2018 LEAP 2025 Grades 3-8
Operational Technical Report
English Language Arts and Mathematics

Submitted to the
Louisiana Department of Education

July 2019

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

Table of Contents ii
Executive Summary 5
E.1 Overview of This Report 5
E.2 Administration 6
E.3 Student Performance 7
E.4 Validity and Test Scores 7
Chapter 1: Introduction 9
1.1 Background 9
1.2 Purpose of the LEAP 2025 10
1.3 Design of the LEAP 2025 10
Chapter 2: The Uses of Test Scores 13
2.1 Uses of Test Scores 13
2.2 Test-Level Scores 13
2.2.1 Scale Scores 14
2.2.2 Levels of Achievement 14
2.2.3 Use of Test-Level Scores 14
2.3 Claim-,Subclaim-, and Subcategory- Level Subscores 15
2.3.1 Use of the Claim-Level and Subclaim-Level Ratings 15
Chapter 3: Test Content Development 16
3.1.1 Defining the Specific Test Blueprint 18
3.1.2 English Language Arts Test Blueprints and Test Designs 18
3.1.3 Mathematics Test Blueprints and Test Designs 29
3.1.4 Item Development and Selection 40
3.1.5 Considerations of Test Fairness in Item Development 40
3.1.6 Item Reviews 40
3.1.7 Louisiana Item Alignment Review 41
3.2 Operational Test Selection 42
3.2.1 English Language Arts Item and Passage Selection Process and Criteria 42
3.2.2 ELA Content Review and Forms Development 43
3.2.3 Review of the English Language Arts Items and Forms 43
3.2.4 Mathematics Item Selection Process and Criteria 44
3.2.5 Mathematics Content Review and Forms Development 45
3.2.6 Item-Selection Options for Special Cases 47
3.2.7 Psychometric Review 47
3.3 Universal Design 54
3.4 Accommodations and Designated Supports 55
3.5 Item and Task Specifications 57
3.6 Summary 57
Chapter 4: Test Administration 59
4.1.1 Return Material Forms and Guidelines 65
4.1.2 Security Checklists 65
4.1.3 Interpretive Guides 68
4.2 Test Security Measures 68
4.2.1 Data Forensic Analyses 68

Copyright © 2019 by Louisiana Department of Education.

4.2.2 Response Change Analysis ... 68
4.2.3 Score Fluctuation Analysis ... 68
4.2.4 Web Monitoring ... 69
4.2.5 Plagiarism Detection ... 69
4.3 Test Administration ... 69
4.3.1 Time ... 69
4.3.2 Accommodations ... 70
4.4 Summary ... 75
Chapter 5: Constructed-Response and Technology-Enhanced Scoring ... 77
5.1 Constructed-Response Item Scoring Process ... 77
5.1.1 Selection of Scoring Evaluators ... 78
5.1.2 Security ... 79
5.1.3 Handscoring Training Process ... 79
5.1.4 Monitoring the Scoring Process ... 83
5.2 Inter-Rater Reliability ... 86
5.3 Technology-Enhanced Item Scoring Process ... 92
5.4 Multiple-Choice and Multiple-Select Item Scoring Process ... 92
5.5 Summary ... 93
Chapter 6: Operational Data Analyses ... 94
6.1.1 Test-Level Statistics ... 94
6.1.2 Item-Level Statistics ... 96
6.2 Item Reponse Theory ... 122
6.3 Calibration and Linking ... 127
6.3.1 Calibration of 2018 LEAP 2025 Tests ... 128
6.3.2 Linking 2018 LEAP 2025 Grades 3–8 to PARCC Scale ... 132
6.4 Summary ... 180
Chapter 7: Test Results ... 181
7.1 Current Administration Data ... 187
7.1.1 Description of Each Type of Report ... 190
Chapter 8: Performance-Level Setting ... 193
8.1 PARCC Performance-Level Setting Process for English Language Arts and Mathematics ... 193
8.2 Cut Scores ... 193
8.2.1 Claim Cut Scores ... 194
8.3 Summary ... 195
Chapter 9: Evidence of Validity ... 196
9.1 Construct-Irrelevant Variance and Construct Underrepresentation ... 197
9.2 Reliability ... 197
9.2.1 Test Reliability ... 198
9.2.2 Standard Error of Measurement ... 200
9.2.3 Conditional Standard Error of Measurement ... 200
9.2.4 Classification Accuracy and Consistency ... 203
9.2.1 Convergent Validity ... 207
9.3 Principal Components Analysis ... 208
9.4 Analyses by Claims, Subclaims, and Subcategories ... 210
9.4.1 Correlations among Claims, Subclaims, and Subcategories ... 210

Copyright © 2019 by Louisiana Department of Education.

9.4.2 Reliability of Claims, Subclaims, or Subcategories 215
9.4.3 Standard Error of Measurement of Claims, Subclaims, or Subcategories 216
9.5 Divergent (Discriminant) Validity 220
9.6 Regression of LEAP 2025 from 2017 to 2018 220
9.7 Summary 224
Chapter 10: Fairness 225
10.1 Minimizing Bias through Careful Test Development 226
10.2 Evaluating Bias through Differential Item Functioning (DIF) Statistics 227
10.3 Evaluating Bias through Impact Analysis 235
10.3.1 Reliability 235
10.3.2 Effect Size 243
10.4 Mode Effect Study 259
10.4.1 Sampling Using Propensity Score Matching 259
10.5 Summary 261
References 263
Appendix A—Forms Development Process for Spanish TransADAPTation Forms 265
Appendix B—Accommodated Print Form Creation 267
Appendix C—Item Alignment Review Process 269
Appendix D – LEAP 2025 Spring 2018 Handscoring/AI Documentation 275

Copyright © 2019 by Louisiana Department of Education.

EXECUTIVE SUMMARY

This report is a technical summary of the 2018 administration of the Louisiana Educational Assessment Program (LEAP 2025) in English Language Arts (ELA) and mathematics for grades 3 through 8. The LEAP 2025 is a summative assessment in ELA and mathematics administered in grades 3 through 8 and high school. These tests are designed to measure students' readiness for the next grade or course of study and proficiency in ELA and mathematics. The ELA and mathematics test forms were constructed 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 2018 operational technical report.

E.1 Overview of This Report

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

The Uses of Test Scores (Chapter 2)

Chapter 2 of the technical report discusses the concept of validity evidence. This technical report is composed of evidence that supports the intended uses of the LEAP 2025 test scores, and Chapter 2 discusses some of 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 2018 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 2018 LEAP 2025 assessments, which include the following major parts: the classical item analysis; calibration, scaling, and linking using item response theory (IRT) models; and student

Copyright © 2019 by Louisiana Department of Education.

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 2018 LEAP 2025 assessments. 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 test 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 test scores.

Fairness (Chapter 10)

Chapter 10 of the technical report discusses fairness and how the LEAP 2025 assessments 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.

E.2 Administration

In the spring of 2018, Louisiana administered the LEAP 2025 summative assessments in ELA and mathematics to students in grades 3–8. A paper-based test (PBT) and a computer-based test (CBT) were administered in grades 3–4, and a CBT was administered in grades 5–8. The CBTs were administered from April 9 to May 4, 2018. The PBTs were administered from April 30 to May 4, 2018. Test administration is discussed in Chapter 4 of this report.

Eighty eight 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.

Copyright © 2019 by Louisiana Department of Education.

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

Grade	Accountable in ELA	Percentage Reportable in ELA	Accountable in Mathematics	Percentage Reportable in Mathematics*
3	≥55,370	99.85%	≥55,990	99.85%
4	≥55,800	99.82%	≥56,340	99.87%
5	≥55,280	99.93%	≥55,290	99.92%
6	≥52,810	99.86%	≥52,820	99.87%
7	≥51,590	99.76%	≥51,590	99.77%
8	≥51,090	99.71%	≥51,100	99.72%

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

E.3 Student Performance

Tables E.2 and E.3 present the percentage of students who were classified in each of the 2018 achievement levels for ELA and mathematics.

Table E.2 Percentage of Students Classified in 2018 Achievement Levels Using 2018 Census Data: English Language Arts

Grade	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
3	14.2	18.2	22.3	39.8	5.6
4	10.8	17.0	28.7	34.8	8.8
5	8.8	17.7	30.4	39.3	3.7
6	9.3	24.6	31.5	30.3	4.4
7	10.7	19.2	26.8	31.4	11.9
8	10.8	17.4	26.6	36.9	8.4

Table E.3 Percentage of Students Classified in 2018 Achievement Levels Using 2018 Census Data: Mathematics

Grade	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
3	10.3	19.7	28.1	34.6	7.3
4	8.6	22.8	30.3	34.4	3.9
5	10.2	25.8	34.0	25.7	4.2
6	11.6	29.0	32.0	24.8	2.6
7	9.9	29.0	35.7	22.9	2.4
8	20.9	27.4	23.7	26.1	1.9

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

E.4 Validity and Test Scores

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

Copyright © 2019 by Louisiana Department of Education.

2025 test scores. Chapter 3 discusses the test development process used to create the LEAP 2025 tests, which is important to the content-related validity of the LEAP 2025 test scores. Chapter 4 presents information on test administration. Chapter 5 discusses the scoring process and the results of the inter-rater reliability studies. Chapter 6 presents the test scaling and linking procedures, student scoring methodology, and the results of other operational data analyses. Chapter 7 reviews the results of the 2018 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 processes used and the process for evaluating bias to ensure fairness of the LEAP 2025 for all examinees.

Copyright © 2019 by Louisiana Department of Education.

CHAPTER 1: INTRODUCTION

The LEAP 2025 aseessment system is designed to measure students' knowledge of ELA, mathematics, science, and social studies. This report provides a technical overview of the LEAP 2025 ELA and mathematics assessments administered in grades 3 through 8 in the spring of 2018 and presents evidence for the validity of the 2018 LEAP 2025 ELA and mathematics assessment scores.

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

1.1 Background

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

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

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

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

In March 2016, BESE approved the Louisiana Student Standards in ELA and mathematics. In the 2016–2017 and 2017–2018 school years, students in grades 3–8, except those qualifying for an alternate assessment for students with the most significant cognitive disabilities (the LAA 1 in 2016–2017 or LEAP Connect in 2017-2018), were administered forms for ELA and mathematics that consisted of PARCC assessment items while developing some Louisiana-owned items to enhance the PARCC item bank. This allowed for the continued comparability to forms administered in the 2014–2015 and 2015–2016 school years.

The information that follows describes the technical aspects of the 2018 LEAP 2025 ELA and mathematics assessments and provides information about how to read and interpret the data.

Copyright © 2019 by Louisiana Department of Education.

1.2 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 the full range of student performance, including the performance of high- and low-performing students and providing information for educators and parents about student readiness and whether students are "on track" for college and careers.

1.3 Design of the LEAP 2025

Students in grades 3–8 were administered computer-based tests (CBTs) in both ELA and mathematics; some school systems opted to administer paper-based tests (PBTs) to students in grades 3 and 4. All mathematics assessments were translated into Spanish forms. Additionally, a braille form was available for each grade and content area. The braille form was based on the PBT in grades 3 and 4 and was based on the CBT in grades 5–8. Online tools allowed students to magnify assessment items, as needed, and students with visual impairments could also take large-print versions of the PBTs. See Chapter 3, Section 3.4 for more information about the accommodations and designated supports available for LEAP 2025 testers.

PARCC operational blueprints are the basis of the design of the 2018 LEAP 2025 test blueprints and test design. The 2018 LEAP 2025 test blueprints and test design for ELA and mathematics differ from PARCC in order to reduce testing time but maintain full coverage and include a variety of standards.

The 2018 LEAP 2025 ELA blueprints kept a similar design as PARCC; which includes both performance-based tasks and standalone passage sets, and a higher percentage of reading points to writing points. However; only two of the three types of performance tasks—Research Simulation Task (RST) and Literary Analysis Task (LAT) **or** Narrative Writing Task (NWT)—are included on each of the grade-level tests. All three task types are represented across grades 3–8, which allows Louisiana flexibility in the choice of the tasks administered for each grade from year to year and encourages teachers to focus equally on all three writing types. Besides having two (instead of three) performance tasks, the 2018 LEAP 2025 Spring ELA blueprints are also different from PARCC with respect to testing time and percentage of reading and writing points. Since the choice of Literary Analysis Task or Narrative Writing Task is determined during the forms construction process, alternative blueprints—one with a Literary Analysis Task and a Research Simulation Task and the other with a Research Simulation Task and a Narrative Writing Task—were created for each grade.

The passages chosen for the 2018 LEAP 2025 ELA assessments contain a variety of text types, including texts that diverse populations will find engaging and that have a balance of gender and ethnicity among authors. Chosen passages are authentic, contain a variety of different genres and varying degrees of text complexity, and are content-rich, engaging, high-quality, and challenging. Additionally, paired passages are selected with careful consideration of the purpose of the standards that require the use of more than one text to be assessed with that passage. This combination of criteria during passage selection allows students to demonstrate their ability to

Copyright © 2019 by Louisiana Department of Education.

read and comprehend a range of complex texts. With respect to an overall passage set and form, the goal is to ensure as much coverage of standards as possible.

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

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

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

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

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

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

Copyright © 2019 by Louisiana Department of Education.

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

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

Additional details about the design of the ELA and mathematics assessments can be found in *Chapter 3, Test Content Development*.

Copyright © 2019 by Louisiana Department of Education.

CHAPTER 2: THE USES OF TEST SCORES

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

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

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

2.1 Uses of Test Scores

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

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

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

2.2 Test-Level Scores

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

The 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

Copyright © 2019 by Louisiana Department of Education.

response theory (IRT) scales for ELA and mathematics and were reviewed and approved for use by LDOE content experts and committees of Louisiana educators. Braille and large-print forms, in addition to Spanish translations of mathematics forms, were also developed. See *Chapter 3, Test Content Development*, for additional details about the processes used to develop these test forms.

The following sections discuss two types of test-level scores that are reported to indicate a student's achievement on the LEAP 2025 assessments: (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 assessments. The overall scale score for a content area quantifies the achievement being measured by the ELA or mathematics assessments. In other words, the scale score represents the student's level of achievement, where higher scale scores indicate higher levels of achievement on the test and lower scale scores indicate lower levels of achievement. For all LEAP 2025 test forms, the lowest obtainable scale score (LOSS) is 650 and the highest obtainable scale score (HOSS) is 850.

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

2.2.2 **Levels of Achievement**

A student's performance on the ELA or mathematics LEAP 2025 assessments is reported in one of five levels of achievement: *Advanced, Mastery, Basic, Approaching Basic,* or *Unsatisfactory*. The cut scores for the ELA and mathematics achievement levels were established by PARCC using the Evidence-Based Standard Setting (EBSS) method (Beimers, Way, McClarty, & Miles, 2012) for the PARCC Performance-Level Setting (PLS) process. Details regarding the PLS process can be found in the *Performance Level Setting Technical Report* (Pearson, 2015) (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 LEAP *Interpretive Guide* (see Chapter 7).

2.2.3 **Use of Test-Level Scores**

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

Copyright © 2019 by Louisiana Department of Education.

understand, and are able to do relative to the Louisiana Student Standards in ELA and mathematics.

2.3 Claim-,Subclaim-, and Subcategory- Level Subscores

A student's performance on the ELA claims (i.e., reading and writing) is reported by one of three ratings: *Strong*, *Moderate*, or *Weak*. Additionally, performance on the subclaims is reported at the student level for ELA and mathematics. ELA has three subclaims for reading and two subclaims for writing, as described in Table 3.1, *ELA Claims and Subclaims*. Mathematics has four subclaims, as described in Table 3.9, *Overview of LEAP 2025 Mathematics Task Types and Reporting Categories*. Subclaim performance is reported in one of three ratings: *Strong*, *Moderate*, or *Weak*.

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

2.3.1 Use of the Claim-, Subclaim-, and Subcategory-Level Ratings

The purpose of reporting claim-, subclaim-, or subcategory-level performance ratings on LEAP 2025 assessments is to show, for each student, the relationship between the overall achievement being measured and the skills in each of the areas defined by the claims in ELA, the subclaims in ELA and mathematics, and the subcategories in mathematics. Teachers may use these ratings 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 performance ratings. Chapter 9 of this technical report provides evidence of construct-related validity that further supports the use of these performance ratings.

Copyright © 2019 by Louisiana Department of Education.

CHAPTER 3: TEST CONTENT DEVELOPMENT

Content-related validity in achievement tests is evidenced by a correspondence between test content and the range of knowledge and skills that compose the construct the assessment is designed to measure, i.e., the ELA or mathematics Louisiana Student Standards. Content-related validity can be demonstrated through consistent adherence to test blueprints, through a high-quality test development process that includes review of items for accessibility to English learners and students with disabilities, and through alignment studies performed by independent groups. This 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 2018 LEAP 2025 assessments were expected to cover.

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

Standard 4.0 states the following:

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

Standard 4.1 states the following:

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

The 2018 LEAP 2025 test specifications consisted of a test blueprint and a test design for each grade and content area. The 2018 blueprints and test designs were closely aligned to the PARCC flagship blueprint that was used for their spring 2018 test administrations. The specific content area and grade-level test blueprints for the 2018 LEAP 2025 ELA assessments for grades 3–8 were designed with the goal for all students to read, understand, and express understanding of complex, grade-level texts. The specific content area and grade-level test blueprints for the 2018 LEAP 2025 mathematics assessments for grades 3–8 were designed with the goal of supporting students to become mathematically proficient by focusing on three components of rigor: conceptual understanding, procedural skill and fluency, and application. The 2018 LEAP 2025 ELA and mathematics assessments for grades 3–8 provide questions that have been reviewed by Louisiana educators to ensure their alignment to the Louisiana Student Standards and appropriateness for Louisiana students; measure the full range of student performance, including the performance of high- and low-performing students; and inform educators and parents about student readiness in ELA and mathematics and whether students are “on track” for college and

Copyright © 2019 by Louisiana Department of Education.

careers. For ELA and mathematics, the 2018 LEAP 2025 assessments for grades 3–8 use the same reporting categories that were used in spring 2017. Subcategories in mathematics were introduced for spring 2018 in response to requests from school systems. In ELA, the type and/or number of reading literary and informational passage sets has changed from the 2017 LEAP 2025 assessments to the 2018 LEAP 2025 ELA assessments to reflect a similar change made in the PARCC blueprints.

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. DRC contracted with PARCC and was provided access to the entire bank of items and related passages that could potentially be used on operational forms. The acquired items/passages make up the available item pool for the 2018 LEAP 2025 forms construction. Please refer to the PARCC Model Content Frameworks for ELA/Literacy (Grades 3-11) and the PARCC Model Content Frameworks for Mathematics (Grades 3-11) for additional information about the development of item specifications and blueprints for the PARCC assessments. These resources can be accessed via the Educator Resources page of the PARCC website. Once at the website, for ELA click on PARCC Model Content Frameworks for ELA/Literacy (Grades 3-11) and for mathematics click on PARCC Model Content Frameworks for Mathematics (Grades 3-11). LDOE and DRC confirmed that all items selected for use on the LEAP 2025 forms were appropriate for use on Louisiana assessments by convening a committee review, by which Louisiana educators reviewed and approved items from the PARCC item bank prior to form selection. Educators reviewed alignment to content standards and appropriateness of item content.

The ELA and mathematics LEAP 2025 assessments for grades 3–8 were developed based on the requirements of RFP #678PUR-LEAP 2016 Mathematics and ELA as follows:

The assessments shall be

- aligned to the ELA and mathematics Louisiana Student Standards;
- designed to be accessible for use by the widest possible range of students, including, but not limited to, students with disabilities and students with limited English proficiency;
- constructed to yield valid and reliable test results;
- constructed to report student performance using achievement level policy definitions and reporting categories which are comparable to a significant number of other states and, for grades 3 through 8 assessments, to Louisiana's 2015–2017 assessments;
- constructed to use Louisiana's grades 3 through 8 ELA and mathematics assessments as the baseline scale[1] to report test results for grades 3 through 8 students;
- developed to limit the amount of testing time required and to be in compliance with state law regarding testing time;
- developed and reviewed with Louisiana educator involvement;

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

Copyright © 2019 by Louisiana Department of Education.

- non-computer adaptive;
- used in assessing students' readiness to successfully transition to postsecondary education and the workplace; and
- administered, scored, and reported through a separate administration contract in both paper- and computer-based formats.

The products of the above requirements are dual-mode assessments—paper-based tests (PBTs) and computer-based tests (CBTs)—composed of PARCC test items aligned to the Louisiana Student Standards. Louisiana had access to the complete PARCC item bank from which to select for the alignment review more than enough items to build the forms needed for the 2018 LEAP 2025 ELA and mathematics assessments. For grades 3 and 4, the contract with PARCC provided for the use of enough items and related passages that had been approved during Item Alignment Review to create one complete operational test form for each content area and grade that can be administered in a dual-mode testing environment (i.e., PBT and CBT). For grades 5–8, Louisiana selected one CBT form per grade from the content that was reviewed during Item Alignment Review. These items and passages became the available item pool used to construct the 2018 forms. DRC and LDOE content experts scrutinized each final blueprint to ensure optimal content coverage and prudent use of time and resources. In general, the blueprints represent content sampling proportions that reflect intended emphasis in instruction and mastery at each grade level and are comparable to PARCC 2018 test blueprints. The test specifications provide the numbers of items by reporting category, assessment focus, and item type, and they demonstrate the desired proportions within test delivery and available item pool constraints. These specifications can be found in the *2017-2018 LEAP 2025 3-8 English Language Arts and Mathematics Assessment Frameworks*. All assessments were fixed forms, in that all students who received the same form were administered the same set of items, as the forms were not adaptive.

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.

3.1.2 **English Language Arts Test Blueprints and Test Designs**

The ELA test was administered during a CBT testing window (April 9 – May 4, 2018) and during a PBT testing window (April 30 – May 4, 2018). Only two of the three types of performance tasks—Research Simulation Task, Literary Analysis Task, and Narrative Writing Task—were included on each of the Louisiana grade-level tests; however, all three types were represented across grades 3 through 8. This allows Louisiana to rotate the tasks given for each grade from administration to administration and encourages educators to focus on all three performance task types. As the choice of Literary Analysis Task or Narrative Writing Task would be made during the forms construction process, alternative blueprints—one with a Literary Analysis Task and a Research Simulation Task and the other with a Research Simulation Task and a Narrative Writing Task—were created for each grade. During forms construction, the Narrative Writing Task was selected for grades 3, 5, 6, and 7 and the Literary Analysis Task was selected for grades 4 and 8, based on item performance and the quality of the available passage sets for each performance task.

Copyright © 2019 by Louisiana Department of Education.

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

Table 3.1 ELA Claims and Subclaims

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 use details from provided texts to compose well-developed, organized, clear writing.
	Knowledge and Use of Language Conventions	Students use the rules of standard English (grammar, mechanics, and usage) to compose writing.

These reporting categories are the same as the reporting categories on the Spring 2015-2017 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 subclaim, which may help identify when students need additional support or more challenging work in reading and writing; and
- how well schools and school systems help students achieve expectations.

The session testing times shown in the ELA test blueprints (see Tables 3.2 through 3.7) are based on PARCC testing times proportioned to be comparable based on the passage type. 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.

Copyright © 2019 by Louisiana Department of Education.

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

Section	Task/ Item Set	Number of Passages	Claim/Subclaim	Number of two-point SR Items	Number of Points from two-point SR Items	Number of PCR Items	Number of Points from PCR Items	Total Items	Total Points	Assessable ELA Student Standards (by subclaim)	Testing Time (minutes)
1	Research Simulation Task	2	Reading: Reading Informational Text/Reading Vocabulary*	6	12	1	3	6	15	RI standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	75
			Writing: Written Expression	0	0		9	1	9	Writing standards W.1-2, 7-8, 10	
			Writing: Knowledge of Language and Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Totals	2		6	12	1	15	7	27		
2	Narrative Writing Task	1	Reading: Reading Literary Text/Reading Vocabulary*	4	8	1	0	4	8	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	75
			Writing: Written Expression	0	0		9	1	9	Writing standards W.3, 10	
			Writing: Knowledge of Language and Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Reading Literary/ Informational Texts	1	Reading: Reading Literary Text/Reading Informational Text/Reading Vocab*	4	8	0	0	4	8	RL Standards 1-3, 5-10; RI standards 1-3, 5-10; vocabulary standards RL.4, RI.4, L.4, L.5	
	Totals	2		8	16	1	12	9	28		
3	Reading Literary Texts	2	Reading: Reading Literary Text/Reading Vocabulary*	8	16	0	0	8	16	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	60**
	Reading Informational Texts		Reading: Reading Informational Text/Reading Vocabulary*							RI standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	
	Totals	2		8	16	0	0	8	16		
Grade 3 Totals		6	Reading: Reading Literary Text/Reading Vocab*	22	44	2	0	22	47	47	210
			Reading: Reading Informational Text/Reading Vocab*				3				
			Writing: Written Expression	0	0		18	2	18	24	
			Writing: Knowledge of Language and Conventions	0	0		6		6		
			Total	22	44	2	27	24	71	71	

**There must be at least eight points of reading vocabulary items on the test.*
***The time in session 3 includes an additional passage set that is being field tested.*

Copyright © 2019 by Louisiana Department of Education.

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

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

**There must be at least eight points of reading vocabulary items on the test.*
***The time in session 3 includes an additional passage set that is being field tested.*

Copyright © 2019 by Louisiana Department of Education.

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

Section	Task/ Item Set	Number of Passages	Claim/Subclaim	Number of two-point SR Items	Number of Points from two-point SR Items	Number of PCR Items	Number of Points from PCR Items	Total Items	Total Points	Assessable ELA Student Standards (by subclaim)	Testing Time (minutes)
1	Research Simulation Task	3	Reading: Reading Informational Text/Reading Vocabulary*	8	16	1	4	8	20	RI standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	90
			Writing: Written Expression	0	0		12	1	12	Writing standards W.1-2, 4, 7-10,	
			Writing: Knowledge of Language and Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Totals	3		8	16	1	19	9	35		
2	Narrative Writing Task	1	Reading: Reading Literary Text/Reading Vocabulary*	4	8	1	0	4	8	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	90
			Writing: Written Expression	0	0		9	1	9	Writing standards W.3, 4, 10,	
			Writing: Knowledge of Language and Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Reading Informational Texts	1-2	Reading (Reading Informational Text/Reading Vocabulary)	6	12	0	0	6	12	RI standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	
	Totals	2-3		10	20	1	12	11	32		
3	Reading Literary Texts	2	Reading: Reading Literary Text/Reading Vocabulary*	8	16	0	0	8	16	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	60**
	Totals	2		8	16	0	0	8	16		
Grade 5 Totals		**7-8**	Reading: Reading Literary Text/Reading Vocab*	12	24	2	0	12	24	56	240
			Reading: Reading Informational Text/Reading Vocab*	14	28		4	14	32		
			Writing: Written Expression	0	0		21	2	21	27	
			Writing: Knowledge of Language and Conventions	0	0		6		6		
			Total	26	52	2	31	28	83	83	

**There must be at least eight points of reading vocabulary items on the test.*
***The time in session 3 includes an additional passage set that is being field tested.*

Copyright © 2019 by Louisiana Department of Education.

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

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

**There must be at least eight points of reading vocabulary items on the test.*
***The time in session 3 includes an additional passage set that is being field tested.*

Table 3.6 Grade 7 English Language Arts Test Blueprint and Test Design

Copyright © 2019 by Louisiana Department of Education.

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

**There must be at least eight points of reading vocabulary items on the test.*
***The time in session 3 includes an additional passage set that is being field tested.*

Copyright © 2019 by Louisiana Department of Education.

Table 3.7 Grade 8 English Language Arts Test Blueprint and Test Design

Section	Task/ Item Set	Number of Passages	Claim/Subclaim	Number of two-point SR Items	Number of Points from two-point SR Items	Number of PCR Items	Number of Points from PCR Items	Total Items	Total Points	Assessable ELA Student Standards (by subclaim)	Testing Time (minutes)
1	Literary Analysis Task	2	Reading: Reading Literary Text/Reading Vocabulary*	6	12	1	4	6	16	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	90
			Writing: Written Expression	0	0		12	1	12	Writing standards W.1-2, 4, 9, 10,	
			Writing: Knowledge of Language and Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Reading Literary / Informational Texts	1	Reading (Reading Literary Text/Reading Informational Text/Reading Vocabulary)	4	8	0	0	4	8	RL Standards 1-3, 5-10; RI standards 1-3, 5-10; vocabulary standards RL.4, RI.4, L.4, L.5	
	Totals	3		10	20	1	19	11	39		
2	Research Simulation Task	3	Reading: Reading Informational Text/ Reading Vocabulary*	8	16	1	4	8	20	RI standards 1-3, 5-10; vocabulary standards RI.4, L.4, L.5	90
			Writing: Written Expression	0	0		12	1	12	Writing standards W.1-2, 4, 7- 10,	
			Writing: Knowledge of Language and Conventions	0	0		3		3	Convention standards L.1, 2, plus language skills from previous grades	
	Totals	3		8	16	1	19	9	35		
3	Reading Literary Texts	2	Reading: Reading Literary Text/Reading Vocabulary*	10	20	0	0	10	20	RL Standards 1-3, 5-10; vocabulary standards RL.4, L.4, L.5	80**
	Reading Informational Texts		Reading: Reading Informational Text/Reading Vocab*			0	0			RI standards 1-3, 5, 7-10; vocabulary standards RI.4, L.4, L.5	
	Totals	2		10	20	0	0	10	20		
Grade 8 Totals		**8**	Reading: Reading Literary Text/Reading Vocab*	28	56	2	4	28	64	64	260
			Reading: Reading Informational Text/Reading Vocab*				4				
			Writing: Written Expression	0	0		24	2	24	30	
			Writing: Knowledge 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.*
***The time in session 3 includes an additional passage set that is being field tested.*

Copyright © 2019 by Louisiana Department of Education.

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

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

The following item types were included in the 2018 LEAP 2025 ELA assessments:

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

Copyright © 2019 by Louisiana Department of Education.

ways to show understanding that parallels classroom instructional techniques teachers use to determine whether students are able to comprehend grade-level, complex text.

TE Items may involve any of the following:

- Highlighting text: requires a student to select text-based answer(s) from within a larger text
- Drag and drop: requires a student to move draggable elements (e.g., words, phrases, or sentences) into one or more drop boxes (e.g., cells within a table or part[s] of a diagram)
- Dropdown menu: requires a student to select from one or more dropdown menus to complete a phrase or sentence
- Match interaction table: requires a student to select a checkbox in each row from two or more columns to classify statements presented in each row

- Prose constructed response – PCR: This item type appears at the end of each of the tasks and asks students to create an extended and complete written response. It elicits evidence that students have understood a text or texts they have read and can communicate that understanding well, both in terms of written expression and knowledge of language and conventions.

A variety of item types allows for the measurement of the full range of student performance, including the performance of high- and low-performing students. Items and tasks should be clearly aligned with specific standards. Some items and tasks may ask students to draw evidence from one specific standard, while others may ask students to draw evidence from several standards.

Copyright © 2019 by Louisiana Department of Education.

The following table details the number of items and points by session and item type for each of the PBT (grades 3 and 4) and CBT (grades 3–8) forms.

Table 3.8 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 -Based Test (PBT)	ELA	3	1. Research Simulation Task	6	12					1	15	71
			2. Narrative Writing Task/Reading Passage	7	14	1	2			1	12	
			3. Reading Literary/Informational Texts	8	16							
	ELA	4	1. Literary Analysis Task/Reading Passage	9	18	1	2			1	19	86
			2. Research Simulation Task	7	14	1	2			1	19	
			3. Reading Literary/Informational Texts	6	12							
Computer-Based Tests (CBT)	ELA	3	1. Research Simulation Task	5	10			1	2	1	15	71
			2. Narrative Writing Task/Reading Passage	5	10	1	2	2	4	1	12	
			3. Reading Literary/Informational Texts	7	14			1	2			
	ELA	4	1. Literary Analysis Task/Reading Passage	6	12	1	2	3	6	1	19	86
			2. Research Simulation Task	6	12	1	2	1	2	1	19	
			3. Reading Literary/Informational Texts	5	10			1	2			
	ELA	5	1 Research Simulation Task	5	10	3	6			1	19	83
			2. Narrative Writing Task/Reading Passage	6	12	1	2	3	6	1	12	
			3. Reading Literary/Informational Texts	3	6	3	6	2	4			
	ELA	6	1. Research Simulation Task	7	14	1	2			1	19	90
			2. Narrative Writing Task/Reading Passage	5	10	2	4	3	6	1	15	
			3. Reading Literary/Informational Texts	3	6	4	8	3	6			
	ELA	7	1. Research Simulation Task	6	12	1	2	1	2	1	19	90
			2. Narrative Writing Task/Reading Passage	5	10	1	2	4	8	1	15	
			3. Reading Literary/Informational Texts	5	10	4	8	1	2			
	ELA	8	1. Literary Analysis Task/Reading Passage	4	8	3	6	3	6	1	19	94
			2. Research Simulation Task	4	8	2	4	2	4	1	19	
			3. Reading Literary/Informational Texts	6	12	4	8					

Copyright © 2019 by Louisiana Department of Education.

3.1.3 Mathematics Test Blueprints and Test Designs

The mathematics assessments were administered during a CBT testing window (April 9 – May 4, 2018) or during a PBT testing window (April 30 – May 4, 2018). The 2018 mathematics assessment had a similar structure to the 2017 assessment: each test session included the four mathematics subclaims, using the three mathematics task types (see Table 3.9).

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: Major Content, Additional and Supporting Content, Expressing Mathematical Reasoning, and Modeling and Application. Each task type is designed to align with at least one of the Standards for Mathematical Practice (MP).

Table 3.9 Overview of LEAP 2025 Mathematics Task Types and Reporting Categories

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

These reporting categories are the same as the reporting categories on the spring 2017 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 subclaim, 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.

Copyright © 2019 by Louisiana Department of Education.

Table 3.10 provides the distribution of operational points by subclaim, or reporting category, by grade.

Table 3.10 Distribution of Points by Reporting Category—Mathematics

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

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

Table 3.11 Major Content Subcategories by Grade

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

Copyright © 2019 by Louisiana Department of Education.

The resulting 2018 LEAP 2025 mathematics test blueprints are shown in Tables 3.12–3.17.

Table 3.12 Grade 3 Mathematics Test Blueprint

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

Copyright © 2019 by Louisiana Department of Education.

Table 3.13 Grade 4 Mathematics Test Blueprint

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

Table 3.14 Grade 5 Mathematics Test Blueprint

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

Copyright © 2019 by Louisiana Department of Education.

Table 3.15 Grade 6 Mathematics Test Blueprint

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

Table 3.16 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**	

Copyright © 2019 by Louisiana Department of Education.

Table 3.17 Grade 8 Mathematics Test Blueprint

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

Unlike the ELA test blueprints, which were organized by test sessions one through three, the mathematics test blueprints were organized by reporting categories, so it was necessary to define the general structure of the test forms by test session. The design goal was to have balanced test sessions with a variety of task types and equivalent testing times. For all forms in grades 3–5, students are prohibited from using calculators, except for those students with a calculator accommodation. For session one of the mathematics test in grades 6–8, students are prohibited from using calculators, except for those students with a calculator accommodation. Calculators are allowed to be used by all students in grades 6–8 in sessions two and three. The general test structures (see Tables 3.18–3.23) guided test form sequencing and design. The LEAP 2025 Calculator Policy provided the basis for calculator designation of tasks and items.

Copyright © 2019 by Louisiana Department of Education.

Table 3.18 General Mathematics Test Structure—Grade 3

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

Table 3.19 General Mathematics Test Structure—Grade 4

Reporting Category	Test Session						TOTAL (Operational Only)	
	Session 1 No Calculator		Session 2 No Calculator		Session 3 No Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Major Content	9–10	10	8–10	10	10	10	27–30	30
Additional & Supporting Content	3–4	4	2–4	4	2	2	7–10	10
Expressing Mathematical Reasoning	1	4	1	3	1	3	3	10
Modeling & Application	1	3	1	3	1	6	3	12
Field Test	3	3-4	1	3-6	2	2-3	6	8-13
TOTAL (Operational Only)	**15**	**21**	**14**	**20**	**14**	**21**	**43**	**62**
Test Duration (minutes)	**75**		**85**		**75**		**235**	

Copyright © 2019 by Louisiana Department of Education.

Table 3.20 General Mathematics Test Structure—Grade 5

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

Table 3.21 General Mathematics Test Structure—Grade 6

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

Copyright © 2019 by Louisiana Department of Education.

Table 3.22 General Mathematics Test Structure—Grade 7

Reporting Category	Test Session						TOTAL (Operational Only)	
	Session 1 No Calculator		Session 2 Calculator		Session 3 Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Major Content	16–20	20	3–5	5	3–5	5	26–30	30
Additional & Supporting Content	0	0	3–5	5	3–5	5	6–10	10
Expressing Mathematical Reasoning	0	0	2	7	2	7	4	14
Modeling & Application	0	0	2	9	1	3	3	12
Field Test	2	2-3	3	3-4	1	3-6	6	8-13
TOTAL (Operational Only)	**16–20**	**20**	**12–14**	**26**	**11–13**	**20**	**43**	**66**
Test Duration (minutes)	**60**		**90**		**90**		**240**	

Table 3.23 General Mathematics Test Structure—Grade 8

Reporting Category	Test Session						TOTAL (Operational Only)	
	Session 1 No Calculator		Session 2 Calculator		Session 3 Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Major Content	13–18	18	3–6	6	4–6	6	25–30	30
Additional & Supporting Content	2–4	4	2–3	3	2–3	3	5–10	10
Expressing Mathematical Reasoning	0	0	2	7	2	7	4	14
Modeling & Application	0	0	2	9	1	3	3	12
Field Test	3	3	2	2-3	1	3-6	6	8-12
TOTAL (Operational Only)	**15–20**	**22**	**10–13**	**25**	**10–12**	**19**	**42**	**66**
Test Duration (minutes)	**60**		**90**		**90**		**240**	

Copyright © 2019 by Louisiana Department of Education.

The following item types were used in the 2018 LEAP 2025 mathematics assessments:

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

Copyright © 2019 by Louisiana Department of Education.

- Dropdown menu: requires a student to select from one or more dropdown menus to complete a sentence, phrase, or expression/equation/inequality
- Hot spot: requires a student to select one or more responses by choosing selectable areas on the screen
- Match interaction table: requires a student to select a checkbox in each row from two or more columns
- Graph input: requires a student to enter a response on a coordinate grid
- Number line input: requires a student to enter a response on a number line
- Line plot: requires a student to complete a line plot with "X" as the input

A variety of item types allows for the measurement of the full range of student performance, including that of high- and low-performing students.

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

Table 3.1 Distribution of Mathematics Tasks and Points by Task Type

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

Copyright © 2019 by Louisiana Department of Education.

3.1.4 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 2018 LEAP 2025 ELA and mathematics assessments came from the PARCC consortium's item bank.

The PARCC items selected for use on the 2018 LEAP forms were used to equate to the PARCC scale. Operational forms were selected based on LEAP 2025 test blueprint specifications, which were supported by statistical data from PARCC operational testing.

3.1.5 Considerations of Test Fairness in Item Development

Standard 3.2 is particularly relevant to fairness in item development:

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

Bias and sensitivity guidelines used 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, race, gender, regional background, native language, religion, sexual orientation, and socioeconomic status. DRC relied strongly on the bias and sensitivity guidelines in the development of the assessments, particularly in item selection and review. To be included in the assessments, items had to comply with the bias and sensitivity guidelines and be approved by Louisiana educators involved in the Louisiana Item Alignment Review.

3.1.6 Item Reviews

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

- text reviews of all passages (during which participants review and edit passages independently and then discuss content and bias concerns as a grade level group),
- item reviews (during which committees review and edit items for adherence to PARCC foundational documents, basic principles of Universal Design, PARCC Accessibility Guidelines, selected metadata fields, and the PARCC Style Guide),

Copyright © 2019 by Louisiana Department of Education.

- bias and sensitivity reviews (during which educators and community members review items and tasks to confirm the absence of issues relating to bias, fairness, and sensitivity to ensure that items and tasks do not unfairly advantage or disadvantage any student subgroup over another subgroup),
- editorial reviews (during which the review committee completes a copy edit review and records their comments), and
- data reviews (during which educators evaluate item-level statistics to determine eligibility of items and tasks to move forward to the operational assessments).

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

Standard 3.1 states the following:

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

Standard 3.2 states the following:

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

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

3.1.7 **Louisiana Item Alignment Review**

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

DRC, with guidance from LDOE, conducted the Louisiana Item Alignment Review in September 2017 with committees of Louisiana educators. Grade-level committees met for one (mathematics) or two (ELA) days to provide feedback on the alignment and appropriateness of

Copyright © 2019 by Louisiana Department of Education.

items that made up the PARCC item bank. To the extent possible, each committee included educators from different parts of Louisiana, who represent all Louisiana students (e.g., special education, English Learners, students with disabilities, etc.). Committee members are also representative of the diverse demographics of the state.

As described in the preceding sections, items presented at this review had already been through a rigorous review process before and after the items were field tested by PARCC to ensure quality and appropriateness. They were then selected for inclusion in the form selection pool, imported into IDEAS, DRC's item banking system, and formatted for use on a Louisiana test form. The items were then placed on mock test forms to allow them to be reviewed as students would see the items, tasks, and passages. Louisiana educators then reviewed these items to confirm that the items were acceptable for use in a Louisiana assessment. Educators reviewed items individually to verify that each item aligned to the Louisiana Student Standard(s) for that item prior to discussing the items as a group. In addition, educators reviewed item keys and discussed the cognitive complexity for each item and task. The group then came to a consensus regarding the status of each item: Accepted with Current Alignment, Accepted with Realignment, or Rejected. Items that were accepted were determined to appropriately measure the intended standard(s) and be free of issues of bias, fairness, or sensitivity that could impact student responses to the item. For a detailed description of the process followed during the Item Alignment Reviews, including results and descriptions of the demographics of each committee, see Appendix C, *Item Alignment Review Process*.

3.2 Operational Test Selection

Operational item selection for 2018 took place from October 2017 through January 2018 by LDOE and DRC. The PARCC item pool was used to select fixed LEAP 2025 ELA and mathematics forms.

The LEAP 2025 assessments were given in two modalities: computer-based test (CBT) or paper-based test (PBT). For both ELA and mathematics, students in grades 3 through 8 took the CBTs; some school systems elected to administer the PBTs to students in grades 3 and 4. For ELA, the dual-mode forms were identical except for a small quantity (four to five items) of technology-enhanced items (TE) in each CBT. Items used on PBTs as replacements for the TE items were evidence-based selected-response items that addressed the same content standards and were of similar rigor as the TE items, when possible. For mathematics, short-answer (SA) items were reformatted as gridded-response (GR) items for use on PBTs.

3.2.1 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 were 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 2016 and 2017 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.

Copyright © 2019 by Louisiana Department of Education.

Item Selection Guidelines

- Using the PARCC pool of items, ELA content specialists select performance tasks and passage sets to match the blueprint. The performance tasks and passage 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
  - meeting the requirements of the item specifications;
  - clear and correct answer(s);
  - clear and concise language;
  - grammatical correctness;
  - appropriate range of difficulty; and
  - content that is not offensive, inappropriate, or biased.

3.2.2 **ELA Content Review and Forms Development**

After ELA performance 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 (i.e., inadvertently provide or clue the answer for another item).
- Clang association should be avoided. Clang association occurs 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 (i.e., contain a balance of authors by gender and ethnicity, and appeal to a diverse student population, etc.).
- Forms should include a wide range of topics and a variety of question types.
- Correct answer distributions should follow best practices (e.g., 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.

3.2.3 **Review of the English Language Arts Items and Forms**

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

Copyright © 2019 by Louisiana Department of Education.

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 verified all items against their designated content codes and metadata, both to evaluate the correctness of the coding and to ensure that the given item measures what it purports to measure.

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

- have the appropriate number of correct answer(s) based on item type;
- only one clear, correct answer for each part of an evidence-based selected response (EBSR) item with only four answer options 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 the item 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 assessed;
- be based on content that is accurate and current;
- when appropriate, contain stimulus material that is clear and concise and provides all the information needed,
- contain graphics that are clearly labeled, when appropriate;
- contain answer choices that are reasonably parallel in length and structure;
- contain answer choices that are plausible and reasonable in terms of the requirements of the question as well as the student's grade-level expectations;
- contain distractors that relate to the question in the same way and can be supported by a rationale;
- reflect current teaching and learning practices for the content area; and
- be free of gender, ethnic, racial, cultural, socioeconomic, regional, and other forms of bias.

3.2.4 **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 were the only

Copyright © 2019 by Louisiana Department of Education.

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 2016 and 2017 targets. Although the PARCC item pool was limited, items that were determined to be very difficult (IRT difficulty parameter b > 2.0) and/or not discriminating (IRT discrimination parameter 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
  - meeting the requirements of the item specifications;
  - clear and correct answer(s);
  - clear and concise language;
  - grammatical correctness;
  - appropriate range of difficulty; and
  - content that is not offensive, inappropriate, or biased.

3.2.5 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 (i.e., inadvertently provide or clue the answer for another item). If it is necessary to use items with clueing issues due to item pool limitations, then these items should be positioned in separate testing sessions.
- Clang association should be avoided. Clang association occurs 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 question types.

Copyright © 2019 by Louisiana Department of Education.

- 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.

3.2.5.1. Review of the Mathematics Items and Forms

DRC and LDOE mathematics content specialists also ensured the items were in compliance with the guidelines provided by LDOE, including alignment to the content standards and appropriateness for Louisiana students. As 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 verified all items against their designated content codes and metadata, both to evaluate the correctness of the coding and to ensure that the given item measures what it purports to measure.

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

Guidelines for Reviewing Items Selected for Forms

A good item should

- 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 assessed;
- be based on content that is accurate and current;
- when appropriate, contain stimulus material that is clear and concise and provides all the necessary information;
- when appropriate, contain graphics that are clearly labeled;
- contain answer choices that are plausible and reasonable in terms of the requirements of the question as well as the student's level of knowledge;

Copyright © 2019 by Louisiana Department of Education.

- contain distractors that relate to the question in the same way and can be supported by a rationale;
- reflect current teaching and learning practices in the content area; and
- be free of gender, ethnic, racial, cultural, socioeconomic, regional, and other forms of bias.

3.2.6 Item-Selection Options for Special Cases

While every effort is made to select a test form that meets all psychometric guidelines for excellence, it may not be possible to comply with all the psychometric criteria for item/form difficulty due to item pool limitations. In these cases, critical psychometric guidelines are followed while allowing some tolerance on less critical item-selection guidelines. The tolerance of meeting target characteristics, the relative exposure of previously used operational items, and other considerations (such as content coverage) may possibly be affected in such cases.

3.2.7 Psychometric Review

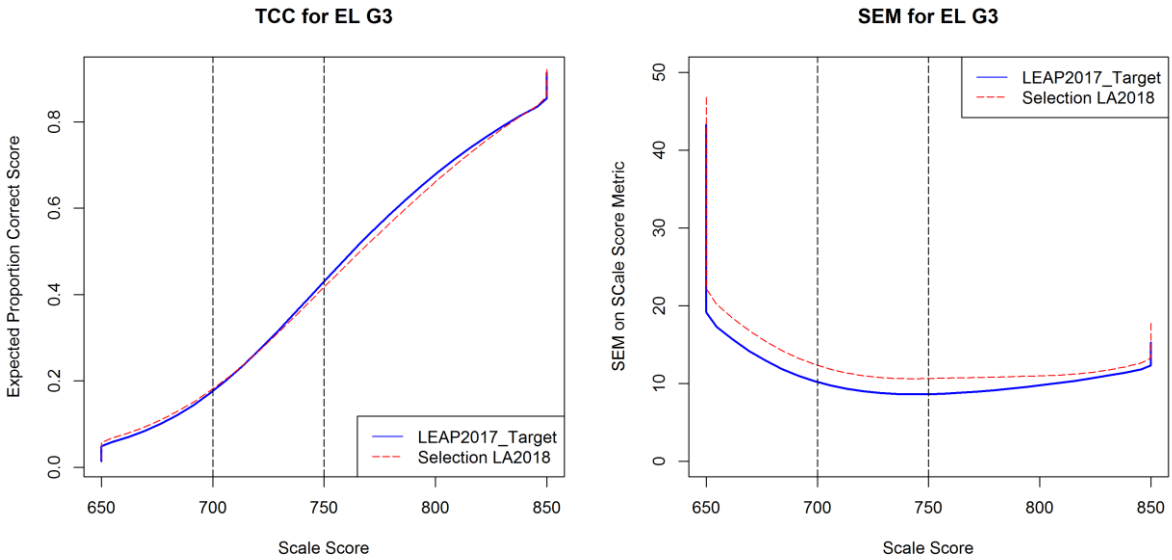
The psychometric evaluation of each selection was centered on reviewing the PARCC items with operational item parameters.

3.2.7.1. Selecting Targets

The PARCC 2017 operational form was used to build a new PARCC 2017 scale and to select the items in the PARCC 2018 operational form. The rationale for the choice of the targets was that each 2018 LEAP 2025 form should be on the PARCC scale and be closely comparable to PARCC assessments. Figures 3.1 through 3.6 for ELA and Figures 3.7 through 3.12 for mathematics show the test characteristic curves (TCCs) and standard errors of measurement (SEMs) of the final forms compared to those of the target forms. The left line graph displays the TCC of the target and the selected 2018 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 2018 form. This summarizes the amount of measurement error surrounding a scale score.

Copyright © 2019 by Louisiana Department of Education.

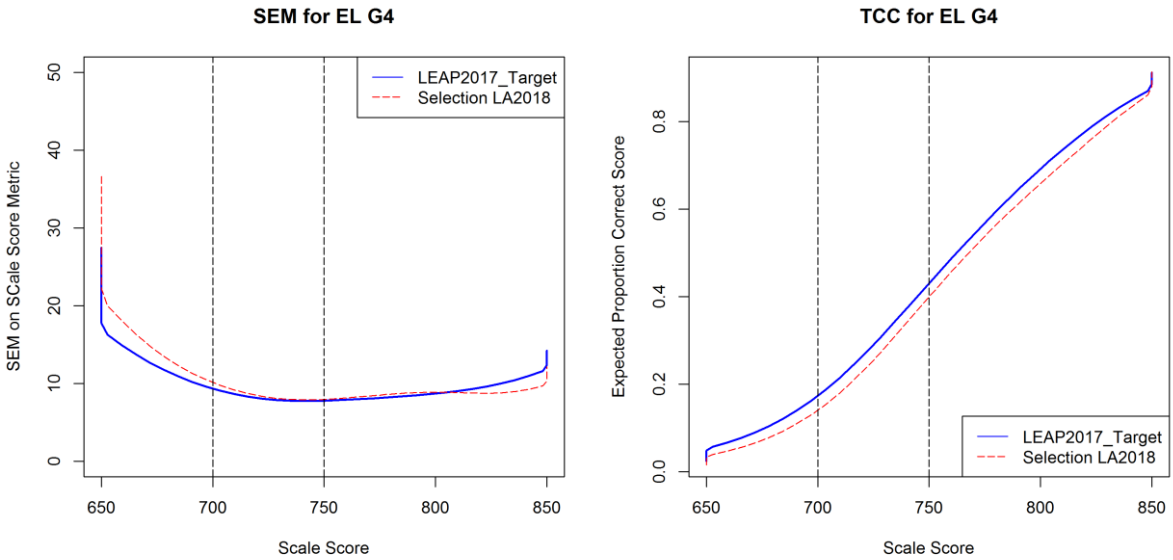
Figure 3.1 2018 ELA Form Evaluation—Grade 3



NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

Figure 3.2 2018 ELA Form Evaluation—Grade 4

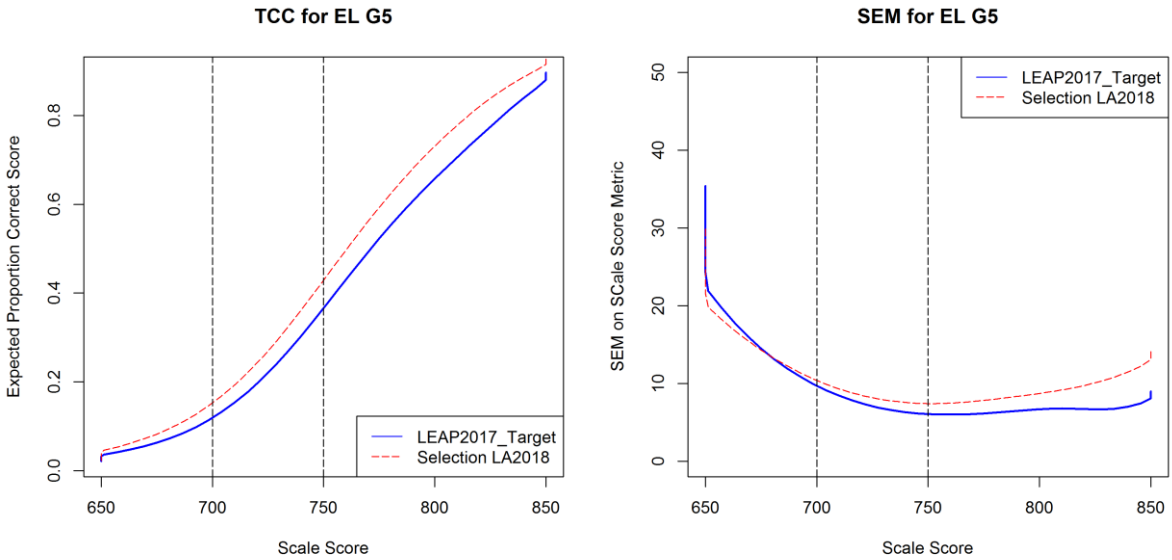


NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

Copyright © 2019 by Louisiana Department of Education.

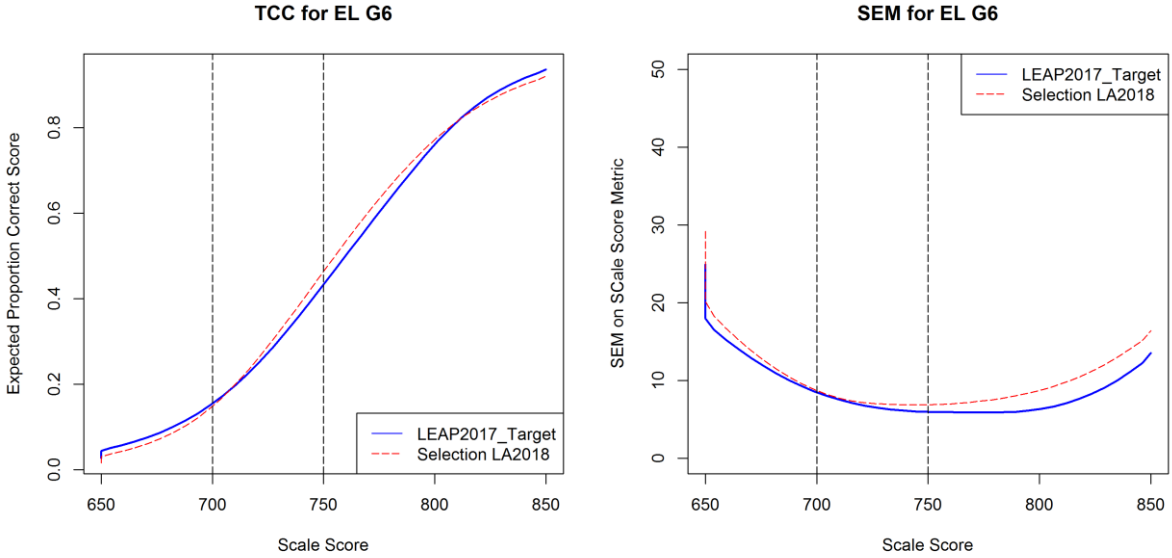
Figure 3.3 2018 ELA Form Evaluation—Grade 5



NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

Figure 3.4 2018 ELA Form Evaluation—Grade 6

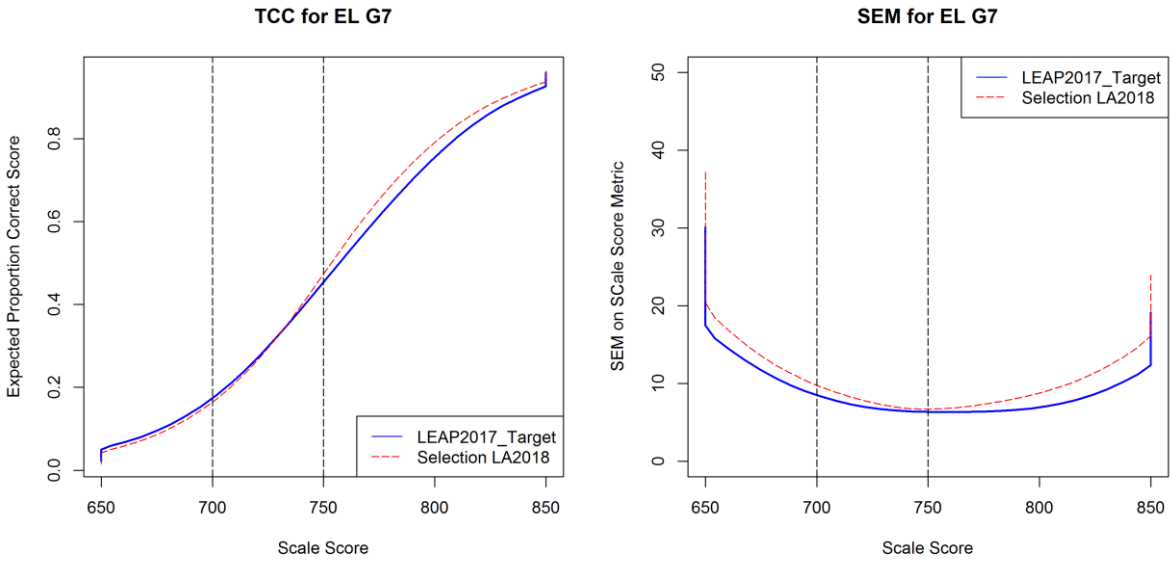


NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

Copyright © 2019 by Louisiana Department of Education.

Figure 3.5 2018 ELA Form Evaluation—Grade 7

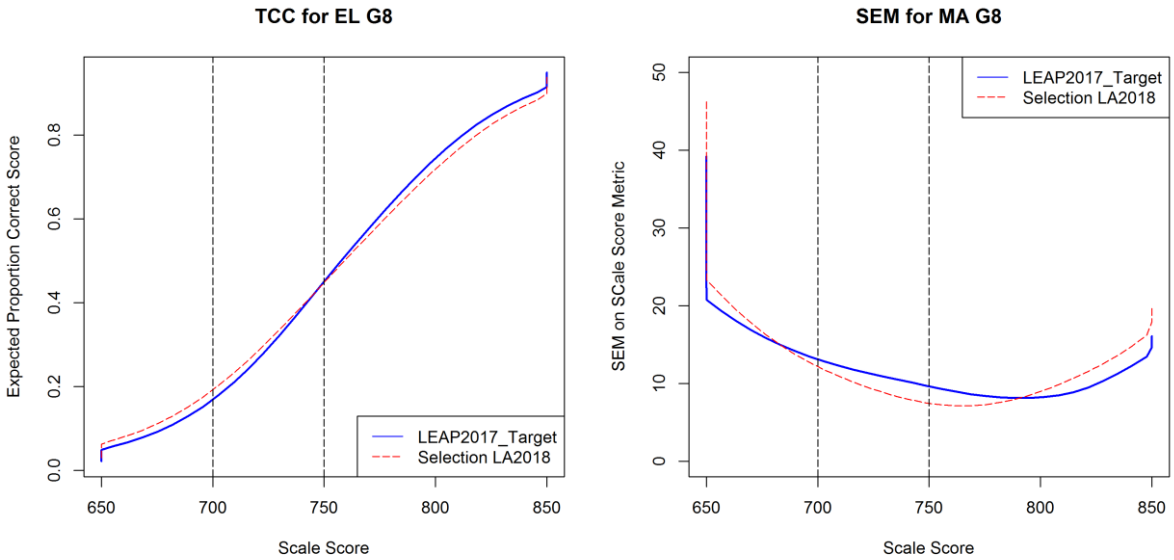


NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

Copyright © 2019 by Louisiana Department of Education.

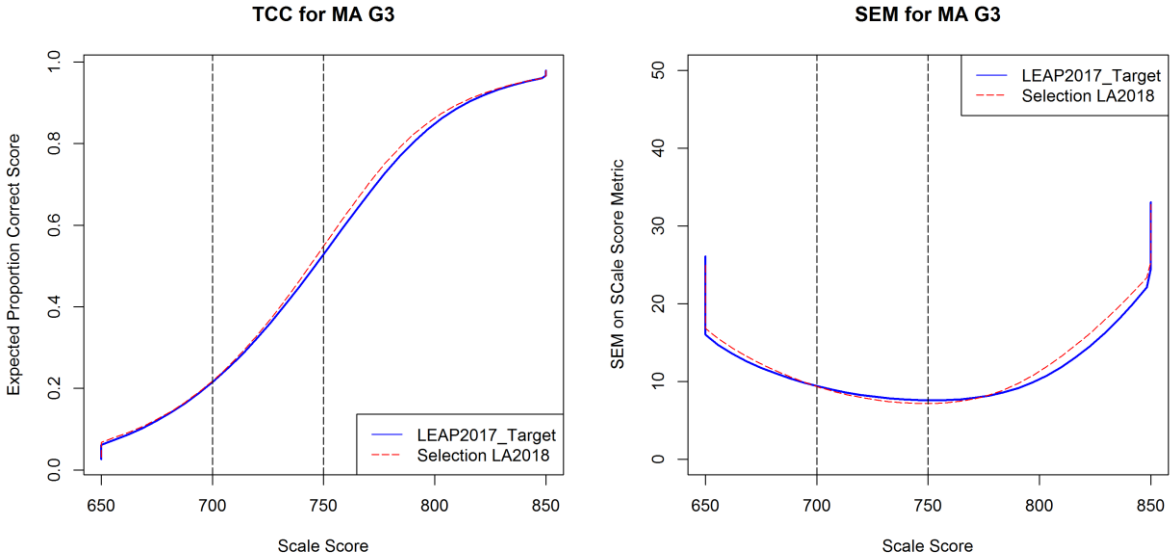
Figure 3.6 2018 ELA Form Evaluation—Grade 8



NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

Figure 3.7 2018 Mathematics Form Evaluation—Grade 3

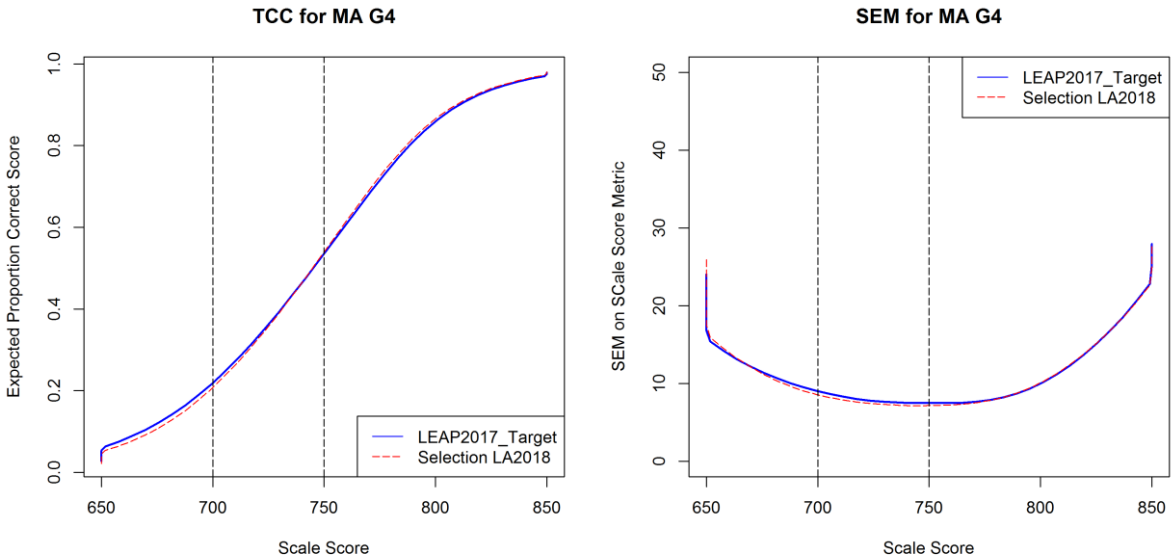


NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

Copyright © 2019 by Louisiana Department of Education.

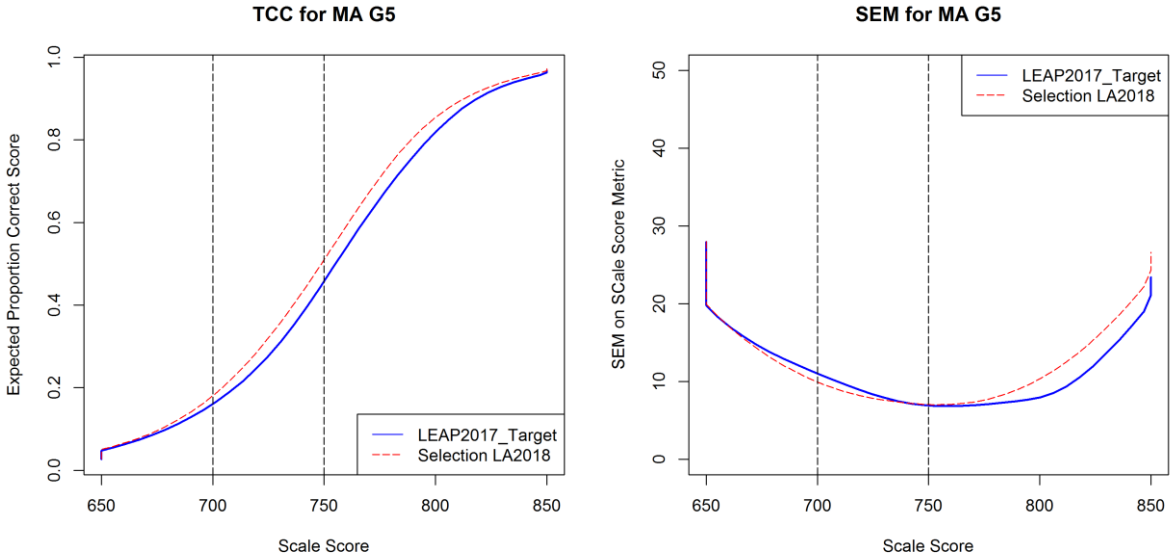
Figure 3.8 2018 Mathematics Form Evaluation—Grade 4



NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

Figure 3.9 2018 Mathematics Form Evaluation—Grade 5



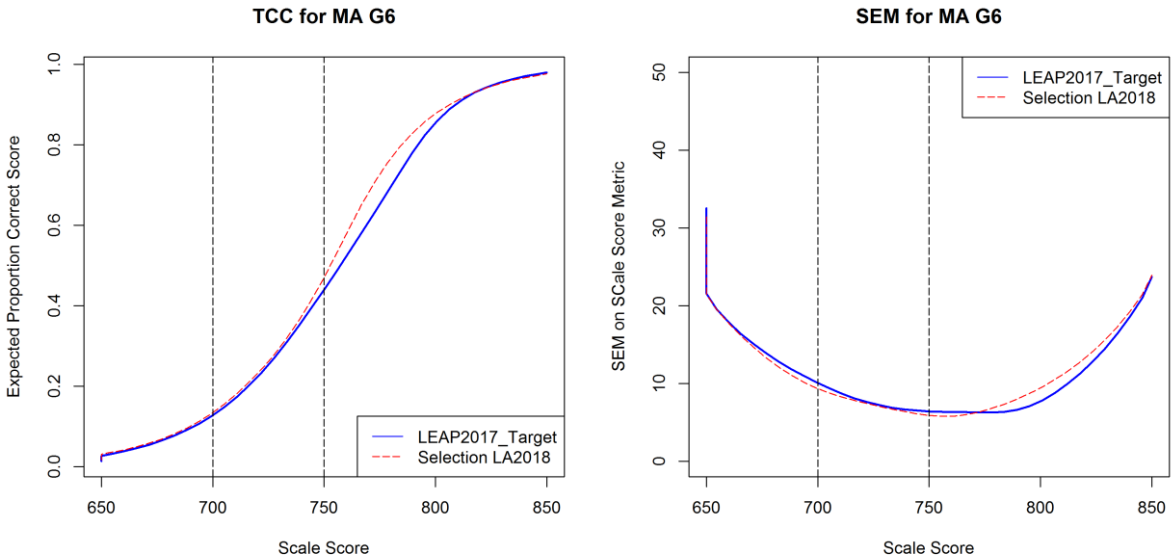
NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

.

Copyright © 2019 by Louisiana Department of Education.

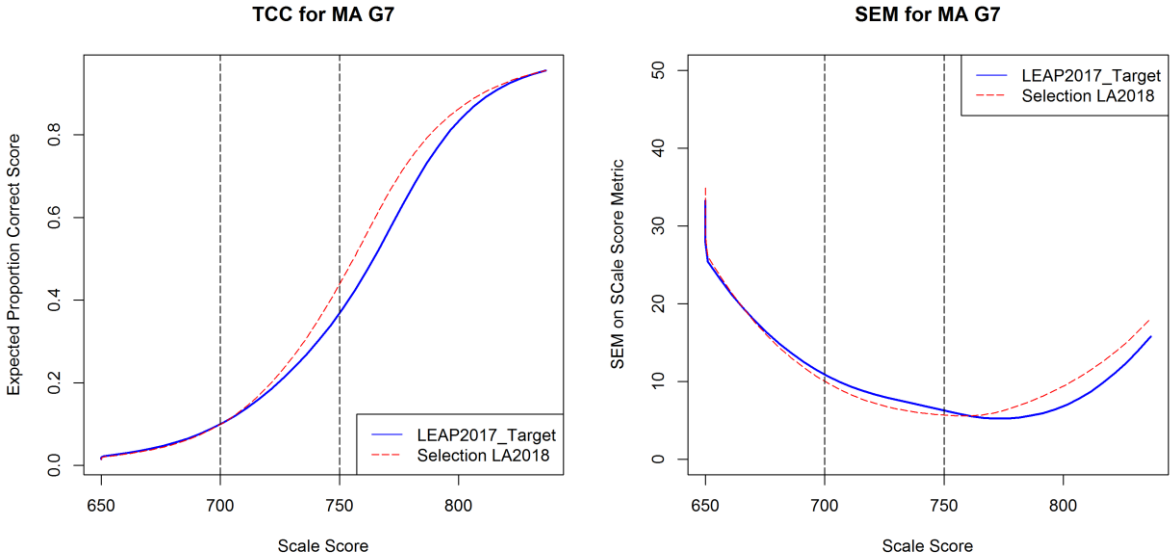
Figure 3.10 2018 Mathematics Form Evaluation—Grade 6



NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

Figure 3.11 2018 Mathematics Form Evaluation—Grade 7

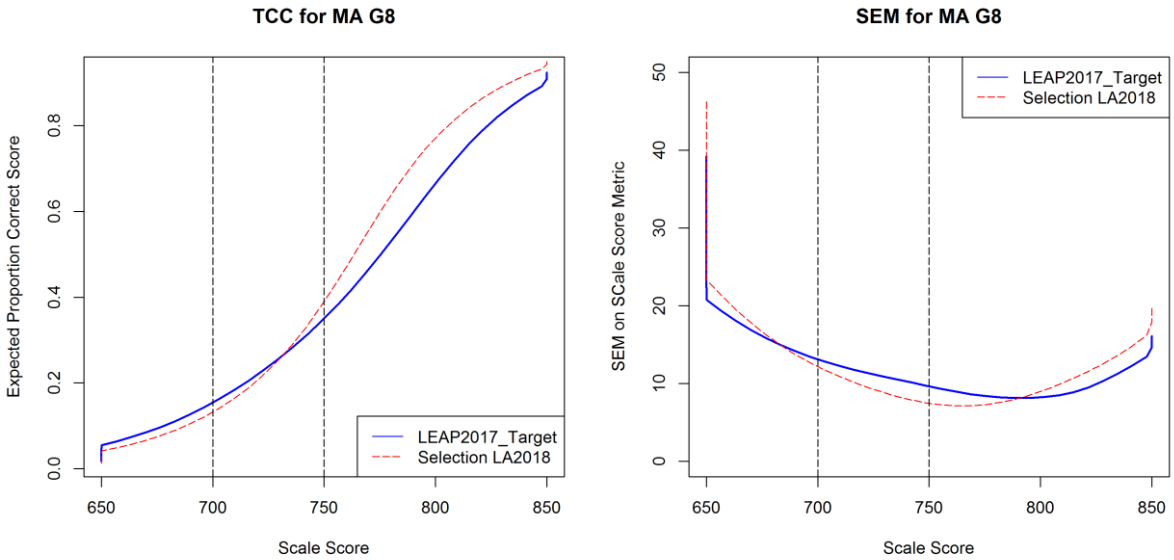


NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

Copyright © 2019 by Louisiana Department of Education.

Figure 3.12 2018 Mathematics Form Evaluation—Grade 8



NOTE:

- *LEAP2017_Target is the 2017 LEAP 2025 intact test form.*
- *Selection LA2018 is the selected 2018 LEAP 2025 test form.*

3.2.7.2. 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 2018 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.

3.3 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 3.24 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 2018 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

Copyright © 2019 by Louisiana Department of Education.

> for the widest possible range of individuals and relevant subgroups in the intended population. (63)

Universal design requires that grade-level assessments measure the performance of students with a wide range of abilities and skills, ensuring that students with diverse learning needs receive opportunities to demonstrate competence on the same content. To ensure that students can access the tests, the LEAP 2025 assessments include simple, clear, and intuitive instructions and procedures; maximum readability and comprehensibility; and maximum legibility. All these design components were addressed primarily through the CBTs. The online test specifications define how directions and test items are formatted online, including the spacing between an item stem and answer choices, and other page elements (such as online tools and Help files) to ensure consistent, clean visual appearance of CBTs. Test directions at the beginning of each test session were 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.

Table 3.24 Elements of Universal Design

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

3.4 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, students with 504 plans, and English learners (ELs) may be provided test administration accommodations as documented on their accommodation plan. More

Copyright © 2019 by Louisiana Department of Education.

information on accommodations can be found in Section 4.3.2 of Chapter 4. Accommodation code definitions can be found in the *Paper-Based Test Administration Manual*.

Accommodated Print forms were developed in grades 5–8 of ELA and mathematics for those students who were unable to participate in an online administration. For a detailed description of the process used to develop the Accommodated Print forms and how to modify technology-enhanced items for use in an Accommodated Print form, see Appendix B, *Accommodated Print Form Creation*.

Braille and large-print test forms were constructed for each grade and content area to enable students with visual impairments to participate in the LEAP 2025 assessments. Braille and large-print forms for grades 3 and 4 of ELA and mathematics were based on the standard-print forms. Braille forms for grades 5–8 of ELA and mathematics were based on the accommodated print forms. There are no large-print versions of the grades 5–8 accommodated print forms. Instead, students needing a large-print version in grades 5–8 use appropriate-size monitors and/or the magnification features of the online testing system. All online test content has been developed to scale in relation to the available area on larger monitors while maintaining the correct aspect ratio. Specific recommendations on how to transcribe items into braille were provided by the braille publisher to produce the braille version of the LEAP 2025 assessments and the test administrator's notes that accompany the braille forms. The goal was to maximize the number of items on the braille form so it would be possible to transcribe all items into braille.

The following assessment features were available to all students and do not require any documentation either prior to or during the assessment: blank scratch paper and graph paper, calculators (to be used in the calculator section only), color overlay, contrasting colors/reverse colors, directions in native language, equation builder, bookmark, general administration directions clarified, general administration directions read aloud and repeated as necessary, general masking, headphones, highlighters, line guides, magnifiers/variable zoom, measurement tools, redirect student to the test, specialized furniture or equipment, sticky note/notepad, strikethrough, and writing/formatting tools (for writing items only).

The following accessibility features were available for all students with the particular need documented in their IEP, IAP, EL plan, or PNP: individual testing, small group testing, student reads assessment aloud to himself or herself, adaptive and specialized equipment or furniture, and math read aloud (text-to-speech, Kurzweil, recorded voice, or human reader).

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

Copyright © 2019 by Louisiana Department of Education.

3.5 Item and Task Specifications

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

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

The item and task specifications are designed to ensure that the assessment items measure the assessment's claims. The purpose of the item and task specifications is to define the characteristics of the items and tasks that will provide the evidence to support one or more claims. To do this, the item and task specifications delineate the types of evidence, or targets, that should be elicited for each reporting category within a grade level. Then, the specifications provide explicit guidance on how to write items to elicit the desired evidence. To address LEAP 2025 assessment comparability goals with PARCC 2018, 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 reporting category. 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 underscores DRC's efforts to select items that are accessible to the widest range of students possible; in other words, 2018 LEAP 2025 items were selected according to the elements of universal design.

3.6 Summary

In summary, the overall purpose of this chapter is to explicate the procedures used in the development of the 2018 LEAP 2025 grade-level assessments. The efforts by LDOE and DRC in developing the LEAP 2025 assessments are in alignment with multiple best practices of the test industry but, in particular, support the following AERA, APA, & NCME (2014) standards:

> Standard 3.1 Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (63)
>
> Standard 3.2 Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests being affected by construct-

Copyright © 2019 by Louisiana Department of Education.

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)

Copyright © 2019 by Louisiana Department of Education.

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 2018 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.

To ensure that the LEAP 2025 assessments are administered in accordance with the department's mandates, the LDOE takes a primary role in communicating with and training school system personnel. The development of the assessments is a collaborative effort between LDOE and DRC. The LDOE conveys to school systems the purpose of the assessments and the importance of test administration being consistent with test industry standards. The tests and 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, the LDOE provides train-the-trainer opportunities for the school system test coordinators, who, in turn, administer test-administration training to schools within their school systems. The LDOE conducts quality assurance visits during testing to ensure 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 the LDOE and their school systems. The LDOE also provides assistance with and interpretation of assessment data and test results.

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

For the Spring 2018 administration of the LEAP 2025 assessments, DRC produced the following administration manuals: *LEAP 2025 Grades 3 – 4 Paper-Based Test Administration Manual* and *LEAP 2025 Grades 3 – 8 Computer-Based Test Administration Manual* (TAMs). DRC also produced the following district Test Coordinators Manuals: *LEAP 2025 Computer-Based Test Coordinators Manual* and *LEAP 2025 Paper-Based Test Coordinators Manual* (TCMs). LDOE assessment administration and development staff review these manuals, provide feedback, and give final approval. The TCM includes ELA, mathematics, social studies, and science in grades 3

Copyright © 2019 by Louisiana Department of Education.

through 8. It provides detailed instructions for school system and school test coordinators' responsibilities for distributing and collecting test materials 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

Copyright © 2019 by Louisiana Department of Education.

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

Paper Administration Table of Contents

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 3–4
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

Copyright © 2019 by Louisiana Department of Education.

10.6. English Learner (EL) Students
11. Hand-coded Consumable Test Booklets and Answer Documents
12. Students Absent from Testing
13. Consumable Test Booklet and Answer Document Coding
13.1. Coding the Demographic Section
13.2. Sample Grade 3 English Language Arts Consumable Test Booklet
14. General Instructions for English Language Arts and Mathematics
14.1. Student Marking/Erasing on Consumable Test Booklet
14.2. Reading Directions to Students
14.3. Special Instructions
15. Directions for Administering LEAP 2025: English Language Arts and Mathematics
16. General Instructions for Science
16.1. Student Marking/Erasing on Answer Document
16.2. Reading Directions to Students
16.3. Special Instructions
17. Directions for Administering LEAP/*i*LEAP: Science
18. Post-test Procedures
18.1. Test Administrator Oath of Security and Confidentiality Statement
18.2. Science Test Booklets
18.3. Used and Unused Answer Documents and Consumable Test Booklets (Defined)
18.4. Transferring Student Responses
18.5. Returning Test Materials to the School Test Coordinator
19. Index

Online 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. Reporting Testing Irregularities and Security Breaches
4.4. 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

Copyright © 2019 by Louisiana Department of Education.

9.4. Testing Times for Grades 3–8
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. English Learner (EL) Students
11. Students Absent from Testing
12. Directions for Administering the Grades 3–8 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.

Copyright © 2019 by Louisiana Department of Education.

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

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

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

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

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

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

The test administration manuals present instructions for post-test activities to ensure that online tests are submitted and printed test materials are handled properly to maintain the integrity of student information and test scores. Detailed instructions guide test examiners in submitting all online test records. For students who were administered a large-print or braille test form, examiners are instructed to transcribe students' responses from the large-print test or braille test form into 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

Copyright © 2019 by Louisiana Department of Education.

of test security are so noted. Detailed information about test security procedures are presented under “Test Security” in the test administration manuals.

4.1.1 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.

4.1.2 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 secure test materials to DRC, district test coordinators are required to complete and submit a materials accountability form that details the number of consumable test booklets or secure accommodated test materials returned. This materials accountability form also requires that school systems document nonstandard situations, including lost, damaged, destroyed, extra, or missing test books. This form ensures all materials are accounted for. Any material not accounted for on this form is place on a missing materials list which is used by DRC and LDOE to follow up with all districts to ensure security of all materials. A sample accountability form is shown in Figure 4.1.

Copyright © 2019 by Louisiana Department of Education.

Figure 4.1 Sample Accountability Form

Administration District School (All)

Show

Enter Counts | Summary | Status Report

Accountability Form Data for District 999 has been completed. You may continue making changes through the end of the accountability form window.

Reference the *Instructional Text* below for the reasons for any return material discrepancies.

Instructions

This form may be updated throughout the testing window, but it MUST be completed by the end of the testing window when all materials have been returned to Data Recognition Corporation.

All secure materials received from Data Recognition Corporation should be included in the box counts provided in the "Returned to DRC" column.

Any secure documents (test booklets, answer documents, or consumable test booklets) soiled with bodily fluids must be listed in the"Record reasons for discrepancies here:" field to ensure they are not reported as missing materials. Always provide both the security barcode number AND the date the document was destroyed.

Accountability Form for

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

Accountability Form for

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

Record reasons for discrepancies here:

Enter Counts | Summary | Status Report

Instructions

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

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

Copyright © 2019 by Louisiana Department of Education.

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

Copyright © 2019 by Louisiana Department of Education.

4.1.3 Interpretive Guides

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

4.2 Test Security Measures

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

4.2.1 Data Forensic Analyses

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

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

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

4.2.2 Response Change Analysis

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

4.2.3 Score Fluctuation Analysis

Copyright © 2019 by Louisiana Department of Education.

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

4.2.4 **Web Monitoring**

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

4.2.5 **Plagiarism Detection**

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

4.3 Test Administration

The 2018 assessments were administered to students within the state testing window of April 9 through May 4, 2018. The paper testing window was April 30 through May 4, 2018. Each session of the assessment within each content area of the LEAP 2025 assessments was required to be administered in one block of time.

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

Copyright © 2019 by Louisiana Department of Education.

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

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

4.3.2 **Accommodations**

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

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

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

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

Copyright © 2019 by Louisiana Department of Education.

Tables 4.2 through 4.5 summarize the numbers of reportable students receiving accommodations by accommodation type for the 2018 LEAP 2025. Accommodation assignment guidance is provided in the LEAP 2025 Accommodations and Accessibility User Guide. Accommodations are grouped into four sections: special education accommodation, English learner status accommodation, Section 504 status accommodation, and online accommodation. The analyses are based on census data and the number includes only those students who received accommodations and received a scale score on the ELA or mathematics LEAP 2025 assessments. The percentage represents the percentage of the census population receiving that accommodation.

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	≥2,130	3.93%	≥2,110	3.9%
3	Braille	<50	NR	<50	NR
3	Large Print	<50	NR	<50	NR
3	Answers Recorded	≥570	1.06%	≥570	1.06%
3	Extended Time	≥4,770	8.79%	≥4,730	8.71%
3	Transferred Answers	≥210	0.39%	≥210	0.4%
3	Individual/Small Group Administration	≥4,690	8.63%	≥4,650	8.57%
3	Tests Read Aloud	≥3,420	6.31%	≥3,900	7.19%
4	No Accommodation	≥2,250	4.46%	≥2,230	4.42%
4	Braille	<50	NR	<50	NR
4	Large Print	<50	NR	<50	NR
4	Answers Recorded	≥490	0.97%	≥480	0.96%
4	Extended Time	≥4,680	9.25%	≥4,680	9.28%
4	Transferred Answers	≥240	0.48%	≥240	0.49%
4	Individual/Small Group Administration	≥4,500	8.91%	≥4,510	8.94%
4	Tests Read Aloud	≥3,270	6.48%	≥3,810	7.54%

Copyright © 2019 by Louisiana Department of Education.

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

EL Accommodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Number	Percentage	Number	Percentage
3	No Accommodation	≥530	0.98%	≥500	0.92%
3	Extended Time	≥1,950	3.6%	≥1,920	3.55%
3	Individual/Small Group Administration	≥1,620	2.99%	≥1,590	2.94%
3	English/Native Language Word-to-Word Dictionary	≥300	0.55%	≥270	0.51%
3	Directions Read Aloud/Clarified in Native Language	≥90	0.17%	<50	NR
4	No Accommodation	≥380	0.76%	≥330	0.66%
4	Extended Time	≥1,630	3.22%	≥1,620	3.21%
4	Individual/Small Group Administration	≥1,300	2.59%	≥1,300	2.59%
4	English/Native Language Word-to-Word Dictionary	≥370	0.74%	≥360	0.72%
4	Directions Read Aloud/Clarified in Native Language	≥130	0.26%	<50	NR

Copyright © 2019 by Louisiana Department of Education.

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	≥370	0.68%	≥4,330	7.98%
3	Large Print	<50	NR	<50	NR
3	Answers Recorded	≥120	0.24%	≥130	0.24%
3	Extended Time	≥3,880	7.14%	≥3,880	7.15%
3	Transferred Answers	<50	NR	<50	NR
3	Individual/Small Group Administration	≥3,280	6.04%	≥3,250	6.00%
3	Tests Read Aloud	≥1,430	2.63%	≥60	0.13%
4	No Accommodation	≥370	0.73%	≥4,790	9.48%
4	Large Print	<50	NR	<50	NR
4	Answers Recorded	≥120	0.24%	≥105	0.21%
4	Extended Time	≥4,330	8.58%	≥4,340	8.6%
4	Transferred Answers	≥60	0.12%	≥70	0.14%
4	Individual/Small Group Administration	≥3,580	7.08%	≥3,590	7.11%
4	Tests Read Aloud	≥1,520	3.00%	≥70	0.15%

Copyright © 2019 by Louisiana Department of Education.

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
3	Text-to-Speech	≥50	5.58%	≥150	14.37%
3	Human Read Aloud	<50	NR	<50	NR
3	Native Language Word-to-Word Dictionary	<50	NR	<50	NR
3	Directions in Native Language	<50	NR	<50	NR
3	Transferred Answers	<50	NR	<50	NR
3	Answers Recorded	<50	NR	<50	NR
3	Extended Time	≥150	15.01%	≥160	15.90%
3	Individual/Small Group Administration	≥160	16.07%	≥170	16.76%
4	Text-to-Speech	≥430	8.50%	≥930	18.17%
4	Human Read Aloud	≥160	3.17%	≥230	4.49%
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	≥100	2.05%	≥100	2.07%
4	Extended Time	≥1,170	22.78%	≥1,180	22.90%
4	Individual/Small Group Administration	≥1,040	20.26%	≥1,050	20.40%
5	Text-to-Speech	≥5,290	9.56%	≥8,540	15.48%
5	Human Read Aloud	≥1,280	2.32%	≥1,540	2.80%
5	Native Language Word-to-Word Dictionary	≥380	0.70%	≥350	0.64%
5	Directions in Native Language	≥80	0.16%	≥60	0.13%
5	Transferred Answers	≥200	0.37%	≥200	0.37%
5	Answers Recorded	≥460	0.84%	≥460	0.84%
5	Extended Time	≥11,870	21.46%	≥11,860	21.49%
5	Individual/Small Group Administration	≥9,770	17.67%	≥9,750	17.67%
6	Text-to-Speech	≥5,110	9.69%	≥7,490	14.23%
6	Human Read Aloud	≥660	1.26%	≥800	1.52%
6	Native Language Word-to-Word Dictionary	≥490	0.94%	≥410	0.78%
6	Directions in Native Language	≥100	0.20%	≥60	0.12%
6	Transferred Answers	≥110	0.21%	≥110	0.22%
6	Answers Recorded	≥240	0.46%	≥240	0.46%
6	Extended Time	≥10,980	20.8%	≥10,910	20.73%
6	Individual/Small Group Administration	≥8,070	15.28%	≥8,010	15.22%

Copyright © 2019 by Louisiana Department of Education.

Online Accommodation Type (continued)					
		English Language Arts		Mathematics	
Grade	Accommodation	Number	Percentage	Number	Percentage
7	Text-to-Speech	≥4,940	9.59%	≥7,070	13.77%
7	Human Read Aloud	≥520	1.02%	≥630	1.23%
7	Native Language Word-to-Word Dictionary	≥600	1.18%	≥510	1.01%
7	Directions in Native Language	≥90	0.18%	≥50	0.10%
7	Transferred Answers	≥90	0.18%	≥90	0.18%
7	Answers Recorded	≥170	0.35%	≥180	0.35%
7	Extended Time	≥10,620	20.62%	≥10,560	20.54%
7	Individual/Small Group Administration	≥7,230	14.03%	≥7,190	13.99%
8	Text-to-Speech	≥4,590	9.01%	≥6,630	14.76%
8	Human Read Aloud	≥490	0.97%	≥580	1.31%
8	Native Language Word-to-Word Dictionary	≥650	1.28%	≥550	1.23%
8	Directions in Native Language	≥120	0.25%	≥80	0.18%
8	Transferred Answers	≥80	0.17%	≥80	0.18%
8	Answers Recorded	≥140	0.29%	≥130	0.31%
8	Extended Time	≥10,160	19.92%	≥9,850	21.93%
8	Individual/Small Group Administration	≥6,770	13.28%	≥6,580	14.66%

4.4 Summary

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

Copyright © 2019 by Louisiana Department of Education.

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)

Copyright © 2019 by Louisiana Department of Education.

CHAPTER 5: CONSTRUCTED-RESPONSE AND TECHNOLOGY-ENHANCED SCORING

In this chapter, the scoring process used for the 2018 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. Handscoring and AI processing rules are detailed in Appendix D. Five ELA items across grades 5-8 ELA (noted in the table below) were Artificial Intelligence (AI) scored by Pearson's Intelligent Essay Assessor (IEA) using scoring models previously developed by Pearson. Human scorers provided second reads to 10% of these responses as well as handscoring supervisory reviews of all responses that IEA was not able to score.

Table 5.1 Constructed-response Scoring

Course	Handscoring Only	AI Scoring	AI Vendor
ELA grade 3	Q7, Q12	N/A	
ELA grade 4	Q7, Q20	N/A	
ELA grade 5	Q9	Q14	Pearson
ELA grade 6	Q9	Q14	Pearson
ELA grade 7	N/A	Q9, Q14	Pearson
ELA grade 8	Q20	Q7	Pearson
Math grades 3-8	All CRs	N/A	

Copyright © 2019 by Louisiana Department of Education.

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.

The Recruitment and Interview Process

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

All readers hired by DRC to score 2018 LEAP 2025 ELA and mathematics test 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 New Meridian and conveyed by Pearson described in the following two paragraphs:

Copyright © 2019 by Louisiana Department of Education.

> 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 student responses.
>
> 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 student responses.

5.1.2 **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.

5.1.3 **Handscoring Training Process**

Standard 6.9 specifies:

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

Copyright © 2019 by Louisiana Department of Education.

Training Material Development

DRC scoring supervisors trained scorers using PARCC-approved training materials provided by PARCC. These materials were developed according to processes described in PARCC Technical Reports and include 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. LDOE visits the scoring centers to review training materials and oversee the training process. An explanation of the training and qualification procedures follows.

DRC used the 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.[2] Readers who did not qualify on the prototype item training were not allowed to continue the training.

After qualifying on the associated prototype item training, 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.

---

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

Copyright © 2019 by Louisiana Department of Education.

Table 5.2 Mathematics Training Set Composition

Set Type	Prototype Item Training	Abbreviated Item Training	Annotated
Anchor Set	3 responses per score point (Composite items had 3 responses per composite score.)	3 responses per score point (Composite items had 3 responses per composite score.)	Yes
Practice Set 1	10 responses representing the range of responses	10 responses representing the range of responses	Yes
Practice Set 2	10 responses representing the range of responses	10 responses representing the range of responses	Yes
Qualifying Set 1	10 responses comparable to the anchor set responses		No
Qualifying Set 2	10 responses comparable to the anchor set responses		No
Qualifying Set 3	10 responses comparable to the anchor set responses		No

Copyright © 2019 by Louisiana Department of Education.

Table 5.3 ELA Training Set Composition

Set Type	Prototype Item Training	Abbreviated Item Training	Annotated
Anchor Set*	3 responses per score point	16 responses per item: • Anchor Sets for abbreviated RST and LAT item training included scores for the combined trait Reading Comprehension and Written Expression (RCWE). • 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 • the Reading Comprehension and Written Expression (RCWE) trait (for LAT and RST items) • the Written Expression trait (for NWT items)	10 responses representing the range of responses for 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.

Copyright © 2019 by Louisiana Department of Education.

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

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

Table 5.4 Mathematics Qualifying Standards

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

Table 5.5 Mathematics Qualifying Standards (Composite Items)*

Composite (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 5.6 ELA Qualifying Standards

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

ELA readers were required to meet all three of the qualifications listed in Table 5.5. 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).

5.1.4 **Monitoring the Scoring Process**

Standard 6.8 states:

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

Copyright © 2019 by Louisiana Department of Education.

Section 5.1.4 explains the monitoring procedures that DRC uses to ensure that handscoring evaluators follow established scoring criteria while items are being scored. Detailed scoring rubrics, which specify the criteria for scoring, are available for 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 concerns were apparent among individual scorers, team leaders dealt with those issues on an individual basis. If a scorer appeared to need clarification of the scoring rules, DRC supervisors typically monitored one out of five of the scorer's readings, making adjustments to that ratio as needed. If a supervisor disagreed with a reader's scores during monitoring, 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

Copyright © 2019 by Louisiana Department of Education.

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

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

Table 5.7 Expectations for Validity and Inter-Rater Reliability

Agreement Rate Requirements for Validity and Inter-Rater Reliability			
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. In addition to the Reader Feedback Logs, DRC provides LDOE with handscoring quality control reports for review throughout the scoring window. Further detail about these reports can be found in Appendix D.

Copyright © 2019 by Louisiana Department of Education.

5.2 Inter-Rater Reliability

A minimum of 10% of the constructed responses in ELA and mathematics were scored independently by a second reader. This is the case regardless of the first reader being human or AI. 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 first and second score 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 5.8 for ELA items, Table 5.9 for mathematics items, and Table 5.10 for Spanish mathematics items.

As shown in Table 5.8, raters demonstrated at least 97% perfect and adjacent agreement for all ELA handscored items. As shown in Table 5.9, raters demonstrated at least 96% perfect and adjacent agreement for mathematics items. As shown in Table 5.10, raters demonstrated 100% perfect and adjacent agreement for Spanish mathematics items.

Copyright © 2019 by Louisiana Department of Education.

Table 5.8 Inter-Rater Agreement, English Language Arts Items

Grade	Task Type	Question	Trait	Total Reads	Read 2x	Inter-Rater Reliability %		
						EX	AD	EX + AD
3	Research Simulation	7	Reading Comprehension and Written Expression	≥63,090	≥14,910	78	21	99
			Knowledge of Language and Conventions	≥63,090	≥14,910	77	22	99
3	Narrative Writing	12	Written Expression	≥62,260	≥13,240	80	19	99
			Knowledge of Language and Conventions	≥62,260	≥13,240	77	22	99
4	Literary Analysis	7	Reading Comprehension and Written Expression	≥62,120	≥12,190	83	17	100
			Knowledge of Language and Conventions	≥62,120	≥12,190	81	19	100
4	Research Simulation	20	Reading Comprehension and Written Expression	≥62,300	≥12,550	75	25	100
			Knowledge of Language and Conventions	≥62,300	≥12,550	75	25	100
5	Research Simulation	9	Reading Comprehension and Written Expression	≥61,150	≥11,550	71	27	98
			Knowledge of Language and Conventions	≥61,150	≥11,550	69	29	98
5	Narrative Writing (AI)	14	Written Expression	≥61,310	≥11,700	75	25	100
			Knowledge of Language and Conventions	≥61,310	≥11,700	68	30	98* (na =1)
6	Research Simulation	9	Reading Comprehension and Written Expression	≥58,640	≥11,400	69	28	97* (na =2)
			Knowledge of Language and Conventions	≥58,640	≥11,400	69	29	98
6	Narrative Writing (AI)	14	Written Expression	≥58,770	≥11,760	74	25	99
			Knowledge of Language and Conventions	≥58,770	≥11,760	71	28	99
7	Research Simulation (AI)	9	Reading Comprehension and Written Expression	≥57,550	≥11,790	68	31	99
			Knowledge of Language and Conventions	≥57,550	≥11,790	66	32	98
7	Narrative Writing (AI)	14	Written Expression	≥57,320	≥11,530	73	26	99* (na =2)
			Knowledge of Language and Conventions	≥57,320	≥11,530	70	29	99
8	Literary Analysis (AI)	7	Reading Comprehension and Written Expression	≥57,030	≥12,090	73	26	99
			Knowledge of Language and Conventions	≥57,030	≥12,090	76	24	100
8	Research Simulation	20	Reading Comprehension and Written Expression	≥56,370	≥10,680	69	29	98
			Knowledge of Language and Conventions	≥56,370	≥10,680	73	26	99

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

Copyright © 2019 by Louisiana Department of Education.

Table 5.9 Inter-Rater Agreement, Mathematics Items

Grade	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
3	17	Part A (CBT)	≥1,190	≥240	95	5	100
		Part B (CBT)	≥1,190	≥240	94	6	100
3	17	Part A (PBT)	≥60,320	≥11,550	88	8	100
		Part B (PBT)	≥60,320	≥11,550	95	5	100
3	18	N/A	≥61,500	≥11,820	90	10	100
3	32	Part C (CBT)	≥1,210	≥270	95	5	100
3	32	Part C (PBT)	≥60,340	≥11,590	92	8	100
3	33	N/A	≥61,310	≥11,450	93	7	100
3	48	N/A	≥61,390	≥11,570	88	12	100
3	49	Part A	≥61,320	≥11,420	90	10	100
		Part B	≥61,320	≥11,420	95	5	100
		Part C	≥61,320	≥11,420	93	7	100
4	17	Part C (CBT)	≥5,830	≥1,230	96	4	100
4	17	Part C (PBT)	≥56,150	≥10,770	95	5	100
4	18	N/A	≥61,740	≥11,700	95	5	100
4	32	N/A	≥61,900	≥11,930	91	8	99
4	33	N/A	≥61,820	≥11,750	89	10	99
4	48	Part B (CBT)	≥5,850	≥1,290	94	6	100
4	48	Part B (PBT)	≥56,310	≥11,080	95	5	100
4	49	Part C (CBT)	≥5,720	≥1,080	96	4	100
4	49	Part C (PBT)	≥56,110	≥10,690	95	4	99* (na = 0)
5	17	Part A	≥60,850	≥11,180	89	11	100
		Part B	≥60,850	≥11,180	88	11	99* (na = 0)
5	18	N/A	≥59,660	≥11,460	93	6	99
5	32	Part A	≥60,970	≥11,440	85	15	100
		Part B	≥60,970	≥11,440	94	5	99* (na = 0)
5	33	N/A	≥60,610	≥11,710	91	9	100
5	48	Part B	≥61,030	≥11,260	91	9	100

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

Copyright © 2019 by Louisiana Department of Education.

Table 5.9 Inter-Rater Agreement, Mathematics Items, continued

Grade	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
6	31	Part A	≥57,940	≥11,280	82	17	99* (na = 0)
		Part B	≥57,940	≥11,280	93	7	100
6	35	N/A	≥57, 680	≥11,570	91	8	99* (na = 0)
6	36	Part A	≥57,390	≥11,090	89	7	96
		Part B	≥57,390	≥11,090	93	5	98
6	37	N/A	≥56,130	≥11,800	93	7	100
6	47	N/A	≥58,060	≥11,440	74	22	96
6	48	Part A	≥58,110	≥11,070	85	15	100
		Part B	≥58,110	≥11,070	90	10	100
6	49	Part A	≥57,600	≥11,530	91	8	99
7	30	Part A	≥56,480	≥10,560	96	4	100
		Part B	≥56,480	≥10,560	96	4	100
7	35	Part A	≥56,670	≥11,050	96	4	100
		Part B	≥56,670	≥11,050	98	2	100
7	36	N/A	≥55,570	≥11,120	97	3	100
7	37	N/A	≥54,500	≥11,080	96	3	99* (na = 0)
7	47	N/A	≥56,410	≥11,210	94	5	99* (na = 0)
7	48	Part A	≥56,450	≥10,880	98	2	100
		Part B	≥56,450	≥10,880	98	2	100
7	49	N/A	≥55,760	≥11,490	95	4	99* (na = 0)
8	29	Part A	≥49,310	≥9,510	90	10	100
		Part B	≥49,310	≥9,510	91	8	100
8	34	N/A	≥48,660	≥10,180	92	8	100
8	35	N/A	≥48,830	≥10,020	92	8	100
8	36	N/A	≥48,480	≥9,870	96	4	100
8	46	N/A	≥49,280	≥10,080	94	6	100
8	47	Part B	≥49,700	≥9,050	96	4	100

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

Copyright © 2019 by Louisiana Department of Education.

Table 5.10 Inter-Rater Agreement, Spanish Mathematics Items

Grade	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
3	17	Part A (CBT)	<10	<10	NR	NR	NR
		Part B (CBT)	<10	<10	NR	NR	NR
3	17	Part A (PBT)	≥50	≥20	100	0	100
		Part B (PBT)	≥50	≥20	100	0	100
3	18	N/A	≥50	≥20	90	10	100
3	32	Part C (CBT)	<10	<10	NR	NR	NR
3	32	Part C (PBT)	≥50	≥10	100	0	100
3	33	N/A	≥50	≥10	88	13	101*
3	48	N/A	≥50	≥10	100	0	100
3	49	Part A	≥50	≥10	100	0	100
		Part B	≥50	≥10	100	0	100
		Part C	≥50	≥10	100	0	100
4	17	Part C (CBT)	<10	<10	NR	NR	NR
4	17	Part C (PBT)	≥50	≥10	100	0	100
4	18	N/A	≥50	≥10	100	0	100
4	32	N/A	≥50	≥10	100	0	100
4	33	N/A	≥50	≥10	100	0	100
4	48	Part B (CBT)	<10	<10	NR	NR	NR
4	48	Part B (PBT)	≥50	≥10	100	0	100
4	49	Part C (CBT)	<10	<10	NR	NR	NR
4	49	Part C (PBT)	≥40	<10	NR	NR	NR
5	17	Part A	≥60	≥20	100	0	100
		Part B	≥60	≥20	100	0	100
5	18	N/A	≥60	≥10	100	0	100
5	32	Part A	≥60	≥20	100	0	100
		Part B	≥60	≥20	90	10	100
5	33	N/A	≥60	≥20	100	0	100
5	48	Part B	≥60	≥10	86	14	100

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

Copyright © 2019 by Louisiana Department of Education.

Table 5.100 Inter-Rater Agreement, Spanish Mathematics Items, continued

Grade	Question	Part(s)	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
6	31	Part A	≥120	≥20	100	0	100
		Part B	≥120	≥20	100	0	100
6	35	N/A	≥130	≥30	100	0	100
6	36	Part A	≥120	≥20	100	0	100
		Part B	≥120	≥20	100	0	100
6	37	N/A	≥130	≥30	100	0	100
6	47	N/A	≥130	≥40	100	0	100
6	48	Part A	≥130	≥30	100	0	100
		Part B	≥130	≥30	100	0	100
6	49	N/A	≥130	≥30	100	0	100
7	30	Part A	≥120	≥20	100	0	100
		Part B	≥120	≥20	100	0	100
7	35	Part A	≥120	≥20	100	0	100
		Part B	≥120	≥20	100	0	100
7	36	N/A	≥120	≥20	100	0	100
7	37	N/A	≥110	≥20	100	0	100
7	47	N/A	≥120	≥20	100	0	100
7	48	Part A	≥110	≥20	100	0	100
		Part B	≥110	≥20	100	0	100
7	49	N/A	≥120	≥20	100	0	100
8	29	Part A	≥140	≥40	100	0	100
		Part B	≥140	≥40	100	0	100
8	34	N/A	≥140	≥40	95	5	100
8	35	N/A	≥140	≥50	100	0	100
8	36	N/A	≥140	≥40	100	0	100
8	46	N/A	≥140	≥50	100	0	100
8	47	Part B	≥140	≥30	100	0	100

**N/A if an item does not have multiple parts

Copyright © 2019 by Louisiana Department of Education.

5.3 Technology-Enhanced Item Scoring Process

All technology-enhanced items, including EBSR, MPSR, and SA, were processed through DRC's autoscoring engine and scored according to the assigned scoring rules as established during content creation by PARCC or DRC as applicable in conjunction with LDOE. DRC ensured that all rubrics and scoring rules were verified for accuracy before scoring any technology-enhanced items. DRC established an adjudication process for technology-enhanced items and 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 were transcribed into the online system by a test administrator.

5.4 Multiple-Choice and Multiple-Select Item Scoring Process

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

Copyright © 2019 by Louisiana Department of Education.

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)

Copyright © 2019 by Louisiana Department of Education.

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 assessments 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 2018 LEAP 2025 analyses, the Louisiana student population was used. In Section 6.3, the characteristics of calibration samples, such as subgroups, are discussed. Chapter 3 presents the test specifications. Information regarding reported data is discussed in detail in Chapter 7.

In this section, summary test statistics for each form, grade, and content area of LEAP 2025 are presented. These statistics are followed by item-level statistics for each grade and content area of LEAP 2025. These statistics were produced using census data.

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 0.35 to 0.45. Average form difficulty for mathematics ranged from 0.30 to 0.50. In general, the 2018 LEAP 2025 tests were relatively difficult tests across all subjects and grades. For ELA, the grades 3 and 4 computer-based test (CBT) was the most difficult, with 0.35 average form difficulty, and the grades 7 and 8 test were the easiest, with 0.45 average form difficulty. For mathematics, the

Copyright © 2019 by Louisiana Department of Education.

grade 8 test was the most difficult, with 0.30 average form difficulty, and the grade 3 paper-based test (PBT) test was the easiest, with 0.52 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	CBT	27	71	24.28 (12.26)	0.35 (0.16)
	3	PBT	27	71	27.10 (13.13)	0.39 (0.12)
	4	CBT	30	86	29.89 (15.34)	0.35 (0.13)
	4	PBT	30	86	32.60 (15.23)	0.38 (0.12)
	5	CBT	31	83	32.31 (14.60)	0.39 (0.14)
	6	CBT	33	90	36.01 (17.46)	0.40 (0.15)
	7	CBT	33	90	40.67 (18.86)	0.45 (0.12)
	8	CBT	34	94	41.98 (16.83)	0.45 (0.14)
Mathematics	3	CBT	43	62	25.77 (11.85)	0.42 (0.22)
	3	PBT	43	62	30.55 (12.71)	0.50 (0.21)
	4	CBT	43	62	27.22 (12.13)	0.44 (0.23)
	4	PBT	43	62	29.21 (12.36)	0.48 (0.23)
	5	CBT	43	62	25.62 (11.81)	0.41 (0.18)
	6	CBT	43	66	23.75 (13.81)	0.36 (0.19)
	7	CBT	43	66	20.88 (12.87)	0.32 (0.17)
	8	CBT	42	66	19.51 (12.70)	0.30 (0.16)

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

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

Copyright © 2019 by Louisiana Department of Education.

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

				Item *p*-Value				Average Item-Total Correlation			
Content	Grade	Mode	N of Items	Mean	Std. Dev.	Min.	Max	Mean	Std. Dev.	Min.	Max
ELA	3	CBT	27	0.39	0.15	0.17	0.68	0.44	0.13	0.12	0.65
	3	PBT	27	0.42	0.13	0.16	0.67	0.44	0.14	0.11	0.69
	4	CBT	30	0.39	0.14	0.17	0.72	0.48	0.17	0.24	0.78
	4	PBT	30	0.41	0.14	0.17	0.67	0.45	0.17	0.18	0.76
	5	CBT	31	0.42	0.15	0.14	0.79	0.42	0.16	0.18	0.76
	6	CBT	33	0.44	0.16	0.23	0.73	0.48	0.13	0.21	0.74
	7	CBT	33	0.47	0.14	0.24	0.76	0.48	0.15	0.15	0.81
	8	CBT	34	0.46	0.16	0.23	0.89	0.44	0.19	0.10	0.82
Mathematics	3	CBT	43	0.48	0.24	0.11	0.98	0.43	0.14	0.17	0.72
	3	PBT	43	0.56	0.22	0.18	0.98	0.46	0.12	0.18	0.73
	4	CBT	43	0.52	0.22	0.16	0.89	0.45	0.12	0.24	0.70
	4	PBT	43	0.56	0.22	0.19	0.91	0.44	0.12	0.26	0.71
	5	CBT	43	0.46	0.19	0.14	0.78	0.41	0.13	0.18	0.67
	6	CBT	43	0.43	0.20	0.14	0.93	0.47	0.14	0.15	0.70
	7	CBT	43	0.37	0.17	0.07	0.76	0.46	0.13	0.19	0.70
	8	CBT	42	0.34	0.17	0.10	0.68	0.46	0.13	0.13	0.68

6.1.2 Item-Level Statistics

Tables 6.3–6.10 present the item statistics for each operational item included in regular test forms by grade for ELA. Tables 6.11–6.18 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 items with multiple responses (PBT only), omitted responses, responses that were not scored, or an item that has received a non-score code), 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 0.39 in grades 3 and 4 CBT

Copyright © 2019 by Louisiana Department of Education.

forms to 0.47 in grade 7. The average *p*-values associated with the mathematics forms range from 0.34 in grade 8 CBT to 0.56 in grades 3 and 4 PBT.

It is important that one examines the range of *p*-values, not just the average *p*-value, to determine whether a test measures well. It is desirable for 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 0.14 to the 0.89 (see Tables 6.3–6.9) across all grade levels. The *p*-values on the mathematics forms range from the 0.07 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 speediness issues on tests. A test may be speeded if students do not have adequate time to answer all questions on the test. In general, an item is said to have a high omit rate if more than 5% of students failed to respond to the item. Evidence of speededness is considered a threat to validity as student test scores may not reflect their ability. Additionally, content validity may be threatened as the not-reached items are needed to fulfill content blueprint specifications (Lu & Sireci, 2007).

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

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

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

The results presented in Tables 6.3–6.16 show that the omit rates for most of the items on the LEAP 2025 regular forms are less than 5%, suggesting that the majority of students were able to complete the test in the prescribed amount of time. There is not an omit rate higher than 7%, and the omit rates for the last items in the tests do not exceed 4%. These omit rates indicate that 96% of the students completed the test. Lu & Sireci (2007) report that the Education Testing Service

Copyright © 2019 by Louisiana Department of Education.

has used an approach where a test was considered unspeeded if at least 80% of the examinees reach the last item and if everyone reaches at least 75% of the items. The reported omit rates fall within these ranges.

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

ELA Grade 3 Computer-Based Test Administration										
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**
868494	ESR	≥1,030	≥1,030	0.54	0.50	0.29	29.84	31.28	38.59	
868496	ESR	≥1,030	≥1,030	0.39	0.42	0.48	50.82	19.63	29.07	
868495	ESR	≥1,030	≥1,030	0.17	0.21	0.48	75.17	14.82	9.53	
868497	ESR	≥1,030	≥1,030	0.36	0.46	0.67	46.78	32.63	19.92	
945390	TE	≥1,030	≥1,030	0.68	0.41	0.58	22.23	19.06	58.13	
937200	ESR	≥1,030	≥1,030	0.31	0.26	0.67	58.52	19.25	21.56	
868500O2	CR	≥1,030	≥1,010	0.17	0.65	1.44	54.57	37.34	5.29	0.19
868500O3	CR	≥1,030	≥1,010	0.22	0.61	1.44	45.43	40.71	10.20	1.06
936916	MS	≥1,030	≥1,030	0.23	0.37	0.10	64.87	23.20	11.84	
913494	ESR	≥1,030	≥1,030	0.40	0.44	0.10	54.96	10.68	34.26	
913495	TE	≥1,030	≥1,030	0.64	0.48	0.67	14.73	42.64	41.96	
913493	ESR	≥1,030	≥1,030	0.37	0.37	0.38	54.19	16.27	29.16	
913497O2	CR	≥1,030	≥990	0.22	0.61	0.96	43.41	42.64	8.47	1.35
913497O3	CR	≥1,030	≥990	0.29	0.60	0.96	28.97	52.26	12.80	1.83
913514	ESR	≥1,030	≥1,030	0.51	0.42	0.10	32.63	33.11	34.17	
913516	ESR	≥1,030	≥1,030	0.38	0.33	0.29	52.17	20.12	27.43	
913517	TE	≥1,030	≥1,020	0.62	0.48	0.96	23.20	27.91	47.93	
936920	ESR	≥1,030	≥1,030	0.43	0.48	0.38	49.09	15.98	34.55	
934821	ESR	≥1,030	≥1,030	0.33	0.28	0.10	56.21	21.17	22.52	
934823	ESR	≥1,030	≥1,030	0.54	0.39	0.10	21.08	49.28	29.55	
934822	TE	≥1,030	≥1,030	0.59	0.53	0.38	28.59	24.64	46.39	
934802	ESR	≥1,030	≥1,030	0.52	0.48	0.38	33.88	27.24	38.50	
870596	ESR	≥1,030	≥1,030	0.45	0.46	0.48	50.53	8.66	40.33	
870597	ESR	≥1,030	≥1,030	0.39	0.38	0.67	51.68	17.13	30.51	
870594	ESR	≥1,030	≥1,030	0.19	0.12	0.48	75.26	11.26	12.99	
870595	ESR	≥1,030	≥1,030	0.46	0.51	0.58	44.66	18.19	36.57	

Copyright © 2019 by Louisiana Department of Education.

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

ELA Grade 3 Paper-Based Test Administration										
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**
868494	ESR	≥54,350	≥54,110	0.60	0.49	0.44	24.30	30.69	44.57	
868496	ESR	≥54,350	≥54,100	0.44	0.39	0.46	46.22	18.82	34.49	
868495	ESR	≥54,350	≥54,060	0.17	0.30	0.54	75.87	13.77	9.83	
868497	ESR	≥54,350	≥54,060	0.42	0.44	0.53	40.78	33.50	25.19	
868498	ESR	≥54,350	≥53,870	0.45	0.45	0.87	50.30	8.30	40.53	
937200	ESR	≥54,350	≥53,720	0.35	0.35	1.15	55.04	19.16	24.65	
868500P2	CR	≥54,350	≥53,560	0.29	0.69	1.04	34.19	45.13	16.87	2.36
868500P3	CR	≥54,350	≥53,560	0.36	0.64	1.04	23.08	46.84	25.28	3.34
936916	MS	≥54,350	≥54,060	0.30	0.37	0.53	59.53	20.39	19.55	
913494	ESR	≥54,350	≥53,920	0.47	0.39	0.78	47.82	9.74	41.66	
913496	ESR	≥54,350	≥54,090	0.67	0.55	0.47	27.52	10.85	61.16	
913493	ESR	≥54,350	≥54,000	0.41	0.36	0.64	53.40	11.10	34.86	
913497P2	CR	≥54,350	≥52,980	0.28	0.59	0.76	31.29	50.83	13.82	1.54
913497P3	CR	≥54,350	≥52,980	0.36	0.59	0.76	16.22	59.03	20.68	1.56
913514	ESR	≥54,350	≥53,020	0.54	0.44	2.45	25.07	39.56	32.93	
913516	ESR	≥54,350	≥52,820	0.41	0.39	2.81	47.61	20.02	29.55	
936923	ESR	≥54,350	≥53,080	0.31	0.25	2.33	56.49	22.29	18.90	
936920	ESR	≥54,350	≥52,840	0.50	0.51	2.78	41.11	14.52	41.59	
934821	ESR	≥54,350	≥53,940	0.35	0.25	0.75	54.95	18.94	25.36	
934823	ESR	≥54,350	≥53,850	0.61	0.38	0.92	14.86	48.44	35.77	
934806	ESR	≥54,350	≥54,080	0.45	0.44	0.50	52.88	4.58	42.04	
934802	ESR	≥54,350	≥53,940	0.60	0.52	0.75	28.85	22.26	48.14	
870596	ESR	≥54,350	≥53,600	0.49	0.46	1.37	45.33	10.40	42.91	
870597	ESR	≥54,350	≥53,120	0.45	0.41	2.26	45.98	15.14	36.62	
870594	ESR	≥54,350	≥53,110	0.16	0.11	2.28	78.57	7.41	11.74	
870595	ESR	≥54,350	≥52,640	0.52	0.51	3.15	38.02	17.18	41.65	

Copyright © 2019 by Louisiana Department of Education.

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

ELA Grade 4 Computer-Based Test Administration											
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
913561	ESR	≥5,160	≥5,160	0.58	0.45	0.02	33.15	18.27	48.56		
913562	TE	≥5,160	≥5,150	0.72	0.56	0.25	14.63	26.90	58.22		
913563	ESR	≥5,160	≥5,160	0.48	0.42	0.12	41.47	20.81	37.60		
946024	TE	≥5,160	≥5,150	0.25	0.44	0.17	63.25	23.42	13.16		
913564	ESR	≥5,160	≥5,160	0.55	0.49	0.06	39.75	10.45	49.74		
913566	MS	≥5,160	≥5,160	0.45	0.46	0.12	42.83	25.02	32.03		
913567O2	CR	≥5,160	≥5,110	0.26	0.78	0.31	29.59	40.66	25.20	3.29	0.17
913567O3	CR	≥5,160	≥5,110	0.34	0.78	0.31	29.84	40.51	24.99	3.58	
868527	ESR	≥5,160	≥5,140	0.41	0.37	0.35	52.31	12.23	35.11		
868546	TE	≥5,160	≥5,130	0.54	0.30	0.58	16.59	57.94	24.89		
868547	ESR	≥5,160	≥5,130	0.28	0.27	0.68	59.98	22.84	16.51		
868529	ESR	≥5,160	≥5,120	0.50	0.49	0.77	42.73	14.13	42.37		
799845	ESR	≥5,160	≥5,160	0.38	0.34	0.04	53.26	17.11	29.59		
799846	ESR	≥5,160	≥5,160	0.29	0.31	0.14	61.89	17.13	20.84		
799847	ESR	≥5,160	≥5,160	0.60	0.46	0.14	22.55	35.24	42.07		
947020	TE	≥5,160	≥5,150	0.18	0.43	0.19	68.57	25.80	5.44		
799849	MS	≥5,160	≥5,160	0.31	0.43	0.12	49.80	38.11	11.98		
799850	ESR	≥5,160	≥5,160	0.17	0.25	0.04	79.93	6.00	14.03		
799851	ESR	≥5,160	≥5,150	0.48	0.33	0.15	39.87	23.98	36.00		
799852	ESR	≥5,160	≥5,160	0.35	0.42	0.12	42.19	45.46	12.23		
802342O2	CR	≥5,160	≥5,110	0.25	0.77	0.37	30.08	40.55	25.22	2.98	0.17
802342O3	CR	≥5,160	≥5,110	0.34	0.77	0.37	29.92	40.57	25.24	3.27	
915315	ESR	≥5,160	≥5,160	0.63	0.51	0.10	24.54	24.95	50.42		
915319	ESR	≥5,160	≥5,150	0.21	0.24	0.19	64.29	28.47	7.04		
915322	ESR	≥5,160	≥5,150	0.37	0.37	0.17	57.87	10.01	31.95		
915325	TE	≥5,160	≥5,150	0.37	0.48	0.23	40.18	45.00	14.59		
915316	ESR	≥5,160	≥5,150	0.37	0.44	0.21	54.25	16.43	29.11		
915317	ESR	≥5,160	≥5,150	0.43	0.51	0.29	48.56	16.57	34.58		

Copyright © 2019 by Louisiana Department of Education.

Table 6.6 Operational Item Statistics—English Language Arts Grade 4 PBT Administration

ELA Grade 4 Paper-Based Test Administration											
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
913561	ESR	≥50,590	≥50,400	0.64	0.43	0.38	27.60	15.69	56.32		
946021	ESR	≥50,590	≥50,300	0.46	0.29	0.58	47.74	12.15	39.53		
913563	ESR	≥50,590	≥50,430	0.48	0.42	0.31	42.42	18.00	39.27		
946023	ESR	≥50,590	≥50,300	0.28	0.36	0.58	62.12	19.69	17.61		
913564	ESR	≥50,590	≥50,400	0.50	0.48	0.38	44.94	9.17	45.51		
913566	MS	≥50,590	≥50,400	0.46	0.44	0.38	42.66	22.52	34.45		
913567P2	CR	≥50,590	≥50,180	0.29	0.76	0.59	26.10	35.18	34.62	3.17	0.12
913567P3	CR	≥50,590	≥50,180	0.39	0.75	0.59	24.98	35.97	34.69	3.55	
868527	ESR	≥50,590	≥49,120	0.43	0.33	2.91	49.82	11.02	36.25		
868528	ESR	≥50,590	≥49,130	0.63	0.54	2.88	33.00	5.26	58.86		
868547	ESR	≥50,590	≥48,950	0.33	0.29	3.25	53.04	24.14	19.57		
868529	ESR	≥50,590	≥48,390	0.63	0.53	4.36	30.47	10.75	54.42		
799845	ESR	≥50,590	≥50,430	0.40	0.34	0.32	52.36	14.77	32.54		
799846	ESR	≥50,590	≥50,200	0.32	0.31	0.78	59.42	16.70	23.10		
799847	ESR	≥50,590	≥50,180	0.66	0.45	0.81	17.41	32.32	49.46		
799848	ESR	≥50,590	≥50,340	0.19	0.18	0.50	78.64	4.37	16.49		
799849	MS	≥50,590	≥50,370	0.31	0.41	0.44	52.06	34.25	13.24		
799850	ESR	≥50,590	≥50,280	0.17	0.24	0.62	79.75	5.25	14.39		
799851	ESR	≥50,590	≥50,340	0.53	0.33	0.50	35.64	23.07	40.79		
799852	ESR	≥50,590	≥50,100	0.38	0.41	0.97	39.41	44.18	15.44		
802342P2	CR	≥50,590	≥50,210	0.33	0.75	0.57	17.09	38.27	39.70	4.07	0.12
802342P3	CR	≥50,590	≥50,210	0.44	0.74	0.57	16.71	38.68	39.66	4.19	
915315	ESR	≥50,590	≥50,150	0.67	0.48	0.87	21.54	21.38	56.22		
915319	ESR	≥50,590	≥49,850	0.22	0.24	1.47	63.09	27.44	8.00		
915322	ESR	≥50,590	≥50,160	0.43	0.40	0.86	52.59	8.54	38.01		
915321	ESR	≥50,590	≥49,920	0.36	0.37	1.32	59.90	7.10	31.69		
915316	ESR	≥50,590	≥50,000	0.40	0.42	1.16	51.43	15.11	32.31		
915317	ESR	≥50,590	≥49,740	0.43	0.45	1.69	49.23	13.51	35.56		

Copyright © 2019 by Louisiana Department of Education.

Table 6.7 Operational Item Statistics—English Language Arts Grade 5 CBT Administration

ELA Grade 5 Computer-Based Test Administration											
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
868583	ESR	≥55,310	≥55,300	0.79	0.38	0.03	17.34	7.48	75.16		
868584	MS	≥55,310	≥55,270	0.47	0.29	0.08	11.92	83.03	4.97		
868585	ESR	≥55,310	≥55,210	0.41	0.32	0.19	47.46	22.22	30.13		
868586	ESR	≥55,310	≥55,150	0.46	0.42	0.30	37.76	32.53	29.41		
868587	ESR	≥55,310	≥55,220	0.48	0.50	0.17	39.70	24.11	36.02		
937909	ESR	≥55,310	≥55,210	0.41	0.34	0.19	48.08	21.15	30.57		
868588	MS	≥55,310	≥55,170	0.47	0.41	0.25	22.54	59.76	17.44		
868589	MS	≥55,310	≥55,140	0.27	0.30	0.31	63.10	18.97	17.62		
868590O2	CR	≥55,310	≥54,990	0.28	0.76	0.34	29.85	37.99	24.00	6.18	1.40
868590O3	CR	≥55,310	≥54,990	0.36	0.76	0.34	30.16	37.64	24.13	7.49	
913660	ESR	≥55,310	≥55,290	0.50	0.33	0.03	30.19	39.88	29.90		
913659	ESR	≥55,310	≥55,210	0.30	0.28	0.19	61.29	17.63	20.89		
913662	ESR	≥55,310	≥55,260	0.43	0.37	0.09	53.44	6.69	39.78		
913661	TE	≥55,310	≥55,260	0.60	0.41	0.10	13.34	53.40	33.17		
913663O2	CR	≥55,310	≥54,580	0.32	0.70	0.17	35.50	34.65	24.11	4.42	
913663O3	CR	≥55,310	≥54,580	0.33	0.68	0.17	32.93	38.55	23.52	3.68	
913665	ESR	≥55,310	≥55,220	0.40	0.40	0.16	48.21	23.27	28.36		
913664	ESR	≥55,310	≥55,280	0.40	0.32	0.06	49.67	20.70	29.57		
913666	TE	≥55,310	≥55,250	0.70	0.40	0.12	9.91	41.00	48.97		
913668	ESR	≥55,310	≥55,220	0.53	0.48	0.17	43.96	5.62	50.25		
913667	MS	≥55,310	≥55,240	0.27	0.34	0.12	59.53	27.21	13.13		
913669	TE	≥55,310	≥55,220	0.26	0.46	0.17	60.32	27.82	11.68		
913617	TE	≥55,310	≥55,280	0.14	0.18	0.06	79.70	13.38	6.86		
913611	MS	≥55,310	≥55,260	0.50	0.46	0.09	38.98	22.09	38.84		
913608	ESR	≥55,310	≥55,260	0.54	0.28	0.10	38.40	14.74	46.76		
913614	TE	≥55,310	≥55,260	0.61	0.24	0.10	21.66	35.40	42.85		
868565	ESR	≥55,310	≥55,160	0.68	0.43	0.27	17.41	29.52	52.80		
868564	MS	≥55,310	≥55,060	0.23	0.29	0.46	60.28	32.95	6.31		
868566	MS	≥55,310	≥55,020	0.35	0.49	0.52	49.16	30.88	19.43		
868567	ESR	≥55,310	≥54,950	0.32	0.32	0.66	62.34	9.87	27.13		

Copyright © 2019 by Louisiana Department of Education.

Table 6.8 Operational Item Statistics—English Language Arts Grade 6 CBT Administration

ELA Grade 6 Computer-Based Test Administration											
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
870615	MS	≥52,810	≥52,800	0.71	0.46	0.03	10.66	36.18	53.14		
937914	ESR	≥52,810	≥52,750	0.65	0.37	0.12	24.10	21.36	54.42		
870617	ESR	≥52,810	≥52,670	0.42	0.47	0.28	48.50	18.06	33.16		
870618	ESR	≥52,810	≥52,700	0.23	0.21	0.21	66.15	20.42	13.22		
870619	ESR	≥52,810	≥52,710	0.56	0.43	0.19	37.44	13.07	49.29		
870620	ESR	≥52,810	≥52,730	0.43	0.38	0.17	41.42	30.09	28.33		
870621	ESR	≥52,810	≥52,670	0.31	0.27	0.27	65.30	6.23	28.20		
892471	ESR	≥52,810	≥52,710	0.50	0.52	0.21	45.47	9.81	44.51		
870622O2	CR	≥52,810	≥52,570	0.29	0.72	0.27	25.08	41.48	25.94	5.87	1.18
870622O3	CR	≥52,810	≥52,570	0.38	0.72	0.27	26.36	40.12	26.23	6.84	
913690	MS	≥52,810	≥52,800	0.40	0.41	0.02	39.95	40.17	19.85		
913691	TE	≥52,810	≥52,750	0.38	0.43	0.12	48.00	28.20	23.69		
913692	MS	≥52,810	≥52,770	0.28	0.35	0.09	56.83	29.73	13.35		
913693	TE	≥52,810	≥52,760	0.24	0.41	0.11	66.17	18.78	14.93		
913694O2	CR	≥52,810	≥52,300	0.25	0.74	0.36	41.31	24.23	25.47	6.27	1.76
913694O3	CR	≥52,810	≥52,300	0.35	0.74	0.36	30.28	38.98	23.75	6.02	
870630	ESR	≥52,810	≥52,700	0.70	0.48	0.21	22.53	14.01	63.24		
870633	ESR	≥52,810	≥52,750	0.57	0.55	0.12	38.40	9.88	51.60		
870634	ESR	≥52,810	≥52,750	0.48	0.47	0.12	48.27	7.01	44.60		
870632	TE	≥52,810	≥52,730	0.37	0.42	0.16	35.20	54.78	9.85		
870636	ESR	≥52,810	≥52,700	0.60	0.59	0.21	35.74	7.34	56.70		
870635	ESR	≥52,810	≥52,710	0.64	0.58	0.20	30.82	10.67	58.31		
913752	ESR	≥52,810	≥52,790	0.39	0.44	0.05	52.53	17.15	30.28		
913753	TE	≥52,810	≥52,770	0.69	0.53	0.09	20.77	19.88	59.26		
913754	TE	≥52,810	≥52,740	0.73	0.41	0.13	24.38	4.43	71.06		
913755	ESR	≥52,810	≥52,700	0.39	0.41	0.21	54.33	13.40	32.05		
913757	MS	≥52,810	≥52,730	0.34	0.51	0.16	57.18	18.26	24.40		
913756	MS	≥52,810	≥52,730	0.24	0.34	0.16	67.87	15.90	16.08		
913749	TE	≥52,810	≥52,610	0.65	0.51	0.39	14.05	41.10	44.47		
913748	MS	≥52,810	≥52,540	0.28	0.39	0.53	58.22	26.59	14.66		
913750	MS	≥52,810	≥52,480	0.39	0.45	0.63	37.23	46.85	15.28		
913751	ESR	≥52,810	≥52,440	0.51	0.47	0.70	33.96	28.55	36.78		

Copyright © 2019 by Louisiana Department of Education.

Table 6.9 Operational Item Statistics—English Language Arts Grade 7 CBT Administration

ELA Grade 7 Computer-Based Test Administration											
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
868689	ESR	≥51,540	≥51,530	0.76	0.49	0.03	20.24	8.30	71.43		
868690	ESR	≥51,540	≥51,480	0.57	0.42	0.11	31.91	21.83	46.15		
868691	ESR	≥51,540	≥51,490	0.66	0.50	0.10	28.60	11.66	59.64		
868715	ESR	≥51,540	≥51,480	0.55	0.46	0.12	25.31	39.50	35.07		
868693	ESR	≥51,540	≥51,510	0.55	0.37	0.06	41.29	7.62	51.04		
868696	TE	≥51,540	≥51,480	0.41	0.38	0.11	45.97	26.53	27.39		
868716	ESR	≥51,540	≥51,430	0.24	0.15	0.21	72.88	5.83	21.08		
937929	MS	≥51,540	≥51,480	0.44	0.46	0.12	35.77	39.99	24.12		
868704O2	CR	≥51,540	≥51,100	0.42	0.77	0.28	11.69	31.04	36.43	16.74	3.24
868704O3	CR	≥51,540	≥51,100	0.56	0.76	0.28	11.19	29.76	37.30	20.89	
913840	TE	≥51,540	≥51,510	0.31	0.47	0.06	54.66	27.73	17.55		
913839	ESR	≥51,540	≥51,460	0.62	0.45	0.15	28.19	20.03	51.64		
913841	MS	≥51,540	≥51,500	0.28	0.41	0.08	50.86	41.50	7.56		
913838	TE	≥51,540	≥51,510	0.68	0.57	0.06	18.80	27.23	53.91		
913842O2	CR	≥51,540	≥50,890	0.38	0.81	0.48	35.04	12.46	25.25	17.99	8.00
913842O3	CR	≥51,540	≥50,890	0.48	0.81	0.48	26.48	22.90	29.24	20.11	
913807	ESR	≥51,540	≥51,480	0.63	0.40	0.12	33.78	5.92	60.18		
913808	ESR	≥51,540	≥51,500	0.57	0.49	0.07	33.66	18.72	47.54		
913811	ESR	≥51,540	≥51,490	0.36	0.37	0.09	53.63	21.49	24.79		
913810	ESR	≥51,540	≥51,470	0.31	0.43	0.14	58.75	19.53	21.59		
913812	TE	≥51,540	≥51,480	0.40	0.48	0.11	41.51	36.08	22.30		
913809	TE	≥51,540	≥51,470	0.26	0.45	0.14	58.09	31.97	9.79		
874774	ESR	≥51,540	≥51,510	0.41	0.34	0.06	35.29	46.58	18.07		
874773	ESR	≥51,540	≥51,500	0.42	0.39	0.09	55.95	3.88	40.09		
874776	MS	≥51,540	≥51,470	0.54	0.53	0.14	32.65	26.82	40.40		
874775	MS	≥51,540	≥51,430	0.43	0.45	0.21	38.77	35.35	25.68		
874778	ESR	≥51,540	≥51,460	0.54	0.30	0.16	41.68	9.12	49.04		
874779	MS	≥51,540	≥51,480	0.33	0.31	0.11	54.51	24.17	21.21		
913787	ESR	≥51,540	≥51,400	0.75	0.43	0.26	16.94	16.58	66.22		
913788	TE	≥51,540	≥51,390	0.43	0.48	0.29	45.99	22.28	31.44		
913790	MS	≥51,540	≥51,360	0.32	0.43	0.35	54.67	25.74	19.24		
913791	ESR	≥51,540	≥51,340	0.47	0.55	0.39	43.82	18.40	37.40		

Copyright © 2019 by Louisiana Department of Education.

Table 6.10 Operational Item Statistics—English Language Arts Grade 8 CBT Administration

ELA Grade 8 Computer-Based Test Administration											
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
913952	ESR	≥51,020	≥50,970	0.36	0.30	0.10	55.40	17.45	27.06		
913953	ESR	≥51,020	≥50,930	0.39	0.37	0.18	48.55	25.43	25.84		
913954	MS	≥51,020	≥50,950	0.46	0.43	0.15	47.44	12.99	39.43		
913955	TE	≥51,020	≥50,860	0.65	0.37	0.32	10.04	48.96	40.68		
913956	ESR	≥51,020	≥50,960	0.47	0.41	0.13	48.47	8.97	42.44		
913957	TE	≥51,020	≥50,920	0.39	0.36	0.21	33.03	55.30	11.46		
913958O2	CR	≥51,020	≥50,180	0.37	0.82	0.76	17.68	31.96	35.46	12.12	1.14
913958O3	CR	≥51,020	≥50,180	0.52	0.80	0.76	13.75	30.88	39.69	14.04	
913894	MS	≥51,020	≥50,620	0.55	0.47	0.80	31.27	27.19	40.74		
913898	MS	≥51,020	≥50,430	0.30	0.44	1.16	55.74	27.80	15.30		
913891	TE	≥51,020	≥50,100	0.23	0.31	1.80	63.25	23.86	11.09		
913896	ESR	≥51,020	≥50,040	0.34	0.42	1.92	51.26	26.49	20.33		
868744	MS	≥51,020	≥51,010	0.63	0.34	0.03	21.42	31.73	46.82		
868738	MS	≥51,020	≥50,970	0.54	0.40	0.10	16.48	59.53	23.89		
868742	ESR	≥51,020	≥50,940	0.29	0.19	0.17	59.36	23.20	17.27		
868739	ESR	≥51,020	≥50,960	0.33	0.37	0.12	57.86	17.66	24.36		
868743	ESR	≥51,020	≥50,940	0.34	0.41	0.16	55.80	19.19	24.84		
868740	ESR	≥51,020	≥50,960	0.52	0.18	0.12	43.94	8.63	47.31		
938100	TE	≥51,020	≥50,930	0.41	0.40	0.19	38.19	42.08	19.54		
868745	TE	≥51,020	≥50,900	0.27	0.45	0.23	48.11	50.31	1.35		
868741O2	CR	≥51,020	≥50,410	0.47	0.81	0.72	10.33	24.72	37.23	20.58	5.94
868741O3	CR	≥51,020	≥50,410	0.59	0.80	0.72	11.62	24.28	36.92	25.98	
872710	ESR	≥51,020	≥51,010	0.85	0.46	0.03	9.85	11.20	78.92		
872712	ESR	≥51,020	≥50,980	0.84	0.45	0.08	9.51	13.25	77.17		
872713	ESR	≥51,020	≥50,980	0.89	0.37	0.08	9.25	3.37	87.30		
872715	MS	≥51,020	≥50,970	0.40	0.26	0.11	39.84	39.23	20.81		
913974	ESR	≥51,020	≥50,970	0.39	0.39	0.11	53.78	13.53	32.58		
913970	MS	≥51,020	≥50,980	0.51	0.27	0.09	21.98	53.26	24.68		
913971	ESR	≥51,020	≥50,940	0.24	0.10	0.17	68.82	13.24	17.77		
913972	MS	≥51,020	≥50,920	0.37	0.41	0.21	37.36	51.31	11.12		
913973	ESR	≥51,020	≥50,910	0.34	0.37	0.22	52.15	28.29	19.34		
913975	MS	≥51,020	≥50,920	0.52	0.41	0.20	29.09	38.43	32.28		

Copyright © 2019 by Louisiana Department of Education.

Table 6.11 Operational Item Statistics—Mathematics Grade 3 CBT Administration

Mathematics Grade 3 Computer-Based Test Administration													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
914045	MC	≥1,040	≥1,040	0.92	0.24	0.10							
913997	SA	≥1,040	≥1,030	0.42	0.57	1.15	57.57	41.28					
870666	SA	≥1,040	≥1,040	0.63	0.46	0.29	37.16	62.55					
914024	SA	≥1,040	≥1,030	0.46	0.30	0.48	53.26	46.26					
914012	MC	≥1,040	≥1,040	0.14	0.19	0.19							
914038	SA	≥1,040	≥1,040	0.40	0.42	0.38	59.67	39.94					
870311	MC	≥1,040	≥1,040	0.90	0.24	0.00							
870686	SA	≥1,040	≥1,030	0.66	0.44	0.86	33.62	65.52					
870648	MC	≥1,040	≥1,040	0.98	0.17	0.00							
870679	MS	≥1,040	≥1,040	0.41	0.55	0.29	59.00	40.71					
870660	SA	≥1,040	≥1,030	0.74	0.23	0.48	26.25	73.28					
914025	SA	≥1,040	≥1,030	0.11	0.45	0.48	86.88	4.31	8.33				
913987	MC	≥1,040	≥1,040	0.64	0.41	0.29							
914008	CR	≥1,040	≥1,000	0.30	0.56	1.82	36.69	25.00	18.97	12.16	3.64		
914048	CR	≥1,040	≥970	0.26	0.58	3.45	50.19	23.95	10.34	9.20			
914033	MC	≥1,040	≥1,030	0.80	0.41	0.57							
914058	SA	≥1,040	≥1,030	0.82	0.25	0.48	18.01	81.51					
914019	MS	≥1,040	≥1,040	0.40	0.51	0.10	60.25	39.66					
916044	SA	≥1,040	≥1,030	0.38	0.57	0.57	42.72	37.36	19.35				
914027	SA	≥1,040	≥1,040	0.64	0.34	0.19	35.82	63.98					
870663	MS	≥1,040	≥1,040	0.19	0.30	0.19	81.23	18.58					
870650	MC	≥1,040	≥1,030	0.48	0.42	0.57							
916048	MC	≥1,040	≥1,040	0.43	0.22	0.38							
914031	SA	≥1,040	≥1,030	0.49	0.64	0.48	51.15	48.37					
914049	MPS	≥1,040	≥1,040	0.34	0.30	0.19	41.19	48.47	10.15				
914007	SA	≥1,040	≥1,030	0.25	0.60	0.57	74.81	24.62					
914028	MC	≥1,040	≥1,040	0.23	0.36	0.29							
798301	CR	≥1,040	≥1,030	0.31	0.67	0.48	43.39	25.77	21.26	8.33			
868619	CR	≥1,040	≥980	0.13	0.53	4.21	73.47	10.54	3.64	6.32			
878270	SA	≥1,040	≥1,040	0.60	0.55	0.00	39.56	60.44					
914035	SA	≥1,040	≥1,030	0.71	0.20	0.57	28.93	70.50					
914017	SA	≥1,040	≥1,040	0.64	0.58	0.10	35.92	63.98					
891472	MC	≥1,040	≥1,040	0.68	0.33	0.29							
914040	MS	≥1,040	≥1,040	0.64	0.41	0.10	36.21	63.70					
870683	SA	≥1,040	≥1,040	0.17	0.46	0.29	82.76	16.95					
913991	MC	≥1,040	≥1,030	0.54	0.45	0.48							
914001	MS	≥1,040	≥1,040	0.30	0.37	0.38	70.02	29.60					
870670	SA	≥1,040	≥1,030	0.17	0.56	0.67	82.66	16.67					
870689	MC	≥1,040	≥1,040	0.85	0.39	0.10							

Copyright © 2019 by Louisiana Department of Education.

Mathematics Grade 3 Computer-Based Test Administration (continued)													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
868618	MS	≥1,040	≥1,040	0.56	0.40	0.10	44.16	55.75					
914036	MS	≥1,040	≥1,040	0.37	0.47	0.38	62.55	37.07					
914039	CR	≥1,040	≥990	0.37	0.52	1.82	28.83	31.80	31.13	3.74			
914037	CR	≥1,040	≥1,010	0.29	0.72	1.44	28.93	20.50	17.62	13.89	7.85	5.75	2.20

Copyright © 2019 by Louisiana Department of Education.

Table 6.12 Operational Item Statistics—Mathematics Grade 3 PBT Administration

Mathematics Grade 3 Paper-Based Test Administration													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
914045	MC	≥54,320	≥53,930	0.94	0.26	0.31							
913997	SA	≥54,320	≥53,440	0.52	0.58	1.61	47.04	51.34					
870666	SA	≥54,320	≥53,070	0.72	0.49	2.30	27.30	70.40					
914024	SA	≥54,320	≥53,060	0.48	0.30	2.32	50.76	46.92					
914012	MC	≥54,320	≥53,470	0.25	0.38	0.60							
914038	SA	≥54,320	≥53,380	0.53	0.42	1.73	46.40	51.87					
870311	MC	≥54,320	≥53,370	0.94	0.27	1.16							
870686	SA	≥54,320	≥52,900	0.76	0.52	2.60	23.62	73.78					
870648	MC	≥54,320	≥54,070	0.98	0.18	0.41							
870679	MS	≥54,320	≥53,420	0.55	0.57	1.66	44.29	54.06					
870660	SA	≥54,320	≥53,560	0.76	0.27	1.39	23.98	74.63					
914025	SA	≥54,320	≥53,610	0.18	0.49	1.30	77.49	7.74	13.46				
913987	MC	≥54,320	≥52,990	0.74	0.45	0.71							
914008	CR	≥54,320	≥53,810	0.35	0.56	0.86	28.60	28.26	18.07	20.60	3.54		
914048	CR	≥54,320	≥51,240	0.34	0.60	5.13	33.75	30.19	24.55	5.84			
914033	MC	≥54,320	≥54,060	0.87	0.40	0.30							
914058	SA	≥54,320	≥53,320	0.83	0.31	1.83	16.87	81.31					
914019	MS	≥54,320	≥53,440	0.52	0.52	1.62	47.33	51.05					
916044	SA	≥54,320	≥53,880	0.45	0.58	0.80	35.50	39.06	24.65				
914027	SA	≥54,320	≥53,360	0.72	0.36	1.77	27.17	71.06					
870663	MS	≥54,320	≥53,400	0.30	0.41	1.69	69.18	29.13					
870650	MC	≥54,320	≥51,810	0.57	0.48	2.25							
916048	MC	≥54,320	≥53,400	0.46	0.28	1.20							
914031	SA	≥54,320	≥53,330	0.61	0.63	1.81	38.77	59.42					
914049	MPS	≥54,320	≥53,380	0.40	0.40	1.72	34.96	47.66	15.66				
914007	SA	≥54,320	≥53,360	0.37	0.60	1.77	62.07	36.17					
914028	MC	≥54,320	≥53,730	0.36	0.48	0.82							
798301	CR	≥54,320	≥54,030	0.43	0.65	0.54	26.81	25.02	40.06	7.58			
868619	CR	≥54,320	≥50,540	0.21	0.54	6.77	65.99	8.76	5.51	12.77			
878270	SA	≥54,320	≥53,530	0.69	0.56	1.44	30.20	68.36					
914035	SA	≥54,320	≥53,740	0.70	0.26	1.06	29.85	69.09					
914017	SA	≥54,320	≥52,780	0.73	0.56	2.82	25.97	71.21					
891472	MC	≥54,320	≥53,800	0.82	0.36	0.86							
914040	MS	≥54,320	≥53,670	0.72	0.43	1.20	27.66	71.14					
870683	SA	≥54,320	≥52,880	0.20	0.46	2.65	77.72	19.63					
913991	MC	≥54,320	≥53,580	0.61	0.48	1.31							
914001	MS	≥54,320	≥53,170	0.45	0.45	2.12	54.12	43.76					
870670	SA	≥54,320	≥52,800	0.25	0.57	2.80	72.42	24.78					
870689	MC	≥54,320	≥53,470	0.92	0.36	1.29							

Copyright © 2019 by Louisiana Department of Education.

Mathematics Grade 3 Paper-Based Test Administration (continued)													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
868618	MS	≥54,320	≥53,650	0.66	0.43	1.24	33.49	65.27					
914036	MS	≥54,320	≥53,560	0.48	0.52	1.39	51.31	47.30					
914039	CR	≥54,320	≥53,530	0.47	0.53	1.24	18.11	27.59	45.92	6.92			
914037	CR	≥54,320	≥53,720	0.38	0.73	0.96	20.92	20.23	16.50	14.48	10.22	12.92	3.63

Copyright © 2019 by Louisiana Department of Education.

Table 6.13 Operational Item Statistics—Mathematics Grade 4 CBT Administration

Mathematics Grade 4 Computer-Based Test Administration													
Item ID	Item Type	Total N	Adj. N	*p*-Value	Pbis	Omit Rate	% at 0	% at 1	% at 2	% at 3	% at 4	% at 5	% at 6
914131	MC	≥5,160	≥5,150	0.73	0.37	0.08							
870744	MC	≥5,160	≥5,150	0.69	0.34	0.10							
914130	SA	≥5,160	≥5,150	0.82	0.26	0.08	17.73	82.20					
914104	SA	≥5,160	≥5,140	0.23	0.39	0.43	76.46	23.11					
870702	SA	≥5,160	≥5,150	0.58	0.38	0.19	41.73	58.08					
868632	SA	≥5,160	≥5,140	0.43	0.52	0.29	57.01	42.70					
870730	MC	≥5,160	≥5,150	0.50	0.61	0.08							
870732	MC	≥5,160	≥5,150	0.70	0.50	0.19							
944080	MC	≥5,160	≥5,150	0.65	0.38	0.12							
870743	MC	≥5,160	≥5,150	0.58	0.43	0.19							
914080	MS	≥5,160	≥5,150	0.73	0.48	0.08	26.83	73.09					
870717	MPS	≥5,160	≥5,150	0.29	0.27	0.15	53.82	33.94	12.09				
914070	SA	≥5,160	≥5,140	0.41	0.58	0.41	59.05	40.55					
914084	CR	≥5,160	≥5,140	0.35	0.68	0.08	24.14	32.10	26.73	14.32	2.32		
914086	CR	≥5,160	≥5,000	0.18	0.60	1.76	65.85	19.33	3.29	8.56			
914101	MC	≥5,160	≥5,160	0.78	0.39	0.02							
914102	SA	≥5,160	≥5,150	0.81	0.28	0.08	18.56	81.36					
870713	MC	≥5,160	≥5,160	0.57	0.40	0.00							
868627	SA	≥5,160	≥5,140	0.29	0.40	0.39	51.45	37.64	10.52				
914095	MC	≥5,160	≥5,160	0.35	0.53	0.04							
914121	SA	≥5,160	≥5,150	0.74	0.38	0.10	26.06	73.85					
914088	MC	≥5,160	≥5,150	0.50	0.51	0.08							
914078	MC	≥5,160	≥5,150	0.53	0.50	0.06							
878669	SA	≥5,160	≥5,150	0.58	0.45	0.12	17.57	48.33	33.98				
914105	MS	≥5,160	≥5,150	0.60	0.35	0.17	40.31	59.51					
914126	MC	≥5,160	≥5,150	0.83	0.40	0.15							
914079	SA	≥5,160	≥5,150	0.68	0.38	0.10	32.16	67.75					
870723	CR	≥5,160	≥5,040	0.28	0.59	0.81	58.21	12.03	13.73	13.77			
870317	CR	≥5,160	≥5,070	0.25	0.63	0.68	57.11	14.39	20.86	5.87			
914124	MC	≥5,160	≥5,150	0.69	0.48	0.15							
914073	MS	≥5,160	≥5,160	0.89	0.35	0.00	11.47	88.53					
914122	MC	≥5,160	≥5,160	0.89	0.27	0.04							
868631	SA	≥5,160	≥5,140	0.33	0.59	0.31	67.16	32.53					
914135	MC	≥5,160	≥5,150	0.84	0.36	0.08							
868622	SA	≥5,160	≥5,150	0.36	0.32	0.12	64.39	35.49					
870736	MC	≥5,160	≥5,150	0.28	0.24	0.10							
914099	SA	≥5,160	≥5,150	0.26	0.53	0.21	73.50	26.29					
870731	MC	≥5,160	≥5,150	0.49	0.44	0.21							

Copyright © 2019 by Louisiana Department of Education.

Mathematics Grade 4 Computer-Based Test Administration (continued)													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
870725	MC	≥5,160	≥5,150	0.60	0.42	0.08							
914108	MS	≥5,160	≥5,150	0.41	0.44	0.12	59.28	40.60					
870722	CR	≥5,160	≥5,150	0.18	0.56	0.12	57.61	31.93	7.44	2.83			
870724	CR	≥5,160	≥5,130	0.16	0.70	0.62	54.57	19.86	11.78	6.99	3.10	1.69	1.39

Copyright © 2019 by Louisiana Department of Education.

Table 6.14 Operational Item Statistics—Mathematics Grade 4 PBT Administration

Mathematics Grade 4 Paper-Based Test Administration													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
914131	MC	≥50,520	≥50,280	0.77	0.35	0.35							
870744	MC	≥50,520	≥50,030	0.75	0.34	0.45							
914130	SA	≥50,520	≥49,790	0.82	0.28	1.44	17.48	81.08					
914104	SA	≥50,520	≥49,280	0.23	0.37	2.46	74.96	22.58					
870702	SA	≥50,520	≥49,350	0.61	0.39	2.33	38.25	59.42					
868632	SA	≥50,520	≥49,430	0.45	0.53	2.16	54.10	43.74					
870730	MC	≥50,520	≥50,100	0.55	0.61	0.77							
870732	MC	≥50,520	≥49,890	0.75	0.49	1.19							
944080	MC	≥50,520	≥50,080	0.68	0.40	0.83							
870743	MC	≥50,520	≥49,920	0.58	0.44	1.12							
914080	MS	≥50,520	≥50,100	0.80	0.44	0.83	19.80	79.37					
870717	MPS	≥50,520	≥49,980	0.34	0.31	1.07	48.10	35.25	15.58				
914070	SA	≥50,520	≥49,260	0.45	0.59	2.49	53.26	44.25					
914084	CR	≥50,520	≥50,310	0.37	0.68	0.41	21.92	30.36	29.14	15.12	3.05		
914086	CR	≥50,520	≥47,890	0.24	0.60	4.99	53.03	24.78	7.65	9.34			
914101	MC	≥50,520	≥50,190	0.81	0.36	0.41							
914102	SA	≥50,520	≥49,810	0.83	0.28	1.41	16.66	81.94					
870713	MC	≥50,520	≥49,870	0.65	0.42	1.26							
868627	SA	≥50,520	≥50,020	0.31	0.37	0.98	50.07	36.45	12.50				
914095	MC	≥50,520	≥50,020	0.37	0.57	0.93							
914121	SA	≥50,520	≥49,620	0.76	0.39	1.78	23.69	74.53					
914088	MC	≥50,520	≥49,450	0.57	0.47	0.53							
914078	MC	≥50,520	≥50,240	0.57	0.51	0.44							
878669	SA	≥50,520	≥50,330	0.62	0.41	0.39	14.25	47.55	37.82				
914105	MS	≥50,520	≥49,760	0.63	0.33	1.50	36.35	62.15					
914126	MC	≥50,520	≥49,930	0.85	0.38	1.11							
914079	SA	≥50,520	≥49,480	0.71	0.39	2.06	28.41	69.53					
870723	CR	≥50,520	≥49,660	0.32	0.57	1.44	53.93	12.14	15.26	16.96			
870317	CR	≥50,520	≥49,250	0.35	0.62	2.28	48.05	13.43	18.74	17.25			
914124	MC	≥50,520	≥50,070	0.72	0.48	0.84							
914073	MS	≥50,520	≥50,010	0.89	0.27	1.02	11.00	87.98					
914122	MC	≥50,520	≥50,210	0.91	0.26	0.60							
868631	SA	≥50,520	≥49,380	0.36	0.59	2.27	62.76	34.98					
914135	MC	≥50,520	≥50,220	0.88	0.34	0.56							
868622	SA	≥50,520	≥49,360	0.42	0.35	2.30	56.64	41.06					
870736	MC	≥50,520	≥49,770	0.29	0.29	1.43							
914099	SA	≥50,520	≥49,350	0.28	0.53	2.32	69.98	27.69					

Copyright © 2019 by Louisiana Department of Education.

Mathematics Grade 4 Paper-Based Test Administration (continued)													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
870731	MC	≥50,520	≥49,670	0.54	0.44	1.11							
870706	SA	≥50,520	≥49,020	0.58	0.50	2.96	40.95	56.09					
870725	MC	≥50,520	≥50,000	0.66	0.40	0.92							
914108	MS	≥50,520	≥49,750	0.42	0.42	1.53	57.06	41.40					
870722	CR	≥50,520	≥50,260	0.20	0.56	0.50	54.59	33.47	8.52	2.90			
870724	CR	≥50,520	≥49,990	0.19	0.71	1.06	47.68	21.19	12.20	8.76	4.45	2.26	2.41

Copyright © 2019 by Louisiana Department of Education.

Table 6.15 Operational Item Statistics—Mathematics Grade 5 CBT Administration

Mathematics Grade 5 Computer-Based Test Administration													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
914204	MC	≥55,200	≥55,130	0.60	0.44	0.11							
870764	MPS	≥55,200	≥55,130	0.56	0.35	0.11	28.47	31.37	40.05				
914162	TE	≥55,200	≥55,130	0.23	0.42	0.12	77.22	22.67					
870749	MS	≥55,200	≥55,030	0.22	0.47	0.30	77.95	21.75					
870767	MC	≥55,200	≥55,150	0.73	0.33	0.08							
914168	MC	≥55,200	≥55,060	0.27	0.19	0.25							
870770	SA	≥55,200	≥54,630	0.17	0.54	1.03	82.53	16.44					
914183	TE	≥55,200	≥55,120	0.45	0.52	0.14	55.23	44.63					
870789	MC	≥55,200	≥55,130	0.64	0.42	0.13							
870785	MC	≥55,200	≥55,030	0.27	0.33	0.30							
870758	SA	≥55,200	≥54,820	0.58	0.46	0.68	41.75	57.57					
914209	MC	≥55,200	≥55,110	0.49	0.43	0.16							
914219	MC	≥55,200	≥55,050	0.66	0.26	0.26							
914193	CR	≥55,200	≥54,960	0.46	0.60	0.43	16.25	25.22	27.13	21.60	9.37		
914148	CR	≥55,200	≥53,280	0.25	0.67	2.82	55.33	21.71	8.16	11.32			
914150	MC	≥55,200	≥55,100	0.76	0.27	0.17							
870795	MC	≥55,200	≥55,150	0.63	0.24	0.09							
870787	SA	≥55,200	≥55,130	0.67	0.50	0.11	33.13	66.76					
914190	SA	≥55,200	≥54,960	0.57	0.45	0.42	42.57	57.02					
914182	SA	≥55,200	≥55,070	0.18	0.18	0.22	82.08	17.70					
914145	SA	≥55,200	≥55,070	0.28	0.52	0.22	72.21	27.57					
870794	MC	≥55,200	≥55,160	0.31	0.26	0.07							
914187	MPS	≥55,200	≥55,180	0.51	0.51	0.03	32.69	31.83	35.45				
914215	MC	≥55,200	≥55,150	0.62	0.48	0.09							
870781	SA	≥55,200	≥55,090	0.46	0.28	0.19	53.52	46.29					
870784	MC	≥55,200	≥55,120	0.56	0.52	0.13							
914166	MPS	≥55,200	≥55,150	0.52	0.33	0.09	15.91	63.34	20.66				
870325	CR	≥55,200	≥54,750	0.24	0.52	0.43	44.17	41.49	10.26	3.28			
868655	CR	≥55,200	≥54,090	0.14	0.53	1.32	69.32	19.71	5.16	3.80			
914201	MC	≥55,200	≥55,170	0.57	0.31	0.05							
914171	SA	≥55,200	≥55,140	0.64	0.48	0.10	35.53	64.37					
914161	TE	≥55,200	≥55,110	0.21	0.24	0.15	79.10	20.75					
914149	SA	≥55,200	≥55,110	0.67	0.29	0.15	32.85	67.01					
914192	MC	≥55,200	≥55,140	0.45	0.40	0.10							
914164	SA	≥55,200	≥55,050	0.32	0.30	0.26	67.58	32.16					
870751	MC	≥55,200	≥55,140	0.52	0.48	0.10							
914184	SA	≥55,200	≥55,080	0.69	0.33	0.20	31.42	68.37					

Copyright © 2019 by Louisiana Department of Education.

Mathematics Grade 5 Computer-Based Test Administration (continued)													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
914198	SA	≥55,200	≥55,100	0.38	0.59	0.17	61.77	38.07					
914144	MS	≥55,200	≥55,170	0.31	0.32	0.05	69.43	30.52					
914203	MC	≥55,200	≥55,130	0.64	0.28	0.12							
914210	SA	≥55,200	≥55,140	0.78	0.32	0.11	21.77	78.12					
914195	CR	≥55,200	≥55,130	0.28	0.66	0.11	49.43	27.40	13.69	9.36			
868654	CR	≥55,200	≥54,590	0.24	0.65	0.64	42.72	13.86	15.22	15.96	6.03	2.69	2.42

Copyright © 2019 by Louisiana Department of Education.

Table 6.16 Item Statistics—Mathematics Grade 6 Computer-Based Test Administration

Mathematics Grade 6 Computer-Based Test Administration													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
914266	MC	≥52,670	≥52,650	0.93	0.19	0.03							
914288	SA	≥52,670	≥52,350	0.54	0.35	0.60	45.52	53.87					
914223	TE	≥52,670	≥52,620	0.59	0.47	0.08	41.43	58.49					
914260	SA	≥52,670	≥52,390	0.39	0.39	0.53	61.10	38.38					
914252	MC	≥52,670	≥52,610	0.60	0.34	0.11							
916476	SA	≥52,670	≥52,580	0.36	0.31	0.16	63.64	36.20					
914246	TE	≥52,670	≥52,570	0.47	0.39	0.19	53.04	46.77					
878298	MS	≥52,670	≥52,610	0.55	0.47	0.10	44.59	55.31					
870806	MC	≥52,670	≥52,590	0.56	0.52	0.14							
878302	MC	≥52,670	≥52,600	0.64	0.47	0.12							
914259	TE	≥52,670	≥52,590	0.57	0.42	0.14	42.52	57.35					
870809	MS	≥52,670	≥52,570	0.23	0.28	0.18	77.30	22.52					
914237	TE	≥52,670	≥52,590	0.67	0.40	0.15	33.36	66.49					
870835	MS	≥52,670	≥52,590	0.62	0.45	0.15	38.27	61.58					
870827	MC	≥52,670	≥52,490	0.77	0.42	0.34							
870833	SA	≥52,670	≥52,080	0.29	0.26	1.11	69.95	28.93					
870834	MC	≥52,670	≥52,460	0.32	0.17	0.39							
914253	TE	≥52,670	≥52,410	0.52	0.50	0.48	47.48	52.04					
914228	MC	≥52,670	≥52,350	0.58	0.48	0.61							
914257	TE	≥52,670	≥52,360	0.71	0.51	0.58	28.77	70.65					
878637	MC	≥52,670	≥52,630	0.77	0.37	0.07							
870811	SA	≥52,670	≥52,390	0.20	0.50	0.52	79.29	20.20					
870814	MS	≥52,670	≥52,630	0.27	0.51	0.06	73.33	26.61					
870830	MC	≥52,670	≥52,600	0.43	0.52	0.12							
914272	SA	≥52,670	≥52,370	0.31	0.61	0.56	69.05	30.38					
914248	SA	≥52,670	≥52,310	0.25	0.60	0.67	64.00	21.86	13.47				
914286	CR	≥52,670	≥51,620	0.14	0.60	1.21	60.62	23.62	9.97	3.25	0.55		
914264	MC	≥52,670	≥52,520	0.33	0.33	0.28							
945486	SA	≥52,670	≥52,460	0.21	0.60	0.38	71.61	14.23	13.77				
868818	CR	≥52,670	≥51,040	0.16	0.59	2.03	71.50	10.53	7.42	7.47			
868798	CR	≥52,670	≥51,190	0.18	0.67	2.09	65.86	6.26	7.20	3.92	3.14	5.26	5.56
870825	CR	≥52,670	≥49,130	0.18	0.64	5.14	53.41	31.68	6.18	2.01			
914249	MC	≥52,670	≥52,580	0.54	0.33	0.16							
914271	SA	≥52,670	≥52,500	0.31	0.58	0.32	68.43	31.25					
914273	SA	≥52,670	≥52,540	0.48	0.62	0.24	27.19	49.61	22.97				
914233	MS	≥52,670	≥52,610	0.20	0.51	0.10	79.62	20.28					

Copyright © 2019 by Louisiana Department of Education.

Mathematics Grade 6 Computer-Based Test Administration (continued)													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
878297	TE	≥52,670	≥52,590	0.45	0.70	0.14	28.40	52.23	19.24				
878300	MS	≥52,670	≥52,620	0.19	0.38	0.08	80.95	18.97					
868796	SA	≥52,670	≥52,550	0.25	0.58	0.23	74.73	25.04					
914280	CR	≥52,670	≥51,560	0.37	0.67	1.11	42.83	18.35	19.16	17.57			
914231	CR	≥52,670	≥52,050	0.34	0.62	0.64	41.41	12.67	21.25	14.63	8.87		
914244	CR	≥52,670	≥50,990	0.25	0.66	2.04	62.08	12.42	8.03	14.27			

Copyright © 2019 by Louisiana Department of Education.

Table 6.17 Item Statistics—Mathematics Grade 7 Computer-Based Test Administration

Mathematics Grade 7 Computer-Based Test Administration													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
915100	MC	≥51,420	≥51,370	0.76	0.42	0.08							
914294	MS	≥51,420	≥51,260	0.21	0.50	0.31	78.69	21.00					
870846	SA	≥51,420	≥51,260	0.47	0.61	0.31	52.46	47.23					
870848	MC	≥51,420	≥51,310	0.57	0.49	0.21							
914299	MC	≥51,420	≥51,330	0.38	0.38	0.16							
914324	SA	≥51,420	≥51,060	0.55	0.42	0.69	44.98	54.32					
892610	MS	≥51,420	≥51,350	0.24	0.58	0.13	75.54	24.33					
914323	TE	≥51,420	≥51,030	0.45	0.61	0.75	54.96	44.28					
914359	SA	≥51,420	≥51,240	0.51	0.53	0.34	48.42	51.23					
914293	MS	≥51,420	≥51,310	0.31	0.56	0.21	68.49	31.30					
891483	MC	≥51,420	≥51,260	0.39	0.30	0.31							
868840	SA	≥51,420	≥51,210	0.41	0.46	0.40	58.80	40.80					
802365	MC	≥51,420	≥51,290	0.58	0.45	0.24							
914340	MC	≥51,420	≥51,240	0.46	0.45	0.34							
868829	MS	≥51,420	≥51,130	0.18	0.56	0.55	81.39	18.05					
914352	TE	≥51,420	≥51,020	0.19	0.34	0.77	80.30	18.93					
915099	MC	≥51,420	≥51,050	0.56	0.40	0.72							
870871	MC	≥51,420	≥50,970	0.48	0.21	0.86							
914311	MS	≥51,420	≥50,800	0.17	0.47	1.20	81.67	17.13					
915098	MS	≥51,420	≥50,780	0.13	0.32	1.24	85.92	12.84					
914330	MS	≥51,420	≥51,350	0.70	0.30	0.12	30.42	69.46					
870337	MC	≥51,420	≥51,290	0.46	0.28	0.24							
870854	MC	≥51,420	≥51,340	0.42	0.46	0.15							
868846	MS	≥51,420	≥51,350	0.18	0.19	0.13	81.69	18.18					
870875	SA	≥51,420	≥51,220	0.23	0.43	0.39	61.25	31.09	7.27				
914362	CR	≥51,420	≥50,730	0.12	0.61	0.90	80.80	2.53	2.04	3.72	2.23	2.53	4.81
914316	TE	≥51,420	≥51,210	0.35	0.45	0.40	64.50	35.11					
914304	MPS	≥51,420	≥51,340	0.47	0.59	0.15	34.65	35.91	29.29				
914301	MC	≥51,420	≥51,290	0.43	0.24	0.25							
914333	CR	≥51,420	≥50,500	0.33	0.64	0.98	26.56	29.86	27.65	11.41	2.74		
868849	CR	≥51,420	≥49,160	0.15	0.58	3.21	65.95	21.70	3.38	4.58			
878705	CR	≥51,420	≥48,000	0.07	0.45	5.25	79.97	8.29	4.37	0.72			
914342	MC	≥51,420	≥51,350	0.51	0.38	0.12							
914319	SA	≥51,420	≥51,370	0.34	0.51	0.09	45.36	42.04	12.51				
870873	MC	≥51,420	≥51,350	0.62	0.40	0.12							
870878	SA	≥51,420	≥51,270	0.20	0.47	0.28	79.82	19.90					
892612	TE	≥51,420	≥51,310	0.50	0.47	0.20	50.13	49.68					
915096	TE	≥51,420	≥51,330	0.16	0.22	0.17	84.04	15.79					
914335	MPS	≥51,420	≥51,360	0.35	0.46	0.11	44.70	40.90	14.28				

Copyright © 2019 by Louisiana Department of Education.

Mathematics Grade 7 Computer-Based Test Administration (continued)													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
914337	SA	≥51,420	≥51,260	0.49	0.52	0.30	50.85	48.85					
914336	CR	≥51,420	≥50,000	0.34	0.66	1.62	34.92	31.82	24.75	5.74			
914339	CR	≥51,420	≥50,390	0.18	0.58	1.21	63.52	9.47	19.23	2.21	3.57		
914329	CR	≥51,420	≥48,980	0.30	0.70	3.17	46.00	21.45	18.31	9.50			

Copyright © 2019 by Louisiana Department of Education.

Table 6.18 Item Statistics—Mathematics Grade 8 Computer-Based Test Administration

Mathematics Grade 8 Computer-Based Test Administration													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
914415	MC	≥44,910	≥44,510	0.38	0.16	0.89							
878745	MPS	≥44,910	≥44,860	0.68	0.39	0.12	11.07	42.36	46.44				
914372	SA	≥44,910	≥44,210	0.23	0.55	1.56	75.83	22.61					
868878	MC	≥44,910	≥44,820	0.63	0.35	0.20							
914405	TE	≥44,910	≥44,510	0.22	0.42	0.89	77.29	21.82					
870886	MS	≥44,910	≥44,850	0.29	0.51	0.14	70.67	29.19					
914367	MC	≥44,910	≥44,760	0.40	0.29	0.34							
870900	SA	≥44,910	≥43,820	0.24	0.49	2.43	74.05	23.51					
870901	MS	≥44,910	≥44,820	0.35	0.56	0.20	64.73	35.07					
870915	MC	≥44,910	≥44,770	0.29	0.28	0.31							
870890	SA	≥44,910	≥44,600	0.64	0.30	0.70	35.82	63.49					
914427	MS	≥44,910	≥44,580	0.37	0.53	0.75	62.71	36.54					
914403	MC	≥44,910	≥44,510	0.27	0.13	0.89							
868871	MS	≥44,910	≥44,630	0.24	0.36	0.62	75.38	24.00					
914371	TE	≥44,910	≥44,400	0.10	0.43	1.14	89.31	9.56					
868886	SA	≥44,910	≥43,390	0.33	0.56	3.40	65.09	31.51					
914431	SA	≥44,910	≥43,610	0.24	0.58	2.89	73.47	23.63					
878747	MC	≥44,910	≥44,250	0.47	0.34	1.48							
944912	MPS	≥44,910	≥44,220	0.39	0.48	1.55	40.26	39.80	18.39				
914370	MS	≥44,910	≥43,840	0.29	0.55	2.40	69.36	28.25					
914411	SA	≥44,910	≥44,660	0.50	0.51	0.56	49.86	49.58					
878742	MC	≥44,910	≥44,840	0.54	0.34	0.16							
868873	MS	≥44,910	≥44,890	0.18	0.36	0.05	82.33	17.62					
870347	SA	≥44,910	≥44,010	0.12	0.49	2.01	81.98	8.48	7.54				
892626	CR	≥44,910	≥43,890	0.27	0.63	1.62	47.22	14.20	7.89	9.17	5.73	5.36	8.16
914378	MC	≥44,910	≥44,830	0.51	0.35	0.18							
914399	MPS	≥44,910	≥44,880	0.50	0.57	0.07	28.91	41.46	29.56				
914410	SA	≥44,910	≥44,620	0.10	0.44	0.65	89.03	10.32					
800273	CR	≥44,910	≥42,510	0.20	0.66	3.77	54.01	23.65	6.89	3.32	6.77		
914376	CR	≥44,910	≥42,810	0.16	0.62	3.23	65.79	17.47	6.73	5.33			
878156	CR	≥44,910	≥42,590	0.12	0.49	3.81	74.08	8.56	10.70	1.50			
914429	MC	≥44,910	≥44,890	0.57	0.32	0.05							
870918	SA	≥44,910	≥44,800	0.21	0.55	0.25	79.10	20.65					
914436	SA	≥44,910	≥44,640	0.30	0.68	0.61	56.33	26.03	17.03				
914396	MC	≥44,910	≥44,850	0.49	0.48	0.14							
914397	MC	≥44,910	≥44,850	0.34	0.52	0.13							
914424	TE	≥44,910	≥44,870	0.60	0.39	0.10	39.49	60.41					
914430	MS	≥44,910	≥44,860	0.26	0.55	0.12	73.83	26.05					

Copyright © 2019 by Louisiana Department of Education.

Mathematics Grade 8 Computer-Based Test Administration (continued)													
Item ID	**Item Type**	**Total N**	**Adj. N**	***p*-Value**	**Pbis**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
914395	TE	≥44,910	≥44,720	0.61	0.53	0.42	39.11	60.47					
914381	CR	≥44,910	≥43,260	0.16	0.64	2.20	58.79	15.59	19.46	2.00	0.49		
868882	CR	≥44,910	≥44,850	0.23	0.60	0.13	42.62	47.82	8.25	1.18			
914400	CR	≥44,910	≥42,070	0.11	0.55	3.95	71.30	14.02	7.35	0.99			

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

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

- it has a *p*-value less than 0.25 or above .95,
- an incorrect option had greater than the percentage choosing the correct (keyed) response for the sub-group of students who scored at the top 20%, based on the total score points,
- the item-total correlation of the keyed response is less than 0.20,
- any of the incorrect answer options yields a positive distractor-total correlation, or
- the percentage of omitting or not reaching each item is 5% or greater.

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

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

Copyright © 2019 by Louisiana Department of Education.

6.2 Item Reponse Theory

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

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

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

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

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

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

The software 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.

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)

Sample data was used for grades 3 and 4 PBT calibration. All student data available at the time of calibration was used for grades 3 to 8 CBT calibration, resulting in a near-census data file. Tables 6.19 and 6.20 show the representativeness of the calibration samples compared to the census data. These tables demonstrate that the calibration sample was representative of the state.

Copyright © 2019 by Louisiana Department of Education.

Table 6.19 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	≥34,460	100.00%	≥55,540	100.00%	0.00%
	Gender					
	Male	≥17,560	50.95%	≥28,340	51.03%	0.08%
	Female	≥16,900	49.05%	≥27,180	48.93%	(0.11%)
	Race Ethnicity					
	Hispanic/Latino	≥2,960	8.61%	≥4,360	7.86%	(0.75%)
	American Indian or Alaska Native	≥180	0.53%	≥340	0.61%	0.09%
	Asian	≥560	1.64%	≥770	1.39%	(0.25%)
	Black or African American	≥13,980	40.57%	≥24,400	43.93%	3.36%
	Native Hawaiian or Other Pacific	≥20	0.08%	≥40	0.08%	0.00%
	White	≥15,670	45.48%	≥23,830	42.92%	(2.57%)
	Two or More Races	≥1,060	3.09%	≥1,720	3.11%	0.02%
4	All Students	≥45,320	100.00%	≥55,940	100.00%	0.00%
	Gender					
	Male	≥23,040	50.85%	≥28,470	50.89%	0.05%
	Female	≥22,280	49.15%	≥27,430	49.03%	(0.12%)
	Race Ethnicity					
	Hispanic/Latino	≥3,590	7.93%	≥4,290	7.67%	(0.26%)
	American Indian or Alaska Native	≥280	0.63%	≥30	0.61%	(0.02%)
	Asian	≥750	1.67%	≥880	1.59%	(0.08%)
	Black or African American	≥18,560	40.95%	≥24,580	43.95%	2.99%
	Native Hawaiian or Other Pacific	≥40	0.09%	≥50	0.10%	0.01%
	White	≥20,660	45.59%	≥24,040	42.98%	(2.61%)
	Two or More Races	≥1,420	3.13%	≥1,660	2.98%	(0.16%)
5	All Students	≥55,490	100.00%	≥55,540	100.00%	0.00%
	Gender					
	Male	≥28,690	51.71%	≥28,260	50.88%	(0.82%)
	Female	≥26,800	48.29%	≥27,270	49.12%	0.82%
	Race Ethnicity					
	Hispanic/Latino	≥3,960	7.14%	≥3,940	7.10%	(0.04%)
	American Indian or Alaska Native	≥340	0.62%	≥350	0.64%	0.02%
	Asian	≥800	1.45%	≥820	1.48%	0.03%
	Black or African American	≥24,690	44.49%	≥24,110	43.42%	(1.06%)
	Native Hawaiian or Other Pacific	≥40	0.08%	≥40	0.09%	0.00%
	White	≥24,190	43.59%	≥24,630	44.36%	0.76%
	Two or More Races	≥1,440	2.60%	≥1,590	2.87%	0.27%

Copyright © 2019 by Louisiana Department of Education.

Calibration and Census Data: English Language Arts (continued)						
		Calibration Sample		Census Data		
Grade		N	%	N	%	(Calib % - Census %)
6	All Students	≥52,970	100.00%	≥53,050	100.00%	0.00%
	Gender					
	Male	≥27,170	51.29%	≥27,210	51.29%	0.01%
	Female	≥25,800	48.71%	≥25,840	48.71%	(0.01%)
	Race Ethnicity					
	Hispanic/Latino	≥3,590	6.79%	≥3,590	6.77%	(0.03%)
	American Indian or Alaska Native	≥330	0.64%	≥330	0.64%	(0.00%)
	Asian	≥810	1.54%	≥810	1.53%	(0.00%)
	Black or African American	≥23,440	44.25%	≥23,480	44.27%	0.03%
	Native Hawaiian or Other Pacific	≥40	0.09%	≥40	0.09%	(0.00%)
	White	≥23,390	44.16%	≥23,410	44.14%	(0.02%)
	Two or More Races	≥1,330	2.52%	≥1,330	2.52%	(0.00%)
7	All Students	≥51,330	100.00%	≥51,800	100.00%	0.00%
	Gender					
	Male	≥26,360	51.35%	≥26,650	51.45%	0.10%
	Female	≥24,970	48.65%	≥25,140	48.55%	(0.10%)
	Race Ethnicity					
	Hispanic/Latino	≥3,880	7.57%	≥3,390	6.55%	(1.02%)
	American Indian or Alaska Native	≥350	0.69%	≥370	0.72%	0.03%
	Asian	≥790	1.55%	≥790	1.54%	(0.01%)
	Black or African American	≥22,760	44.33%	≥22,710	43.84%	(0.49%)
	Native Hawaiian or Other Pacific	≥30	0.07%	≥20	0.04%	(0.03%)
	White	≥22,220	43.28%	≥23,360	45.10%	1.82%
	Two or More Races	≥1,270	2.49%	≥1,120	2.16%	(0.33%)
8	All Students	≥51,290	100.00%	≥51,380	100.00%	0.00%
	Gender					
	Male	≥26,550	51.76%	≥26,020	50.64%	(1.12%)
	Female	≥24,740	48.24%	≥25,350	49.36%	1.12%
	Race Ethnicity					
	Hispanic/Latino	≥3,590	7.01%	≥3,270	6.37%	(0.65%)
	American Indian or Alaska Native	≥310	0.62%	≥380	0.75%	0.13%
	Asian	≥820	1.60%	≥870	1.71%	0.11%
	Black or African American	≥22,930	44.71%	≥22,540	43.88%	(0.84%)
	Native Hawaiian or Other Pacific	≥40	0.08%	≥30	0.06%	(0.01%)
	White	≥22,190	43.26%	≥23,240	45.23%	1.97%
	Two or More Races	≥1,380	2.70%	≥1,010	1.97%	(0.73%)

Copyright © 2019 by Louisiana Department of Education.

Table 6.20 Summary of Calibration and Census Data: Mathematics

Calibration and Census Data: Mathematics						
		Calibration Sample		Census Data		
Grade		N	%	N	%	(Calib % - Census %)
3	All Students	≥44,710	100.00%	≥55,560	100.00%	0.00%
	Gender					
	Male	≥22,840	51.09%	≥28,350	51.02%	(0.07%)
	Female	≥21,860	48.91%	≥27,190	48.93%	0.02%
	Race Ethnicity					
	Hispanic/Latino	≥3,630	8.13%	≥4,360	7.86%	(0.27%)
	American Indian or Alaska Native	≥300	0.68%	≥340	0.62%	(0.06%)
	Asian	≥660	1.48%	≥770	1.39%	(0.09%)
	Black or African American	≥17,790	39.80%	≥24,410	43.93%	4.13%
	Native Hawaiian or Other Pacific	≥20	0.06%	≥40	0.08%	0.01%
	White	≥20,770	46.46%	≥23,840	42.90%	(3.56%)
	Two or More Races	≥1,510	3.38%	≥1,720	3.11%	(0.27%)
4	All Students	≥55,010	100.00%	≥55,910	100.00%	0.00%
	Gender					
	Male	≥27,960	50.84%	≥28,470	50.92%	0.07%
	Female	≥27,040	49.16%	≥27,410	49.03%	(0.12%)
	Race Ethnicity					
	Hispanic/Latino	≥4,140	7.53%	≥4,280	7.66%	0.13%
	American Indian or Alaska Native	≥330	0.62%	≥340	0.61%	(0.01%)
	Asian	≥870	1.60%	≥880	1.59%	(0.01%)
	Black or African American	≥24,180	43.96%	≥24,570	43.96%	(0.01%)
	Native Hawaiian or Other Pacific	≥50	0.10%	≥50	0.10%	0.00%
	White	≥23,750	43.19%	≥24,040	43.01%	(0.18%)
	Two or More Races	≥1,640	3.00%	≥1,660	2.98%	(0.02%)
5	All Students	≥55,170	100.00%	≥55,480	100.00%	0.00%
	Gender					
	Male	≥28,520	51.70%	≥28,220	50.87%	(0.82%)
	Female	≥26,650	48.30%	≥27,250	49.13%	0.82%
	Race Ethnicity					
	Hispanic/Latino	≥3,940	7.15%	≥3,940	7.10%	(0.04%)
	American Indian or Alaska Native	≥340	0.62%	≥350	0.64%	0.02%
	Asian	≥800	1.46%	≥820	1.49%	0.03%
	Black or African American	≥24,530	44.46%	≥24,070	43.39%	(1.07%)
	Native Hawaiian or Other Pacific	≥40	0.09%	≥40	0.09%	0.00%
	White	≥24,060	43.62%	≥24,620	44.38%	0.77%
	Two or More Races	≥1,430	2.60%	≥1,590	2.87%	0.28%

Copyright © 2019 by Louisiana Department of Education.

Calibration and Census Data: Mathematics (continued)						
		Calibration Sample		Census Data		
Grade		N	%	N	%	(Calib % - Census %)
6	All Students	≥52,800	100.00%	≥53,030	100.00%	0.00%
	Gender					
	Male	≥27,060	51.24%	≥27,200	51.31%	0.06%
	Female	≥25,740	48.76%	≥25,820	48.69%	(0.06%)
	Race Ethnicity					
	Hispanic/Latino	≥3,750	7.11%	≥3,580	6.77%	(0.34%)
	American Indian or Alaska Native	≥320	0.61%	≥330	0.64%	0.03%
	Asian	≥790	1.51%	≥810	1.53%	0.02%
	Black or African American	≥23,350	44.23%	≥23,460	44.25%	0.02%
	Native Hawaiian or Other Pacific	≥30	0.07%	≥40	0.09%	0.01%
	White	≥23,140	43.83%	≥23,420	44.16%	0.34%
	Two or More Races	≥1,380	2.62%	≥1,330	2.52%	(0.10%)
7	All Students	≥51,570	100.00%	≥51,800	100.00%	0.00%
	Gender					
	Male	≥26,480	51.35%	≥26,650	51.45%	0.10%
	Female	≥25,090	48.65%	≥25,140	48.55%	(0.10%)
	Race Ethnicity					
	Hispanic/Latino	≥3,900	7.58%	≥3,390	6.55%	(1.03%)
	American Indian or Alaska Native	≥350	0.69%	≥370	0.72%	0.03%
	Asian	≥800	1.56%	≥790	1.53%	(0.02%)
	Black or African American	≥22,850	44.32%	≥22,710	43.85%	(0.47%)
	Native Hawaiian or Other Pacific	≥30	0.07%	≥20	0.04%	(0.03%)
	White	≥22,310	43.27%	≥23,360	45.10%	1.83%
	Two or More Races	≥1,290	2.50%	≥1,120	2.17%	(0.34%)
8	All Students	≥45,020	100.00%	≥45,380	100.00%	0.00%
	Gender					
	Male	≥23,330	51.83%	≥23,350	51.46%	(0.37%)
	Female	≥21,680	48.17%	≥22,020	48.54%	0.37%
	Race Ethnicity					
	Hispanic/Latino	≥3,130	6.97%	≥2,910	6.42%	(0.55%)
	American Indian or Alaska Native	≥280	0.63%	≥350	0.79%	0.16%
	Asian	≥720	1.61%	≥620	1.37%	(0.24%)
	Black or African American	≥20,180	44.84%	≥21,210	46.74%	1.91%
	Native Hawaiian or Other Pacific	≥30	0.08%	≥30	0.07%	(0.01%)
	White	≥19,430	43.16%	≥19,330	42.60%	(0.56%)
	Two or More Races	≥1,220	2.71%	≥890	1.98%	(0.73%)

Copyright © 2019 by Louisiana Department of Education.

All 2018 LEAP 2025 item calibration and linking were performed based on IRT.

Calibration and linking methodology used for the Spring 2018 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 Technical Report for 2016 Administration* (see https://parcc-assessment.org/wp-content/uploads/2018/02/PARCC-2016-Tech-Report.pdf).

There were two test forms, CBT and PBT, for the 2018 LEAP 2025 grades 3 and 4 ELA and mathematics assessments. Only CBT forms were administered for the grades 5 through 8 ELA and mathematics assessments. In general, a school administered the same test mode for ELA and mathematics. Table 6.21 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 2018 LEAP 2025 tests on the 2017 LEAP 2025 scale, which are on 2016 PARCC baseline scale:

1. Calibrate the 2018 LEAP 2025 tests.
2. Link 2018 LEAP 2025 tests, except for the grade 6 ELA test, to the LEAP 2025 scale under the non-equivalent common item design. The grade 6 ELA test was directly linked to 2016 PARCC baseline scale.

PARCC established a new baseline scale using 2016 PARCC Spring tests. LEAP 2016 and 2017 were directly linked to this new PARCC 2016 baseline scale using PARCC item parameters as anchor item parameters. Therefore, LEAP 2016 and 2017 were placed on the PARCC scale. Since LEAP 2016 and 2017 tests were calibrated with Louisiana students, the scale for LEAP 2016 and 2017 tests will be referred to as LEAP 2025 scale, although its scale was placed on PARCC scales built with PARCC associated states' data. Spring 2018 LEAP 2025 forms were linked to the LEAP 2025 scale using LEAP items, which were administered in LEAP 2025 forms

Copyright © 2019 by Louisiana Department of Education.

in 2016-2018 as anchors by the Stocking & Lord procedure. Since the 2018 anchor items are on PARCC scale, 2018 LEAP 2025 forms continue to be considered on the PARCC scale.

6.3.1 Calibration of 2018 LEAP 2025 Tests

For 2018 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 6.21 shows the number of students in the calibration samples and number of calibration items by mode. About 97% of grade 3 students took the PBT, and about 3% of grade 3 students took the CBT. About 89% of grade 4 students took PBT, and about 11% of grade 4 students took 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 in the N-Items columns because calibration does not include reading item scores; it only includes WE item scores. A reading item score and a WE item score for the same writing prompt are the same. There were 27~34 ELA items and 42~43 mathematics items across grades.

Table 6.21 Summary of Student Count in Calibration Sample and Item Count by Test Mode

		N			Percentage		N-Items	
Content	Grade	All	CBT	PBT	CBT	PBT	CBT	PBT
ELA	3	≥34,460	≥1,070	≥33,380	3.12	96.88	27	27
	4	≥45,320	≥5,210	≥40,110	11.50	88.50	30	30
	5	≥55,490	≥55,490	*	100.00	*	31	*
	6	≥52,970	≥52,970	*	100.00	*	33	*
	7	≥51,330	≥51,330	*	100.00	*	33	*
	8	≥51,290	≥51,290	*	100.00	*	34	*
Mathematics	3	≥44,710	≥1,080	≥43,630	2.42	97.58	43	43
	4	≥55,010	≥5,210	≥49,790	9.47	90.53	43	43
	5	≥55,170	≥55,170	*	100.00	*	43	*
	6	≥52,800	≥52,800	*	100.00	*	43	*
	7	≥51,570	≥51,570	*	100.00	*	43	*
	8	≥45,020	≥45,020	*	100.00	*	42	*

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

6.3.1.1. Concurrent Calibration for PBT and CBT

For 2018 LEAP 2025 calibration, CBT and PBT were combined and calibrated together for grades 3 and 4 based on mode effect study (section 10.4).

6.3.1.2. Separate Calibration for ELA Prose Constructed-Response Tasks

To address the issue of local independence for ELA prose-constructed response (PCR) tasks, the sparse matrix method was applied for grades 5 to 8. Each ELA test consisted of two PCR tasks; each task had a written expression (WE) and a written knowledge and use of the language (WKL) trait. As can be seen in Table 6.22, a single calibration was performed for grades 5 to 8 by randomly splitting the students into two groups. Almost half of the data set included responses to other items and responses to two WE traits, and the other calibration data set included the same responses to other items and responses to two WKL traits. Therefore, WE

Copyright © 2019 by Louisiana Department of Education.

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

Group	Other Items	WE	WKL
I	XXXXXXXX	XX	
II	XXXXXXXX		XX

For ELA grades 3 and 4, the same sample repeated design was applied to the WE and WKL calibration due to the small number of students in maximum score points of WE items. For each grade, two datasets of responses were generated. One calibration dataset included two WE responses and responses of the other items and excluded the WKL responses. The other calibration dataset included two WKL responses and the same responses of the other items and excluded the WE responses. In these datasets, the responses of the other items, except for WE and WKL, were the same for both forms, and each dataset included either WE or WKL responses. After each dataset was separately calibrated, the item parameters with WKL responses were equated to those with WE responses using all common items as anchor items. Table 6.23 illustrates this design.

Table 6.23 Calibration Data Structure for ELA WE and WKL Traits with the Same Sample Repeated Design

Dataset	Group	Other Items	WE	WKL
1	All Students	XXXXXXXX	XX	
2	All Students	XXXXXXXX		XX

6.3.1.3. IRT Item Fit

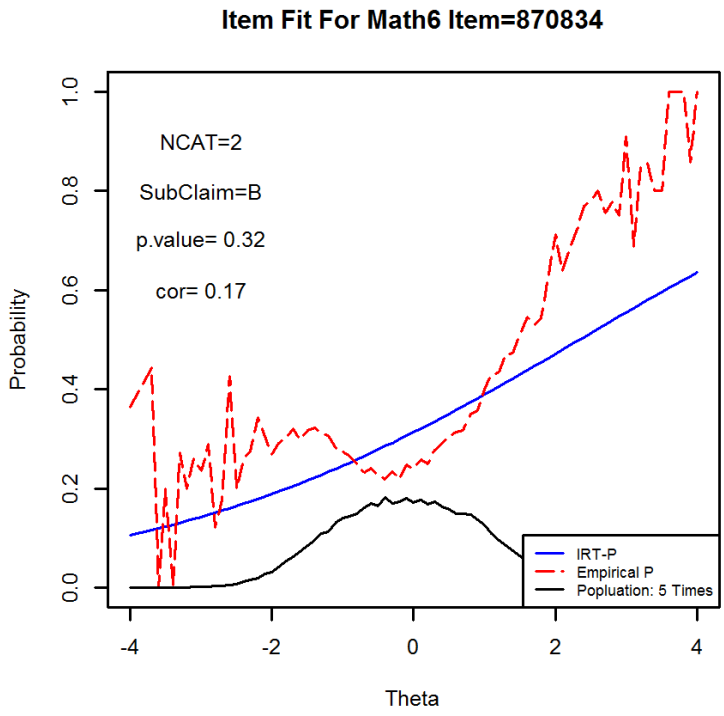
The usefulness of IRT models is dependent on the extent to which they effectively reflect the data. Hambleton, Swaminathan, and Rogers (1991) explain, “The advantages of item response models can be obtained only when the fit between the model and the test data of interest is satisfactory. A poorly fitting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

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

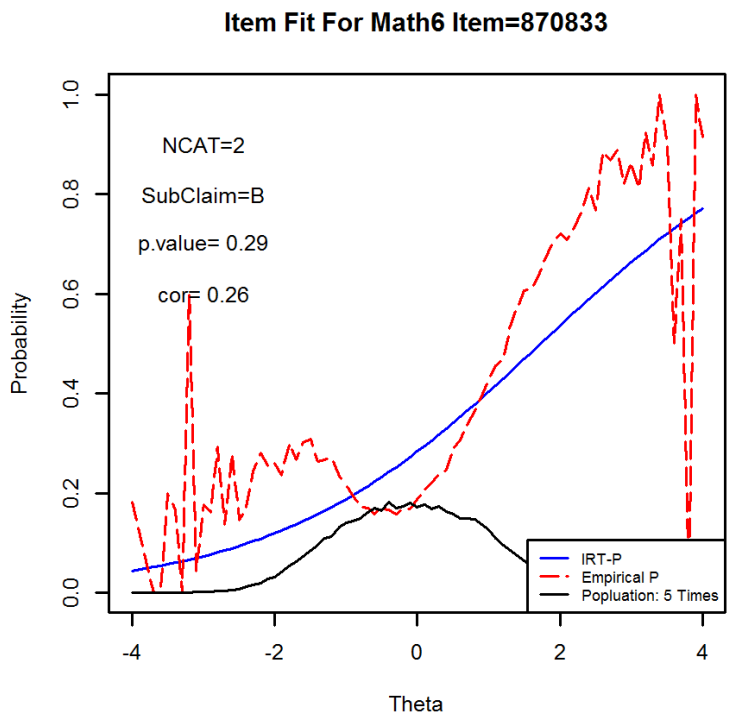
After convergence was achieved for each IRT data set, an item characteristic curve (ICC) for each item was plotted together with empirical students’ performances from theta ability -4 to 4. No items were suppressed from calibration due to poor fit. However, three items were removed from the anchor sets due to poor fit. The fit plots for the two items removed from the grade 6 mathematics anchor set and the grade 8 anchor set are seen in Figure 6.1.

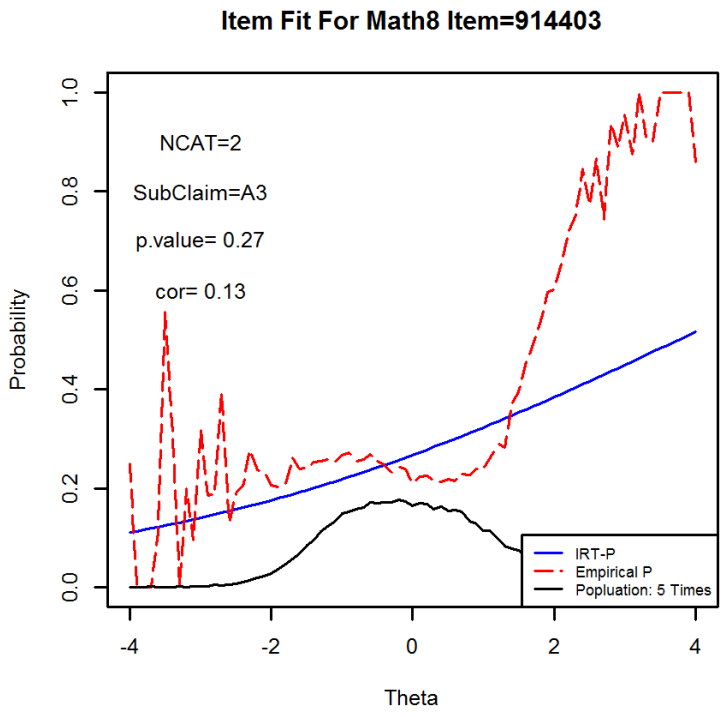
Copyright © 2019 by Louisiana Department of Education.

Figure 6.1 Item Fit Plots of Items Removed from Anchor Sets



Copyright © 2019 by Louisiana Department of Education.





After calibration, the IRT model fit 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).

Copyright © 2019 by Louisiana Department of Education.

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.

Tables 6.24 and 6.25 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 0.16 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 0.28 for mathematics grade 7.

Table 6.24 Summary of Adjusted Fit for ELA

Grade	Mode	No. Items	Mean	Std. Dev.	Min.	Max.	No. Flagged Items	No. Not Available
3	CBT/PBT	30	0.16	0.04	0.08	0.23	0	0
4	CBT/PBT	33	0.15	0.05	0.07	0.27	0	0
5	CBT	30	0.14	0.05	0.07	0.23	0	0
6	CBT	32	0.14	0.05	0.05	0.24	0	0
7	CBT	32	0.13	0.05	0.06	0.23	0	0
8	CBT	32	0.14	0.05	0.05	0.24	0	0

Table 6.25 Summary of Adjusted Fit for Mathematics

Grade	Mode	No. Items	Mean	Std. Dev.	Min.	Max.	No. Flagged Items	No. Not Available
3	CBT/PBT	43	0.09	0.03	0.04	0.17	0	0
4	CBT/PBT	43	0.10	0.04	0.04	0.23	0	0
5	CBT	43	0.10	0.04	0.04	0.18	0	0
6	CBT	43	0.09	0.04	0.04	0.20	0	0
7	CBT	43	0.10	0.05	0.03	0.28	0	0
8	CBT	42	0.08	0.04	0.04	0.17	0	0

6.3.2 **Linking 2018 LEAP 2025 Grades 3–8 to PARCC Scale**

The 2016 and 2017 LEAP 2025 forms were linked to the PARCC scale using intact PARCC items embedded into the LEAP 2025 forms by using the Stocking & Lord procedure (1983). Therefore, these item parameters were placed on PARCC scale. However, these equated

Copyright © 2019 by Louisiana Department of Education.

Louisiana item parameters are based on only Louisiana students' responses while intact PARCC item parameters were estimated based on PARCC associated states' responses. To distinguish these two item parameters, item parameters based on only Louisiana students' responses will be called LEAP item parameters and its scale as the LEAP scale.

Three anchor sets were created for the 2018 Spring LEAP 2025 ELA and mathematics assessments equating process. Anchor 1 items were intact PARCC items embedded in the 2018 LEAP 2025 form. Anchor 2 items were common items between the 2017 and 2018 LEAP 2025 spring forms, and their item parameters were the 2016 or 2017 LEAP 2025 item parameters. Anchor 3 item parameters consisted of all Anchor 2 item parameters and Anchor 1 item parameters were used for the remaining items. Anchor 2 was used in the operational analyses to link to the LEAP 2025 scale, which is the same as the PARCC scale, and Anchor 1 and Anchor 3 were used to help evaluate drift from the PARCC scale. Table 6.26 provides the Stocking & Lord transformation constants that were used to link to scale. Table 6.27 summarizes the number and score points of the initial anchor item selection before equating. Table 6.27 also summarizes the number and score points of the final anchor item selections. The difference between the initial number of anchor items and the final number of anchor items is the number of anchor items dropped.

Table 6.26 Stocking & Lord Transformation Constants

Content	Grade	Slope	Intercept
ELA	3	0.998	0.255
	4	0.961	0.118
	5	0.930	0.178
	6	0.954	0.001
	7	0.959	0.043
	8	0.940	0.041
Mathematics	3	0.969	-0.009
	4	0.988	0.043
	5	0.953	-0.201
	6	0.979	-0.177
	7	0.979	-0.089
	8	0.981	0.029

Copyright © 2019 by Louisiana Department of Education.

Table 6.27 Number and Score Points of Initial and Final Anchor Item Sets

Content	Grade		Anchor 1		Anchor 2		Anchor 3	
			Number of Items	Score Points	Number of Items	Score Points	Number of Items	Score Points
ELA	3	Initial	23	60	12	30	23	60
		Final	17	41	9	17	15	30
	4	Initial	21	62	14	37	21	62
		Final	19	58	13	35	15	49
	5	Initial	29	77	13	37	29	77
		Final	23	65	9	18	19	39
	6	Initial	32	86	15	41	32	86
		Final	22	66	14	38	26	74
	7	Initial	32	86	15	41	32	86
		Final	23	57	12	34	22	66
	8	Initial	31	84	13	37	31	84
		Final	24	61	9	19	24	70
Mathematics	3	Initial	43	62	16	19	43	62
		Final	31	42	11	12	35	48
	4	Initial	43	56	16	19	43	56
		Final	35	43	14	15	32	42
	5	Initial	43	62	15	20	43	62
		Final	34	51	11	14	34	53
	6	Initial	41	64	13	17	41	64
		Final	30	49	11	13	32	52
	7	Initial	43	66	15	20	43	66
		Final	34	55	11	14	32	52
	8	Initial	41	65	15	21	41	65
		Final	35	58	12	15	32	49

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

Figures 6.2 to 6.13 are test characteristic curves for anchor items, corresponding 2018 LEAP 2025 estimated anchor items (EQ_ANC), 2017 LEAP 2025 operational items (LEAP2017), and all 2018 LEAP 2025 estimated items (EQ_ALL) for ELA and mathematics after applying the Stocking & Lord equating procedure. The blue solid line illustrates the anchor items, the red dotted line is the 2018 LEAP 2025 equated anchor items, the black dash line is the 2017 LEAP 2025 operational items, and the green dotted line is for all 2018 LEAP 2025 equated items. Anchor items for each anchor set 1, 2, and 3 are different as mentioned above. For most ELA and mathematics grades, the TCCs for anchor items and the corresponding 2018 estimated anchor items were overlapped across most ability levels.

For ELA, the four TCCs were overlapped over most ability levels across all five grades. The overlap of TCCs indicates the same equating results from three different equating anchor sets. For ELA grade 6, the TCC for all LEAP items showed some differences compared to the other

Copyright © 2019 by Louisiana Department of Education.

two TCCs at the cut score of 750. For mathematics, the three TCCs were also overlapped over most ability levels across all six grades. There were some differences at the extreme ranges, such as low ability or high ability.

Figures 6.14 to 6.25 present scatter plots of slope item parameters and difficulty item parameters for ELA and mathematics and their correlation after linking 2018 LEAP 2025 to the PARCC 2016 scale.

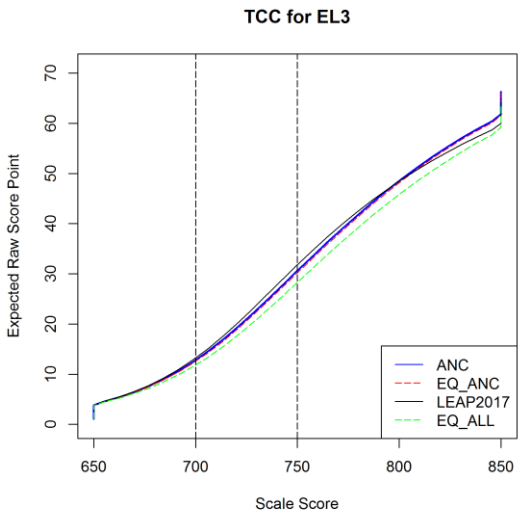
As can be seen in the ELA slope parameter plots, most parameters were around the identity line except for two items in ELA grade 5. The correlation between PARCC 2017 and 2018 LEAP 2025 anchors was 0.766 for ELA grade 5. For the other grades, correlations ranged from 0.905 to 0.985. For mathematics, most item slope parameters were around the identity line, and the correlations ranged from 0.795 to 0.933.

For ELA, most item difficulty parameters were around the identity line, and the correlations ranged from 0.966 to 0.991. Compared to ELA item difficulty parameters, mathematics item difficulty parameters were a little scattered from the identity line for most grades. Correlations ranged from 0.906 to 0.977 across grades. It is common to find higher correlations for difficulty parameters than those for slope parameters.

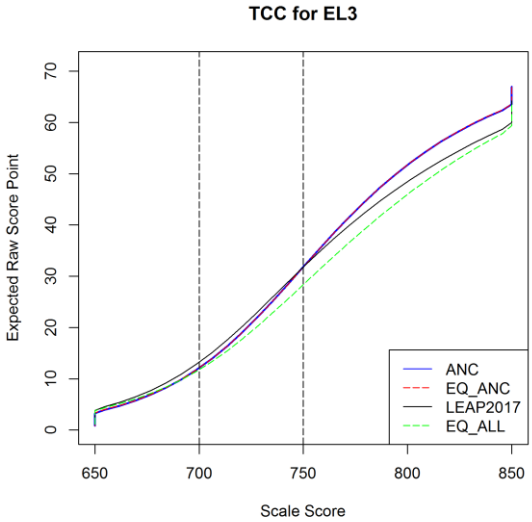
Copyright © 2019 by Louisiana Department of Education.

Figure 6.2 ELA Grade 3 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

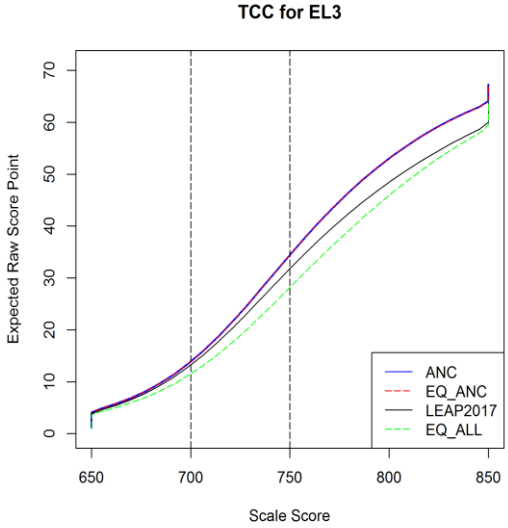
Anchor 1



Anchor 2



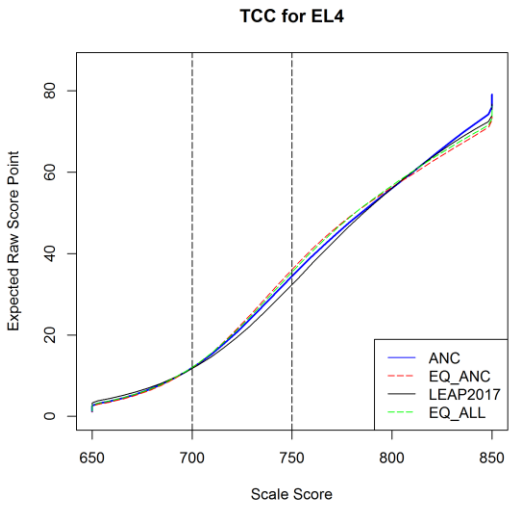
Anchor 3



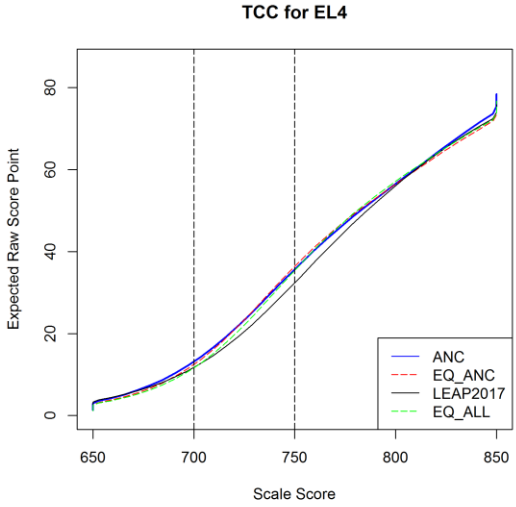
Copyright © 2019 by Louisiana Department of Education.

Figure 6.3 ELA Grade 4 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

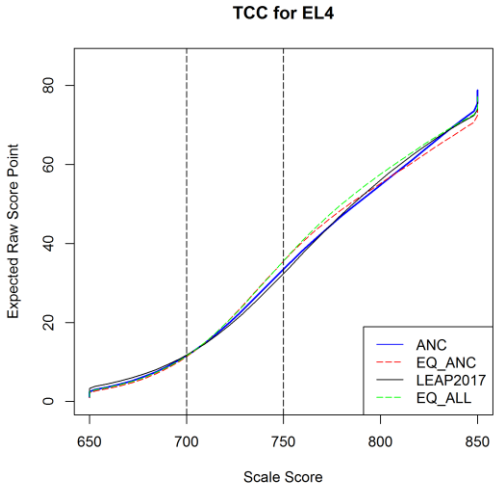
Anchor 1



Anchor 2



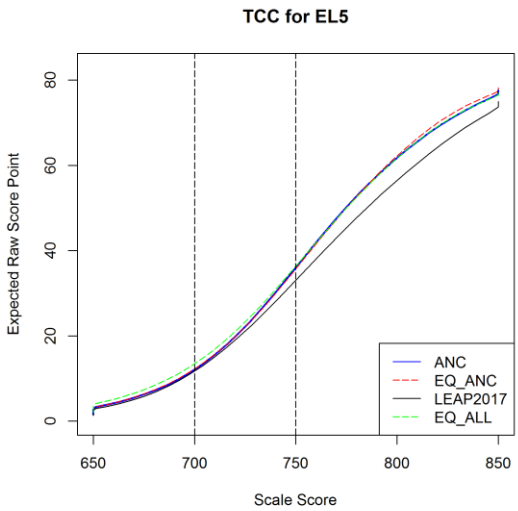
Anchor 3



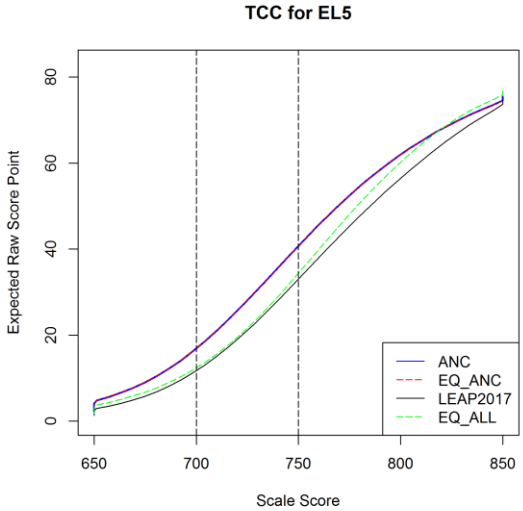
Copyright © 2019 by Louisiana Department of Education.

Figure 6.4 ELA Grade 5 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

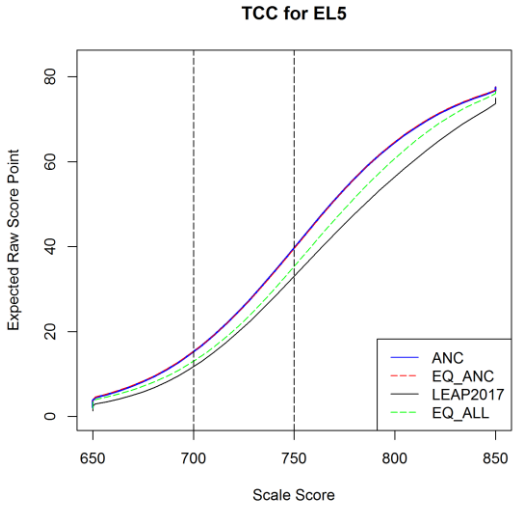
Anchor 1



Anchor 2



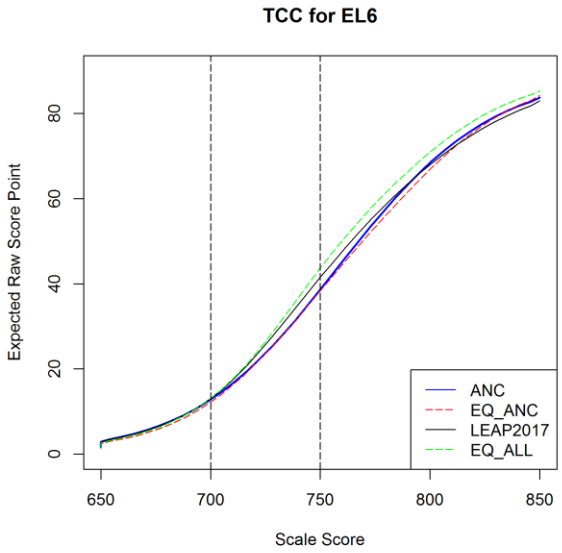
Anchor 3



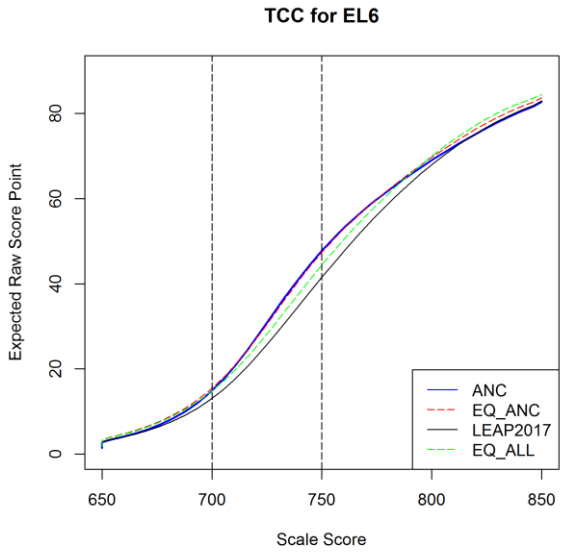
Copyright © 2019 by Louisiana Department of Education.

Figure 6.5 ELA Grade 6 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

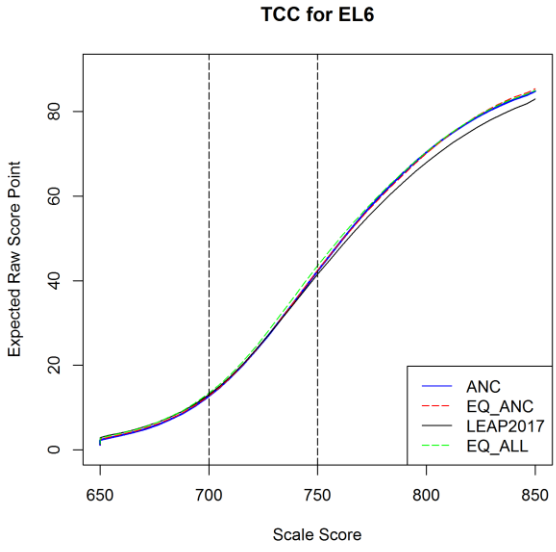
Anchor 1



Anchor 2



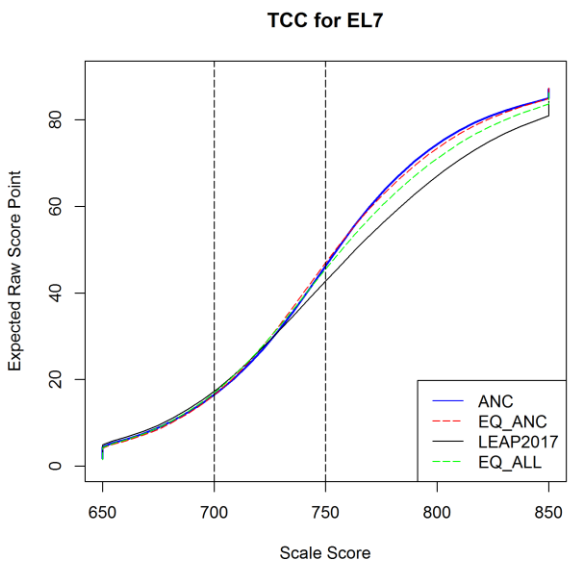
Anchor 3



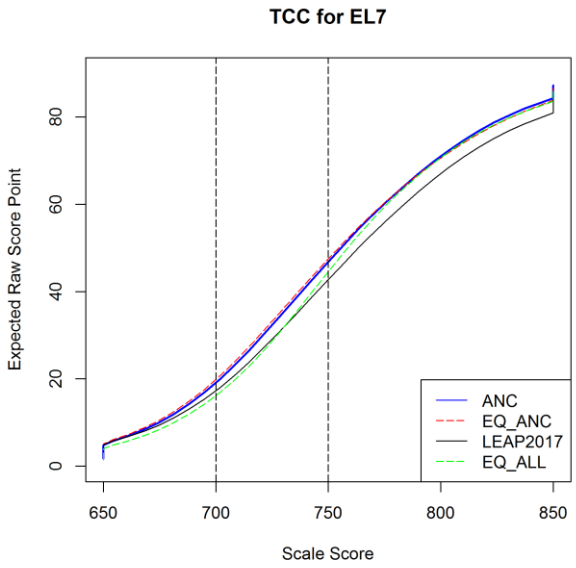
Copyright © 2019 by Louisiana Department of Education.

Figure 6.6 ELA Grade 7 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

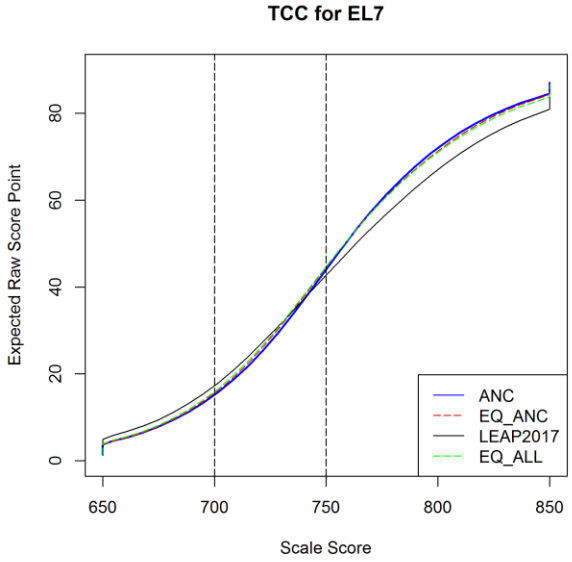
Anchor 1



Anchor 2



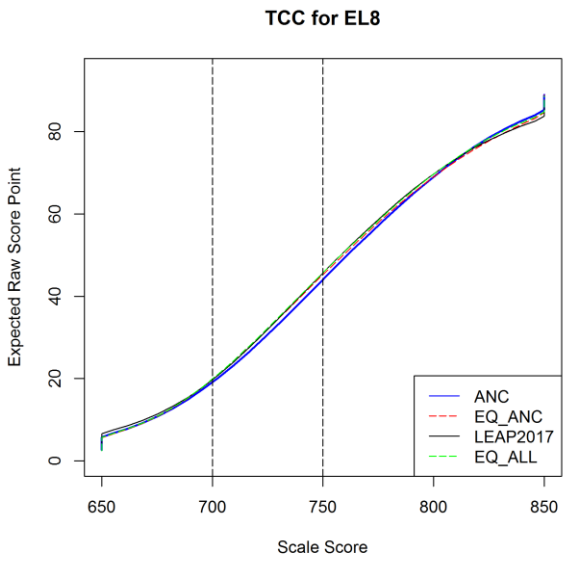
Anchor 3



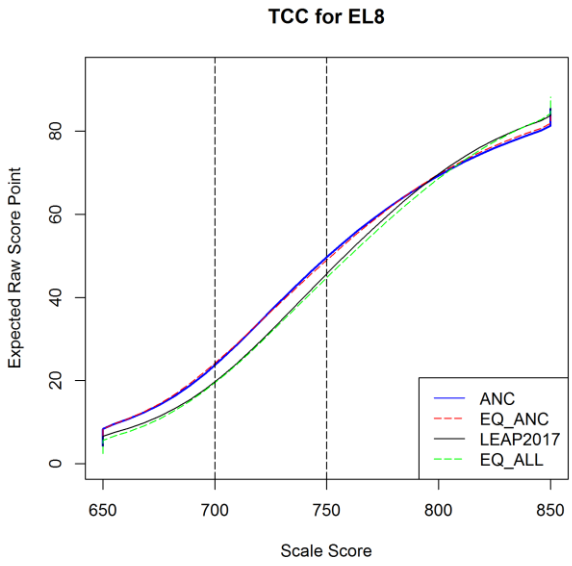
Copyright © 2019 by Louisiana Department of Education.

Figure 6.7 ELA Grade 8 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

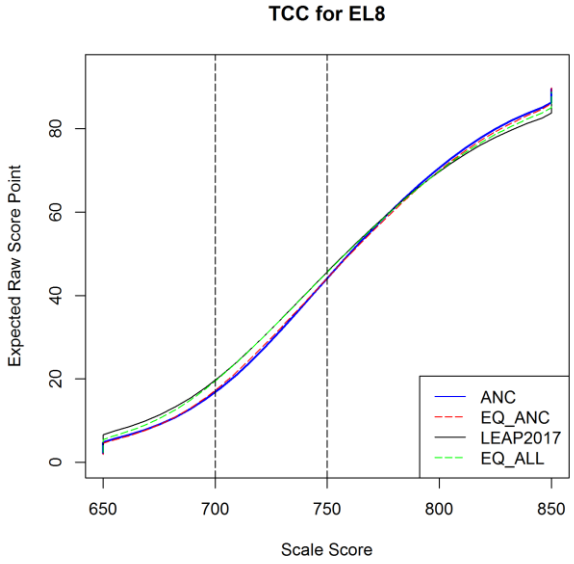
Anchor 1



Anchor 2



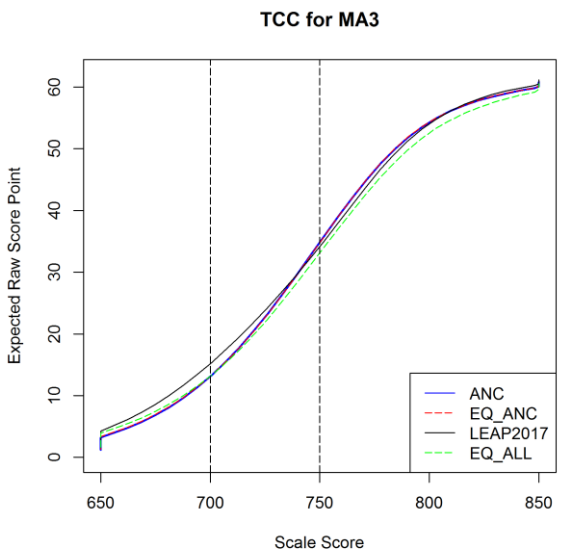
Anchor 3



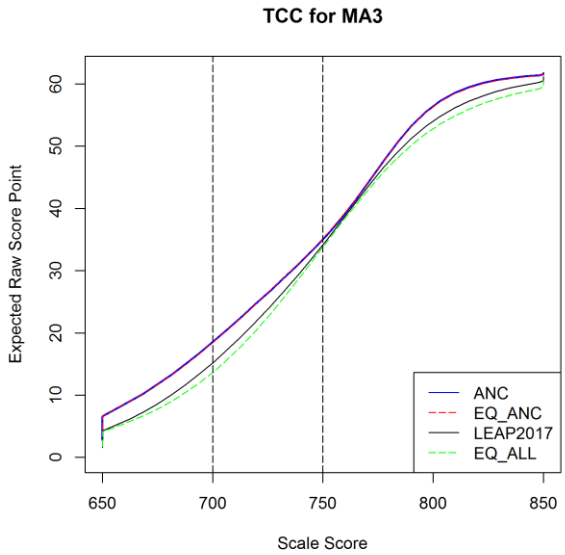
Copyright © 2019 by Louisiana Department of Education.

Figure 6.8 Mathematics Grade 3 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

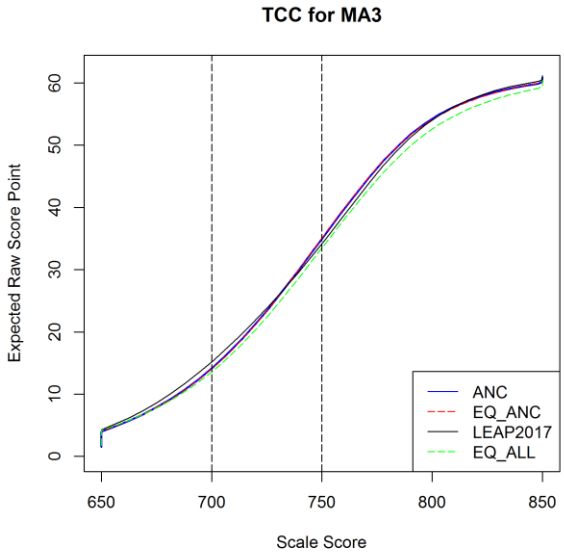
Anchor 1



Anchor 2



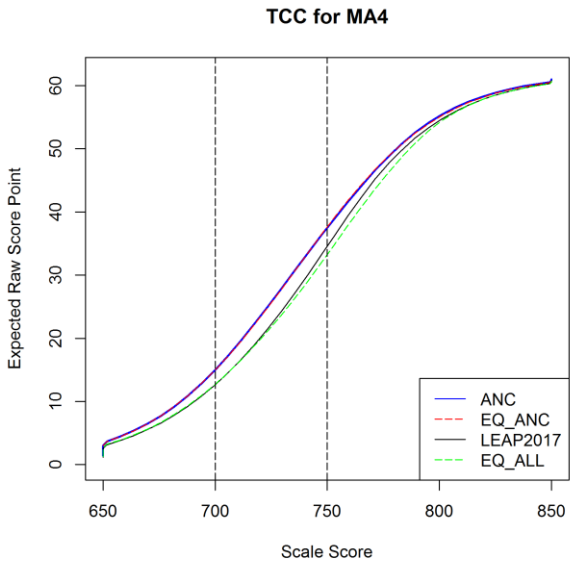
Anchor 3



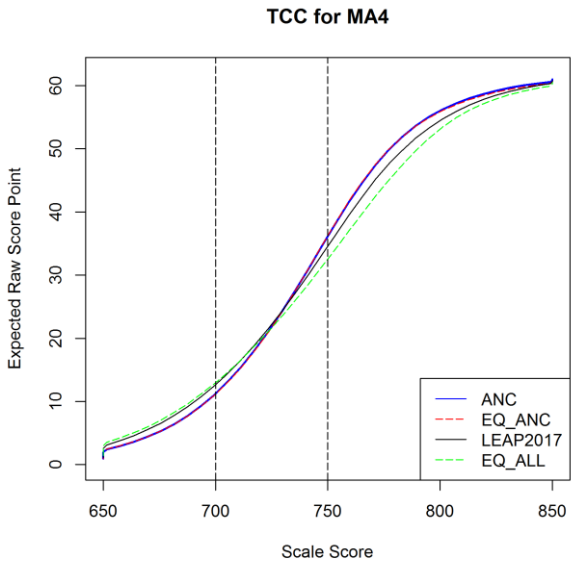
Copyright © 2019 by Louisiana Department of Education.

Figure 6.9 Mathematics Grade 4 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

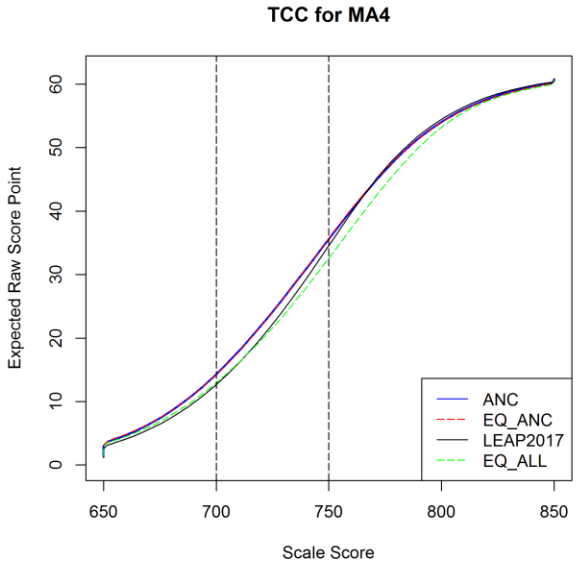
Anchor 1



Anchor 2



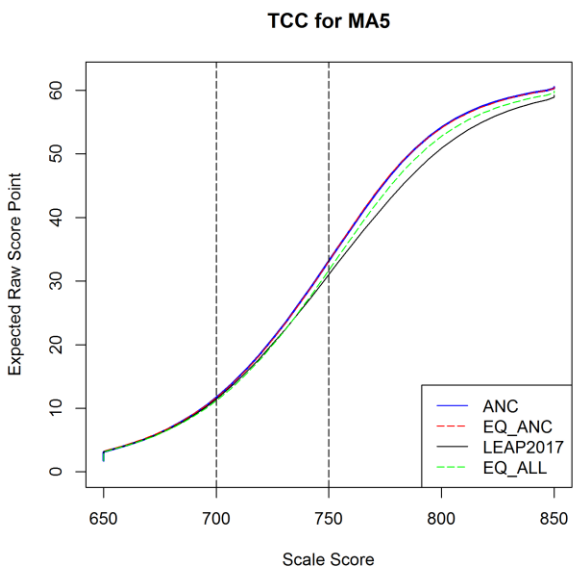
Anchor 3



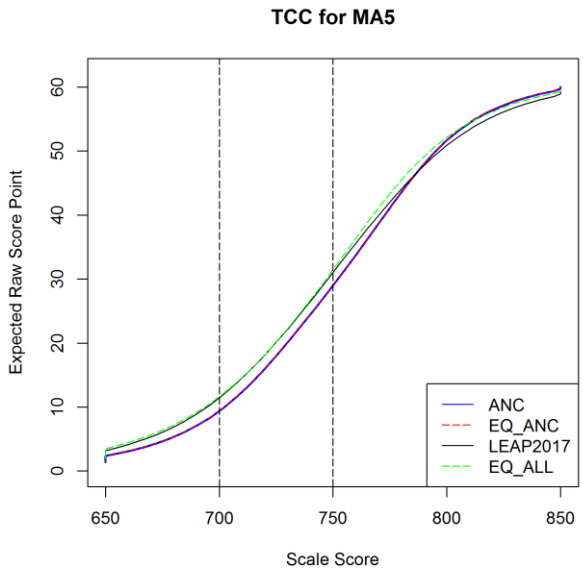
Copyright © 2019 by Louisiana Department of Education.

Figure 6.10 Mathematics Grade 5 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

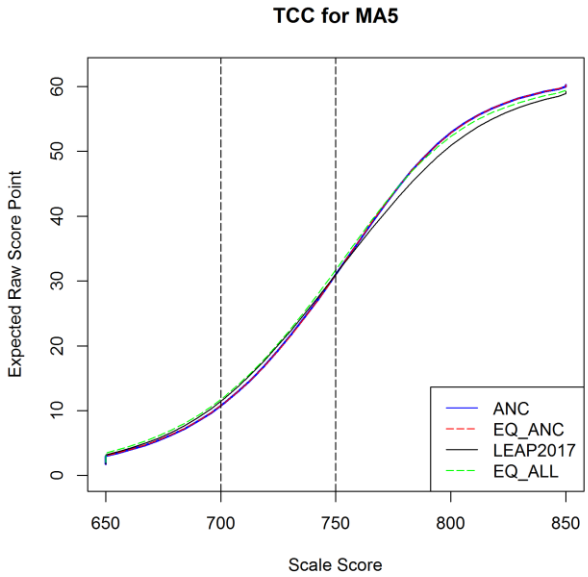
Anchor 1



Anchor 2



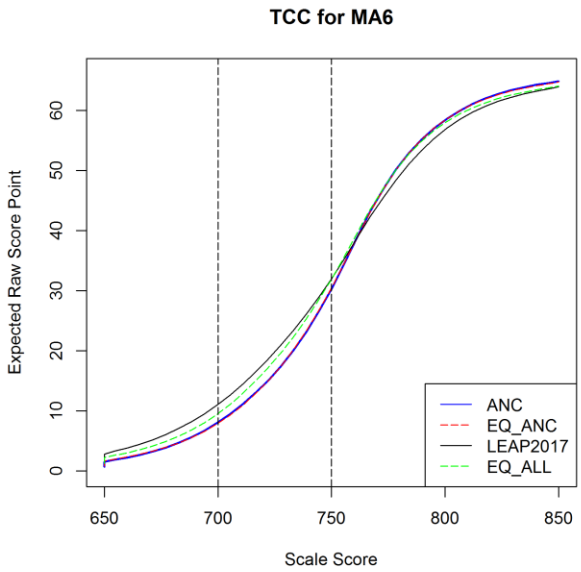
Anchor 3



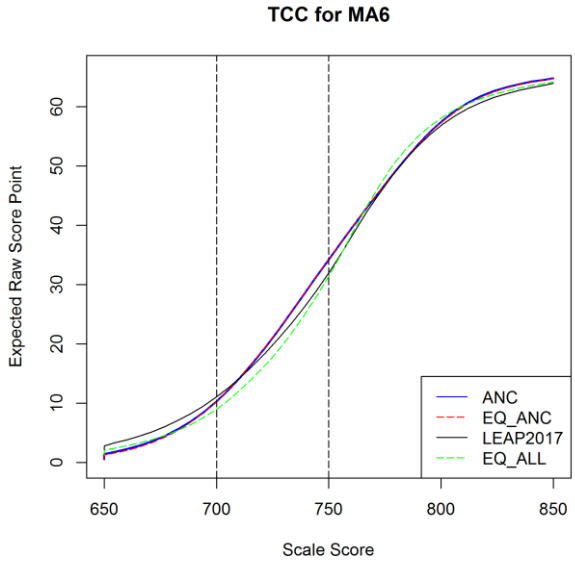
Copyright © 2019 by Louisiana Department of Education.

Figure 6.11 Mathematics Grade 6 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

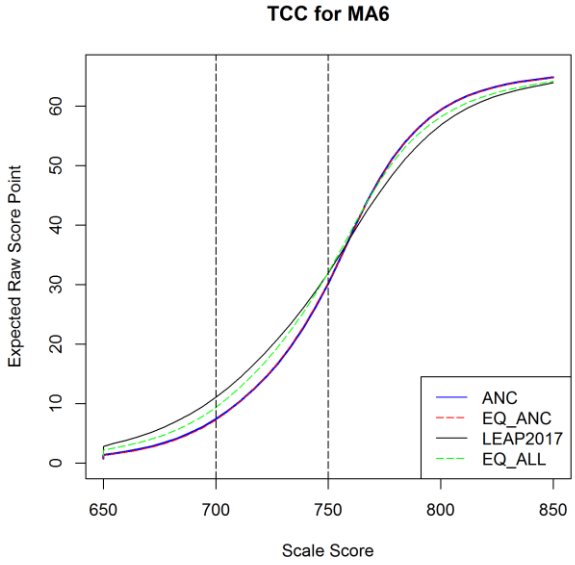
Anchor 1



Anchor 2



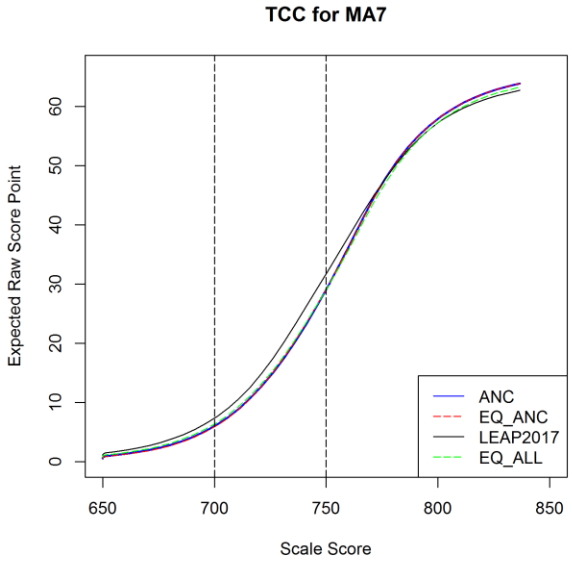
Anchor 3



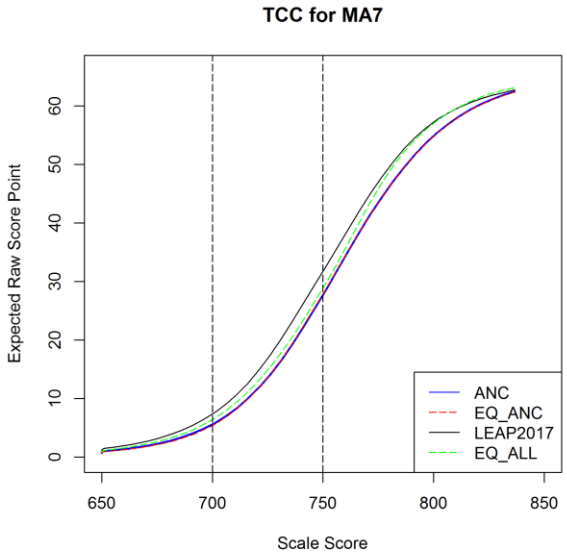
Copyright © 2019 by Louisiana Department of Education.

Figure 6.12 Mathematics Grade 7 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

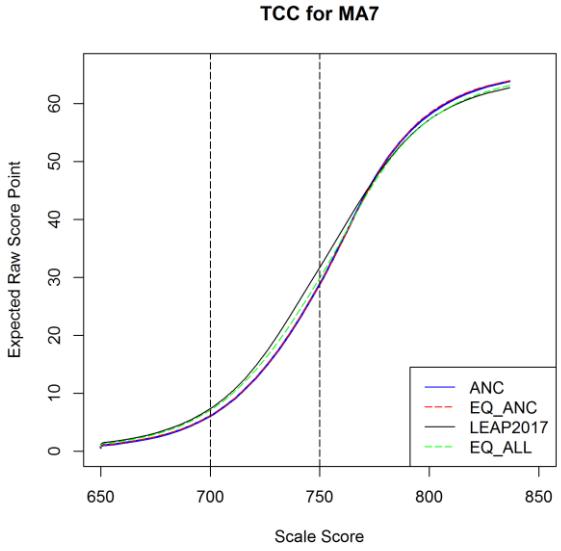
Anchor 1



Anchor 2



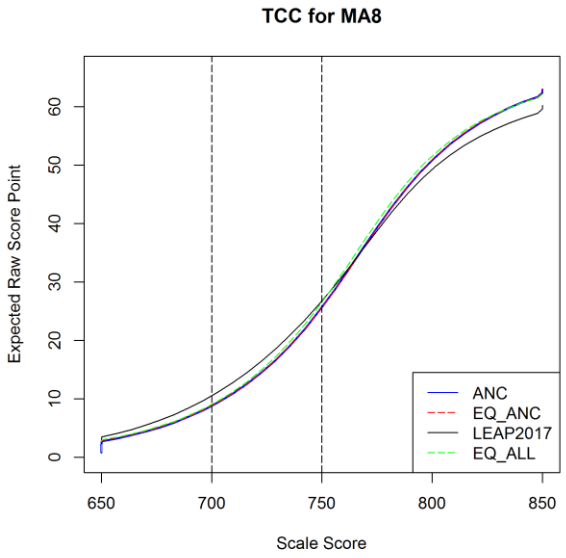
Anchor 3



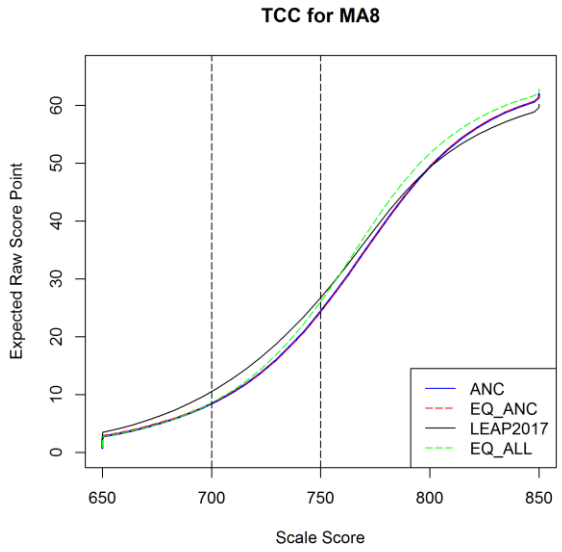
Copyright © 2019 by Louisiana Department of Education.

Figure 6.13 Mathematics Grade 8 TCC between Anchor, LEAP 2018 Equated Anchor, LEAP 2017, and ALL LEAP 2018 Items

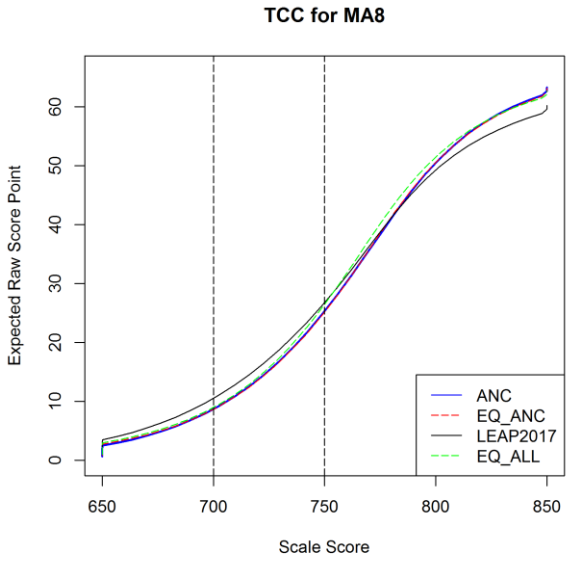
Anchor 1



Anchor 2



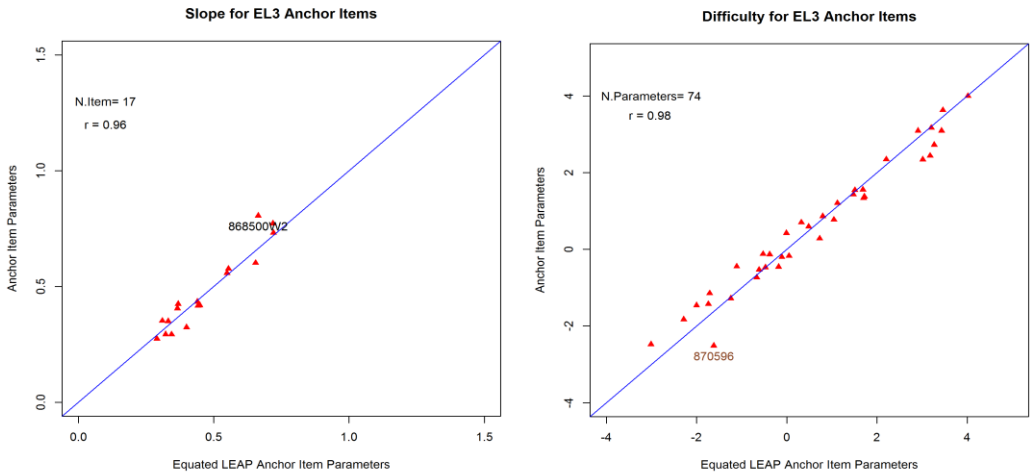
Anchor 3



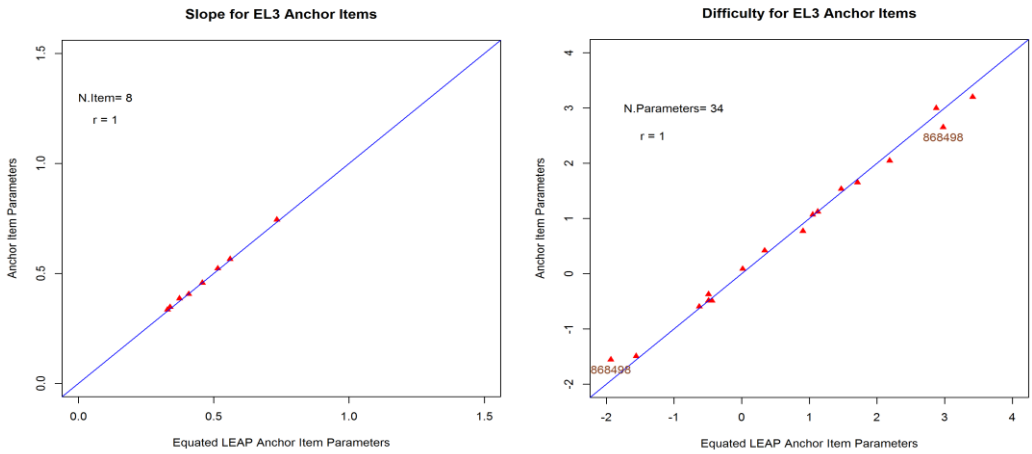
Copyright © 2019 by Louisiana Department of Education.

Figure 6.14 ELA Grade 3 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

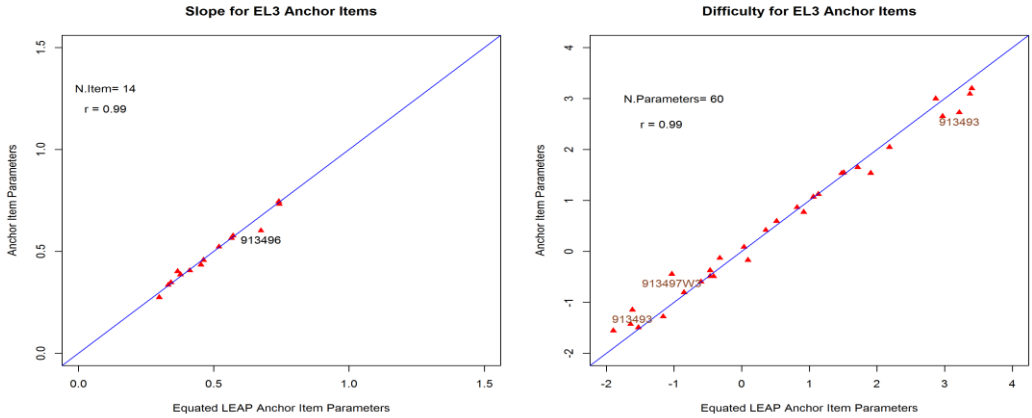
Anchor 1



Anchor 2



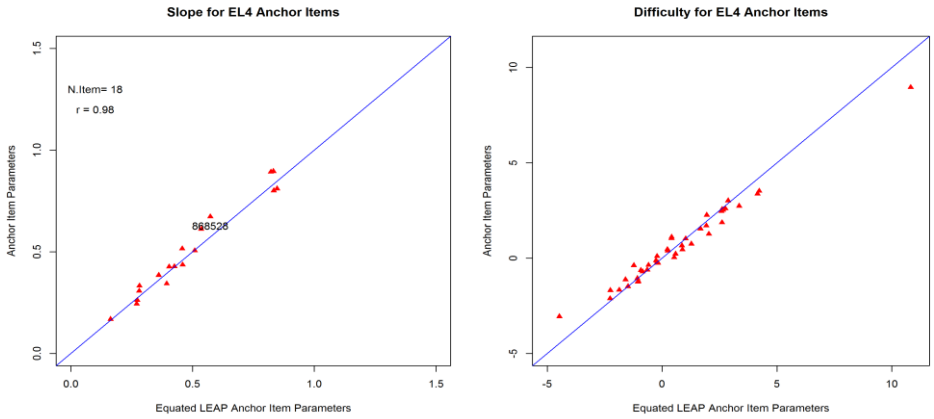
Anchor 3



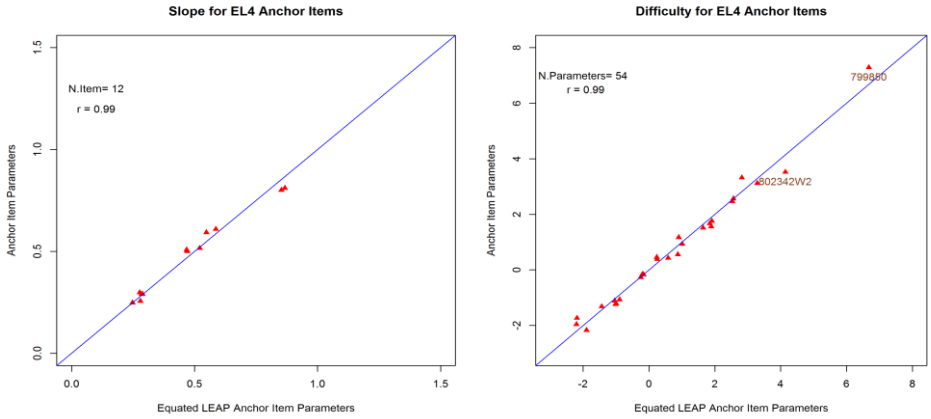
Copyright © 2019 by Louisiana Department of Education.

Figure 6.15 ELA Grade 4 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

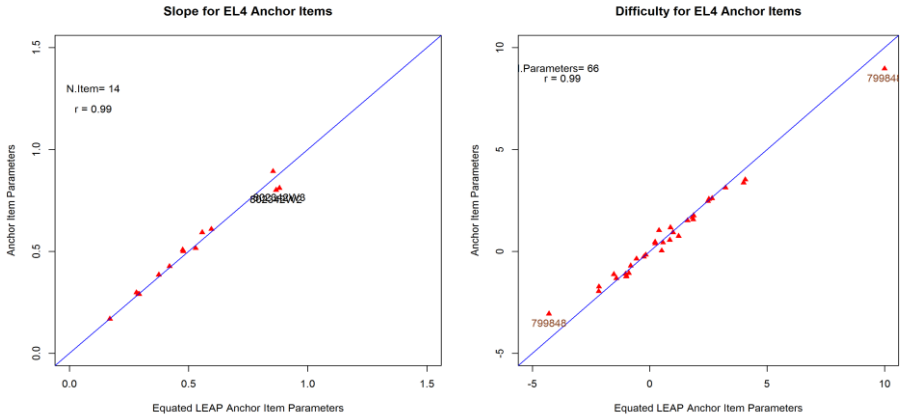
Anchor 1



Anchor 2



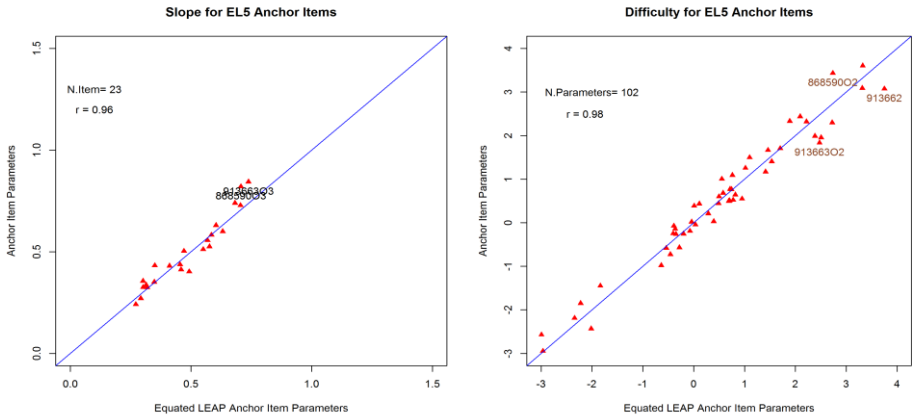
Anchor 3



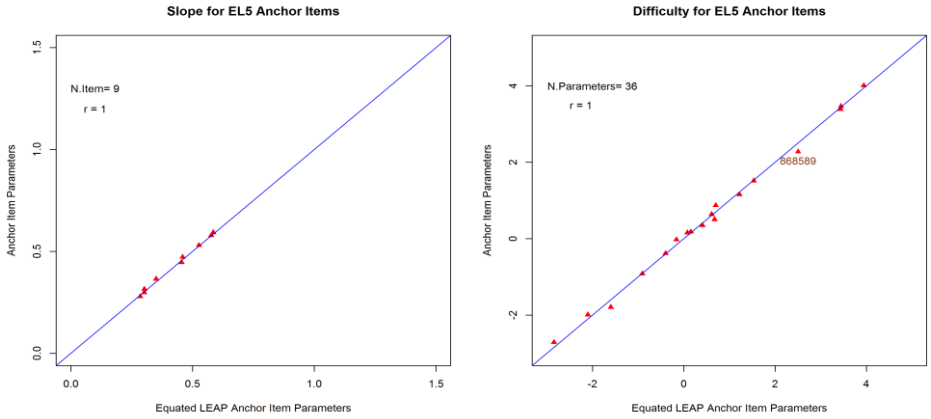
Copyright © 2019 by Louisiana Department of Education.

Figure 6.16 ELA Grade 5 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

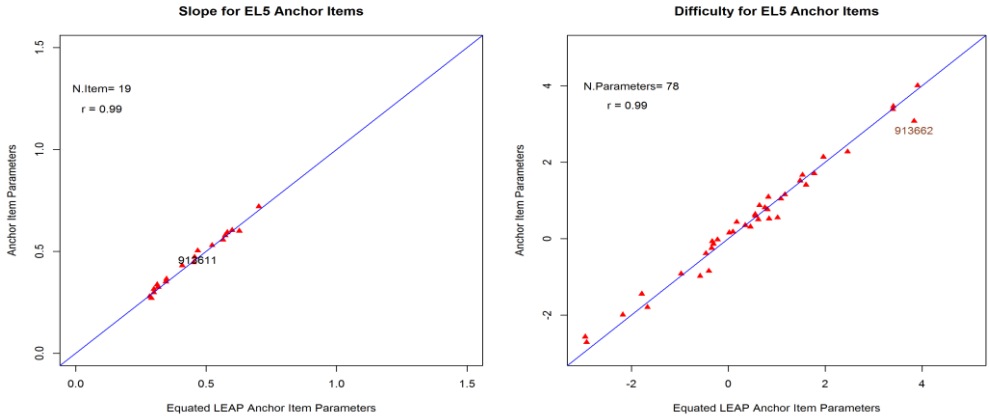
Anchor 1



Anchor 2



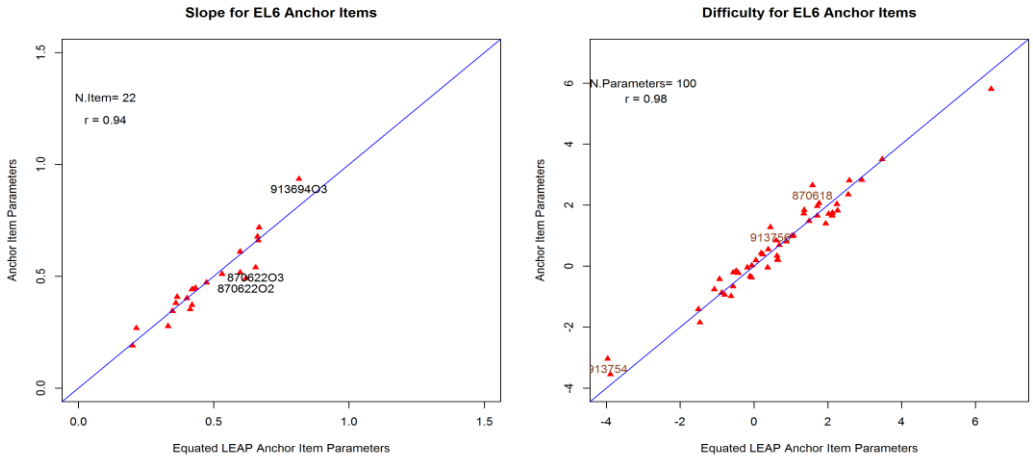
Anchor 3



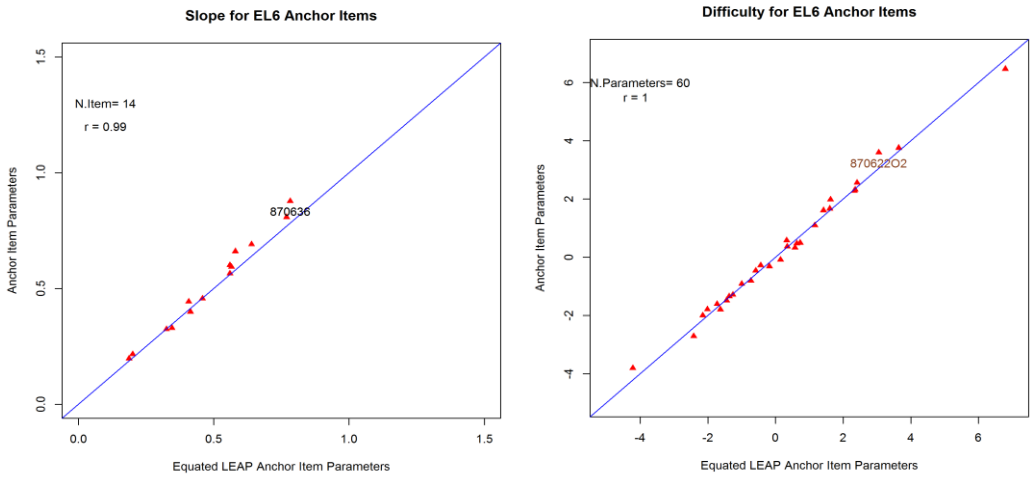
Copyright © 2019 by Louisiana Department of Education.

Figure 6.17 ELA Grade 6 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

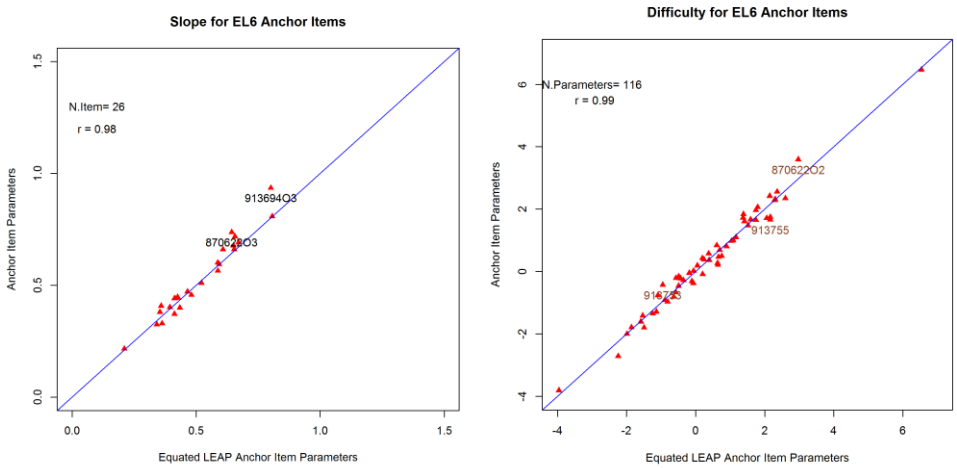
Anchor 1



Anchor 2



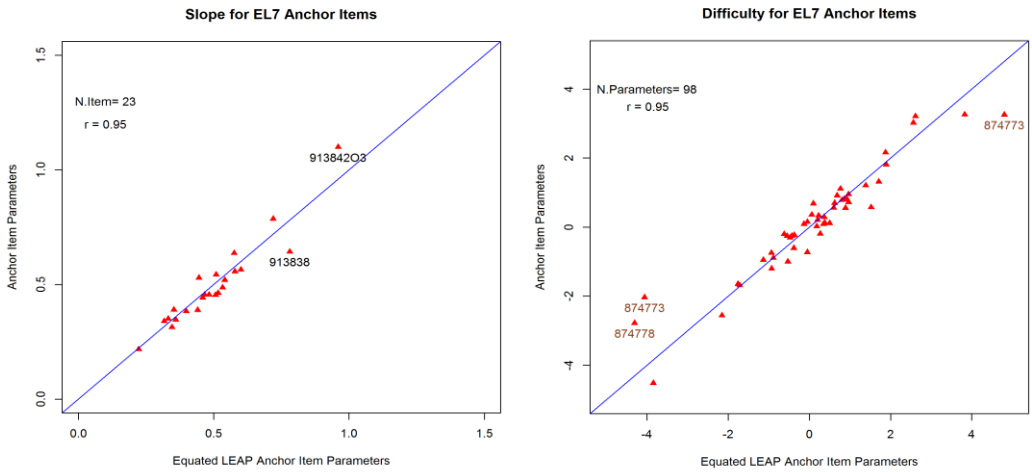
Anchor 3



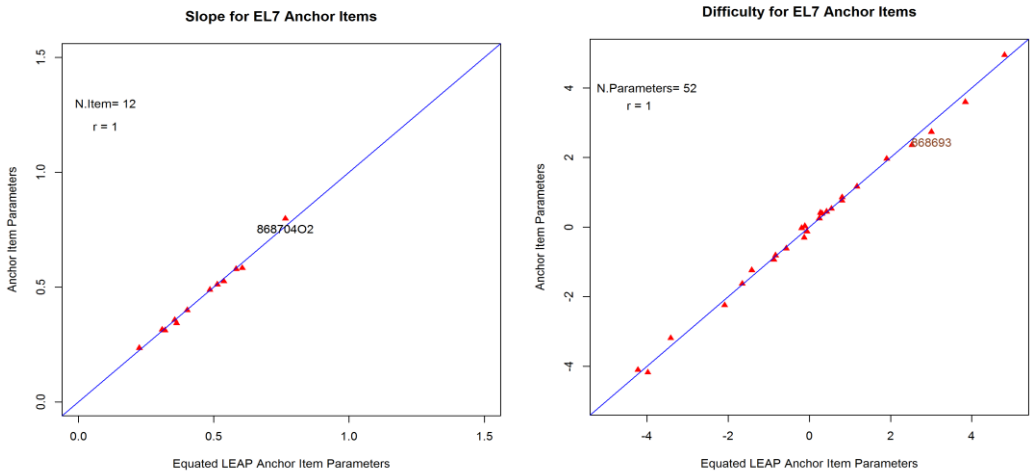
Copyright © 2019 by Louisiana Department of Education.

Figure 6.18 ELA Grade 7 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

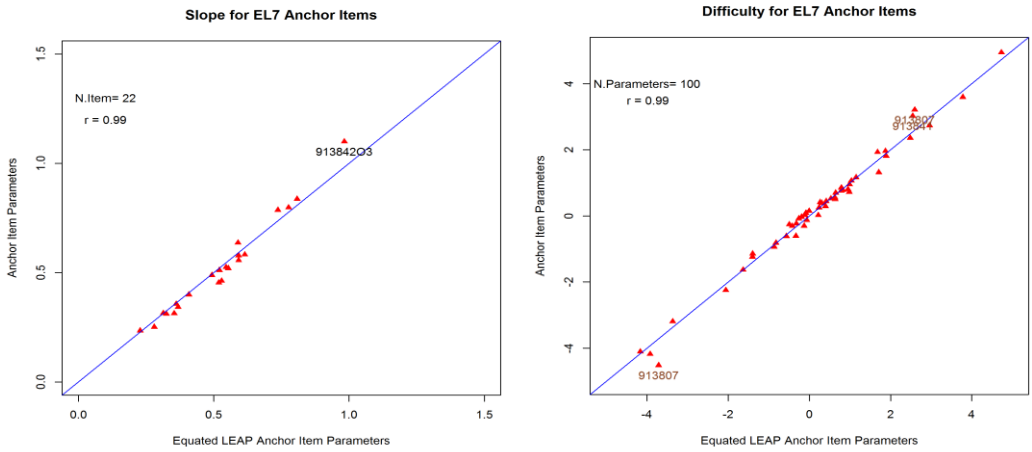
Anchor 1



Anchor 2



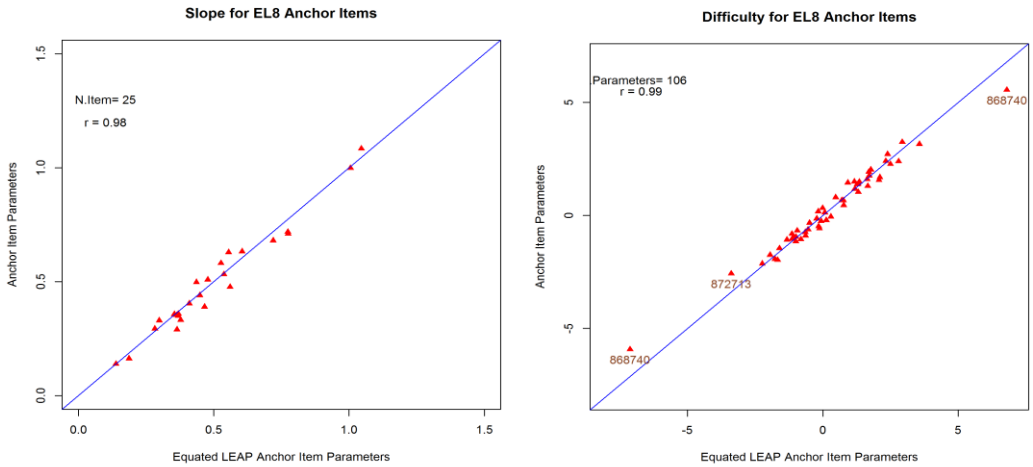
Anchor 3



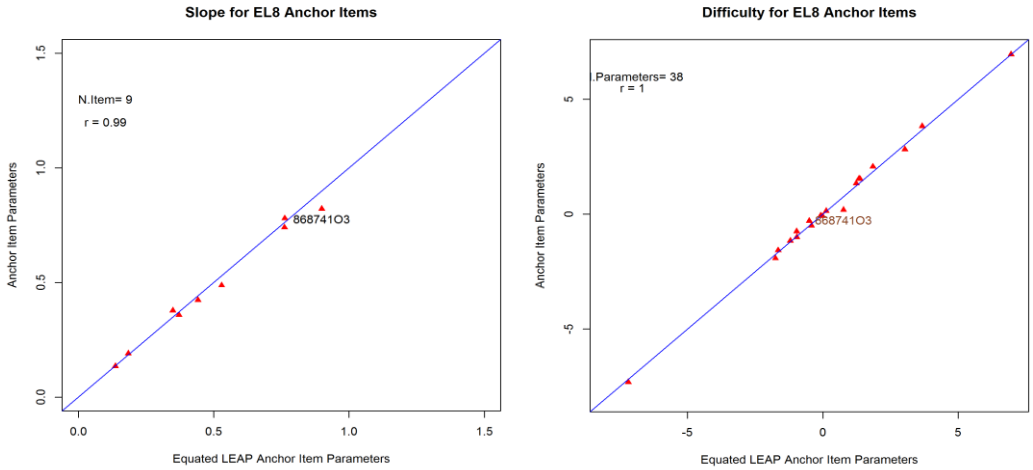
Copyright © 2019 by Louisiana Department of Education.

Figure 6.19 ELA Grade 8 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

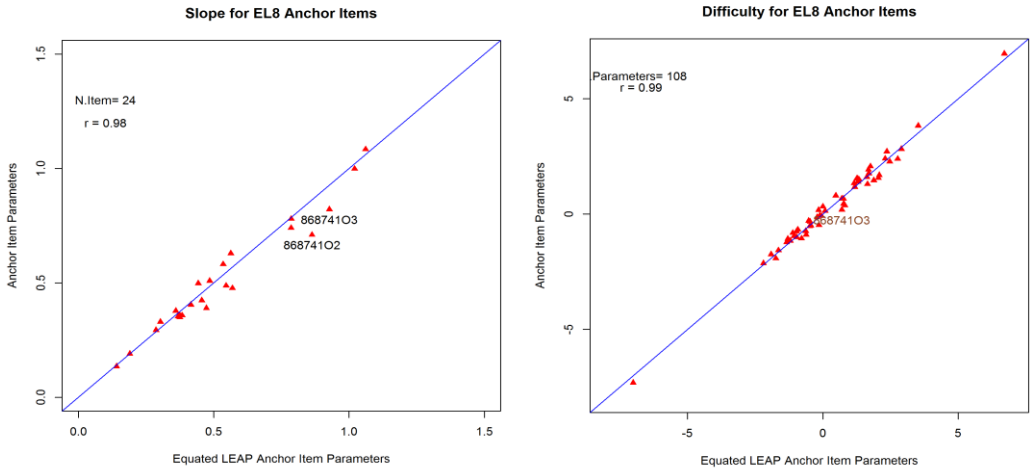
Anchor 1



Anchor 2



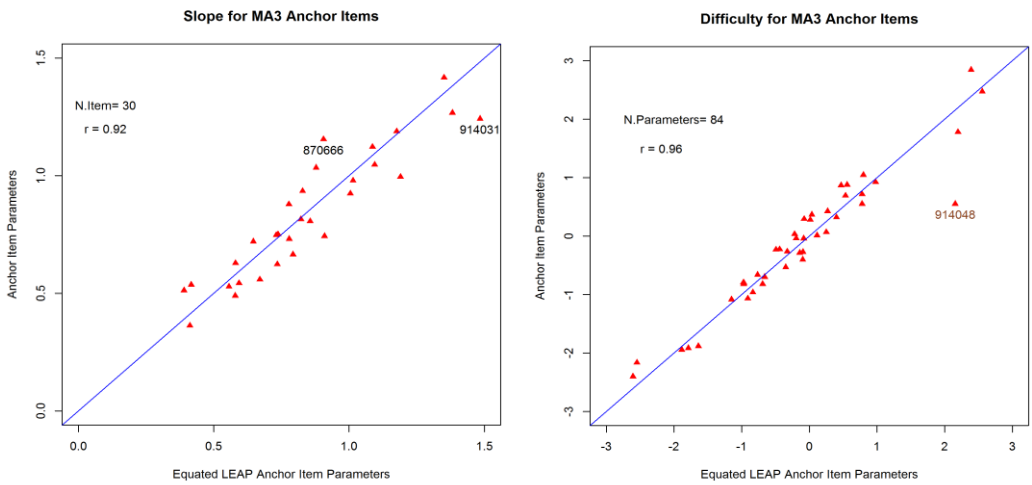
Anchor 3



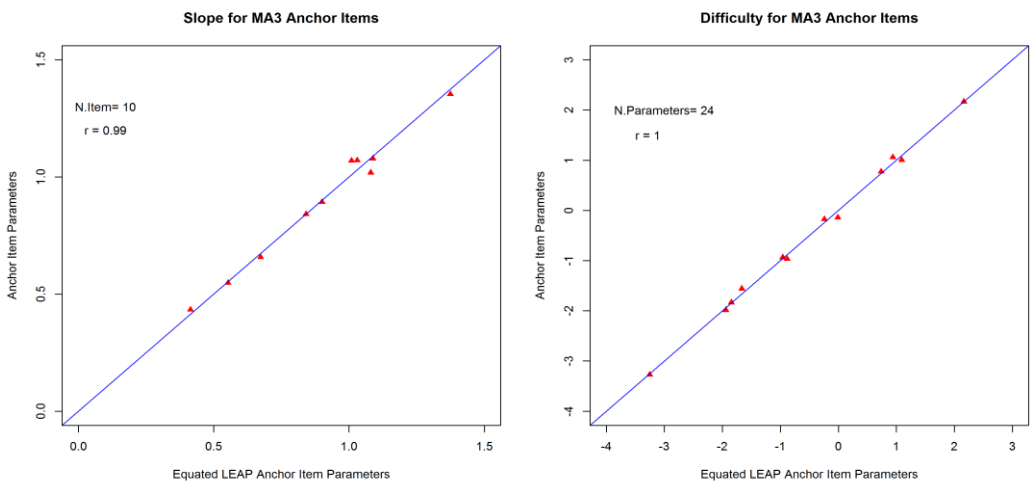
Copyright © 2019 by Louisiana Department of Education.

Figure 6.20 Mathematics Grade 3 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

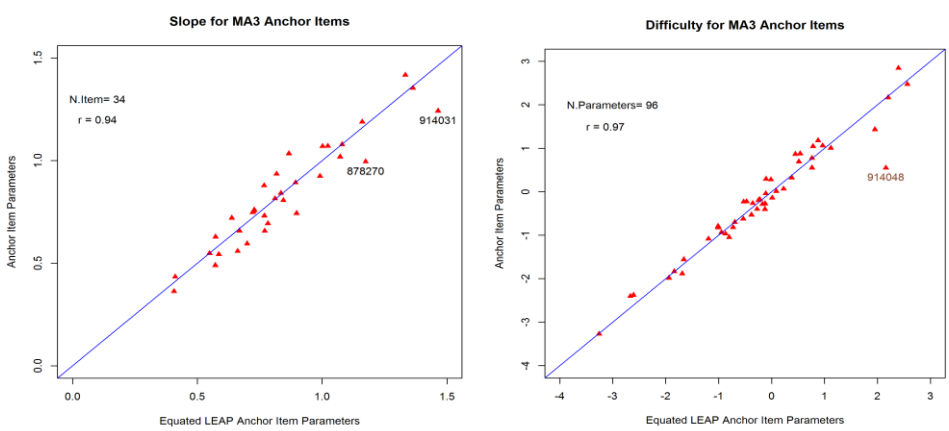
Anchor 1



Anchor 2



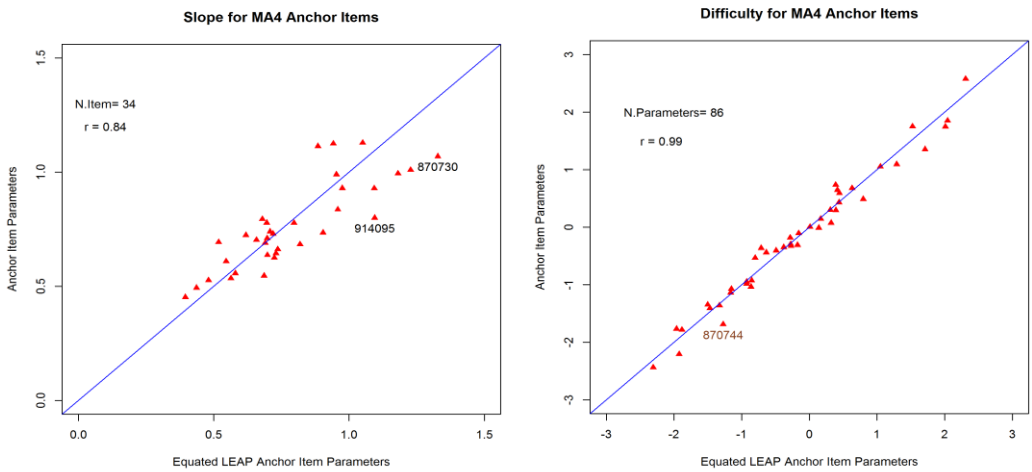
Anchor 3



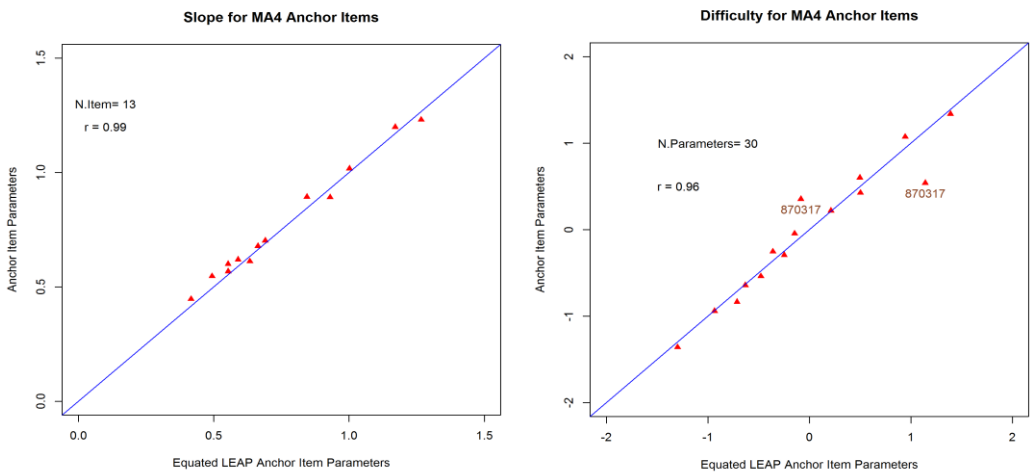
Copyright © 2019 by Louisiana Department of Education.

Figure 6.21 Mathematics Grade 4 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

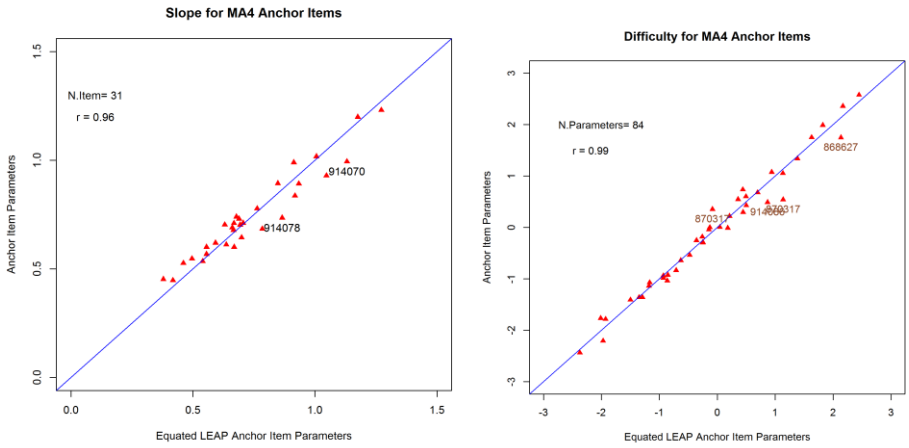
Anchor 1



Anchor 2



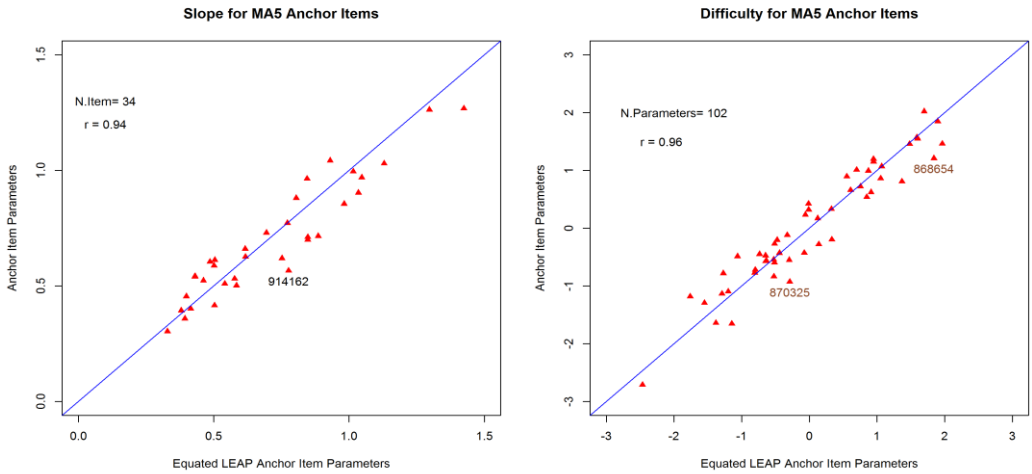
Anchor 3



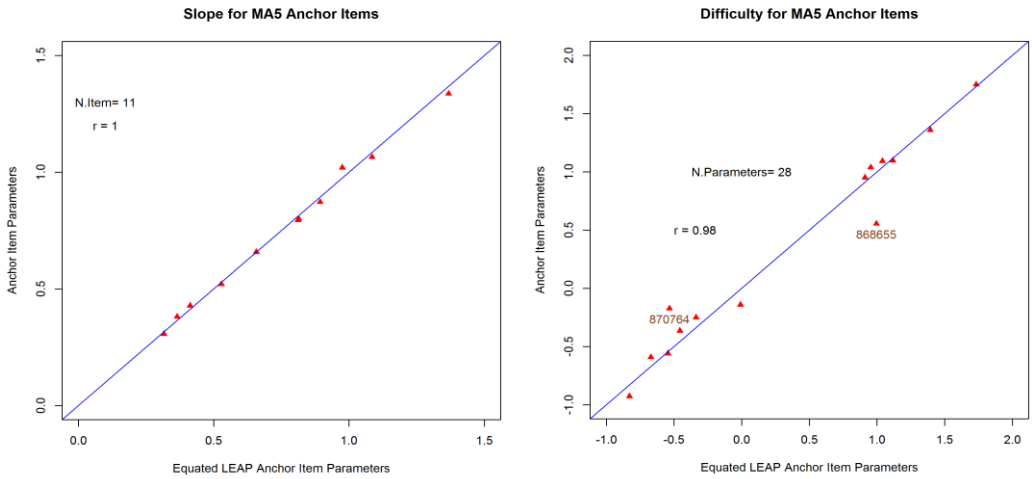
Copyright © 2019 by Louisiana Department of Education.

Figure 6.22 Mathematics Grade 5 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

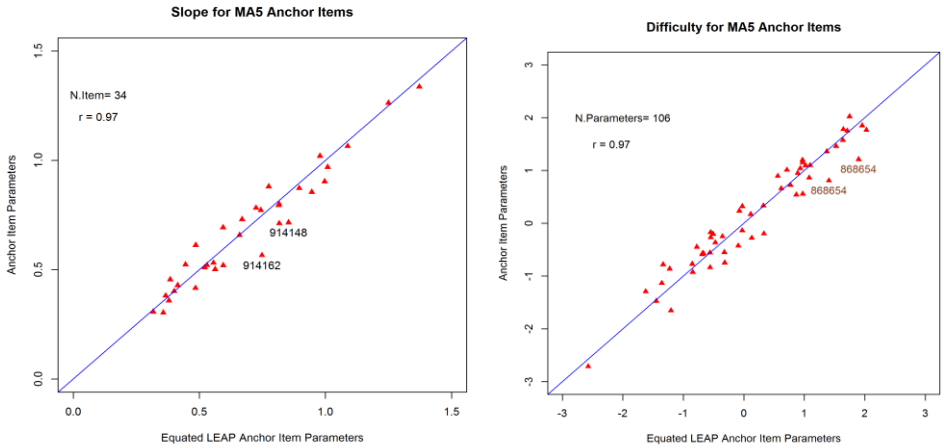
Anchor 1



Anchor 2



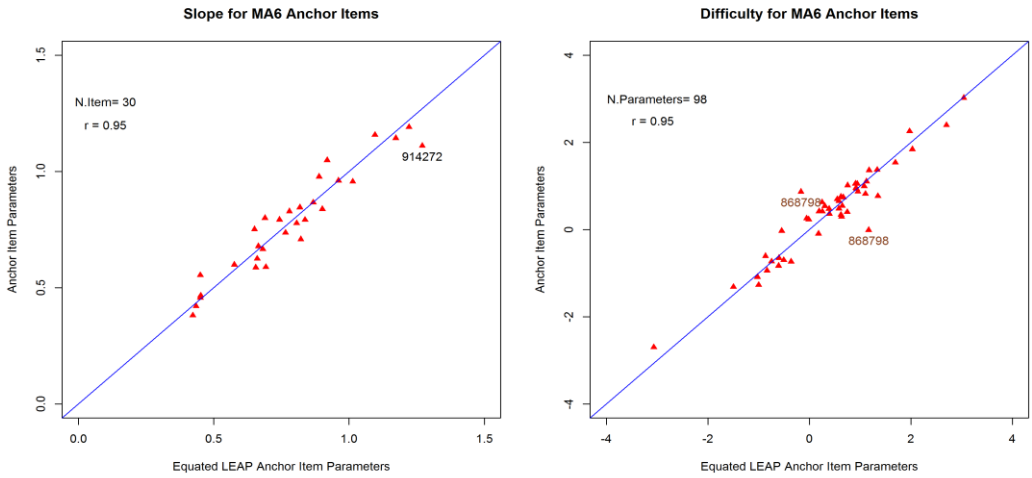
Anchor 3



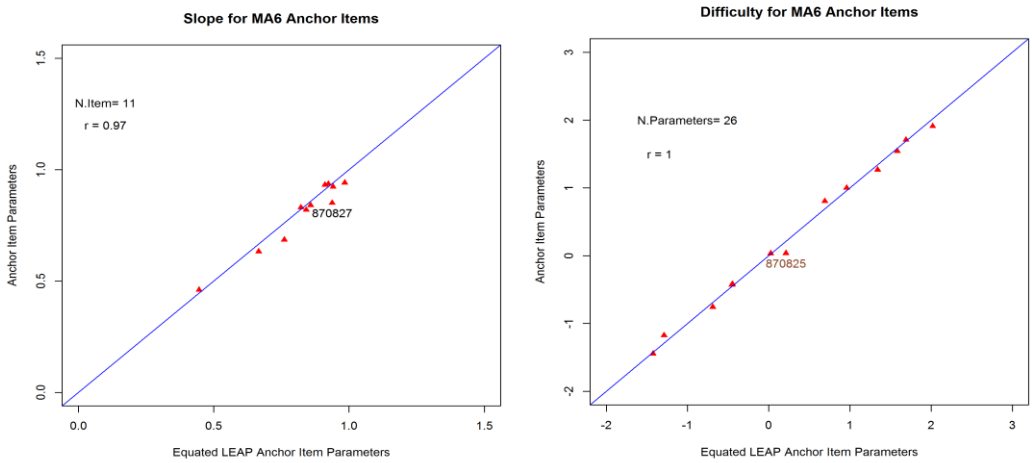
Copyright © 2019 by Louisiana Department of Education.

Figure 6.23 Mathematics Grade 6 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

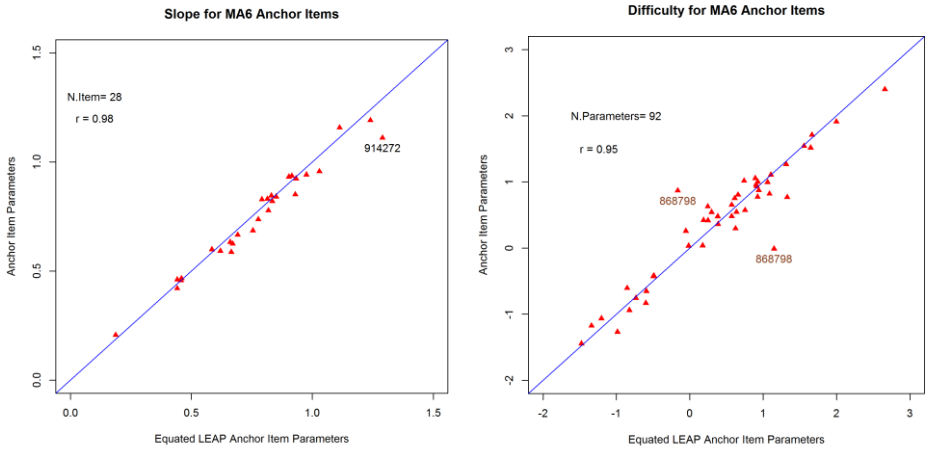
Anchor 1



Anchor 2



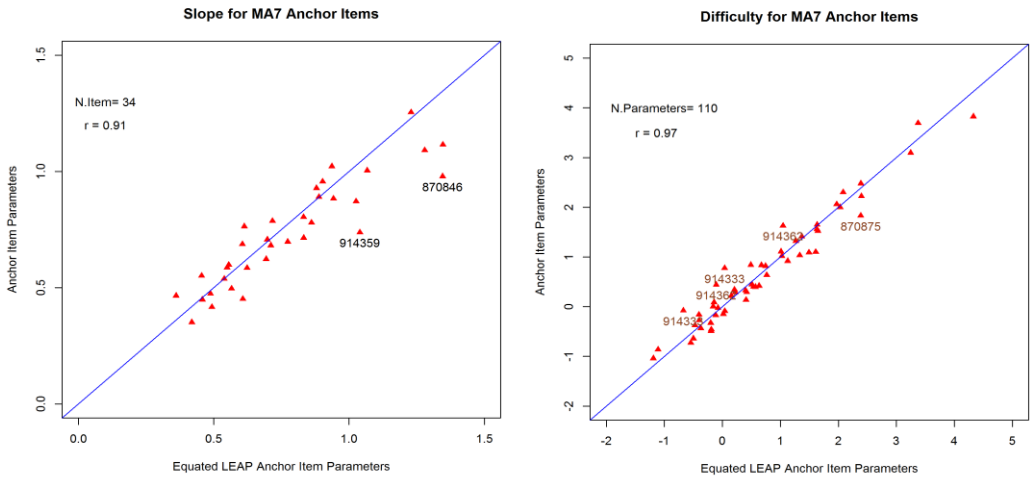
Anchor 3



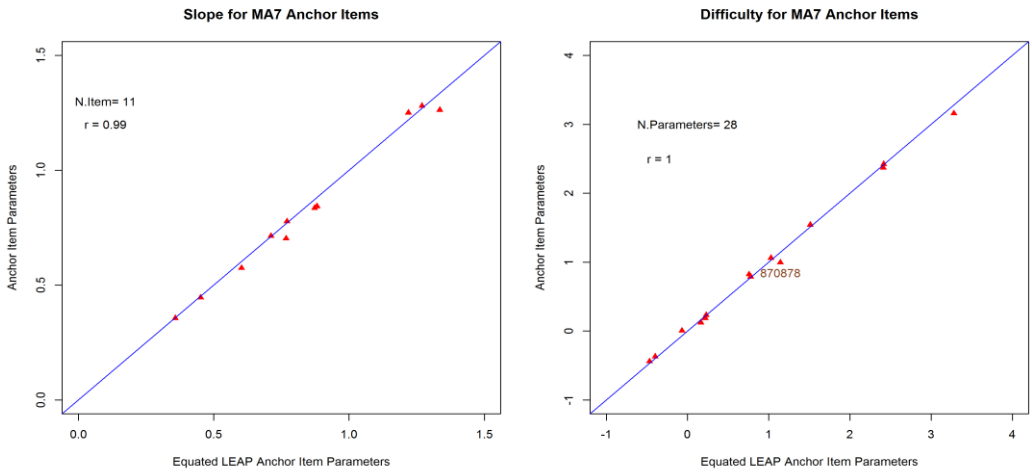
Copyright © 2019 by Louisiana Department of Education.

Figure 6.24 Mathematics Grade 7 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

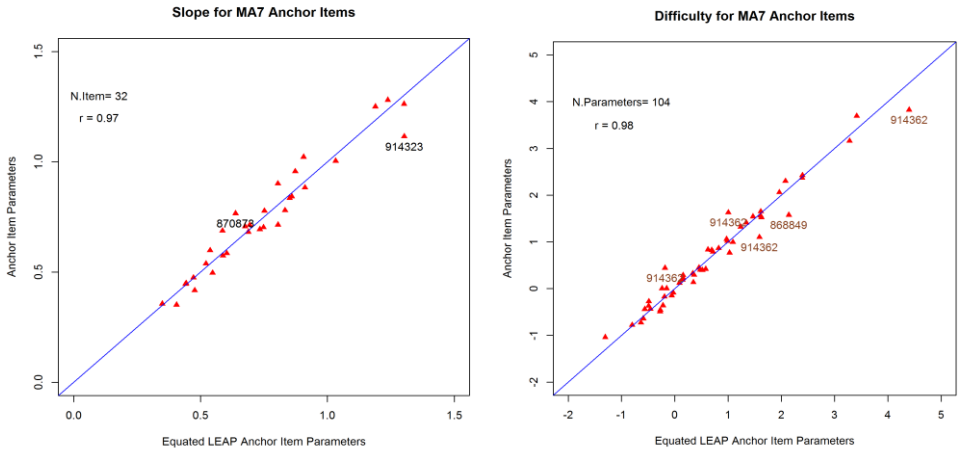
Anchor 1



Anchor 2



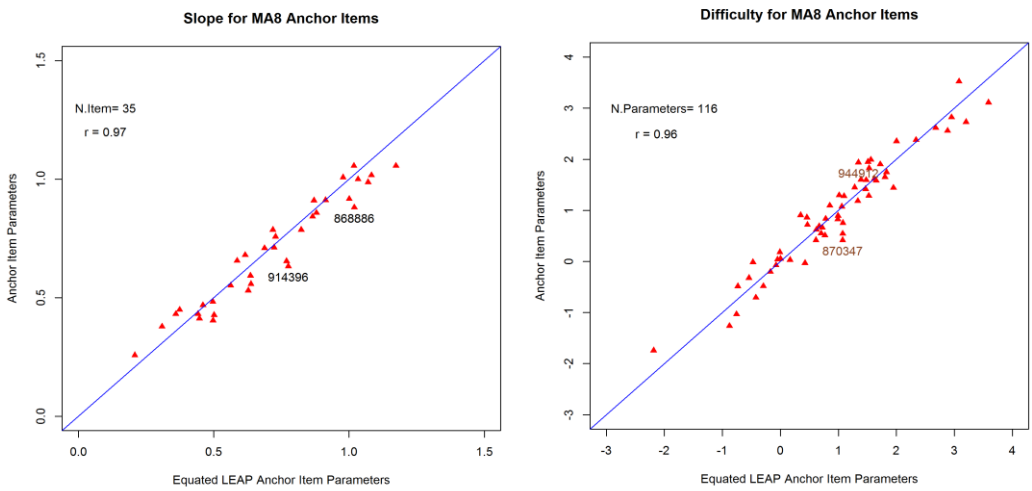
Anchor 3



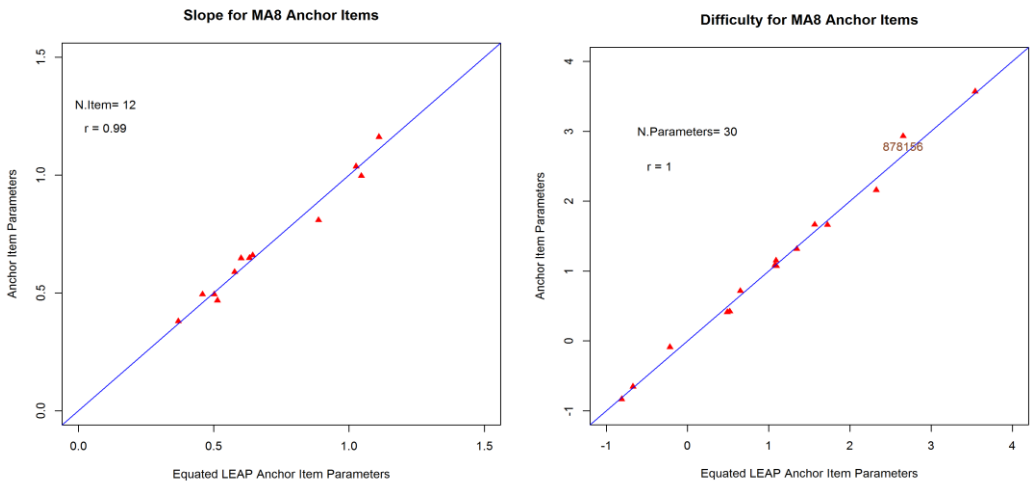
Copyright © 2019 by Louisiana Department of Education.

Figure 6.25 Mathematics Grade 8 Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters

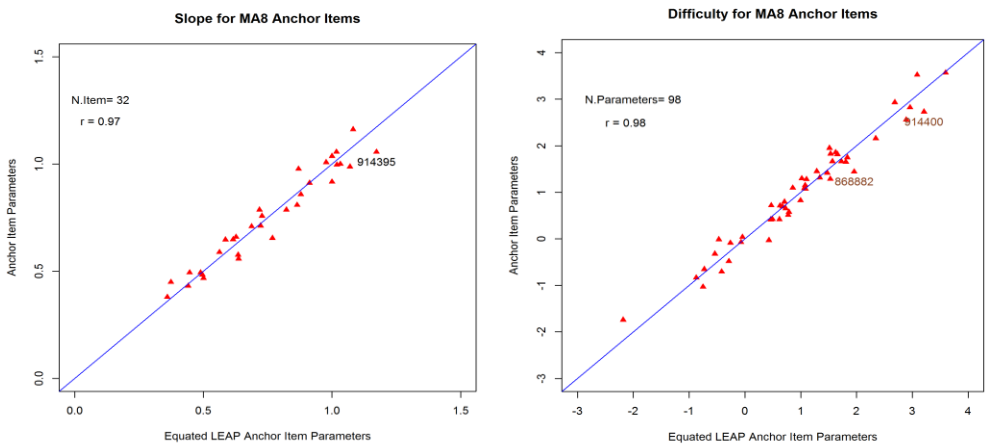
Anchor 1



Anchor 2



Anchor 3



Copyright © 2019 by Louisiana Department of Education.

6.3.2.1. 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 impact the validity of the results. Before an anchor item may be dropped from an anchor set, the item characteristics, adequacy of the content coverage, and impact to the size of the anchor set should be evaluated.

Outliers of anchor items were reviewed with the Robust *Z* (Huynh & Meyer, 2010) and the weighted root mean square difference (WRMSD) method in addition to 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. 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 Robust Z method and remove flagged items from anchor set using the criterion value of |1.96|
- Run STUIRT using the remaining items after Robust Z is applied, 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.

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

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

Copyright © 2019 by Louisiana Department of Education.

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

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

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

where *Q* represents a quadrature point (i.e., node), *W* represents its weight given quadrature point *Q* from PARSCALE PH2 output, *INTACT* represents intact item parameters, and *EST* represents estimated item parameters corresponding to intact item parameters. Table 6.28 summarizes WRMSD flagging criteria for inspection and possible removal of linking items.

Table 6.28 PARCC WRMSD Flagging Criteria

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

6.3.2.2. Lowest and Highest Obtainable Scale Scores

A maximum likelihood (MML) procedure cannot produce scale score estimates for students with perfect scores or scores below the level expected 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 2018 LEAP 2025 used the same LOSS and HOSS values. The LOSS value was 650, and the HOSS value was 850.

6.3.2.3. Claim-Level,Subclaim-Level, and Subcategory 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 ratings 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).

Copyright © 2019 by Louisiana Department of Education.

Mathematics has four subclaims and four subcategories. Subclaim and Subcategory 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, subclaim, or subcategory 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 *Approaching Basic* achievement levels, a *Moderate* rating requires similar knowledge and ability as the *Basic* achievement level, and a *Strong* rating requires similar knowledge and ability as the *Mastery* or *Advanced* achievement levels.

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

The 2018 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

Subcategory

1. Mathematics Subcategory A.1
2. Mathematics Subcategory A.2
3. Mathematics Subcategory A.3
4. Mathematics Subcategory A.4

In mathematics, the subclaim 1, Major Content, is further reported at either three to four subcategories, depending on the grade. Grades 4, 6, and 7 have three subcategories and grades 3, 5, and 8 had four subcategories.

Copyright © 2019 by Louisiana Department of Education.

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 6.28 for cut scores for summatives, claims, and subclaims).

For subclaims in both ELA and mathematics, 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 Θ_{L3} and Θ_{L4} are cut scores for the subclaim. This is also illustrated in Table 6.29.

For Subcategory, mathematics subclaim A is split as 4 subcategories, A.1, A.2, A.3, and A.4. As like subclaim, only performance-level ratings were reported. The same scoring procedure and the cuts for subclaims are applied to generate a raw-score-to-theta scoring tables.

Table 6.29 Cut Scores for Summative, Claim, and Subclaim

Performance Level	Summative Test	Claim		Subclaim/Subcategory
		Reading	Writing	
1				
2	700	30	25	
3	725	40	30	Θ_{L3}
4	750	50	35	Θ_{L4}
5	Around 800			

**Subclaim thetas are those from summative tests (i.e., 725 & 750).*
***Yellow highlight shows cut scores for claim and subclaim.*

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.

Copyright © 2019 by Louisiana Department of Education.

In the spring of 2018, 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 6.30 through Table 6.41 provide scale scores at selected percentiles that can be used to compare the distributional characteristics of the Spring 2018 forms to previous administrations. Although these scale scores are rounded values, there were differences in the scale-score values for a given percentile across the forms. These variations could arise for several reasons: (1) differences in 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 6.30 Comparisons of Scale Scores at Selected Percentiles—Grade 3 ELA

	2016	2017	2018
Percentile	Form A	Form B	Form C
99	822	839	842
95	796	810	810
90	783	793	797
85	774	784	788
80	768	775	779
75	762	770	773
70	757	762	768
65	751	757	762
60	746	752	757
55	741	748	752
50	738	743	746
45	732	739	741
40	727	734	736
35	721	727	730
30	715	723	724
25	712	718	715
20	706	710	708
15	695	701	701
10	687	695	692
5	676	679	676
1	654	655	650

Copyright © 2019 by Louisiana Department of Education.

Table 6.31 Comparisons of Scale Scores at Selected Percentiles—Grade 4 ELA

	2016	2017	2018
Percentile	Form A	Form B	Form C
99	816	818	821
95	794	796	800
90	785	785	789
85	777	777	778
80	769	771	774
75	765	765	767
70	760	761	763
65	755	756	757
60	751	752	753
55	746	748	749
50	744	744	744
45	740	741	740
40	735	737	736
35	731	733	731
30	727	728	727
25	722	724	721
20	715	717	714
15	709	711	707
10	701	702	698
5	691	691	687
1	666	670	668

Copyright © 2019 by Louisiana Department of Education.

Table 6.32 Comparisons of Scale Scores at Selected Percentiles—Grade 5 ELA

	2016	2017	2018
Percentile	Form A	Form B	Form C
99	816	813	817
95	792	793	795
90	782	782	782
85	774	775	777
80	767	769	769
75	763	763	765
70	758	758	760
65	754	754	756
60	749	750	753
55	745	747	749
50	740	743	746
45	738	739	740
40	733	735	736
35	728	731	732
30	723	727	728
25	720	721	724
20	714	716	716
15	708	709	711
10	701	701	702
5	692	691	691
1	675	673	676

Copyright © 2019 by Louisiana Department of Education.

Table 6.33 Comparisons of Scale Scores at Selected Percentiles—Grade 6 ELA

	2016	2017	2018
Percentile	Form A	Form B	Form C
99	813	814	808
95	792	790	789
90	780	779	777
85	772	770	770
80	765	763	763
75	760	759	758
70	756	754	753
65	752	748	749
60	748	745	746
55	745	741	742
50	741	736	737
45	737	733	735
40	734	729	730
35	730	724	726
30	727	721	721
25	723	716	718
20	718	711	713
15	713	705	707
10	706	698	700
5	696	689	691
1	676	671	675

Copyright © 2019 by Louisiana Department of Education.

Table 6.34 Comparisons of Scale Scores at Selected Percentiles—Grade 7 ELA

	2016	2017	2018
Percentile	Form A	Form B	Form C
99	825	826	831
95	800	800	801
90	787	786	789
85	777	778	780
80	771	770	774
75	766	765	767
70	761	759	762
65	756	756	757
60	751	751	752
55	747	745	749
50	742	742	744
45	740	737	740
40	735	733	735
35	730	728	730
30	726	723	726
25	721	717	719
20	714	711	713
15	706	702	707
10	697	692	697
5	683	675	685
1	655	654	662

Copyright © 2019 by Louisiana Department of Education.

Table 6.35 Comparisons of Scale Scores at Selected Percentiles—Grade 8 ELA

	2016	2017	2018
Percentile	Form A	Form B	Form C
99	825	834	824
95	804	806	801
90	790	791	789
85	781	782	781
80	775	776	774
75	770	770	768
70	764	764	764
65	759	758	758
60	754	754	754
55	752	749	751
50	747	745	745
45	743	740	741
40	739	734	737
35	735	731	732
30	731	725	726
25	727	719	722
20	721	714	716
15	714	707	708
10	706	696	699
5	693	681	683
1	670	651	657

Copyright © 2019 by Louisiana Department of Education.

Table 6.36 Comparisons of Scale Scores at Selected Percentiles—Grade 3 Mathematics

	2016	2017	2018
Percentile	Form A	Form B	Form C
99	824	822	817
95	802	796	793
90	789	786	783
85	781	776	775
80	775	772	771
75	770	765	764
70	765	761	759
65	760	756	755
60	756	752	750
55	751	747	746
50	746	743	742
45	741	738	740
40	738	733	735
35	733	728	731
30	728	725	726
25	722	720	719
20	716	715	713
15	710	706	708
10	703	699	698
5	692	689	686
1	672	667	664

Copyright © 2019 by Louisiana Department of Education.

Table 6.37 Comparisons of Scale Scores at Selected Percentiles—Grade 4 Mathematics

	2016	2017	2018
Percentile	**Form A**	**Form B**	**Form C**
99	819	812	812
95	797	792	790
90	786	779	780
85	777	774	772
80	771	767	768
75	766	762	762
70	761	756	757
65	756	752	753
60	752	748	749
55	747	744	744
50	743	740	740
45	738	736	735
40	732	732	733
35	728	727	728
30	723	722	723
25	718	717	718
20	713	712	715
15	708	706	710
10	703	700	700
5	693	693	689
1	677	674	670

Copyright © 2019 by Louisiana Department of Education.

Table 6.38 Comparisons of Scale Scores at Selected Percentiles—Grade 5 Mathematics

	2016	2017	2018
Percentile	Form A	Form B	Form C
99	819	808	810
95	792	784	784
90	779	774	774
85	771	767	765
80	766	760	759
75	759	755	755
70	754	751	749
65	749	747	745
60	745	742	743
55	740	740	738
50	735	735	734
45	731	730	729
40	728	728	727
35	722	723	722
30	720	720	720
25	714	715	714
20	711	709	711
15	705	706	705
10	699	699	698
5	691	691	689
1	678	675	672

Copyright © 2019 by Louisiana Department of Education.

Table 6.39 Comparisons of Scale Scores at Selected Percentiles—Grade 6 Mathematics

	2016	2017	2018
Percentile	Form A	Form B	Form C
99	803	808	800
95	783	781	780
90	771	771	770
85	765	762	762
80	758	757	757
75	753	752	752
70	747	746	748
65	744	742	743
60	740	738	739
55	735	734	736
50	731	732	732
45	729	727	728
40	724	724	723
35	722	719	721
30	717	717	716
25	714	711	713
20	709	708	707
15	706	701	704
10	699	697	696
5	692	688	686
1	679	671	672

Copyright © 2019 by Louisiana Department of Education.

Table 6.40 Comparisons of Scale Scores at Selected Percentiles—Grade 7 Mathematics

	2016	2017	2018
Percentile	Form A	Form B	Form C
99	797	796	797
95	779	777	777
90	768	766	766
85	760	760	759
80	754	754	755
75	750	749	750
70	746	746	745
65	742	741	742
60	738	737	739
55	734	734	735
50	730	731	731
45	728	727	729
40	723	723	725
35	721	721	721
30	719	717	718
25	714	712	713
20	712	709	710
15	706	706	706
10	703	699	702
5	695	694	693
1	678	673	679

Copyright © 2019 by Louisiana Department of Education.

Table 6.41 Comparisons of Scale Scores at Selected Percentiles—Grade 8 Mathematics

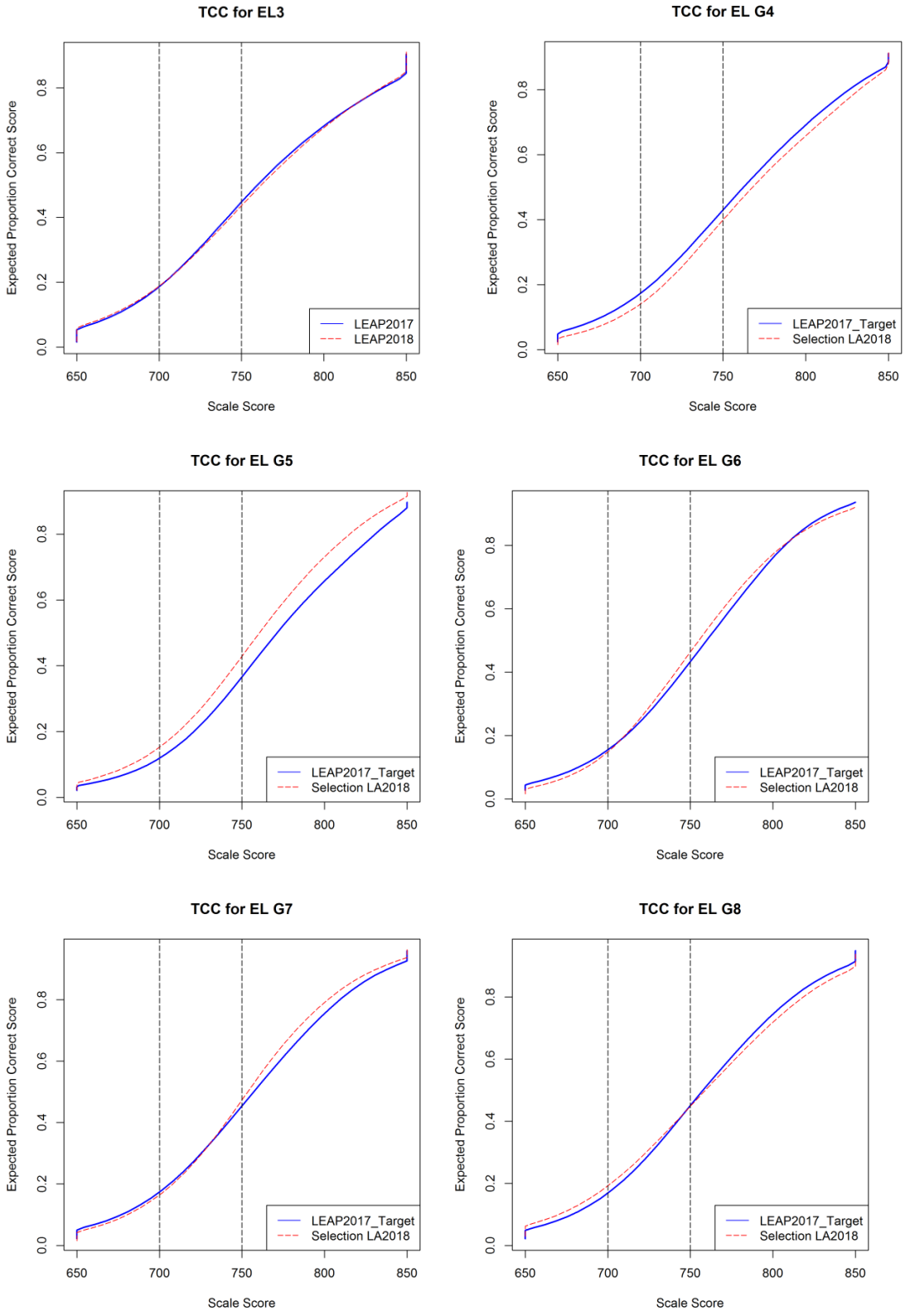
	2016	2017	2018
Percentile	Form A	Form B	Form C
99	808	809	807
95	787	784	784
90	775	771	773
85	766	763	764
80	761	757	757
75	753	751	752
70	749	746	746
65	744	741	742
60	737	736	737
55	734	730	732
50	731	727	727
45	727	724	721
40	724	718	718
35	720	714	715
30	712	710	707
25	708	706	702
20	704	698	697
15	699	693	691
10	695	687	684
5	684	674	676
1	663	656	654

Additional evidence of comparability can be found by reviewing the test characteristic curves (TCCs) for the 2017 and 2018 administration of LEAP 2025, see figures 6.26 and 6.27. For most contents and grades, the TCCs for the two years were similar across ability (scale score) ranges. For ELA grade 5, 6, and 7 the 2018 forms were slightly easier that the 2017 forms for high performing students. For mathematics grades 6, 7, and 8 the 2018 forms were slightly more difficult than the 2017 forms for low performing students. Note that this different difficulty is adjusted by reporting different scale scores given raw scores: A scale score of difficult form is higher than that of easy form given the same raw score.

Figures 6.28 and 6.29 show SEMs for the 2017 and 2018 LEAP 2025 assessments. For most contents and grades, the SEMs were similar across ability (scale score) ranges, especially in the middle ability ranges.

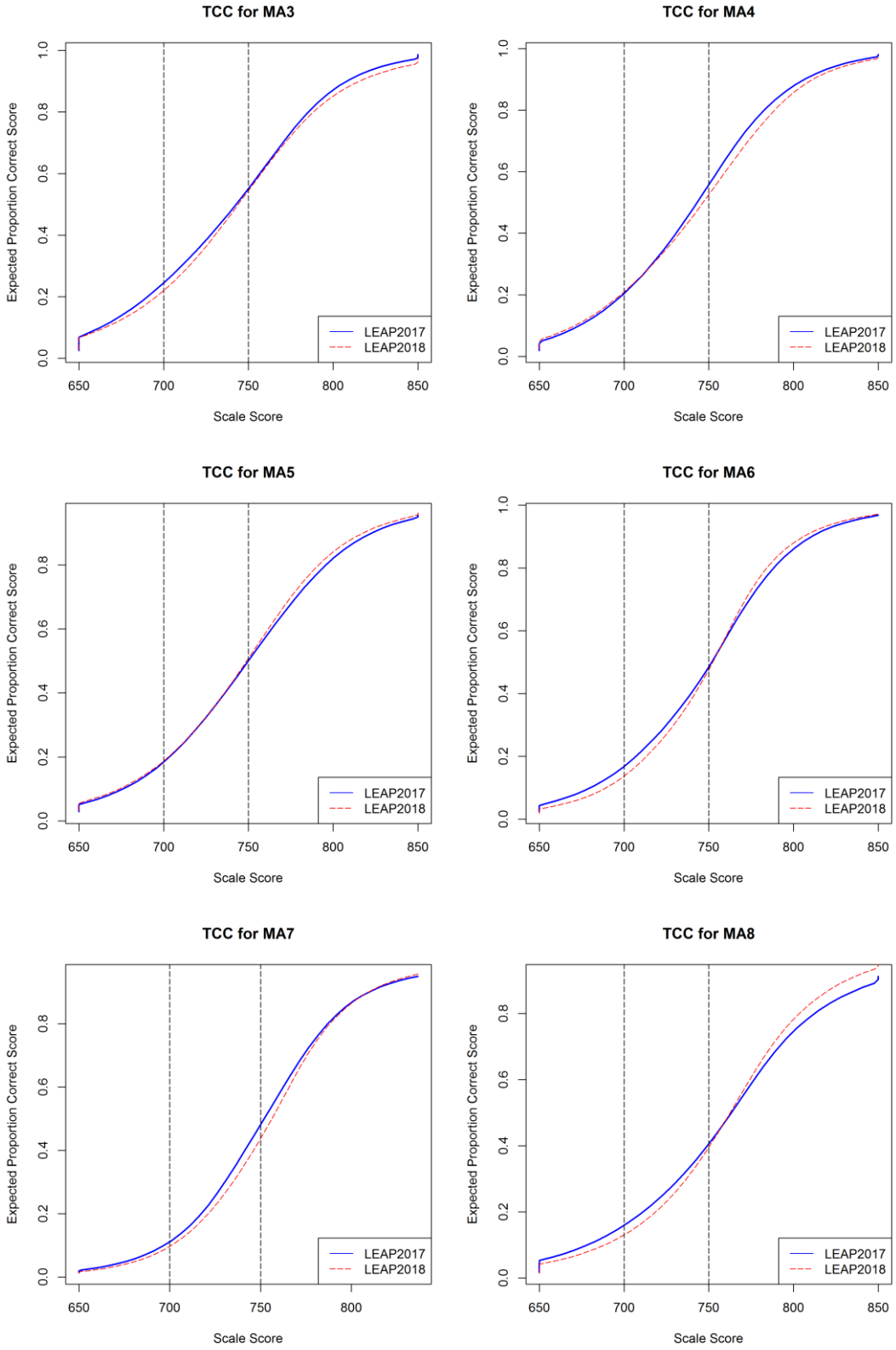
Copyright © 2019 by Louisiana Department of Education.

Figure 6.26 TCCs Across Years: ELA



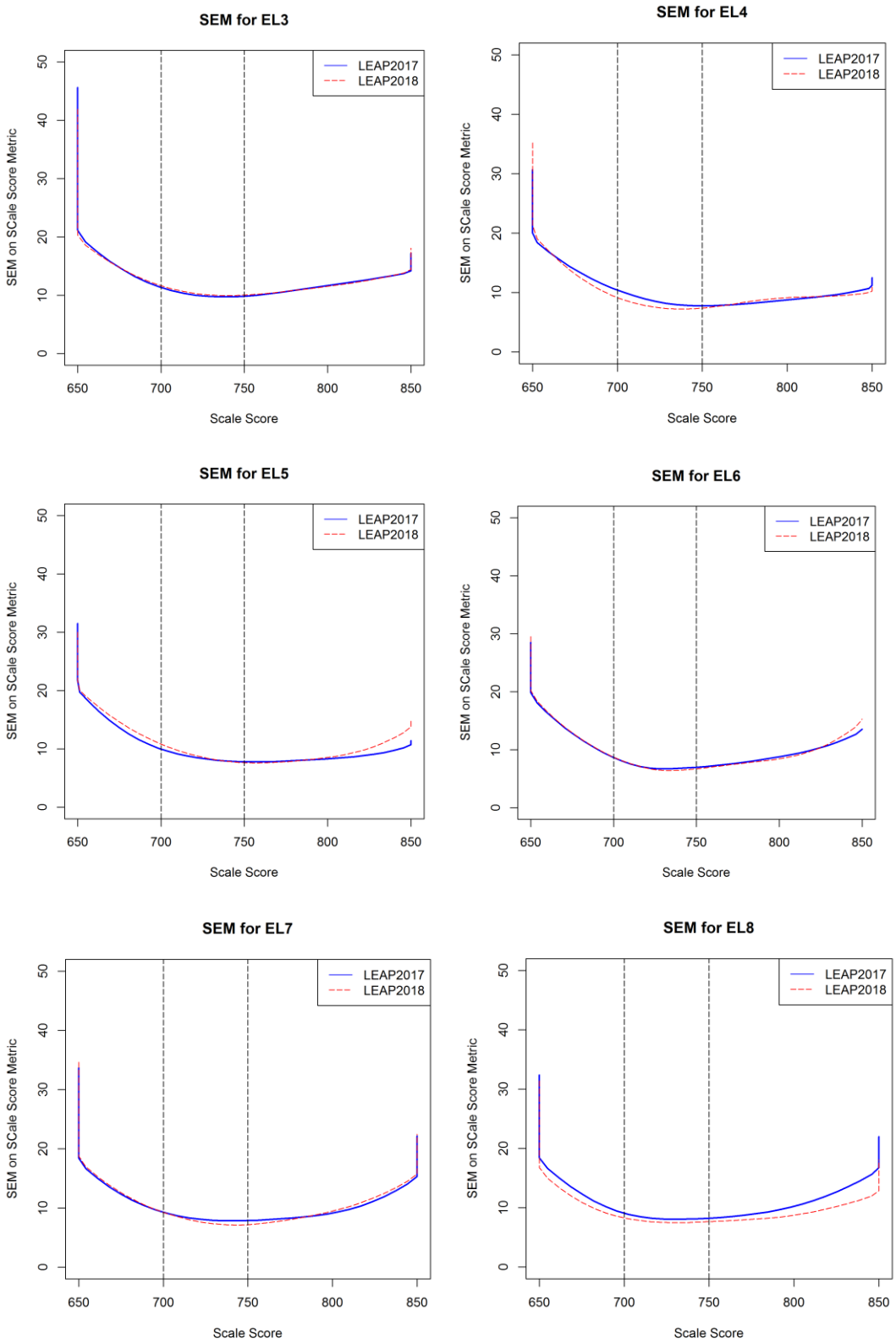
Copyright © 2019 by Louisiana Department of Education.

Figure 6.27 TCCs Across Years: Mathematics



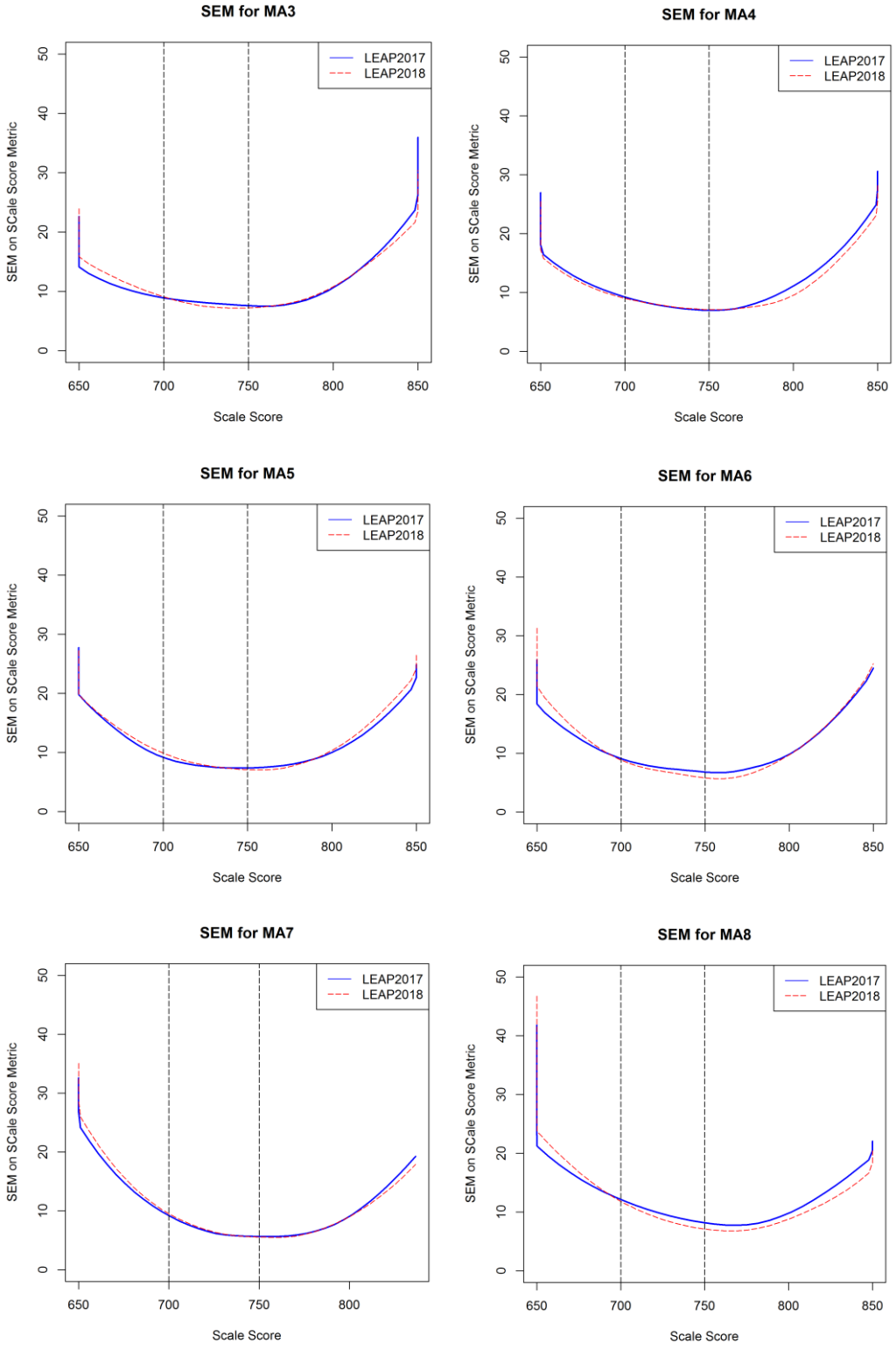
Copyright © 2019 by Louisiana Department of Education.

Figure 6.28 SEM Across Years: ELA



Copyright © 2019 by Louisiana Department of Education.

Figure 6.29 SEM Across Years: Mathematics



Copyright © 2019 by Louisiana Department of Education.

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)

Copyright © 2019 by Louisiana Department of Education.

CHAPTER 7: TEST RESULTS

This chapter of the technical report contains information on the results of the Spring 2018 LEAP 2025 ELA and mathematics assessments. The scale score results and Achievement Level information are presented here. Presenting the results by achievement level translates the quantitative scale provided through scale scores into a qualitative description of student achievement: *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 in this chapter are based on census data. The results presented here may differ slightly from the official state summary report of all student populations due to ongoing resolution of test materials and student information. The results in the tables in this chapter are presented as evidence of reliability and validity of the scores from the LEAP 2025 assessments and should not be used for state accountability purposes.

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

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

Copyright © 2019 by Louisiana Department of Education.

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**	≥55,370	99.85%	≥55,990	99.85%
	Gender				
	Female	≥27,100	99.86%	≥27,360	99.87%
	Male	≥28,250	99.86%	≥28,570	99.88%
	Ethnicity				
	Hispanic/Latino	≥4,180	99.79%	≥4,260	99.79%
	American Indian or Alaska Native	≥340	99.71%	≥350	99.72%
	Asian	≥730	100.00%	≥740	100.00%
	Black or African American	≥24,430	99.83%	≥24,740	99.85%
	Native Hawaiian or Other Pacific	≥40	100.00%	≥40	100.00%
	White	≥23,830	99.91%	≥23,980	99.92%
	Two or More Races	≥1,720	99.88%	≥1,730	99.94%
	Education Classification				
	Regular	≥49,320	99.86%	≥49,870	99.85%
	Special	≥6,050	99.77%	≥6,120	99.85%
	Economic Status				
	Economically Disadvantaged	≥39,870	99.85%	≥40,060	99.87%
	Not Economically Disadvantaged	≥14,630	99.90%	≥14,660	99.90%
	English Learner Status				
	Non-EL	≥52,880	99.85%	≥53,440	99.85%
	EL	≥2,490	99.72%	≥2,550	99.84%
	Migrant Status				
	Nonmigrant	≥55,230	99.85%	≥55,850	99.85%
	Migrant	≥140	99.30%	≥140	100.00%
	Section 504 Status				
	Non-Section 504	≥50,520	99.85%	≥51,070	99.85%
	Section 504	≥55,370	99.85%	≥55,990	99.85%

Copyright © 2019 by Louisiana Department of Education.

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**	≥55,800	99.82%	≥56,340	99.87%
	Gender				
	Female	≥27,370	99.82%	≥27,610	99.88%
	Male	≥28,370	99.83%	≥28,640	99.90%
	Ethnicity				
	Hispanic/Latino	≥4,100	99.85%	≥4,150	99.86%
	American Indian or Alaska Native	≥340	99.41%	≥340	99.42%
	Asian	≥860	99.88%	≥870	99.89%
	Black or African American	≥24,640	99.76%	≥24,930	99.85%
	Native Hawaiian or Other Pacific	≥50	100.00%	≥50	100.00%
	White	≥24,050	99.89%	≥24,180	99.94%
	Two or More Races	≥1,660	99.94%	≥1,670	99.94%
	Education Classification				
	Regular	≥49,630	99.83%	≥50,100	99.87%
	Special	≥6,170	99.68%	≥6,230	99.82%
	Economic Status				
	Economically Disadvantaged	≥39,990	99.80%	≥40,140	99.88%
	Not Economically Disadvantaged	≥15,010	99.91%	≥15,030	99.93%
	English Learner Status				
	Non-EL	≥53,510	99.82%	≥54,010	99.87%
	EL	≥2,280	99.87%	≥2,320	99.87%
	Migrant Status				
	Nonmigrant	≥55,670	99.82%	≥56,220	99.87%
	Migrant	≥120	100.00%	≥120	100.00%
	Section 504 Status				
	Non-Section 504	≥50,080	99.83%	≥50,570	99.85%
	Section 504	≥5,710	99.72%	≥5,770	100.00%

Copyright © 2019 by Louisiana Department of Education.

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**	≥55,280	99.93%	≥55,290	99.92%
	Gender				
	Female	≥27,140	99.94%	≥27,140	99.94%
	Male	≥28,140	99.93%	≥28,140	99.90%
	Ethnicity				
	Hispanic/Latino	≥3,800	99.92%	≥3,800	99.92%
	American Indian or Alaska Native	≥350	100.00%	≥350	100.00%
	Asian	≥800	100.00%	≥800	100.00%
	Black or African American	≥24,060	99.95%	≥24,070	99.92%
	Native Hawaiian or Other Pacific	≥40	100.00%	≥40	100.00%
	White	≥24,590	99.92%	≥24,590	99.92%
	Two or More Races	≥1,590	99.87%	≥1,590	99.87%
	Education Classification				
	Regular	≥49,170	99.94%	≥49,180	99.92%
	Special	≥6,100	99.89%	≥6,100	99.89%
	Economic Status				
	Economically Disadvantaged	≥38,900	99.93%	≥38,910	99.91%
	Not Economically Disadvantaged	≥15,940	99.96%	≥15,940	99.96%
	English Learner Status				
	Non-EL	≥53,590	99.93%	≥53,600	99.92%
	EL	≥1,680	99.94%	≥1,680	99.94%
	Migrant Status				
	Nonmigrant	≥55,190	99.93%	≥55,200	99.92%
	Migrant	≥80	100.00%	≥80	100.00%
	Section 504 Status				
	Non-Section 504	≥50,070	99.94%	≥50,080	99.92%
	Section 504	≥5,200	99.88%	≥5,200	99.88%

Copyright © 2019 by Louisiana Department of Education.

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**	≥52,810	99.86%	≥52,820	99.87%
	Gender				
	Female	≥25,730	99.89%	≥25,730	99.90%
	Male	≥27,080	99.83%	≥27,080	99.85%
	Ethnicity				
	Hispanic/Latino	≥3,430	99.94%	≥3,430	99.94%
	American Indian or Alaska Native	≥330	99.71%	≥330	99.71%
	Asian	≥790	100.00%	≥790	100.00%
	Black or African American	≥23,460	99.85%	≥23,460	99.87%
	Native Hawaiian or Other Pacific	≥40	100.00%	≥40	100.00%
	White	≥23,380	99.86%	≥23,390	99.86%
	Two or More Races	≥1,330	99.93%	≥1,330	99.93%
	Education Classification				
	Regular	≥47,270	99.87%	≥47,280	99.89%
	Special	≥5,530	99.75%	≥5,540	99.75%
	Economic Status				
	Economically Disadvantaged	≥36,940	99.84%	≥36,950	99.85%
	Not Economically Disadvantaged	≥15,470	99.93%	≥15,470	99.93%
	English Learner Status				
	Non-EL	≥51,610	99.86%	≥51,610	99.87%
	EL	≥1,200	99.92%	≥1,200	99.92%
	Migrant Status				
	Nonmigrant	≥52,740	99.86%	≥52,750	99.87%
	Migrant	≥70	100.00%	≥70	100.00%
	Section 504 Status				
	Non-Section 504	≥47,600	99.86%	≥47,610	99.87%
	Section 504	≥5,210	99.87%	≥5,210	99.87%

Copyright © 2019 by Louisiana Department of Education.

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**	≥51,590	99.76%	≥51,590	99.77%
	Gender				
	Female	≥25,040	99.80%	≥25,040	99.80%
	Male	≥26,550	99.72%	≥26,550	99.74%
	Ethnicity				
	Hispanic/Latino	≥3,210	99.84%	≥3,210	99.84%
	American Indian or Alaska Native	≥370	99.73%	≥370	99.73%
	Asian	≥780	100.00%	≥780	100.00%
	Black or African American	≥22,710	99.70%	≥22,710	99.72%
	Native Hawaiian or Other Pacific	≥20	100.00%	≥20	100.00%
	White	≥23,340	99.80%	≥23,350	99.80%
	Two or More Races	≥1,120	99.64%	≥1,120	99.73%
	Education Classification				
	Regular	≥46,410	99.78%	≥46,420	99.79%
	Special	≥5,170	99.54%	≥5,170	99.59%
	Economic Status				
	Economically Disadvantaged	≥35,390	99.72%	≥35,390	99.74%
	Not Economically Disadvantaged	≥15,840	99.86%	≥15,840	99.85%
	English Learner Status				
	Non-EL	≥50,460	99.75%	≥50,470	99.77%
	EL	≥1,120	99.91%	≥1,120	99.91%
	Migrant Status				
	Nonmigrant	≥51,510	99.76%	≥51,520	99.77%
	Migrant	≥70	100.00%	≥70	100.00%
	Section 504 Status				
	Non-Section 504	≥46,420	99.75%	≥46,430	99.76%
	Section 504	≥5,160	99.85%	≥5,160	99.88%

Copyright © 2019 by Louisiana Department of Education.

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**	≥51,090	99.71%	≥51,100	99.72%
	Gender				
	Female	≥25,180	99.82%	≥25,190	99.82%
	Male	≥25,900	99.61%	≥25,900	99.63%
	Ethnicity				
	Hispanic/Latino	≥3,030	99.77%	≥3,030	99.77%
	American Indian or Alaska Native	≥380	99.74%	≥380	99.74%
	Asian	≥860	100.00%	≥860	100.00%
	Black or African American	≥22,540	99.62%	≥22,550	99.64%
	Native Hawaiian or Other Pacific	≥30	100.00%	≥30	100.00%
	White	≥23,200	99.78%	≥23,200	99.78%
	Two or More Races	≥1,000	99.70%	≥1,000	99.70%
	Education Classification				
	Regular	≥46,240	99.71%	≥46,250	99.73%
	Special	≥4,840	99.69%	≥4,850	99.69%
	Economic Status				
	Economically Disadvantaged	≥34,660	99.63%	≥34,670	99.65%
	Not Economically Disadvantaged	≥15,960	99.91%	≥15,960	99.91%
	English Learner Status				
	Non-EL	≥50,010	99.71%	≥50,020	99.72%
	EL	≥1,070	99.63%	≥1,070	99.63%
	Migrant Status				
	Nonmigrant	≥51,020	99.71%	≥51,030	99.72%
	Migrant	≥60	100.00%	≥60	100.00%
	Section 504 Status				
	Non-Section 504	≥46,090	99.70%	≥46,100	99.72%
	Section 504	≥4,990	99.78%	≥4,990	99.74%

*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 7.2 through 7.13 show the percentage of students in each achievement level based on the state population for the 2018 administration of the ELA and mathematics assessments. Previous year results are presented as well for comparison purposes.

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥56,800	13.4	17.8	24.7	38.9	5.1
2018	≥55,390	14.2	18.2	22.3	39.8	5.6

Copyright © 2019 by Louisiana Department of Education.

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥56,230	8.8	18.3	29.3	36.2	7.3
2018	≥55,760	10.8	17.0	28.7	34.8	8.8

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥53,300	8.7	18.8	31.1	37.9	3.4
2018	≥55,310	8.8	17.7	30.4	39.3	3.7

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥52,370	10.4	24.9	29.8	29.4	5.5
2018	≥52,810	9.3	24.6	31.5	30.3	4.4

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥51,930	13.2	19.2	26.5	30.3	10.8
2018	≥51,540	10.7	19.2	26.8	31.4	11.9

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥50,450	11.4	17.4	27.0	35.1	9.0
2018	≥51,020	10.8	17.4	26.6	36.9	8.4

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥56,800	11.1	18.4	27.1	36.2	7.1
2018	≥55,360	10.3	19.7	28.1	34.6	7.3

Copyright © 2019 by Louisiana Department of Education.

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥56,230	8.2	23.2	29.7	35.0	3.8
2018	≥55,680	8.6	22.8	30.3	34.4	3.9

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥53,310	11.1	24.9	32.4	27.7	3.9
2018	≥55,200	10.2	25.8	34.0	25.7	4.2

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥52,350	12.6	30.8	29.2	23.7	3.7
2018	≥52,670	11.6	29.0	32.0	24.8	2.6

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥51,800	11.2	28.9	35.2	22.6	2.1
2018	≥51,420	9.9	29.0	35.7	22.9	2.4

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

Year	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
2017	≥44,710	20.3	28.2	25.0	24.7	1.8
2018	≥44,910	20.9	27.4	23.7	26.1	1.9

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

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

Copyright © 2019 by Louisiana Department of Education.

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. For more information about the test, parents are provided A Parent Guide to the LEAP 2025 Student Reports. In the 2018 administration year, student reports for each school were posted by grade, then downloaded and printed from eDIRECT by the school system and school. eDIRECT is DRC's secure online system that provides schools and districts access to manage student tests and reports.

7.1.1 Description of Each Type of Report

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

In compliance with AERA, APA, & NCME (2014) Standard 12.18, the LEAP 2025 score reports provide clear information about 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, subclaim, and subcategory performance ratings are also reported for students. At the school level, the percentage of students at each achievement level and rating by claim,

Copyright © 2019 by Louisiana Department of Education.

subclaim, and subcategory 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, subclaim, and subcategory performance indicators are reported. Achievement-level descriptors and the percentage of students in each achievement level by school, school system, and the state, which allows comparisons of the student's overall achievement level to those of 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.

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.

The *LEAP Interpretive Guide* (see https://www.louisianabelieves.com/docs/default-source/assessment/leap-2025-gr-3-8-spring-2018-interpretive-guide.pdf?sfvrsn=32c79e1f_4) 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 Interpretive Guide* was developed collaboratively by DRC and LDOE staff. LDOE staff had opportunities to review the guide, provide feedback, and give final approval.

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

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

> ***Standard 5.1*** Test users should be provided with clear explanations of the characteristics, meaning, and intended interpretation of scale scores, as well as their limitations. (102)

Copyright © 2019 by Louisiana Department of Education.

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)

Copyright © 2019 by Louisiana Department of Education.

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 assessments measuring different content and constructs than previous tests were administered to Louisiana students. The new tests were built using the PARCC item bank and were fully aligned to the Louisiana Student Standards. The new tests were reported on 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 test scores, it is essential to understand that descriptors and cut scores are established in a collaborative and participatory process. The descriptors clearly establish, in plain language, the proper frame of reference for understanding how to interpret test scores, particularly cut scores.

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

According to the *Performance Level Setting Technical Report* (Pearson, 2015), PARCC used the evidence-based standard setting (EBSS) method (Beimers, Way, McClarty, & Miles, 2012) for the PARCC performance-level setting (PLS) process. The EBSS method is used to combine various considerations into the process for setting performance levels, including policy considerations, content standards, 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.

Copyright © 2019 by Louisiana Department of Education.

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

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

Copyright © 2019 by Louisiana Department of Education.

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

Table 8.3 Achievement-Level Scale-Score Ranges

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

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

Copyright © 2019 by Louisiana Department of Education.

CHAPTER 9: EVIDENCE OF VALIDITY

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

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

Source of Validity	Related Information	Related Chapter/Source
Evidence Based on Test Content	Item Development Process	Chapter 3 2018-2019 LEAP 3-8 ELA and Mathematics Assessment Frameworks
	Test Blueprint and Item Alignment to Curriculum and Standards	Chapter 3 2018-2019 LEAP 3-8 ELA and Mathematics Assessment Frameworks
	Item Bias, Sensitivity, and Content Appropriateness	Chapter 3
	Accommodations	Chapters 3 and 4
Evidence Based on Response Processes	Data Review	2018-2019 LEAP 3-8 ELA and Mathematics Assessment Frameworks
	Classical Item analysis	Chapter 6
Evidence Based on Internal Structure	Differential Item Functioning	Chapter 10
	Reliability and Standard Errors of Measurement	Chapter 9
Evidence Based on Relationships to Other Variables	Divergent Validity	Chapter 9
	Regression of LEAP 2025 from 2017 to 2018	Chapter 9
Evidence Based on the Consequences of Testing	Scale Score and Performance Level Information	Chapter 7
	Test Interpretive Guide	Chapter 4

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.

Copyright © 2019 by Louisiana Department of Education.

9.1 Construct-Irrelevant Variance and Construct Underrepresentation

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

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

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

9.2 Reliability

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

The 2014 *Standards* indicates:

> The 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.

Copyright © 2019 by Louisiana Department of Education.

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 Sections 9.4.2 and 9.4.3. The SEM of the total score is discussed in Section 9.2.2.

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

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

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

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

9.2.1 **Test Reliability**

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

Copyright © 2019 by Louisiana Department of Education.

Cronbach's coefficient alpha was computed using the formula

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

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

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

The reliability coefficients for the LEAP 2025 are reported in Table 9.2. These reliability coefficients were computed using the census data. The reliability statistics ranged from 0.87 to 0.91 for all ELA forms. The ELA forms have have one writing component (RI or RL) that is the same score of another component (WE); the item score for the RI/RL component was excluded from the reliablility computation. For mathematics, the reliabilities ranged from 0.90 to 0.92. These results indicate acceptable reliability coefficients for the LEAP 2025 tests.

Table 9.2 Reliability in English Language Arts and Mathematics

Content	Grade	Mode	Number of Items	Number of Score Points	SEM	Cronbach's Alpha	N-Count
ELA	3	CBT	26	71	4.36	0.87	≥1,070
ELA	3	PBT	26	71	4.65	0.87	≥54,370
ELA	4	CBT	28	86	5.01	0.89	≥5,210
ELA	4	PBT	28	86	5.27	0.88	≥50,640
ELA	5	CBT	30	83	5.13	0.88	≥55,530
ELA	6	CBT	32	90	5.14	0.91	≥53,040
ELA	7	CBT	32	90	5.58	0.91	≥51,800
ELA	8	CBT	32	94	5.69	0.89	≥51,370
Mathematics	3	CBT	43	62	3.55	0.91	≥1,080
Mathematics	3	PBT	43	62	3.63	0.92	≥54,340
Mathematics	4	CBT	43	62	3.46	0.92	≥5,200
Mathematics	4	PBT	43	62	3.59	0.92	≥50,580
Mathematics	5	CBT	43	62	3.66	0.90	≥55,410
Mathematics	6	CBT	43	66	3.83	0.92	≥52,910
Mathematics	7	CBT	43	66	3.62	0.92	≥51,680
Mathematics	8	CBT	42	66	3.67	0.92	≥45,250

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

Copyright © 2019 by Louisiana Department of Education.

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 9.3.

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

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

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 $q_{ij}(\theta_i) = 1 - p_{ij}(\theta_i)$.

Copyright © 2019 by Louisiana Department of Education.

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 9.3.

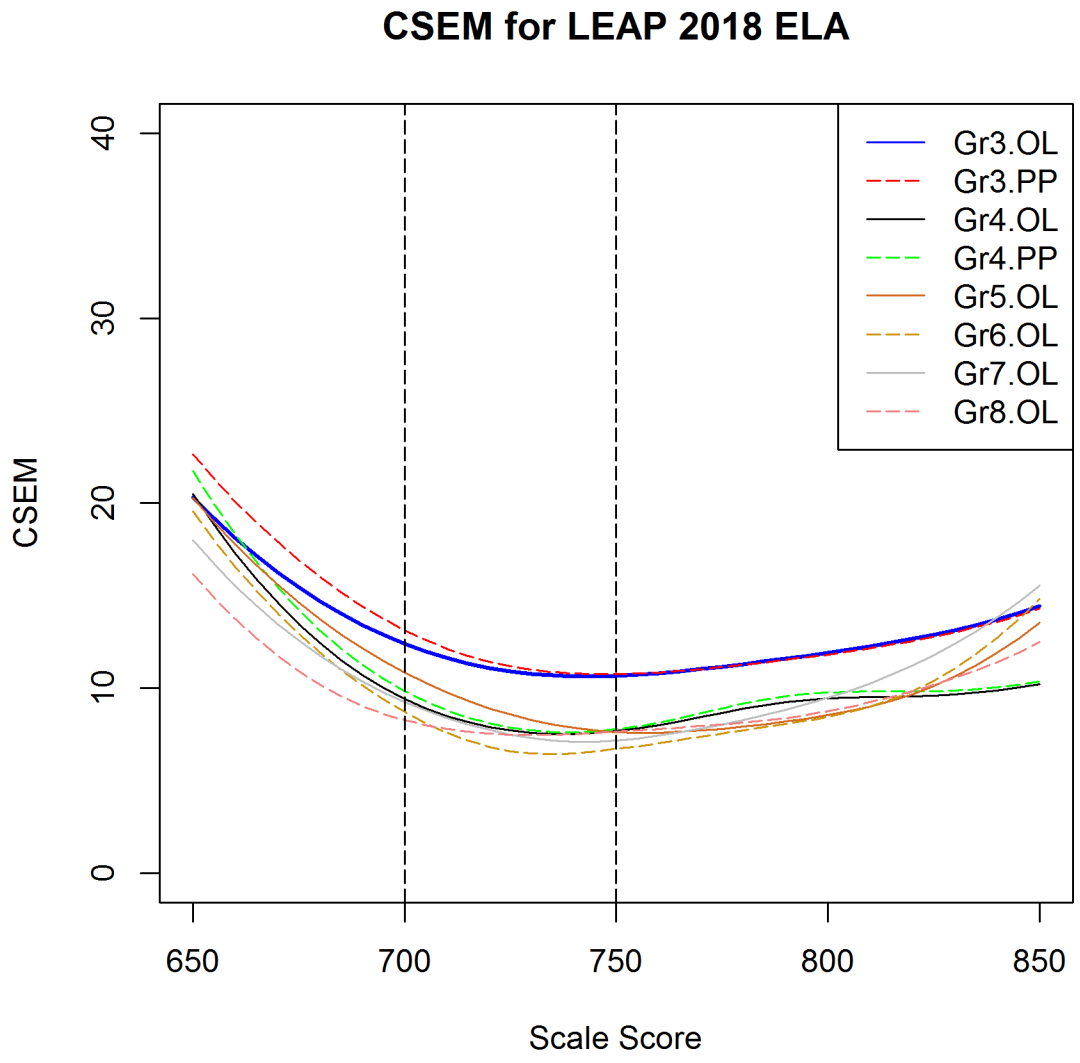
Table 9.3 Conditional Standard Errors of Measurement at the *Approaching Basic*, *Basic*, *Mastery*, and *Advanced* Cut Scores

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

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.

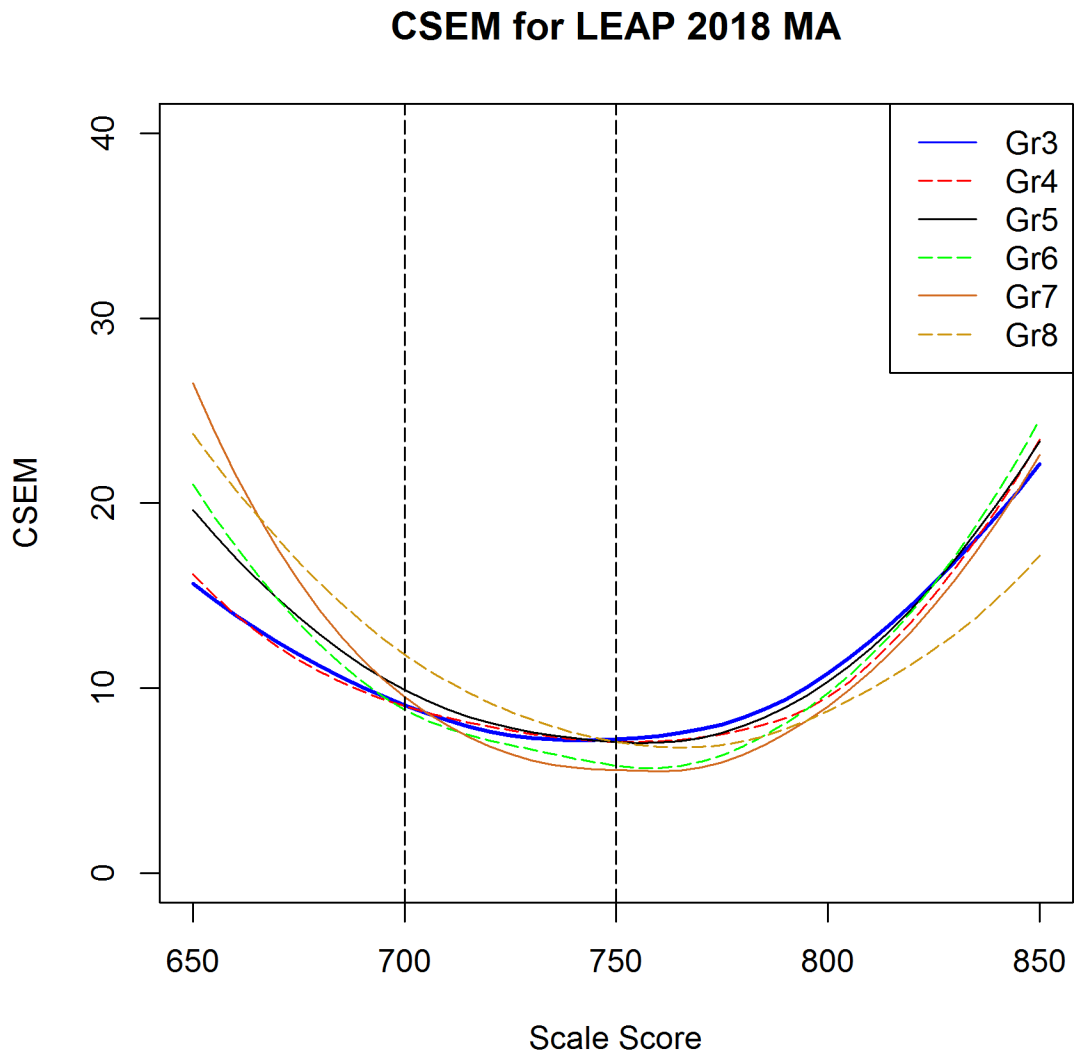
Copyright © 2019 by Louisiana Department of Education.

Figure 9.1 CSEM Curves for ELA Grades 3 through 8



Copyright © 2019 by Louisiana Department of Education.

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.

Copyright © 2019 by Louisiana Department of Education.

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 2018 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 9.4 and Table 9.5) and on cut score (see Table 9.6 and Table 9.7) are presented for the 2018 LEAP 2025 in this section of the report. The magnitude of classification consistency and accuracy measures is influenced by several key features of the test design, including the number of items, the location and number of cut scores, the score distribution, and the reliability and associated SEM. As can be seen in Table 9.4, classification accuracy conditioned on achievement level ranged from 0.37 to 0.84 for ELA and 0.59 to 0.86 for mathematics. Classification consistency (see Table 9.5) conditioned on achievement level ranged from 0.39 to 0.76 for ELA and 0.48 to 0.80 for mathematics. Table 9.6 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 9.7) conditioned at achievement cut points ranged from 0.85 to 0.97 for ELA and 0.84 to 0.98 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.

Copyright © 2019 by Louisiana Department of Education.

Table 9.4 Decision Accuracy Conditioned on Level of Achievement

	Decision Accuracy						
Content Area	Grade	Mode	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
ELA	3	CBT	0.78	0.60	0.62	0.79	0.37
ELA	3	PBT	0.71	0.59	0.62	0.83	0.53
ELA	4	CBT	0.74	0.67	0.70	0.78	0.59
ELA	4	PBT	0.69	0.66	0.70	0.79	0.53
ELA	5	CBT	0.61	0.64	0.69	0.84	0.53
ELA	6	CBT	0.64	0.76	0.77	0.82	0.59
ELA	7	CBT	0.73	0.68	0.73	0.78	0.74
ELA	8	CBT	0.70	0.66	0.67	0.80	0.61
Mathematics	3	CBT	0.77	0.70	0.75	0.80	0.62
Mathematics	3	PBT	0.76	0.68	0.75	0.84	0.62
Mathematics	4	CBT	0.70	0.71	0.76	0.85	0.64
Mathematics	4	PBT	0.66	0.69	0.75	0.86	0.61
Mathematics	5	CBT	0.57	0.70	0.78	0.80	0.68
Mathematics	6	CBT	0.62	0.70	0.77	0.86	0.66
Mathematics	7	CBT	0.56	0.73	0.78	0.84	0.66
Mathematics	8	CBT	0.76	0.59	0.68	0.86	0.59

Table 9.5 Decision Consistency Conditioned on Level of Achievement

	Decision Consistency						
Content Area	Grade	Mode	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
ELA	3	CBT	0.72	0.47	0.47	0.72	0.39
ELA	3	PBT	0.67	0.47	0.47	0.74	0.52
ELA	4	CBT	0.71	0.52	0.57	0.67	0.57
ELA	4	PBT	0.66	0.52	0.58	0.67	0.51
ELA	5	CBT	0.57	0.49	0.58	0.76	0.52
ELA	6	CBT	0.64	0.64	0.65	0.74	0.54
ELA	7	CBT	0.68	0.57	0.60	0.68	0.69
ELA	8	CBT	0.69	0.52	0.53	0.70	0.58
Mathematics	3	CBT	0.69	0.59	0.62	0.74	0.62
Mathematics	3	PBT	0.70	0.57	0.62	0.76	0.62
Mathematics	4	CBT	0.65	0.60	0.64	0.79	0.63
Mathematics	4	PBT	0.62	0.59	0.62	0.79	0.61
Mathematics	5	CBT	0.55	0.55	0.65	0.74	0.66
Mathematics	6	CBT	0.52	0.59	0.66	0.79	0.62
Mathematics	7	CBT	0.48	0.60	0.69	0.78	0.66
Mathematics	8	CBT	0.63	0.50	0.55	0.80	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

Copyright © 2019 by Louisiana Department of Education.

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 9.6 shows the classification accuracy and Table 9.7 shows the consistency estimates when conditioned on LEAP 2025 cut scores. The classification accuracy statistics are at or above 0.88, 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 9.6 Decision Accuracy at Achievement Cut Points

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

Copyright © 2019 by Louisiana Department of Education.

Table 9.7 Decision Consistency at Achievement Cut Points

	Decision Consistency					
Content Area	Grade	Mode	Unsatisfactory/ Approaching Basic	Approaching Basic/ Basic	Basic/ Mastery	Mastery/ Advanced
ELA	3	CBT	0.87	0.86	0.87	0.97
ELA	3	PBT	0.91	0.87	0.85	0.94
ELA	4	CBT	0.91	0.88	0.87	0.94
ELA	4	PBT	0.93	0.89	0.86	0.91
ELA	5	CBT	0.93	0.87	0.85	0.96
ELA	6	CBT	0.93	0.89	0.89	0.96
ELA	7	CBT	0.93	0.89	0.88	0.92
ELA	8	CBT	0.93	0.88	0.85	0.93
Mathematics	3	CBT	0.89	0.87	0.90	0.98
Mathematics	3	PBT	0.94	0.89	0.88	0.95
Mathematics	4	CBT	0.92	0.88	0.90	0.98
Mathematics	4	PBT	0.93	0.88	0.88	0.97
Mathematics	5	CBT	0.90	0.86	0.89	0.97
Mathematics	6	CBT	0.88	0.87	0.91	0.98
Mathematics	7	CBT	0.89	0.87	0.91	0.98
Mathematics	8	CBT	0.87	0.86	0.87	0.97

9.2.1 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 0.91 to 0.92 for mathematics forms, indicating items on the 2018 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.

Copyright © 2019 by Louisiana Department of Education.

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 9.8) and for mathematics (see Table 9.9). 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 9.8 and 9.9). 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.

Copyright © 2019 by Louisiana Department of Education.

Table 9.8 Principal Component Analysis for English Language Arts

Grade	Mode	Components	Eigenvalue	Percentage of Variance Explained	Cumulative Percentage of Variance Explained
3	CBT	First Component	6.57	25.28	25.28
3	CBT	Second Component	1.23	4.73	30.02
3	CBT	Ratio (First/Second)	5.34	.	.
3	PBT	First Component	6.62	25.46	25.46
3	PBT	Second Component	1.22	4.69	30.15
3	PBT	Ratio (First/Second)	5.43	.	.
4	CBT	First Component	7.62	27.21	27.21
4	CBT	Second Component	1.33	4.76	31.98
4	CBT	Ratio (First/Second)	5.71	.	.
4	PBT	First Component	7.02	25.07	25.07
4	PBT	Second Component	1.40	5.00	30.07
4	PBT	Ratio (First/Second)	5.02	.	.
5	CBT	First Component	6.90	23.00	23.00
5	CBT	Second Component	1.39	4.64	27.64
5	CBT	Ratio (First/Second)	4.95	.	.
6	CBT	First Component	8.94	27.94	27.94
6	CBT	Second Component	1.55	4.84	32.78
6	CBT	Ratio (First/Second)	5.77	.	.
7	CBT	First Component	8.97	28.03	28.03
7	CBT	Second Component	1.32	4.12	32.15
7	CBT	Ratio (First/Second)	6.80	.	.
8	CBT	First Component	7.61	23.79	23.79
8	CBT	Second Component	1.49	4.67	28.45
8	CBT	Ratio (First/Second)	5.10	.	.

Copyright © 2019 by Louisiana Department of Education.

Table 9.9 Principal Component Analysis for Mathematics

Grade	Mode	Components	Eigenvalue	Percentage of Variance Explained	Cumulative Percentage of Variance Explained
3	CBT	First Component	10.07	23.41	23.41
3	CBT	Second Component	1.84	4.27	27.68
3	CBT	Ratio (First/Second)	5.48	.	.
3	PBT	First Component	11.11	25.84	25.84
3	PBT	Second Component	1.88	4.36	30.20
3	PBT	Ratio (First/Second)	5.92	.	.
4	CBT	First Component	10.56	24.55	24.55
4	CBT	Second Component	2.17	5.04	29.59
4	CBT	Ratio (First/Second)	4.87	.	.
4	PBT	First Component	10.47	24.35	24.35
4	PBT	Second Component	1.98	4.59	28.95
4	PBT	Ratio (First/Second)	5.30	.	.
5	CBT	First Component	9.45	21.98	21.98
5	CBT	Second Component	1.81	4.22	26.20
5	CBT	Ratio (First/Second)	5.21	.	.
6	CBT	First Component	11.61	26.99	26.99
6	CBT	Second Component	1.65	3.83	30.83
6	CBT	Ratio (First/Second)	7.04	.	.
7	CBT	First Component	11.11	25.83	25.83
7	CBT	Second Component	1.59	3.70	29.53
7	CBT	Ratio (First/Second)	6.98	.	.
8	CBT	First Component	11.19	26.64	26.64
8	CBT	Second Component	1.34	3.20	29.84
8	CBT	Ratio (First/Second)	8.34	.	.

9.4 Analyses by Claims, Subclaims, and Subcategories

Three sets of analyses were conducted at the claim and subclaim level for ELA and at the subclaim and subcategory for 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 in ELA and between the subclaim scores and subcategory scores in mathematics 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, Subclaims, and Subcategories

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

Copyright © 2019 by Louisiana Department of Education.

presented above the diagonal portion of the matrix, and the reliability coefficients used are shown in Tables 9.10–9.13.

The uncorrected PPM in Tables 9.10–9.13 should be interpreted in the context of the reliability coefficient. In general, lower PPM coefficients are expected between variables that are less reliable. In most cases, the PPM coefficients show that performance on one 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 9.1). 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 9.10–9.13 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.

Copyright © 2019 by Louisiana Department of Education.

Table 9.10 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	CBT	1	Reading	22	.	0.84
	CBT	2	Writing	4	0.68	.
	PBT	1	Reading	22	.	0.82
	PBT	2	Writing	4	0.67	.
4	CBT	1	Reading	24	.	0.86
	CBT	2	Writing	4	0.76	.
	PBT	1	Reading	24	.	0.85
	PBT	2	Writing	4	0.73	.
5	CBT	1	Reading	26	.	0.79
	CBT	2	Writing	4	0.68	.
6	CBT	1	Reading	28	.	0.80
	CBT	2	Writing	4	0.71	.
7	CBT	1	Reading	28	.	0.83
	CBT	2	Writing	4	0.75	.
8	CBT	1	Reading	28	.	0.86
	CBT	2	Writing	4	0.76	.

Table 9.11 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	CBT	1	Reading Literary Text	10	.	0.99	1.00	0.95	0.81
	CBT	2	Reading Information Text	7	0.65	.	1.01	1.10	0.90
	CBT	3	Reading Vocabulary	5	0.65	0.60	.	0.91	0.83
	CBT	4	Written Expression	4	0.58	0.61	0.50	.	1.25
	CBT	5	Knowledge & Use of Language	2	0.55	0.56	0.52	0.73	.
	PBT	1	Reading Literary Text	10	.	1.00	0.97	0.88	0.78
	PBT	2	Reading Information Text	7	0.66	.	1.01	1.11	0.94
	PBT	3	Reading Vocabulary	5	0.63	0.61	.	0.88	0.78
	PBT	4	Written Expression	4	0.54	0.64	0.50	.	1.27
	PBT	5	Knowledge & Use of Language	2	0.54	0.60	0.50	0.76	.

Copyright © 2019 by Louisiana Department of Education.

Subclaim Uncorrected and Corrected Correlation Coefficients: English Language Arts (continued)									
Grade	**Mode**	**No.**	**Subclaim**	**N Items**	**1**	**2**	**3**	**4**	**5**
4	CBT	1	Reading Literary Text	7	.	1.03	1.03	1.03	1.02
	CBT	2	Reading Information Text	9	0.70	.	1.01	0.99	0.98
	CBT	3	Reading Vocabulary	8	0.68	0.68	.	0.87	0.86
	CBT	4	Written Expression	4	0.72	0.70	0.61	.	1.34
	CBT	5	Knowledge & Use of Language	2	0.71	0.70	0.60	0.99	.
	PBT	1	Reading Literary Text	7	.	1.04	1.02	1.02	1.01
	PBT	2	Reading Information Text	9	0.67	.	1.03	0.99	0.98
	PBT	3	Reading Vocabulary	8	0.66	0.66	.	0.87	0.86
	PBT	4	Written Expression	4	0.69	0.66	0.58	.	1.39
	PBT	5	Knowledge & Use of Language	2	0.68	0.65	0.58	0.99	.
5	CBT	1	Reading Literary Text	11	.	0.98	0.95	0.80	0.79
	CBT	2	Reading Information Text	10	0.67	.	1.01	0.96	0.95
	CBT	3	Reading Vocabulary	5	0.53	0.60	.	0.83	0.83
	CBT	4	Written Expression	4	0.53	0.68	0.48	.	1.34
	CBT	5	Knowledge & Use of Language	2	0.53	0.68	0.48	0.93	.
6	CBT	1	Reading Literary Text	13	.	0.96	0.97	0.81	0.79
	CBT	2	Reading Information Text	9	0.73	.	1.02	1.00	0.98
	CBT	3	Reading Vocabulary	6	0.68	0.65	.	0.84	0.82
	CBT	4	Written Expression	4	0.61	0.70	0.54	.	1.30
	CBT	5	Knowledge & Use of Language	2	0.61	0.70	0.54	0.93	.
7	CBT	1	Reading Literary Text	10	.	0.99	0.96	0.87	0.83
	CBT	2	Reading Information Text	13	0.75	.	0.97	0.94	0.89
	CBT	3	Reading Vocabulary	5	0.63	0.67	.	0.84	0.81
	CBT	4	Written Expression	4	0.64	0.72	0.56	.	1.18
	CBT	5	Knowledge & Use of Language	2	0.63	0.71	0.56	0.93	.
8	CBT	1	Reading Literary Text	9	.	1.10	1.02	0.96	0.94
	CBT	2	Reading Information Text	10	0.72	.	1.00	0.99	0.97
	CBT	3	Reading Vocabulary	9	0.65	0.63	.	0.78	0.79
	CBT	4	Written Expression	4	0.71	0.72	0.55	.	1.16
	CBT	5	Knowledge & Use of Language	2	0.69	0.71	0.56	0.95	.

Copyright © 2019 by Louisiana Department of Education.

Table 9.12 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	CBT	1	Major Content	27	.	0.98	0.94	0.97
	CBT	2	Additional & Supporting Con	10	0.75	.	0.99	1.01
	CBT	3	Expressing Mathematical Rea	3	0.70	0.66	.	1.03
	CBT	4	Modeling & Application	3	0.72	0.67	0.68	.
	PBT	1	Major Content	27	.	0.98	0.97	1.01
	PBT	2	Additional & Supporting Con	10	0.78	.	1.00	0.99
	PBT	3	Expressing Mathematical Rea	3	0.73	0.68	.	1.03
	PBT	4	Modeling & Application	3	0.76	0.68	0.66	.
4	CBT	1	Major Content	29	.	0.96	0.92	0.86
	CBT	2	Additional & Supporting Con	8	0.74	.	1.00	0.97
	CBT	3	Expressing Mathematical Rea	3	0.70	0.68	.	1.00
	CBT	4	Modeling & Application	3	0.69	0.68	0.71	.
	PBT	1	Major Content	29	.	0.95	0.92	0.89
	PBT	2	Additional & Supporting Con	8	0.72	.	1.00	0.98
	PBT	3	Expressing Mathematical Rea	3	0.70	0.66	.	1.01
	PBT	4	Modeling & Application	3	0.71	0.68	0.70	.
5	CBT	1	Major Content	28	.	0.95	1.01	0.94
	CBT	2	Additional & Supporting Con	9	0.69	.	1.00	0.97
	CBT	3	Expressing Mathematical Rea	3	0.73	0.63	.	1.04
	CBT	4	Modeling & Application	3	0.70	0.63	0.68	.
6	CBT	1	Major Content	27	.	0.99	0.97	0.95
	CBT	2	Additional & Supporting Con	9	0.74	.	0.97	0.96
	CBT	3	Expressing Mathematical Rea	4	0.78	0.66	.	1.04
	CBT	4	Modeling & Application	3	0.73	0.63	0.72	.
7	CBT	1	Major Content	27	.	0.97	0.97	1.01
	CBT	2	Additional & Supporting Con	9	0.70	.	0.98	1.04
	CBT	3	Expressing Mathematical Rea	4	0.79	0.65	.	1.06
	CBT	4	Modeling & Application	3	0.72	0.61	0.69	.
8	CBT	1	Major Content	27	.	1.02	0.97	1.05
	CBT	2	Additional & Supporting Con	8	0.79	.	1.02	1.08
	CBT	3	Expressing Mathematical Rea	4	0.77	0.73	.	1.07
	CBT	4	Modeling & Application	3	0.70	0.64	0.65	.

Copyright © 2019 by Louisiana Department of Education.

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

Grade	Mode	No.	Subcategory	N Items	1	2	3	4
03	CBT	1	A1	9	.	1.11	0.80	0.89
	CBT	2	A2	6	0.85	.	0.84	0.93
	CBT	3	A3	9	0.56	0.56	.	0.92
	CBT	4	A4	7	0.57	0.57	0.51	.
	PBT	1	A1	9	.	1.10	0.83	0.86
	PBT	2	A2	6	0.86	.	0.85	0.92
	PBT	3	A3	9	0.63	0.62	.	0.90
	PBT	4	A4	7	0.59	0.59	0.57	.
04	CBT	1	A1	5	.	0.74	0.99	.
	CBT	2	A2	9	0.49	.	0.79	.
	CBT	3	A3	9	0.67	0.55	.	.
	PBT	1	A1	5	.	0.79	1.00	.
	PBT	2	A2	9	0.52	.	0.82	.
	PBT	3	A3	9	0.66	0.58	.	.
05	CBT	1	A1	6	.	0.93	0.87	0.90
	CBT	2	A2	8	0.54	.	0.92	0.82
	CBT	3	A3	6	0.49	0.54	.	0.66
	CBT	4	A4	7	0.55	0.52	0.41	.
06	CBT	1	A1	6	.	0.93	0.92	.
	CBT	2	A2	8	0.64	.	0.98	.
	CBT	3	A3	13	0.68	0.72	.	.
07	CBT	1	A1	10	.	0.98	0.99	.
	CBT	2	A2	11	0.73	.	1.01	.
	CBT	3	A3	6	0.67	0.71	.	.
08	CBT	1	A1	7	.	0.91	0.93	0.91
	CBT	2	A2	9	0.58	.	0.88	0.97
	CBT	3	A3	6	0.57	0.61	.	0.88
	CBT	4	A4	5	0.53	0.64	0.55	.

9.4.2 **Reliability of Claims, Subclaims, or Subcategories**

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 9.14–9.17 for ELA and mathematics. Reliability indices, such as Cronbach's coefficient alpha (and resulting SEM), are a function of the number of 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.

Copyright © 2019 by Louisiana Department of Education.

9.4.3 Standard Error of Measurement of Claims, Subclaims, or Subcategories

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

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

Grade	Mode	Claim	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
3	CBT	Reading	23	47	19.44	9.02	3.61	0.84
	CBT	Writing	4	24	4.84	4.18	1.94	0.79
	PBT	Reading	23	47	19.90	9.47	3.76	0.84
	PBT	Writing	4	24	7.20	4.74	2.09	0.80
4	CBT	Reading	26	56	21.81	10.33	3.95	0.85
	CBT	Writing	4	30	8.09	5.94	1.78	0.91
	PBT	Reading	26	56	22.79	10.42	4.19	0.84
	PBT	Writing	4	30	9.81	5.83	1.85	0.90
5	CBT	Reading	27	56	24.07	9.58	3.88	0.84
	CBT	Writing	4	27	8.24	6.30	2.15	0.88
6	CBT	Reading	29	60	27.33	12.12	4.02	0.89
	CBT	Writing	4	30	8.68	6.63	2.23	0.89
7	CBT	Reading	29	60	28.11	11.97	4.10	0.88
	CBT	Writing	4	30	12.57	8.16	2.44	0.91
8	CBT	Reading	30	64	28.84	10.59	4.25	0.84
	CBT	Writing	4	30	13.14	7.30	1.93	0.93

Copyright © 2019 by Louisiana Department of Education.

Table 9.15 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 Pts.	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
3	CBT	Reading Literary Text	10	20	8.44	4.39	2.34	0.71
	CBT	Reading Information Text	8	17	6.46	3.39	2.15	0.60
	CBT	Reading Vocabulary	5	10	4.54	2.54	1.64	0.59
	CBT	Written Expression	2	18	3.36	3.21	2.24	0.51
	CBT	Knowledge & Use of Language	2	6	1.48	1.22	0.72	0.66
	PBT	Reading Literary Text	10	20	8.49	4.50	2.43	0.71
	PBT	Reading Information Text	8	17	6.42	3.74	2.33	0.61
	PBT	Reading Vocabulary	5	10	4.99	2.57	1.62	0.60
	PBT	Written Expression	2	18	5.07	3.70	2.51	0.54
	PBT	Knowledge & Use of Language	2	6	2.12	1.27	0.74	0.66
4	CBT	Reading Literary Text	8	18	7.42	3.72	2.18	0.66
	CBT	Reading Information Text	10	22	6.99	4.07	2.26	0.69
	CBT	Reading Vocabulary	8	16	7.39	3.83	2.23	0.66
	CBT	Written Expression	2	24	6.07	4.47	2.31	0.73
	CBT	Knowledge & Use of Language	2	6	2.02	1.49	0.76	0.74
	PBT	Reading Literary Text	8	18	7.54	4.01	2.37	0.65
	PBT	Reading Information Text	10	22	7.52	4.04	2.45	0.63
	PBT	Reading Vocabulary	8	16	7.73	3.80	2.27	0.64
	PBT	Written Expression	2	24	7.34	4.39	2.40	0.70
	PBT	Knowledge & Use of Language	2	6	2.47	1.46	0.78	0.71
5	CBT	Reading Literary Text	11	22	9.01	4.08	2.43	0.64
	CBT	Reading Information Text	11	24	9.69	4.59	2.41	0.72
	CBT	Reading Vocabulary	5	10	5.38	2.37	1.70	0.49
	CBT	Written Expression	2	21	6.19	4.82	2.68	0.69
	CBT	Knowledge & Use of Language	2	6	2.05	1.55	0.85	0.70
6	CBT	Reading Literary Text	13	26	11.79	6.02	2.53	0.82
	CBT	Reading Information Text	10	22	9.28	4.66	2.53	0.71
	CBT	Reading Vocabulary	6	12	6.26	2.80	1.80	0.59
	CBT	Written Expression	2	24	6.50	5.14	2.84	0.69
	CBT	Knowledge & Use of Language	2	6	2.18	1.58	0.81	0.73
7	CBT	Reading Literary Text	10	20	9.13	4.36	2.30	0.72
	CBT	Reading Information Text	14	30	12.80	6.38	2.94	0.79
	CBT	Reading Vocabulary	5	10	6.17	2.55	1.62	0.59
	CBT	Written Expression	2	24	9.48	6.39	3.14	0.76
	CBT	Knowledge & Use of Language	2	6	3.09	1.87	0.81	0.81
8	CBT	Reading Literary Text	10	22	9.05	4.29	2.48	0.67
	CBT	Reading Information Text	11	24	9.24	4.28	2.57	0.64
	CBT	Reading Vocabulary	9	18	10.55	3.42	2.13	0.61
	CBT	Written Expression	2	24	9.85	5.62	2.39	0.82
	CBT	Knowledge & Use of Language	2	6	3.28	1.74	0.74	0.82

Copyright © 2019 by Louisiana Department of Education.

Table 9.16 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	CBT	Major Content	27	30	14.45	5.84	2.23	0.85
	CBT	Additional & Supporting Content	10	10	5.40	2.28	1.28	0.68
	CBT	Expressing Mathematical Reasoning	3	6	3.13	2.39	1.40	0.66
	CBT	Modeling & Application	3	8	2.80	2.78	1.65	0.65
	PBT	Major Content	27	30	16.51	6.36	2.23	0.88
	PBT	Additional & Supporting Content	10	10	6.16	2.34	1.24	0.72
	PBT	Expressing Mathematical Reasoning	3	6	4.08	2.33	1.41	0.63
	PBT	Modeling & Application	3	8	3.79	3.07	1.81	0.65
4	CBT	Major Content	29	30	17.88	6.22	2.24	0.87
	CBT	Additional & Supporting Content	8	10	4.41	2.36	1.35	0.67
	CBT	Expressing Mathematical Reasoning	3	5	2.74	2.33	1.32	0.68
	CBT	Modeling & Application	3	7	2.19	2.68	1.36	0.74
	PBT	Major Content	29	30	18.77	6.17	2.22	0.87
	PBT	Additional & Supporting Content	8	10	4.59	2.33	1.36	0.66
	PBT	Expressing Mathematical Reasoning	3	5	2.99	2.38	1.37	0.67
	PBT	Modeling & Application	3	7	2.86	2.99	1.57	0.73
5	CBT	Major Content	28	30	15.10	5.98	2.38	0.84
	CBT	Additional & Supporting Content	9	10	4.58	2.29	1.40	0.63
	CBT	Expressing Mathematical Reasoning	3	3	3.36	2.30	1.39	0.63
	CBT	Modeling & Application	3	9	2.58	2.77	1.59	0.67
6	CBT	Major Content	27	30	13.73	6.71	2.23	0.89
	CBT	Additional & Supporting Content	9	10	4.29	2.34	1.41	0.63
	CBT	Expressing Mathematical Reasoning	4	10	3.47	3.24	1.70	0.73
	CBT	Modeling & Application	3	9	2.26	3.04	1.75	0.67
7	CBT	Major Content	27	30	12.10	6.76	2.29	0.88
	CBT	Additional & Supporting Content	9	10	3.64	2.10	1.34	0.59
	CBT	Expressing Mathematical Reasoning	4	10	3.40	2.90	1.49	0.74
	CBT	Modeling & Application	3	9	1.74	2.57	1.66	0.58
8	CBT	Major Content	27	30	11.23	6.23	2.26	0.87
	CBT	Additional & Supporting Content	8	10	3.54	2.42	1.34	0.69
	CBT	Expressing Mathematical Reasoning	4	14	2.16	2.71	1.38	0.74
	CBT	Modeling & Application	3	7	2.58	2.75	1.94	0.50

Copyright © 2019 by Louisiana Department of Education.

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

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	CBT	A1	9	9	5.17	2.70	1.18	0.81
	CBT	A2	6	8	3.59	2.24	1.16	0.73
	CBT	A3	9	9	4.21	1.80	1.11	0.62
	CBT	A4	7	8	3.63	1.65	1.15	0.51
	PBT	A1	9	9	5.94	2.63	1.13	0.82
	PBT	A2	6	8	4.18	2.29	1.16	0.74
	PBT	A3	9	9	5.04	2.08	1.13	0.71
	PBT	A4	7	8	3.86	1.82	1.20	0.57
4	CBT	A1	5	5	3.19	1.42	0.84	0.64
	CBT	A2	9	10	4.58	2.43	1.34	0.69
	CBT	A3	9	9	5.97	2.21	1.20	0.71
	PBT	A1	5	5	3.34	1.35	0.84	0.61
	PBT	A2	9	10	4.90	2.47	1.35	0.70
	PBT	A3	9	9	6.24	2.16	1.17	0.70
5	CBT	A1	6	6	3.31	1.61	1.06	0.56
	CBT	A2	8	8	3.82	1.91	1.20	0.60
	CBT	A3	6	7	2.52	1.61	1.06	0.57
	CBT	A4	7	8	4.82	2.18	1.26	0.67
6	CBT	A1	6	6	3.39	1.82	1.01	0.69
	CBT	A2	8	9	4.07	2.22	1.25	0.68
	CBT	A3	13	15	6.27	3.45	1.53	0.80
7	CBT	A1	10	11	4.49	2.65	1.41	0.72
	CBT	A2	11	13	5.45	3.22	1.52	0.78
	CBT	A3	6	6	2.16	1.61	0.97	0.64
8	CBT	A1	7	7	2.48	1.68	1.11	0.56
	CBT	A2	9	9	4.16	2.39	1.27	0.72
	CBT	A3	6	7	1.52	1.70	0.98	0.67
	CBT	A4	5	7	3.07	1.74	1.10	0.60

Copyright © 2019 by Louisiana Department of Education.

9.5 Divergent (Discriminant) Validity

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

To assess the divergent validity of the LEAP 2025 assessments, correlations were computed between the ELA, mathematics, and social studies scale scores for students who took more than one LEAP 2025 content-area test in 2018. Science was not included as it was not administered operationally in 2018. These correlations are based on the census data, and the results are shown in Table 9.18. The correlation coefficients ranged from 0.68 (between mathematics and social studies in grade 6) to 0.83 (between ELA and social studies in grade 8). The correlation coefficients suggest that individual student scores across subjects are moderately related, indicating that these tests measure a similar knowledge base or general underlying ability but still measure some different traits as planned.

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

Grade	ELA/ Mathematics	ELA/ Social Studies	Mathematics/ Social Studies
3	0.72	0.77	0.71
4	0.71	0.79	0.72
5	0.70	0.77	0.68
6	0.77	0.81	0.76
7	0.79	0.82	0.78
8	0.72	0.83	0.74

9.6 Regression of LEAP 2025 from 2017 to 2018

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

This study required the collection of data from adjacent grades for each content area. For this purpose, matched longitudinal LEAP 2025 test data from spring 2017 and spring 2018 were used. For example, grade 3 students were matched with grade 4 students, and only matched students were used to estimate correlation and preform linear regression from 2017 to 2018. Table 9.19 summarizes the correlation and regression results for 2017 and 2018 LEAP 2025. For ELA, the correlation ranged from 0.75 to 0.84, and for mathematics, the correlation ranged from 0.80 to 0.85. That is, correlations for mathematics were slightly higher than those for ELA. Correlations for both content areas can be considered moderate, which can be often found in

state assessments. R2 indicates how much of 2017 performance can explain 2018. For example, 0.59 for ELA 2017 grade 3 and 2018 grade 4 means that 2017's grade 3 performance can explain (predict) about 59% of 2018's grade 4 performance. This R2 value is generally the power of 2 for the matching correlation. The R2 values for ELA range from 0.59 to 0.70, and those for mathematics range from 0.63 to 0.72. These also show the moderate relationships between adjacent grades for both ELA and mathematics.

Table 9.19 Correlation and Regression Summary for 2017 and 2018 LEAP 2025

Content	2017 Grade	2018 Grade	N	Correlation	R2	Slope	Intercept
ELA	3	4	≥52,130	0.77	0.59	0.68	238.73
	4	5	≥52,160	0.75	0.56	0.73	201.86
	5	6	≥49,410	0.79	0.62	0.76	170.45
	6	7	≥48,170	0.84	0.70	0.96	33.34
	7	8	≥47,650	0.82	0.67	0.79	162.38
Mathematics	3	4	≥52,110	0.80	0.64	0.76	177.43
	4	5	≥52,090	0.80	0.63	0.77	164.61
	5	6	≥49,330	0.80	0.65	0.80	144.92
	6	7	≥48,090	0.85	0.72	0.76	179.64
	7	8	≥41,670	0.82	0.67	1.18	-129.51

Figures 9.3 and 9.4 show regression line and scatter plots for ELA and mathematics. The linear lines in the plots are linear regression lines from 2017 to 2018. In general, the length of band given the linear regression line shows the strength of correlation. If the band is narrow, the correlation is high, and if the band is large, the correlation is low. Every plot shows some moderate linear relationships between 2017 and 2018 adjacent grades for both ELA and mathematics.

Figure 9.3 Regression Line and Scatter Plots: ELA

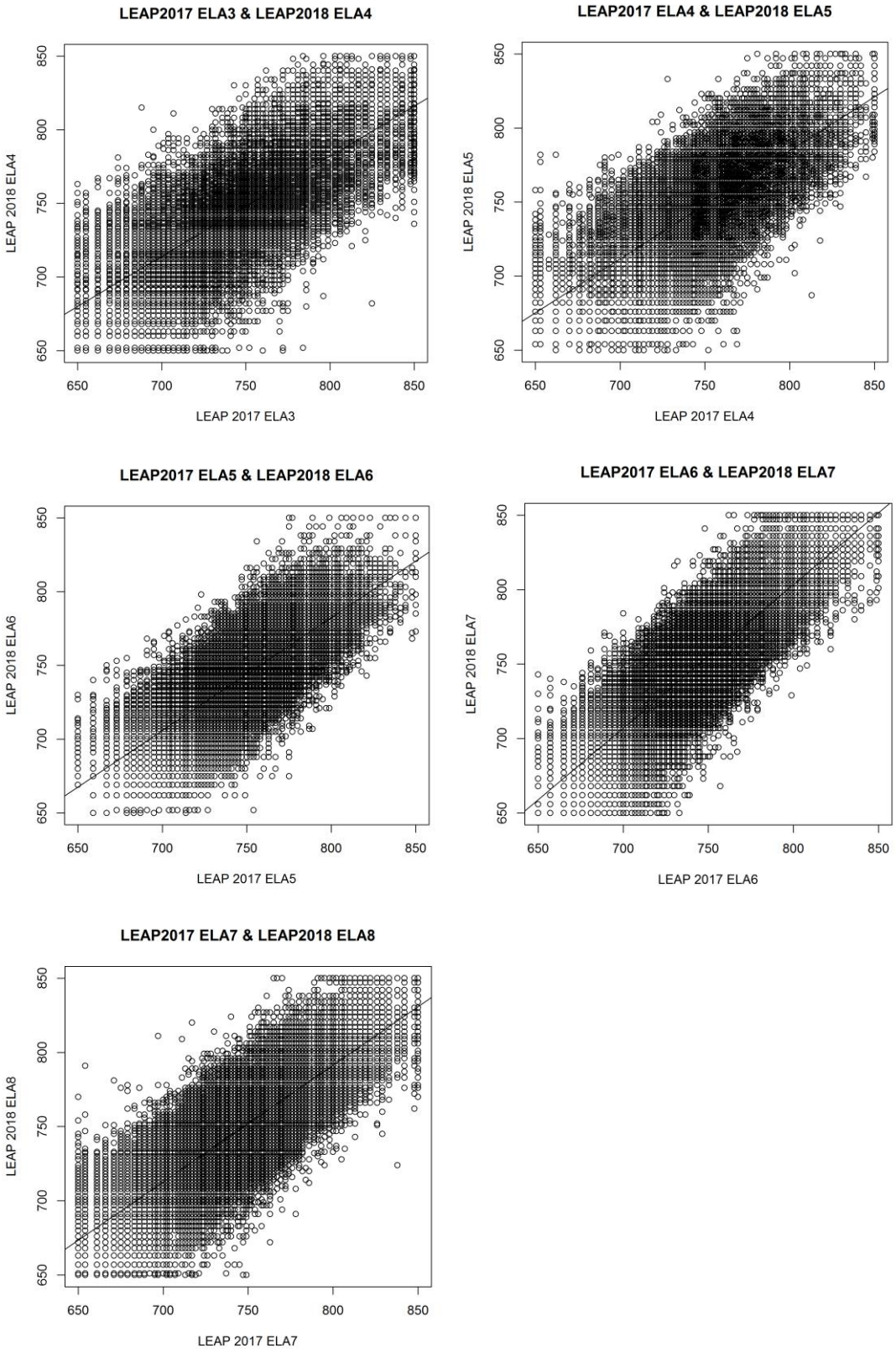
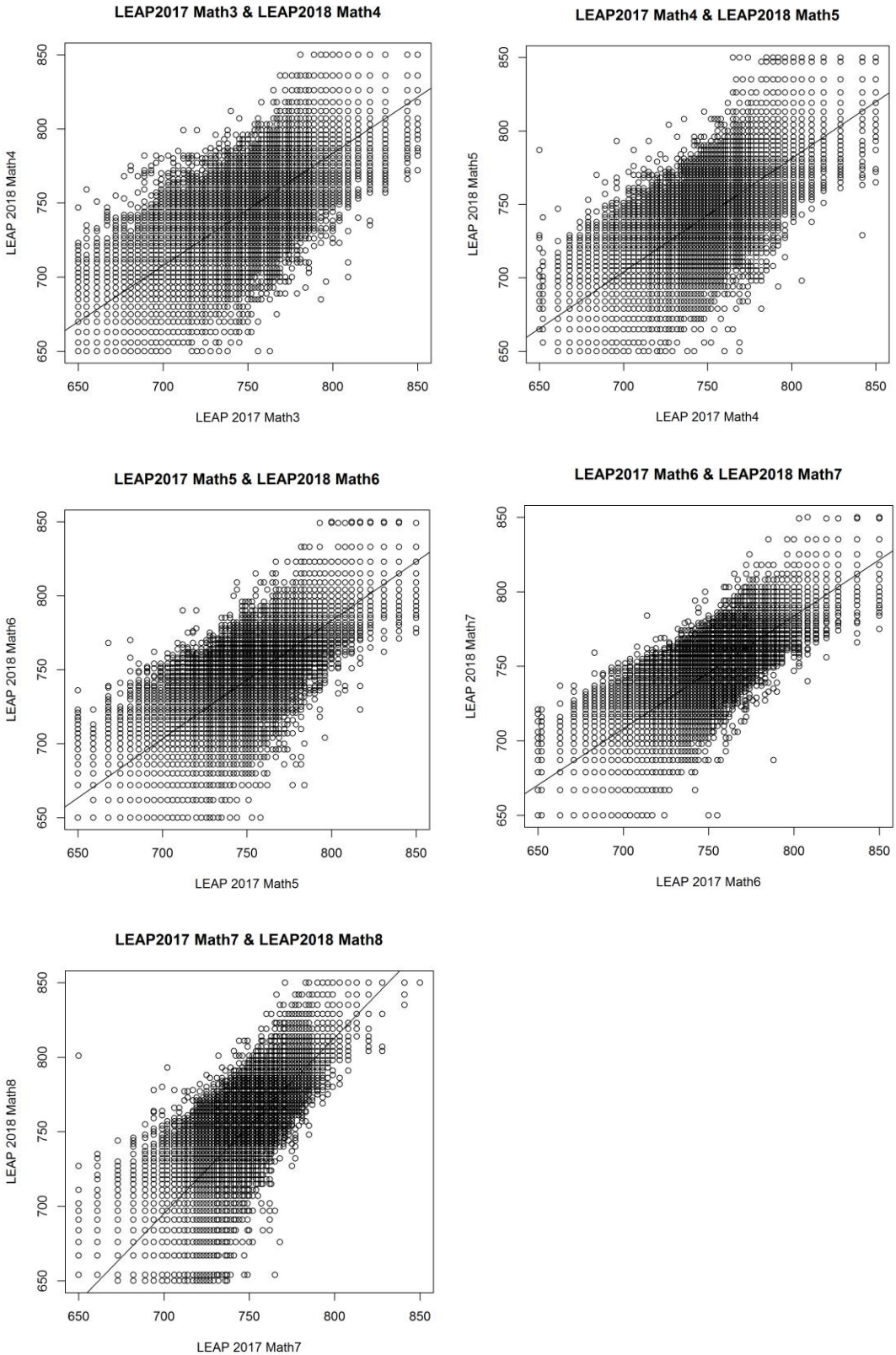


Figure 9.4 Regression Line and Scatter Plots: Mathematics



9.7 Summary

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

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

CHAPTER 10: FAIRNESS

As noted in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), there are varying definitions of fairness. This chapter examines fairness as it relates to minimizing bias on a test. This chapter also discusses test performance among varying subgroups assessed by LEAP 2025 assessments. It should be noted that 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)

Test scores of examinee subgroups differing in meaning 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, along with Louisiana educators, conduct content and bias reviews on items during the selection process, as explained in Chapter 3. These practices adhere to Standard 3.3, which states, "Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test" (64).

The PARCC consortium 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 2018 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 PARCC Fairness and Sensitivity Guidelines for reducing or eliminating bias. DRC test development staff reviewed all items and other testing materials with these guidelines in mind. Internal editorial reviews were conducted by at least three different people: a content editor who directly supervised the item writers, a style editor, and a content supervisor. The final test was again reviewed by 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 kth level of the matching variable. Note that the MH statistic is sensitive to N such that larger sample sizes increase the value of chi-square.

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

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

where N_{r1k} is the number of correct responses in the reference group at ability level k, N_{f0k} is the number of incorrect responses in the focal group at ability level k, N_k is the total number of responses, N_{f1k} is the number of correct responses in the focal group at ability level k, and N_{r0k} is the number of incorrect responses in the reference group at ability level k. MH-D DIF is then computed as follows:

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

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

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

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

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

where p_{Fk} = proportion of the focal group members at the kth 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.

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

Economic Status: Focal group is students who are classified as economically disadvantaged; reference group is all others.

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 CBT 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 Black or African American subgroup: it exhibited moderate positive DIF. Two items each were flagged for the Hispanic/Latino and Asian subgroups; one in each subgroup showed moderate negative DIF and one showed large negative DIF. Two items were flagged for the EL subgroup for moderate negative DIF.

Table 10.1 2018 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	CBT	26	Gender	Female	0	0	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	0	0	1	0
			Ethnicity	Two or More Races	NR	NR	NR	NR
			Education Classification	Special	NR	NR	NR	NR
			EL Status	EL	NR	NR	NR	NR
			Economic Status	Economically Disadvantaged	NR	NR	NR	NR
			Section 504 Status	Section 504	NR	NR	NR	NR
3	PBT	26	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	1	0	1	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	1	0	1	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	2	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0

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	28	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	1	0	0	0
			Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
			Ethnicity	Asian	NR	NR	NR	NR
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	NR	NR	NR	NR
			Education Classification	Special	0	0	0	0
			EL Status	EL	2	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
	PBT	28	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	1	1	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
5	CBT	30	Gender	Female	0	2	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	1	0	0	0
			EL Status	EL	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	32	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	1	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	1	0	0	0
			EL Status	EL	1	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
7	CBT	32	Gender	Female	1	2	0	2
			Ethnicity	Hispanic/Latino	0	0	1	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	0	0	0
			EL Status	EL	2	0	1	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
8	CBT	32	Gender	Female	0	2	0	0
			Ethnicity	Hispanic/Latino	1	0	1	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	1	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Education Classification	Special	0	0	0	0
			EL Status	EL	1	0	2	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0

Table 10.2 2018 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	CBT	43	Gender	Female	1	1	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	1	1	1	0
			Ethnicity	Two or More Races	NR	NR	NR	NR
			Education Classification	Special	NR	NR	NR	NR
			EL Status	EL	NR	NR	NR	NR
			Economic Status	Economically Disadvantaged	NR	NR	NR	NR
			Section 504 Status	Section 504	NR	NR	NR	NR
3	PBT	43	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	1	2	0	0
			Ethnicity	Black or African American	1	1	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Ed. Classification	Special	2	1	2	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
4	CBT	43	Gender	Female	1	0	0	0
			Ethnicity	Hispanic/Latino	2	0	0	0
			Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
			Ethnicity	Asian	NR	NR	NR	NR
			Ethnicity	Black or African American	1	0	1	0
			Ethnicity	Two or More Races	NR	NR	NR	NR
			Ed. Classification	Special	5	2	3	3
			EL Status	EL	1	0	0	0
			Economic Status	Economically Disadvantaged	1	0	0	0
			Section 504 Status	Section 504	0	0	0	0
	PBT	43	Gender	Female	1	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	2	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Ed. Classification	Special	4	1	0	3
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0

DIF Statistics: Mathematics (continued)					Count of Items at DIF Magnitude			
					Moderate		Large	
Grade	Mode	Number of Items	Category	Group	B-	B+	C-	C+
5	CBT	43	Gender	Female	0	1	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	1	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	1	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Ed. Classification	Special	4	2	1	4
			EL Status	EL	1	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	1	0	0
6	CBT	43	Gender	Female	1	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Ed. Classification	Special	1	0	1	2
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
7	CBT	43	Gender	Female	1	0	0	1
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	1	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or More Races	0	0	0	0
			Ed. Classification	Special	1	0	0	0
			EL Status	EL	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			Section 504 Status	Section 504	0	0	0	0
8	CBT	42	Gender	Female	0	1	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	1	0	0	0
			Ethnicity	Asian	0	1	0	0
			Ethnicity	Black or African American	0	0	1	0
			Ethnicity	Two or More Races	0	0	0	0
			Ed. Classification	Special	1	0	0	0
			EL Status	EL	0	0	1	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 10.3–10.10 show the regular test form reliability coefficients and SEM by student gender, ethnicity, education classification, EL status, economic status, and 504 status. The reliability coefficients for English Language Arts forms ranged from 0.82 to 0.93. 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 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	≥1,030	0.87	4.39	≥1,040	0.91	3.56
Gender						
Female	≥510	0.87	4.45	≥510	0.91	3.58
Male	≥520	0.87	4.31	≥530	0.91	3.54
Ethnicity						
Hispanic/Latino	≥70	0.90	4.32	≥70	0.91	3.55
American Indian or Alaska Native	<10	NR	NR	<10	NR	NR
Asian	≥10	0.93	4.43	≥10	0.91	3.96
Black or African American	≥480	0.84	4.35	≥480	0.90	3.39
Native Hawaiian or Other Pacific	<10	NR	NR	<10	NR	NR
White	≥420	0.87	4.41	≥420	0.91	3.67
Two or More Races	≥30	0.89	4.57	≥30	0.92	3.69
Education Classification						
Regular	≥950	0.87	4.41	≥950	0.91	3.58
Special	≥80	0.86	4.03	≥90	0.91	3.22
English Learner Status						
Non-EL	≥1,000	0.87	4.39	≥1,000	0.91	3.57
EL	≥30	0.87	4.47	≥30	0.91	3.46
Economic Status						
Economically Disadvantaged	≥810	0.87	4.35	≥820	0.90	3.47
Not Economically Disadvantaged	≥210	0.86	4.54	≥210	0.91	3.79
Section 504 Status						
Non-Section 504	≥990	0.87	4.39	≥990	0.91	3.58
Section 504	≥40	0.82	4.17	≥40	0.84	3.11

Table 10.4 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	≥54,350	0.87	4.65	≥54,320	0.92	3.64
Gender						
Female	≥26,590	0.87	4.72	≥26,570	0.91	3.63
Male	≥27,740	0.87	4.55	≥27,720	0.92	3.63
Ethnicity						
Hispanic/Latino	≥4,270	0.88	4.59	≥4,230	0.92	3.59
American Indian or Alaska Native	≥320	0.87	4.71	≥320	0.91	3.68
Asian	≥750	0.89	4.87	≥750	0.91	3.68
Black or African American	≥23,850	0.86	4.58	≥23,850	0.91	3.50
Native Hawaiian or Other Pacific	≥40	0.91	4.41	≥40	0.93	3.77
White	≥23,370	0.86	4.73	≥23,360	0.91	3.68
Two or More Races	≥1,680	0.86	4.73	≥1,680	0.91	3.66
Education Classification						
Regular	≥48,400	0.87	4.67	≥48,380	0.92	3.65
Special	≥5,950	0.85	4.37	≥5,930	0.92	3.39
English Learner Status						
Non-EL	≥51,650	0.87	4.66	≥51,660	0.92	3.64
EL	≥2,690	0.85	4.51	≥2,660	0.92	3.51
Economic Status						
Economically Disadvantaged	≥39,170	0.86	4.60	≥39,130	0.91	3.57
Not Economically Disadvantaged	≥14,320	0.86	4.78	≥14,320	0.91	3.66
Section 504 Status						
Non-Section 504	≥49,540	0.87	4.67	≥49,520	0.92	3.64
Section 504	≥4,800	0.84	4.45	≥4,800	0.90	3.48

Table 10.5 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	≥5,160	0.89	5.02	≥5,160	0.92	3.46
Gender						
Female	≥2,570	0.89	5.09	≥2,570	0.91	3.48
Male	≥2,580	0.90	4.91	≥2,580	0.92	3.43
Ethnicity						
Hispanic/Latino	≥580	0.90	5.10	≥580	0.91	3.42
American Indian or Alaska Native	≥20	0.94	4.66	≥20	0.92	3.48
Asian	≥90	0.90	5.30	≥90	0.91	3.86
Black or African American	≥2,130	0.87	4.91	≥2,130	0.90	3.24
Native Hawaiian or Other Pacific	<10	NR	NR	<10	NR	NR
White	≥2,180	0.89	5.06	≥2,180	0.92	3.57
Two or More Races	≥130	0.86	5.26	≥130	0.91	3.46
Education Classification						
Regular	≥4,560	0.89	5.05	≥4,550	0.92	3.47
Special	≥600	0.89	4.45	≥600	0.91	3.22
English Learner Status						
Non-EL	≥4,830	0.89	5.02	≥4,830	0.92	3.47
EL	≥330	0.84	5.07	≥320	0.90	3.26
Economic Status						
Economically Disadvantaged	≥3,800	0.87	4.99	≥3,800	0.91	3.33
Not Economically Disadvantaged	≥1,320	0.88	5.05	≥1,320	0.91	3.63
Section 504 Status						
Non-Section 504	≥4,780	0.89	5.03	≥4,770	0.92	3.47
Section 504	≥380	0.83	4.80	≥380	0.88	3.17

Table 10.6 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	≥50,590	0.88	5.28	≥50,520	0.92	3.59
Gender						
Female	≥24,750	0.88	5.32	≥24,730	0.91	3.61
Male	≥25,790	0.88	5.20	≥25,760	0.92	3.57
Ethnicity						
Hispanic/Latino	≥3,670	0.89	5.30	≥3,630	0.92	3.56
American Indian or Alaska Native	≥300	0.88	5.16	≥300	0.91	3.70
Asian	≥790	0.89	5.27	≥790	0.91	3.79
Black or African American	≥22,370	0.86	5.31	≥22,370	0.91	3.38
Native Hawaiian or Other Pacific	≥50	0.89	5.23	≥50	0.92	3.78
White	≥21,800	0.86	5.24	≥21,800	0.90	3.69
Two or More Races	≥1,520	0.86	5.30	≥1,520	0.91	3.63
Education Classification						
Regular	≥45,050	0.88	5.27	≥44,960	0.91	3.61
Special	≥5,540	0.86	5.02	≥5,560	0.90	3.28
English Learner Status						
Non-EL	≥48,410	0.88	5.28	≥48,380	0.92	3.60
EL	≥2,170	0.85	5.29	≥2,140	0.91	3.36
Economic Status						
Economically Disadvantaged	≥36,270	0.87	5.29	≥36,210	0.91	3.50
Not Economically Disadvantaged	≥13,590	0.86	5.22	≥13,580	0.90	3.71
Section 504 Status						
Non-Section 504	≥45,280	0.88	5.28	≥45,210	0.92	3.61
Section 504	≥5,300	0.85	5.18	≥5,310	0.90	3.35

Table 10.7 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	≥55,310	0.88	5.13	≥55,200	0.90	3.66
Gender						
Female	≥27,170	0.87	5.22	≥27,110	0.90	3.67
Male	≥28,140	0.88	4.97	≥28,080	0.91	3.65
Ethnicity						
Hispanic/Latino	≥3,920	0.88	5.08	≥3,870	0.90	3.62
American Indian or Alaska Native	≥330	0.85	5.24	≥330	0.89	3.65
Asian	≥820	0.90	5.23	≥820	0.91	3.78
Black or African American	≥24,000	0.86	4.99	≥23,960	0.88	3.47
Native Hawaiian or Other Pacific	≥40	0.86	5.49	≥40	0.91	3.72
White	≥24,560	0.86	5.23	≥24,550	0.90	3.75
Two or More Races	≥1,580	0.87	5.26	≥1,590	0.91	3.69
Education Classification						
Regular	≥49,220	0.87	5.16	≥49,110	0.90	3.68
Special	≥6,090	0.86	4.50	≥6,080	0.87	3.29
English Learner Status						
Non-EL	≥53,450	0.88	5.14	≥53,390	0.90	3.67
EL	≥1,860	0.83	4.78	≥1,800	0.88	3.37
Economic Status						
Economically Disadvantaged	≥39,100	0.87	5.05	≥39,000	0.89	3.57
Not Economically Disadvantaged	≥15,850	0.86	5.29	≥15,840	0.90	3.77
Section 504 Status						
Non-Section 504	≥50,100	0.88	5.14	≥49,990	0.90	3.67
Section 504	≥5,200	0.84	4.84	≥5,200	0.87	3.45

Table 10.8 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	≥52,810	0.91	5.14	≥52,670	0.92	3.83
Gender						
Female	≥25,710	0.91	5.18	≥25,630	0.92	3.89
Male	≥27,100	0.91	5.03	≥27,030	0.93	3.76
Ethnicity						
Hispanic/Latino	≥3,580	0.92	5.07	≥3,460	0.92	3.80
American Indian or Alaska Native	≥330	0.90	5.26	≥330	0.91	3.95
Asian	≥810	0.93	5.22	≥810	0.93	4.32
Black or African American	≥23,330	0.90	5.05	≥23,300	0.91	3.44
Native Hawaiian or Other Pacific	≥40	0.93	5.12	≥40	0.93	3.94
White	≥23,350	0.90	5.22	≥23,350	0.92	4.08
Two or More Races	≥1,330	0.90	5.21	≥1,330	0.92	3.92
Education Classification						
Regular	≥47,290	0.91	5.17	≥47,140	0.92	3.89
Special	≥5,520	0.88	4.53	≥5,520	0.90	2.96
English Learner Status						
Non-EL	≥51,420	0.91	5.15	≥51,390	0.92	3.84
EL	≥1,390	0.85	4.68	≥1,280	0.90	3.10
Economic Status						
Economically Disadvantaged	≥37,090	0.90	5.08	≥36,950	0.91	3.61
Not Economically Disadvantaged	≥15,380	0.90	5.23	≥15,380	0.91	4.18
Section 504 Status						
Non-Section 504	≥47,600	0.91	5.15	≥47,470	0.92	3.87
Section 504	≥5,210	0.88	4.93	≥5,200	0.90	3.37

Table 10.9 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	≥51,540	0.91	5.57	≥51,420	0.92	3.62
Gender						
Female	≥25,020	0.91	5.61	≥24,960	0.92	3.67
Male	≥26,510	0.91	5.41	≥26,450	0.92	3.56
Ethnicity						
Hispanic/Latino	≥3,370	0.93	5.50	≥3,260	0.92	3.60
American Indian or Alaska Native	≥350	0.90	5.70	≥350	0.90	3.59
Asian	≥790	0.93	5.54	≥790	0.93	4.27
Black or African American	≥22,550	0.90	5.52	≥22,550	0.90	3.22
Native Hawaiian or Other Pacific	≥20	0.92	6.00	≥20	0.93	3.85
White	≥23,300	0.90	5.64	≥23,300	0.92	3.87
Two or More Races	≥1,110	0.90	5.72	≥1,110	0.92	3.72
Education Classification						
Regular	≥46,390	0.91	5.58	≥46,270	0.92	3.68
Special	≥5,150	0.88	4.93	≥5,140	0.88	2.75
English Learner Status						
Non-EL	≥50,210	0.91	5.59	≥50,190	0.92	3.63
EL	≥1,330	0.86	4.97	≥1,220	0.88	3.01
Economic Status						
Economically Disadvantaged	≥35,500	0.90	5.54	≥35,380	0.91	3.38
Not Economically Disadvantaged	≥15,740	0.90	5.60	≥15,740	0.92	4.00
Section 504 Status						
Non-Section 504	≥46,380	0.91	5.58	≥46,260	0.92	3.65
Section 504	≥5,150	0.89	5.39	≥5,150	0.90	3.24

Table 10.10 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	≥51,020	0.89	5.68	≥44,910	0.92	3.67
Gender						
Female	≥25,190	0.88	5.63	≥21,810	0.91	3.75
Male	≥25,830	0.88	5.59	≥23,100	0.92	3.58
Ethnicity						
Hispanic/Latino	≥3,250	0.91	5.65	≥2,770	0.92	3.57
American Indian or Alaska Native	≥370	0.89	5.75	≥340	0.92	3.76
Asian	≥870	0.90	5.79	≥610	0.93	4.12
Black or African American	≥22,410	0.87	5.62	≥21,070	0.90	3.32
Native Hawaiian or Other Pacific	≥30	0.87	6.14	≥30	0.94	3.98
White	≥23,050	0.87	5.74	≥19,160	0.91	3.91
Two or More Races	≥1,000	0.88	5.73	≥890	0.92	3.79
Education Classification						
Regular	≥46,180	0.88	5.67	≥40,170	0.91	3.73
Special	≥4,840	0.85	5.19	≥4,730	0.88	2.82
English Learner Status						
Non-EL	≥49,690	0.88	5.68	≥43,710	0.92	3.68
EL	≥1,330	0.84	5.28	≥1,200	0.90	2.92
Economic Status						
Economically Disadvantaged	≥34,820	0.88	5.64	≥32,140	0.91	3.49
Not Economically Disadvantaged	≥15,880	0.86	5.71	≥12,480	0.92	3.98
Section 504 Status						
Non-Section 504	≥46,040	0.88	5.70	≥40,110	0.92	3.71
Section 504	≥4,980	0.87	5.48	≥4,800	0.90	3.27

10.3.2 **Effect Size**

One way to evaluate the magnitude of the standardized mean difference (SMD) is to calculate the ES. Cohen's *d* was used 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 10.11–10.18. 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-EL and EL students in EL status, and non-migrant and migrant students in migrant status.

Table 10.11 Impact Analysis, Grade 3 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	≥1,030	730.69	38.62		≥1,500	730.09	30.75	
Gender								
Male	≥520	725.86	37.70		≥530	727.63	31.13	
Female	≥510	735.71	38.96	-0.26	≥510	732.66	30.16	-0.16
Ethnicity								
White	≥420	739.84	38.02		≥420	737.95	30.21	
Hispanic/Latino	≥70	723.87	42.58	0.41	≥70	732.92	30.12	0.17
American Indian or Alaska Native	<10	NR	NR	NR	<10	NR	NR	NR
Asian	≥10	748.38	53.13	-0.22	≥10	750.06	33.39	-0.40
Black or African American	≥480	721.99	35.35	0.49	≥480	721.83	28.87	0.55
Native Hawaiian or Other Pacific	<10	NR	NR	NR	<10	NR	NR	NR
Two or More Races	≥30	745.91	41.69	-0.16	≥30	735.35	32.51	0.09
Education Classification								
Regular	≥950	732.78	38.24		≥950	732.01	30.13	
Special	≥80	708.11	35.61	0.65	≥90	710.72	30.38	0.71
Economic Status								
Not Economically Disadvantaged	≥210	747.47	36.73		≥210	743.03	32.26	
Economically Disadvantaged	≥810	726.22	37.76	0.57	≥820	726.69	29.34	0.55
English Learner Status								
Non-EL	≥1,000	731.22	38.50		≥1,000	730.08	30.83	
EL	≥30	717.05	39.61	0.37	≥30	730.41	28.84	-0.01
Migrant Status								
Nonmigrant	≥1,030	730.69	38.62		≥1,040	730.09	30.75	
Migrant	<10	NR	NR	NR	<10	NR	NR	NR
Section 504 Status								
Non-Section 504	≥990	731.33	38.81		≥1,000	730.78	30.93	
Section 504	≥40	716.58	31.37	0.38	≥40	714.78	21.53	0.52

Table 10.12 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	≥54,350	744.70	40.72		≥53,880	742.09	32.71	
Gender								
Male	≥27,740	738.95	39.70		≥27,390	740.99	33.53	
Female	≥26,590	750.72	40.89	-0.29	≥26,310	743.25	31.79	-0.07
Ethnicity								
White	≥23,370	757.66	38.27		≥23,200	752.69	30.89	
Hispanic/Latino	≥4,270	737.45	41.16	0.52	≥4,180	739.97	31.74	0.41
American Indian or Alaska Native	≥320	748.52	40.72	0.24	≥310	743.69	30.45	0.29
Asian	≥750	766.39	45.83	-0.23	≥740	767.30	34.27	-0.47
Black or African American	≥23,850	732.14	38.51	0.66	≥23,540	730.96	30.72	0.71
Native Hawaiian or Other Pacific	≥40	737.40	48.01	0.53	≥40	748.88	37.40	0.12
Two or More Races	≥1,680	751.54	39.13	0.16	≥1,670	745.48	31.24	0.23
Education Classification								
Regular	≥48,400	747.73	40.07		≥47,840	744.26	32.10	
Special	≥5,950	720.10	37.50	0.69	≥5,860	724.43	32.25	0.62
Economic Status								
Not Economically Disadvantaged	≥14,320	765.64	37.99		≥14,290	758.68	30.50	
Economically Disadvantaged	≥39,170	737.52	38.94	0.73	≥38,940	736.26	31.29	0.72
English Learner Status								
Non-EL	≥51,650	745.82	40.50		≥51,080	742.53	32.67	
EL	≥2,690	723.31	38.92	0.56	≥2,620	733.70	32.25	0.27
Migrant Status								
Nonmigrant	≥54,200	744.76	40.70		≥53,560	742.14	32.69	
Migrant	≥140	723.66	40.92	0.52	≥140	725.99	34.76	0.49
Section 504 Status								
Non-Section 504	≥49,540	746.38	40.78		≥48,970	743.36	32.74	
Section 504	≥4,800	727.42	35.74	0.47	≥4,730	728.95	29.32	0.44

Table 10.13 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	≥5,160	736.93	34.82		≥5,590	735.92	30.82	
Gender								
Male	≥2,580	734.19	34.83		≥2,580	735.33	31.66	
Female	≥2,570	739.68	34.60	-0.16	≥2,570	736.50	29.94	-0.04
Ethnicity								
White	≥2,180	747.80	33.54		≥2,180	745.16	30.71	
Hispanic/Latino	≥580	734.15	35.88	0.40	≥580	736.75	29.53	0.28
American Indian or Alaska Native	≥20	739.67	42.03	0.24	≥20	732.33	31.38	0.42
Asian	≥90	760.83	37.79	-0.39	≥90	765.47	32.91	-0.66
Black or African American	≥2,130	725.11	31.52	0.70	≥2,130	725.01	26.96	0.70
Native Hawaiian or Other Pacific	<10	NR	NR	NR	<10	NR	NR	NR
Two or More Races	≥130	744.37	30.42	0.10	≥130	737.12	30.48	0.26
Education Classification								
Regular	≥4,560	740.24	33.63		≥4,560	737.87	30.66	
Special	≥600	711.73	33.37	0.85	≥600	721.10	27.84	0.55
Economic Status								
Not Economically Disadvantaged	≥1,320	758.00	33.32		≥1,320	754.79	30.13	
Economically Disadvantaged	≥3,800	729.78	32.23	0.87	≥3,800	729.47	28.20	0.88
English Learner Status								
Non-EL	≥4,830	738.27	34.71		≥4,830	736.52	30.92	
EL	≥330	717.25	30.19	0.61	≥330	727.03	27.82	0.31
Migrant Status								
Nonmigrant	≥5,150	736.98	34.81		≥5,150	735.95	30.82	
Migrant	<10	NR	NR	NR	<10	NR	NR	NR
Section 504 Status								
Non-Section 504	≥4,780	738.13	35.11		≥4,770	736.93	31.10	
Section 504	≥380	721.86	26.71	0.47	≥380	723.32	23.71	0.44

Table 10.14 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	≥50,590	744.58	33.84		≥50,160	741.15	31.02	
Gender								
Male	≥25,790	741.40	33.34		≥25,490	740.58	31.50	
Female	≥24,750	747.95	34.00	-0.19	≥24,480	741.75	30.51	-0.04
Ethnicity								
White	≥21,800	755.77	31.37		≥21,650	751.71	29.29	
Hispanic/Latino	≥3,670	739.15	35.92	0.52	≥3,610	739.00	31.24	0.43
American Indian or Alaska Native	≥300	747.29	32.30	0.27	≥290	745.16	30.54	0.22
Asian	≥790	763.89	37.36	-0.26	≥790	767.33	32.84	-0.53
Black or African American	≥22,370	733.59	31.87	0.70	≥22,050	729.92	28.31	0.76
Native Hawaiian or Other Pacific	≥50	756.05	36.17	-0.01	≥50	756.15	34.30	-0.15
Two or More Races	≥1,520	749.09	31.59	0.21	≥1,500	743.99	29.77	0.26
Education Classification								
Regular	≥45,050	747.56	32.89		≥44,490	743.26	30.79	
Special	≥5,540	720.38	31.61	0.83	≥5,480	724.07	27.38	0.63
Economic Status								
Not Economically Disadvantaged	≥13,590	762.22	31.39		≥13,560	757.04	29.56	
Economically Disadvantaged	≥36,270	738.34	32.29	0.75	≥36,040	735.36	29.39	0.74
English Learner Status								
Non-EL	≥48,410	745.53	33.62		≥47,850	741.67	31.00	
EL	≥2,170	723.49	31.65	0.66	≥2,120	729.58	29.06	0.39
Migrant Status								
Nonmigrant	≥50,470	744.62	33.82		≥49,860	741.18	31.01	
Migrant	≥110	728.27	36.34	0.48	≥110	729.57	33.92	0.37
Section 504 Status								
Non-Section 504	≥45,280	746.37	33.85		≥44,750	742.70	31.04	
Section 504	≥5,300	729.36	29.67	0.51	≥5,220	727.93	27.52	0.48

Table 10.15 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	≥55,310	744.01	30.94		≥55,310	735.34	29.45	
Gender								
Male	≥28,140	739.94	30.56		≥28,100	734.22	30.41	
Female	≥27,170	748.23	30.76	-0.27	≥27,140	736.51	28.38	-0.08
Ethnicity								
White	≥24,560	753.17	28.95		≥24,550	744.36	29.02	
Hispanic/Latino	≥3,920	739.98	32.53	0.45	≥3,920	731.89	29.52	0.43
American Indian or Alaska Native	≥330	744.31	28.29	0.31	≥330	734.69	26.83	0.33
Asian	≥820	766.41	34.39	-0.45	≥820	763.92	33.57	-0.67
Black or African American	≥24,000	734.16	29.25	0.65	≥23,950	725.48	25.91	0.69
Native Hawaiian or Other Pacific	≥40	755.44	29.80	-0.08	≥40	745.17	31.60	-0.03
Two or More Races	≥1,580	749.43	30.11	0.13	≥1,580	738.49	30.16	0.20
Education Classification								
Regular	≥49,220	747.36	29.55		≥49,150	737.46	29.35	
Special	≥6,090	716.97	28.51	1.03	≥6,090	718.20	24.20	0.67
Economic Status								
Not Economically Disadvantaged	≥15,850	758.39	29.06		≥15,830	750.08	29.50	
Economically Disadvantaged	≥39,100	738.32	29.73	0.68	≥39,040	729.53	27.23	0.74
English Learner Status								
Non-EL	≥53,450	744.82	30.72		≥53,380	735.91	29.39	
EL	≥1,860	720.74	27.93	0.79	≥1,860	719.10	26.62	0.57
Migrant Status								
Nonmigrant	≥55,230	744.03	30.93		≥55,160	735.35	29.45	
Migrant	≥80	734.58	31.79	0.31	≥80	728.73	28.97	0.22
Section 504 Status								
Non-Section 504	≥50,100	745.52	30.93		≥50,040	736.57	29.65	
Section 504	≥5,200	729.48	26.92	0.52	≥5,200	723.54	24.53	0.45

Table 10.16 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	≥52,810	738.30	29.72		≥52,810	732.14	28.78	
Gender								
Male	≥27,100	732.88	28.56		≥27,080	730.17	29.37	
Female	≥25,710	744.02	29.85	-0.38	≥25,680	734.22	27.99	-0.14
Ethnicity								
White	≥23,350	748.17	28.35		≥23,350	741.90	27.34	
Hispanic/Latino	≥3,580	734.37	30.88	0.48	≥3,570	729.27	29.03	0.46
American Indian or Alaska Native	≥330	742.51	28.48	0.20	≥330	735.56	27.02	0.23
Asian	≥810	757.93	34.45	-0.34	≥810	758.54	33.74	-0.60
Black or African American	≥23,330	728.05	26.96	0.73	≥23,290	721.67	25.83	0.76
Native Hawaiian or Other Pacific	≥40	742.72	35.37	0.19	≥40	729.33	33.05	0.46
Two or More Races	≥1,330	742.69	28.35	0.19	≥1,330	735.67	28.13	0.23
Education Classification								
Regular	≥47,290	741.43	28.72		≥47,250	734.99	27.95	
Special	≥5,520	711.56	24.16	1.06	≥5,520	707.80	23.84	0.99
Economic Status								
Not Economically Disadvantaged	≥15,380	753.21	28.32		≥15,380	747.48	27.38	
Economically Disadvantaged	≥37,090	732.22	28.04	0.75	≥37,040	725.89	26.89	0.80
English Learner Status								
Non-EL	≥51,420	739.08	29.48		≥51,380	732.71	28.66	
EL	≥1,390	709.60	23.60	1.00	≥1,390	711.31	25.30	0.75
Migrant Status								
Nonmigrant	≥52,740	738.31	29.72		≥52,700	732.15	28.78	
Migrant	≥70	733.27	28.67	0.17	≥70	728.65	28.34	0.12
Section 504 Status								
Non-Section 504	≥47,600	739.85	29.77		≥47,570	733.48	28.88	
Section 504	≥5,210	724.15	25.15	0.54	≥5,200	719.96	24.68	0.47

Table 10.17 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	≥51,540	743.60	35.66		≥51,540	732.50	25.79	
Gender								
Male	≥26,510	737.15	35.03		≥26,500	731.25	26.29	
Female	≥25,020	750.44	35.05	-0.38	≥25,010	733.82	25.19	-0.10
Ethnicity								
White	≥23,300	754.38	33.76		≥23,300	741.10	25.10	
Hispanic/Latino	≥3,370	736.35	39.76	0.52	≥3,370	730.04	26.12	0.44
American Indian or Alaska Native	≥350	743.78	32.72	0.31	≥350	731.09	22.98	0.40
Asian	≥790	769.35	41.45	-0.44	≥790	758.08	30.66	-0.67
Black or African American	≥22,550	732.31	32.79	0.66	≥22,540	722.89	22.36	0.77
Native Hawaiian or Other Pacific	≥20	745.09	42.62	0.27	≥20	738.91	29.97	0.09
Two or More Races	≥1,110	750.64	33.73	0.11	≥1,110	736.95	24.83	0.17
Education Classification								
Regular	≥46,390	747.60	33.99		≥46,370	735.15	24.91	
Special	≥5,150	707.60	29.64	1.19	≥5,150	708.66	20.87	1.08
Economic Status								
Not Economically Disadvantaged	≥15,740	761.71	33.45		≥15,740	746.19	25.36	
Economically Disadvantaged	≥35,500	735.72	33.58	0.77	≥35,480	726.53	23.57	0.81
English Learner Status								
Non-EL	≥50,210	744.67	35.19		≥50,190	733.00	25.70	
EL	≥1,330	703.16	29.35	1.18	≥1,330	713.46	21.59	0.76
Migrant Status								
Nonmigrant	≥51,460	743.62	35.66		≥51,440	732.50	25.80	
Migrant	≥70	732.74	33.64	0.30	≥70	727.40	21.56	0.20
Section 504 Status								
Non-Section 504	≥46,380	745.49	35.71		≥46,360	733.76	25.83	
Section 504	≥5,150	726.57	30.31	0.54	≥5,150	721.14	22.41	0.49

Table 10.18 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	≥51,020	744.67	35.29		≥51,020	728.21	33.98	
Gender								
Male	≥25,830	736.81	34.64		≥23,170	725.30	34.28	
Female	≥25,190	752.72	34.13	-0.46	≥21,870	731.30	33.38	-0.18
Ethnicity								
White	≥23,050	755.33	33.20		≥19,160	739.42	32.90	
Hispanic/Latino	≥3,250	734.72	40.65	0.60	≥2,900	723.00	34.26	0.50
American Indian or Alaska Native	≥370	747.56	35.73	0.23	≥340	732.48	34.32	0.21
Asian	≥870	767.69	39.87	-0.37	≥610	755.95	39.66	-0.50
Black or African American	≥22,410	733.86	32.43	0.65	≥21,070	717.58	30.82	0.69
Native Hawaiian or Other Pacific	≥30	756.97	36.27	-0.05	≥30	743.50	39.61	-0.12
Two or More Races	≥1,000	751.89	34.37	0.10	≥890	733.74	33.55	0.17
Education Classification								
Regular	≥46,180	748.40	33.65		≥40,290	731.54	33.00	
Special	≥4,840	709.08	30.41	1.18	≥4,740	700.00	28.66	0.97
Economic Status								
Not Economically Disadvantaged	≥15,880	761.75	32.41		≥12,490	744.68	33.10	
Economically Disadvantaged	≥34,820	737.07	33.73	0.74	≥32,240	721.99	32.08	0.70
English Learner Status								
Non-EL	≥49,690	745.86	34.63		≥43,720	728.96	33.81	
EL	≥1,330	700.49	31.27	1.31	≥1,320	703.38	29.95	0.76
Migrant Status								
Nonmigrant	≥50,950	744.69	35.28		≥44,970	728.22	33.97	
Migrant	≥70	726.07	41.83	0.53	≥60	719.81	36.45	0.25
Section 504 Status								
Non-Section 504	≥46,040	746.39	35.21		≥40,240	729.71	34.05	
Section 504	≥4,980	728.71	31.88	0.51	≥4,790	715.63	30.59	0.42

Additional data for mean scale scores are provided in Tables 10.19 and 10.20. 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 10.19 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	≥980	806.86	27.76	≥54,400	743.31	40.03
	Talented	≥560	782.39	33.89	≥54,820	744.05	40.60
	Autism	≥280	710.88	39.38	≥55,100	744.61	40.66
	Deaf-Blindness	<50	NR	NR	≥55,390	744.44	40.72
	Developmental Delay	≥370	710.06	32.95	≥55,020	744.67	40.67
	Emotional Disturbance	≥100	712.32	37.75	≥55,280	744.50	40.70
	HI—Deaf	<50	NR	NR	≥55,360	744.45	40.72
	HI—Hard-of-Hearing	≥50	716.00	40.37	≥55,330	744.47	40.71
	Mild Mental Disability	≥350	688.69	23.88	≥55,030	744.80	40.56
	Moderate Mental Disability	<50	NR	NR	≥55,370	744.45	40.71
	Orthopedic Impairment	≥50	731.02	38.79	≥55,330	744.45	40.72
	Other Health Impairment	≥870	712.69	34.34	≥54,510	744.95	40.62
	Specific Learning Disability	≥2,190	713.80	31.03	≥53,190	745.70	40.58
	Speech or Language Impairment	≥1,670	742.66	38.39	≥53,710	744.49	40.79
	Traumatic Brain Injury	<50	NR	NR	≥55,380	744.44	40.72
	Visual Impairment	<50	NR	NR	≥55,340	744.45	40.73
	Other	<50	NR	NR	≥55,380	744.44	40.72
	HI—Hearing Impairment	<50	NR	NR	≥55,390	744.44	40.72
	Unknown	<50	NR	NR	≥55,390	744.44	40.72
4	Gifted	≥1,230	794.05	22.76	≥54,520	742.74	33.35
	Talented	≥850	774.18	28.24	≥54,910	743.40	33.87
	Autism	≥310	719.01	36.64	≥55,440	744.01	33.93
	Deaf-Blindness	<50	NR	NR	≥55,760	743.87	34.00
	Developmental Delay	<50	NR	NR	≥55,760	743.87	34.00
	Emotional Disturbance	≥130	712.53	28.32	≥55,620	743.95	33.98
	HI—Deaf	<50	NR	NR	≥55,740	743.89	33.99
	HI—Hard-of-Hearing	≥50	720.14	32.50	≥55,700	743.90	33.99
	Mild Mental Disability	≥360	692.96	20.23	≥55,390	744.21	33.82
	Moderate Mental Disability	<50	NR	NR	≥55,750	743.88	34.00
	Orthopedic Impairment	≥70	733.94	39.58	≥55,680	743.89	33.99
	Other Health Impairment	≥1,090	718.13	30.29	≥54,670	744.39	33.87
	Specific Learning Disability	≥2,690	713.67	26.77	≥53,070	745.40	33.61
	Speech or Language Impairment	≥1,340	739.87	32.64	≥54,410	743.97	34.03
	Traumatic Brain Injury	<50	NR	NR	≥55,750	743.88	34.00
	Visual Impairment	<50	NR	NR	≥55,720	743.88	34.00
	Other	<50	NR	NR	≥55,750	743.88	34.00
	HI—Hearing Impairment	<50	NR	NR	≥55,760	743.87	34.00
	Unknown	<50	NR	NR	≥55,760	743.87	34.00

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	≥1,500	788.96	22.92	≥53,810	742.76	30.19
	Talented	≥1,210	767.86	26.36	≥54,100	743.48	30.82
	Autism	≥290	718.73	29.27	≥55,020	744.15	30.89
	Deaf-Blindness	<50	NR	NR	≥55,310	744.01	30.94
	Developmental Delay	<50	NR	NR	≥55,310	744.01	30.94
	Emotional Disturbance	≥160	715.97	29.26	≥55,150	744.10	30.90
	HI—Deaf	<50	NR	NR	≥55,300	744.03	30.93
	HI—Hard-of-Hearing	≥60	725.85	30.41	≥55,240	744.03	30.93
	Mild Mental Disability	≥350	693.55	17.27	≥54,960	744.34	30.74
	Moderate Mental Disability	<50	NR	NR	≥55,310	744.02	30.93
	Orthopedic Impairment	≥50	731.66	26.69	≥55,260	744.02	30.94
	Other Health Impairment	≥1,120	716.46	26.13	≥54,190	744.58	30.77
	Specific Learning Disability	≥2,960	711.12	23.22	≥52,350	745.87	30.26
	Speech or Language Impairment	≥1,020	740.76	31.96	≥54,280	744.07	30.91
	Traumatic Brain Injury	<50	NR	NR	≥55,300	744.02	30.93
	Visual Impairment	<50	NR	NR	≥55,280	744.02	30.94
	Other	<50	NR	NR	≥55,310	744.01	30.93
	HI—Hearing Impairment	<50	NR	NR	≥55,310	744.01	30.94
	Unknown	<50	NR	NR	≥55,310	744.01	30.94
6	Gifted	≥1,550	782.28	22.55	≥51,260	736.97	28.89
	Talented	≥1,390	760.98	26.13	≥51,420	737.69	29.57
	Autism	≥230	715.06	27.53	≥52,580	738.41	29.69
	Deaf-Blindness	<50	NR	NR	≥52,810	738.30	29.72
	Developmental Delay	<50	NR	NR	≥52,810	738.30	29.72
	Emotional Disturbance	≥190	710.39	24.79	≥52,620	738.40	29.69
	HI—Deaf	<50	NR	NR	≥52,790	738.31	29.72
	HI—Hard-of-Hearing	≥60	719.57	24.31	≥52,750	738.32	29.72
	Mild Mental Disability	≥220	691.46	14.84	≥52,590	738.50	29.61
	Moderate Mental Disability	<50	NR	NR	≥52,810	738.30	29.72
	Orthopedic Impairment	≥60	732.41	29.07	≥52,750	738.31	29.72
	Other Health Impairment	≥1,120	711.81	23.45	≥51,690	738.88	29.58
	Specific Learning Disability	≥2,940	707.58	19.98	≥49,870	740.11	29.20
	Speech or Language Impairment	≥630	732.48	28.33	≥52,180	738.37	29.73
	Traumatic Brain Injury	<50	NR	NR	≥52,800	738.31	29.71
	Visual Impairment	<50	NR	NR	≥52,790	738.31	29.72
	Other	<50	NR	NR	≥52,810	738.30	29.72
	HI—Hearing Impairment	<50	NR	NR	≥52,810	738.30	29.72
	Unknown	<50	NR	NR	≥52,810	738.30	29.72

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	≥1,590	797.42	27.09	≥49,950	741.88	34.55
	Talented	≥1,540	770.75	30.89	≥50,000	742.76	35.47
	Autism	≥250	715.19	36.69	≥51,280	743.74	35.60
	Deaf-Blindness	<50	NR	NR	≥51,540	743.60	35.66
	Developmental Delay	<50	NR	NR	≥51,540	743.60	35.66
	Emotional Disturbance	≥190	711.01	33.15	≥51,340	743.72	35.62
	HI—Deaf	<50	NR	NR	≥51,520	743.61	35.66
	HI—Hard-of-Hearing	≥50	731.91	31.95	≥51,490	743.61	35.67
	Mild Mental Disability	≥220	681.69	18.39	≥51,310	743.87	35.48
	Moderate Mental Disability	<50	NR	NR	≥51,540	743.60	35.66
	Orthopedic Impairment	≥50	725.67	37.30	≥51,490	743.62	35.66
	Other Health Impairment	≥1,080	707.58	29.31	≥50,450	744.38	35.39
	Specific Learning Disability	≥2,800	703.81	24.95	≥48,730	745.89	34.83
	Speech or Language Impairment	≥420	733.46	34.03	≥51,120	743.68	35.67
	Traumatic Brain Injury	<50	NR	NR	≥51,530	743.61	35.66
	Visual Impairment	<50	NR	NR	≥51,510	743.61	35.67
	Other	<50	NR	NR	≥51,540	743.60	35.66
	HI—Hearing Impairment	<50	NR	NR	≥51,540	743.60	35.66
	Unknown	<50	NR	NR	≥51,540	743.60	35.66
8	Gifted	≥1,640	793.41	25.16	≥49,380	743.04	34.41
	Talented	≥1,660	770.99	29.66	≥49,350	743.78	35.13
	Autism	≥220	717.19	35.12	≥50,800	744.79	35.25
	Deaf-Blindness	<50	NR	NR	≥51,020	744.67	35.29
	Developmental Delay	<50	NR	NR	≥51,020	744.67	35.29
	Emotional Disturbance	≥180	705.03	34.66	≥50,840	744.81	35.22
	HI—Deaf	<50	NR	NR	≥51,000	744.68	35.28
	HI—Hard-of-Hearing	≥50	728.86	36.84	≥50,960	744.68	35.29
	Mild Mental Disability	≥210	683.77	22.32	≥50,810	744.92	35.12
	Moderate Mental Disability	<50	NR	NR	≥51,020	744.67	35.29
	Orthopedic Impairment	≥50	742.47	38.67	≥50,970	744.67	35.29
	Other Health Impairment	≥1,020	709.38	29.99	≥50,000	745.39	35.03
	Specific Learning Disability	≥2,720	706.72	26.74	≥48,300	746.80	34.49
	Speech or Language Impairment	≥280	730.39	34.77	≥50,740	744.75	35.28
	Traumatic Brain Injury	<50	NR	NR	≥51,010	744.67	35.29
	Visual Impairment	<50	NR	NR	≥50,980	744.67	35.30
	Other	<50	NR	NR	≥51,020	744.67	35.29
	HI—Hearing Impairment	<50	NR	NR	≥51,020	744.67	35.29
	Unknown	<50	NR	NR	≥51,020	744.67	35.29

Table 10.20 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	≥980	792.68	22.34	≥54,370	740.75	32.20
	Talented	≥560	764.83	27.40	≥54,800	741.44	32.75
	Autism	≥280	722.04	37.69	≥55,080	741.77	32.73
	Deaf-Blindness	<50	NR	NR	≥55,360	741.67	32.78
	Developmental Delay	≥360	718.01	31.69	≥54,990	741.83	32.73
	Emotional Disturbance	≥100	717.55	35.68	≥55,260	741.72	32.76
	HI—Deaf	<50	NR	NR	≥55,340	741.68	32.78
	HI—Hard-of-Hearing	≥50	731.39	32.01	≥55,300	741.68	32.78
	Mild Mental Disability	≥350	694.24	26.59	≥55,000	741.98	32.59
	Moderate Mental Disability	<50	NR	NR	≥55,350	741.69	32.77
	Orthopedic Impairment	≥50	729.19	32.79	≥55,310	741.69	32.78
	Other Health Impairment	≥870	717.55	30.83	≥54,480	742.06	32.67
	Specific Learning Disability	≥2,190	718.94	25.17	≥53,160	742.61	32.72
	Speech or Language Impairment	≥1,670	742.57	31.73	≥53,690	741.65	32.82
	Traumatic Brain Injury	<50	NR	NR	≥55,360	741.68	32.78
	Visual Impairment	<50	NR	NR	≥55,320	741.68	32.78
	Other	<50	NR	NR	≥55,360	741.67	32.78
	HI—Hearing Impairment	<50	NR	NR	≥55,360	741.67	32.78
	Unknown	<50	NR	NR	≥55,360	741.67	32.78
4	Gifted	≥1,230	789.95	22.95	≥54,450	739.36	30.30
	Talented	≥850	763.05	26.33	≥54,830	740.13	31.00
	Autism	≥310	726.61	32.42	≥55,370	740.56	31.04
	Deaf-Blindness	<50	NR	NR	≥55,680	740.48	31.06
	Developmental Delay	<50	NR	NR	≥55,680	740.48	31.06
	Emotional Disturbance	≥130	712.04	26.64	≥55,550	740.55	31.04
	HI—Deaf	<50	NR	NR	≥55,660	740.49	31.07
	HI—Hard-of-Hearing	≥50	733.30	28.95	≥55,630	740.49	31.07
	Mild Mental Disability	≥370	702.99	20.86	≥55,310	740.73	30.97
	Moderate Mental Disability	<50	NR	NR	≥55,670	740.49	31.06
	Orthopedic Impairment	≥70	732.09	30.23	≥55,610	740.49	31.06
	Other Health Impairment	≥1,090	721.40	25.27	≥54,590	740.86	31.05
	Specific Learning Disability	≥2,690	719.03	21.86	≥52,990	741.57	31.07
	Speech or Language Impairment	≥1,340	739.68	31.13	≥54,340	740.50	31.06
	Traumatic Brain Injury	<50	NR	NR	≥55,670	740.48	31.06
	Visual Impairment	<50	NR	NR	≥55,650	740.48	31.06
	Other	<50	NR	NR	≥55,680	740.48	31.06
	HI—Hearing Impairment	<50	NR	NR	≥55,680	740.48	31.06
	Unknown	<50	NR	NR	≥55,680	740.48	31.06

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	≥1,500	785.47	25.91	≥53,690	733.98	28.27
	Talented	≥1,210	754.10	26.65	≥53,980	734.96	29.35
	Autism	≥290	723.83	28.94	≥54,900	735.45	29.42
	Deaf-Blindness	<50	NR	NR	≥55,200	735.38	29.43
	Developmental Delay	<50	NR	NR	≥55,190	735.38	29.43
	Emotional Disturbance	≥160	711.52	24.53	≥55,030	735.45	29.41
	HI—Deaf	<50	NR	NR	≥55,180	735.39	29.43
	HI—Hard-of-Hearing	≥60	725.68	26.62	≥55,130	735.40	29.43
	Mild Mental Disability	≥350	702.78	18.13	≥54,840	735.59	29.37
	Moderate Mental Disability	<50	NR	NR	≥55,190	735.39	29.43
	Orthopedic Impairment	≥50	727.75	27.80	≥55,140	735.39	29.43
	Other Health Impairment	≥1,120	715.93	22.48	≥54,070	735.79	29.42
	Specific Learning Disability	≥2,950	714.66	18.35	≥52,240	736.56	29.50
	Speech or Language Impairment	≥1,020	734.51	31.03	≥54,170	735.40	29.40
	Traumatic Brain Injury	<50	NR	NR	≥55,190	735.39	29.43
	Visual Impairment	<50	NR	NR	≥55,160	735.39	29.43
	Other	<50	NR	NR	≥55,190	735.39	29.43
	HI—Hearing Impairment	<50	NR	NR	≥55,200	735.38	29.43
	Unknown	<50	NR	NR	≥55,200	735.38	29.43
6	Gifted	≥1,550	780.07	21.85	≥51,110	730.78	27.66
	Talented	≥1,390	751.36	24.83	≥51,270	731.71	28.66
	Autism	≥230	714.25	31.20	≥52,430	732.31	28.70
	Deaf-Blindness	<50	NR	NR	≥52,660	732.23	28.74
	Developmental Delay	<50	NR	NR	≥52,670	732.23	28.74
	Emotional Disturbance	≥190	705.39	23.45	≥52,470	732.32	28.71
	HI—Deaf	<50	NR	NR	≥52,640	732.24	28.73
	HI—Hard-of-Hearing	≥60	717.08	24.31	≥52,600	732.24	28.74
	Mild Mental Disability	≥220	687.11	17.02	≥52,440	732.42	28.63
	Moderate Mental Disability	<50	NR	NR	≥52,660	732.23	28.74
	Orthopedic Impairment	≥60	724.17	29.03	≥52,600	732.24	28.74
	Other Health Impairment	≥1,120	707.67	23.42	≥51,540	732.76	28.61
	Specific Learning Disability	≥2,940	703.90	18.91	≥49,720	733.90	28.34
	Speech or Language Impairment	≥620	728.97	27.85	≥52,040	732.27	28.75
	Traumatic Brain Injury	<50	NR	NR	≥52,650	732.23	28.74
	Visual Impairment	<50	NR	NR	≥52,640	732.23	28.74
	Other	<50	NR	NR	≥52,660	732.23	28.74
	HI—Hearing Impairment	<50	NR	NR	≥52,670	732.23	28.74
	Unknown	<50	NR	NR	≥52,670	732.23	28.74

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	≥1,590	775.94	21.20	≥49,820	731.20	24.65
	Talented	≥1,540	749.07	23.64	≥49,880	732.08	25.64
	Autism	≥250	715.59	25.34	≥51,160	732.67	25.72
	Deaf-Blindness	<50	NR	NR	≥51,420	732.59	25.74
	Developmental Delay	<50	NR	NR	≥51,420	732.59	25.74
	Emotional Disturbance	≥190	708.97	25.26	≥51,220	732.67	25.70
	HI—Deaf	<50	NR	NR	≥51,400	732.59	25.74
	HI—Hard-of-Hearing	≥50	725.02	23.95	≥51,360	732.59	25.74
	Mild Mental Disability	≥220	692.82	13.20	≥51,190	732.76	25.65
	Moderate Mental Disability	<50	NR	NR	≥51,410	732.59	25.74
	Orthopedic Impairment	≥50	720.23	25.45	≥51,360	732.60	25.74
	Other Health Impairment	≥1,080	708.88	20.68	≥50,330	733.10	25.60
	Specific Learning Disability	≥2,800	706.04	17.26	≥48,610	734.12	25.31
	Speech or Language Impairment	≥420	726.00	25.38	≥50,990	732.64	25.74
	Traumatic Brain Injury	<50	NR	NR	≥51,400	732.59	25.74
	Visual Impairment	<50	NR	NR	≥51,380	732.59	25.74
	Other	<50	NR	NR	≥51,410	732.58	25.74
	HI—Hearing Impairment	<50	NR	NR	≥51,420	732.59	25.74
	Unknown	<50	NR	NR	≥51,420	732.59	25.74
8	Gifted	≥740	784.52	27.15	≥44,170	727.39	33.23
	Talented	≥1,220	750.28	30.50	≥43,680	727.72	33.82
	Autism	≥210	712.03	34.55	≥44,700	728.42	33.91
	Deaf-Blindness	<50	NR	NR	≥44,910	728.34	33.94
	Developmental Delay	<50	NR	NR	≥44,910	728.34	33.94
	Emotional Disturbance	≥170	698.80	33.62	≥44,730	728.46	33.89
	HI—Deaf	<50	NR	NR	≥44,890	728.34	33.93
	HI—Hard-of-Hearing	≥50	719.02	32.35	≥44,860	728.35	33.94
	Mild Mental Disability	≥200	680.58	16.94	≥44,710	728.56	33.84
	Moderate Mental Disability	<50	NR	NR	≥44,910	728.34	33.93
	Orthopedic Impairment	≥50	720.18	36.47	≥44,860	728.35	33.93
	Other Health Impairment	≥1,000	700.50	28.40	≥43,910	728.97	33.79
	Specific Learning Disability	≥2,680	697.41	25.47	≥42,230	730.31	33.45
	Speech or Language Impairment	≥20	718.25	34.03	≥44,640	728.40	33.93
	Traumatic Brain Injury	<50	NR	NR	≥44,900	728.35	33.93
	Visual Impairment	<50	NR	NR	≥44,870	728.34	33.93
	Other	<50	NR	NR	≥44,910	728.34	33.93
	HI—Hearing Impairment	<50	NR	NR	≥44,910	728.34	33.94
	Unknown	<50	NR	NR	≥44,910	728.34	33.94

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 2018 LEAP 2025 ELA and mathematics tests were administered as both paper-based tests (PBTs) and computer-based tests (CBTs) for grades 3 and 4. The *Standards* indicate that results across different testing modes should be comparable. The mode comparability for the 2018 LEAP 2025 CBT and PBT in grades 3 and 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.
- Similar to PARCC's decision to not apply a mode adjustment, the LDOE also decided to not apply any mode adjustment to the LEAP 2025.

10.4.1 **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 the reference group. That is, all CBT students and their matching PBT students were selected using covariates (matching variables), such as the 2017 LEAP 2025 ELA and mathematics scale scores and the 2018 bio-demographic information, such as gender, ethnicity, economically disadvantaged, and EL. The 2017 LEAP 2025 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 2017 PBT were used in this study. There is no LEAP 2025 grade 2 test. Therefore, school means for the 2017 LEAP grade 3 were used to match with 2018 LEAP 2025 grade 3. All grade 4 students who took the 2018 LEAP 2025 CBT were included, and a sample of matching PBT students was drawn using the R package, MatchIt for PSM.

Table 10.21 shows the number of equivalent CBT and PBT students matched by the PSM method. Grade 3 had a small number of CBT students, making its matching PBT student count small. For mathematics grade 4, there were 4,827 CBT students and 47,036 PBT students who had all PSM covariate information, such as bio-demographics, 2017 ELA and mathematics performance information, and 2018 mode information. Of the 47,036 PBT students, 4,827 were selected (a number equivalent to the number of CBT students) by considering all covariates.

Table 10.21 Number of Students Used for Propensity Score Matching

Content	Grade	CBT	PBT	
		Total	Total	Selected
Mathematics	3	≥930	≥51,980	931
Mathematics	4	≥4,820	≥47,030	4,827
ELA	3	≥930	≥51,840	935
ELA	4	≥4,830	≥46,910	4,832

**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 10.22 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 in grade 3 and one B+ flag item for grade 4, and for ELA, grade 3 had one B- flag and two B+ flags while grade 4 had one B+ flag. The negative sign indicates the CBT item was more difficult than the same PBT item.

Table 10.22 2018 LEAP 2025 Mode DIF Statistics: Number of Flagged Items

			DIF			
Content	Grade	N of Items	C-	C+	B-	B+
Mathematics	3	43		2		
Mathematics	4	43				1
ELA	3	22			1	2
ELA	4	23				1

Scale scores of the CBT and PBT administrations 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 10.23 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 3 was flagged. Factors considered when making the decision to apply a mode adjustment or not included that the student count available for grade 3 made the results somewhat unreliable and that a consistent pattern of which mode performed better was not established.

Table 10.23 Mode Study Scale Score Differences and Effect Size

			PBT		CBT				
Content	Grade	N of Students	Mean	Std. Dev.	Mean	Std. Dev.	Mean Diff PBS-CBT	ES	Flag>|0.2|
Mathematics	3*	≥930	753.5	37.5	741.1	35.4	12.4	0.33	YES
Mathematics	4	≥4,830	737.0	25.7	734.0	25.9	3	0.12	NO
ELA	3*	≥930	735.4	39.4	734.0	37.6	1.4	0.04	NO
ELA	4	≥4,820	738.8	33.8	739.2	34.4	-0.4	-0.01	NO

**Less reliable due to small sample size; using 2017 school scale score mean as independent variable*

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 assessments. The information in this chapter addresses multiple best practices of the testing industry and is particularly related to the following standards:

> ***Standard 3.1*** Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (63)
>
> ***Standard 3.2*** Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (64)
>
> ***Standard 3.3*** Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test. (64)
>
> ***Standard 3.4*** Test takers should receive comparable treatment during the test administration and scoring process. (65)
>
> ***Standard 3.5*** Test developers should specify and document provisions that have been made to test administration and scoring procedures to remove construct-irrelevant barriers for all relevant subgroups in the test-taker population. (65)
>
> ***Standard 3.6*** Where credible evidence indicates that test scores may differ in meaning for relevant subgroups in the intended examinee population, test developers and/or users are responsible for examining the evidence for validity of score interpretations for intended uses for individuals from those subgroups. What constitutes a significant difference in

subgroup scores and what actions are taken in response to such differences may be defined by applicable laws. (65)

Standard 3.16 When credible research indicates that test scores for some relevant subgroups are differentially affected by construct-irrelevant characteristics of the test or of the examinees, when legally permissible, test users should use the test only for those subgroups for which there is sufficient evidence of validity to support score interpretations for the intended uses. (70)

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.

Lu, Y., & Sireci, S. G., (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice, 26*(40), 29-37.

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. Retrieved 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.

APPENDIX A—FORMS DEVELOPMENT PROCESS FOR SPANISH TRANSADAPTATION FORMS

For English Learners, the LDOE offers the mathematics assessments in Spanish for both computer-based tests (CBT) in all grades and paper-based tests (PBT) in grades 3 and 4 only (to mirror the English language forms), as well as text-to-speech (TTS) for CBT and large print and human voice audio CDs for PBT forms. The Spanish versions of the test were developed through transadaptation. Transadaptation takes into consideration grade-level appropriateness of the words and sentence structures used, as well as the linguistic and cultural differences that exist between speakers of two different languages. Accounting for these differences allows experts to ensure that the item will measure the same construct as the English-language version of the item at the same level of rigor and therefore be expected to measure the achievement of English learners in the same way that the English version of the item does for native speakers of English.

Once the operational form was approved in English, DRC provided item IDs for acquired items to New Meridian, who then identified which of those items had previously appeared on a Spanish transadapted form. Once New Meridian identified the items that had previously been transadapted and provided the transadaptations of those items, DRC identified the English version of all items that had not been previously transadapted (either because they were Louisiana-owned items that would appear in field-test positions or because they were acquired items that had not been previously used on a Spanish-language form by PARCC). These items were then provided to the Spanish transadaptation subcontractor for initial transadaptation. DRC's Spanish Test Development Team reviewed the previously transadapted items to ensure consistency between those items transadapted as part of the PARCC assessments and those transadapted specifically for Louisiana and provided guidance to the translator conducting the initial transadaptation in grade-level and culturally appropriate ways. Upon completion of the transadaptation by the subcontractor, DRC's Spanish Test Development team conducted reviews by native Spanish speakers for content and grade-level appropriateness of the transadaptation as well as an editorial review. At least two members of DRC's Spanish Test Development team compared each English item to the Spanish transadaptation to ensure that the transadaptation:

- was accurate;
- contained grade-appropriate wording;
- contained answer choices that were reasonably parallel;
- had not introduced ambiguity into the Spanish version;
- contained graphics that were clearly transadapted;
- did not alter current teaching and learning practices in the content area; and
- remained free of gender, ethnic, cultural, socioeconomic, and regional bias.

The Spanish Test Development team then reconciled any discrepancies and submitted the transadaptations to a senior Spanish Test Development team member for resolution. After

approval by the senior Spanish Test Development team member, the item moved forward to be imported into DRC's item banking system.
Both previously-transadapted items and newly-transadapted items were imported into DRC's item banking system and formatted for online and print (in grades 3 and 4) use. The Spanish item was associated with the corresponding English item in the item bank, and the item was formatted and graphics finalized and review. The final approved transadaptation was then compared to the Spanish item in the DRC assessment system as well as the English version of the item and all changes were verified.

DRC's Spanish Test Development team then used the final, approved Communication Assistance Scripts in English to transadapt descriptions of graphics as necessary. These graphic descriptions were used when preparing the TTS for review and inclusion in the test forms and as the scripts for the recording of human voice audio CDs for PBT forms. Scripting of TTS and reviews of Spanish forms in their final forms was conducted by native Spanish speakers at DRC prior to their submission to LDOE for a translation review by third-party translation vendor.

The vendor reviewed the transadapted forms and provided feedback to LDOE and DRC. Experienced DRC Spanish team members (who are all native Spanish speakers) and the translation vendor resolved any issues and DRC made modifications as necessary. The form was then approved by both DRC and the LDOE translation vendor.

APPENDIX B—ACCOMMODATED PRINT FORM CREATION

Guidelines for Building Accommodated Print Forms

Careful consideration is given to all items selected for a test form, which are to be used for accommodated print (AP) forms and/or brailled. Fairness for all populations, item integrity, and student-item interaction for technology-enhanced (TE) items are factors when selecting items that will appear on an AP form.

TE items used for AP are modified in such a way as to allow the student to interact with the item in a way similar to the online interaction, thereby maintaining both the rigor and the content being assessed. For example:

- Drag and Drop items in the online environment require a student to place the answer options in an interactive table. For the AP form, the student is presented with a table with the same information (column or row headers, any completed cells, and blank spaces) and the answer options are listed below the table (similar to the online form in which the options are listed either below or to the right of the table). The directions are then modified to ask the student to write the information in the table rather than drag and drop the information into the table. Students are also able to circle and draw arrows to where the text should be placed or to add labels to the answer choices and write only the label in the box, as long as their intended response is clear to the test administrator who will transcribe the answers into the online system.
- Match Interaction Table items require a student to select a checkbox in one or more columns for each of multiple rows. In the AP form, the student is provided with a table and asked to mark an X in the correct places.
- Highlight-text items or item parts in the online environment require clicking on the selected text, whereby the selected answer (a word, phrase, or sentence) is highlighted. In the AP form, the text is presented in the same format and the student is asked to circle the answer. Where only certain words or phrases are selectable in the online system, those options are underlined in the AP form to indicate which words and/or phrases the student should select from.
- Drop-down Menu items in the online environment have answer options in a drop-down menu format, oftentimes as part of a complete sentence. The AP form displays the item with a blank line in place of the drop-down menu in the sentence, with all the answer options for the drop-down presented vertically below the sentence. The directions are then modified to ask the student to circle the word/phrase that belongs in the blank.
- Short Answer Items in the online environment allow students to type their answer in a box. In the AP form, a box is provided for the student to write the response.
- Keypad Input Items in the online system allow a student to enter a numeric response including all rational and irrational numbers as well as expressions and equations. In the AP form, a box is provided for the student to write the response.
- Graphing Items, including coordinate planes, number lines, line plots, and bar graphs, in the online system require students to complete a graph by plotting points, adding Xs to create a line plot, or raising/lowering bars to create a bar

graph or histogram. In the AP form, the student is provided with the same coordinate plane, number line, line plot, or bar graph in the online item, including titles, axes labels, and keys, and are asked to complete the graph.

Displaying items similarly in both print and online, and allowing the student to interact with the item in a similar manner, maintain the item integrity by assessing a similar construct in a similar manner, providing students who are unable to access the assessment online with an assessment at the same level of rigor as the online test.

The AP forms are reviewed by DRC and LDOE content experts who review the materials thoroughly to ensure a valid and reliable assessment for students who are unable to participate in the online assessment. These forms are also used as the source files for the creation of braille forms for students in grades 5–8 in ELA and mathematics.

APPENDIX C—ITEM ALIGNMENT REVIEW PROCESS

September 2017 Item Alignment Review Process

1. Committee members attended a general session and received an overview of the process, which included:
    a. Background on items and the review process that took place prior to items being field-tested on forms
    b. Overview of purpose of the alignment review:
        i. To confirm content alignment of each item to designated Louisiana State Standard
        ii. To confirm grade appropriateness of each item
        iii. To confirm items were free of issues of bias, fairness, or sensitivity that could impact student responses to item
    c. Overview of process Itself:
        i. Committee members reviewed each item individually and decided status of each item
            1. Status could be either "accepted with current alignment," "accepted with realignment," or "rejected"
        ii. Group then came to consensus regarding status of each item
        iii. Items that were accepted were determined to appropriately measure the intended standard and were free of issues of bias, fairness, or sensitivity that could impact student responses to item
2. After the general session, committee members broke out into assigned groups.
3. The facilitator went over the review process again by first walking the group through the use of the review spreadsheet.
    a. The ELA spreadsheet contained prepopulated columns of the following information for each item: session, sequence, item ID, passage, key, primary, secondary, and tertiary standards (as applicable)
    b. The Math spreadsheet contained prepopulated columns of the following information for each item: session, sequence, PARCC ID, IDEAS ID, max points, LSSM/LEAP evidence statement, secondary and tertiary standards (as applicable), key, and item type
        i. The facilitator reviewed each column in the spreadsheet so committee members could see what data was prepopulated and what data needed to be filled in by them
            1. The facilitator went over what each column represented to make sure all participants understood the metadata being presented.
            2. The facilitator went over the columns to be filled in individually by each committee member as they reviewed an item (accept with current alignment, accept with realignment, reject), and any comments they might have about the item. If committee members did not agree with an alignment, they were asked to propose a new alignment, if possible.

ii. The facilitator then reviewed the process to be used with each item.
    1. For the first couple of items, the facilitator and committee members worked through reviewing the item and its alignment(s) together. They then came to group consensus as to the status of the items.
    2. Once the group had been through a couple of items and understood the process, committee members reviewed each item individually (facilitator indicated how many items to do before getting back together for group discussion of each item) and deciding if they agree with the content alignment
    3. Once every member finished reviewing the assigned set of items, individuals shared the status they gave each item. Items were discussed one at a time until a status was assigned through group consensus to each item in a given set.
    4. This process continued for all items taken to the item alignment review.
    5. Committee members were provided with an evaluation at the end of the meeting in order that DRC and LDOE could make improvements with future item alignment reviews.

Table C.1 LEAP 2025 September 2017 Item Alignment Review Totals

Content Area	Grade	Total Items Reviewed	Accepted with Current Alignment	Accepted with Realignment	Rejected
ELA	3	120	96	24	0
ELA	4	140	131	9	0
ELA	5	112	106	3	3
ELA	6	116	106	9	1
ELA	7	119	99	19	1
ELA	8	107	74	32	1
Math	3	90	75	12	3
Math	4	75	62	8	5
Math	5	84	68	11	5
Math	6	75	66	0	9
Math	7	82	80	1	1
Math	8	82	78	1	3

Table C.2 LEAP 2025 September 2017 ELA Grade 3 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	White	Teacher of Visually Impaired (VI)
2	Female	White	Teacher
3	Female	African American	Teacher
4	Female	White	Teacher of English Learners (EL)
5	Female	White	Teacher
6	Female	White	District Supervisor
7	Female	White	Special Education (SPED) Teacher

Table C.3 LEAP 2025 September 2017 ELA Grade 4 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	African American	District Supervisor
2	Female	African American	Teacher
3	Male	White	Teacher (SPED)
4	Female	White	Teacher (VI)
5	Female	White	Teacher
6	Female	White	Teacher
7	Male	White	Teacher (VI)

Table C.4 LEAP 2025 September 2017 ELA Grade 5 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	African American	District Supervisor
2	Female	African American	Teacher
3	Female	White	Teacher
4	Female	White	Teacher (VI)
5	Female	White	Teacher
6	Female	African American	Supervisor (EL)

Table C.5 LEAP 2025 September 2017 ELA Grade 6 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	White	Teacher (SPED)
2	Female	African American	Teacher
3	Female	White	Teacher (EL)
4	Female	White	District Supervisor
5	Female	White	Teacher (VI)
6	Female	White	Teacher
7	Female	African American	Teacher
8	Male	African American	Teacher

Table C.6 LEAP 2025 September 2017 ELA Grade 7 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Male	African American	District Supervisor
2	Female	African American	Administrator
3	Female	White	Teacher (VI)
4	Female	White	Teacher
5	Female	African American	Teacher
6	Female	White	Teacher (SPED)
7	Female	White	Teacher (EL)

Table C.7 LEAP 2025 September 2017 ELA Grade 8 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	African American	Teacher
2	Male	White	Teacher (SPED)
3	Female	White	District Supervisor
4	Female	White	Teacher
5	Female	White	Teacher (SPED)
6	Female	African American	Teacher (VI)
7	Female	African American	Teacher
8	Female	White	Teacher
9	Female	Hispanic	Supervisor (EL)

Table C.8 LEAP 2025 September 2017 Mathematics Grade 3 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	White	Teacher (SPED – Hearing Impaired [HI]/VI)
2	Female	White	Teacher (SPED)
3	Female	White	Teacher
4	Female	White	Instructional Lead/Supervisor
5	Female	Hispanic/Latino	Teacher
6	Female	White	Teacher

Table C.9 LEAP 2025 September 2017 Mathematics Grade 4 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	African American	Teacher
2	Male	African American	Teacher (SPED)
3	Male	African American	Teacher
4	Female	White	Instructional Supervisor
5	Female	White	Instructional Lead
6	Female	White	Teacher
7	Female	White	Instructional Lead/Supervisor
8	Male	African American	Teacher (Immersion)

Table C.10 LEAP 2025 September 2017 Mathematics Grade 5 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	White	Teacher (SPED)
2	Female	African American	Teacher (SPED – Gifted)
3	Male	White	Teacher
4	Female	White	Instructional Lead/Supervisor
5	Female	White	Teacher
6	Female	White	Teacher
7	Female	Asian	Teacher
8	Male	White	Teacher
9	Female	White	Teacher

Table C.11 LEAP 2025 September 2017 Mathematics Grade 6 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	African American	Teacher
2	Female	African American	Teacher
3	Female	White	Instructional Supervisor
4	Female	White	Teacher
5	Female	White	Teacher (SPED)
6	Male	White	Teacher
7	Female	White	Teacher
8	Female	White	Teacher

Table C.12 LEAP 2025 September 2017 Mathematics Grade 7 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	White	Teacher
2	Male	White	Instructional Lead/Supervisor
3	Female	White	Instructional Lead/Supervisor
4	Female	White	Teacher
5	Female	Hispanic/Latino	Teacher
6	Female	White	Teacher
7	Female	Middle-Eastern	Teacher (SPED – HI/VI)

Table C.13 LEAP 2025 September 2017 Mathematics Grade 8 Item Alignment Review Committee Make-Up

Member #	Gender	Race/Ethnicity	Background
1	Female	African American	Teacher
2	Female	African American	Teacher
3	Male	White	Instructional Lead/Supervisor
4	Female	White	Instructional Lead/Supervisor
5	Female	Hispanic/Latino	Teacher
6	Female	Middle-Eastern	Teacher (SPED – HI/VI)
7	Female	African American	Teacher (SPED)

APPENDIX D – LEAP 2025 SPRING 2018 HANDSCORING/AI DOCUMENTATION

LEAP 2025 SPRING 2018 HANDSCORING/AI DOCUMENTATION

LEAP 2025 GRADES 3-8

ELA, Math, and Social Studies

LEAP 2025 HIGH SCHOOL AND EOC

Algebra I, Geometry, English I, English II, English III, Biology, and U. S. History

LEAP 2025 SPRING 2018
HANDSCORING/AI DOCUMENTATION

LEAP 2025 GRADES 3-8

ELA, Math, and Social Studies

LEAP 2025 HIGH SCHOOL AND EOC

Algebra I, Geometry, English I, English II, English III, Biology, and U. S. History

Table of Contents

Schedule, Locations, and Staffing 1
Training and Scoring Schedule 1
Scorer Degree Requirements 2
Training 2
Training Materials 3
EOC Biology, LEAP 2025 and EOC U.S. History, and LEAP 2025 Grades 3-8 Social Studies 3
EOC English III 4
LEAP 2025 Algebra I, Geometry, English I, English II, and Grades 3-8 ELA and Math 4
Algebra I, Geometry, and Grades 3-8 Math Training Set Composition 5
English I, English II, and Grades 3-8 ELA Training Set Composition 6
English I Prose Constructed Response (PCR) Items and Associated Training Materials 7
English II Prose Constructed Response (PCR) Items and Associated Training Materials 7
Grades 3-8 ELA PCR Items and Associated Training Materials 8
Algebra I Items and Associated Training Materials 9
Geometry Items and Associated Training Materials 10
Grade 3 Math Items and Associated Training Materials 11
Grade 4 Math Items and Associated Training Materials 11
Grade 5 Math Items and Associated Training Materials 12
Grade 6 Math Items and Associated Training Materials 12
Grade 7 Math Items and Associated Training Materials 13
Grade 8 Math Items and Associated Training Materials 13
Qualifying 14
EOC Biology, English III, and U.S. History 14
Biology 14
English III 14
U.S. History 14
LEAP 2025 Constructed-Response and Extended-Response Items 15
English I, English II, and Grades 3-8 ELA 15

Algebra I, Geometry, and Grades 3-8 Math 15
U.S. History and Grades 3-8 Social Studies 16
Spring 2018 Scoring Plan 16
LEAP 2025 High School and EOC 16
LEAP 2025 Grades 3-8 16
Handscoring Rules 17
AI Scoring 17
Handscoring 17
Reader Monitoring Procedures 18
Team Leader Read-Behinds 18
Validity Sets and Inter-Rater Reliability 18
Calibration Sets 21
Handscoring Quality Control Reports 22
Scoring Summary Report Sample 22
Scoring Summary Report Sample with AI (Reader ID #3) 22
Reader Feedback Logs 23
Handling Unusual Responses 23
Nonscore Codes and Definitions 23
Nonscore Code Definitions 24
Nonscore Codes by Grade/Course 24
Alerts 24
Artificial Intelligence Scoring 25
AI Scoring – Measurement, Inc. 25
Model Building 26
Evaluation Metric 27
Scoring Responses with the AI Engine 28
Quality Control of the AI Engine 29
Identifying Responses for Human Review 29
Alert Detection System 30

Identification of Nonscoreables 31
Identifying Copied Text and Plagiarism with the AI Engine 31
AI Scoring – Pearson 34
The Intelligent Essay Assessor 34
How the Intelligent Essay Assessor was Trained 35
Quality Monitoring 37
Scoring (DRC) 38
Rescores (High School Tests Only) 39
Auto Rescores 39
Parent-Requested Rescores 40
Appendix A 42
DRC-MI Streaming Scoring Documentation 42
SECTION 1 – General Information 43
SECTION 2 – SCHEMA SUPPLEMENT 44
SECTION 3 – STATUS CODE INFORMATION 47
Appendix B 48
AI Model Building – English III (Spring 2018) 48
Quadratic Weighted Kappa (QWK) and Inter-rater Reliability (IRR) 48
Score Point Distribution (SPD) 48
AI Model Building – EOC U.S. History ER (Spring 2018) 49
Quadratic Weighted Kappa (QWK), Inter-rater Reliability (IRR), and Score Point Distribution (SPD) 49
AI Model Building – LEAP 2025 U.S. History ERs (Spring 2018) 50
Quadratic Weighted Kappa (QWK), Inter-rater Reliability (IRR), and Score Point Distribution (SPD) 50
AI Model Building – Social Studies Grades 5-8 ERs (Spring 2018) 51
Quadratic Weighted Kappa (QWK), Inter-rater Reliability (IRR), and Score Point Distribution (SPD) 51
PCR Performance – ELA Grades 5-8, English I, and English II (Spring 2018) 52
Reading Comprehension and Written Expression or Written Expression Trait 52
Conventions Trait 52
Spring 2018 LEAP 2025 and EOC Items – IRR and SPD from Previous Administrations 53

Algebra I 53
Geometry 54
Math Grade 3 55
Math Grade 4 56
Math Grade 5 57
Math Grade 6 58
Math Grade 7 59
Math Grade 8 60
English I 61
English II 62
English III (All Report Data Scored by DRC and MI [AI]) 63
ELA Grade 3 64
ELA Grade 4 65
ELA Grade 5 66
ELA Grade 6 66
ELA Grade 7 67
ELA Grade 8 67
Biology 68
U.S. History ERs (EOC) 69
U.S. History ERs and CRs (LEAP 2025) 70
Social Studies Grades 3-8 71

Schedule, Locations, and Staffing

All reader training and handscoring for Spring 2018 LEAP 2025 Grades 3-8 and High School tests and End-of-Course (EOC) High School tests will take place at the DRC scoring center locations noted in the table below.

Training and Scoring Schedule

DRC's Spring 2018 LEAP 2025 reader training and scoring schedule is based on the spring testing windows of April 9, 2018 – May 4, 2018 (LEAP 2025 grades 3-8) and April 23, 2018 – May 18, 2018 (LEAP 2025 and EOC high school). Reader training and scoring locations and the anticipated dates for each are noted below.

Grade/Subject/ Course	DRC Scoring Center Location	Anticipated Staffing	2018 Reader Training and Scoring Window
3 ELA	Columbus, OH	1 Scoring Director, 5 Team Leaders, 55 Readers	May 14 – June 4
4 ELA	Madison, WI	1 Scoring Director, 6 Team Leaders, 57 Readers	May 14 – June 4
5 ELA	Madison, WI	1 Scoring Director, 4 Team Leaders, 34 Readers	April 16 – May 11
6 ELA	Lake Mary, FL	1 Scoring Director, 4 Team Leaders, 35 Readers	April 12 – May 11
7 ELA	Atlanta, GA	1 Scoring Director, 1 Team Leader, 10 Readers	April 16 – May 11
8 ELA	Atlanta, GA	1 Scoring Director, 4 Team Leaders, 35 Readers	April 16 – May 11
3 Math	Plymouth, MN	2 Scoring Directors, 4 Team Leaders, 36 Readers	May 14 – June 4
4 Math	Plymouth, MN	2 Scoring Directors, 4 Team Leaders, 38 Readers	May 14 – June 4
5 Math	Sharonville, OH	2 Scoring Directors, 4 Team Leaders, 38 Readers	April 16 – May 11
6 Math	Atlanta, GA	2 Scoring Directors, 5 Team Leaders, 48 Readers	April 16 – May 11
7 Math	Woodbury, MN	2 Scoring Directors, 5 Team Leaders, 51 Readers	April 16 – May 11
8 Math	Sharonville, OH	2 Scoring Directors, 5 Team Leaders, 51 Readers	April 16 – May 11
3 SS	Atlanta, GA	1 Scoring Director, 4 Team Leaders, 43 Readers	May 14 – June 4
4 SS (CRs)	Atlanta, GA	1 Scoring Director, 3 Team Leaders, 25 Readers	May 14 – June 4
4 SS (ER)	Plymouth, MN	1 Scoring Director, 2 Team Leaders, 18 Readers	May 14 – June 4
5 SS	Plymouth, MN	1 Scoring Director, 2 Team Leaders, 15 Readers	April 9 – May 11
6 SS	Plymouth, MN	1 Scoring Director, 2 Team Leaders, 15 Readers	April 9 – May 11
7 SS	Plymouth, MN	1 Scoring Director, 2 Team Leaders, 15 Readers	April 9 – May 11
8 SS	Plymouth, MN	1 Scoring Director, 2 Team Leaders, 15 Readers	April 9 – May 11
LEAP 2025 Algebra I	Sharonville, OH	2 Scoring Directors, 4 Team Leaders, 37 Readers	April 24 – May 25
LEAP 2025 Geometry	Sharonville, OH	2 Scoring Directors, 2 Team Leaders, 23 Readers	April 24 – May 25
LEAP 2025 English I	Plymouth, MN	1 Scoring Director, 1 Team Leader, 10 Readers	April 23 – May 25
LEAP 2025 English II	Plymouth, MN	1 Scoring Director, 1 Team Leader, 8 Readers	April 23 – May 25
EOC English III	Plymouth, MN	1 Scoring Director, 1 Team Leader, 8 Readers	April 23 – May 25
EOC Biology	Plymouth, MN	1 Scoring Director, 1 Team Leader, 6 Readers	April 19 – May 25
EOC U.S. History	Plymouth, MN	1 Scoring Director, 3 Team Leaders	April 23 – May 25
LEAP 2025 U.S. History	Plymouth, MN	1 Scoring Director, 3 Team Leaders, 21 Readers	
For U.S. History, the same core group of scoring supervisors will be involved with all EOC and LEAP 2025 scoring. In addition, the same team of 21 readers will be involved in the LEAP 2025 ER and CR scoring.			

Each DRC scoring center is a secure facility. 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 will retain these agreements for the duration of the contract. To prevent the unauthorized duplication of secured materials, cell phone/camera use within the scoring rooms is strictly forbidden. Readers only have access to student responses they are qualified to score. Each scorer is assigned a unique username and password to access DRC's imaging system and must qualify before viewing any live student responses. DRC maintains full control of who may access the system and which item each scorer may score. No demographic data is available to scorers at any time.

Scorers will be divided by subjects as detailed in the previous table. Depending on the overall progress of the project, more scorers may be added to some groups. Additionally, depending on the overall progress of the project, some groups may subdivide and work on different items.

Scorer Degree Requirements

DRC readers scoring for Louisiana have at least a four-year college degree. DRC has a Human Resources Director dedicated solely to recruiting and retaining our handscoring staff. In the screening process, preference is given to candidates with previous experience scoring large-scale assessments and with degrees emphasizing the appropriate content areas. During personal interviews, reader candidates are asked to demonstrate their own proficiency in writing by responding to a DRC writing topic and in mathematics by solving word problems with correct work shown. All of this results in a highly educated and diverse workforce. Our personnel files for readers and Team Leaders include evaluations for each project completed. We use these evaluations to place individuals on projects that best fit their professional backgrounds, their college degrees, and their performance on similar projects at DRC.

Training

In preparation for the scoring of all Spring 2018 LEAP 2025 and EOC items, DRC scoring supervisors will train readers using the same training materials that were used by previous vendors for prior administrations of the same items. These training materials originated from the sources noted below.

Reader training materials for the following were provided to DRC by Pacific Metrics and previously approved by LDOE.

- EOC Biology, English III, and U.S. History

Reader training materials for the following were developed by DRC in conjunction with LDOE:

- LEAP 2025 grades 3-8 Social Studies
- LEAP 2025 U.S. History

Reader training materials for the following were provided to DRC by the Partnership for Assessment of Readiness for College and Careers (PARCC) and were approved by PARCC:

- LEAP 2025 grades 3-8 ELA and Math
- LEAP 2025 Algebra I, Geometry, English I, and English II

The materials include:

- Passages, items/prompts, associated stimuli for applicable content areas and item types;
- Rubrics;
- Anchor Sets;
- Training Sets (or Practice Sets); and
- Qualifying Sets.

DRC will start the training with a review of passages, items/prompts, rubrics, and anchor responses, followed by the scoring and discussion of Training/Practice Sets and the scoring and discussion of Qualifying Sets. Once this process has been completed for an item or prompt, qualified readers will be able to start scoring live student responses. A group of scorers will score responses for a particular item until the scoring for that item is complete. Then they may move on to score a different item. Depending on the overall progress of the project and the current quantity of responses available to score for each item, some groups may subdivide and work on different items. Additionally, depending on the overall progress of the project, more scorers may be added to some groups when the groups are ready to score new items.

The following tables detail the composition of the training materials for all of the Spring 2018 LEAP 2025 Grades 3-8 and High School and EOC tests.

Training Materials

EOC Biology, LEAP 2025 and EOC U.S. History, and LEAP 2025 Grades 3-8 Social Studies

Reader training for EOC Biology and U.S. History is conducted using item-specific anchor sets, training sets, and qualifying sets provided by Pacific Metrics. The LEAP 2025 U.S. History and grades 3-8 Social Studies item-specific training materials were developed by DRC in conjunction with LDOE.

Set Type	**Biology, U.S. History, and Grades 3-8 Social Studies Training Materials**	**Annotated**
Anchor Set	Each item-specific Anchor Set contains at least two responses per score point (with at least one of each at the top score points).	Yes
Training Sets	There are at least two Training Sets for each item ● 10 responses per Training Set ● All numeric score points are represented*	No
Qualifying Sets	There are one or more Qualifying Sets for each item ● 10 responses per Qualifying Set ● All numeric score points are represented*	No
*Examples of responses at the top score points may not be present in some Training or Qualifying Sets as there may have been few or no examples found during rangefinding or subsequent field test scoring. In such cases, DRC Scoring Directors will endeavor to find examples of these scores during live scoring to supplement reader training.		

EOC English III

For English III, the Content and Style dimensions will be trained using prompt-specific scoring guides, training sets, and qualifying sets provided by Pacific Metrics.

Set Type	Content and Style Dimension Training Materials	Annotated
Content and Style Anchor Set*	20 responses representing all numeric score points and including nonscore condition codes.	Yes
Content and Style Training Sets 1 – 3	30 responses across the three training sets. All numeric score points are represented in each set. Readers will score each response for both Content and Style.	No
Content and Style Qualifying Sets 1 – 3	40 responses across the three qualifying sets. All numeric score points are represented in each set. Readers will score each response for both Content and Style.	No
* Some responses appear more than once in the Anchor Set to illustrate both a Content and a Style score.		

English III training materials covering the four elements of the Conventions dimension (Sentence Formation [F], Usage [U], Mechanics [M], and Spelling [S]) are made up of sets that include examples from multiple writing prompts. Roughly half of the responses in each set are a score point 0 for each element, and the other half are a score point 1 for each element.

Set Type	Conventions (F, U, M, S) Element Training Materials	Annotated
Conventions Anchor Set*	6 responses for each Conventions element representing both numeric score points (three 0s and three 1s for each of the four elements).	Yes
Conventions Training Sets 1 – 2	20 responses across both training sets. Both numeric score points are represented in each set. Readers will score each response for all four elements.	No
Conventions Qualifying Sets 1 – 2	20 responses across both training sets. Both numeric score points are represented in each set. Readers will score each response for all four elements.	No
* Some responses appear more than once in the Conventions Anchor Set to illustrate scores in more than one element.		

LEAP 2025 Algebra I, Geometry, English I, English II, and Grades 3-8 ELA and Math

For LEAP 2025 Algebra I, Geometry, English I, English II, and grades 3-8 ELA and math, DRC will use the PARCC-approved training and qualifying materials provided by PARCC. Training materials for each item can be grouped into one of two categories: “prototype” item materials or “abbreviated” item materials.

Prototype Item Materials

Some items included in the Louisiana forms are prototype items, meaning they have full sets of associated training materials, including Anchor Sets, Practice Sets, and Qualifying Sets. DRC will start the training with a review of passages/items, rubrics, and anchor responses, followed by the scoring and discussion of Practice Sets and the scoring and discussion of Qualifying Sets. Once this process has been completed for a prototype item included on the Louisiana form, qualified readers will start scoring live student responses for that item.

Abbreviated Item Materials

Unlike prototype items, abbreviated item training materials have only two item-specific Practice Sets and no Qualifying Sets; therefore, abbreviated items require a two-step training/qualifying process. First, scorers will train and qualify as described in the Prototype Item Materials section above using PARCC-approved training materials for an associated prototype item that is similar to the abbreviated one they will be scoring on the Louisiana form.[1] Readers who do not qualify on the prototype item will not be allowed to continue the training.

After qualifying on the associated prototype item, readers receive additional item-specific training on the abbreviated item they are going to score. This consists of an item-specific Anchor Set and two item-specific Practice Sets. After completing the abbreviated item's training, readers may begin scoring live responses for the abbreviated item.

The following tables detail the composition of the training materials provided by PARCC for math and ELA:

Algebra I, Geometry, and Grades 3-8 Math Training Set Composition

Set Type	Mathematics Prototype Item Training	Mathematics Abbreviated Item Training	Annotated
Anchor Set	3 responses per score point (Composite items will have 3 responses per composite score)	3 responses per score point (Composite items will have 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

[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 English items, this determination by PARCC and Pearson was based on grade/course and task type.

English I, English II, and Grades 3-8 ELA Training Set Composition

Set Type	English Prototype Item Training	English Abbreviated Item Training	Annotated
Anchor Set (for the RCWE and WE traits)	3 responses per score point ● Anchor Sets for prototype RST and LAT item training include scores for the combined trait Reading Comprehension and Written Expression (RCWE). ● Anchor sets for prototype NWT item training include scores for Written Expression (WE).	3 responses per score point ● Anchor Sets for abbreviated RST and LAT item training include scores for the combined trait Reading Comprehension and Written Expression (RCWE). ● Anchor Sets for abbreviated NWT item training include scores for Written Expression (WE).	Yes
Practice Set 1	5 responses representing the range of responses for ● the RCWE trait (for LAT and RST items) ● the WE 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 (includes both traits appropriate to the task type)		No
Qualifying Set 2	10 responses comparable to the anchor set responses (includes both traits appropriate to the task type)		No
Qualifying Set 3	10 responses comparable to the anchor set responses (includes 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 (includes both traits appropriate to the task type)	3-5 responses composed entirely or partially of text copied from passage or passages (includes both traits appropriate to the task type)	Yes
*The PARCC-approved Direct Copy sets provide additional annotated sample responses that explain the scoring rationale for responses composed entirely or partially of text copied from the source passage(s) associated with an item. DRC scoring supervisors review these item-specific sets with the readers prior to scoring the associated item.			

Set Type	English Prototype and Abbreviated Item Training	Annotated
Anchor Set (for the Knowledge of Language and Conventions trait)	● There are 3 responses per score point in each set. ● There are two mixed-prompt Anchor Sets per grade level (one set for NWT item training, another set for LAT/RST item training). These sets are not exclusive to specific prototype or abbreviated items; they are intended to familiarize readers with the conventions features appropriate to each task type. ● Subsequent Practice Sets for prototype and abbreviated items will require readers to practice scoring the Knowledge of Language and Conventions trait along with the RCWE trait (for LAT or RST) or with the WE trait (for NWT). ● In addition, readers will be required to qualify on the Knowledge of Language and Conventions trait during each prototype item qualifying session.	Yes

English I Prose Constructed Response (PCR) Items and Associated Training Materials

Question	Task	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
7	LAT	902152	VH017536_2T	abbreviated	QCE09_L_VH037763_2T
20	RST	902161	VH017542_2T	abbreviated	QCE09_R_VH017542_2T
9 [AE form]	RST	902161	VH017542_2T	abbreviated	QCE09_R_VH017542_2T
14 [AE form]	NWT	906152	VH084830	abbreviated	QCE09_N_6139
*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.					

English II Prose Constructed Response (PCR) Items and Associated Training Materials

Question	Task	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
7	LAT	906197	HH428127697	abbreviated	QCE10_L_7233_2T
20	RST	902331	VH004490	abbreviated	QCE10_R_7121_2T
9 [AE form]	RST	902331	VH004490	abbreviated	QCE10_R_7121_2T
14 [AE form]	NWT	902354	7064	abbreviated	QCE10_N_VF908613
*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.					

Grades 3-8 ELA PCR Items and Associated Training Materials

Grade	Question	Task	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
3	7	RST	868500	E03_R_VF581237	abbreviated	E03_R_VF906000
	12	NWT	913497	E03_N_AA431426588	abbreviated	E03_N_VF910093
4	7	LAT	913567	E04_L_VH170170	abbreviated	E04_L_VF925727
	20	RST	802342	E04_R_1805	abbreviated	E04_R_VF653524
5	9	RST	868590	E05_R_C1381	abbreviated	E05_R_2208
	14	NWT	913663	E05_N_CC434542756	abbreviated	E05_N_VH003158
6	9	RST	870622	E06_R_VH082653	abbreviated	E06_R_3538
	14	NWT	913694	E06_N_D1466	abbreviated	E06_N_VH000592
7	9	RST	868704	E07_R_VH069361	abbreviated	E07_R_VH014400
	14	NWT	913842	E07_N_EE430133306	abbreviated	E07_N_4284
8	7	LAT	913958	E08_L_F1460	abbreviated	E08_L_5271
	20	RST	868741	E08_R_5013	abbreviated	E08_R_VH007336

*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.

Algebra I Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item
12	901832	3031-M44083P	Abbreviated	3003_M43111*
13	901882	VH196970	Abbreviated	VH046614*
25	901687	2407-M41752	Prototype	N/A
26	906043	2362-M41568	Abbreviated	VH046614*
37	901851	M41726	Abbreviated	3003_M43111*
38	901705	VF883359	Abbreviated	VH046614*
39	901857	VH046479	Abbreviated	2407_M41752*
12 [AE form]	901832	3031 M44083P	Abbreviated	3003_M43111*
13 [AE form]	901882	VH196970	Abbreviated	VH046614*
26 [AE form]	906043	2362 M41568	Abbreviated	VH046614*
28 [AE form]	901859	3003 M43111	Prototype	N/A
37 [AE form]	901851	M41726	Abbreviated	3003_M43111*
39 [AE form]	901857	VH046479	Abbreviated	2407_M41752*
44 [AE form]	901848	M47287	Abbreviated	M41686*
*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.				

Geometry Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item
13	902012	M41169	Abbreviated	VF935309*
15	902046	M46668	Abbreviated	3042_M44133*
27	902027	M43233	Abbreviated	VH001716*
28	902042	3020-M44058	Abbreviated	3042_M44133*
43	902062	VH150384	Abbreviated	VF613786*
44	902018	3042-M44133	Prototype	N/A
13 [AE form]	902012	M41169	Abbreviated	VF935309*
15 [AE form]	902046	M46668	Abbreviated	3042_M44133*
27 [AE form]	902027	M43233	Abbreviated	VH001716*
28 [AE form]	902036	2904-M43021	Prototype	N/A
43 [AE form]	902047	VH150404	Abbreviated	VF935309*
44 [AE form]	902018	3042-M44133	Prototype	N/A
*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.				

Grade 3 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
17	914008	VH054861	abbreviated	QCM03_M01883
18	914048	M05158	abbreviated	QCM03_M00848
32	798301	VF658085	abbreviated	QCM03_M02345
33	868619	M00848	prototype	N/A
48	914039	M02527	abbreviated	QCM03_M00848
49	914037	4003-M03007	abbreviated	QCM03_M00848
*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.				

Grade 4 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
17	914084	4112-M03491P	abbreviated	QCM04_0081_M00445
18	914086	M04133	abbreviated	QCM04_M03436
32	870723	M02081	abbreviated	QCM04_M02470
33	870317	M02255	abbreviated	QCM04_M02470
48	870722	0220-M00114P	abbreviated	QCM04_0081_M00445
49	870724	0083-M00453	abbreviated	QCM04_M02470
*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.				

Grade 5 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
17	914193	VH055372	abbreviated	QCM05_VH084803
18	914148	M03888	abbreviated	QCM05_VH141466
32	870325	4106-M03478	abbreviated	QCM05_VH084803
33	868655	M03555	prototype	N/A
48	914195	0154-M00796	abbreviated	QCM05_VH084803
49	868654	4134-M03661	abbreviated	QCM05_M03555
*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.				

Grade 6 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
31	914286	5047-M25308	abbreviated	QCM06_VH122131
35	868818	M22225	abbreviated	QCM06_M21577
36	868798	1298-M21432	abbreviated	QCM06_M21577
37	870825	M21477	abbreviated	QCM06_VF655921
47	914231	1740-M23030	abbreviated	QCM06_VH122131
48	914244	5012-M25073	abbreviated	QCM06_M21577
49	914281	M25152	abbreviated	QCM06_VF655921
*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.				

Grade 7 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
30	914362	VH083535	abbreviated	QCM07_VF643181
35	914333	1300-M21438	abbreviated	QCM07_VH151385
36	868849	M22018	prototype	N/A
37	878705	VH095906	abbreviated	QCM07_M22018
47	914336	M25731	abbreviated	QCM07_M20598
48	914339	VH151385	prototype	N/A
49	914329	M25580	abbreviated	QCM07_M20598
*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.				

Grade 8 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
29	892626	VH095800	prototype	N/A
34	800273	M20195	abbreviated	QCM08_M20587
35	914376	VH049778	abbreviated	QCM08_M21063
36	878156	VH057654	abbreviated	QCM08_M20198
46	914381	M25425	abbreviated	QCM08_M21063
47	868882	VH081259	abbreviated	QCM08_M20198
48	914400	M21215	abbreviated	QCM08_VH095800
*An item ID listed in the Associated Prototype column indicates that readers must be qualified on that prototype prior to reviewing the Abbreviated training materials described in the cells to the left.				

As part of their training, readers in all content areas are also instructed on how to recognize alerts such as disturbing content and suspected plagiarism. Alert processing procedures are described in a later section of this document.

Qualifying

Scorers must demonstrate 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 has been scored, the DRC Scoring Director responsible for training the item will lead the scorers in a discussion of the set.

Any scorer who does not qualify by the end of the qualifying process for an item will not be allowed to score actual student work for that item.

In order to maintain scoring comparability with prior administrations of the same items, DRC will use the same qualifying standards for Spring 2018 LEAP 2025 and EOC items as were established by the vendors who scored these items previously.

EOC Biology, English III, and U.S. History

Descriptions of the qualifying standards for the Biology, English III, and U.S. History item types are below. These standards were established by Pacific Metrics.

Biology

Course	Qualifying Standard
Biology	Scorers must qualify with 80% exact agreement or higher on one or more of the qualifying sets in order to score student responses.

English III

Course	Qualifying Standard
English III (Content and Style)	English III scorers will first qualify on the Content and Style dimensions before moving on to qualify in the Conventions dimension. Each reader will complete at least two qualifying sets, and a score of 70% exact agreement or higher is required in each dimension in order to qualify. Since readers complete two sets, they may qualify on one dimension in the first set and the other dimension in the second set.
English III (Conventions)	Once qualified for Content and Style, readers must then qualify for each of the four elements (F, U, M, and S) that make up the Conventions dimension. An exact agreement rate of 80% or higher is required once on each of the individual Conventions elements. A scorer may qualify on some elements in the first qualifying set and the remaining elements in the second qualifying set.

U.S. History

Course	Qualifying Standard
U.S. History	Readers must qualify with 70% exact agreement or higher on one or more of the qualifying sets in order to score student responses.

LEAP 2025 Constructed-Response and Extended-Response Items

For the LEAP 2025 ELA and math CR items, DRC will follow the same qualification standards followed by Pearson for PARCC. A description of these PARCC qualifying standards is below.

English I, English II, and Grades 3-8 ELA

Course	Qualifying Standard	
English I, English II, and Grades 3-8 ELA	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	

Readers of English I, English II, and grades 3-8 ELA responses are required to meet all three of the qualifications listed in the table. Perfect Plus Adjacent Agreement of 96% means that out of the entire pool of scores that a reader gives across the three qualifying sets for an item, no more than 4% of those scores can be non-adjacent. In other words, no more than 2 of the 60 applied scores can be non-adjacent (3 sets x 10 responses/set x 2 traits = 60 applied score).

Algebra I, Geometry, and Grades 3-8 Math

Course	Qualifying Standard		
Algebra I, Geometry, and Grades 3-8 Math	Comprehensive	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

Course	Qualifying Standard		
Algebra I, Geometry, and Grades 3-8 Math	Composite (multi-part) Items*	Perfect Agreement	Perfect Plus Adjacent Agreement
	0, 1 Rubric	90% on two of three sets	100% on two of three sets
	0, 1, 2 Rubric	80% on two of three sets	96% on two of three sets
	0, 1, 2, 3 Rubric	70% on two of three sets	96% on two of three sets
	0, 1, 2, 3, 4 Rubric	70% on two of three sets	95% on two of three sets

*For mathematics composite items, the appropriate qualifying standard should be achieved on each part of the item. For example, if an item has 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.

U.S. History and Grades 3-8 Social Studies

Course and Item Type	Qualifying Standard
Grades 3-8 Social Studies and U.S. History 0-2 point CRs	Scorers must qualify with 80% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
Grade 3 Social Studies 4 point, single dimension ER (Content & Claims)	Scorers must qualify with 70% exact agreement or higher in the Content & Claims trait on one or more of the qualifying sets in order to score student responses.
Grades 4-8 Social Studies and U.S. History 8 point, 2-dimension ERs (Content, 0-4; Claims, 0-4)	Scorers must qualify with 70% exact agreement or higher in both the Content trait and the Claims trait on one or more of the qualifying sets in order to score student responses. Since scorers complete two sets, they may qualify on one trait in the first set and the other trait in the second set.

Spring 2018 Scoring Plan

The charts below provide an overview of the Spring 2018 LEAP 2025 and EOC scoring plan, detailing the types of scoring that will be done for each course/grade.

LEAP 2025 High School and EOC

Course	Handscoring Only	AI Scoring	AI Vendor
LEAP 2025 English I	PCRs (AE form)	PCRs (operational)	Pearson
LEAP 2025 English II	PCRs (AE form)	PCRs (operational)	Pearson
EOC English III	Writing Prompt (AE form)	Writing Prompt (operational)	Measurement Inc.
LEAP 2025 Algebra I	All CRs	N/A	
LEAP 2025 Geometry	All CRs	N/A	
EOC U.S. History	ER (AE form)	ER (operational)	Measurement Inc.
LEAP 2025 U.S. History	All CRs, ER (AE form)	ERs (operational)	Measurement Inc.
EOC Biology	ERs (operational and AE form)	N/A	
Note: All Administrative Error [AE] form items are handscored by DRC scoring supervisors.			

LEAP 2025 Grades 3-8

Course	Handscoring Only	AI Scoring	AI Vendor
ELA grade 3	Both PCRs	N/A	
ELA grade 4	Both PCRs	N/A	
ELA grade 5	RST_C1381	NWT_CC434542756	Pearson
ELA grade 6	RST_VH082653	NWT_D1466	Pearson
ELA grade 7	N/A	Both PCRs	Pearson
ELA grade 8	RST_5013	LAT_F1460	Pearson
Social Studies grades 3 and 4	All CRs and ERs	N/A	
Social Studies grades 5-8	All CRs	All ERs	Measurement Inc.
Math grades 3-8	All CRs	N/A	

Handscoring Rules

AI Scoring

For EOC English III, EOC U.S. History, and the LEAP 2025 U.S. History and grades 5-8 Social Studies ER items, Measurement Incorporated's (MI) Project Essay Grade (PEG) AI scoring system will provide the first score (the score of record). For LEAP 2025 English I, English II, and select PCRs in grades 5-8 ELA (see Spring 2018 Scoring Plan), Pearson's Intelligent Essay Assessor (IEA) will provide the first score (the score of record). DRC's handscoring teams will provide a second read for at least ten percent of these responses in order to capture the inter-rater reliability statistics that will be used to manage scoring consistency of both the AI scoring systems and the handscoring teams. Scoring Directors will also review nonscores, alerts, and flagged responses as required. (The AI scoring process is discussed in-depth later in this document.)

Handscoring

All scores for handscored items (noted as Handscoring Only in the Spring 2018 Scoring Plan above) will be provided by DRC's handscoring team. The first score will be the score of record. Ten percent of the responses will be scored twice to monitor and maintain inter-rater reliability. Scoring Directors will review all nonscores and alerts.

In addition, per PARCC/Pearson rules for ELA and math, if the first two scores are nonadjacent (e.g., 0, 2), a third, independent reading by a Team Leader or Scoring Director will be conducted for additional quality control monitoring. In the unlikely event that a response receives three nonadjacent scores (i.e., 0, 2, 4), a Scoring Director or Project Manager will review the response and provide retraining as needed.

Calculating the Final Score:

- The first score is always the score of record, regardless of how many subsequent scores are applied.
- After handscoring, when the final score processing for the ELA items takes place, the Written Expression trait score is multiplied by 3 (for the Narrative Writing Task). The Reading Comprehension and Written Expression (RCWE) trait score is multiplied by 4 (for the Literary Analysis and Research Simulation tasks), and one fourth of this weighted score will be assigned as the Reading Comprehension score, and three fourths of this weighted score will be assigned as the Written Expression score. The Knowledge of Language and Conventions score is not weighted. (For more information, please see section *2.7.3 Scoring Rules* within the *LEAP 2025 High School and EOC Processing Rules – Scoring* document.)

Reader Monitoring Procedures

Team Leader Read-Behinds

Throughout the handscoring process, DRC Project Managers, Scoring Directors, and Team Leaders will review the statistics that are generated on a daily basis. DRC will assign one Team Leader for every 10–12 readers. If scoring patterns are apparent among individual scorers, Team Leaders will handle these issues on an individual basis. If a scorer appears to need clarification of the scoring rules, DRC supervisors typically monitor one out of five of the scorer's readings, making adjustments to that ratio as needed. If a supervisor disagrees with a reader's scores during monitoring, he or she will correct the score and provide retraining in the form of direct feedback to the reader, using rubric language and applicable training responses. The supervisor's corrected score becomes the score of record; it is not a second read.

DRC will also monitor the inter-rater reliability, which is to be based on the 10% of responses that receive second reads. If a scorer falls below the expected rate of agreement, the Team Leader or Scoring Director will retrain the scorer. If a scorer fails to improve after retraining and feedback, DRC will remove the scorer from the project. In this situation, DRC will remove all unreported scores that were assigned by the scorer during the period in question. These unreported responses with dropped scores will then be re-dealt and rescored.

Validity Sets and Inter-Rater Reliability

In addition to the feedback that supervisors provide to readers based on regular read-behinds and the continuous monitoring of inter-rater reliability and score point distributions, DRC will also conduct validity scoring. For LEAP 2025 Algebra I, Geometry, English I, English II, and grades 3-8 math and ELA, DRC will utilize the same validity responses that Pearson is using. These PARCC-approved validity responses were supplied by PARCC and may include both paper- and computer-based test responses.

Since the previous vendor of the EOC items did not provide validity responses, and because the Biology item is operational at DRC for the first time in the spring of 2018, DRC's Scoring Directors and Team Leaders will need to identify validity responses for Biology during live scoring. DRC scoring supervisors will also identify validity responses during live scoring for all newly operational LEAP 2025 U.S. History and grades 3-8 Social Studies ERs and CRs. DRC Education Project Management staff will post these validity responses and their scores onto the secure LDOE SFTP site for LDOE content staff to review and approve. Validity responses previously identified and approved for English III will be reused.

The validity responses will be added to DRC's image handscoring system prior to the beginning of scoring. The distribution of validity responses will be more frequent at the beginning of the scoring window and will decrease as agreement levels reveal a strong understanding and application of the scoring guidelines by the scorers. Validity reports compare scorers' scores to pre-determined scores and can help detect potential room drift as well as individual scorer drift. This data will be used to make decisions regarding the retraining and/or release of scorers, as well as the rescoring of responses.

To monitor inter-rater reliability, DRC will produce handscoring quality control reports on a daily basis (see the sample on page 22) that provide exact, adjacent, and nonadjacent agreement rates for each reader and item on a daily and cumulative basis. These rates are calculated based on responses that are scored by two readers (or PEG or IEA—the AI scoring systems—and one reader). For English III and the U.S. History and grades 5-8 Social Studies ERs, MI's PEG AI scoring system will provide the first scores (the scores of record). For LEAP 2025 English I, English II, and select PCRs in grades 5-8 ELA (see Spring 2018 Scoring Plan), Pearson's IEA will provide the first score (the score of record). This data will be used in conjunction with scores from human-conducted second reads to calculate inter-rater reliability statistics in these content areas. Metrics and standards associated with the two AI scoring systems and their processes are described in the AI Scoring section starting on page 25. AI scores will be attributed to reader ID number 3 in the appropriate scoring reports. The calculations on these reports are:

- **Percent Exact (%EX)**—total number of responses by reader where scores are the same, divided by the number of responses that were scored twice.

- **Percent Adjacent (%AD)**—total number of responses by reader where scores are one point apart, divided by the number of responses that were scored twice.

- **Percent Non-Adjacent (%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.

DRC will strive to maintain inter-rater and validity exact agreement percentages based on those of previous administrations for the same items. These rates should be at or above the rates specified below for each content area. When a reader's validity or inter-rater agreement falls 5% or more below these expectations, or when non-adjacent percentages exceed 5%, the reader will be flagged for additional monitoring and/or retraining by their Team Leader or Scoring Director. Additionally, for all items which will be AI scored, low inter-rater reliability will need to be investigated to see if it is an indication that the handscorers need retraining or if the AI needs retraining (see the AI Scoring section for details about AI training).

The validity and inter-rater reliability expectations for Biology, English III, U.S. History, and grades 3-8 Social Studies items are shown below:

Agreement Rate Requirements for Validity and Inter-Rater Reliability – Biology, English III, U.S. History, and Grades 3-8 Social Studies			
Subject	**Score Point Range**	**Perfect Agreement**	**Perfect Agreement + Adjacent**
EOC Biology	0-4	80%	95%
EOC English III (Style and Content)	1-4 (each domain)	70%	95%
EOC English III (F, U, M, S)	0-1 (each domain)	80%	95%
EOC U.S. History 4 point ERs	0-4	70%	95%
LEAP 2025 U.S. History, Grades 4-8 Social Studies 8 point, 2 dimension ERs	0-4 (each trait)	70%	95%
LEAP 2025 Grade 3 Social Studies 4 point ERs	0-4	70%	95%
LEAP 2025 U.S. History, Grades 3-8 Social Studies 2 point CRs	0-2	80%	95%

The validity and inter-rater reliability expectations for PARCC mathematics and English items (provided by Pearson) are shown in the following table:

Agreement Rate Requirements for Validity and Inter-Rater Reliability – Algebra I, Geometry, English I, English II, and Grades 3-8 ELA (PARCC Items)			
Subject	**Score Point Range**	**Perfect Agreement**	**Perfect Agreement + Adjacent**
Algebra I, Geometry, Grades 3-8 Math	0-1	90%	95%
Algebra I, Geometry, Grades 3-8 Math	0-2	80%	95%
Algebra I, Geometry, Grades 3-8 Math	0-3	70%	95%
Algebra I, Geometry, Grades 3-8 Math	0-4	65%	95%
English I, English II, Grades 3-8 ELA	Multi-trait	65% (each trait)	96% (each trait)

Each reader will be required to maintain an acceptable level of exact agreement on validity responses and on inter-rater reliability as described above. Additionally, readers will be required to maintain an acceptably low rate of nonadjacent agreement for validity and inter-rater agreement. To monitor this, we will sum each reader's percentages of exact and adjacent agreement rates and require each reader to maintain the levels displayed under "Perfect Agreement + Adjacent" on the tables above.

Calibration Sets

Calibration sets are another means of ensuring consistency in scoring. DRC will use these sets to maintain calibration across the entire scorer population after breaks from scoring (e.g. weekends; down time between scoring periods; when moving between items/prompts). Calibration sets will also be used for an item if trends occur (e.g., low agreement between certain score points, if a certain type of response is missing from initial training).

The responses in these targeted sets help illustrate particular points and familiarize readers with the types of responses commonly seen during operational scoring. They are chosen by DRC scoring supervisors during live scoring (for English III, Biology, U.S. History, and grades 3-8 Social Studies items) or supplied by Pearson (for Algebra I, Geometry, English I, English II, and grades 3-8 ELA and math). After the readers take one of these calibration sets (usually 5-10 responses), the Scoring Director will review it from the front of the room using rubric language and the anchor responses to explain the reasoning behind each response's score. These sets do not have a passing requirement but are designed to help refocus readers on how to properly use the scoring guidelines to score responses. The Scoring Director or Team Leaders will provide individual feedback to any scorers in need of additional clarification based on their performance.

Handscoring Quality Control Reports

Scoring Summary Report Sample

EOC Biology Q000

Totals	Inter-rater Reliability				Score Point Distribution									
	2X	%EX	%AD	%NA	Total	%0	%1	%2	%3	%4	%B	%F	%R	%U
Current Handscore	**11,834**	**87**	**12**	**1**	**58,111**	**47**	**32**	**12**	**6**	**3**	**0**	**0**	**0**	**0**
7435	500	84	14	1	2,437	42	34	11	7	6	0	0	0	0
7443	512	87	13	0	2,751	50	37	10	3	0	0	0	0	0
7762	624	88	11	1	2,968	51	26	7	4	2	0	0	0	0
8103	812	87	13	0	4,237	47	30	10	7	6	0	0	0	0
8339	447	88	11	1	2,410	52	32	9	6	1	0	0	0	0

Scoring Summary Report Sample with AI (Reader ID #3)

G05 Writing Q00 **Conventions**

Totals	Inter-rater Reliability				Score Point Distribution										
	2X	%EX	%AD	%NA	Total	%0	%1	%2	%3	%B	%F	%N	%R	%T	%U
Current Handscore	**11,704**	**68**	**30**	**1**	**61,310**	**34**	**38**	**23**	**4**	**0**	**0**	**0**	**0**	**0**	**0**
3	5,852	68	30	1	55,458	33	39	24	4	0	0	0	0	0	0
27188	6	100	0	0	6	67	33	0	0	0	0	0	0	0	0
61011	383	68	34	1	383	39	40	18	3	0	0	0	0	0	0
64018	503	69	31	1	503	36	37	22	5	0	0	0	0	0	0

Reader Feedback Logs

Reader performance and intervention information will be tracked and updated in bi-weekly Reader Feedback Logs. These Reader Feedback Logs provide at-a-glance information about retraining actions taken with individual readers to ensure scoring consistency in regard to reliability, score point distribution, and validity performance. The logs address the following possible actions:

- Action 1—Includes one or more of the following: increase monitor rate, show and discuss examples of errant scores, pair scorer with a supervisor or stronger reader, provide additional review or training materials/recalibration
- Action 2—Rescoring of responses for which scores have not been handed off for reporting
- Action 3—Removal from scoring item

Below is an example of a Reader Feedback log:

Algebra I Q000									M/D/Yr
Reader	%EX Low	%NA High	Score Point Distribution Skewed	Validity %EX Low	Validity %NA High	Comments	Action 1	Action 2	Action 3
3782				●			●		
12860			●				●		
13296				●			●		
16070	●						●		
18961				●			●		

In addition to the Reader Feedback Logs, DRC will continue to provide the LDOE with handscoring quality control reports (the same cumulative scoring reports that we have provided in the past; samples are provided on page 22). The Scoring Summary reports show inter-rater reliability data and score point distribution information for each item (by part where appropriate).

Handling Unusual Responses

Nonscore Codes and Definitions

Responses that cannot be assigned a score based on the rubric will be assigned a nonscore code. When readers apply nonscore codes, the responses are automatically routed to DRC handscoring supervisors for validation. Responses that receive a nonscore code count as zero points toward student scores that display on reports. The nonscore code will display in the response string that is included in the file provided to the LDOE.

If readers suspect plagiarism but have no concrete evidence, they score the response and alert it for suspected plagiarism. These responses are sent to supervisors for additional investigation. When supervisors find evidence of student-student plagiarism, each of the associated responses is scored according to rubric requirements and processed as an alert. Responses with proven student-internet

plagiarism receive a score of 0 and are also processed as alerts. If supervisors cannot find proof of plagiarism, the scores are left as is and the response is sent on as an alert (see *Alerts* section below).

The nonscore codes that will be used for Spring 2018 scoring are the same codes that were established and used during previous administrations. The non-score codes and the grades/courses to which they apply are described below:

Nonscore Code Definitions

Nonscore Code	Explanation
B	Blank/no response
F	Response is not written in English (Math responses from Spanish forms will be scored by a Spanish-qualified math scorer.)
I	Insufficient amount of original writing to score
N	Don't understand/know
R	Refusal to respond
T	Off-topic
U	Incoherent, unintelligible, or undecipherable

Nonscore Codes by Grade/Course

Grade/Course	B	F	I	N	R	T	U
LEAP 2025 Algebra I, English I, English II, Geometry, and grades 3-8 ELA and math	✓	✓	N/A	✓	✓	✓	✓
EOC Biology	✓	✓	N/A	N/A	✓	N/A	✓
EOC English III	✓	✓	✓	N/A	N/A	N/A	✓
LEAP 2025 and EOC U.S. History, and grades 3-8 social studies ERs	✓	✓	✓	N/A	N/A	N/A	✓
LEAP 2025 and EOC U.S. History, and grades 3-8 social studies CRs	✓	N/A	N/A	N/A	N/A	N/A	N/A

Alerts

Scorers have the ability to apply an alert flag to specific student responses. These are responses that may indicate the possibility of teacher interference, plagiarism, or disturbing content (e.g., possible physical or emotional abuse, suicidal ideation, threats of harm to themselves or others, etc.). After setting the alert flag, which states the reason for the alert, and providing a brief description (as necessary), the reader will score the response according to the specific scoring guidelines for that item.

Likewise, PEG and IEA have the ability to detect specific alerts (described in detail later in the *Artificial Intelligence Scoring* section of this document). All alerted responses (whether identified by a human reader or by AI) are automatically routed to the Scoring Director who reviews the score and forwards

appropriate responses (including grade, subject, lithocode, item number, and reason for alert) to senior project staff and DRC's Project Management Team for review.

If it is concluded that a response warrants an alert, DRC Project Management will contact the LDOE with the student's LASID and post to the SFTP site the response information provided by the scoring staff. If it is determined that a void is required due to plagiarism, the LDOE applies an invalidation to the record in eDIRECT. At no point during this process do scorers, Team Leaders, or Scoring Directors have access to demographic information for any students participating in the assessment. Note that the alert status of responses is not passed on in data files.

Artificial Intelligence Scoring

As part of our comprehensive scoring solution, DRC uses two artificial intelligence (AI) scoring systems. Measurement Incorporated's (MI) Project Essay Grade (PEG) is used to score student extended responses to items in English III, U.S. History, and grades 5-8 Social Studies. Pearson's Intelligent Essay Assessor (IEA) is used to score student responses in English I, English II, and grades 5-8 ELA. Both MI and Pearson are recognized leaders in the field of performance assessment scoring (human scoring) and automated essay scoring.

AI Scoring – Measurement, Inc.

The items in the following table will be AI scored by MI during the Spring 2018 administration. AI scoring models were built for each of these items by MI in the fall of 2016 or fall of 2017 and followed the model-building process described below. (Model-building data for all items included on the spring 2018 test may be found in the Appendix.)

Course	Item Type	IDEAS ID
EOC English III	Writing Prompt	851368
EOC U.S. History	ER	848632
LEAP 2025 U.S. History	ER	894256
LEAP 2025 U.S. History	ER	894104
LEAP 2025 U.S. History	ER	892955
LEAP 2025 U.S. History	ER	894047
LEAP 2025 Grade 5 Social Studies	ER	807728
LEAP 2025 Grade 6 Social Studies	ER	804944
LEAP 2025 Grade 7 Social Studies	ER	805569
LEAP 2025 Grade 8 Social Studies	ER	809006

Model Building

For each model built, PEG analyzed a set of inputs that were randomly pulled from the training set itself, which is made up of approximately 2,500 examples of student field test responses scored by expert human scorers. Specifically, the training set was divided into two independent pieces:

- One set of response data was used to train the AI engine and produce the scoring model. This attributed to 85% of the training set (~2,125 responses).
- The remaining 15% of the training set (~375 responses) was then used to validate the resulting model.

A regression model was built by choosing a set of variables (e.g., grammar, punctuation, style, etc.) and using least squares Linear Regression to find a best-fit relationship based on the training set. An algorithm chose the initial set of variables and added to the set as needed to produce a good fit, by taking into account correlation statistics and multicollinearity. Once the model was built, it was then run against the validation set, so that it could be evaluated for accuracy. Training was complete once PEG's validation set scores agreed with the human scores; however, if this level of accuracy was not met, then further iterations of training (which may involve new parameterizations or new algorithms) were used to produce a different model with higher accuracy. This process was completed for each trait that needed to be scored.

To further understand the importance of the validation set, consider that one of the risks inherent in machine learning is over-fitting the data. This means that it is possible to home in on particular elements of the responses in training data in such a way that the model does not generalize well to unseen data. To mitigate this risk, PEG uses a hold-out validation strategy[2] in which a randomly chosen subset of the initial training data is set aside, never used in training, but used only to evaluate the generalizability of models trained from the remainder of the set.

Validation is implicit in PEG's model training and, so, is complete for any model in production. The essential element of the process is that the models are trained on a larger subset of the training sample (approximately 85%), then validated against an entirely separate smaller subset of the training sample (approximately 15%). What is critical about this process and all validation schemes used in PEG training is that the AI's agreement is always based upon samples the AI has not encountered during training. Put another way, the samples used to train are never the same as the samples used to validate. This maximizes generalizability and minimizes the chance for over-fitting.

---

[2] PEG's agreements are based on a hold-out validation set pattern, as opposed to a cross-validation pattern. Cross-validation was evaluated in the past, but MI has since learned that hold-out validation provides (1) equally valid models with a massive improvement in training time, as well as (2) an easy way to ensure that the validation set remains partitioned from the rest of the training set at all times.

Evaluation Metric

When PEG builds a model, it selects the model elements that maximize scoring accuracy for the data in question. Therefore, it is important to choose an agreement statistic on which PEG can optimize its models in such a way that the final model will exhibit reliable, accurate scoring. The inter-rater reliability of two human raters is often measured via perfect/adjacent agreement or the Pearson product-moment correlation coefficient (Pearson's *r*). However, these two metrics each have significant disadvantages. Perfect/adjacent agreement is highly influenced by the overall scale and underlying distribution of the "true" scores (Williamson &Breyer, 2012), while Pearson's *r* is insensitive to mean difference between raters (Schuster, 2004).

MI has found that using quadratic weighted kappa, which has become the industry standard for AI scoring, as the optimization and evaluation metric leads to the most reliable and accurate scoring. Quadratic weighted kappa as a metric can detect changes in mean difference and variance between raters and is therefore well suited for comparing the accuracy of AI scoring with that of human scoring, as well as measuring the agreement of two independent human raters. For the sake of clarity in the discussion below, the quadratic weighted kappa between PEG and Reader 1 is referred to as κω(PEG, R1) and quadratic weighted kappa between Reader 1 and Reader 2 is referred to as κω(R1, R2).

Even though quadratic weighted kappa performs well as an optimization metric, there are still some deficiencies in using it as an evaluation metric. Quadratic weighted kappa is far less influenced by the overall scale and underlying distribution of the "true" scores than perfect/adjacent agreement, but it does still display some sensitivity to those aspects of the data. In addition, while AI scoring can outperform human scoring with regard to scoring accuracy, the quality of the human scoring data has a significant impact on PEG's ability to accurately model the data. That is, a low κω(R1, R2) will usually lead to a low κω(PEG, R1). Because of these issues with sensitivity to scale and distribution of scores and being bound by the quality of the training data scores themselves, it is difficult to give a fixed number in all scales for what an acceptable value would be for κω(PEG, R1). In cases of four or more levels (e.g. a score ranging from 1-4, or broader) a κω(PEG, R1) of 0.7 has become a rule of thumb as a go-no-go metric. In these broader scales, a κω(PEG, R1) that is less than 0.7 to any significant degree is typically grounds for rejecting the item for AI scoring. In cases where this metric is 0.7 or above, the performance is usually considered satisfactory for AI scoring; however, other metrics such as those discussed in the next paragraph are often considered for additional information.

For instance, where the score range is smaller, such as binary (0-1) or ternary (0-2) ranges, the QWK is of more limited use, as QWK subtracts the rate of chance agreement which is quite high in the binary and ternary cases. In binary and ternary cases, the percent-exact and percent-adjacent agreements can be valuable additional metrics as they are exhaustive in these extremely-limited-range cases. Also useful in such extreme cases is to compare the human-machine agreement with the human-human agreement. In these cases the difference between κω(PEG, R1) and κω(R1, R2) can be used as an additional evaluation metric. MI defines that value as follows:

$$\Delta\kappa = \kappa\omega(\text{PEG, R1}) - \kappa\omega(\text{R1, R2})$$

When Δκ is positive, PEG's scores are more in agreement with Reader 1 than Reader 1's scores are in agreement with Reader 2. When Δκ is negative, the opposite is true; Reader 1 and Reader 2 show higher agreement levels than PEG and Reader 1. Of course, in both cases the absolute value of Δκ maintains its weight as a relative value between the two kappa values. That is, a larger Δκ means more separation between the two kappa values being compared.

The first phase of training is to maximize agreement between the PEG (machine) score and the final expert human score. If high agreement can be reached in this phase (for instance, a quadratic weighted kappa of ≥ 0.7), then the model is considered fit. The PEG team conducts secondary analysis such as this R1 vs. R2 analysis in cases where there is some question as to the fitness of the model – for instance, in a case in which PEG's quadratic weighted kappas are quite low, R1 vs. R2 analysis may be conducted to determine if the lack of agreement is a shortcoming of PEG's training, or if it is implicit in the data. This was not necessary in the current set, with the exception of the binary (i.e. zero-or-one) scores for some English traits. Analysis in this case showed not only that human-human quadratic weighted kappas in the training set were low, but, more to point, that random sets of such binary scoring showed similarly low quadratic weighted kappas. In this case, the low quadratic weighted kappa was simply an artifact of the definition of quadratic weighted kappa itself and no further R1 vs. R2 analysis was necessary.

Δκ is a good metric to quickly show how accurately PEG was able to score a set of data with respect to how accurate human raters are on the same data, but MI also reports other metrics that its clients may be more familiar with, such as perfect/adjacent agreement, Pearson's r, and standard mean difference. However, since PEG was optimized on quadratic weighted kappa, κω and Δκ are the best reflections of actual performance.

Scoring Responses with the AI Engine

The PEG AI scoring engine extracts and uses a large and proprietary set of linguistic feature metrics both during training and during production scoring. During training, PEG's models "learn" to represent the many complex and almost always non-linear relationships found between these linguistic features and the score points assigned by human experts. During production scoring, these same features are extracted from submitted responses. The previously trained models related to the item in question are then used to map these features to their predicted score points.

After PEG has been trained on a scored training set provided by DRC, it is available to receive batches of student responses in a mutually agreed upon format (XML or plain-text). The current preferred scoring method is to exchange XML documents via a web service. No static files are exchanged during this process. The web service supports discovery via Web Service Description Language (WSDL). The file transfer will be encrypted and will satisfy FERPA security requirements. Each record in the batch provides PEG with the student's response and a number of identifiers. The identifiers typically consist of a test ID that uniquely identifies the test, an item ID that uniquely identifies the prompt/item, and a FERPA-compliant student ID that uniquely identifies either the student or the student-test combination. The table on page 4 of the "DRC – Streaming Scoring" document (see Appendix) also contains information on identifiers.

When PEG receives the file, it processes the batch of responses and records the scores. Each record is specific to a student-test-item combination and will contain the item's score or a reason why it could not be scored (most commonly because the response is too short, or does not contain English). After the batch is processed, the scored records will be returned to DRC for reporting.

DRC will send files to MI daily. Scored files will typically be returned to DRC in 2 to 3 days; however, these timeframes are not definite, because they are dependent on numerous variables involved (e.g. number of responses submitted, number of different items, number of traits per item, the average response length, the standard deviation of response lengths, number of unique words submitted in each response, etc.).

Regardless of whether responses are scored by humans or machines, it is inevitable that scoring anomalies requiring human intervention will occur. Built into MI's automated scoring engine are a variety of triggers for identifying alert papers and responses in which it has low confidence. This is detailed later under "Identifying Responses for Human Review."

Quality Control of the AI Engine

The guidelines below are purposefully general as they have proven to be the best practice for training the PEG engine. The PEG team followed this standard procedure in the DRC/Louisiana project and attempted to maximize human-machine quadratic weighted kappa among all holdout sets.

PEG holds out a 15% set of training data for use in validation. This holdout set is not seen by the AI during training. Instead, once training is complete, the holdout set is submitted for test evaluation and PEG's output is compared to the known, human-expert scores. As discussed in "Evaluation Metric" above, the quadratic weighted kappa has proven to be the most valuable agreement metric in PEG's recent history; however, others (e.g., exact, adjacent, and any host of others) are also applicable.

This evaluation was performed along with model building prior to operational scoring, and the results were shared with LDOE and the TAC to demonstrate sufficient scoring accuracy by PEG. For details on these results, please see pages 46-50 in the Appendix.

Once training and model building is complete, the performance of any given model is essentially deterministic (so, for a precise, given input, the output is expected to be identical). The PEG team monitors the services for unexpected events (for instance physical damage to its cloud infrastructure), and handles any data flow issues (for instance, if the client was using a different item number during live scoring than was used during training) but the AI itself does not change during live scoring. When read-behind data becomes available to the PEG team (typically this is on an annual basis), it can be used to re-evaluate and, if necessary, retrain the existing models prior to the next season of use, but such changes do not happen during live scoring. As part of our continuous improvement cycle, the analysis of this data is on-going with no current end date (i.e., items are being reviewed on a rolling basis).

Identifying Responses for Human Review

Built into MI's automated scoring engine are a variety of triggers for identifying responses that require human review, including potential alerts (suspected plagiarism included) and potential nonscoreable

responses (e.g., responses that are primarily copied text, lack proper development, lack enough content to be scored, or are written in an unsupported language). Many of these triggers have client-configurable thresholds. These can be set to standard defaults and then modified as needed. Thresholds are generally deliberately conservative. DRC will work with LDOE content staff and MI to look at the responses that PEG identifies for human review to make sure the high and low copied text and minimum word count settings are set appropriately. (See page 31 for detailed information about these custom thresholds.)

Please note that all responses that are identified in the sections below for human review will be automatically forwarded to a DRC Scoring Director who will determine the correct score or nonscore code to apply to the response. The Scoring Director will provide the final, reported score (or nonscore) for these responses. If the Scoring Director needs assistance in determining the correct score or nonscore, DRC will work with LDOE content staff to ensure that the response is scored correctly.

Alert Detection System

PEG has a robust system for detecting potential alerts, which is described in detail in this section. When PEG detects the presence of alert language, this alone does not indicate that a response is unscorable. Therefore, unless the response is unscorable for some other reason, PEG will return scores as well as the alert status code of 500 (in cases of unscorable alerts, the status code is in the range of 501-599, inclusive). Regardless of the alert flag, any responses returned with a flag to DRC will be evaluated by the handscoring supervisory team who will determine if the response needs to be processed as an alert as described previously in this document (see *Handling Unusual Responses – Alerts*). When it is concluded that a response does warrant an alert, DRC Project Management will contact the LDOE with the student's LASID and post the response information to the SFTP site for LDOE's review.

PEG's Alert flagging system is a pattern-matching system, targeting phrases suggestive of violence towards self or others, drug or alcohol abuse, feelings of anxiety or depression or the use of weapons. This system is rules-based. It responds to concentrations of "alert language" detected within submissions. Typically, these are word counts of particularly violent or profane language often found in actionable alerts. (Such language may also be found in non-alert submissions, but PEG does not attempt to determine "intent" in these cases, rather it flags only the presence of detected verbiage.) PEG currently tracks two types of alert language that differ only in severity (e.g., a statement regarding a person "killing" is considered more severe than a statement regarding a person "beating up," but both are counted as forms of alert language). By default, PEG issues an alert flag if it encounters one instance of severe alert language or two instances of less-severe language. PEG may also issue an alert flag if high counts of profanity are found. By default, this is three instances of severely profane or five instances of less profane verbiage. Although this means that non-actionable alerts may also certainly be flagged, PEG's default settings are purposefully kept highly sensitive to alert language. These levels are configurable, however, so if the rate of return is too high or too low, adjustments can be made. For the responses that it cannot score, PEG returns a condition code to the test delivery system indicating why the response could not be scored (i.e., the response receives a tentative nonscore code that is reviewed by a Scoring Director and corrected if needed). The test delivery system can then route the flagged responses to DRC's performance assessment handscoring system. DRC will perform human handscoring for the limited number of responses that cannot be scored by AI.

With regards to the process and timing, the alerts detection is typically run in series with other essay analysis, so it is no slower (or faster) than a regular scoring. A batch of individually identified extended responses are posted to PEG's Streaming Scoring service, and at that point a response may be flagged as a potential alert. This flag takes the form of a "status code."

The rules are purposefully over-sensitive (they are more likely to give false positives than false negatives), so it is likely that the great majority of ER's flagged with a "5##" status code will not require actual intervention with the student; however, PEG is in no way capable of diagnosing this. Instead PEG just follows rules designed to sense and flag the use of language which has, in the past, been associated with alerts.

Identification of Nonscoreables

PEG's nonscoreable configurability includes the settings listed below, which can flag responses so that they are sent to DRC Scoring Directors who will determine the correct score or nonscore code to apply. These can be set to any threshold, with extreme values effectively disabling any given setting. These are the only non-scorable parameters which can be configured in this way. Each non-scorable setting relates to status codes and general rules surrounding of insufficiency and indecipherability as described below.

1. MIN_WORDS: this controls status code 200 and may correspond to the business concept of "Insufficient" (i.e., too-short response)
2. MIN_CORRECT_WORD: this controls the status code 220 and is similar to the business concept of "Indecipherable" (i.e., foreign words and non-words)
3. Copied Text Low: this control status code 605
4. Copied Text High: this controls status code 610

By adjusting each setting, PEG may impose a reasonable approximation of the scoring rules regarding Insufficiency and/or Indecipherability.

Once the scoring in the cloud is complete, the scores and statuses are sent back to the MI Delivery Service which then returns these scores and codes to DRC.

That entire process typically requires less than 100 hours (~4 days), and quite often takes less than a single day).

Identifying Copied Text and Plagiarism with the AI Engine

Prior to describing the functionality PEG uses to detect copied text and plagiarized responses, an important distinction must be made between what is considered copied and what is considered plagiarized. Copied text is that which a student copies from the directions, prompt, passage(s), or reference sources supplied with an item. A response composed predominately of text copied from item sources will not be alerted for any sort of suspected testing violation, but in most cases, it will receive a lower score (or a nonscore of “I”) depending on the amount of original student writing there is in the response and/or how much text is copied. Responses flagged by PEG for this condition are sent to DRC scoring supervisors for review. Based on this review, English III responses having an insufficient amount

of original writing to score will receive a nonscore of "I." For U.S. History, any response having an insufficient amount of original writing to score, because it is made up entirely or almost entirely of text copied from the directions or reference sources, will receive a score of 0 in the case of EOC ERs, or in the case of LEAP 2025 ERs, a 0 in Content and a 0 in Claims (unless the item-specific rubric makes exceptions for the use of relevant copied text).

Text that a student extracts and uses from a source external to the test itself is considered plagiarized. When PEG detects these responses, they are also sent to DRC scoring supervisors for review, and if they are deemed to warrant an alert for suspected plagiarism, DRC's supervisors route the responses through the same alert process described in an earlier section of this document (Handling Unusual Responses – Alerts).

PEG's copied text and plagiarism detection functionality compares student responses to texts that students may have copied or plagiarized. To do this, per-item reference texts must be provided. For English III, this is the prompt and any associated reading material provided with each test item. For the U.S. History and Social Studies ERs, this includes the prompt and any appropriate source material (including MC/MS items) provided with each test item. In addition to external sources of plagiarism previously provided by LDOE based on results from past administrations, DRC will pre-identify other websites that may be likely sources of external plagiarism. These may include Wikipedia's pages relevant to the topic and/or other "top hit" websites. These external sources will be used by the AI engine to identify potentially plagiarized responses. All of these text references will be added to the appropriate scoring models for each related item.

Upon receiving a response, PEG conducts a high-speed sequence scan of both the reference text and the response. Each sequence is evaluated for both the length and density of copied/plagiarized text. Length is a direct character count, and density is a measure of similarity between sequences. A verbatim copy has a density of 1.0, and a copy that contains some substitutions, additions, or deletions would likely have a density in the ~0.6 - 0.4 range. The product of these two numbers provides a value that is used to flag responses requiring human review due to large amounts of copied/plagiarized text. Clients can configure two thresholds for a low and high flag. For example, the default values for these are 50 and 100 respectively. So, a verbatim copy of 72 characters (~12 prompt words) would be reported as a low match, whereas a verbatim copy of 100 characters (roughly 16 words) would be flagged as a high match. Similarly, a copy (even with some substitutions) of 40 words would still be reported as a high match in the default setting example. The low and high matches will be flagged with status codes. This is similar to the alert flagging above. There will be a three-digit code for low-match (status code 605) and a three-digit code for high-match (status code 610).

Custom thresholds for copied text, plagiarism, and insufficient responses have been established by DRC in consultation with LDOE and were based on recommendations from MI. They are described below:

1. When PEG scans responses for copied text/plagiarism, any text copied from the supplied reference texts (regardless of whether it is contained within quotations marks) will be considered when determining if a response meets or exceeds the thresholds required for it to be routed to DRC for human review. These threshold configurations are noted in 2a–4b below.

2. English III
    a. Copied text thresholds
        i. Low flag (status 605) – 125 characters
        ii. High flag (status 610) – 200 characters
    b. MIN_WORDS (status 200) – 45 words or fewer

3. Grades 5-8 Social Studies
    a. Copied text thresholds
        i. Low flag (status 605) – 125 characters
        ii. High flag (status 610) – 200 characters
    b. MIN_WORDS (status 200) – 25 words or fewer

4. U.S. History
    a. Copied/plagiarized text thresholds
        i. Low flag (status 605) – 85 characters
        ii. High flag (status 610) – 170 characters
    b. MIN_WORDS (status 200) – 25 words or fewer

These settings are deliberately conservative. While some flagged responses are composed exclusively of text copied directly from source/passage material, the majority of responses that PEG flags with status codes 605 and 610 contain a combination of copied text, relevant information cited or paraphrased from the sources, and some amount of original student writing. They are flagged because they meet or exceed the copied text thresholds noted above and need to be checked by DRC scoring supervisors to determine whether they contain a sufficient amount of original student writing to evaluate. Upon review, most will be found to contain enough original writing to be considered scorable. When the supervisor determines that there is sufficient original student writing to score, and there is no evidence of plagiarism, he or she validates the original numeric scores returned by PEG and they are submitted as final scores for that response. On the other hand, if the supervisor determines that the response contains insufficient original student writing to evaluate, he or she will override PEG's original numeric scores and apply the appropriate numeric scores or nonscore as necessary, depending on the content area scoring rules. For Social Studies and U.S. History, flagged responses composed entirely of text copied from item source material (or all copied text combined with an insufficient amount of original student work) are scored as 0s. For English III, such responses are given a nonscore of "I" (Insufficient).

Less frequently, responses will be flagged as potential nonscores for having too little written to be evaluated at all (status code 200). Just as DRC requires all nonscores given by human readers to be reviewed by scoring supervisors, this same requirement holds true when PEG flags responses as potential nonscores. For example, if the DRC supervisor reviews a response flagged by PEG and agrees with PEG's assessment that the response has too little writing to be assessed, the supervisor will validate the AI score of "I," and this nonscore code will be submitted as the final score for that response. On the other hand, if DRC's supervisor reviews the response, and based on the training responses provided in the handscoring training materials, he or she feels that that there is enough original student writing to score, the supervisor scores the response and also overrides PEG's original nonscore, changing PEG's nonscore of "I" to the correct numeric scores. These become the scores of record.

AI Scoring – Pearson

The items in the following table will be AI scored by Pearson during the Spring 2018 LEAP 2025 administration. AI scoring models for each of these items were previously built and used in operational scoring by Pearson during PARCC operational scoring in the spring of 2017. (Model-building data for all items included on the spring 2018 test may be found in the Appendix.)

Course	Item Type	IDEAS ID	PARCC UIN
English I	PCR	902152	LAT_VH017536
English I	PCR	902161	RST_VH017542
English II	PCR	906197	LAT_HH428127697
English II	PCR	902331	RST_VH004490
Grade 5 ELA	PCR	913663	NWT_CC434542756
Grade 6 ELA	PCR	913694	NWT_1466
Grade 7 ELA	PCR	868704	RST_VH069361
Grade 7 ELA	PCR	913842	NWT_EE430133306
Grade 8 ELA	PCR	913958	LAT_F1460

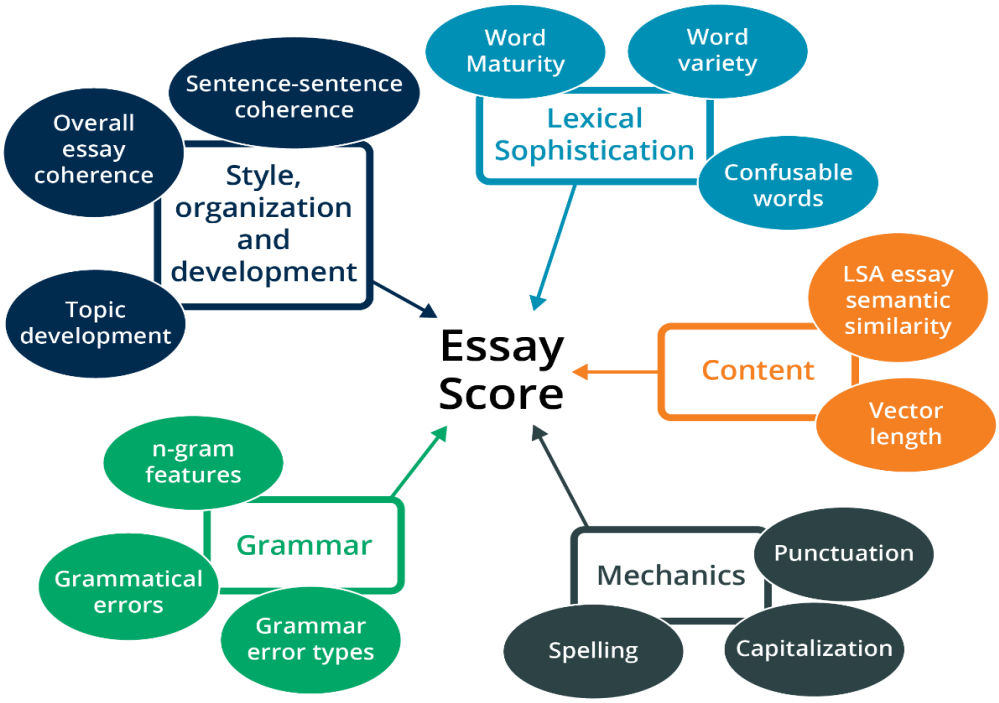
The Intelligent Essay Assessor

Pearson's Intelligent Essay Assessor (IEA) uses a range of machine learning and natural language processing technologies to learn to score based on human-scored responses. One of the hallmarks of IEA is its ability to score constructed responses in content domains using Pearson's unique implementation of Latent Semantic Analysis (LSA), an approach that generates semantic similarity of words and passages by analyzing large bodies of relevant text. LSA can then "understand" the meaning of text much the same as a human scorer.

IEA's background knowledge of English is derived from a collection of texts equivalent to what students are likely to have encountered over the course of their academic career (about 12 million words). Because LSA operates over the semantic representation of texts, rather than at the individual word level, it can evaluate similarity even when texts have few words in common. For example, LSA finds the following two sentences to have a high degree of semantic similarity:

- Surgery is often performed by a team of doctors.
- On many occasions, several physicians are involved in an operation.

The following figure illustrates some of the features used in IEA and how they relate to specific constructs of student writing performance.



Example features used in the Intelligent Essay Assessor. *Like human scorers, IEA evaluates essays for ideas, organization, development, and various grammatical and mechanics errors.*

IEA is trained to associate features extracted from each essay with scores assigned by human scorers. A machine learning-based approach is used to determine the optimal set of features, and the weights for each of those features, to best model the scores for each essay. From these comparisons, IEA derives a prompt- and trait-specific scoring model that predicts the scores human scorers would assign to any new responses.

The automated scoring process mimics the approach that human scorers take when evaluating essays. Human scorers train based on anchors of annotated student responses with agreed-upon scores. Human scorers compare new responses against the anchor set of two to three examples per score point to determine the appropriate score. IEA scores essays similarly, but makes comparisons against a much larger set of examples. Rather than comparing a new essay against the 16-24 examples in an anchor set, it compares against the set of hundreds or thousands of responses on which it was trained.

How the Intelligent Essay Assessor was Trained

For the ELA prompts that will be used by Louisiana, IEA was trained based on operational PARCC responses using Pearson's Continuous Flow approach to training and scoring. When these prompts were first administered, student responses flowed to IEA even before human scoring started. IEA then selected a sample of responses for humans to score first to expedite the creation of automated scoring models. The sample included responses that represented different demographic subgroups to ensure

equity in scoring, as well as responses that were algorithmically selected to likely span the score range. As the human-scored responses flowed back to IEA, the engine automatically built potential scoring models, evaluating them against the industry standards for performance criteria included in the table below.

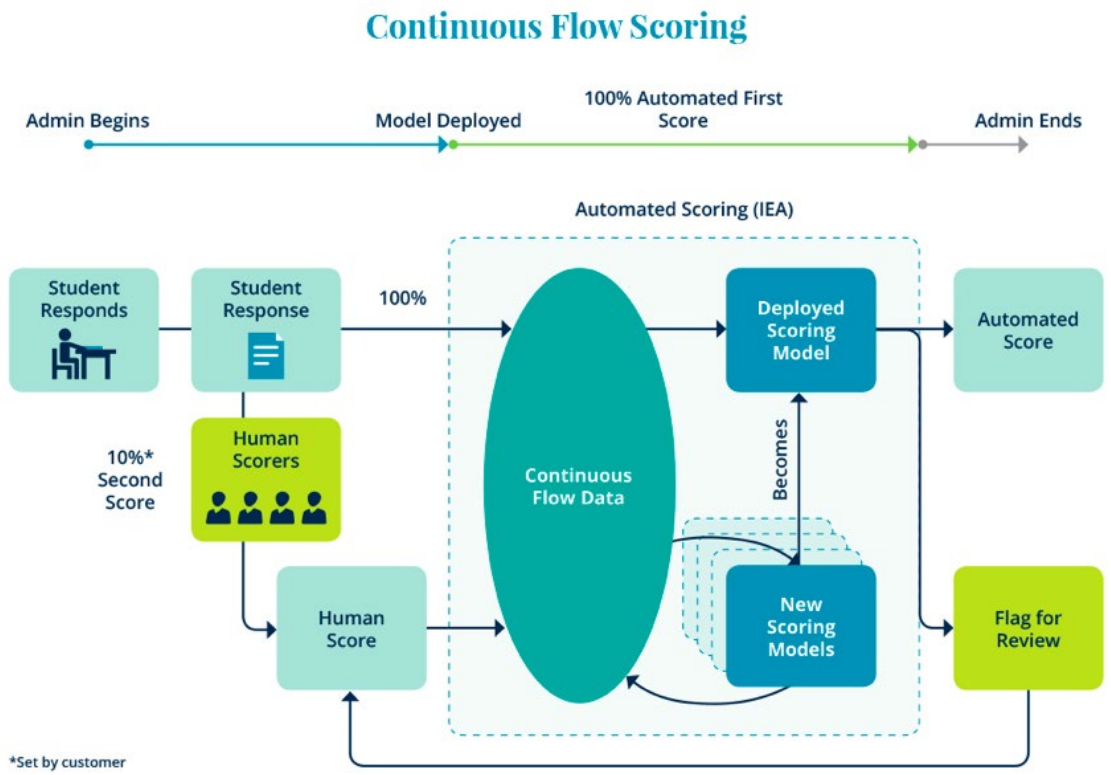
Evaluation of Automated Scoring Systems	
Criterion	**Threshold**
Quadratic weighted kappa (QWK)	Greater than or equal to 0.70
Pearson correlation (r)	Greater than or equal to 0.70
Standardized mean difference (SMD) between human and automated scoring	Less than or equal to|0.15|
Difference in QWK or r from human-human rates	Less than or equal to 0.10
Difference in exact agreement from human-human rates	Less than or equal to 0.05

Evaluating Automated Scoring. *Statistical Criteria for the Evaluation of Automated Scoring Systems based on those used by Williamson et al, Smarter Balanced, and PARCC.*

While the engine was being trained, scoring and psychometrics teams met daily to review progress, quality, and next steps. When IEA met or exceeded the performance criteria for a given constructed response item, it took over as the first scorer for that item.

Responses for which IEA is less confident in its score are routed for additional human scoring. This "smart routing" of responses by the scoring engine occurs when responses fall in a particular score range for which the engine has lower agreement with human scorers, or for responses that are highly unusual or creative.

The figure below depicts the entire Continuous Flow process.



Continuous Flow. *As student responses flowed to IEA, it selected responses for human scorers to score. As the human scores flowed back to IEA, the engine continued to try to build a scoring model that would pass the agreed upon performance criteria. Once the scoring model passed the criteria, it was deployed and began scoring all student responses, with humans applying a second score as a quality check, as well as scoring any responses flagged for review by IEA.*

IEA is also trained to recognize a variety of different non-responses (e.g., non-English language, "don't understand," refusal to answer, off-topic, unintelligible), assigning corresponding condition codes to them or flagging them for human review when less certain. Detection of copying between students is done out of band and accomplished by using Latent Semantic Analysis to compare each student response to every other student response and flagging highly similar responses for human review. Child in danger alerts are also scanned for out of band and flagged for human review.

Quality Monitoring

Human scorers play a key role in maintaining quality throughout the scoring process starting with IEA learning to score based on their scores. Since the models for the 2018 Louisiana items are built and IEA has already established the performance characteristics necessary to accomplish first scoring, DRC human scorers will score 10% of the responses scored by IEA to monitor quality. Should agreement rates between IEA and the human scorers fall below the established agreement rates, the automated scoring model can be examined to determine the appropriate action. This action may include adjusting IEA's confidence threshold to send more responses for human scoring or retraining the scoring engine and rescoring student responses.

Scoring (DRC)

DRC will use human scorers to read behind MI and Pearson's AI engines. Ten percent of the AI-scored student responses will be randomly selected to be read a second time by DRC's handscoring teams. This will provide inter-rater reliability statistics that compare the scores given by PEG and IEA to the scores given by each individual reader. Throughout the handscoring process, DRC Project Managers, Scoring Directors, and Team Leaders will review handscoring reports detailing these results.

If the inter-rater reliability (AI compared to handscoring on the 10% sample) shows exact agreement that is less than desired or nonadjacent agreement that is higher than desired, DRC will investigate and take immediate action. If scoring patterns are apparent among individual readers, Team Leaders will deal with issues of this sort on an individual basis. If a reader appears to need clarification of the scoring rules, DRC supervisors typically monitor one out of five of the scorer's readings, making adjustments to that ratio as needed. If a supervisor disagrees with a reader's scores during monitoring, he or she will provide retraining in the form of direct feedback to the reader, using rubric language and applicable training responses.

If, however, the agreement rates for either PEG or IEA and for large numbers of readers are not as anticipated, DRC scoring experts will need to review the responses that received different scores from the AI engine(s) and from readers. Based on this, the DRC scoring experts will need to determine if they feel that the readers need to be retrained or if they are disagreeing with scores given by AI. In the unlikely scenario that DRC's scoring experts believe that they have detected unexpected trends in the scores given by PEG or IEA, DRC would take examples to LDOE and the appropriate AI vendor to review. Based on this review, if DRC, LDOE, and the vendor determined that the AI modelling was not resulting in sufficiently accurate scores, corrective measures would be put into place. Depending on the nature and timing of the issue and subsequent related LDOE policy decisions, DRC and the AI vendor will enact measures such as updating the AI modeling, providing LDOE with response information (e.g., Item ID, Student IDs, updated scale scores, updated achievement levels), and/or using expert handscorers to determine the final score for subsets of student responses. Time required for alterations of the AI models varies based upon the extent of the necessary correction. A configuration change (for instance in a non-scorable parameter) can typically be deployed within 1-2 days. A complete model retraining (for instance, if it were discovered that all original human training data were invalid) could require 1-2 weeks from receipt of the new training data.

Rescores (High School Tests Only)

The rescoring process includes automatic rescores that occur during the scoring process, as well as parent-requested rescores that take place after the official scoring window.

The rescores for all subjects will be performed by expert readers. If the rescore results in a final score that falls in the *Fair* achievement level for EOC tests or the Approaching Basic achievement level for LEAP 2025 HS tests, the rescore results are sent to LDOE content specialists for review. If LDOE agrees with the rescore, the original score is replaced with the rescore and the score report is produced and sent. Note: The score reports are always held until the rescore process is complete.

Auto Rescores

In Window: Automatic rescores will occur per the criteria listed below when all sessions for a test have been completed.

Post Window: Automatic rescores will occur per the criteria listed below regardless of the number of sessions completed.

Biology, U.S. History (EOC)

Students are eligible for an automatic rescore when both of the following are true for the subject:

Ref. #	Rule/Requirement
1	Reporting Grade = '12'
2	Subject Scale Score within 10 points below the Fair (F) Achievement Level cut score

Algebra I, Geometry, U.S. History (LEAP 2025)

Students are eligible for an automatic rescore when both of the following are true for the subject:

Ref. #	Rule/Requirement
1	Reporting Grade = '12'
2	Subject Scale Score within 10 points below the Approaching Basic (APP) Achievement Level cut score

English III

Students are eligible for an automatic rescore when both of the following are true for the subject:

Ref. #	Rule/Requirement
1	Reporting Grade = '12'
2	Subject Scale Score within 20 points below the Fair (F) Achievement Level cut score

English I and English II

Students are eligible for an automatic rescore when both of the following are true for the subject:

Ref. #	Rule/Requirement
1	Reporting Grade = '12'
2	Subject Scale Score within 20 points below the Approaching Basic (APP) Achievement Level cut score

Students that are automatically rescored will be flagged on the LDESTD file as "<Subject> Rescore Flag" = "P" while the rescore is being processed and "R" after the rescore has been completed. Refer to the **Rescore Flag** section of the LEAP 2025 High School and EOC Processing Rules - Scoring document for the rules to set the Rescore Flag. The original scale score will be maintained in the student record for all rescores that did not result in an achievement level change. The score, original or rescore, approved by LDOE after reviewing the Rescore File will be maintained on the student record for all rescores that resulted in an achievement level change.

All tests that require an automatic rescore will be held back from reporting until the rescore process is either complete with no achievement level change or the rescore process resulted in an achievement level change which was reviewed and approved by LDOE.

Students whose scores have been voided are not eligible for an automatic rescore.

Parent-Requested Rescores

Post Window: Parent-requested rescores are processed by DRC project management.

Biology, U.S. History (EOC)

Students are eligible for a parent-requested rescore when both of the following are true for the subject:

Ref. #	Rule/Requirement
1	Reporting Grade <> '12'
2	Subject Scale Score within 10 points below the Fair (F) Achievement Level cut score

Algebra I, Geometry, U.S. History (LEAP 2025)

Students are eligible for a parent-requested rescore when both of the following are true for the subject:

Ref. #	Rule/Requirement
1	Reporting Grade <> '12'
2	Subject Scale Score within 10 points below the Approaching Basic(APP) Achievement Level cut score

English III

Students are eligible for a parent-requested rescore when both of the following are true for the subject:

Ref. #	Rule/Requirement
1	Reporting Grade <> '12'
2	Subject Scale Score within 20 points below the Fair (F) Achievement Level cut score

English I and English II

Students are eligible for a parent-requested rescore when both of the following are true for the subject:

Ref. #	Rule/Requirement
1	Reporting Grade <> '12'
2	Subject Scale Score within 20 points below the Approaching Basic (APP) Achievement Level cut score

Student responses that are rescored will be flagged on the LDESTD file as "<Subject> Rescore Flag" = "P" while the rescore is being processed and "R" after the rescore has completed. Refer to the **Rescore Flag** section of the LEAP 2025 High School and EOC Processing Rules - Scoring document for the rules to set the Rescore Flag. The original scale score will be maintained in the student record for all rescores that did not result in an achievement level change. The score, original or rescored, approved by LDOE after reviewing the Rescore File will be maintained on the student record for all rescores that resulted in an achievement level change.

Students whose scores have been voided are not eligible for a parent-requested rescore.

Appendix A

DRC-MI Streaming Scoring Documentation

DRC – MI STREAMING SCORING SUBMIT SERVICE DOCUMENTATION

NOTICE: The contents of this document and any references to external resources are intended for review only by representatives of Data Recognition Corporation, Measurement Incorporated, and LDOE, and are considered private. Technical specifications are subject to change.

REVISED: 2015-11-23; *created*

CONTENTS:

SECTION 1 – General Information 43

SECTION 2 – SCHEMA SUPPLEMENT 44-46

SECTION 3 – STATUS CODE INFORMATION 47

SECTION 1 – General Information

1.1 PURPOSE: Submit Service accepts groups ("batches") of constructed responses for processing by the MI Streaming Scoring product.

1.2 SERVICE TYPE: The Submit Service uses a standard SOAP web service interface.

1.3 INTEGRATION: Application-generated service definition (WSDL 1.1) document is available; WCF (Windows Community Foundation) client integration is also possible. The WSDL and WCF URLs for each environment are as follows:

DEVELOPMENT
- **WSDL:**
- **WCF:**

STAGING
- **WSDL:**
- **WCF:**

PRODUCTION:
- **WSDL:**
- **WCF:**

1.4 SERVICE SIGNATURE: The Submit Service provides a single operation **SubmitBatch**. The operation signature – request and response structure – is defined in the WSDL. The structure of each complex type, with field descriptions and expected value ranges is described below.

SECTION 2 – SCHEMA SUPPLEMENT

2.1.1 SUPPLEMENTAL SCHEMA DOCUMENTATION: The following tables are supplemental to the schema for the Submit Service, but are not, themselves, the schema. The service schema is contained within the WSDL, and may be emitted from that source to an XML schema document (XSD) through various means, though this will likely be unnecessary. To reduce confusion in terminology, the following tables will be referred to as the "supplement" or "schema supplement".

2.1.2 TABLE STRUCTURE: Each table documents a specific complex type defined by the Submit Service WSDL, with each row in a table representing a field of that complex type. Column definitions are provided here.

- **Name:** Name of field; note that for complex type fields, the name of the field and the name of the type may, or may not be the same.
- **Type:** Field type; this may be a simple type (string, integer, etc.) or another complex type, which is described in another table.
- **Min:** Minimum expected occurrences (minOccurs). This value with be either 1 or 0 for all fields. For fields with 0 minOccurs, that field may be omitted from the complex type, and it will still be schema-compliant. Omitting a field may still cause an application-level error due to invalid data, refer to the **Range** column for application-level constraints.
- **Max:** Maximum expected occurrences (maxOccurs). This value will usually be 1 or *unbounded.* Unbounded fields/elements may appear multiple time within the complex type, which allows for list-like data structures within the service. While there is no theoretical upper limit to the number of occurrences, some constraints are enforced at the application level. See the **Range** column for more information.
- **Description:** This column defines the field's purpose.
- **Range:** Application-enforced constraints on a field's value are given here. If the field has a minOccurs of 0 in the schema, but is expected to be included by the application, it will be designated *required* in this column. Fields with a maxOccurs of *unbounded* within the schema with an application-enforced limit will be described here. Strings will have their maximum expected length defined here, if any.

2.2.1 SubmitBatch (REQUEST ELEMENT)

Name	Type	Min	Max	Description	Range
request	SubmitBatchRequest	0	1	Application-defined request element	*Required.*

2.2.1 SubmitBatchRequest

Name	Type	Min	Max	Description	Range
BatchId	string	1	1	DRC Batch ID; no validation performed by MI	Max length 50; longer values will be truncated.
ClientId	string	1	1	MI-Assigned client/project identifier; other projects sharing the environment will be assigned separate ClientIds.	Only values provided by MI will be accepted.
ConstructedResponses	ConstructedResponseList	0	1	List of constructed response elements to be scored for this batch	*Required.*

2.2.2 ConstructedResponseList

Name	Type	Min	Max	Description	Range
ConstructedResponse	ConstructedResponse	0	*unbounded*	List of individual CRs to be scored	*Required.* Missing or zero-length lists will not be entered for scoring. Lists exceeding 2000 CRs will also be rejected.

2.2.3 ConstructedResponse

Name	Type	Min	Max	Description	Range
EssayText	string	1	1	Student-generated response text.	This field is technically nillable, though nil or zero-length essays will not be scored. The field also technically has no max length, but essays exceeding 30,000 characters will also not be scored. Description codes will be returned for each of these cases.
ItemId	string	1	1	Identifier for Item/prompt	Responses that do not have a valid ItemId will not be scored; the range and convention for ItemIds are defined by DRC and MI.
ResponseId	string	1	1	DRC constructed response ID; no validation performed by MI	Max length 256; longer values will be truncated.

2.3.1 SubmitBatchResponse (RESPONSE ELEMENT)

Name	Type	Min	Max	Description	Range
SubmitBatchResult	SubmitBatchResult	0	1	Application-defined result element	*Required.*

2.3.2 SubmitBatchResult

Name	Type	Min	Max	Description	Range
BatchId	string	1	1	DRC batch ID as stored by MI (same value given in request)	Value may be truncated if it exceeds 50 characters
ClientId	string	1	1	MI-assigned client identifier (same value given in request)	
MIBatchId	ser:guid	1	1	MI-generated Batch ID	ser:guid is an extension of string, bounding the expected value to a Guid data type. It may be treated as a string or parsed to a Guid by the client.
StatusCode	StatusCode	1	1	Application-generated response code indicating success/failure of operation	

2.3.3 StatusCode

Name	Type	Min	Max	Description	Range
Code	integer	0	1	Numeric status code	*Required.* Will fall in the range 0-999. See **section 3** for more information
Description	string	0	1	Short description of status	*Required.* See **section 3** for more information

SECTION 3 – STATUS CODE INFORMATION

3.1 STATUS CODES: Each SubmitBatch response will contain a status code indicating success or failure in adding the batch to the Streaming Scoring system. Individual CRs processed by Streaming Scoring will also receive similarly structured Status Codes upon delivery, albeit with similar values. Note that lower-level errors will not receive application-generated responses, and therefore will not be given status codes. These types of errors include (but are not limited to): malformed requests (which violate the schema), service unavailable, and TCP/HTTP errors. Expected status codes and their description for the SubmitBatch operation can be found in the following table.

3.2 SubmitBatch STATUS CODES

Code	Description	Notes
0	SUCCESS	Batch successfully accepted and queued for scoring.
100	INVALID_CLIENT_ID	ClientId value in request is not valid.
120	NO_REQUEST_DATA	request element is nil or missing.
140	NO_ESSAY_DATA	ConstructedResponses element is missing or contains zero CRs.
150	BATCH_TOO_LARGE	ConstructedResponses element contains more than 2000 CRs.
190	INTERNAL_ERROR	An unexpected internal error occurred at the application level.

3.3 Individual CR STATUS CODES

Code	Description	Notes
200	too few words (configurable)	blank or extremely short response; response sent to DRC for Supervisor Review
220.	not enough correctly spelled words (configurable)	"Indecipherable" (i.e., foreign words and non-words); response sent to DRC for Supervisor Review
400	unexpected item_id	the item_id is not one of the items PEG AI has modeled; potential set-up issue to be resolved between MI and DRC
500	Alert, otherwise same as 0, above	alerted response sent to DRC for Supervisor Review
520	Alert, otherwise same as 200, above	alerted response sent to DRC for Supervisor Review
522	Alert, otherwise same as 220, above	alerted response sent to DRC for Supervisor Review
530	Alert, otherwise same as 300, above	alerted response sent to DRC for Supervisor Review
540	Alert, otherwise same as 400, above	the item_id is not one of the items PEG AI has modeled; potential set-up issue to be resolved between MI and DRC; alerted response sent to DRC for Supervisor Review
605	copied text low threshold (configurable)	sent to DRC for Supervisor Review
610	copied text high threshold (configurable)	sent to DRC for Supervisor Review
900	timeout	unable to complete essay score prediction within time limits; sent to DRC for Supervisor Review
950	system error processing essay	internal PEG error

Appendix B

AI Model Building – English III (Spring 2018)

Quadratic Weighted Kappa (QWK) and Inter-rater Reliability (IRR)

Course	Form	Item #	Rater	Content				Style				Sentence Formation			Usage			Mechanics			Spelling		
				QWK	EX	ADJ	NON-	QWK	EX	ADJ	NON-	QWK	EX	ADJ	QWK	EX	ADJ	QWK	EX	ADJ	QWK	EX	ADJ
English III	U	851368	H-H*	0.78	65	33	2	0.74	65	34	1	0.26	76	24	0.44	74	26	0.53	77	23	0.47	85	15
			AI-H**		71	29	0		66	33	0		76	24		74	26		82	18		85	15

*Human to human (H-H) inter-rater metrics are from Pacific Metrics EFT scoring.
**Human to AI (AI-H) inter-rater metrics are from the MI 2016 model building results.

Score Point Distribution (SPD)

Course	Form	Item #	Rater	Content				Style				Sentence Formation		Usage		Mechanics		Spelling	
				1%	2%	3%	4%	1%	2%	3%	4%	0%	1%	0%	1%	0%	1%	0%	1%
English III	U	851368	H	12	42	37	9	4	39	46	10	20	79	34	66	26	74	16	83
			AI	12	42	38	8	7	37	46	10	22	78	36	64	26	74	17	83

AI Model Building – EOC U.S. History ER (Spring 2018)

Quadratic Weighted Kappa (QWK), Inter-rater Reliability (IRR), and Score Point Distribution (SPD)

Course	Form	Item #	QWK Score Agreement	IRR Comparison	Inter-Rater Agreement %			SPD Group	Score Point Distribution %				
					EX	ADJ	NON-ADJ		0s	1s	2s	3s	4s
USH	E	848632	0.75	Human to Human*	60	35	4	Human	39	42	15	3	1
				AI to Human**	60	37	4	AI	39	41	15	3	2

*Human to human inter-rater metrics are from Pacific Metrics EFT scoring.
**AI to human inter-rater metrics are from the MI 2016 model building results.

AI Model Building – LEAP 2025 U.S. History ERs (Spring 2018)

Quadratic Weighted Kappa (QWK), Inter-rater Reliability (IRR), and Score Point Distribution (SPD)

Course	IDEAS Item #	# of Responses	Content											Claims										
			QWK	Inter-Rater Agreement %				Score Point Distribution %						QWK	Inter-Rater Agreement %				Score Point Distribution %					
				Comparison	Exact	Adjacent	Nonadjacent	SPD Group	0s	1s	2s	3s	4s		Comparison	Exact	Adjacent	Nonadjacent	SPD Group	0s	1s	2s	3s	4s
USH	894104	2500	0.86	H to H	62	33	5	Human	31	34	22	9	4	0.84	H to H	61	32	7	Human	39	28	21	9	4
		15%		AI to H	70	29	1	AI	30	38	21	8	3		AI to H	63	36	1	AI	39	28	21	8	3
USH	894256	2500	0.85	H to H	66	32	2	Human	28	37	24	8	3	0.86	H to H	64	33	3	Human	34	33	21	8	3
		15%		AI to H	71	29	1	AI	24	37	29	8	2		AI to H	69	30	1	AI	34	34	22	8	3
USH	892955	2500	0.88	H to H	65	32	3	Human	34	29	25	9	3	0.88	H to H	64	32	4	Human	37	26	25	10	3
		15%		AI to H	74	26	0	AI	31	34	24	9	2		AI to H	72	28	0	AI	37	28	22	10	3
USH	894047	2500	0.86	H to H	66	31	3	Human	39	29	19	9	4	0.86	H to H	66	30	4	Human	40	29	19	9	4
		15%		AI to H	71	28	1	AI	40	30	20	7	4		AI to H	73	26	1	AI	44	29	18	5	3

Human to human metrics are from DRC EFT scoring in Spring 2017.
AI to human metrics are from the MI 2017 model building results.

- AI models were built in Fall 2017
- Included 2,500 responses from the Spring 2017 EFT
- Responses scored using DRC developed training materials
- 100% were scored by a second human reader and adjacent scores were resolved

AI Model Building – Social Studies Grades 5-8 ERs (Spring 2018)

Quadratic Weighted Kappa (QWK), Inter-rater Reliability (IRR), and Score Point Distribution (SPD)

Grade	IDEAS Item #	# of Responses	Content											Claims										
			QWK	Inter-Rater Agreement %				Score Point Distribution %						QWK	Inter-Rater Agreement %				Score Point Distribution %					
				Comparison	Exact	Adjacent	Nonadjacent	SPD Group	0s	1s	2s	3s	4s		Comparison	Exact	Adjacent	Nonadjacent	SPD Group	0s	1s	2s	3s	4s
5	807728	2599	0.87	H to H[1]	76	23	1	Human	61	23	13	3	0	0.87	H to H[1]	78	20	2	Human	66	20	11	2	0
		≈500		H to H[2]	65	32	2	Human	19	36	34	10	1		H to H[2]	67	31	3	Human	25	32	35	8	1
		15%		AI to H	70	29	1	AI	32	32	27	7	1		AI to H	70	29	1	AI	39	29	24	7	1
6	804944	2975	0.85	H to H[1]	73	25	1	Human	61	32	7	1	0	0.83	H to H[1]	75	24	1	Human	64	31	5	1	0
		≈500		H to H[2]	68	29	2	Human	19	42	34	5	1		H to H[2]	69	28	3	Human	23	43	29	4	1
		15%		AI to H	77	23	0	AI	39	39	17	3	0		AI to H	75	24	1	AI	44	39	14	3	1
7	805569	2610	0.89	H to H[1]	68	31	1	Human	32	42	21	4	0	0.90	H to H[1]	66	32	1	Human	35	41	20	4	0
		≈500		H to H[2]	76	24	0	Human	1	30	53	15	1		H to H[2]	74	26	1	Human	4	33	50	13	1
		15%		AI to H	75	25	0	AI	30	30	30	9	1		AI to H	74	26	0	AI	32	31	28	7	1
8	809006	2543	0.84	H to H	63	25	2	Human	27	38	24	8	2	0.86	H to H	62	36	2	Human	33	33	24	8	2
		15%		AI to H	69	31	1	AI	26	38	27	7	3		AI to H	71	29	0	AI	34	33	25	8	1

H to H[1] – Human scored 2016 Field Test sample of ≈ 2500 responses per item.

H to H[2] – Human scored 2018 targeted sample of ≈ 500 responses per item used to augment and retrain the AI models. This sample came from spring 2018 operational responses and was intended to find high score points to add to the existing AI models for the purpose of retraining the models prior to spring 2018 operational scoring.

AI – Data based on holdout subsets chosen by stratified random sampling from the full ≈ 3000 per item response count (2016 FT and 2018 sample) and excluded from the training process.

The existing AI scoring model for grade 8 (809006) from 2016 was deemed acceptable; therefore, no augmentation was necessary for spring 2018.

PCR Performance – ELA Grades 5-8, English I, and English II (Spring 2018)

Reading Comprehension and Written Expression or Written Expression Trait

		IEA-Human Agreement						Human-Human Agreement					
Prompt	**Grade**	**Exact**	**SP0**	**SP1**	**SP2**	**SP3**	**SP4**	**Exact**	**SP0**	**SP1**	**SP2**	**SP3**	**SP4**
E05_N_CC434542756	5	**69.9**	**74.5**	**62.2**	**66.5**	**80.8**		67.5	84.6	56.6	61.6	62.2	
E06_N_D1466	6	**74.4**	88.3	**61.3**	**73.6**	**63.5**	**66.0**	67.3	89.9	55.6	68.5	55.4	54.3
E07_R_VH069361	7	**77.9**	**84.4**	**80.9**	**74.7**	60.4	**68.2**	73.5	81.8	73.2	69.1	61.4	66.7
E07_N_EE430133306	7	**73.0**	**88.2**	59.5	**65.9**	**68.0**	**76.7**	66.1	87.4	63.7	57.6	48.4	68.0
E08_L_F1460	8	**68.7**	**79.6**	**68.7**	**61.6**	**65.0**	**56.3**	59.6	71.4	58.1	56.9	48.3	45.7
VH017542_2T	9	**73.9**	**80.7**	**80.1**	**64.9**	**55.9**	**71.9**	64.6	80.5	70.6	52.9	43.0	58.7
VH017536_2T	9	**68.5**	**78.2**	**75.9**	**60.9**	**40.9**	**70.8**	61.9	75.3	67.0	58.1	31.9	51.6
VH004490	10	**74.4**	81.5	**76.9**	**71.7**	**57.9**	**63.1**	70.1	81.7	70.9	66.9	47.1	53.8
HH428127697	10	**73.6**	**83.5**	**75.4**	**71.8**	**59.8**	**60.5**	64.4	81.3	59.1	65.7	51.0	24.0

- Green indicates IEA agreement higher than human agreement
- Red indicates IEA agreement lower than human agreement by > 5.25%
- Source - Pearson

Conventions Trait

		IEA-Human Agreement					Human-Human Agreement				
Prompt	**Grade**	**Exact**	**SP0**	**SP1**	**SP2**	**SP3**	**Exact**	**SP0**	**SP1**	**SP2**	**SP3**
E05_N_CC434542756	5	**64.9**	78.6	**55.8**	**60.7**	**67.9**	60.3	82.2	53.7	47.6	53.8
E06_N_D1466	6	**77.6**	**88.4**	**70.3**	**73.2**	**76.9**	70.3	85.5	64.3	61.6	67.7
E07_R_VH069361	7	**79.4**	**85.8**	**78.6**	**73.3**	**76.9**	72.2	80.3	67.9	70.2	69.2
E07_N_EE430133306	7	**69.5**	**79.6**	**53.0**	**63.7**	**81.2**	65.5	74.4	62.5	48.9	74.3
E08_L_F1460	8	**70.5**	**76.2**	**69.9**	**64.2**	**73.9**	62.8	65.3	59.2	58.5	72.9
VH017542_2T	9	**74.5**	**80.7**	**74.7**	**57.2**	**83.4**	68.6	80.5	62.9	51.1	74.2
VH017536_2T	9	**67.9**	**70.9**	**69.2**	58.5	**78.2**	62.8	69.4	57.3	60.9	67.6
VH004490	10	**74.9**	**85.1**	**71.9**	**67.9**	**73.4**	67.0	84.3	63.1	58.5	53.2
HH428127697	10	**74.8**	82.7	**72.9**	**72.5**	**72.5**	68.7	84.9	63.6	63.1	63.9

- Green indicates IEA agreement higher than human agreement
- Red indicates IEA agreement lower than human agreement by > 5.25%
- Source - Pearson

Spring 2018 LEAP 2025 and EOC Items – IRR and SPD from Previous Administrations

Algebra I

IDEAS ID	Spring 2018 Form	PARCC UIN	Responses Available	Trait	Score Points	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
901882	Op, AE	VH196970	9,586	Part A	0,1	1,950	97.7%	99.9%	71.1%	12.5%				16.4%	Pearson Spring 2016
901882	Op, AE	VH196970	9,586	Part B	0,1,2	1,950	89.8%	97.1%	66.2%	6.7%	3.9%			23.1%	Pearson Spring 2016
901882	Op, AE	VH196970	8,522	Part A	0,1	1,940	99.0%	100.0%	94.0%	3.0%				4.0%	DRC Fall 2017, Op
901882	Op, AE	VH196970	8,522	Part B	0,1,2	1,940	99.0%	100.0%	94.0%	2.0%	1.0%			4.0%	DRC Fall 2017, Op
901859	AE	3003-M43111	253,395	Part C	0,1,2,3	48,917	92.1%	99.3%	43.7%	5.4%	14.9%	25.9%		10.1%	Pearson Spring 2016
901859	AE	3003-M43111	8,485	Part C	0,1,2,3	2,504	98.0%	100.0%	73.0%	4.0%	7.0%	13.0%		2.0%	DRC Fall 2017, Op
901832	Op, AE	3031-M44083P	95,907	Part B	0,1,2	18,835	91.3%	99.9%	29.9%	45.1%	11.9%			13.1%	Pearson Spring 2016
901848	AE	M47287	17,146	Overall	0,1,2,3,4	3,335	97.2%	99.6%	70.1%	9.2%	0.9%	0.4%	0.2%	19.1%	Pearson Spring 2016
901848	AE	M47287	8,445	Overall	0,1,2,3,4	2,796	100.0%	100.0%	78.0%	4.0%	0.0%	0.0%	0.0%	17.0%	DRC Fall 2017, Op
906043	Op, AE	2362-M41568	77,509	Part A	0,1,2	14,705	97.4%	99.9%	78.2%	3.4%	5.9%			12.5%	Pearson Spring 2017
906043	Op, AE	2362-M41568	77,509	Part B	0,1,2	14,705	91.2%	99.8%	28.0%	49.8%	4.5%			17.7%	Pearson Spring 2017
901687	Op	2407-M41752	83,197	Part A	0,1,2	15,543	96.7%	99.5%	63.3%	4.3%	23.3%			9.0%	Pearson Spring 2017
901687	Op	2407-M41752	83,197	Part B	0,1,2	15,543	92.6%	99.7%	68.3%	10.6%	7.0%			14.1%	Pearson Spring 2017
901687	Op	2407-M41752	83,197	Part C	0,1,2	15,543	95.5%	99.5%	71.4%	5.8%	4.1%			18.7%	Pearson Spring 2017
901857	Op, AE	VH046479	78,418	Part A	0,1,2	13,963	88.2%	99.8%	51.2%	36.3%	2.9%			9.6%	Pearson Spring 2017
901857	Op, AE	VH046479	78,418	Part B	0,1	13,963	91.8%	99.7%	68.9%	19.0%				12.1%	Pearson Spring 2017
901857	Op, AE	VH046479	8,686	Part A	0,1,2	2,258	94.0%	100.0%	77.0%	13.0%	1.0%			9.0%	DRC Fall 2017, Op
901857	Op, AE	VH046479	8,686	Part B	0,1	2,258	97.0%	100.0%	86.0%	5.0%				9.0%	DRC Fall 2017, Op
901851	Op, AE	M41726	77,612	Comp-rehensive	0,1,2,3	14,776	88.4%	99.1%	52.4%	10.8%	19.3%	9.9%		7.6%	Pearson Spring 2017
901705	Op	VF883359	147,898	Part A	0,1,2,3	29,794	96.8%	99.6%	83.9%	2.5%	1.5%	1.2%		10.9%	Pearson Spring 2015
901705	Op	VF883359	147,898	Part B	0,1	29,794	91.1%	99.8%	68.3%	12.3%				19.4%	Pearson Spring 2015

Geometry

IDEAS ID	Spring 2018 Form	PARCC UIN	Responses Available	Trait	Score Points	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
902036	AE	2904-M43021	42,708	Part A	0,1,2	8,216	95.9%	99.5%	53.9%	6.8%	33.2%			6.1%	Pearson Spring 2016
902036	AE	2904-M43021	42,708	Part B	0,1,2	8,216	94.7%	99.2%	61.0%	6.5%	24.1%			8.5%	Pearson Spring 2016
902036	AE	2904-M43021	42,708	Part C	0,1,2	8,216	94.9%	98.4%	75.2%	3.5%	4.2%			17.2%	Pearson Spring 2016
902036	AE	2904-M43021	6,800	Part A	0,1,2	1,518	99.0%	100.0%	69.0%	9.0%	20.0%			2.0%	DRC Fall 2017, Op
902036	AE	2904-M43021	6,800	Part B	0,1,2	1,518	96.0%	99.0%	68.0%	10.0%	19.0%			2.0%	DRC Fall 2017, Op
902036	AE	2904-M43021	6,800	Part C	0,1,2	1,518	97.0%	99.0%	91.0%	3.0%	3.0%			2.0%	DRC Fall 2017, Op
902042	Op	3020-M44058	45,304	Part A	0,1,2,3	8,509	94.5%	99.7%	47.9%	29.7%	7.3%	4.1%		11.0%	Pearson Spring 2016
902042	Op	3020-M44058	45,304	Part B	0,1	8,509	96.1%	99.8%	61.4%	21.9%				16.7%	Pearson Spring 2016
902042	Op	3020-M44058	45,304	Part C	0,1,2	8,509	94.8%	97.7%	61.2%	4.7%	12.2%			21.9%	Pearson Spring 2016
902047	AE	VH150404	47,576	Part A	0,1,2	8,713	97.9%	99.6%	64.1%	6.6%	2.8%			26.5%	Pearson Spring 2016
902047	AE	VH150404	47,576	Part B	0,1,2	8,713	92.1%	99.7%	52.6%	19.2%	7.0%			21.2%	Pearson Spring 2016
902047	AE	VH150404	6,775	Part A	0,1,2	1,636	98.0%	100.0%	81.0%	10.0%	4.0%			5.0%	DRC Fall 2017, Op
902047	AE	VH150404	6,775	Part B	0,1,2	1,636	96.0%	100.0%	74.0%	15.0%	6.0%			5.0%	DRC Fall 2017, Op
902012	Op, AE	M41169	90,471	Overall	0,1,2,3	16,723	87.1%	98.7%	46.2%	12.3%	14.6%	7.0%		19.9%	Pearson Spring 2016
902046	Op, AE	M46668	42,630	Overall	0,1,2,3	7,622	92.9%	98.9%	70.0%	8.5%	5.2%	0.5%		15.8%	Pearson Spring 2016
902046	Op, AE	M46668	6,821	Overall	0,1,2,3	1,880	97.0%	100.0%	78.0%	9.0%	3.0%	0.0%		9.0%	DRC Fall 2017, Op
902062	Op	VH150384	2,581	Overall	0,1,2,3,4	542	88.6%	97.4%	56.6%	6.1%	4.1%	1.5%	0.8%	30.9%	Pearson Spring 2016
902018	Op, AE	3042-M44133	48,083	Part A	0,1,2,3	9,072	86.1%	98.2%	45.8%	4.1%	17.4%	17.3%		15.3%	Pearson Spring 2017
902018	Op, AE	3042-M44133	48,083	Part B	0,1,2,3	9,072	91.6%	98.2%	62.9%	5.0%	5.1%	6.4%		20.7%	Pearson Spring 2017
902018	Op, AE	3042-M44133	6,775	Part A	0,1,2,3	1,762	95.0%	100.0%	61.0%	4.0%	13.0%	16.0%		7.0%	DRC Fall 2017, Op
902018	Op, AE	3042-M44133	6,775	Part B	0,1,2,3	1,762	98.0%	2.0%	82.0%	4.0%	4.0%	4.0%		7.0%	DRC Fall 2017, Op
902027	Op, AE	M43233	84,614	Overall	0,1,2,3,4	15,944	88.2%	97.7%	51.8%	12.5%	9.5%	5.0%	5.4%	15.9%	Pearson Spring 2017

Math Grade 3

Grade	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
3	914037	4003-M03007	7,194	Part A	0,1,2	668	92.0%	99.7%	34.6%	41.4%	20.8%			3.2%	Pearson Spring 2017
3	914037	4003-M03007	7,194	Part B	0,1,2	668	97.4%	100.0%	83.2%	4.5%	7.0%			5.3%	Pearson Spring 2017
3	914037	4003-M03007	7,194	Part C	0,1,2	668	97.5%	98.2%	0.7%	0.4%	91.5%			7.4%	Pearson Spring 2017
3	914008	VH054861	103,690	Part A	0,1,2	9,200	84.1%	94.7%	42.2%	13.3%	43.5%			1.0%	Pearson Spring 2017
3	914008	VH054861	103,690	Part B	0,1	9,207	94.4%	100.0%	87.0%	11.3%				1.7%	Pearson Spring 2017
3	914039	M02527	7,113	Comprehensive	0,1,2,3	699	92.7%	98.7%	37.9%	30.2%	23.4%	1.7%		6.8%	Pearson Spring 2017
3	914048	M05158	79,640	Comprehensive	0,1,2,3	7,819	92.4%	99.3%	52.8%	18.5%	11.5%	15.7%		1.5%	Pearson Spring 2017
3	798301	VF658085	57,049	Part C	0,1	11,648	87.0%	100.0%	80.0%	18.0%				2.0%	DRC Spring 2017
3	798301	VF658085	57,699	Part C	0,1	11,918	87.0%	100.0%	60.0%	37.0%				2.0%	DRC Spring 2016
3	798301	VF658085	55,234	Part C	0,1	10,885	88.5%	99.8%	78.7%	14.4%				6.9%	Pearson Spring 2015
3	868619	M00848	57,049	Comprehensive	0,1,2,3	11,716	93.0%	99.0%	64.0%	9.0%	5.0%	13.0%		8.0%	DRC Spring 2017

Math Grade 4

Grade	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Condition Code %	Report Source
4	870722	0220-M00114P	45,036	Part B	0,1,2	5	91.3%	99.6%	82.1%	10.3%	2.6%			5.0%	Pearson Spring 2017
4	870722	0220-M00114P	1,965	Part B	0,1,2	472	100.0%	100.0%	60.0%	41.0%	0.0%			0.0%	DRC 2017 [online]
4	870722	0220-M00114P	54,550	Part B	0,1,2	11,387	95.0%	100.0%	89.0%	7.0%	2.0%			2.0%	DRC 2017 [paper]
4	914084	4112-M03491P	383,723	Part C	0,1,2	37,737	94.9%	99.9%	65.0%	28.5%	2.5%			4.0%	Pearson Spring 2017
4	870317	M02255	191,643	Overall	0,1,2,3	38,068	82.4%	99.1%	46.0%	21.9%	25.6%	5.7%		0.8%	Pearson Spring 2016
4	870317	M02255	56,517	Overall	0,1,2,3	11,474	85.0%	99.0%	43.0%	20.0%	18.0%	14.0%		4.0%	DRC Spring 2017
4	914086	M04133	107,359	Comprehensive	0,1,2,3	10,670	91.1%	99.4%	53.1%	23.9%	7.4%	14.9%		0.7%	Pearson Spring 2017
4	870724	0083-M00453	79,205	Part C	0,1,2,3,4	16,182	94.4%	99.4%	56.7%	18.1%	14.7%	3.6%	2.6%	4.3%	Pearson Spring 2015
4	870723	M02081	79,454	Comprehensive	0,1,2,3	15,743	87.0%	98.1%	58.2%	13.1%	13.5%	10.8%		4.5%	Pearson Spring 2015

Math Grade 5

Grade	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
5	870325	4106-M03478	320,279	Part A	0,1	31,015	87.8%	100.0%	34.3%	64.6%				1.1%	Pearson Spring 2015
5	870325	4106-M03478	320,279	Part B	0,1,2	31,015	94.0%	99.8%	73.3%	16.2%	6.1%			4.4%	Pearson Spring 2015
5	870325	4106-M03478	53,396	Part A	0,1	11,104	79.0%	100.0%	45.0%	54.0%				0.0%	DRC Spring 2017
5	870325	4106-M03478	53,396	Part B	0,1,2	11,104	93.0%	100.0%	75.0%	19.0%	5.0%			0.0%	DRC Spring 2017
5	868654	4134-M03661	320,279	Part A	0,1,2,3	31,719	81.5%	97.4%	70.3%	9.6%	9.9%	9.4%		0.8%	Pearson Spring 2017
5	868654	4134-M03661	320,279	Part B	0,1,2,3	31,719	83.2%	98.6%	30.0%	14.3%	21.9%	31.0%		2.7%	Pearson Spring 2017
5	914139	VH055372	46,453	Part A	0,1	4,627	93.4%	100.0%	84.6%	12.9%				2.4%	Pearson Spring 2017
5	914139	VH055372	46,453	Part B	0,1,2	4,627	89.9%	99.4%	46.3%	26.0%	24.1%			3.6%	Pearson Spring 2017
5	914195	0154-M00796	92,904	Part B	0,1,2	9,282	95.9%	99.8%	80.4%	8.3%	6.4%			4.8%	Pearson Spring 2017
5	868655	M03555	4,052	Comprehensive	0,1,2,3	408	84.4%	99.8%	74.5%	15.1%	3.3%	2.7%		4.5%	Pearson Spring 2017
5	868655	M03555	53,396	Comprehensive	0,1,2,3	11,338	84.0%	99.0%	61.0%	26.0%	6.0%	5.0%		1.0%	DRC Spring 2017
5	914148	M03888	72,736	Comprehensive	0,1,2,3	7,272	86.7%	98.9%	39.9%	27.5%	13.2%	18.7%		0.8%	Pearson Spring 2017

Math Grade 6

Grade	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
6	868798	1298-M21432	2,653	Part A	0,1,2,3	266	96.2%	98.9%	81.2%	3.1%	2.4%	5.6%		7.7%	Pearson Spring 2017
6	868798	1298-M21432	2,653	Part B	0,1,2,3	266	96.6%	99.2%	82.6%	2.1%	1.9%	3.5%		9.8%	Pearson Spring 2017
6	868798	1298-M21432	52,395	Part A	0,1,2,3	10,800	91.0%	98.0%	71.0%	7.0%	7.0%	15.0%		0.0%	DRC Spring 2017
6	868798	1298-M21432	52,395	Part B	0,1,2,3	10,800	93.0%	98.0%	77.0%	3.0%	6.0%	13.0%		0.0%	DRC Spring 2017
6	914244	5012-M25073	112,484	Part A	0,1,2	11,249	88.4%	99.6%	41.5%	17.7%	39.4%			1.3%	Pearson Spring 2017
6	914244	5012-M25073	112,484	Part B	0,1,2	11,249	92.0%	99.6%	59.6%	17.1%	19.6%			3.7%	Pearson Spring 2017
6	914286	5047-M25308	38,543	Part A	0,1,2	3,854	87.6%	99.5%	77.4%	15.4%	1.2%			5.9%	Pearson Spring 2017
6	914286	5047-M25308	38,543	Part B	0,1,2	3,854	96.7%	99.4%	75.9%	14.2%	2.5%			7.4%	Pearson Spring 2017
6	914231	1740-M23030	89,916	Comprehensive	0,1,2,3	8,905	70.5%	96.2%	40.2%	18.0%	20.4%	19.0%		2.3%	Pearson Spring 2017
6	870825	M21477	38,264	Comprehensive	0,1,2,3	3,822	94.1%	99.6%	63.3%	25.4%	4.2%	2.3%		4.8%	Pearson Spring 2017
6	870825	M21477	52,418	Comprehensive	0,1,2,3	11,165	93.0%	100.0%	53.0%	36.0%	6.0%	3.0%		2.0%	DRC Spring 2017
6	868818	M22225	166,058	Comprehensive	0,1,2,3	16,588	88.3%	98.9%	61.7%	10.7%	10.1%	14.6%		3.0%	Pearson Spring 2017
6	868818	M22225	52,418	Comprehensive	0,1,2,3	11,080	90.0%	99.0%	68.0%	11.0%	9.0%	10.0%		1.0%	DRC Spring 2017
6	914281	M25152	112,484	Comprehensive	0,1,2,3	11,247	89.0%	99.0%	53.8%	14.1%	12.3%	17.2%		2.6%	Pearson Spring 2017

Math Grade 7

Grade	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
7	914333	1300-M21438	111,058	Part A	0,1,2	10,988	92.0%	99.8%	52.4%	34.4%	7.1%			6.0%	Pearson Spring 2017
7	914333	1300-M21438	111,058	Part B	0,1,2	10,988	93.9%	99.7%	48.2%	35.3%	9.0%			7.5%	Pearson Spring 2017
7	914339	VH151385	88,725	Part A	0,1,2	8,838	95.4%	99.4%	66.9%	7.8%	20.9%			4.4%	Pearson Spring 2017
7	914339	VH151385	88,725	Part B	0,1,2	8,838	95.5%	99.7%	77.2%	5.7%	9.7%			7.4%	Pearson Spring 2017
7	868849	M22018	325,765	Comprehensive	0,1,2,3	32,435	91.9%	99.3%	54.2%	23.8%	7.8%	11.4%		2.8%	Pearson Spring 2017
7	868849	M22018	51,790	Comprehensive	0,1,2,3	10,988	97.0%	100.0%	65.0%	20.0%	6.0%	8.0%		1.0%	DRC Spring 2017
7	914329	M25580	111,124	Comprehensive	0,1,2,3	11,013	87.9%	99.1%	47.6%	19.8%	18.4%	10.0%		4.3%	Pearson Spring 2017
7	914336	M25731	88,721	Comprehensive	0,1,2,3	8,591	84.7%	99.2%	26.2%	31.4%	30.6%	8.8%		3.0%	Pearson Spring 2017
7	878705	VH095906	13,001	Comprehensive	0,1,2,3	1,279	90.5%	98.0%	70.8%	10.0%	7.7%	3.6%		7.9%	Pearson Spring 2017
7	878705	VH095906	51,790	Comprehensive	0,1,2,3	10,720	96.0%	99.0%	84.0%	9.0%	5.0%	1.0%		2.0%	DRC Spring 2017
7	914362	VH083535	100,577	Part A	0,1,2,3	19,892	90.4%	98.5%	74.7%	5.1%	4.9%	12.5%		2.8%	Pearson Spring 2016
7	914362	VH083535	100,577	Part B	0,1,2,3	19,892	90.3%	98.8%	70.8%	5.6%	5.6%	13.8%		4.3%	Pearson Spring 2016

Math Grade 8

Grade	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
8	892626	VH095800	29,600	Part A	0,1,2,3	2,775	92.0%	99.3%	68.5%	8.4%	6.3%	10.7%		6.1%	Pearson Spring 2017
8	892626	VH095800	29,600	Part B	0,1,2,3	2,774	94.1%	99.1%	78.9%	4.2%	3.1%	5.4%		8.4%	Pearson Spring 2017
8	892626	VH095800	44,761	Part A	0,1,2,3	9,566	87.0%	99.0%	61.0%	12.0%	7.0%	18.0%		2.0%	DRC Spring 2017
8	892626	VH095800	44,761	Part B	0,1,2,3	9,566	92.0%	100.0%	76.0%	8.0%	5.0%	9.0%		2.0%	DRC Spring 2017
8	914400	M21215	87,197	Comprehensive	0,1,2,3	8,580	87.3%	98.9%	64.9%	14.1%	10.1%	3.5%		7.4%	Pearson Spring 2017
8	914381	M25425	69,637	Comprehensive	0,1,2,3,4	6,943	90.8%	99.1%	52.2%	12.5%	26.2%	1.9%	1.0%	6.2%	Pearson Spring 2017
8	914376	VH049778	285,447	Comprehensive	0,1,2,3	28,266	84.0%	98.9%	59.1%	19.5%	10.9%	6.2%		4.4%	Pearson Spring 2017
8	878156	VH057654	99,440	Comprehensive	0,1,2,3	9,865	89.6%	98.5%	74.6%	9.6%	9.0%	2.5%		4.2%	Pearson Spring 2017
8	878156	VH057654	44,810	Comprehensive	0,1,2,3	9,714	93.0%	100.0%	77.0%	9.0%	11.0%	2.0%		2.0%	DRC Spring 2017
8	868882	VH081259	29,600	Part B	0,1,2	2,954	94.5%	99.7%	79.5%	7.4%	1.4%			11.7%	Pearson Spring 2017
8	800273	M20195	44,810	Comprehensive	0,1,2,3,4	9,878	90.0%	100.0%	56.0%	26.0%	7.0%	3.0%	6.0%	2.0%	DRC Spring 2017
8	800273	M20195	47,273	Comprehensive	0,1,2,3,4	10,152	88.0%	99.0%	46.0%	27.0%	8.0%	4.0%	8.0%	6.0%	DRC Spring 2016

English I

Task	IDEAS ID	Spring 2018 Form	PARCC UIN	Responses Available	Trait	Score Points	Human 1st Score Count	Human 2nd Score Count	AI 1st & 2nd Score Count	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
LAT	902152	Op	VH017536_2T	126,939	RCWE	0,1,2,3,4	8,443	14,816	114,377	23,741	73.6%	99.3%	36.7%	36.2%	16.6%	5.3%	1.8%	3.4%	Pearson Spring 2017
LAT	902152	Op	VH017536_2T	126,939	Conv	0,1,2,3	8,443	14,816	114,377	23,741	73.4%	99.6%	31.5%	37.1%	20.4%	7.5%		3.4%	Pearson Spring 2017
NWT	906152	AE	VH084830	61,936	WE	0,1,2,3,4	3,125	7,776	53,955	10,498	73.3%	98.7%	30.3%	21.8%	27.3%	8.7%	4.2%	7.6%	Pearson Spring 2017
NWT	906152	AE	VH084830	61,936	Conv	0,1,2,3	3,125	7,776	53,955	10,498	74.4%	99.4%	28.1%	27.7%	25.4%	11.2%		7.6%	Pearson Spring 2017
NWT	906152*	AE	VH084830	4,663	WE	0,1,2,3,4	n/a	n/a	n/a	1,076	81.0%	99.0%	22.0%	34.0%	29.0%	10.0%	1.0%	2.0%	DRC Fall 2017, Op
NWT	906152*	AE	VH084830	4,663	Conv	0,1,2,3	n/a	n/a	n/a	1,076	78.0%	99.0%	25.0%	36.0%	26.0%	10.0%		2.0%	DRC Fall 2017, Op
RST	902161	Op, AE	VH017542_2T	123,860	RCWE	0,1,2,3,4	2,656	13,063	116,406	23,334	76.1%	99.5%	22.2%	33.3%	24.3%	12.4%	3.7%	4.1%	Pearson Spring 2017
RST	902161	Op, AE	VH017542_2T	123,860	Conv	0,1,2,3	2,656	13,063	116,407	23,334	76.1%	99.6%	23.1%	32.7%	23.5%	16.6%		4.1%	Pearson Spring 2017
RST	902161	Op, AE	VH017614_2T	3,179	RCWE	0,1,2,3,4	3,012	317	0	620	82.6%	99.4%	43.9%	33.3%	12.0%	2.5%	0.8%	7.6%	Pearson Spring 2017
RST	902161	Op, AE	VH017614_2T	3,179	Conv	0,1,2,3	3,012	317	0	620	79.4%	99.7%	47.8%	30.2%	11.7%	2.8%		7.6%	Pearson Spring 2017
RST	902161*	Op, AE	VH017542_2T	4,674	RCWE	0,1,2,3,4	n/a	n/a	n/a	982	78.0%	99.0%	12.0%	34.0%	40.0%	13.0%	0.0%	0.0%	DRC Fall 2017, Op
RST	902161*	Op, AE	VH017542_2T	4,674	Conv	0,1,2,3	n/a	n/a	n/a	982	78.0%	99.0%	14.0%	32.0%	38.0%	15.0%		0.0%	DRC Fall 2017, Op

*Handscored by DRC in Fall of 2017

English II

Task	IDEAS ID	Spring 2018 Form	PARCC UIN	Responses Available	Trait	Score Points	Human 1st Score Count	Human 2nd Score Count	AI 1st & 2nd Score Count	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
NWT	902354	AE	7064	4,409	WE	0,1,2,3,4	4,189	435	0	844	84.5%	100.0%	42.5%	19.5%	13.5%	6.0%	2.0%	16.5%	Pearson Spring 2017
NWT	902354	AE	7064	4,409	Conv	0,1,2,3	4,189	435	0	844	85.1%	100.0%	39.7%	21.7%	15.1%	7.0%		16.5%	Pearson Spring 2017
NWT	902354*	AE	7064	9,271	WE	0,1,2,3,4	n/a	n/a	n/a	2,098	81.0%	100.0%	46.0%	17.0%	19.0%	12.0%	2.0%	2.0%	DRC Fall 2017, Op
NWT	902354*	AE	7064	9,271	Conv	0,1,2,3	n/a	n/a	n/a	2,098	83.0%	100.0%	43.0%	18.0%	22.0%	14.0%		2.0%	DRC Fall 2017, Op
RST	902331	Op, AE	VH004490**	2,605	RCWE	0,1,2,3,4	1,915	263	646	827	81.9%	99.3%	51.5%	28.1%	7.2%	0.9%	0.1%	12.3%	Pearson Spring 2017
RST	902331	Op, AE	VH004490**	2,605	Conv	0,1,2,3	1,915	263	646	827	84.8%	99.8%	53.3%	26.3%	7.4%	0.7%		12.3%	Pearson Spring 2017
RST	902331*	Op, AE	VH004490**	9,305	RCWE	0,1,2,3,4	n/a	n/a	n/a	2,020	79.0%	100.0%	37.0%	24.0%	25.0%	11.0%	2.0%	2.0%	DRC Fall 2017, Op
RST	902331*	Op, AE	VH004490**	9,305	Conv	0,1,2,3	n/a	n/a	n/a	2,020	77.0%	99.0%	35.0%	23.0%	27.0%	14.0%		2.0%	DRC Fall 2017, Op
LAT	906197	Op	HH428127697	57,407	RCWE	0,1,2,3,4	28,405	6,463	27,217	14,673	75.4%	99.2%	27.0%	30.6%	27.3%	9.0%	1.3%	4.8%	Pearson Spring 2017
LAT	906197	Op	HH428127697	57,407	Conv	0,1,2,3	28,405	6,463	27,218	14,673	74.8%	99.1%	27.1%	30.0%	27.5%	10.7%		4.8%	Pearson Spring 2017
Task	IDEAS ID	Spring 2018 Form	PARCC UIN	Responses Available	Trait	Score Points	Score Count	n/a	n/a	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	CC %	Report Source
RST	902331	Op, AE	VH004490**	126,270	RCWE	0,1,2,3,4	121,660	n/a	n/a	16,036	76.6%	99.7%	22.7%	34.8%	23.4%	8.3%	2.0%	8.8%	Pearson Spring 2016
RST	902331	Op, AE	VH004490**	126,270	Conv	0,1,2,3	121,507	n/a	n/a	16,003	76.9%	99.7%	25.1%	32.6%	24.0%	9.5%		8.8%	Pearson Spring 2016

*Handscored by DRC in Fall of 2017

** Pearson – Statistics from 2017 and 2016 are included for 902331/VH004490. Volumes were significantly higher in 2016, but reports in 2016 did not split out human and AI scoring, so the 2016 and 2017 column headers are different.

English III (All Report Data Scored by DRC and MI [AI])

IDEAS ID	Spring 2018 Form	Total Reads	Trait	Score Points	Read 2x	Exact%	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	NS%	Source of Data
851368	Form U, Op	7,783	Content	1, 2, 3, 4	2,204	83%	17%	0%		17%	39%	38%	5%	2%	Op Fall 2016
		7,783	Style	1, 2, 3, 4	2,204	83%	17%	0%		12%	32%	44%	11%	2%	
		7,783	Formation (F)	0, 1	2,204	83%	17%		37%	61%				2%	
		7,783	Usage (U)	0, 1	2,204	78%	22%		21%	78%				2%	
		7,783	Mechanics (M)	0, 1	2,204	89%	11%		22%	77%				2%	
		7,783	Spelling (S)	0, 1	2,204	83%	17%		25%	73%				2%	
851368	Form U, Op	686	Content	1, 2, 3, 4	486	94%	6%	0%		56%	26%	2%	0%	15%	Op Summer 2017
		686	Style	1, 2, 3, 4	486	94%	6%	0%		48%	30%	6%	1%	15%	
		686	Formation (F)	0, 1	486	95%	5%		57%	27%				15%	
		686	Usage (U)	0, 1	486	96%	4%		71%	14%				15%	
		686	Mechanics (M)	0, 1	486	93%	7%		57%	28%				15%	
		686	Spelling (S)	0, 1	486	93%	7%		35%	50%				15%	
851370*	Form R, AE	8,382	Content	1, 2, 3, 4	3,672	91%	9%	0%		18%	34%	28%	11%	10%	Op Fall 2017
		8,382	Style	1, 2, 3, 4	3,672	91%	9%	0%		14%	34%	33%	9%	10%	
		8,382	Formation (F)	0, 1	3,672	94%	6%		27%	63%				10%	
		8,382	Usage (U)	0, 1	3,672	93%	7%		46%	44%				10%	
		8,382	Mechanics (M)	0, 1	3,672	93%	7%		28%	62%				10%	
		8,382	Spelling (S)	0, 1	3,672	96%	4%		21%	69%				10%	
851370*	Form R, AE	182	Content	1, 2, 3, 4	62	81%	19%	0%		5%	43%	46%	5%	1%	AE Fall 2016
		182	Style	1, 2, 3, 4	62	87%	13%	0%		5%	40%	51%	3%	1%	
		182	Formation (F)	0, 1	62	87%	13%		13%	86%				1%	
		182	Usage (U)	0, 1	62	90%	10%		30%	69%				1%	
		182	Mechanics (M)	0, 1	62	94%	6%		13%	86%				1%	
		182	Spelling (S)	0, 1	62	87%	13%		14%	85%				1%	
Form U (851368), will be AI scored by MI with human backreads by DRC.															
Form R (851370), the AE form, will be handscored by DRC supervisors in Spring 2018.															

ELA Grade 3

Task	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Human 1st Score Count	Human 2nd Score Count	AI 1st & 2nd Score Count	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
NWT	913497	AA431426588	118,416	WE	0,1,2,3	34,298	13,546	84,911	27,299	71.2%	98.6%	30.0%	56.0%	10.6%	1.6%		1.8%	Pearson Spring 2017
NWT	913497	AA431426588	118,416	Conventions	0,1,2,3	34,298	13,546	84,910	27,299	68.6%	98.6%	33.3%	46.8%	16.0%	2.1%		1.8%	Pearson Spring 2017
RST	868500	VF581237	66,149	RCWE	0,1,2,3	62,975	6,643	3,282	13,959	70.8%	98.5%	42.8%	41.0%	10.9%	1.3%		3.9%	Pearson Spring 2017
RST	868500	VF581237	66,149	Conventions	0,1,2,3	62,975	6,643	3,282	13,959	68.8%	98.3%	37.1%	41.2%	15.7%	2.1%		3.9%	Pearson Spring 2017
RST	868500	VF581237	57,098	RCWE	0,1,2,3	n/a	n/a	n/a	12,140	75.0%	99.0%	41.0%	43.0%	13.0%	1.0%		2.0%	DRC Spring 2017
RST	868500	VF581237	57,098	Conventions	0,1,2,3	n/a	n/a	n/a	12,140	76.0%	100.0%	25.0%	46.0%	24.0%	3.0%		2.0%	DRC Spring 2017

ELA Grade 4

Task	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Human 1st Score Count	Human 2nd Score Count	AI 1st & 2nd Score Count	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
RST	802342	1805	4,565	RCWE	0,1,2,3,4	n/a	n/a	n/a	878	75.9%	97.3%	41.1%	37.9%	10.8%	1.1%	0.1%	9.0%	Pearson Spring 2016
RST	802342	1805	4,565	Conventions	0,1,2,3	n/a	n/a	n/a	878	74.3%	97.7%	39.6%	40.0%	10.1%	1.2%		9.0%	Pearson Spring 2016
RST	802342	1805_AT	65	RCWE	0,1,2,3,4	n/a	n/a	n/a	42	100.0%	100.0%	46.3%	30.5%	14.6%	0.0%	0.0%	8.5%	Pearson Spring 2016
RST	802342	1805_AT	65	Conventions	0,1,2,3	n/a	n/a	n/a	42	81.0%	100.0%	43.9%	36.6%	11.0%	0.0%		8.5%	Pearson Spring 2016
RST	802342	1805_TTS	9,851	RCWE	0,1,2,3,4	n/a	n/a	n/a	1,951	80.6%	97.5%	46.4%	38.0%	6.7%	0.4%	0.1%	8.4%	Pearson Spring 2016
RST	802342	1805_TTS	9,851	Conventions	0,1,2,3	n/a	n/a	n/a	1,951	78.7%	97.7%	45.3%	39.9%	5.8%	0.6%		8.4%	Pearson Spring 2016
RST	802342	1805	54,392	RCWE	0,1,2,3,4	n/a	n/a	n/a	14,620	73%	99%	15%	45%	33%	4%	0%	2%	DRC Spring 2016
RST	802342	1805	54,392	Conventions	0,1,2,3	n/a	n/a	n/a	14,620	72%	98%	14%	44%	35%	4%		2%	DRC Spring 2016
RST	802342	1805	2,605	RCWE	0,1,2,3,4	n/a	n/a	n/a	525	75.6%	97.7%	46.7%	32.6%	6.0%	0.6%	0.1%	14.0%	Pearson Spring 2017
RST	802342	1805	2,605	Conventions	0,1,2,3	n/a	n/a	n/a	525	75.6%	97.3%	44.1%	35.9%	5.2%	0.7%		14.0%	Pearson Spring 2017
LAT	913567	VH170170	121,461	RCWE	0,1,2,3,4	35,658	13,901	87,168	28,425	67.3%	98.5%	33.5%	40.0%	21.2%	3.5%	0.5%	1.3%	Pearson Spring 2017
LAT	913567	VH170170	121,461	Conventions	0,1,2,3	35,659	13,893	87,168	28,418	69.3%	99.1%	28.1%	45.7%	21.0%	3.9%		1.3%	Pearson Spring 2017

ELA Grade 5

Task	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Human 1st Score Count	Human 2nd Score Count	AI 1st & 2nd Score Count	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
NWT	913663	CC434542756	130,346	WE	0,1,2,3	34,715	4,660	96,145	20,437	71.5%	98.8%	30.9%	29.1%	27.5%	11.5%		0.9%	Pearson Spring 2017
NWT	913663	CC434542756	130,346	Conventions	0,1,2,3	34,710	4,660	96,143	20,440	69.1%	98.7%	32.0%	31.2%	26.3%	9.6%		0.9%	Pearson Spring 2017
RST	868590	C1381	13,923	RCWE	0,1,2,3,4	13,559	1,295	0	2,584	66.8%	96.9%	29.2%	42.0%	21.9%	4.2%	0.3%	2.4%	Pearson Spring 2017
RST	868590	C1381	13,923	Conventions	0,1,2,3	13,559	1,295	0	2,584	67.1%	97.3%	27.9%	39.3%	25.2%	5.2%		2.4%	Pearson Spring 2017
RST	868590	C1381	53,560	RCWE	0,1,2,3,4	n/a	n/a	n/a	12,210	73.0%	99.0%	28.0%	47.0%	20.0%	4.0%	0.0%	0.0%	DRC Spring 2017
RST	868590	C1381	535,360	Conventions	0,1,2,3	n/a	n/a	n/a	12,210	72.0%	99.0%	23.0%	45.0%	25.0%	6.0%		0.0%	DRC Spring 2017

ELA Grade 6

Task	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Human 1st Score Count	Human 2nd Score Count	AI 1st & 2nd Score Count	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
NWT	913694	D1466	127,628	WE	0,1,2,3,4	34,718	14,034	93,800	29,433	75.9%	99.4%	40.3%	20.6%	22.9%	10.0%	4.3%	1.8%	Pearson Spring 2017
NWT	913694	D1466	127,628	Conventions	0,1,2,3	34,718	14,034	93,800	29,433	75.0%	99.6%	33.4%	30.2%	23.3%	11.3%		1.8%	Pearson Spring 2017
RST	870622	VH082653	52,594	RCWE	0,1,2,3,4	n/a	n/a	n/a	10,928	72.0%	99.0%	24.0%	40.0%	29.0%	6.0%	1.0%	0.0%	DRC Spring 2017
RST	870622	VH082653	52,594	Conventions	0,1,2,3	n/a	n/a	n/a	10,928	73.0%	99.0%	22.0%	38.0%	33.0%	6.0%		0.0%	DRC Spring 2017

ELA Grade 7

Task	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Human 1st Score Count	Human 2nd Score Count	AI 1st & 2nd Score Count	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
NWT	913842	EE430133306	128,845	WE	0,1,2,3,4	37,606	14,582	91,555	30,289	72.7%	98.6%	34.2%	12.7%	20.4%	17.9%	12.6%	2.2%	Pearson Spring 2017
NWT	913842	EE430133306	128,845	Conventions	0,1,2,3	37,605	14,582	91,555	30,289	71.5%	99.0%	28.5%	20.9%	23.8%	24.7%		2.2%	Pearson Spring 2017
RST	868704	VH069361	158,086	RCWE	0,1,2,3,4	n/a	n/a	n/a	27,575	74.6%	99.5%	20.9%	33.7%	27.8%	12.2%	3.5%	1.9%	Pearson Spring 2016
RST	868704	VH069361	158,086	Conventions	0,1,2,3	n/a	n/a	n/a	27,471	75.7%	99.5%	21.3%	33.0%	28.2%	15.6%		1.9%	Pearson Spring 2016
RST	868704	VH069361	52,091	RCWE	0,1,2,3,4	n/a	n/a	n/a	10,916	72.0%	99.0%	16.0%	33.0%	32.0%	15.0%	3.0%	0.0%	DRC Spring 2017
RST	868704	VH069361	52,091	Conventions	0,1,2,3	n/a	n/a	n/a	10,916	73.0%	99.0%	17.0%	31.0%	33.0%	18.0%		0.0%	DRC Spring 2017

ELA Grade 8

Task	IDEAS ID	PARCC UIN	Responses Available	Trait	Score Points	Human 1st Score Count	Human 2nd Score Count	AI 1st & 2nd Score Count	Reliability Read Count	Exact IRR %	Exact + Adj IRR %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Cond Code %	Report Source
LAT	913958	F1460	128,084	RCWE	0,1,2,3,4	36,606	4,234	89,633	19,154	69.6%	99.8%	26.4%	31.2%	25.9%	10.9%	3.1%	2.4%	Pearson Spring 2017
LAT	913958	F1460	128,084	Conventions	0,1,2,3	36,606	4,234	89,634	19,154	71.9%	99.7%	22.9%	30.7%	29.2%	14.7%		2.4%	Pearson Spring 2017
RST	868741	5013	7,826	RCWE	0,1,2,3,4	7,573	783	0	1,547	74.5%	99.4%	52.8%	26.4%	11.1%	2.4%	0.5%	6.8%	Pearson Spring 2017
RST	868741	5013	7,826	Conventions	0,1,2,3	7,573	783	0	1,547	77.5%	99.2%	53.1%	26.5%	10.5%	3.0%		6.8%	Pearson Spring 2017
RST	868741	5013	50,673	RCWE	0,1,2,3,4	n/a	n/a	n/a	11,002	72.0%	99.0%	12.0%	21.0%	29.0%	25.0%	12.0%	1.0%	DRC Spring 2017
RST	868741	5013	50,673	Conventions	0,1,2,3	n/a	n/a	n/a	11,002	78.0%	99.0%	12.0%	21.0%	29.0%	38.0%		1.0%	DRC Spring 2017

Biology

IDEAS ID	Spring 2018 Form	Total Reads	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	NS%	Report Source	Source of Data
845340	Form J, Op	n/a	0-4	This item has not previously been scored operationally by DRC. Was AE item in spring of 2017 with 9 total responses.											
845226	Form K, AE	12,560	0-4	3,680	91%	9%	0%	45%	24%	11%	6%	3%	12%	Op Fall 2016	DRC - handscore
845226	Form K, AE	2,952	0-4	1,128	98%	2%	0%	61%	16%	2%	0%	0%	20%	Op Sum 2017	DRC - handscore
These items will be handscored by DRC in Spring 2018.															

U.S. History ERs (EOC)

IDEAS ID	Spring 2018 Form	Total Reads	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	NS%	Source of Data
848632	Form E, Spring 2018 Op	9,847	0-4	3,086	89%	11%	0%	53%	30%	12%	4%	1%	1%	Fall 2016, Op
848632	Form E, Spring 2018 Op	1,609	0-4	1,158	95%	5%	1%	85%	12%	2%	0%	0%	1%	Summer 2017, Op
848291*	Form C, Spring 2018 Admin Error	216	0-4	86	79%	21%	0%	6%	13%	28%	30%	23%	0%	AE Fall 2016
All report data scored by DRC - AI [MI]														
Form E (848632) will be AI scored by MI with DRC human backreads.														
Form C (848291), the AE form, will be handscored by DRC supervisors.														

U.S. History ERs and CRs (LEAP 2025)

IDEAS ID	Item Type	Spring 2018 Form	Total Reads	Trait	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Source of Data
894104*	ER	Op, AE	5,000	Content	0-4	5,000	62%	33%	5%	31%	34%	22%	9%	4%	Spring 2017 FT
894104*	ER	Op, AE	5,000	Claims	0-4	5,000	61%	32%	7%	39%	28%	21%	9%	4%	Spring 2017 FT
894256*	ER	Op	5,000	Content	0-4	5,000	66%	32%	2%	28%	37%	24%	8%	3%	Spring 2017 FT
894256*	ER	Op	5,000	Claims	0-4	5,000	64%	33%	3%	34%	33%	21%	8%	3%	Spring 2017 FT
892955*	ER	Op	5,000	Content	0-4	5,000	65%	32%	3%	34%	29%	25%	9%	3%	Spring 2017 FT
892955*	ER	Op	5,000	Claims	0-4	5,000	64%	32%	4%	37%	26%	25%	10%	3%	Spring 2017 FT
894047*	ER	Op	5,000	Content	0-4	5,000	66%	31%	3%	39%	29%	19%	9%	4%	Spring 2017 FT
894047*	ER	Op	5,000	Claims	0-4	5,000	66%	30%	4%	40%	29%	19%	9%	4%	Spring 2017 FT
894225**	CR	Op	1,660		0-2	320	71%	29%	0%	44%	34%	22%			Spring 2017 FT
892994**	CR	Op	1,659		0-2	318	68%	31%	1%	13%	43%	44%			Spring 2017 FT
894188**	CR	Op, AE	6,150		0-2	1,190	85%	15%	0%	53%	33%	14%			Fall 2017 Op
894188**	CR	Op, AE	1,655		0-2	310	74%	25%	1%	33%	38%	28%			Spring 2017 FT
894149**	CR	Op, AE	6,056		0-2	1,132	87%	13%	0%	68%	18%	14%			Fall 2017 Op
894149**	CR	Op, AE	1,653		0-2	306	76%	21%	3%	44%	26%	30%			Spring 2017 FT

All data from DRC handscoring. Nonscores excluded.

*AI scoring models built by MI using 2500 FT responses, scored 2x and resolved. Op ER responses will be AI scored by MI with DRC human backreads. (See U.S. History 5-level model building data earlier in Appendix.)

**All other Spring 2018 USH items (operational 2-point CRs and all AE form items) will be handscored.

Social Studies Grades 3-8

Gr	IDEAS ID	Item Type	Total Reads	Trait	Score Points	Read 2x	Exact %	Adj %	Non-Adj %	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	Source of Data
3	801223	ER	5,428	Content & Claims	0-4	5,428	74	24	2	61	30	5	1	0	Spring 2016 FT
3	801185	CR	63,014		0-2	11,542	82	17	1	23	35	37			Spring 2017 Op
3	801185	CR	1,651		0-2	302	79	19	2	25	48	22			Spring 2016 FT
3	890671	CR	1,650		0-2	300	73	26	1	47	34	14			Spring 2017 FT
4	801572	ER	4,048	Content	0-4	4,048	74	25	2	51	39	9	1	0	Spring 2016 FT [online]
4	801572	ER	4,048	Claims	0-4	4,048	74	25	1	51	39	9	1	0	Spring 2016 FT [online]
4	801572	ER	4,124	Content	0-4	4,124	69	28	3	46	39	11	2	0	Spring 2016 FT [paper]
4	801572	ER	4,124	Claims	0-4	4,124	69	28	2	46	39	11	2	0	Spring 2016 FT [paper]
4	890774	CR	1,657		0-2	314	81	19	0	13	49	36			Spring 2017 FT
4	801582	CR	62,365		0-2	11,578	83	17	0	41	48	6			Spring 2017 Op
4	801582	CR	1,657		0-2	314	78	21	2	42	41	9			Spring 2016 FT
5	807728*	ER	5,198	Content	0-4	5,198	76	23	1	61	23	13	3	0	Spring 2016 FT
5	807728*	ER	5,198	Claims	0-4	5,198	78	20	2	66	20	11	2	0	Spring 2016 FT
5	807709	CR	57,809		0-2	10,582	82	18	0	65	31	4			Spring 2017 Op
5	807709	CR	1,566		0-2	280	81	19	0	61	34	4			Spring 2016 FT
5	890929	CR	1,653		0-2	306	73	27	0	60	31	10			Spring 2017 FT
6	804944*	ER	5,949	Content	0-4	5,949	73	25	1	61	32	7	1	0	Spring 2016 FT
6	804944*	ER	5,949	Claims	0-4	5,949	75	24	1	64	31	5	1	0	Spring 2016 FT
6	891206	CR	1,651		0-2	302	79	20	1	56	26	18			Spring 2017 FT
6	891161	CR	1,654		0-2	308	81	19	0	37	47	16			Spring 2017 FT
7	805569*	ER	5,220	Content	0-4	5,220	68	31	1	32	42	21	4	0	Spring 2016 FT
7	805569*	ER	5,220	Claims	0-4	5,220	66	32	1	35	41	20	4	0	Spring 2016 FT
7	891311	CR	1,654		0-2	308	73	27	1	34	42	24			Spring 2017 FT
7	891331	CR	1,659		0-2	318	74	25	1	48	40	12			Spring 2017 FT
8	809006*	ER	5,086	Content	0-4	5,086	63	35	2	27	38	24	8	2	Spring 2016 FT
8	809006*	ER	5,086	Claims	0-4	5,086	62	36	2	33	33	24	8	2	Spring 2016 FT
8	892278	CR	1,656		0-2	312	79	20	1	43	41	15			Spring 2017 FT
8	892336	CR	1,668		0-2	336	70	29	1	23	46	30			Spring 2017 FT

All data from DRC handscoring. Nonscores excluded.
*AI scoring models built by MI using 2500 FT responses, scored 2x and resolved. Grades 5-8 Op ER responses will be AI scored by MI with DRC human backreads. (See Grades 5-8 SS model building data earlier in Appendix.)
**All other Spring 2018 SS items (operational 2-point CRs and grade 3 and 4 ER items) will be handscored.