0 | 0 | 2 | 0 |
| | | | Economic Status | Economically Disadvantaged | 0 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 0 | 0 | 0 | 0 |
| 8 | CBT | 36 | Gender | Female | 1 | 1 | 0 | 3 |
| | | | Ethnicity | Hispanic/Latino | 1 | 0 | 1 | 0 |
| | | | Ethnicity | American Indian or Alaska Native | 0 | 0 | 0 | 0 |
| | | | Ethnicity | Asian | 2 | 2 | 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 | 2 | 0 |
| | | | LEP Status | LEP | 1 | 0 | 2 | 0 |
| | | | Economic Status | Economically Disadvantaged | 0 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 0 | 0 | 0 | 0 |

**Table 10.2 2017 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 | Group | B- | B+ | C- | C+ |
| 3 | PBT | 45 | Gender | Female | 0 | 1 | 0 | 0 |
| | | | Ethnicity | Hispanic/Latino | 1 | 0 | 0 | 0 |
| | | | Ethnicity | American Indian or Alaska Native | 0 | 0 | 0 | 0 |
| | | | Ethnicity | Asian | 0 | 1 | 0 | 0 |
| | | | Ethnicity | Black or African American | 2 | 1 | 0 | 0 |
| | | | Ethnicity | Two or More Races | 0 | 0 | 0 | 0 |
| | | | Education Classification | Special | 2 | 0 | 1 | 0 |
| | | | LEP Status | LEP | 0 | 1 | 0 | 0 |
| | | | Economic Status | Economically Disadvantaged | 1 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 0 | 0 | 0 | 0 |
| 4 | CBT | 44 | Gender | Female | 0 | 0 | 0 | 1 |
| | | | Ethnicity | Hispanic/Latino | NR | NR | NR | NR |
| | | | Ethnicity | American Indian or Alaska Native | NR | NR | NR | NR |
| | | | Ethnicity | Asian | NR | NR | NR | NR |
| | | | Ethnicity | Black or African American | 0 | 1 | 1 | 0 |
| | | | Ethnicity | Two or More Races | NR | NR | NR | NR |
| | | | Education Classification | Special | 6 | 1 | 1 | 5 |
| | | | LEP Status | LEP | NR | NR | NR | NR |
| | | | Economic Status | Economically Disadvantaged | 1 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | NR | NR | NR | NR |
| | PBT | 44 | Gender | Female | 0 | 1 | 0 | 0 |
| | | | Ethnicity | Hispanic/Latino | 0 | 0 | 0 | 0 |
| | | | Ethnicity | American Indian or Alaska Native | 0 | 0 | 0 | 0 |
| | | | Ethnicity | Asian | 0 | 0 | 0 | 1 |
| | | | Ethnicity | Black or African American | 0 | 0 | 1 | 0 |
| | | | Ethnicity | Two or More Races | 0 | 0 | 0 | 0 |
| | | | Education Classification | Special | 3 | 1 | 1 | 4 |
| | | | LEP Status | LEP | 1 | 0 | 0 | 0 |
| | | | Economic Status | Economically Disadvantaged | 1 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 1 | 0 | 0 | 0 |

| DIF Statistics: Mathematics (continued) | | | | | Count of Items at DIF Magnitude | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Moderate | | Large | |
| Grade | Mode | Number of Items | Category | Group | B- | B+ | C- | C+ |
| 5 | CBT | 44 | 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 | 1 | 0 | 0 | 0 |
| | | | Ethnicity | Two or More Races | 0 | 0 | 0 | 0 |
| | | | Education Classification | Special | 5 | 1 | 1 | 4 |
| | | | LEP Status | LEP | 0 | 0 | 0 | 0 |
| | | | Economic Status | Economically Disadvantaged | 1 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 0 | 0 | 0 | 0 |
| 6 | CBT | 44 | Gender | Female | 1 | 0 | 0 | 0 |
| | | | Ethnicity | Hispanic/Latino | 0 | 0 | 0 | 0 |
| | | | Ethnicity | American Indian or Alaska Native | 2 | 1 | 0 | 1 |
| | | | Ethnicity | Asian | 0 | 0 | 1 | 0 |
| | | | Ethnicity | Black or African American | 0 | 0 | 0 | 0 |
| | | | Ethnicity | Two or More Races | 3 | 0 | 3 | 3 |
| | | | Education Classification | Special | 1 | 0 | 0 | 0 |
| | | | LEP Status | LEP | 1 | 0 | 0 | 0 |
| | | | Economic Status | Economically Disadvantaged | 0 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 1 | 1 | 0 | 0 |
| 7 | CBT | 42 | Gender | Female | 3 | 1 | 0 | 0 |
| | | | Ethnicity | Hispanic/Latino | 0 | 0 | 0 | 0 |
| | | | Ethnicity | American Indian or Alaska Native | 0 | 0 | 0 | 0 |
| | | | Ethnicity | Asian | 2 | 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 | 1 | 1 | 0 |
| | | | LEP Status | LEP | 1 | 0 | 0 | 0 |
| | | | Economic Status | Economically Disadvantaged | 0 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 0 | 0 | 0 | 0 |
| 8 | CBT | 45 | Gender | Female | 0 | 1 | 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 | 0 | 1 | 1 | 0 |
| | | | LEP Status | LEP | 1 | 0 | 0 | 0 |
| | | | Economic Status | Economically Disadvantaged | 1 | 0 | 0 | 0 |
| | | | Section 504 Status | Section 504 | 0 | 0 | 0 | 0 |

## 10.3 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 10.3–10.16 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.

### 10.3.1 Reliability

Tables 10.3–10.9 show the regular test form reliability coefficients and SEM by student gender, ethnicity, education classification, LEP status, economic status, and 504 status. The reliability coefficients for English Language Arts forms ranged from 0.82 to 0.93, except for the CBT in grade 4, which had a reliability coefficient of 0.77 for the LEP group. For mathematics the reliability coefficients ranged from 0.84 to 0.93. This analysis shows that the test reliability is of acceptable magnitude for all the subgroups. Note that the reliability coefficients are NR for subgroups smaller than 10 students.

**Table 10.3 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** | ≥ 56,800 | 0.89 | 4.83 | ≥ 56,800 | 0.91 | 3.73 |
| **Gender** | | | | | | |
| Female | ≥ 27,840 | 0.89 | 4.89 | ≥ 27,840 | 0.91 | 3.76 |
| Male | ≥ 28,920 | 0.89 | 4.75 | ≥ 28,920 | 0.92 | 3.70 |
| **Ethnicity** | | | | | | |
| Hispanic/Latino | ≥ 4,230 | 0.90 | 4.80 | ≥ 4,220 | 0.91 | 3.68 |
| American Indian or Alaska Native | ≥ 320 | 0.88 | 4.94 | ≥ 320 | 0.92 | 3.79 |
| Asian | ≥ 890 | 0.90 | 4.94 | ≥ 890 | 0.91 | 3.86 |
| Black or African American | ≥ 25,150 | 0.87 | 4.79 | ≥ 25,150 | 0.91 | 3.50 |
| Native Hawaiian or Other Pacific | ≥ 70 | 0.92 | 4.67 | ≥ 70 | 0.93 | 3.82 |
| White | ≥ 24,440 | 0.88 | 4.87 | ≥ 24,430 | 0.90 | 3.82 |
| Two or More Races | ≥ 1,630 | 0.87 | 4.86 | ≥ 1,630 | 0.91 | 3.73 |
| **Education Classification** | | | | | | |
| Regular | ≥ 50,450 | 0.88 | 4.84 | ≥ 50,440 | 0.91 | 3.76 |
| Special | ≥ 6,350 | 0.88 | 4.59 | ≥ 6,350 | 0.92 | 3.38 |
| **LEP Status** | | | | | | |
| Non-LEP | ≥ 54,100 | 0.89 | 4.84 | ≥ 54,110 | 0.91 | 3.74 |
| LEP | ≥ 2,690 | 0.87 | 4.68 | ≥ 2,690 | 0.91 | 3.51 |
| **Economic Status** | | | | | | |
| Economically Disadvantaged | ≥ 41,240 | 0.88 | 4.80 | ≥ 41,200 | 0.91 | 3.63 |
| Not Economically Disadvantaged | ≥ 14,690 | 0.87 | 4.90 | ≥ 14,700 | 0.90 | 3.83 |
| **Section 504 Status** | | | | | | |
| Non-Section 504 | ≥ 51,890 | 0.89 | 4.84 | ≥ 51,890 | 0.91 | 3.75 |
| Section 504 | ≥ 4,900 | 0.86 | 4.63 | ≥ 4,900 | 0.91 | 3.42 |

**Table 10.4 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** | ≥ 1,930 | 0.88 | 5.19 | ≥ 1,930 | 0.92 | 3.53 |
| **Gender** | | | | | | |
| Female | ≥ 960 | 0.88 | 5.28 | ≥ 960 | 0.91 | 3.57 |
| Male | ≥ 970 | 0.88 | 5.05 | ≥ 970 | 0.92 | 3.47 |
| **Ethnicity** | | | | | | |
| Hispanic/Latino | ≥ 150 | 0.86 | 5.21 | ≥ 150 | 0.91 | 3.46 |
| American Indian or Alaska Native | < 10 | NR | NR | < 10 | NR | NR |
| Asian | ≥ 20 | 0.85 | 5.72 | ≥ 20 | 0.92 | 3.59 |
| Black or African American | ≥ 540 | 0.83 | 5.13 | ≥ 540 | 0.90 | 3.32 |
| Native Hawaiian or Other Pacific | < 10 | NR | NR | < 10 | NR | NR |
| White | ≥ 1,140 | 0.88 | 5.21 | ≥ 1,130 | 0.91 | 3.56 |
| Two or More Races | ≥ 60 | 0.88 | 5.35 | ≥ 60 | 0.91 | 3.59 |
| **Education Classification** | | | | | | |
| Regular | ≥ 1,700 | 0.88 | 5.24 | ≥ 1,700 | 0.92 | 3.53 |
| Special | ≥ 230 | 0.88 | 4.60 | ≥ 230 | 0.90 | 3.26 |
| **LEP Status** | | | | | | |
| Non-LEP | ≥ 1,860 | 0.88 | 5.20 | ≥ 1,860 | 0.92 | 3.54 |
| LEP | ≥ 70 | 0.77 | 5.26 | ≥ 70 | 0.89 | 3.24 |
| **Economic Status** | | | | | | |
| Economically Disadvantaged | ≥ 1,210 | 0.86 | 5.11 | ≥ 1,210 | 0.91 | 3.41 |
| Not Economically Disadvantaged | ≥ 680 | 0.88 | 5.34 | ≥ 680 | 0.91 | 3.62 |
| **Section 504 Status** | | | | | | |
| Non-Section 504 | ≥ 1,760 | 0.88 | 5.21 | ≥ 1,760 | 0.92 | 3.54 |
| Section 504 | ≥ 170 | 0.86 | 4.95 | ≥ 170 | 0.89 | 3.30 |

**Table 10.5 Grade 4 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** | ≥ 54,300 | 0.87 | 5.13 | ≥ 54,300 | 0.91 | 3.73 |
| **Gender** | | | | | | |
| Female | ≥ 26,590 | 0.87 | 5.21 | ≥ 26,600 | 0.91 | 3.76 |
| Male | ≥ 27,680 | 0.88 | 5.00 | ≥ 27,660 | 0.92 | 3.69 |
| **Ethnicity** | | | | | | |
| Hispanic/Latino | ≥ 3,750 | 0.88 | 5.13 | ≥ 3,740 | 0.91 | 3.72 |
| American Indian or Alaska Native | ≥ 340 | 0.85 | 5.30 | ≥ 340 | 0.90 | 3.75 |
| Asian | ≥ 790 | 0.89 | 5.44 | ≥ 790 | 0.92 | 3.58 |
| Black or African American | ≥ 24,140 | 0.85 | 5.07 | ≥ 24,140 | 0.90 | 3.63 |
| Native Hawaiian or Other Pacific | ≥ 50 | 0.88 | 5.34 | ≥ 50 | 0.92 | 3.77 |
| White | ≥ 23,710 | 0.86 | 5.21 | ≥ 23,720 | 0.90 | 3.71 |
| Two or More Races | ≥ 1,460 | 0.86 | 5.28 | ≥ 1,460 | 0.91 | 3.73 |
| **Education Classification** | | | | | | |
| Regular | ≥ 48,340 | 0.87 | 5.17 | ≥ 48,340 | 0.91 | 3.72 |
| Special | ≥ 5,950 | 0.86 | 4.65 | ≥ 5,950 | 0.89 | 3.48 |
| **LEP Status** | | | | | | |
| Non-LEP | ≥ 52,140 | 0.87 | 5.14 | ≥ 52,140 | 0.91 | 3.73 |
| LEP | ≥ 2,150 | 0.82 | 5.05 | ≥ 2,150 | 0.90 | 3.58 |
| **Economic Status** | | | | | | |
| Economically Disadvantaged | ≥ 38,860 | 0.86 | 5.09 | ≥ 38,830 | 0.90 | 3.70 |
| Not Economically Disadvantaged | ≥ 14,670 | 0.87 | 5.24 | ≥ 14,660 | 0.91 | 3.68 |
| **Section 504 Status** | | | | | | |
| Non-Section 504 | ≥ 48,690 | 0.88 | 5.16 | ≥ 48,690 | 0.91 | 3.73 |
| Section 504 | ≥ 5,600 | 0.84 | 4.84 | ≥ 5,600 | 0.89 | 3.55 |

**Table 10.6 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** | ≥ 53,300 | 0.88 | 5.12 | ≥ 53,310 | 0.90 | 3.59 |
| **Gender** | | | | | | |
| Female | ≥ 25,890 | 0.87 | 5.21 | ≥ 25,890 | 0.89 | 3.58 |
| Male | ≥ 27,410 | 0.88 | 4.98 | ≥ 27,410 | 0.90 | 3.58 |
| **Ethnicity** | | | | | | |
| Hispanic/Latino | ≥ 3,530 | 0.89 | 5.08 | ≥ 3,530 | 0.90 | 3.60 |
| American Indian or Alaska Native | ≥ 320 | 0.87 | 5.17 | ≥ 320 | 0.88 | 3.60 |
| Asian | ≥ 820 | 0.90 | 5.28 | ≥ 820 | 0.92 | 3.74 |
| Black or African American | ≥ 23,670 | 0.86 | 5.04 | ≥ 23,670 | 0.88 | 3.41 |
| Native Hawaiian or Other Pacific | ≥ 50 | 0.88 | 5.26 | ≥ 50 | 0.90 | 3.67 |
| White | ≥ 23,580 | 0.87 | 5.19 | ≥ 23,570 | 0.89 | 3.67 |
| Two or More Races | ≥ 1,320 | 0.87 | 5.28 | ≥ 1,320 | 0.89 | 3.60 |
| **Education Classification** | | | | | | |
| Regular | ≥ 47,420 | 0.87 | 5.14 | ≥ 47,430 | 0.90 | 3.60 |
| Special | ≥ 5,880 | 0.85 | 4.63 | ≥ 5,870 | 0.86 | 3.25 |
| **LEP Status** | | | | | | |
| Non-LEP | ≥ 51,640 | 0.88 | 5.13 | ≥ 51,650 | 0.90 | 3.59 |
| LEP | ≥ 1,660 | 0.83 | 4.89 | ≥ 1,660 | 0.87 | 3.36 |
| **Economic Status** | | | | | | |
| Economically Disadvantaged | ≥ 38,110 | 0.87 | 5.07 | ≥ 38,120 | 0.89 | 3.50 |
| Not Economically Disadvantaged | ≥ 14,860 | 0.86 | 5.23 | ≥ 14,850 | 0.89 | 3.69 |
| **Section 504 Status** | | | | | | |
| Non-Section 504 | ≥ 47,650 | 0.88 | 5.14 | ≥ 47,650 | 0.90 | 3.60 |
| Section 504 | ≥ 5,650 | 0.84 | 4.84 | ≥ 5,650 | 0.86 | 3.38 |

**Table 10.7 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** | ≥ 52,370 | 0.91 | 5.37 | ≥ 52,350 | 0.90 | 3.91 |
| **Gender** | | | | | | |
| Female | ≥ 25,380 | 0.91 | 5.42 | ≥ 25,380 | 0.90 | 3.95 |
| Male | ≥ 26,990 | 0.91 | 5.24 | ≥ 26,960 | 0.91 | 3.86 |
| **Ethnicity** | | | | | | |
| Hispanic/Latino | ≥ 3,300 | 0.92 | 5.33 | ≥ 3,300 | 0.91 | 3.86 |
| American Indian or Alaska Native | ≥ 350 | 0.90 | 5.37 | ≥ 350 | 0.90 | 3.87 |
| Asian | ≥ 800 | 0.93 | 5.50 | ≥ 800 | 0.92 | 4.43 |
| Black or African American | ≥ 23,130 | 0.89 | 5.27 | ≥ 23,110 | 0.88 | 3.53 |
| Native Hawaiian or Other Pacific | ≥ 30 | 0.93 | 5.53 | ≥ 30 | 0.92 | 3.86 |
| White | ≥ 23,610 | 0.90 | 5.47 | ≥ 23,600 | 0.90 | 4.13 |
| Two or More Races | ≥ 1,130 | 0.91 | 5.45 | ≥ 1,130 | 0.90 | 3.99 |
| **Education Classification** | | | | | | |
| Regular | ≥ 46,810 | 0.91 | 5.40 | ≥ 46,790 | 0.90 | 3.95 |
| Special | ≥ 5,560 | 0.87 | 4.70 | ≥ 5,550 | 0.87 | 3.11 |
| **LEP Status** | | | | | | |
| Non-LEP | ≥ 51,080 | 0.91 | 5.38 | ≥ 51,050 | 0.90 | 3.92 |
| LEP | ≥ 1,290 | 0.84 | 4.87 | ≥ 1,290 | 0.87 | 3.17 |
| **Economic Status** | | | | | | |
| Economically Disadvantaged | ≥ 37,040 | 0.90 | 5.30 | ≥ 37,020 | 0.89 | 3.70 |
| Not Economically Disadvantaged | ≥ 15,000 | 0.90 | 5.49 | ≥ 15,000 | 0.90 | 4.23 |
| **Section 504 Status** | | | | | | |
| Non-Section 504 | ≥ 46,540 | 0.91 | 5.39 | ≥ 46,510 | 0.90 | 3.95 |
| Section 504 | ≥ 5,830 | 0.88 | 5.14 | ≥ 5,830 | 0.88 | 3.46 |

**Table 10.8 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** | ≥ 51,930 | 0.89 | 6.06 | ≥ 51,800 | 0.91 | 3.88 |
| **Gender** | | | | | | |
| Female | ≥ 25,450 | 0.89 | 6.08 | ≥ 25,380 | 0.91 | 3.93 |
| Male | ≥ 26,480 | 0.89 | 5.91 | ≥ 26,420 | 0.91 | 3.81 |
| **Ethnicity** | | | | | | |
| Hispanic/Latino | ≥ 3,150 | 0.91 | 5.99 | ≥ 3,150 | 0.92 | 3.84 |
| American Indian or Alaska Native | ≥ 390 | 0.88 | 6.02 | ≥ 390 | 0.90 | 3.87 |
| Asian | ≥ 850 | 0.90 | 6.27 | ≥ 840 | 0.93 | 4.06 |
| Black or African American | ≥ 23,040 | 0.87 | 6.00 | ≥ 23,000 | 0.89 | 3.63 |
| Native Hawaiian or Other Pacific | ≥ 40 | 0.91 | 6.10 | ≥ 40 | 0.93 | 4.05 |
| White | ≥ 23,460 | 0.87 | 6.17 | ≥ 23,370 | 0.91 | 3.97 |
| Two or More Races | ≥ 970 | 0.88 | 6.08 | ≥ 960 | 0.91 | 3.93 |
| **Education Classification** | | | | | | |
| Regular | ≥ 46,730 | 0.88 | 6.08 | ≥ 46,600 | 0.91 | 3.90 |
| Special | ≥ 5,200 | 0.85 | 5.36 | ≥ 5,200 | 0.87 | 3.12 |
| **LEP Status** | | | | | | |
| Non-LEP | ≥ 50,650 | 0.89 | 6.07 | ≥ 50,510 | 0.91 | 3.88 |
| LEP | ≥ 1,280 | 0.84 | 5.36 | ≥ 1,280 | 0.87 | 3.23 |
| **Economic Status** | | | | | | |
| Economically Disadvantaged | ≥ 36,240 | 0.88 | 6.01 | ≥ 36,180 | 0.90 | 3.74 |
| Not Economically Disadvantaged | ≥ 15,340 | 0.87 | 6.20 | ≥ 15,260 | 0.91 | 4.01 |
| **Section 504 Status** | | | | | | |
| Non-Section 504 | ≥ 46,380 | 0.89 | 6.08 | ≥ 46,240 | 0.91 | 3.90 |
| Section 504 | ≥ 5,550 | 0.86 | 5.75 | ≥ 5,550 | 0.89 | 3.52 |

**Table 10.9 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** | ≥ 50,450 | 0.87 | 6.37 | ≥ 44,710 | 0.89 | 3.74 |
| **Gender** | | | | | | |
| Female | ≥ 24,610 | 0.86 | 6.36 | ≥ 21,610 | 0.89 | 3.82 |
| Male | ≥ 25,830 | 0.88 | 6.21 | ≥ 23,090 | 0.90 | 3.66 |
| **Ethnicity** | | | | | | |
| Hispanic/Latino | ≥ 2,960 | 0.90 | 6.26 | ≥ 2,680 | 0.89 | 3.67 |
| American Indian or Alaska Native | ≥ 350 | 0.85 | 6.21 | ≥ 320 | 0.89 | 3.87 |
| Asian | ≥ 820 | 0.88 | 6.41 | ≥ 570 | 0.92 | 4.17 |
| Black or African American | ≥ 22,270 | 0.85 | 6.25 | ≥ 21,080 | 0.87 | 3.44 |
| Native Hawaiian or Other Pacific | ≥ 40 | 0.82 | 6.61 | ≥ 30 | 0.89 | 4.10 |
| White | ≥ 23,130 | 0.86 | 6.49 | ≥ 19,250 | 0.89 | 3.95 |
| Two or More Races | ≥ 860 | 0.86 | 6.51 | ≥ 760 | 0.89 | 3.79 |
| **Education Classification** | | | | | | |
| Regular | ≥ 45,820 | 0.86 | 6.39 | ≥ 40,170 | 0.89 | 3.80 |
| Special | ≥ 4,630 | 0.84 | 5.64 | ≥ 4,540 | 0.84 | 2.97 |
| **LEP Status** | | | | | | |
| Non-LEP | ≥ 49,250 | 0.87 | 6.38 | ≥ 43,530 | 0.89 | 3.76 |
| LEP | ≥ 1,200 | 0.83 | 5.52 | ≥ 1,180 | 0.88 | 3.05 |
| **Economic Status** | | | | | | |
| Economically Disadvantaged | ≥ 34,850 | 0.86 | 6.30 | ≥ 32,480 | 0.88 | 3.59 |
| Not Economically Disadvantaged | ≥ 15,290 | 0.85 | 6.49 | ≥ 11,930 | 0.89 | 4.04 |
| **Section 504 Status** | | | | | | |
| Non-Section 504 | ≥ 45,360 | 0.87 | 6.39 | ≥ 39,820 | 0.89 | 3.78 |
| Section 504 | ≥ 5,080 | 0.84 | 6.12 | ≥ 4,890 | 0.87 | 3.36 |

### 10.3.2 Effect Size

One way to evaluate the magnitude of the SMD is to calculate the ES. Cohen's *d* was used 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 10.10–10.16. Results are NR for subgroups with fewer than ten students. On the ELA test in most grades, there are medium differences in mean test scores between females and males where females outperform males. Although there were no ESs larger than a small ES, |0.20|, for mathematics, females tend to perform better than males in general. 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 and special education students in education classification, between not economically disadvantaged and economically disadvantaged in economic status, non-LEP and LEP students in LEP status, and non-migrant and migrant students in migrant status.

**Table 10.10 Impact Analysis, Grade 3 Paper-Based Test Administration**

| | ELA | | | | Mathematics | | | |
|---|---|---|---|---|---|---|---|---|
| **Group** | **N** | **Scale Score Mean** | **Scale Score Std. Dev.** | **Effect Size** | **N** | **Scale Score Mean** | **Scale Score Std. Dev.** | **Effect Size** |
| **All Students** | ≥ 56,800 | 743.41 | 38.70 | | ≥ 56,800 | 742.52 | 33.28 | |
| **Gender** | | | | | | | | |
| Male | ≥ 28,920 | 738.62 | 37.81 | | ≥ 28,920 | 741.25 | 33.72 | |
| Female | ≥ 27,840 | 748.42 | 38.97 | -0.26 | ≥ 27,840 | 743.85 | 32.76 | -0.08 |
| **Ethnicity** | | | | | | | | |
| White | ≥ 24,440 | 755.92 | 37.12 | | ≥ 24,430 | 753.59 | 31.75 | |
| Hispanic/Latino | ≥ 4,230 | 736.38 | 40.35 | 0.52 | ≥ 4,220 | 741.34 | 32.78 | 0.38 |
| American Indian or Alaska Native | ≥ 320 | 747.55 | 38.61 | 0.23 | ≥ 3280 | 746.53 | 34.00 | 0.22 |
| Asian | ≥ 890 | 762.79 | 44.11 | -0.18 | ≥ 890 | 767.95 | 34.86 | -0.45 |
| Black or African American | ≥ 25,150 | 731.35 | 35.53 | 0.68 | ≥ 25,150 | 730.86 | 30.47 | 0.73 |
| Native Hawaiian or Other Pacific | ≥ 70 | 754.79 | 44.83 | 0.03 | ≥ 70 | 756.95 | 40.83 | -0.11 |
| Two or More Races | ≥ 1,630 | 748.88 | 36.69 | 0.19 | ≥ 1,630 | 744.81 | 32.26 | 0.28 |
| **Education Classification** | | | | | | | | |
| Regular | ≥ 50,450 | 746.17 | 38.18 | | ≥ 50,440 | 744.92 | 32.58 | |
| Special | ≥ 6,350 | 721.57 | 35.72 | 0.65 | ≥ 6,350 | 723.47 | 32.64 | 0.66 |
| **Economic Status** | | | | | | | | |
| Not Economically Disadvantaged | ≥ 14,690 | 762.73 | 37.27 | | ≥ 14,700 | 759.28 | 31.89 | |
| Economically Disadvantaged | ≥ 41,240 | 736.84 | 36.80 | 0.70 | ≥ 41,200 | 736.89 | 31.66 | 0.71 |
| **LEP Status** | | | | | | | | |
| Non-LEP | ≥ 54,100 | 744.53 | 38.49 | | ≥ 54,110 | 742.97 | 33.31 | |
| LEP | ≥ 2,690 | 721.00 | 35.99 | 0.61 | ≥ 2,690 | 733.47 | 31.30 | 0.29 |
| **Migrant Status** | | | | | | | | |
| Nonmigrant | ≥ 56,660 | 743.48 | 38.68 | | ≥ 56,660 | 742.56 | 33.27 | |
| Migrant | ≥ 130 | 717.21 | 38.23 | 0.68 | ≥ 130 | 722.91 | 33.29 | 0.59 |
| **Section 504 Status** | | | | | | | | |
| Non-Section 504 | ≥ 51,890 | 744.99 | 38.81 | | ≥ 51,890 | 743.94 | 33.25 | |
| Section 504 | ≥ 4,900 | 726.80 | 33.20 | 0.47 | ≥ 4,900 | 727.52 | 29.67 | 0.50 |

**Table 10.11 Impact Analysis, Grade 4 Computer-Based Test Administration**

| | ELA | | | | Mathematics | | | |
|---|---|---|---|---|---|---|---|---|
| **Group** | **N** | **Scale Score Mean** | **Scale Score Std. Dev.** | **Effect Size** | **N** | **Scale Score Mean** | **Scale Score Std. Dev.** | **Effect Size** |
| **All Students** | ≥ 1,930 | 740.53 | 34.38 | | ≥ 1,930 | 736.56 | 29.63 | |
| **Gender** | | | | | | | | |
| Male | ≥ 970 | 735.89 | 33.50 | | ≥ 970 | 736.44 | 29.90 | |
| Female | ≥ 960 | 745.23 | 34.64 | -0.27 | ≥ 960 | 736.68 | 29.36 | -0.01 |
| **Ethnicity** | | | | | | | | |
| White | ≥ 1,140 | 748.25 | 33.35 | | ≥ 1,130 | 742.63 | 29.19 | |
| Hispanic/Latino | ≥ 150 | 735.68 | 32.26 | 0.38 | ≥ 150 | 731.18 | 27.40 | 0.40 |
| American Indian or Alaska Native | < 10 | NR | NR | NR | < 10 | NR | NR | NR |
| Asian | ≥ 20 | 760.58 | 31.53 | -0.37 | ≥ 20 | 761.88 | 32.39 | -0.66 |
| Black or African American | ≥ 540 | 723.64 | 30.62 | 0.76 | ≥ 540 | 723.86 | 26.17 | 0.66 |
| Native Hawaiian or Other Pacific | < 10 | NR | NR | NR | < 10 | NR | NR | NR |
| Two or More Races | ≥ 60 | 751.31 | 32.92 | -0.09 | ≥ 60 | 742.92 | 29.50 | -0.01 |
| **Education Classification** | | | | | | | | |
| Regular | ≥ 1,700 | 744.08 | 33.13 | | ≥ 1,700 | 738.80 | 29.43 | |
| Special | ≥ 230 | 714.76 | 32.30 | 0.89 | ≥ 230 | 720.31 | 25.82 | 0.64 |
| **Economic Status** | | | | | | | | |
| Not Economically Disadvantaged | ≥ 680 | 753.85 | 33.90 | | ≥ 680 | 748.50 | 29.39 | |
| Economically Disadvantaged | ≥ 1,210 | 732.77 | 32.20 | 0.64 | ≥ 1,210 | 729.58 | 27.49 | 0.67 |
| **LEP Status** | | | | | | | | |
| Non-LEP | ≥ 1,860 | 741.38 | 34.37 | | ≥ 1,860 | 737.21 | 29.66 | |
| LEP | ≥ 70 | 719.72 | 27.62 | 0.63 | ≥ 70 | 720.47 | 24.00 | 0.57 |
| **Migrant Status** | | | | | | | | |
| Nonmigrant | ≥ 1,930 | 740.57 | 34.41 | | ≥ 1,930 | 736.52 | 29.64 | |
| Migrant | < 10 | NR | NR | NR | < 10 | NR | NR | NR |
| **Section 504 Status** | | | | | | | | |
| Non-Section 504 | ≥ 1,760 | 742.13 | 34.20 | | ≥ 1760 | 737.98 | 29.66 | |
| Section 504 | ≥ 170 | 724.25 | 31.91 | 0.53 | ≥ 170 | 722.05 | 25.16 | 0.54 |

**Table 10.12 Impact Analysis, Grade 4 Paper-Based Test Administration**

| | ELA | | | | Mathematics | | | |
|---|---|---|---|---|---|---|---|---|
| **Group** | **N** | **Scale Score Mean** | **Scale Score Std. Dev.** | **Effect Size** | **N** | **Scale Score Mean** | **Scale Score Std. Dev.** | **Effect Size** |
| **All Students** | ≥ 54,300 | 744.31 | 31.85 | | ≥ 54,300 | 740.31 | 30.70 | |
| **Gender** | | | | | | | | |
| Male | ≥ 27,680 | 740.08 | 31.55 | | ≥ 27,660 | 739.55 | 30.98 | |
| Female | ≥ 26,500 | 748.73 | 31.54 | -0.27 | ≥ 26,600 | 741.13 | 30.38 | -0.05 |
| **Ethnicity** | | | | | | | | |
| White | ≥ 23,710 | 755.06 | 30.05 | | ≥ 23,720 | 750.85 | 29.17 | |
| Hispanic/Latino | ≥ 3,750 | 737.40 | 33.47 | 0.58 | ≥ 3,740 | 737.23 | 30.26 | 0.46 |
| American Indian or Alaska Native | ≥ 340 | 745.61 | 30.23 | 0.31 | ≥ 340 | 740.73 | 27.53 | 0.35 |
| Asian | ≥ 790 | 762.31 | 35.86 | -0.24 | ≥ 790 | 765.91 | 33.27 | -0.51 |
| Black or African American | ≥ 24,140 | 733.95 | 29.36 | 0.71 | ≥ 24,140 | 729.44 | 28.02 | 0.75 |
| Native Hawaiian or Other Pacific | ≥ 50 | 744.96 | 34.56 | 0.34 | ≥ 50 | 742.61 | 31.42 | 0.28 |
| Two or More Races | ≥ 1,460 | 749.08 | 30.90 | 0.20 | ≥ 1,460 | 743.29 | 30.01 | 0.26 |
| **Education Classification** | | | | | | | | |
| Regular | ≥ 48,340 | 746.80 | 31.28 | | ≥ 48,340 | 742.47 | 30.46 | |
| Special | ≥ 5,950 | 724.10 | 29.07 | 0.73 | ≥ 5,950 | 722.79 | 26.77 | 0.65 |
| **Economic Status** | | | | | | | | |
| Not Economically Disadvantaged | ≥ 14,670 | 760.42 | 30.14 | | ≥ 14,660 | 755.49 | 29.77 | |
| Economically Disadvantaged | ≥ 38,860 | 738.53 | 30.27 | 0.72 | ≥ 38,830 | 734.91 | 29.04 | 0.70 |
| **LEP Status** | | | | | | | | |
| Non-LEP | ≥ 52,140 | 745.19 | 31.62 | | ≥ 52,140 | 740.81 | 30.67 | |
| LEP | ≥ 2,150 | 722.97 | 29.91 | 0.70 | ≥ 2,150 | 728.25 | 28.80 | 0.41 |
| **Migrant Status** | | | | | | | | |
| Nonmigrant | ≥ 54,180 | 744.34 | 31.83 | | ≥ 54,180 | 740.33 | 30.70 | |
| Migrant | ≥ 110 | 727.27 | 34.11 | 0.54 | ≥ 110 | 734.33 | 28.68 | 0.20 |
| **Section 504 Status** | | | | | | | | |
| Non-Section 504 | ≥ 48,690 | 745.75 | 31.97 | | ≥ 48,690 | 741.87 | 30.76 | |
| Section 504 | ≥ 5,600 | 731.77 | 27.75 | 0.44 | ≥ 5,600 | 726.78 | 26.59 | 0.50 |

**Table 10.13 Impact Analysis, Grade 5 Computer-Based Test Administration**

| | ELA | | | | Mathematics | | | |
|---|---|---|---|---|---|---|---|---|
| **Group** | **N** | **Scale Score Mean** | **Scale Score Std. Dev.** | **Effect Size** | **N** | **Scale Score Mean** | **Scale Score Std. Dev.** | **Effect Size** |
| **All Students** | ≥ 53,300 | 742.37 | 30.77 | | ≥ 53,310 | 736.04 | 29.17 | |
| **Gender** | | | | | | | | |
| Male | ≥ 27,410 | 738.62 | 30.58 | | ≥ 27,410 | 734.65 | 29.86 | |
| Female | ≥ 25,890 | 746.34 | 30.47 | -0.25 | ≥ 25,890 | 737.51 | 28.34 | -0.10 |
| **Ethnicity** | | | | | | | | |
| White | ≥ 23,580 | 751.16 | 29.18 | | ≥ 23,570 | 745.50 | 28.07 | |
| Hispanic/Latino | ≥ 3,530 | 738.62 | 33.01 | 0.42 | ≥ 3,530 | 734.01 | 29.54 | 0.41 |
| American Indian or Alaska Native | ≥ 320 | 746.00 | 29.42 | 0.18 | ≥ 320 | 736.14 | 26.40 | 0.33 |
| Asian | ≥ 820 | 761.28 | 35.99 | -0.34 | ≥ 820 | 761.78 | 35.20 | -0.57 |
| Black or African American | ≥ 23,670 | 733.23 | 28.94 | 0.62 | ≥ 23,670 | 725.91 | 26.25 | 0.72 |
| Native Hawaiian or Other Pacific | ≥ 50 | 745.72 | 32.17 | 0.19 | ≥ 50 | 739.85 | 28.44 | 0.20 |
| Two or More Races | ≥ 1,320 | 746.46 | 29.61 | 0.16 | ≥ 1,320 | 738.12 | 27.39 | 0.26 |
| **Education Classification** | | | | | | | | |
| Regular | ≥ 47,420 | 745.77 | 29.32 | | ≥ 47,430 | 738.39 | 28.91 | |
| Special | ≥ 5,880 | 714.89 | 28.33 | 1.06 | ≥ 5,870 | 717.07 | 23.86 | 0.75 |
| **Economic Status** | | | | | | | | |
| Not Economically Disadvantaged | ≥ 14,860 | 756.95 | 28.81 | | ≥ 14,850 | 750.67 | 28.66 | |
| Economically Disadvantaged | ≥ 38,110 | 736.80 | 29.63 | 0.69 | ≥ 38,120 | 730.46 | 27.32 | 0.73 |
| **LEP Status** | | | | | | | | |
| Non-LEP | ≥ 51,640 | 743.16 | 30.53 | | ≥ 51,650 | 736.52 | 29.12 | |
| LEP | ≥ 1,660 | 717.61 | 27.90 | 0.84 | ≥ 1,660 | 720.95 | 26.42 | 0.54 |
| **Migrant Status** | | | | | | | | |
| Nonmigrant | ≥ 53,200 | 742.39 | 30.77 | | ≥ 53,210 | 736.05 | 29.17 | |
| Migrant | ≥ 100 | 732.41 | 31.15 | 0.32 | ≥ 100 | 727.70 | 26.77 | 0.29 |
| **Section 504 Status** | | | | | | | | |
| Non-Section 504 | ≥ 47,650 | 744.10 | 30.78 | | ≥ 47,650 | 737.47 | 29.35 | |
| Section 504 | ≥ 5,650 | 727.78 | 26.55 | 0.54 | ≥ 5,650 | 723.96 | 24.42 | 0.47 |

**Table 10.14 Impact Analysis, Grade 6 Computer-Based Test Administration**

| | ELA | | | | Mathematics | | | |
|---|---|---|---|---|---|---|---|---|
| Group | N | Scale Score Mean | Scale Score Std. Dev. | Effect Size | N | Scale Score Mean | Scale Score Std. Dev. | Effect Size |
| **All Students** | ≥ 52,370 | 737.88 | 31.02 | | ≥ 52,350 | 731.91 | 29.28 | |
| **Gender** | | | | | | | | |
| Male | ≥ 26,990 | 733.06 | 30.13 | | ≥ 26,960 | 730.25 | 29.92 | |
| Female | ≥ 25,380 | 743.00 | 31.14 | -0.32 | ≥ 25,380 | 733.69 | 28.48 | -0.12 |
| **Ethnicity** | | | | | | | | |
| White | ≥ 23,610 | 748.19 | 29.78 | | ≥ 23,600 | 741.13 | 28.47 | |
| Hispanic/Latino | ≥ 3,300 | 732.67 | 33.16 | 0.51 | ≥ 3,300 | 728.51 | 30.45 | 0.44 |
| American Indian or Alaska Native | ≥ 350 | 740.56 | 28.55 | 0.26 | ≥ 350 | 731.41 | 28.03 | 0.34 |
| Asian | ≥ 800 | 757.84 | 37.53 | -0.32 | ≥ 800 | 759.42 | 36.07 | -0.64 |
| Black or African American | ≥ 23,130 | 727.06 | 27.64 | 0.74 | ≥ 23,110 | 721.83 | 25.80 | 0.71 |
| Native Hawaiian or Other Pacific | ≥ 30 | 738.54 | 37.09 | 0.32 | ≥ 30 | 730.21 | 33.31 | 0.38 |
| Two or More Races | ≥ 1,130 | 744.07 | 30.07 | 0.14 | ≥ 1,130 | 736.30 | 28.32 | 0.17 |
| **Education Classification** | | | | | | | | |
| Regular | ≥ 46,810 | 741.19 | 29.98 | | ≥ 46,790 | 734.73 | 28.57 | |
| Special | ≥ 5,560 | 710.00 | 25.02 | 1.06 | ≥ 5,550 | 708.21 | 24.02 | 0.94 |
| **Economic Status** | | | | | | | | |
| Not Economically Disadvantaged | ≥ 15,000 | 754.16 | 30.16 | | ≥ 15,000 | 747.35 | 28.98 | |
| Economically Disadvantaged | ≥ 37,040 | 731.40 | 28.88 | 0.78 | ≥ 37,020 | 725.80 | 26.99 | 0.78 |
| **LEP Status** | | | | | | | | |
| Non-LEP | ≥ 51,080 | 738.69 | 30.75 | | ≥ 51,050 | 732.51 | 29.12 | |
| LEP | ≥ 1,290 | 706.03 | 24.13 | 1.07 | ≥ 1,290 | 708.53 | 25.61 | 0.83 |
| **Migrant Status** | | | | | | | | |
| Nonmigrant | ≥ 52,290 | 737.90 | 31.02 | | ≥ 52,260 | 731.92 | 29.28 | |
| Migrant | ≥ 80 | 728.42 | 29.96 | 0.31 | ≥ 80 | 726.08 | 25.91 | 0.20 |
| **Section 504 Status** | | | | | | | | |
| Non-Section 504 | ≥ 46,540 | 739.65 | 31.21 | | ≥ 46,510 | 733.40 | 29.48 | |
| Section 504 | ≥ 5,830 | 723.80 | 25.47 | 0.52 | ≥ 5,830 | 720.03 | 24.55 | 0.46 |

**Table 10.15 Impact Analysis, Grade 7 Computer-Based Test Administration**

| | ELA | | | | Mathematics | | | |
|---|---|---|---|---|---|---|---|---|
| Group | N | Scale Score Mean | Scale Score Std. Dev. | Effect Size | N | Scale Score Mean | Scale Score Std. Dev. | Effect Size |
| **All Students** | ≥ 51,930 | 740.64 | 36.84 | | ≥ 51,800 | 731.92 | 26.03 | |
| **Gender** | | | | | | | | |
| Male | ≥ 26,480 | 734.44 | 36.43 | | ≥ 26,420 | 730.16 | 26.14 | |
| Female | ≥ 25,450 | 747.08 | 36.15 | -0.35 | ≥ 25,380 | 733.76 | 25.80 | -0.14 |
| **Ethnicity** | | | | | | | | |
| White | ≥ 23,460 | 752.10 | 34.69 | | ≥ 23,370 | 740.43 | 25.23 | |
| Hispanic/Latino | ≥ 3,150 | 731.17 | 42.15 | 0.59 | ≥ 3,150 | 728.88 | 27.00 | 0.45 |
| American Indian or Alaska Native | ≥ 390 | 742.14 | 34.73 | 0.29 | ≥ 390 | 732.57 | 24.46 | 0.31 |
| Asian | ≥ 850 | 762.57 | 41.73 | -0.30 | ≥ 840 | 755.17 | 29.59 | -0.58 |
| Black or African American | ≥ 23,040 | 729.11 | 33.92 | 0.67 | ≥ 23,000 | 722.66 | 22.96 | 0.74 |
| Native Hawaiian or Other Pacific | ≥ 40 | 745.79 | 42.97 | 0.18 | ≥ 40 | 737.38 | 31.09 | 0.12 |
| Two or More Races | ≥ 970 | 747.83 | 34.81 | 0.12 | ≥ 960 | 735.51 | 25.27 | 0.20 |
| **Education Classification** | | | | | | | | |
| Regular | ≥ 46,730 | 744.70 | 35.06 | | ≥ 46,600 | 734.46 | 25.29 | |
| Special | ≥ 5,200 | 704.13 | 31.98 | 1.17 | ≥ 5,200 | 709.21 | 21.16 | 1.01 |
| **Economic Status** | | | | | | | | |
| Not Economically Disadvantaged | ≥ 15,340 | 759.27 | 33.72 | | ≥ 15,260 | 745.57 | 25.63 | |
| Economically Disadvantaged | ≥ 36,240 | 732.94 | 35.16 | 0.76 | ≥ 36,180 | 726.29 | 23.98 | 0.79 |
| **LEP Status** | | | | | | | | |
| Non-LEP | ≥ 50,650 | 741.80 | 36.20 | | ≥ 50,510 | 732.46 | 25.91 | |
| LEP | ≥ 1,280 | 694.69 | 31.88 | 1.30 | ≥ 1,280 | 710.82 | 21.73 | 0.84 |
| **Migrant Status** | | | | | | | | |
| Nonmigrant | ≥ 51,840 | 740.66 | 36.82 | | ≥ 51,710 | 731.93 | 26.03 | |
| Migrant | ≥ 80 | 724.99 | 41.25 | 0.43 | ≥ 80 | 725.46 | 26.06 | 0.25 |
| **Section 504 Status** | | | | | | | | |
| Non-Section 504 | ≥ 46,380 | 742.73 | 36.81 | | ≥ 46,240 | 733.35 | 26.10 | |
| Section 504 | ≥ 5,550 | 723.16 | 32.17 | 0.54 | ≥ 5,550 | 720.02 | 22.14 | 0.52 |

**Table 10.16 Impact Analysis, Grade 8 Computer-Based Test Administration**

| | ELA | | | | Mathematics | | | |
|---|---|---|---|---|---|---|---|---|
| Group | N | Scale Score Mean | Scale Score Std. Dev. | Effect Size | N | Scale Score Mean | Scale Score Std. Dev. | Effect Size |
| **All Students** | ≥ 50,450 | 744.26 | 37.52 | | ≥ 44,710 | 728.40 | 33.09 | |
| **Gender** | | | | | | | | |
| Male | ≥ 25,830 | 736.44 | 37.35 | | ≥ 23,090 | 726.13 | 33.52 | |
| Female | ≥ 24,610 | 752.45 | 35.92 | -0.44 | ≥ 21,610 | 730.83 | 32.45 | -0.14 |
| **Ethnicity** | | | | | | | | |
| White | ≥ 23,130 | 755.56 | 36.03 | | ≥ 19,250 | 738.88 | 32.17 | |
| Hispanic/Latino | ≥ 2,960 | 731.96 | 43.41 | 0.64 | ≥ 2,680 | 722.20 | 34.60 | 0.51 |
| American Indian or Alaska Native | ≥ 350 | 747.17 | 33.60 | 0.23 | ≥ 320 | 733.21 | 31.91 | 0.18 |
| Asian | ≥ 820 | 770.12 | 41.40 | -0.40 | ≥ 570 | 759.18 | 41.19 | -0.63 |
| Black or African American | ≥ 22,270 | 732.90 | 33.87 | 0.65 | ≥ 21,080 | 718.56 | 29.92 | 0.66 |
| Native Hawaiian or Other Pacific | ≥ 40 | 756.44 | 31.37 | -0.02 | ≥ 30 | 748.42 | 31.87 | -0.30 |
| Two or More Races | ≥ 860 | 749.97 | 36.58 | 0.15 | ≥ 760 | 731.65 | 32.28 | 0.22 |
| **Education Classification** | | | | | | | | |
| Regular | ≥ 45,820 | 747.93 | 36.05 | | ≥ 40,170 | 731.58 | 32.21 | |
| Special | ≥ 4,630 | 707.90 | 31.91 | 1.12 | ≥ 4,540 | 700.30 | 26.93 | 0.99 |
| **Economic Status** | | | | | | | | |
| Not Economically Disadvantaged | ≥ 15,290 | 762.13 | 35.10 | | ≥ 11,930 | 743.88 | 32.62 | |
| Economically Disadvantaged | ≥ 34,850 | 736.64 | 35.73 | 0.72 | ≥ 32,480 | 722.82 | 31.37 | 0.66 |
| **LEP Status** | | | | | | | | |
| Non-LEP | ≥ 49,250 | 745.46 | 36.80 | | ≥ 43,530 | 729.11 | 32.87 | |
| LEP | ≥ 1,20 | 694.82 | 32.92 | 1.38 | ≥ 1,180 | 702.34 | 30.54 | 0.82 |
| **Migrant Status** | | | | | | | | |
| Nonmigrant | ≥ 50,370 | 744.28 | 37.51 | | ≥ 44,640 | 728.41 | 33.08 | |
| Migrant | ≥ 70 | 727.43 | 38.93 | 0.45 | ≥ 70 | 726.64 | 37.25 | 0.05 |
| **Section 504 Status** | | | | | | | | |
| Non-Section 504 | ≥ 45,360 | 746.26 | 37.50 | | ≥ 39,820 | 730.09 | 33.15 | |
| Section 504 | ≥ 5,080 | 726.36 | 32.62 | 0.54 | ≥ 4,890 | 714.67 | 29.18 | 0.47 |

Additional data for mean scale scores are provided in Tables 10.17 and 10.18. 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.

**Table 10.17 Special Education Classification Scale-Score Means and Standard Deviations: English Language Arts**

| Special Education Classification Scale-Score Means and Standard Deviations: English Language Arts | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Yes | | | No | | |
| Grade | Group | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| 3 | Gifted | ≥ 1,080 | 803.72 | 28.16 | ≥ 55,710 | 742.24 | 37.94 |
| | Talented | ≥ 510 | 783.22 | 33.06 | ≥ 56,280 | 743.05 | 38.56 |
| | Autism | ≥ 310 | 714.43 | 37.56 | ≥ 56,490 | 743.57 | 38.64 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 56,800 | 743.41 | 38.70 |
| | Developmental Delay | ≥ 480 | 712.28 | 31.37 | ≥ 56,320 | 743.68 | 38.65 |
| | Emotional Disturbance | ≥ 90 | 716.67 | 31.02 | ≥ 56,710 | 743.46 | 38.69 |
| | HI—Deaf | < 50 | NR | NR | ≥ 56,780 | 743.43 | 38.69 |
| | HI—Hard-of-Hearing | ≥ 50 | 714.75 | 38.53 | ≥ 56,740 | 743.44 | 38.69 |
| | Mild Mental Disability | ≥ 340 | 694.12 | 22.67 | ≥ 56,450 | 743.72 | 38.58 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 56,780 | 743.43 | 38.69 |
| | Orthopedic Impairment | ≥ 70 | 737.36 | 36.85 | ≥ 56,720 | 743.42 | 38.70 |
| | Other Health Impairment | ≥ 900 | 717.79 | 32.85 | ≥ 55,890 | 743.83 | 38.65 |
| | Specific Learning Disability | ≥ 2,160 | 714.83 | 28.89 | ≥ 54,630 | 744.55 | 38.60 |
| | Speech or Language Impairment | ≥ 1,810 | 740.65 | 38.37 | ≥ 54,980 | 743.51 | 38.71 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 56,790 | 743.42 | 38.70 |
| | Visual Impairment | < 50 | NR | NR | ≥ 56,760 | 743.41 | 38.70 |
| | Other | < 50 | NR | NR | ≥ 56,800 | 743.42 | 38.70 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 56,800 | 743.41 | 38.70 |
| | Unknown | < 50 | NR | NR | ≥ 56,800 | 743.41 | 38.70 |
| 4 | Gifted | ≥ 1,420 | 792.01 | 23.17 | ≥ 54,810 | 742.94 | 31.18 |
| | Talented | ≥ 860 | 769.88 | 26.77 | ≥ 55,370 | 743.78 | 31.85 |
| | Autism | ≥ 290 | 721.76 | 30.37 | ≥ 55,940 | 744.30 | 31.91 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 56,230 | 744.18 | 31.94 |
| | Developmental Delay | < 50 | NR | NR | ≥ 56,210 | 744.19 | 31.94 |
| | Emotional Disturbance | ≥ 140 | 717.46 | 29.16 | ≥ 56,090 | 744.25 | 31.92 |
| | HI—Deaf | < 50 | NR | NR | ≥ 56,220 | 744.19 | 31.94 |
| | HI—Hard-of-Hearing | ≥ 60 | 725.76 | 28.38 | ≥ 56,170 | 744.20 | 31.94 |
| | Mild Mental Disability | ≥ 350 | 701.47 | 19.69 | ≥ 55,880 | 744.45 | 31.82 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 56,230 | 744.18 | 31.94 |
| | Orthopedic Impairment | ≥ 60 | 732.74 | 32.57 | ≥ 56,170 | 744.19 | 31.94 |
| | Other Health Impairment | ≥ 1,020 | 722.16 | 27.19 | ≥ 55,210 | 744.59 | 31.88 |
| | Specific Learning Disability | ≥ 2,680 | 718.99 | 24.39 | ≥ 53,550 | 745.44 | 31.75 |
| | Speech or Language Impairment | ≥ 1,430 | 740.28 | 32.68 | ≥ 54,800 | 744.28 | 31.92 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 56,230 | 744.18 | 31.94 |
| | Visual Impairment | < 50 | NR | NR | ≥ 56,190 | 744.18 | 31.94 |
| | Other | < 50 | NR | NR | ≥ 56,230 | 744.18 | 31.94 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 56,230 | 744.18 | 31.94 |
| | Unknown | < 50 | NR | NR | ≥ 56,230 | 744.18 | 31.94 |

| Special Education Classification Scale-Score Means and Standard Deviations: English Language Arts (continued) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Yes | | | No | | |
| Grade | Group | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| 5 | Gifted | ≥ 1,520 | 788.15 | 23.09 | ≥ 51,780 | 741.02 | 29.93 |
| | Talented | ≥ 1,120 | 764.22 | 25.33 | ≥ 52,180 | 741.90 | 30.71 |
| | Autism | ≥ 270 | 716.11 | 31.62 | ≥ 53,030 | 742.50 | 30.71 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 53,300 | 742.37 | 30.77 |
| | Developmental Delay | ≥ 60 | 700.38 | 27.03 | ≥ 53,240 | 742.42 | 30.74 |
| | Emotional Disturbance | ≥ 160 | 713.23 | 26.24 | ≥ 53,140 | 742.46 | 30.74 |
| | HI—Deaf | < 50 | NR | NR | ≥ 53,280 | 742.38 | 30.76 |
| | HI—Hard-of-Hearing | ≥ 60 | 726.78 | 32.89 | ≥ 53,240 | 742.39 | 30.76 |
| | Mild Mental Disability | ≥ 310 | 690.18 | 17.44 | ≥ 52,990 | 742.68 | 30.56 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 53,300 | 742.37 | 30.77 |
| | Orthopedic Impairment | ≥ 60 | 737.64 | 32.97 | ≥ 53,240 | 742.37 | 30.77 |
| | Other Health Impairment | ≥ 1,080 | 715.64 | 26.60 | ≥ 52,220 | 742.92 | 30.60 |
| | Specific Learning Disability | ≥ 2,780 | 708.90 | 22.83 | ≥ 50,520 | 744.21 | 30.08 |
| | Speech or Language Impairment | ≥ 980 | 737.34 | 30.65 | ≥ 52,320 | 742.46 | 30.76 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 53,290 | 742.37 | 30.77 |
| | Visual Impairment | < 50 | NR | NR | ≥ 53,270 | 742.38 | 30.77 |
| | Other | < 50 | NR | NR | ≥ 53,300 | 742.37 | 30.77 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 53,300 | 742.37 | 30.77 |
| | Unknown | < 50 | NR | NR | ≥ 53,300 | 742.37 | 30.77 |
| 6 | Gifted | ≥ 1,610 | 787.56 | 25.17 | ≥ 50,760 | 736.30 | 29.87 |
| | Talented | ≥ 1,340 | 762.51 | 28.43 | ≥ 51,030 | 737.23 | 30.82 |
| | Autism | ≥ 270 | 716.27 | 31.07 | ≥ 52,100 | 738.00 | 30.98 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 52,370 | 737.88 | 31.02 |
| | Developmental Delay | < 50 | NR | NR | ≥ 52,340 | 737.91 | 31.01 |
| | Emotional Disturbance | ≥ 200 | 711.69 | 26.27 | ≥ 52,170 | 737.98 | 31.00 |
| | HI—Deaf | < 50 | NR | NR | ≥ 52,360 | 737.89 | 31.02 |
| | HI—Hard-of-Hearing | ≥ 60 | 722.45 | 24.82 | ≥ 52,310 | 737.90 | 31.03 |
| | Mild Mental Disability | ≥ 250 | 689.05 | 14.95 | ≥ 52,120 | 738.12 | 30.89 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 52,370 | 737.88 | 31.02 |
| | Orthopedic Impairment | ≥ 50 | 725.56 | 32.71 | ≥ 52,320 | 737.89 | 31.02 |
| | Other Health Impairment | ≥ 1,110 | 710.56 | 23.52 | ≥ 51,260 | 738.48 | 30.90 |
| | Specific Learning Disability | ≥ 2,840 | 705.86 | 20.46 | ≥ 49,530 | 739.72 | 30.52 |
| | Speech or Language Impairment | ≥ 650 | 729.63 | 30.21 | ≥ 51,720 | 737.98 | 31.02 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 52,370 | 737.89 | 31.02 |
| | Visual Impairment | < 50 | NR | NR | ≥ 52,340 | 737.89 | 31.02 |
| | Other | < 50 | NR | NR | ≥ 52,370 | 737.88 | 31.02 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 52,370 | 737.88 | 31.02 |
| | Unknown | < 50 | NR | NR | ≥ 52,370 | 737.88 | 31.02 |

| Special Education Classification Scale-Score Means and Standard Deviations: English Language Arts (continued) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Yes | | | No | | |
| Grade | Group | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| 7 | Gifted | ≥ 1,680 | 791.65 | 27.08 | ≥ 50,250 | 738.93 | 35.89 |
| | Talented | ≥ 1,550 | 767.49 | 30.67 | ≥ 50,370 | 739.81 | 36.70 |
| | Autism | ≥ 240 | 713.54 | 35.33 | ≥ 51,690 | 740.77 | 36.80 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 51,930 | 740.64 | 36.84 |
| | Developmental Delay | < 50 | NR | NR | ≥ 51,930 | 740.64 | 36.84 |
| | Emotional Disturbance | ≥ 150 | 704.51 | 32.13 | ≥ 51,770 | 740.75 | 36.80 |
| | HI—Deaf | < 50 | NR | NR | ≥ 51,910 | 740.66 | 36.82 |
| | HI—Hard-of-Hearing | ≥ 60 | 719.60 | 41.09 | ≥ 51,870 | 740.66 | 36.83 |
| | Mild Mental Disability | ≥ 240 | 678.06 | 21.28 | ≥ 51,690 | 740.93 | 36.64 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 51,930 | 740.64 | 36.83 |
| | Orthopedic Impairment | ≥ 50 | 735.79 | 42.14 | ≥ 51,870 | 740.64 | 36.83 |
| | Other Health Impairment | ≥ 1,060 | 705.19 | 30.57 | ≥ 50,860 | 741.38 | 36.59 |
| | Specific Learning Disability | ≥ 2,800 | 699.74 | 28.16 | ≥ 49,130 | 742.97 | 35.89 |
| | Speech or Language Impairment | ≥ 470 | 728.45 | 34.77 | ≥ 51,460 | 740.75 | 36.84 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 51,920 | 740.65 | 36.83 |
| | Visual Impairment | < 50 | NR | NR | ≥ 51,880 | 740.64 | 36.84 |
| | Other | < 50 | NR | NR | ≥ 51,930 | 740.64 | 36.84 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 51,930 | 740.64 | 36.84 |
| | Unknown | < 50 | NR | NR | ≥ 51,930 | 740.64 | 36.84 |
| 8 | Gifted | ≥ 1,710 | 796.77 | 29.30 | ≥ 48,730 | 742.41 | 36.42 |
| | Talented | ≥ 1,590 | 771.37 | 32.31 | ≥ 48,860 | 743.37 | 37.35 |
| | Autism | ≥ 220 | 717.71 | 36.22 | ≥ 50,220 | 744.38 | 37.48 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 50,450 | 744.26 | 37.52 |
| | Developmental Delay | < 50 | NR | NR | ≥ 50,450 | 744.26 | 37.52 |
| | Emotional Disturbance | ≥ 200 | 701.09 | 31.96 | ≥ 50,250 | 744.43 | 37.44 |
| | HI—Deaf | < 50 | NR | NR | ≥ 50,430 | 744.27 | 37.51 |
| | HI—Hard-of-Hearing | ≥ 70 | 717.36 | 37.26 | ≥ 50,380 | 744.29 | 37.51 |
| | Mild Mental Disability | ≥ 190 | 680.70 | 23.44 | ≥ 50,260 | 744.50 | 37.36 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 50,440 | 744.26 | 37.51 |
| | Orthopedic Impairment | ≥ 70 | 733.08 | 41.02 | ≥ 50,370 | 744.27 | 37.51 |
| | Other Health Impairment | ≥ 930 | 708.73 | 32.52 | ≥ 49,520 | 744.92 | 37.29 |
| | Specific Learning Disability | ≥ 2,510 | 705.19 | 27.96 | ≥ 47,940 | 746.30 | 36.83 |
| | Speech or Language Impairment | ≥ 320 | 729.92 | 34.16 | ≥ 50,120 | 744.35 | 37.52 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 50,440 | 744.26 | 37.52 |
| | Visual Impairment | < 50 | NR | NR | ≥ 50,400 | 744.27 | 37.51 |
| | Other | < 50 | NR | NR | ≥ 50,450 | 744.26 | 37.52 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 50,450 | 744.26 | 37.52 |
| | Unknown | < 50 | NR | NR | ≥ 50,450 | 744.26 | 37.52 |

**Table 10.18 Special Education Classification Scale-Score Means and Standard Deviations: Mathematics**

| Special Education Classification Scale-Score Means and Standard Deviations: Mathematics | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Yes | | | No | | |
| Grade | Group | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| 3 | Gifted | ≥ 1,080 | 795.38 | 22.74 | ≥ 55,710 | 741.49 | 32.61 |
| | Talented | ≥ 510 | 770.84 | 25.99 | ≥ 56,280 | 742.26 | 33.23 |
| | Autism | ≥ 310 | 721.68 | 37.03 | ≥ 56,490 | 742.63 | 33.22 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 56,800 | 742.52 | 33.28 |
| | Developmental Delay | ≥ 480 | 715.21 | 31.22 | ≥ 56,310 | 742.75 | 33.20 |
| | Emotional Disturbance | ≥ 90 | 719.09 | 30.36 | ≥ 56,710 | 742.55 | 33.27 |
| | HI—Deaf | < 50 | NR | NR | ≥ 56,780 | 742.53 | 33.28 |
| | HI—Hard-of-Hearing | ≥ 50 | 728.04 | 30.17 | ≥ 56,740 | 742.53 | 33.28 |
| | Mild Mental Disability | ≥ 340 | 694.51 | 25.74 | ≥ 56,450 | 742.81 | 33.11 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 56,780 | 742.54 | 33.26 |
| | Orthopedic Impairment | ≥ 70 | 733.22 | 32.15 | ≥ 56,720 | 742.53 | 33.28 |
| | Other Health Impairment | ≥ 910 | 717.79 | 30.17 | ≥ 55,890 | 742.92 | 33.17 |
| | Specific Learning Disability | ≥ 2,160 | 717.25 | 25.11 | ≥ 54,630 | 743.52 | 33.17 |
| | Speech or Language Impairment | ≥ 1,810 | 742.12 | 33.22 | ≥ 54,980 | 742.53 | 33.28 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 56,790 | 742.52 | 33.27 |
| | Visual Impairment | < 50 | NR | NR | ≥ 56,760 | 742.52 | 33.28 |
| | Other | < 50 | NR | NR | ≥ 56,800 | 742.52 | 33.28 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 56,800 | 742.52 | 33.28 |
| | Unknown | < 50 | NR | NR | ≥ 56,800 | 742.52 | 33.28 |
| 4 | Gifted | ≥ 1,420 | 786.38 | 23.04 | ≥ 54,810 | 738.98 | 29.91 |
| | Talented | ≥ 860 | 760.80 | 27.03 | ≥ 55,370 | 739.86 | 30.61 |
| | Autism | ≥ 290 | 724.43 | 29.98 | ≥ 55,940 | 740.27 | 30.65 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 56,230 | 740.18 | 30.67 |
| | Developmental Delay | < 50 | NR | NR | ≥ 56,210 | 740.20 | 30.67 |
| | Emotional Disturbance | ≥ 150 | 716.23 | 27.79 | ≥ 56,080 | 740.25 | 30.65 |
| | HI—Deaf | < 50 | NR | NR | ≥ 56,220 | 740.19 | 30.67 |
| | HI—Hard-of-Hearing | ≥ 60 | 731.24 | 28.04 | ≥ 56,170 | 740.19 | 30.67 |
| | Mild Mental Disability | ≥ 350 | 702.62 | 19.82 | ≥ 55,880 | 740.42 | 30.58 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 56,230 | 740.19 | 30.67 |
| | Orthopedic Impairment | ≥ 60 | 726.18 | 30.28 | ≥ 56,170 | 740.20 | 30.67 |
| | Other Health Impairment | ≥ 1,020 | 718.78 | 23.73 | ≥ 55,210 | 740.58 | 30.64 |
| | Specific Learning Disability | ≥ 2,680 | 718.12 | 20.48 | ≥ 53,550 | 741.29 | 30.68 |
| | Speech or Language Impairment | ≥ 1,430 | 739.07 | 31.26 | ≥ 54,800 | 740.21 | 30.66 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 56,230 | 740.19 | 30.67 |
| | Visual Impairment | < 50 | NR | NR | ≥ 56,190 | 740.19 | 30.67 |
| | Other | < 50 | NR | NR | ≥ 56,230 | 740.19 | 30.67 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 56,230 | 740.18 | 30.67 |
| | Unknown | < 50 | NR | NR | ≥ 56,230 | 740.18 | 30.67 |

| Special Education Classification Scale-Score Means and Standard Deviations: Mathematics (continued) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Yes | | | No | | |
| Grade | Group | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| 5 | Gifted | ≥ 1,520 | 783.27 | 25.18 | ≥ 51,790 | 734.65 | 28.10 |
| | Talented | ≥ 1,120 | 753.30 | 26.08 | ≥ 52,190 | 735.67 | 29.12 |
| | Autism | ≥ 270 | 720.04 | 28.68 | ≥ 53,040 | 736.12 | 29.15 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 53,310 | 736.04 | 29.17 |
| | Developmental Delay | ≥ 60 | 707.97 | 21.79 | ≥ 53,250 | 736.07 | 29.16 |
| | Emotional Disturbance | ≥ 160 | 711.28 | 22.38 | ≥ 53,150 | 736.11 | 29.15 |
| | HI—Deaf | < 50 | NR | NR | ≥ 53,280 | 736.05 | 29.17 |
| | HI—Hard-of-Hearing | ≥ 60 | 731.22 | 25.77 | ≥ 53,240 | 736.04 | 29.17 |
| | Mild Mental Disability | ≥ 310 | 699.94 | 17.43 | ≥ 52,990 | 736.25 | 29.09 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 53,310 | 736.04 | 29.17 |
| | Orthopedic Impairment | ≥ 60 | 729.67 | 28.94 | ≥ 53,240 | 736.05 | 29.17 |
| | Other Health Impairment | ≥ 1,080 | 716.10 | 22.47 | ≥ 52,220 | 736.45 | 29.15 |
| | Specific Learning Disability | ≥ 2,780 | 713.35 | 18.85 | ≥ 50,520 | 737.29 | 29.12 |
| | Speech or Language Impairment | ≥ 980 | 733.17 | 28.65 | ≥ 52,320 | 736.09 | 29.18 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 53,300 | 736.04 | 29.17 |
| | Visual Impairment | < 50 | NR | NR | ≥ 53,280 | 736.05 | 29.17 |
| | Other | < 50 | NR | NR | ≥ 53,310 | 736.04 | 29.17 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 53,310 | 736.04 | 29.17 |
| | Unknown | < 50 | NR | NR | ≥ 53,310 | 736.04 | 29.17 |
| 6 | Gifted | ≥ 1,610 | 782.25 | 24.73 | ≥ 50,740 | 730.32 | 27.97 |
| | Talented | ≥ 1,340 | 750.85 | 26.52 | ≥ 51,000 | 731.42 | 29.18 |
| | Autism | ≥ 270 | 715.33 | 31.15 | ≥ 52,070 | 732.00 | 29.24 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 52,350 | 731.91 | 29.28 |
| | Developmental Delay | < 50 | NR | NR | ≥ 52,310 | 731.94 | 29.27 |
| | Emotional Disturbance | ≥ 200 | 705.61 | 25.63 | ≥ 52,140 | 732.02 | 29.25 |
| | HI—Deaf | < 50 | NR | NR | ≥ 52,330 | 731.92 | 29.28 |
| | HI—Hard-of-Hearing | ≥ 60 | 723.73 | 30.69 | ≥ 52,280 | 731.92 | 29.28 |
| | Mild Mental Disability | ≥ 250 | 687.17 | 17.73 | ≥ 52,090 | 732.13 | 29.16 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 52,350 | 731.91 | 29.28 |
| | Orthopedic Impairment | ≥ 50 | 716.36 | 31.13 | ≥ 52,290 | 731.93 | 29.27 |
| | Other Health Impairment | ≥ 1,110 | 708.23 | 21.82 | ≥ 51,230 | 732.43 | 29.21 |
| | Specific Learning Disability | ≥ 2,840 | 704.81 | 19.30 | ≥ 49,510 | 733.47 | 28.99 |
| | Speech or Language Impairment | ≥ 650 | 727.05 | 28.65 | ≥ 51,700 | 731.98 | 29.28 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 52,340 | 731.92 | 29.28 |
| | Visual Impairment | < 50 | NR | NR | ≥ 52,310 | 731.92 | 29.28 |
| | Other | < 50 | NR | NR | ≥ 52,340 | 731.91 | 29.28 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 52,350 | 731.91 | 29.28 |
| | Unknown | < 50 | NR | NR | ≥ 52,350 | 731.91 | 29.28 |

| Special Education Classification Scale-Score Means and Standard Deviations: Mathematics (continued) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Yes | | | No | | |
| Grade | Group | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| 7 | Gifted | ≥ 1,630 | 774.06 | 20.45 | ≥ 50,160 | 730.55 | 25.03 |
| | Talented | ≥ 1,550 | 747.83 | 22.80 | ≥ 50,240 | 731.43 | 25.97 |
| | Autism | ≥ 240 | 715.78 | 25.16 | ≥ 51,550 | 732.00 | 26.02 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 51,800 | 731.92 | 26.03 |
| | Developmental Delay | < 50 | NR | NR | ≥ 51,800 | 731.92 | 26.03 |
| | Emotional Disturbance | ≥ 150 | 708.94 | 22.78 | ≥ 51,640 | 731.99 | 26.01 |
| | HI—Deaf | < 50 | NR | NR | ≥ 51,770 | 731.93 | 26.03 |
| | HI—Hard-of-Hearing | ≥ 60 | 720.30 | 26.35 | ≥ 51,740 | 731.93 | 26.03 |
| | Mild Mental Disability | ≥ 240 | 693.78 | 14.13 | ≥ 51,550 | 732.10 | 25.95 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 51,790 | 731.92 | 26.03 |
| | Orthopedic Impairment | ≥ 50 | 728.55 | 29.17 | ≥ 51,740 | 731.92 | 26.03 |
| | Other Health Impairment | ≥ 1,060 | 709.32 | 20.20 | ≥ 50,730 | 732.40 | 25.93 |
| | Specific Learning Disability | ≥ 2,800 | 706.37 | 17.97 | ≥ 48,990 | 733.38 | 25.66 |
| | Speech or Language Impairment | ≥ 470 | 724.75 | 25.44 | ≥ 51,320 | 731.99 | 26.03 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 51,780 | 731.93 | 26.03 |
| | Visual Impairment | < 50 | NR | NR | ≥ 51,750 | 731.92 | 26.04 |
| | Other | < 50 | NR | NR | ≥ 51,790 | 731.92 | 26.03 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 51,800 | 731.92 | 26.03 |
| | Unknown | < 50 | NR | NR | ≥ 51,800 | 731.92 | 26.03 |
| 8 | Gifted | ≥ 760 | 782.28 | 30.21 | ≥ 43,950 | 727.47 | 32.36 |
| | Talented | ≥ 1,180 | 747.37 | 29.44 | ≥ 43,530 | 727.89 | 33.03 |
| | Autism | ≥ 210 | 710.93 | 35.54 | ≥ 44,500 | 728.49 | 33.06 |
| | Deaf-Blindness | < 50 | NR | NR | ≥ 44,710 | 728.40 | 33.09 |
| | Developmental Delay | < 50 | NR | NR | ≥ 44,710 | 728.40 | 33.09 |
| | Emotional Disturbance | ≥ 200 | 696.60 | 26.40 | ≥ 44,510 | 728.55 | 33.05 |
| | HI—Deaf | < 50 | NR | NR | ≥ 44,690 | 728.41 | 33.09 |
| | HI—Hard-of-Hearing | ≥ 70 | 706.71 | 30.36 | ≥ 44,640 | 728.44 | 33.08 |
| | Mild Mental Disability | ≥ 190 | 681.06 | 18.04 | ≥ 44,520 | 728.61 | 32.99 |
| | Moderate Mental Disability | < 50 | NR | NR | ≥ 44,710 | 728.41 | 33.09 |
| | Orthopedic Impairment | ≥ 70 | 707.37 | 29.96 | ≥ 44,640 | 728.44 | 33.09 |
| | Other Health Impairment | ≥ 910 | 700.76 | 27.37 | ≥ 43,800 | 728.98 | 32.95 |
| | Specific Learning Disability | ≥ 2,470 | 698.01 | 23.67 | ≥ 42,240 | 730.18 | 32.70 |
| | Speech or Language Impairment | ≥ 300 | 720.51 | 31.31 | ≥ 44,400 | 728.46 | 33.10 |
| | Traumatic Brain Injury | < 50 | NR | NR | ≥ 44,700 | 728.41 | 33.09 |
| | Visual Impairment | < 50 | NR | NR | ≥ 44,670 | 728.42 | 33.08 |
| | Other | < 50 | NR | NR | ≥ 44,710 | 728.41 | 33.09 |
| | HI—Hearing Impairment | < 50 | NR | NR | ≥ 44,710 | 728.40 | 33.09 |
| | Unknown | < 50 | NR | NR | ≥ 44,710 | 728.40 | 33.09 |

## 10.4 Mode Effect Study

It is also important to evaluate fairness in test administration in addition to evaluating fairness by examining performance among subgroups. The 2017 LEAP 2025 ELA and Mathematics tests were administered as both paper-based tests (PBTs) and CBTs for grade 4. The *Standards* indicate that results across different testing modes should be comparable. The mode comparability for the 2017 LEAP 2025 CBT and PBT in grade 4 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.
- Following PARCC's decision to not apply mode adjustment, LDOE and DRC decided to not apply any mode adjustment to the LEAP 2025.

### 10.4.1 Sampling Using Propensity Score Matching

The CBT was administered to a smaller number of students than the PBT; therefore, the CBT was designated as the focal group for PSM (Rosenbaum & Rubin, 1983) and the PBT was considered to be the reference group. That is, all CBT students and their matching PBT students were selected using covariates (matching variables), such as LEAP 2016 ELA and Mathematics raw scores and 2017 bio-demographic information, such as gender, ethnicity, free-lunch, and LEP. The LEAP 2016 CBT and PBT were administered to students in grade 3, and most students took the PBT. Only scale scores of the grade 3 students who took the 2016 PBT were used in this study. All grade 4 students who took the 2017 LEAP 2025 CBT were included, and a sample of matching PBT students was drawn using the R package, MatchIt for PSM.

Table 10.19 shows the number of equivalent CBT and PBT students matched by the PSM method. Grade 4 had a small number of CBT students, making its matching PBT student count small. For mathematics grade 4, there were 1,739 CBT students and 51,852 PBT students who had all PSM covariate information, such as bio-demographics, 2016 ELA and mathematics performance information, and 2017 mode information. Of the 51,852 PBT students, 1,739 were selected (a number equivalent to the number of CBT students) by considering all covariates.

**Table 10.19 Number of Students Used for Propensity Score Matching**

| Content | Grade | CBT | PBT | |
|---|---|---|---|---|
| | | Total | Total | Selected |
| Mathematics | 4 | ≥ 1,730 | ≥ 51,850 | ≥ 1,730 |
| ELA | 4 | ≥ 1,730 | ≥ 51,740 | ≥ 1,730 |

**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 four unique items in each ELA CBT and PBT, and these items were dropped from analysis. Table 10.20 shows the number of mode DIF items flagged using the same rules that are used in NAEP. For mathematics, there were two C- flag items and one B- flag item for grade 4, and for ELA, there were two C- flag items for grade 5. The negative sign indicates the CBT item was more difficult than the same PBT item.

**Table 10.20 2017 LEAP 2025 Mode DIF Statistics: Number of Flagged Items**

| | | | DIF | | | |
|---|---|---|---|---|---|---|
| Content | Grade | N of Items | C- | C+ | B- | B+ |
| Mathematics | 4 | 43 | 2 | 0 | 1 | 0 |
| ELA | 5 | 26 | 2 | 0 | 0 | 0 |

Scale scores of the CBT and PBT were estimated using the item parameters for score reporting, and their difference scores were calculated. ESs of the difference scores were calculated as follows:

ES = (CBT Mean – PBT Mean)/$\sqrt{(CBT\ VAR + PBT\ VAR)/2}$, where $VAR = SD^2$.

Table 10.21 shows the mean scale scores and standard deviations of the CBT and PBT. When the mean scale scores were compared, the CBT appeared slightly more difficult than the PBT for mathematics. When a flag criterion, |0.2|, which can be considered a small difference criterion, was applied, mathematics grade 4 was flagged.

**Table 10.21 Mode Study Scale Score Differences and Effect Size**

| | | | CBT | | PBT | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Content | Grade | N of Items | Mean | Std. Dev. | Mean | Std. Dev. | Mean Diff. | ES | Flag > \|0.2\| |
| Mathematics | 4 | 42 | 730.0 | 31.4 | 747.1 | 31.1 | 6.6 | 0.21 | Yes |
| ELA | 4 | 30 | 741.0 | 33.9 | 736.6 | 31.3 | 6.1 | 0.19 | No |

## 10.5 Summary

In summary, the overall purpose of this chapter is to address fairness concerns that are relevant to the administration of LEAP 2025. 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)

## 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

Camilli, G., & Shepard, A. L. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publication.

Center for Assessment. (June 2017). *LEAP 2017: English language arts -grade 6 summary – comparability with PARCC performance standards* [Memorandum]. Dove, NH.

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.

Dorans, N. J., & Schmitt, M. P. (1991). *Constructed response and differential item functioning: A pragmatic approach*. Princeton, NJ: Educational Testing Service.

Data Recognition Corporation. (2016). *Interpretive guide: Grades 3–8 ELA and math* Maple Grove, MN.

Educational Testing Service, Pearson, & Measured Progress. (2016). *Final technical report for 2015 administration*. PARCC. Retrieved from https://parcc-assessment.org/wp-content/uploads/2018/02/PARCC-2015-Tech-Report.pdf

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.Wainerand H. I. Braun (Eds.), Test Validity, pp. 129-145. Hillsdale, NJ:Erlbaum.

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.

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

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719–748.

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

Muraki, E., & Bock, R. D. (2003). *PARSCALE 4: IRT item analysis and test scoring for rating-scale data* [Computer software]. Chicago, IL: Scientific Software.

Pearson. (2015). *Performance level setting technical report*. PARCC. Retrieved from https://parcc-assessment.org/parcc_pls_techreport_011316_toparcc-final/.

Pearson. (2017). *PARCC: Final technical report for 2016 administration.* PARCC. Retrived from https://parcc-assessment.org/wp-content/uploads/2018/02/PARCC-2016-Tech-Report.pdf.

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, 479.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, 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*, 233–251.