

**2018-2019 LEAP 2025 High School  
Operational Technical Report  
English I, English II, Algebra I, and Geometry**

Submitted to the  
Louisiana Department of Education

November 2019



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

## Table of Contents

---

Executive Summary.....	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 .....	8
1.1 Background .....	8
1.2 Purpose of the LEAP 2025 .....	8
1.3 Design of the LEAP 2025 .....	8
Chapter 2: The Uses of Test Scores.....	11
2.1 Uses of Test Scores.....	11
2.2 Test-Level Scores.....	11
2.3 Scale Scores.....	12
2.4 Levels of Achievement .....	12
2.5 Use of Test-Level Scores.....	12
2.6 Category- and Subcategory-Level Subscores .....	12
2.7 Use of the Category- and Subcategory-Level Ratings .....	13
Chapter 3: Test Content Development.....	14
3.1 Defining the Specific Test Blueprint.....	16
3.2 Louisiana Item Development and Item Review .....	33
3.3 Guidelines on Bias, Fairness, and Sensitivity.....	36
3.3 Operational Test Selection .....	37
3.4 Universal Design.....	43
3.5 Accommodations and Designated Supports .....	44
3.6 Item and Task Specifications .....	46
3.7 Summary .....	46
Chapter 4: Test Administration.....	48
4.1 Return Material Forms and Guidelines .....	52
4.2 Security Checklists.....	53
4.3 Interpretive Guides .....	53
4.4 Test Security Measures .....	53

4.5	Test Administration.....	54
4.6	Summary .....	60
Chapter 5: Scoring of Constructed-Response and Technology-Enhanced Items.....		62
5.1	Constructed-Response Item Scoring Process.....	62
5.2	Inter-Rater Reliability .....	71
5.3	Technology-Enhanced Item Scoring Process.....	84
5.4	Multiple-Choice and Multiple-Select Item Scoring Process .....	85
5.5	Summary .....	85
Chapter 6: Operational Data Analyses.....		86
6.1	Classical Item Statistics.....	86
6.2	Calibration Sample .....	91
6.3	Calibration and Linking.....	96
6.4	Comparability: Form Equating .....	113
6.5	Summary .....	119
Chapter 7: Test Results .....		121
7.1	Student Participation .....	121
7.2	Reports.....	143
7.3	Description of Each Type of Report.....	144
7.4	Summary .....	145
Chapter 8: Performance-Level Setting.....		146
8.1	PARCC Performance-Level Setting Process for English Language Arts and Mathematics .....	146
8.2	Cut Scores.....	146
8.3	Category Cut Scores .....	147
8.4	Achievement-Level Definitions .....	147
8.5	Summary .....	147
Chapter 9: Evidence of Construct-Related Validity.....		149
9.1	Construct-Irrelevant Variance and Construct Underrepresentation .....	150
9.2	Reliability.....	150
9.3	Test Reliability .....	151
9.4	Standard Error of Measurement.....	152
9.5	Conditional Standard Error of Measurement .....	153
9.6	Classification Accuracy and Consistency.....	156
9.7	Convergent Validity.....	159
9.8	Principal Components Analysis .....	159

9.9	Analyses by Reporting Categories and Subcategories .....	162
9.10	Correlations among Reporting Categories and Subcategories .....	162
9.11	Reliability of Reporting Categories, or Subcategories.....	167
9.12	Standard Error of Measurement of Reporting Categories or Subcategories.....	167
9.13	Divergent (Discriminant) Validity .....	171
9.14	Summary .....	172
Chapter 10: Fairness .....		173
10.1	Minimizing Bias through Careful Test Development .....	174
10.2	Evaluating Bias through Differential Item Functioning (DIF) Statistics .....	174
10.3	Evaluating Bias through Impact Analysis.....	183
10.4	Reliability.....	183
10.5	Effect Size .....	192
10.6	Summary .....	210
Appendix A—Text Complexity Placemat Template .....		211
Appendix B—Item Content and Bias Review .....		212
Appendix C—Accommodated Print Form Creation .....		219
Appendix D—Item Alignment Review Process .....		221
Appendix E—Transadaptation Process for Spanish Mathematics Forms.....		224
Appendix F—LEAP 2025 Spring 2019 Handscoring/AI Documentation.....		226
Appendix G—Classical Item Statistics.....		307
Appendix H—Student Participation Counts.....		323
References .....		339

## Executive Summary

---

This report is a technical summary of the 2018–2019 administrations of the Louisiana Educational Assessment Program (LEAP 2025) in English Language Arts (ELA) and mathematics for high school. 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 developed by Data Recognition Corporation (DRC) test development staff using the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium's item bank as well as items from the Louisiana Department of Education's own item bank. Items taken from these banks were on pre-established item response theory (IRT) scales. This section provides a summary of the 2018–2019 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 those uses.

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

Chapter 3 of the technical report provides a summary of the test development activities that occurred to create the 2018–2019 operational test forms.

#### *Test Administration (Chapter 4)*

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

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

Chapter 5 of the technical report describes the processes used to score constructed-response and technology-enhanced items. This chapter discusses how scorers are trained and the measures used to ensure consistency among scorers. Finally, this chapter presents the results of the inter-rater reliability studies.

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

Chapter 6 of the technical report includes a detailed description of the operational data analyses of the 2018–2019 LEAP 2025 ELA and mathematics assessments, which include the following major parts: the classical item analysis; calibration, scaling, and linking using IRT models; and student scoring. This chapter also describes the demographics of the calibration samples and compares them to state census data. It reports the results of the classical item analysis and the results of the calibration, scaling, and linking processes.

### *Test Results (Chapter 7)*

Chapter 7 of the technical report contains information on the results of the 2018–2019 LEAP 2025 assessments. Detailed summary statistics based on scale scores and achievement-levels are also provided. Finally, this chapter presents information on the score reports sent to school systems.

### *Performance-Level Setting (Chapter 8)*

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

### *Evidence of Construct-Related Reliability (Chapter 9)*

Chapter 9 of the technical report provides evidence of the reliability and validity of the LEAP 2025 test scores. This chapter provides detailed evidence of the reliability of the tests and 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 designed to determine whether large differences exist within the test results of different demographic groups in Louisiana. The results of the administration mode study are also summarized.

## **E.2 Administration**

Louisiana administered the LEAP 2025 summative assessments in ELA and mathematics to high school students in 2018–2019. Computer-based tests (CBT) were administered during the following three testing windows: November 28 through December 14, 2018; April 15 through May 17, 2019; and June 17–21, 2019. Test administration is discussed in Chapter 4 of this report.

Ninety-eight school systems and twenty-six charter schools administered the ELA and mathematics LEAP 2025 high school tests across the three administrations.

### E.3 Student Performance

Table E.1 presents the percentage of students in 2018–2019 who were classified in each of the achievement levels for each subject. In general students that make up the population for each administration are:

- Fall: students from schools with block schedules and students retesting
- Spring: students from schools with block and regular schedules, as well as students retesting
- Summer: primarily students retesting

**Table E.1 Percentage of Students Classified in Achievement Levels Using 2018–2019 Census Data: English Language Arts and Mathematics**

Administration	Subject	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
Fall	English I	23.8	20.9	21.2	27.3	6.9
	English II	34.9	21.0	17.0	19.0	8.1
	Algebra I	20.8	36.8	19.7	21.1	1.6
	Geometry	11.8	29.4	28.5	26.5	3.9
Spring	English I	12.4	18.5	26.7	34.5	8.0
	English II	18.4	16.3	20.5	32.3	12.5
	Algebra I	11.0	23.5	28.4	32.7	4.4
	Geometry	4.8	28.0	34.0	28.1	5.1
Summer	English I	53.2	36.5	9.1	1.2	0.0
	English II	69.0	23.0	6.7	1.2	0.2
	Algebra I	32.8	46.2	18.3	2.7	0.0
	Geometry	33.5	51.6	7.6	2.2	5.1

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

### E.4 Validity and Test Scores

Most sections of this technical report are designed to provide validity evidence to support the use of the LEAP 2025 test scores. Chapter 2 discusses the uses of the LEAP 2025 test scores. Chapter 3 discusses the test development process used to create the LEAP 2025 tests, which is important to the content-related validity of the LEAP 2025 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 2018–2019 administrations and gives an overview of the score reports that were electronically delivered to the school systems for distribution to schools and parents. Chapter 8 highlights the procedures for performance-level setting implemented by PARCC, which were used because PARCC's standards and achievement levels were used for the LEAP 2025. Chapter 9 discusses reliability and construct-related validity. Chapter 10 gives an overview of the statistical processes used to evaluate bias to ensure the fairness of the LEAP 2025 for all examinees.



## Chapter 1: Introduction

---

The LEAP 2025 assessment system is designed to measure students' knowledge of ELA, mathematics, science, and social studies. This report provides a technical overview of the LEAP 2025 ELA and Mathematics high school assessments administered in the 2018–2019 academic year and presents evidence for the validity of the 2018–2019 LEAP 2025 ELA and mathematics high school assessment scores.

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

### 1.1 Background

In 2010, the Board of Elementary and Secondary Education (BESE) approved the Common Core State Standards (CCSS) in ELA and mathematics. After adopting the CCSS, Louisiana became a governing member of PARCC, a group of states working to develop high-quality assessments that measure the full range of the CCSS. Beginning in 2015, students in grades 3–8 began taking these newly aligned assessments.

In 2016, Louisiana ELA and mathematics academic content standards underwent a review process resulting in the adoption of the Louisiana Student Standards in English language arts and mathematics. In spring 2017, ELA and mathematics students in grades 3–8, except those qualifying for the LEAP Alternate Assessment Level 1 (LAA 1), took the LEAP 2025 assessments.

Beginning in the 2017–2018 school year, the Louisiana Department of Education (LDOE) transitioned to LEAP 2025 ELA and mathematics high school assessments, which were aligned to the Louisiana Student Standards in ELA and mathematics. The five-performance-level LEAP 2025 high school assessments replaced the four-performance-level End-of-Course (EOC) tests. Students enrolled in English I, English II, Algebra I, and Geometry in grades 9–12 took the LEAP 2025 high school assessments.

The information that follows describes the technical aspects of the 2018–2019 LEAP 2025 ELA and Mathematics assessments and provides information about how to read and interpret the data on the 2018–2019 assessment reports.

### 1.2 Purpose of the LEAP 2025

The BESE and the LDOE are committed to ensuring that every student is on track to be successful in either postsecondary education or the workforce 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 for college and careers.

### 1.3 Design of the LEAP 2025

High school students were administered computer-based tests (CBTs) in both ELA and mathematics. Additionally, a braille form was available for each course and content area. Online tools allowed students to magnify assessment items as needed. See Section 3.5 in Chapter 3 for more information about the accommodations and designated supports available for students taking the LEAP 2025. All mathematics assessments were translated into Spanish forms.

The operational blueprints for the PARCC flagship form are the basis of the design of the 2018–2019 LEAP 2025 test blueprints and test design. The 2018–2019 LEAP 2025 test blueprints and test design for ELA and mathematics differ from the PARCC blueprints and design in order to reduce testing time while maintaining full coverage and including a variety of standards.

The 2018–2019 LEAP 2025 ELA blueprints kept a similar design as the design of PARCC’s flagship form, including both performance-based tasks and stand-alone passage sets; 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 LEAP 2025 English I and English II assessments. All three task types are represented across administrations, which encourages teachers to focus equally on all three writing types. Besides having two (instead of three) performance tasks, the 2018–2019 LEAP 2025 ELA blueprints are also different from the PARCC blueprint 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 course’s assessment.

The passages chosen for the 2018–2019 LEAP 2025 English I and English II assessments contain a variety of texts of different genres with a balance of authors by gender and ethnicity. The assessments also contain texts that appeal to a diverse student population. Chosen passages are authentic and contain a variety of different types of text with varying degrees of text complexity. They are rich in content, engaging, high-quality, and challenging. Additionally, paired passages are selected with careful consideration of the standards that require the use of more than one text. This combination of criteria during passage selection allows students to demonstrate their ability to read and comprehend a range of complex texts and helps ensure as much coverage of the standards as possible.

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

The 2018–2019 LEAP 2025 mathematics blueprints kept a similar design as the design of PARCC’s flagship form, with a few notable exceptions:

- Both LEAP 2025 and PARCC have three sessions, with Session 1 split into non-calculator and calculator sections. However, PARCC has three sessions that last 90 minutes each (for a total of 270 minutes), while LEAP 2025 has three sessions that last 80 minutes each (for a total of 240 minutes).
- In Algebra I, PARCC and LEAP 2025 have the same number of Type II items worth 4 points. The LEAP 2025 design uses 1 more Type I item worth 1 point, 2 fewer Type I items worth 2 points, 1 fewer Type I item worth 4 points, 1 fewer Type II item worth 3 points, 1 more Type III item worth 3 points, and 1 fewer Type III item worth 6 points.
- In Geometry, PARCC and LEAP 2025 have the same number of Type II items worth 4 points. The LEAP 2025 design uses 1 fewer Type I item worth 1 point, 1 fewer Type I item worth 2 points, 1 fewer Type I item worth 4 points, 1 fewer Type II item worth 3 points, 1 more Type III item worth 3 points, and 1 fewer Type III item worth 6 points.

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

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

The LSSM support students in becoming 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 solving problems in a relevant and 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 Algebra I and Geometry assessments 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 tests are aligned directly to the LSSM for all reporting categories.

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

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

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

## Chapter 2: The Uses of Test Scores

---

Validity is the central component of 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 use 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 uses of test-level scores (i.e., scale scores and achievement levels), category-level scores and achievement-level classifications, and subcategory-level scores and achievement-level classifications.

### 2.2 Test-Level Scores

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

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

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

## 2.3 Scale Scores

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

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

## 2.4 Levels of Achievement

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

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

## 2.5 Use of Test-Level Scores

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

## 2.6 Category- and Subcategory-Level Subscores

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

Additionally, subcategory subscores are reported at the student level for ELA and mathematics. ELA has three subcategories for reading and two subcategories for writing, as described in Table 3.1, ELA Reporting Categories and Subcategories. Mathematics has four subcategories, as described in Table 3.6, Overview of LEAP 2025 Mathematics Task Types and Reporting Categories. Subcategory performance is reported in one of three ratings: *Strong*, *Moderate*, or *Weak*.

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

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

The purpose of reporting category- or subcategory-level subscores 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 reporting categories and subcategories. 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 category- or subcategory-level subscores. Chapter 9 of this technical report provides evidence of construct-related validity that further supports the use of these subscores.

## Chapter 3: Test Content Development

---

Content-related validity in achievement tests is evidenced by a correspondence between test content and the range of knowledge and skills that compose the construct the assessment is designed to measure (i.e., the ELA or mathematics Louisiana Student Standards). Content-related validity can be demonstrated through consistent adherence to test blueprints, through a high-quality test development process that includes review of items for accessibility to English Learners and students with disabilities, and through alignment studies performed by independent groups. This chapter provides a detailed discussion of the test development process. In particular, it shows how rigorous procedures were followed to construct tests that reflect the full range of content that the 2018–2019 LEAP 2025 high school 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 2018–2019 LEAP 2025 high school test specifications consisted of a blueprint and a design for each of the following tests: English I, English II, Algebra I, and Geometry. The 2018–2019 blueprints and test designs were closely aligned to the PARCC flagship blueprints that were used for the PARCC 2018–2019 test administrations. The test blueprints for the 2018–2019 LEAP 2025 high school ELA assessments were designed with the goal of having all students read, understand, and express their understanding of complex, grade-level texts. The test blueprints for the 2018–2019 LEAP 2025 mathematics assessments 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 2018–2019 LEAP 2025 high school ELA and mathematics assessments provided questions that were reviewed by Louisiana educators to ensure their alignment to the Louisiana Student Standards and appropriateness for Louisiana students; measured the full range of student performance, including the performance of high- and low-performing students; and informed educators and parents about student readiness in ELA and mathematics and whether students are “on track” for college and careers. For ELA and mathematics, the 2018–2019 LEAP 2025 assessments use the same blueprints and reporting categories and subcategories that were used in 2017–2018.

To construct the assessments following the LDOE-approved test blueprints and test designs, LDOE and DRC collaborated to use items from the PARCC- and Louisiana-owned item banks. Both item banks are comprised of items aligned to the Louisiana Student Standards. DRC contracted with PARCC and was provided access to the entire bank of items and passage sets that could potentially be used on operational forms. The acquired items and passages and the Louisiana-owned items and passage sets made up the available item pool used for the 2018–2019 LEAP 2025 high school 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 [New Meridian website](#). LDOE and DRC confirmed that all items selected for use on the LEAP 2025 forms were appropriate for use on Louisiana assessments by convening committees of Louisiana educators who reviewed and approved items from the item banks prior to form selection.

The ELA and mathematics LEAP 2025 assessments for high school were developed based on the requirements of “RFP #678PUR-LEAP 2025 English Language Arts and Mathematics Assessment System” as follows:

The assessments shall be

- aligned to the ELA and mathematics Louisiana Student Standards;
- designed to be accessible for use by the widest possible range of students, including, but not limited to, students with disabilities and students with limited English proficiency;
- constructed to yield valid and reliable test results;
- constructed to report student performance using achievement level policy definitions and reporting categories that are comparable to a significant number of other states;
- developed to limit the amount of testing time required and to be in compliance with all state laws regarding testing time;
- developed and reviewed with Louisiana educator involvement;
- 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.

The products of the above requirements are computer-based tests (CBTs) made of PARCC and Louisiana-owned items aligned to the Louisiana Student Standards. Louisiana had access to the complete PARCC item bank when selecting items to build the forms needed for the 2018–2019 LEAP 2025 high school ELA and mathematics assessments. Items and passage sets were deemed appropriate for use on Louisiana assessments by Louisiana educators during an item alignment review. These items and passage sets were approved because they aligned to the Louisiana Student Standards and/or Louisiana Evidence Statements for mathematics and because they were free of issues related to bias, fairness, and sensitivity. These items and passage sets became the available item pool used to construct the 2018–2019 forms. For each course, three test forms were selected from the available pool of items and/or passage sets. 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 in each course and are comparable to the PARCC 2018–2019 flagship 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 *2018–2019 LEAP 2025 High School English Language Arts and Mathematics*



*Assessment Frameworks.* All assessments were fixed forms, which means that all students who received the same form were administered the same set of items, as the forms were not adaptive.

The LEAP 2025 high school assessments are administered in fall, spring, and summer each school year. For fall and summer administrations, two forms are administered: an operational form and an administrative error form, which is used only if there is an administrative testing error (see Chapter 4 for additional details regarding the administrative error form). For spring administrations, two operational forms and one administrative error form are administered. In addition, spring administrations also include a senior-only form to allow students who will be graduating to receive their assessment results earlier than students who take the operational forms. The forms are administered on a rotating schedule, so they are not the same from administration to administration.

### 3.1 Defining the Specific Test Blueprint

The test blueprints for the 2018–2019 assessments were designed based on two primary factors: (1) the content requirements of the Louisiana Student Standards and (2) the reporting needs of the assessments.

#### English I and English II Test Blueprints and Test Designs

The English I and English II tests were administered during operational testing windows: November 28 to December 14, 2018; April 15 to May 17, 2019; and June 17 to 21, 2019. 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 tests. All three types were represented across administrations (fall, spring, and summer), which encourages educators to focus on all three writing types. 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—are created for each course.

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

**Table 3.1 ELA Categories and Subcategories**

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

These reporting categories provide parents and educators with valuable information about

- overall student performance, including readiness to continue further studies in English language arts;
- student performance broken down by subcategory, which may help identify when students need additional support or more challenging work in reading and writing; and

- how well schools and school systems help students achieve expectations.

The session testing times shown in the ELA test designs (see Tables 3.2 and 3.3) are based on PARCC testing times proportioned to be comparable based on the passage type being tested. The passage set that comes after the Narrative Writing Task or the Literary Analysis Task is designed to balance the reading load between the Narrative Writing Task or the Literary Analysis Task and the Research Simulation Task and to provide consistent timing in sessions 1 and 2.

**Table 3.2 English I and English II Test Design—Literary Analysis Task and Research Simulation Task**

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

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

\*\*The time in session 3 allows for additional passage set that is a field-test or placeholder passage set.

Table 3.3 English I and English II Test Design—Research Simulation Task and Narrative Writing Task

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

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

\*\*The time in session 3 allows for additional passage set that is a field-test or placeholder passage set.

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

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

The following item types were included in the 2018–2019 LEAP 2025 high school 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 with opportunities to make explicit the evidence that supports their close analysis of a specific text.
  - **Multiple select (MS):** This item type requires students to select more than one correct answer and may appear as a one-part question or as part of an EBSR item. This type of item allows for the assessment of students' ability to identify multiple pieces of evidence to support a claim.
  - **Technology enhanced (TE):** This item type allows measurement of learning that may not be sufficiently measured by traditional multiple-choice items. TE items can measure the ordering of ideas within a summary; ordering of steps in a process; sorting, classifying, and categorizing ideas; matching of two themes/ideas to their unique evidence, etc. The technology used in TE items offers students additional ways to show understanding that parallels the classroom instructional techniques that teachers use to determine whether

students are able to comprehend complex, grade-level text. TE Items may involve any of the following:

- Highlighting text: requires students to select text-based answer(s) from within a larger text
- Drag and drop: requires students to move draggable elements (e.g., words, phrases, or sentences) into one or more drop boxes (e.g., cells within a table or part[s] of a diagram)
- Drop-down menu: requires students to select from one or more drop-down menus to complete a phrase or sentence
- Match interaction table: requires students 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 task and asks students to create an extended, complete written response. It elicits evidence that students have understood a text or texts they have read and can communicate that understanding well, both in terms of written expression and in terms of knowledge 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 to 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.

The following tables detail the number of items and points by session and item type for English I and English II forms.

**Table 3.4 Distribution of English I Items and Points by Session and Item Type**

Form	Administration	Session	EBSR		MS		TE		PCR		Total No. of Pts.
			No. of Items	No. of Pts.	No. of Items	No. of Pts.	No. of Items	No. of Pts.	No. of Items	No. of Pts.	
A	Spring (SR) Summer (OP)	1. Research Simulation Task	6	12	0	0	2	4	1	19	90
		2. Narrative Writing Task/ Reading Passage	5	10	2	4	3	6	1	15	
		3. Reading Literary/ Informational Texts	8	16	1	2	1	2	0	0	
B	Fall (OP)	1. Literary Analysis Task/ Reading Passage	8	16	0	0	2	4	1	19	94
		2. Research Simulation Task	6	12	0	0	2	4	1	19	
		3. Reading Literary/ Informational Texts	6	12	1	2	3	6	0	0	
C	Fall (AE) Spring (AE) Summer (AE)	1. Research Simulation Task	6	12	1	2	1	2	1	19	90
		2. Narrative Writing Task/ Reading Passage	5	10	3	6	2	4	1	15	
		3. Reading Literary/ Informational Texts	6	12	3	6	1	2	0	0	
D	Spring (OP)	1. Literary Analysis Task/ Reading Passage	9	18	0	0	1	2	1	19	94
		2. Research Simulation Task	4	8	2	4	2	4	1	19	
		3. Reading Literary/ Informational Texts	8	16	0	0	2	4	0	0	
E	Spring (OP)	1. Research Simulation Task	4	8	2	4	2	4	1	19	90
		2. Narrative Writing Task/ Reading Passage	4	8	2	4	4	8	1	15	
		3. Reading Literary/ Informational Texts	6	12	1	2	3	6	0	0	

OP: Operational  
AE: Administrative Error  
SR: Senior

Table 3.5 Distribution of English II Items and Points by Session and Item Type

Form	Administration	Session	EBSR		MS		TE		PCR		Total No. of Pts.
			No. of Items	No. of Pts.	No. of Items	No. of Pts.	No. of Items	No. of Pts.	No. of Items	No. of Pts.	
A	Spring (SR)	1. Research Simulation Task	5	10	2	4	1	2	1	19	90
		2. Narrative Writing Task/ Reading Passage	2	4	7	14	1	2	1	15	
		3. Reading Literary/ Informational Texts	5	10	4	8	1	2	0	0	
B	Fall (OP) Summer (OP)	1. Literary Analysis Task/ Reading Passage	4	8	3	6	3	6	1	19	94
		2. Research Simulation Task	5	10	2	4	1	2	1	19	
		3. Reading Literary/ Informational Texts	4	8	4	8	2	4	0	0	
C	Fall (AE) Spring (AE) Summer (AE)	1. Literary Analysis Task/ Reading Passage	8	16	0	0	2	4	1	19	94
		2. Research Simulation Task	5	10	1	2	2	4	1	19	
		3. Reading Literary/ Informational Texts	2	4	4	8	4	8	0	0	
D	Spring (OP)	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	1	2	3	6	0	0	
E	Spring (OP)	1. Research Simulation Task	4	8	2	4	2	4	1	19	90
		2. Narrative Writing Task/ Reading Passage	7	14	1	2	2	4	1	15	
		3. Reading Literary/ Informational Texts	2	4	5	10	3	6	0	0	

OP: Operational

AE: Administrative Error

SR: Senior



## Mathematics Test Blueprints and Test Designs

The mathematics assessments were administered during operational testing windows: November 28 to December 14, 2018; April 15 to May 17, 2019; and June 17 to 21, 2019. The 2018–2019 mathematics assessments included three test sessions, and each test session included the four mathematics subcategories and the three mathematics task types. See Table 3.6 for details about categories and task types.

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

**Table 3.6 Overview of LEAP 2025 Mathematics Task Types and Reporting Categories**

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

These reporting categories provide parents and educators with valuable information about

- overall student performance, including readiness to continue further studies in mathematics;
- student performance broken down by mathematics subcategories, 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.

Tables 3.7 and 3.8 provide the distribution of operational points by reporting category and by form for each mathematics course.

**Table 3.7 Distribution of Points by Reporting Category—Algebra**

Reporting Category	Form			
	AR	BR	D	E
Major Content	28	28	28	28
Additional & Supporting Content	13*	14	14	14
Expressing Mathematical Reasoning	11	11	11	11
Modeling & Application	15	15	15	15
<b>Total</b>	<b>67</b>	<b>68</b>	<b>68</b>	<b>68</b>

\* A one-point item within the Additional & Supporting Content reporting category was dropped from scoring in Form AR.

**Table 3.8 Distribution of Points by Reporting Category—Geometry**

Reporting Category	Form			
	AR	BR	D	E
Major Content	26	26	26	26
Additional & Supporting Content	16	16	16	16
Expressing Mathematical Reasoning	11	11	11	11
Modeling & Application	15	15	15	15
<b>Total</b>	<b>68</b>	<b>68</b>	<b>68</b>	<b>68</b>

The Major Content category for mathematics is broken into subcategories by course as follows:

**Table 3.9 Major Content Subcategories by Course**

Course	Major Content Subcategories
Algebra	<ul style="list-style-type: none"> <li>• Interpreting Functions</li> <li>• Solving Algebraically</li> <li>• Solving Graphically/Rate of Change</li> </ul>

Course	Major Content Subcategories
Geometry	<ul style="list-style-type: none"> <li>• Congruence Transformations/Similarity</li> <li>• Similarity in Trigonometry/Modeling &amp; Applying</li> </ul>

The resulting 2018–2019 LEAP 2025 mathematics test blueprints are shown in Tables 3.10 and 3.11.

**Table 3.10 Algebra I Test Blueprint**

Reporting Category	Major Content	Additional & Supporting Content	Expressing Mathematical Reasoning	Modeling & Application
Task Type	Type I: I.1 (24 items, 24 points) I.2 (7 items, 14 points) I.4 (1 item, 4 points) Total: 32 items, 42 points (62% of total)		Type II: II.3 (1 item, 3 points) II.4 (2 items, 8 points) Total: 3 items, 11 points (16% of total)	Type III: III.3 (3 items, 9 points) III.6 (1 item, 6 points) Total 4 items, 15 points (22% of total)
Total OP Points	28 (41% of total)	14 (21% of total)	11 (16 % of total)	15 (22% of total)
Assessable Content	A1: A-APR.A.1 A1: A-CED.A.3 A1: A-CED.A.4 A1: A-REI.B.3 A1: A-REI.B.4a A1: A-REI.B.4b A1: A-REI.D.10 A1: A-REI.D.11 A1: A-REI.D.12 A1: A-SSE.A.1a A1: A-SSE.A.1b A1: A-SSE.A.2 A1: F-IF.A.1 A1: F-IF.A.2 A1: F-IF.B.4 A1: F-IF.B.5 A1: F-IF.B.6 LEAP.I.A1.1 LEAP.I.A1.2 LEAP.I.A1.3 LEAP.I.A1.4 LEAP.I.A1.5 LEAP.I.A1.6	A1: A-APR.B.3 A1: A-REI.C.6 A1: A-SSE.B.3a A1: A-SSE.B.3b A1: A-SSE.B.3c A1: F-BF.B.3 A1: F-IF.C.7a A1: F-IF.C.7b A1: F-IF.C.8a A1: F-IF.C.9 A1: F-LE.A.2 A1: S-ID.B.5 LEAP.I.A1.7	LEAP.II.A1.1 LEAP.II.A1.2 LEAP.II.A1.3 LEAP.II.A1.4 LEAP.II.A1.5 LEAP.II.A1.6 LEAP.II.A1.7 LEAP.II.A1.8 LEAP.II.A1.9 LEAP.II.A1.10	LEAP.III.A1.1 LEAP.III.A1.2 LEAP.III.A1.3 LEAP.III.A1.4

**Table 3.11 Geometry Test Blueprint**

Reporting Category	Major Content	Additional & Supporting Content	Expressing Mathematical Reasoning	Modeling & Application
Task Type	Type I: I.1 (24 items, 24 points) I.2 (7 items, 14 points) I.4 (1 item, 4 points) Total: 32 items, 42 points (62% of total)		Type II: II.3 (1 item, 3 points) II.4 (2 items, 8 points) Total: 3 items, 11 points (16% of total)	Type III: III.3 (3 items, 9 points) III.6 (1 item, 6 points) Total 4 items, 15 points (22% of total)
Total OP Points	26 (38% of total)	16 (24% of total)	11 (16% of total)	15 (22% of total)
Assessable Content	GM: G-CO.B.6 GM: G-GPE.B.6 GM: G-SRT.A.1a GM: G-SRT.A.1b GM: G-SRT.A.2 GM: G-SRT.B.5 GM: G-SRT.C.6 GM: G-SRT.C.7 GM: G-SRT.C.8 LEAP.I.GM.1 LEAP.I.GM.2	GM: G-C.A.2 GM: G-CO.A.1 GM: G-CO.A.3 GM: G-CO.A.5 GM: G-GMD.A.1 GM: G-GMD.A.3 GM: G-GMD.B.4 GM: G-GPE.A.1 LEAP.I.GM.3 LEAP.I.GM.4 LEAP.I.GM.5	LEAP.II.GM.1 LEAP.II.GM.2 LEAP.II.GM.3 LEAP.II.GM.4	LEAP.III.GM.1 LEAP.III.GM.2 LEAP.III.GM.3 LEAP.III.GM.4 LEAP.III.GM.5

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 session 1a of the mathematics assessments, students were prohibited from using calculators, except those students with a calculator accommodation. Calculators were allowed to be used by all students in sessions 1b, 2, and 3. The general test structures (see Tables 3.12 and 3.13) guided test form sequencing and design. The [LEAP 2025 Calculator Policy](#) provided the basis for calculator designation of tasks and items.

**Table 3.12 Algebra I Testing Sessions**

Reporting Category	Session 1a: No Calculator	Session 1b: Calculator	Session 2: Calculator	Session 3: Calculator	Total
Major Content (points)	5	5	9	9	28
Additional & Supporting Content (points)	4	2	4	4	14
Expressing Mathematical Reasoning (points)	0	3	4	4	11
Modeling & Application (points)	0	3	6	6	15
Total Operational Points	9	13	23	23	68
Test Duration *(minutes)	25	55	80	80	240
# of Operational Items	I.1: 5 I.2: 2 I.4: 0 II.3: 0 II.4: 0 III.3: 0 III.6: 0	I.1: 3 I.2: 0 I.4: 1 II.3: 1 II.4: 0 III.3: 1 III.6: 0	I.1: 9 I.2: 2 I.4: 0 II.3: 0 II.4: 1 III.3: 0 III.6: 1	I.1: 7 I.2: 3 I.4: 0 II.3: 0 II.4: 1 III.3: 2 III.6: 0	I.1: 24 I.2: 7 I.4: 1 II.3: 1 II.4: 2 III.3: 3 III.6: 1

\* Six Embedded Field Test Items were included throughout the assessment; they are included in the total time.

**Table 3.13 Geometry Testing Sessions**

Reporting Category	Session 1a: No Calculator	Session 1b: Calculator	Session 2: Calculator	Session 3: Calculator	Total
Major Content (points)	5	5	8	8	26
Additional & Supporting Content (points)	4	2	5	5	16
Expressing Mathematical Reasoning (points)	0	3	4	4	11
Modeling & Application (points)	0	3	6	6	15
Total Operational Points	9	13	23	23	68
Test Duration *(minutes)	25	55	80	80	240
# of Operational Items	I.1: 5 I.2: 2 I.4: 0 II.3: 0 II.4: 0 III.3: 0 III.6: 0	I.1: 3 I.2: 0 I.4: 1 II.3: 1 II.4: 0 III.3: 1 III.6: 0	I.1: 7 I.2: 3 I.4: 0 II.3: 0 II.4: 1 III.3: 0 III.6: 1	I.1: 9 I.2: 2 I.4: 0 II.3: 0 II.4: 1 III.3: 2 III.6: 0	I.1: 24 I.2: 7 I.4: 1 II.3: 1 II.4: 2 III.3: 3 III.6: 1

\* Six Embedded Field Test Items were included throughout the assessment; they are included in the total time.

The following item types were used in the 2018–2019 LEAP 2025 mathematics assessments:

- Multiple-choice: This item type requires students to select one correct answer from four answer choices. It may appear as a one-part question, as part of a two-part question, or as part of a constructed-response item. The multiple-choice items are worth one point.
- Multiple select: This item type requires students to select more than one correct answer from more than four answer choices. It may appear as a one-part question, as part of a two-part question, or as part of a constructed-response item. The multiple select items are worth one point. Students must choose all correct answers and no incorrect answers to receive credit.
- Short answer: This item type requires students to enter a numeric response by typing from the keyboard. It may appear as a one-part question, as part of a two-part question, or as part of a constructed-response item. The short answer items are worth one point. Unless specified in the question, students will earn credit for an answer that is equivalent to the correct numerical answer. Proper rounding may be required. Answers to short answer items can be positive or negative and must be entered in integer or decimal form.
- Keypad input items: This item type requires students to enter a mathematical response using a customized pallet of numbers, operations, variables, and/or mathematical symbols; allows the use of all rational and irrational numbers, 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 part of a constructed-response item.
- Constructed-response items: This item type requires students to respond to an open-ended question, which must be typed into a response box; students may use the equation builder tool (specific to the course) 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, multistep contextual problems. Students 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 items: 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 part of a constructed-response item. The technology-enhanced items are worth one point. Technology-enhanced items may involve any of the following:
  - Bar graph: requires students to complete a bar graph or histogram by raising/lowering each bar to a value
  - Drag and drop: requires students to move draggable elements into one or more drop boxes
  - Drop-down menu: requires students to select from one or more drop-down menus to complete a sentence, phrase, or expression/equation/inequality
  - Hot spot: requires students to select one or more responses by choosing selectable areas on the screen
  - Match interaction table: requires students to select a checkbox in each row from two or more columns
  - Graph input: requires students to enter a response on a coordinate grid
  - Number line input: requires students to enter a response on a number line
  - Line plot: requires students to complete a line plot with “X” as the input

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

The following table details the number of items by point value and task type and the number of points per task type for each form.

**Table 3.14 Distribution of Mathematics Tasks and Points by Task Type**

Form	Content Area	Type I				Type II			Type III			Total Number of Points
		1 pt. Tasks	2 pt. Tasks	4 pt. Tasks	Points	3 pt. Tasks	4 pt. Tasks	Points	3 pt. Tasks	6 pt. Tasks	Points	
AR	Algebra	23*	7	1	41	1	2	11	3	1	15	67*
BR	Algebra	24	7	1	42	1	2	11	3	1	15	68
D	Algebra	24	7	1	42	1	2	11	3	1	15	68
E	Algebra	24	7	1	42	1	2	11	3	1	15	68
AR	Geometry	24	7	1	42	1	2	11	1	2	15	68
BR	Geometry	24	7	1	42	1	2	11	1	2	15	68
D	Geometry	24	7	1	42	1	2	11	3	1	15	68
E	Geometry	24	7	1	42	1	2	11	3	1	15	68

\* One Type I item worth 1 point was dropped from scoring in form AR.



## Item Development and Selection

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

Standard 4.7 states the following:

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

The items used in the 2018–2019 LEAP 2025 high school ELA and mathematics assessments came from the PARCC consortium’s item bank and the Louisiana-owned item bank.

The items selected for use on the 2018–2019 LEAP high school forms were used to equate to the LEAP 2025 scale, which is comparable to the PARCC scale. Operational forms were selected based on LEAP 2025 high school test blueprint specifications, which were supported by statistical data from PARCC operational testing.

## Considerations of Test Fairness in Item Development

Standard 3.2 is particularly relevant to fairness in item development:

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

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

## PARCC 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 that included items and passages available for selection on Louisiana forms. These reviews include

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

Additional information on PARCC’s item review processes and procedures can be found at the [New Meridian Resource Center](#). Only items that have been approved by expert reviewers during text reviews (ELA only), item reviews, bias and sensitivity reviews, and editorial reviews are moved forward for field-testing 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 criteria set forth by the *Standards*.

Standard 3.1 states the following:

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

Standard 3.2 states the following:

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

Independent studies of 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 New Meridian’s website: <https://resources.newmeridiancorp.org/research/>.

## 3.2 Louisiana Item Development and Item Review

### ELA Development Process

Item development for ELA began with a detailed analysis of the acquired item bank to determine the needs of the pool. This analysis resulted in development targets for each course beginning with the selection of passages. Development targets indicated whether passage sets should be short, long, or paired. They also determined whether passages should be literary or informational. Additional traits for each target, such as text complexity, standards that should be assessed, genre coverage, gender representation, topic variety, and ways to add diversity to the pool of passages available, were also provided. Once the targets were identified and approved by LDOE ELA content experts, DRC’s ELA test development team worked to provide options for each target for LDOE review. Experienced passage finders recommended authentic texts, including permissioned and/or public domain passages. These initial selections were reviewed by DRC’s ELA test development team members, who then analyzed the text complexity of each passage. The passages, any associated graphics, and the results of text complexity analyses were provided to LDOE. LDOE’s ELA team reviewed the options and provided feedback to ensure that two options for each target were ready for review by Louisiana educators.

### Text Complexity Specifications for Field Test Passages

As part of the passage development process, a passage’s text complexity is analyzed so that an appropriate grade-level placement for each passage can be made. The analysis of the passage’s text complexity was captured on a placemat. (Please see Appendix A for a sample placemat.) DRC used a process that includes (1) a quantitative evaluation of the text and (2) a qualitative evaluation of the text. Passages and their respective placemats are submitted to LDOE during initial passage reviews. In addition, a third component, matching reader to text and task, is also taken into consideration during passage evaluation.

## Passage Review

In June 2018, in conjunction with the alignment reviews of items from the acquired item bank, passage reviews were conducted by Louisiana educator committees. During the review process, the committees, which represented a variety of perspectives, reviewed the proposed literary and informational passages to ensure the texts used to develop passage sets on the LEAP 2025 ELA tests were fair and appropriate for all students and would allow an opportunity for students to demonstrate their knowledge and skills in ELA. Educators reviewed the passages and provided feedback and a consensus decision about the status of each passage. The status identified whether the passage was acceptable to move forward with development or not acceptable to move forward with development. Educators also provided individual rankings of the preference for passages of each target type. Upon conclusion of the passage reviews, members of the LDOE and DRC's ELA test development teams met to discuss the results. Decisions regarding which passages would move forward to item development were made at that time.

Table 3.15 provides the count of passages brought to passage review and the status of those passages after passage review.

**Table 3.15. LEAP 2025 June 2018 Passage Review**

Course	Number of Passages Reviewed	Accepted	Rejected
English I	7	7	0
English II	8*	7	1

\*Of the 8 passages, 4 were paired sets; 1 passage in a pair was rejected.

## ELA Item Writing and Review

Once passage sets were approved for development by Louisiana educators and content experts from the LDOE, the passage sets were provided to experienced item writers for development. Item writers participated in item writing training with Pearson, DRC's subcontractor, prior to developing items. The training involved discussion of the following:

- Passage set quality
  - Passage sets have value.
  - Passages sets are cohesive.
  - Items are text dependent.
  - Items are aligned to and reflect the rigor of the Louisiana Student Standards.
- Item type descriptions and examples of each
  - Evidence-Based Selected Response
  - Multiple Select
  - Technology Enhanced
- Resources to support item writing
  - PARCC practice tests
  - PARCC released items
- Training on universal design and bias, fairness, and sensitivity
- Training on security and confidentiality

Once the items were written, they were revised as necessary by Pearson prior to delivery to the LDOE. The LDOE reviewed each batch of items and provided feedback that was implemented prior to the passage sets being prepared in mock test forms for review by committees of Louisiana educators at item content and bias review meetings.

At the item content and bias review meetings, course-level committees of ELA educators met to provide feedback on the alignment and appropriateness of items for use on Louisiana assessments. Louisiana educators reviewed items for alignment to content standards; grade appropriateness; issues of bias, fairness, and sensitivity; difficulty and cognitive complexity; and clarity of language. The discussion about difficulty and cognitive complexity included not only approving the cognitive complexity levels assigned to each item but also ensuring that the difficulty and cognitive complexity were appropriate for the course. Louisiana educators reviewed the items to ensure they represented a range of difficulty and cognitive complexity. Louisiana educators edited items as needed to ensure they were appropriate for use on Louisiana assessments, which allowed the items to move forward for possible field-testing. Any items deemed inappropriate were rejected if educators were not able to revise or recommend appropriate revisions for those items. Items that successfully passed through the item content and bias reviews were then embedded within operational test forms for field-testing, and data was collected on each field test item. For a detailed description of the process followed during the content and bias review meetings, see Appendix B.

### Mathematics Item Development

To determine the mathematics item development needs for field-testing in the Spring 2019 administration, the LDOE determined the count of items needed per course and then DRC content experts analyzed the item pool to determine the number of type II or type III items and the evidence statements/standards based on that analysis. DRC content experts reviewed standards coverage on the previous year's test by looking at the number and types of items used to cover each content standard, the difficulty range, the level of cognitive complexity covered by each content standard, and the topic/material presented in items (to ensure a variety of engaging topics are included). DRC determined gaps or holes in coverage, based on these criteria, to create an item development plan for the number and types of items to be newly developed for possible field-testing in spring 2019. DRC presented the item development plan to LDOE content experts, who then provided feedback to DRC. DRC and the LDOE collaborated to finalize the item development plan. DRC used Pearson as a subcontractor to have items written. Item writers participated in item writing training with Pearson and LDOE prior to developing items. The training included

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

These items were reviewed by LDOE and revised by Pearson. Once items were approved by LDOE, they became part of the set of items that were taken to item content and bias reviews with Louisiana educators in summer 2018. Refer to Appendix C, "Item Development Plans", for counts of the items developed for content and bias reviews and field-testing.

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

### 3.3 Guidelines on Bias, Fairness, and Sensitivity

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

#### **Types of Bias:**

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

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

### Louisiana Item Alignment Review

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

DRC, with guidance from LDOE, conducted the Louisiana item alignment reviews in June 2017 with committees of Louisiana educators. Course-specific committees met for two days for mathematics and three days for ELA to provide feedback on the alignment and appropriateness of items that made up the PARCC item bank. To the extent possible, each committee included educators from different parts of Louisiana to represent all Louisiana students (e.g., special education students, English Learners, students with disabilities). Committee members were also representative of the diverse demographics of the state.

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

### 3.3 Operational Test Selection

Operational item selection for the 2018–2019 administration took place from June through September 2018 by LDOE and DRC. The PARCC and Louisiana item pools were used to select fixed LEAP 2025 ELA and mathematics high school forms.

The LEAP 2025 high school assessments were given as computer-based tests (CBTs). For students unable to participate in a CBT administration, accommodated print forms were available for secure download and

printing by authorized users. Test administrators transcribed all student responses into the appropriate CBT test form. (See Chapter 4 for additional details.)

### Item and Passage Selection Process and Criteria

The item and passage selection process used for forms construction was a content-focused, collaborative process between the LDOE and DRC ELA and mathematics content specialists, and it was followed by a psychometric evaluation of each selection. The critical psychometric consideration, other than individual item performance, was the degree to which the selected items reflected the 2018–2019 LEAP 2025 targets, which were supposed to match the Spring 2018 LEAP 2025 operational forms. Although the item pool was limited, items that were determined to be very difficult (i.e., IRT difficulty parameter  $b > 2.0$ ) and/or not discriminating (i.e., IRT discrimination parameter  $a < 0.3$ ) were avoided when possible.

#### *Item Selection Guidelines*

- Using the acquired pool of items, content-area assessment specialists select ELA passage sets and tasks that consist of quality texts displaying diversity in topics and authors and mathematics tasks that match the blueprint. The sets and tasks include items that cover a range of Louisiana Student Standards and/or Evidence Statements (mathematics only) and address the appropriate reporting categories.
- Content-area assessment specialists and research analysts verify that each item meets psychometric guidelines for excellence as available item-performance data allows.
- Forms include adequate content coverage, as required by the detailed test blueprint.
- Each form contains an anchor set that includes passage sets or tasks from a previous administration. The anchor set, which is a mini-blueprint of the form, ensures comparability between the 2018 form and the 2019 form. The remaining sets or tasks selected for a form complete the blueprint requirements.
- No item in a form should “clue” (or provide the answer to) another item on that same form.
- Clang association should be avoided. Clang is when a distractor can be associated with a stem word by sound rather than meaning (e.g., rhyming, alliteration).
- Passage sets in ELA forms should be diverse.
- Forms should be diverse, including a variety of text types, including texts that appeal to a diverse student population.
- Forms should include a wide range of topics and a variety of questions.
- Correct answer distributions should follow best practice (no more than 3 keys of the same answer option in a row).
- Forms **must not** contain any items that have been released to the public.

### Review of the ELA Items and Forms

DRC and LDOE ELA content specialists and members of educator committees verified that the items were in compliance with the guidelines provided by LDOE, including alignment to the content standards and appropriateness for Louisiana students. Because establishing content validity is one of the most important aspects in the legal defensibility of a test, the alignment of the items to the content standards must be reviewed and verified at every stage of the test development process. As a result, it is essential that an item selected for a form link directly to the content standard that it purports to measure. The ELA content specialists also verified all items against their designated content codes and metadata, both to evaluate the correctness of the coding and to ensure that the given item measures what it purports to measure.

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

### *Guidelines for Reviewing Items Selected for Forms*

A good item should

- have the appropriate number of correct answer(s) based on the item type;
- have only one clear, correct answer for each part of an evidence-based selected response (ESR) item that has only four answer options in each part;
- have only the indicated number of correct answers for a multiple select (MS) item or item part;
- 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 and the students' 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.

### *Review of the Mathematics Items and Forms*

DRC and LDOE mathematics content specialists also ensured the items were in compliance with the guidelines provided by LDOE, including alignment to the content standards and appropriateness for Louisiana students. Since establishing content validity is one of the most important aspects in the legal defensibility of a test, the alignment of the items to the content standards must be reviewed and verified at every stage of the test development process. As a result, it is essential that an item selected for a form link directly to the content standard that it purports to measure. The mathematics content specialists also verified all items against their designated content codes and metadata, both to evaluate the accuracy of the coding and to ensure that the given item measures what it purports to measure.



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

### *Guidelines for Reviewing Items Selected for Forms*

A good item should

- contain answer choices that are reasonably parallel in length and structure;
- have the appropriate number of correct answer(s) based on item type:
  - only one clear, correct answer for a multiple-choice (MC) item
  - only the indicated number of correct answers for a multiple select (MS) item;
- have a correctly assigned content code (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 and the student’s level of knowledge;
- contain distractors that relate to the question in the same way and can be supported by a rationale;
- reflect current teaching and learning practices in the content area; and
- be free of gender, ethnic, racial, cultural, socioeconomic, regional, and other forms of bias.

### Item-Selection Options for Special Cases

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

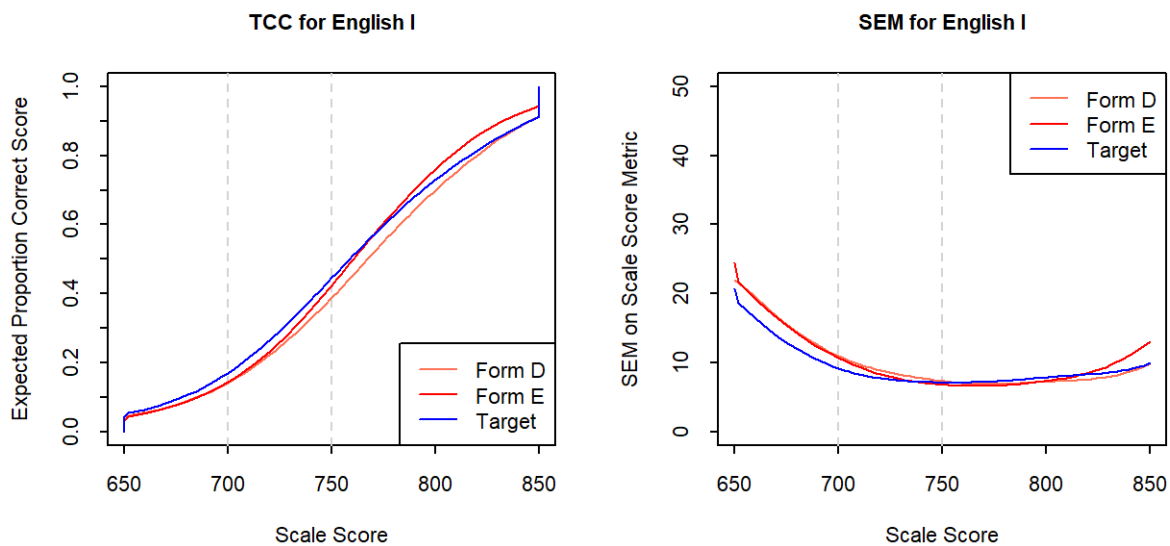
### Psychometric Review

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

## Selecting Targets

The spring 2018 LEAP 2025 operational form was selected to be the target form in 2018–2019 LEAP 2025 form construction. The rationale for the choice of the targets was that each 2017–2018 LEAP 2025 form should be on the PARCC scale and be closely comparable to PARCC assessments. Figure 3.1 and Figure 3.2 for English I and II and Figure 3.3 and Figure 3.4 for Algebra I and Geometry show the test characteristic curves (TCCs) and standard errors of measurement (SEMs) of the final forms compared to those of the target forms. The left line graph displays the TCC of the target form and the selected 2018–2019 forms, 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 2018–2019 forms. This summarizes the amount of measurement error surrounding a scale score.

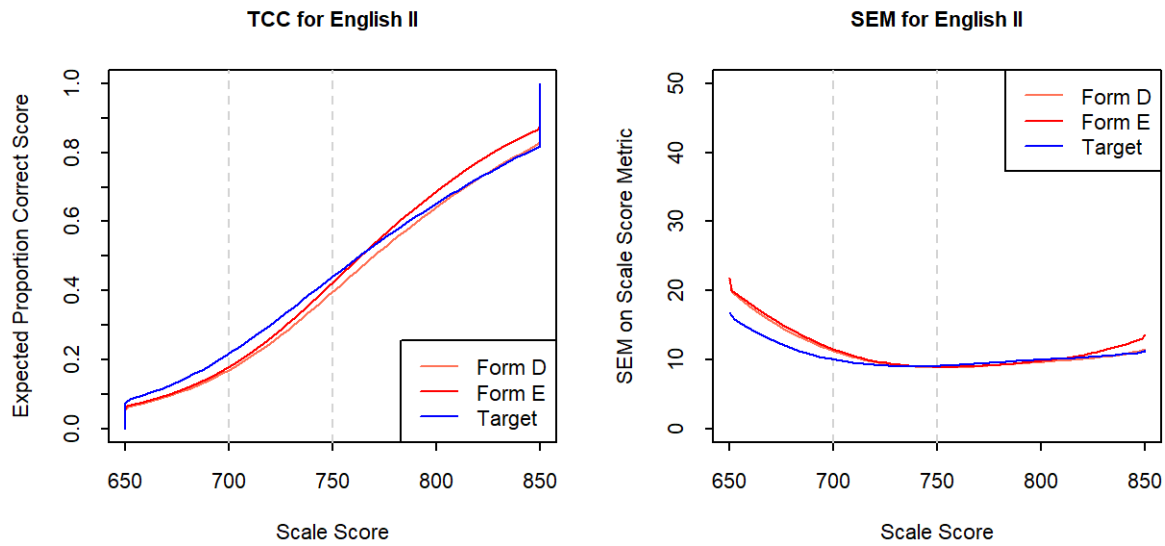
**Figure 3.1 Spring English I Form-Building Evaluation for 2018–2019 Administrations**



Notes:

- The target form is the Spring 2018 LEAP 2025 HS test form.
- Forms D and E are Spring 2019 LEAP 2025 HS test forms.

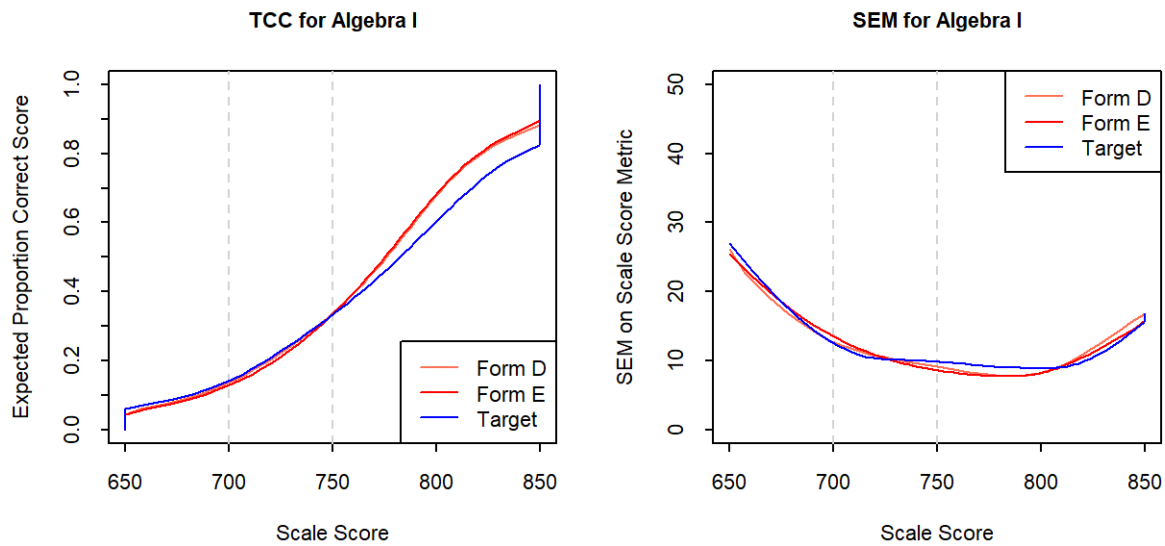
**Figure 3.2 Spring English II Form-Building Evaluation for 2018–2019 Administrations**



Notes:

- The target form is the Spring 2018 LEAP 2025 HS test form.
- Forms D and E are Spring 2019 LEAP 2025 HS test forms.

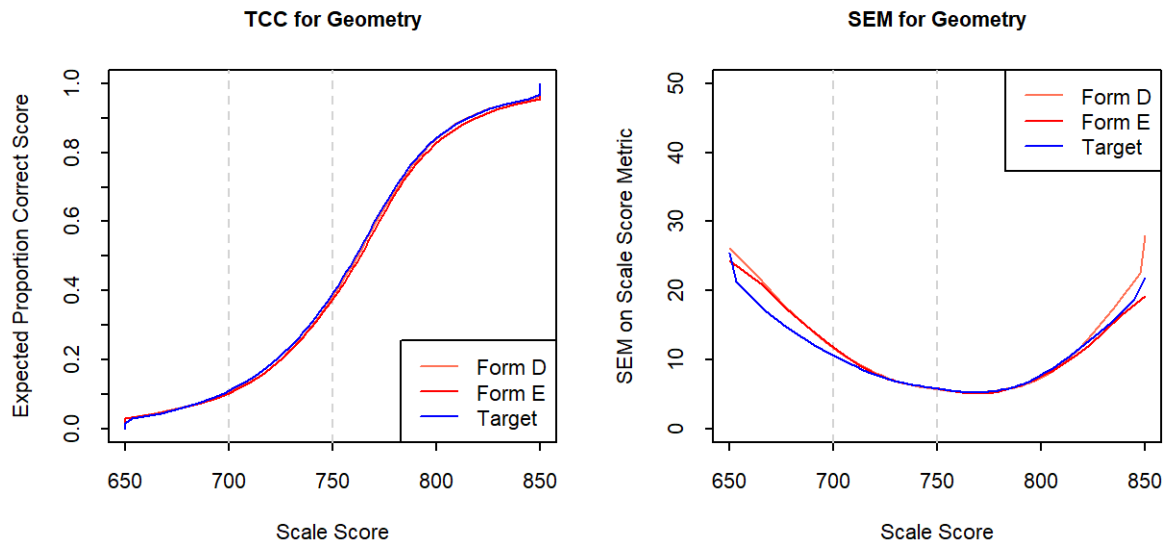
**Figure 3.3 Algebra I Form-Building Evaluation for 2018–2019 Administrations**



Notes:

- The target form is the Spring 2018 LEAP 2025 HS test form.
- Forms D and E are Spring 2019 LEAP 2025 HS test forms.

**Figure 3.4 Geometry Form-Building Evaluation for 2018–2019 Administrations**



Notes:

- The target form is the Spring 2018 LEAP 2025 HS test form.
- Forms D and E are Spring 2019 LEAP 2025 HS test forms.

### Selecting Anchors

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

## 3.4 Universal Design

Course-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 3.16 presents the elements of universal design (Thompson & Thurlow, 2002). The elements of universal design are relevant to both item development and form construction. This section describes how the elements of universal design were addressed in the construction of the 2018–2019 test forms in compliance with AERA, APA, & NCME (2014) Standard 3.1, which states the following:

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

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

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

**Table 3.16 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.

### 3.5 Accommodations and Designated Supports

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

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

Students with disabilities, students with 504 plans, and English Learners (ELs) may be provided test administration accommodations based on their accommodation plan. More information on accommodations can be found in Chapter 4. Accommodation coding instructions can be found in the *Test Coordinators Manual*.

Accommodated print forms were developed for the high school ELA and mathematics tests 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 C, "Accommodated Print Form Creation."

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

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

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

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

Accommodations were available for students who have an IEP, IAP, or EL plan. The following accommodations were available: braille test materials, calculation device and math tools for non-calculator sections of mathematics assessments, transferred answers, recorded answers, mathematics Spanish read aloud, translated mathematics test, and test read aloud (text-to-speech). For details on these accessibility features and accommodations, 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 E, "Forms Development Process for Spanish Translations Forms."

## 3.6 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. The specifications provide explicit guidance on how to write items to elicit the desired evidence. To address 2018–2019 LEAP 2025 high school assessment comparability goals with PARCC, 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 and guidelines on how to create the items that are specific to each assessment target and reporting category. In ELA and mathematics, item specifications describe the knowledge, skills, and processes being measured by each item type aligned to particular standards.

These item specifications were developed for each course and standard to delineate the expectations of knowledge and skill to be included on test questions. In addition, the ELA and mathematics item and stimulus specifications provide guidance on determining the appropriateness of task and stimulus materials (i.e., the materials that a student must refer to when working on a test question). The stimulus specifications also provide information on the characteristics of stimuli or activities that should be avoided because they are not important to the knowledge, skill, or process being measured. This underscores DRC's efforts to select items that are accessible to the widest range of students possible; in other words, 2018–2019 LEAP 2025 items were selected according to the elements of universal design.

## 3.7 Summary

In summary, the overall purpose of this chapter is to explicate the procedures used in the development of the 2018–2019 LEAP 2025 high school assessments. The efforts by LDOE and DRC in developing the LEAP 2025 high school assessments are in alignment with multiple best practices of the test industry but, in particular, support the following AERA, APA, & NCME (2014) standards:

*Standard 3.1* Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (63)

*Standard 3.2* Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (64)

*Standard 3.9* Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that

otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs. (67)

*Standard 4.0* Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population. (85)

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

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

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



## Chapter 4: Test Administration

---

Chapter 4 of the technical report describes the processes implemented and the information disseminated to help ensure standardized test administration procedures and, thus, uniform test administration conditions for students. According to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), “The usefulness and interpretability of test scores require that a test be administered and scored according to the test developer’s instructions” (111). This chapter examines how test administration procedures implemented for the 2018–2019 Louisiana Education Assessment Program (LEAP 2025) strengthen and support the intended score interpretations and reduce construct-irrelevant variance that could threaten the validity of score interpretations.

Chapter 4 demonstrates how the LEAP 2025 assessments adhere to AERA, APA, & NCME (2014) Standards 4.15, 6.1, 6.2, 6.3, 6.4, 6.6, and 6.7. Each standard will be explicated in 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 administration standards must also meet the State Board of Elementary and Secondary Education policies and the mandates of both state and federal legislation.

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

The school system test coordinators are responsible for the schools within their school systems. They disseminate information to each school, assist with test administration, and serve as liaisons between the LDOE and the schools in their system. 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 administration of the LEAP 2025 High School assessments, DRC produced the following test administration manuals (TAMs): *High School Test Administration Manual: LEAP 2025 and EOC, Fall 2018*; *High School Test Administration Manual, Spring 2019*; *High School Test Administration Manual, Summer 2019*. DRC also produced the following district test coordinators manuals (TCMs): *Test Coordinators Manual: LEAP 2025 and EOC, Fall 2018*; *Test Coordinators Manual: LEAP 2025 and EOC, Spring 2019*; *Test Coordinators Manual: LEAP 2025 and EOC, Summer 2019*. LDOE assessment administration and development staff review these manuals, provide feedback, and give final approval. Each TCM includes information about LEAP 2025 HS and EOC ELA, mathematics, U.S. history, and biology. It provides detailed instructions for school system and school test coordinators on distributing and collecting test materials and for returning them to DRC as outlined in its table of contents.

*Test Coordinators Manual* Table of Contents

1. Key Dates
2. Resources Available in eDIRECT
3. LEAP 2025 and EOC High School Alerts
4. Pre-Administration Oath of Security and Confidentiality Statement
5. Post-Administration Oath of Security and Confidentiality Statement
6. General Information
  - 6.1. eDIRECT and INSIGHT
7. LEAP 2025/EOC High School
  - 7.1. Testing Requirements
8. Test Security
  - 8.1. Key Definitions
  - 8.2. Violations of Test Security
  - 8.3. Testing Guidelines
  - 8.4. Testing Conditions
  - 8.5. Testing Schedule
  - 8.6. Extended Time for Testing
  - 8.7. Extended Breaks
  - 8.8. Makeup Testing
  - 8.9. LEAP 2025 High School and End-of-Course Testing Times
9. Roles and Responsibilities
  - 9.1. District Test Coordinator
  - 9.2. School Test Coordinator
  - 9.3. Chief Technology Officer
10. Managing Test Tickets
  - 10.1. Student Transfers
  - 10.2. Locked Test Tickets
  - 10.3. Technical Issues
  - 10.4. Invalidating Test Tickets
11. Resources for Online Testing
  - 11.1. High School Test Administration Manual
  - 11.2. eDIRECT User Guide
  - 11.3. LEAP 2025 Accommodations and Accessibility User Guide
  - 11.4. INSIGHT Technology User Guide
  - 11.5. Student Tutorials
  - 11.6. Online Tools Training (OTT)
12. Post-administration Rescoring Process for LEAP 2025/EOC Tests
13. Request for Rescoring
14. Void Notification

The TAMs provide detailed instructions for administering the LEAP 2025 assessments. The manuals include instructions for test security, test preparation, administration of tests, and post-test procedures. Information included in the TAMs is listed below.

*Test Administrators Manual* Table of Contents

1. Notes and Reminders
2. Pre-administration Oath and Security Confidentiality Statement
3. Post-administration Oath and Security Confidentiality Statement
4. Overview
5. Test Security
  - 5.1. Secure Test Materials
  - 5.2. Testing Irregularities and Security Breaches
  - 5.3. Testing Environment
  - 5.4. Violations of Test Security
  - 5.5. Voiding Student Tests
6. Test Administrator Responsibilities
  - 6.1. Software Tools and Features for Test Administrators
7. Test Administration Checklists
  - 7.1. Before Testing
  - 7.2. During Testing
  - 7.3. After Testing (Daily)
  - 7.4. After Testing (Last Day)
8. Test Materials
  - 8.1. Receipt of Test Materials
9. Testing Guidelines
  - 9.1. Testing Eligibility
  - 9.2. Testing Schedule
  - 9.3. LEAP 2025 Testing Time
  - 9.4. EOC Testing Time
  - 9.5. Extended Time for Testing
  - 9.6. Makeup Test Procedures
  - 9.7. Testing Conditions
  - 9.8. Accessibility Features
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 Students who are Deaf or Hearing Impaired
  - 10.6. English Learners (ELs)
11. Directions for Administering the LEAP 2025 Tests
12. LEAP 2025 Testing Times
13. General Information for LEAP 2025
  - 13.1. LEAP 2025 English I and English II
  - 13.2. LEAP 2025 Algebra I and Geometry
  - 13.3. LEAP 2025 Biology
  - 13.4. LEAP 2025 U.S. History
14. Directions for Administering End-of-Course Tests
15. End-of-Course Suggested Testing Times

16. General Instructions for EOC
  - 16.1. End-of-Course English III
  - 16.2. End-of-Course Biology
17. Post-Test Procedures
  - 17.1. Test Administrator Post-Administration Oath of Security and Confidentiality Statement
  - 17.2. Returning Test Materials to the School Test Coordinator
18. Index

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

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

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

The LEAP 2025 Test Administration Manuals provide instructions for activities conducted before, during, and after testing with sufficient detail and clarity to support reliable test administrations by qualified test administrators. To ensure uniform administration conditions throughout the state, instructions in the manuals describe the following: general rules of 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. 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 the 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 manuals also contain the schedule for timing each test session. The test timing schedule is presented in Table 4.1.

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

The LDOE staff administer reports on testing concerns that describe a wide range of improper activities that may occur during testing, including the following: copying and reviewing test questions with students; cueing students during testing, verbally or with written materials on the classroom walls; cueing students nonverbally, such as by tapping or nodding the head; allowing students to use 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 Individualized

Education Program (IEP); allowing accommodations for students who do not have an IEP; or defining terms on the test.

Each administration includes an administrative error retest, which provides an opportunity for students to retake a test that was voided during the regular test window because of improper activities that occurred during testing (e.g., the student was not given enough time to complete the test, the student was not provided proper accommodations during the testing time, the teacher or administrator provided information or answers that resulted in the test being voided).

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

The test administration manuals outline the steps that teachers should take to prepare classroom environments 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 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 that 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 braille test form, examiners are instructed to transcribe students' responses from the braille test form into the online testing system (INSIGHT) exactly as the responses appear in the original form.

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

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

#### 4.1 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 that the secure test materials are properly accounted for and organized appropriately for return shipment.

## 4.2 Security Checklists

As soon as printed test materials are received by a school system, the district test coordinator 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 test materials to be sent to schools. District test coordinators are required to return communication assistance scripts (CAS) and braille test materials to DRC. School systems are required to document nonstandard situations, including lost, damaged, destroyed, extra, or missing materials. Any material not accounted for is placed on a missing materials list, which is used by DRC and LDOE to follow up with all districts to ensure security of all materials.

## 4.3 Interpretive Guides

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

## 4.4 Test Security Measures

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

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

### Data Forensic Analyses

Due to the importance of the LEAP 2025 assessments, 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 in ensuring that assessment results are valid, data forensic analyses are conducted to help separate meaningful gains from spurious gains. It is important to note that although the results may be used to identify potential problems within a school, the identification of a problem is not an accusation of misconduct.

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

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

### 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. 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.

### Score-Change Analysis

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

### Web Monitoring

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

### Plagiarism Detection

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

## 4.5 Test Administration

The 2018–2019 assessments were administered to students within the state testing windows of November 28 to December 14, 2018; April 15 to May 17, 2019; and June 17–21, 2019. Each session of the LEAP 2025 assessments was required to be administered in one block of time.

### Time

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

**Table 4.1 LEAP 2025 Administration Schedule Timing by Session**

Course	Session	Minutes
English I	1	90
	2	90
	3	80
English II	1	90
	2	90
	3	80
Algebra I	1a	25
	1b	55
	2	80
	3	80
Geometry	1a	25
	1b	55
	2	80
	3	80

## 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 Manuals contain 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.

Tables 4.2 through 4.4 summarize the numbers of reportable students receiving accommodations or designated features by type for the 2018–2019 LEAP 2025 HS administrations. Accommodation assignment guidance is provided in the *LEAP 2025 Accommodations and Accessibility Features User Guide*. The analyses



are based on census data and include only those students who received accommodations or designated features and received a scale score on the ELA or Mathematics LEAP 2025 high school assessments. The percentage represents the percentage of the census population receiving that accommodation or designated feature.

**Table 4.2 Fall 2018 Number and Percentage of Students Receiving Accommodations by Accommodation/Designated Feature Type, as identified in eDIRECT**

Content	Accommodation/Designated Feature	Number	Percentage
English I	Text-to-Speech	≥930	14.0
	Accommodated Print	<50	NR
	Human Read Aloud	≥150	2.3
	Native Language Word-to-Word Dictionary	≥280	4.2
	Directions in Native Language	<50	NR
	Communication Assistance	<50	NR
	Transferred Answers	<50	NR
	Answers Recorded	<50	NR
	Extended Time	≥1830	27.4
	Individual/Small Group Administration	≥960	14.5
	Braille	<50	NR
English II	Text-to-Speech	≥1520	15.9
	Accommodated Print	<50	NR
	Human Read Aloud	≥320	3.4
	Native Language Word-to-Word Dictionary	≥470	4.9
	Directions in Native Language	≥100	1.1
	Communication Assistance	<50	NR
	Transferred Answers	<50	NR
	Answers Recorded	<50	NR
	Extended Time	≥2830	29.6
	Individual/Small Group Administration	≥1500	15.7
	Braille	<50	NR
Algebra I	Text-to-Speech	≥1040	18.4
	Accommodated Print	<50	NR
	Human Read Aloud	≥160	2.9
	Native Language Word-to-Word Dictionary	≥300	5.3
	Directions in Native Language	<50	NR
	Communication Assistance	<50	NR
	Transferred Answers	<50	NR
	Answers Recorded	<50	NR
	Calculator	≥910	16.2
	Extended Time	≥1680	29.7
	Individual/Small Group Administration	≥870	15.5
	Braille	<50	NR

**Table 4.3 Fall 2018 Number and Percentage of Students Receiving Accommodations by Accommodation/Designated Feature Type, as identified in eDIRECT (Continued)**

Content	Accommodation/Designated Feature	Number	Percentage
Geometry	Text-to-Speech	≥510	9.6
	Accommodated Print	<50	NR
	Human Read Aloud	<50	NR
	Native Language Word-to-Word Dictionary	≥170	3.3
	Directions in Native Language	<50	NR
	Communication Assistance	<50	NR
	Transferred Answers	<50	NR
	Answers Recorded	<50	NR
	Calculator	≥390	7.4
	Extended Time	≥900	16.8
	Individual/Small Group Administration	≥420	7.9
	Braille	<50	NR

**Table 4.4 Spring 2019 Number and Percentage of Students Receiving Accommodations by Accommodation/Designated Feature Type, as identified in eDIRECT**

Accommodation/Designated Feature Type: Spring 2019					
Content	Accommodation/Designated Feature	Form D		Form E	
		Number	Percentage	Number	Percentage
English I	Text-to-Speech	≥4,360	16.88	<50	NR
	Accommodated Print	<50	NR	<50	NR
	Human Read Aloud	≥360	1.40	<50	NR
	Native Language Word-to-Word Dictionary	≥610	2.38	≥320	1.51
	Directions in Native Language	≥100	0.39	≥50	0.26
	Communication Assistance	<50	NR	<50	NR
	Transferred Answers	<50	NR	<50	NR
	Answers Recorded	<50	NR	<50	NR
	Extended Time	≥6,620	25.61	≥2,240	10.47
	Individual/Small Group Administration	≥4,110	15.91	≥990	4.64
	Braille	<50	NR	<50	NR
English II	Text-to-Speech	≥3,280	14.36	<50	NR
	Accommodated Print	<50	NR	<50	NR
	Human Read Aloud	≥300	1.31	<50	NR
	Native Language Word-to-Word Dictionary	≥390	1.72	≥260	1.37
	Directions in Native Language	≥60	0.30	≥50	0.28
	Communication Assistance	<50	NR	<50	NR
	Transferred Answers	<50	NR	<50	NR
	Answers Recorded	<50	NR	<50	NR
	Extended Time	≥5,180	22.69	≥1,950	9.94
	Individual/Small Group Administration	≥3,360	14.73	≥910	4.69
	Braille	<50	NR	<50	NR
Algebra I	Text-to-Speech	≥5,090	18.98	<50	NR
	Accommodated Print	<50	NR	<50	NR
	Human Read Aloud	≥400	1.52	<50	NR
	Native Language Word-to-Word Dictionary	≥680	2.56	≥150	0.73
	Directions in Native Language	≥80	0.32	<50	NR
	Communication Assistance	<50	NR	<50	NR
	Transferred Answers	<50	NR	<50	NR
	Answers Recorded	≥60	0.23	<50	NR
	Calculator	≥4,360	16.26	≥570	2.67
	Extended Time	≥6,910	25.75	≥1,900	8.79
	Individual/Small Group Administration	≥4,330	16.15	≥830	3.84
	Braille	<50	NR	<50	NR

Accommodation/Designated Feature Type: Spring 2019 (continued)					
Content	Accommodation/Designated Feature	Form D		Form E	
		Number	Percentage	Number	Percentage
Geometry	Text-to-Speech	≥2,080	11.10	<50	NR
	Accommodated Print	<50	NR	<50	NR
	Human Read Aloud	≥190	1.04	<50	NR
	Native Language Word-to-Word Dictionary	≥310	1.66	≥100	0.65
	Directions in Native Language	<50	NR	<50	NR
	Communication Assistance	<50	NR	<50	NR
	Transferred Answers	<50	NR	<50	NR
	Answers Recorded	<50	NR	<50	NR
	Calculator	≥1,800	9.60	≥310	1.85
	Extended Time	≥3,200	17.04	≥1,110	6.67
	Individual/Small Group Administration	≥2,000	10.65	≥480	2.88
	Braille	<50	NR	<50	NR

**Table 4.5 Summer 2019 Number and Percentage of Students Receiving Accommodations by Accommodation/Designated Feature Type, as identified in eDIRECT**

Accommodation/Designated Feature Type: Summer 2019			
Content	Accommodation/Designated Feature	Number	Percentage
English I	Text-to-Speech	≥620	33.09
	Accommodated Print	<50	NR
	Human Read Aloud	≥130	6.84
	Native Language Word-to-Word Dictionary	≥130	6.84
	Directions in Native Language	<50	NR
	Communication Assistance	<50	NR
	Transferred Answers	<50	NR
	Answers Recorded	<50	NR
	Extended Time	≥910	48.03
	Individual/Small Group Administration	≥530	28.30
	Braille	<50	NR
English II	Text-to-Speech	≥510	30.21
	Accommodated Print	<50	NR
	Human Read Aloud	≥130	8.08
	Native Language Word-to-Word Dictionary	≥150	8.91
	Directions in Native Language	≥50	2.95
	Communication Assistance	<50	NR
	Transferred Answers	<50	NR
	Answers Recorded	<50	NR
	Extended Time	≥820	48.50
	Individual/Small Group Administration	≥450	26.55
	Braille	<50	NR

Accommodation/Designated Feature Type: Summer 2019			
Content	Accommodation/Designated Feature	Number	Percentage
Algebra I	Text-to-Speech	≥540	27.67
	Accommodated Print	<50	NR
	Human Read Aloud	≥150	8.13
	Native Language Word-to-Word Dictionary	≥70	3.79
	Directions in Native Language	<50	NR
	Communication Assistance	<50	NR
	Transferred Answers	<50	NR
	Answers Recorded	<50	NR
	Calculator	≥530	27.11
	Extended Time	≥790	40.41
	Individual/Small Group Administration	≥410	21.13
	Braille	<50	NR
Geometry	Text-to-Speech	≥50	20.36
	Accommodated Print	<50	NR
	Human Read Aloud	<50	NR
	Native Language Word-to-Word Dictionary	<50	NR
	Directions in Native Language	<50	NR
	Communication Assistance	<50	NR
	Transferred Answers	<50	NR
	Answers Recorded	<50	NR
	Calculator	≥50	18.91
	Extended Time	≥70	27.64
	Individual/Small Group Administration	≥50	18.91
	Braille	<50	NR

## 4.6 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 through appropriate accommodations, and maintaining the 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)

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

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

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

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

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

---

In this chapter, the scoring process used for the 2018–2019 LEAP 2025 high school ELA and mathematics assessments is described, with a particular focus on the handscoring of constructed-response items and the automated scoring of technology-enhanced items. At the end of this section, the results of the inter-rater reliability for the handscoring of the 2018–2019 LEAP 2025 constructed-response items are presented.

Chapter 5 demonstrates how the LEAP 2025 assessments adhere to the American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME, 2014) Standards 4.18, 4.20, 6.8, and 6.9. Each standard is presented in the pertinent section of this chapter. Standard 4.18 provides some general guidance for Chapter 5:

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

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

### 5.1 Constructed-Response Item Scoring Process

Constructed-response items were scored by human raters who were trained by DRC. Handscoring and Artificial Intelligence (AI) processing rules are detailed in Appendix F. Seven different ELA items across English I and English II (noted in Table 5.1) were scored by an AI engine, Pearson’s Intelligent Essay Assessor (IEA), using scoring models previously developed by Pearson. Second reads of 10% of these responses were completed by human scorers; handscoring supervisors also reviewed the responses that IEA was not able to score.

**Table 5.1 Constructed-Response Operational Scoring**

Administration	Course	Handscoring Only	AI Scoring	AI Vendor
Fall 2018	English I	N/A	902152, 902161	Pearson
	English II	N/A	902331, 906197	Pearson
	Algebra I	All CRs	N/A	
	Geometry	All CRs	N/A	
Spring 2019	English I	902161, 906152, 983215	902152, 914552	Pearson
	English II	902331, 902354, 983688	906197, 983642	Pearson
	Algebra I	All CRs	N/A	
	Geometry	All CRs	N/A	
Summer 2019	English I	N/A	902161, 906152	Pearson
	English II	N/A	902331, 906197	Pearson
	Algebra I	All CRs	N/A	
	Geometry	All CRs	N/A	

### Selection of Scoring Evaluators

Standard 4.20 states the following:

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

The following sections explain how scorers were selected and trained for the LEAP 2025 ELA and Mathematics handscoring process and how 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 the 2018–2019 LEAP 2025 high school ELA and mathematics test responses had at least a four-year college degree.

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



## 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 all readers 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 items each scorer may score. No demographic data is available to scorers at any time.

## Handscoring Training Process

Standard 6.9 specifies:

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

## Training Material Development

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

1. 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)
  
2. Math training materials developed by DRC in conjunction with and approved by LDOE. These materials were made for use with DRC-developed math items (which were newly operational in the spring of 2019) according to processes described in DRC's response to the LDOE's "REQUEST FOR PROPOSALS For LEAP 2025 Assessment Administration (RFP #: 815200-20150723001)".
  - Prompts
  - Rubrics
  - Anchor sets
  - Practice sets
  - Qualifying sets (for all DRC-developed items)

## Training and Qualifying Procedures

Handscoring involves training and qualifying team leaders and evaluators, monitoring scoring accuracy and production, and ensuring security of both the test materials and the scoring facilities. 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. Note that, like the PARCC “prototype” items for math, full sets of training and qualifying materials were also developed for all DRC-developed math items. The training and qualifying procedures DRC used for these items was the same process outlined below for PARCC-approved “prototype” math items.

### Prototype Items

A small number of items (two each for Algebra I and Geometry and one for ELA) 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.<sup>1</sup> 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.

---

<sup>1</sup> 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 evidence statements. For ELA items, this determination by PARCC and Pearson was based on grade and task type.

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

**Table 5.2 Mathematics Training Set Composition**

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

**Table 5.3 ELA Training Set Composition**

<b>Set Type</b>	<b>Prototype Item Training</b>	<b>Abbreviated Item Training</b>	<b>Annotated</b>
Anchor Set*	3 responses per score point	16 responses per item: <ul style="list-style-type: none"> <li>Anchor Sets for abbreviated RST and LAT item training included scores for the combined trait Reading Comprehension and Written Expression (RCWE).</li> <li>Anchor Sets for abbreviated NWT item training included scores for Written Expression (WE).</li> </ul>	Yes
Practice Set 1	5 responses representing the range of responses for <ul style="list-style-type: none"> <li>the Reading Comprehension and Written Expression (RCWE) trait (for LAT and RST items)</li> <li>the Written Expression trait (for NWT items)</li> </ul>	10 responses representing the range of responses for the trait appropriate to the task type	Yes
Practice Set 2	5 responses representing the range of responses for the Knowledge and Use of Language Conventions trait	10 responses representing the range of responses for the conventions trait	Yes
Practice Set 3	10 responses representing the range of responses for both traits appropriate to the task type		Yes
Practice Set 4	10 responses representing the range of responses for both traits appropriate to the task type		Yes
Qualifying Set 1	10 responses comparable to the anchor set responses (included both traits appropriate to the task type)		No
Qualifying Set 2	10 responses comparable to the anchor set responses (included both traits appropriate to the task type)		No
Qualifying Set 3	10 responses comparable to the anchor set responses (included both traits appropriate to the task type)		No
Direct Copy Set**	3-5 responses composed entirely or partially of text copied from passage or passages (included both traits appropriate to the task type)	3-5 responses composed entirely or partially of text copied from passage or passages (included both traits appropriate to the task type)	Yes

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

\*\*These 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.

Some items selected for use on the spring 2019 administration were previously only field-tested by PARCC. Consequently, the abbreviated training materials available for use with these items were abridged versions of typical abbreviated sets of materials. They consisted of:

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

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

### Qualifying Standards

DRC followed the same qualification standards that Pearson used for PARCC. Scorers demonstrated their ability to apply the scoring criteria by qualifying (i.e., scoring with acceptable agreement with true scores on qualifying sets). After each qualifying set was scored, the DRC scoring director responsible for training led the scorers in a discussion of the set. Any scorer who did not qualify by the end of the qualifying process for an item was not allowed to score live student responses.

**Table 5.4 Mathematics Qualifying Standards**

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

**Table 5.5 Mathematics Qualifying Standards (Composite Items)\***

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

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

**Table 5.6 ELA Qualifying Standards**

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

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

### 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)

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

### Reader Monitoring Procedures

Throughout the handscoring process, DRC project managers, scoring directors, and team leaders reviewed the statistics that were generated on a daily basis. DRC used one team leader for every 10 to 12 readers, which was the same ratio that Pearson used for PARCC. 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 and LDOE-approved validity responses identified by DRC scoring supervisors during live scoring for newly operational DRC-developed math items and PARCC field-test only items. Validity responses were inserted among the live student responses.

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

Approximately 10% of all live student responses were scored by a second reader to establish inter-rater reliability statistics for all constructed-response items. This procedure is called a "double-blind read" because the second reader does not know the first reader's score. DRC monitored inter-rater reliability based on the

responses that were scored by two readers. If a scorer fell below the expected rate of agreement, the team leader or scoring director retrained the scorer. If a scorer failed to improve after retraining and feedback, DRC removed the scorer from the project. In this situation, DRC removed all scores assigned by the scorer in question. The responses were then reassigned and rescored.

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

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

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

**Table 5.7 Agreement Rate Requirements for Validity and Inter-Rater Reliability**

Subject	Score Point Range	Perfect Agreement	Perfect Agreement + Adjacent
Mathematics	0–1	90%	100%
	0–2	80%	95%
	0–3	70%	95%
	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 responses for all operational items that came from the PARCC item pool. DRC pulled calibration responses for DRC-developed math items as well as additional responses for field-test only items from PARCC. DRC used these sets to perform calibration across the entire scorer population for an item if trends were detected (e.g., low agreement between certain score points if a certain type of response was missing from initial training). These calibrations were designed to help refocus scorers on how to properly use the scoring guidelines. They were selected to help illustrate particular points and familiarize scorers with the types of responses commonly seen during operational scoring. After readers scored a calibration set, the scoring director reviewed it from the front of the room, using rubric language and scoring concepts exemplified by the anchor responses to explain the reasoning behind each response’s score.

## Reports and Reader Feedback

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

## 5.2 Inter-Rater Reliability

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

A total of 79 operational items were scored by human readers across all LEAP 2025 high school ELA and mathematics assessments. The inter-rater reliability rates and the total numbers of reads are shown in Tables 5.8–5.10 for ELA items, Tables 5.11–5.13 for mathematics items, Tables 5.14–5.16 for Spanish mathematics items, and Table 5.17 for mathematics field test items.

As shown in Tables 5.8–5.10, raters demonstrated at least 99% perfect and adjacent agreement for all ELA handscored items. As shown in Tables 5.11–5.13, raters demonstrated at least 99% perfect and adjacent agreement for mathematics items. As shown in Tables 5.14–5.16, raters demonstrated 100% perfect and adjacent agreement for Spanish mathematics items. As shown in Table 5.17, raters demonstrated 100% perfect and adjacent agreement for mathematics field test items.



**Table 5.8 Inter-Rater Agreement, English Language Arts Items, Fall 2018**

Course	Task Type	Question	Trait	Total Reads	Read 2x	Inter-Rater Reliability %		
						EX	AD	EX + AD
English I	Research Simulation (AI)	902161	Reading Comprehension and Written Expression	≥7,410	≥1,840	84	16	100
			Knowledge and Use of Language Conventions	≥7,410	≥1,840	83	17	100
	Literary Analysis (AI)	902152	Written Expression	≥7,600	≥2,120	86	14	100
			Knowledge and Use of Language Conventions	≥7,600	≥2,120	84	16	100
English II	Research Simulation (AI)	902331	Reading Comprehension and Written Expression	≥10,670	≥2,780	83	17	100
			Knowledge and Use of Language Conventions	≥10,670	≥2,780	81	19	100
	Literary Analysis (AI)	906197	Written Expression	≥10,660	≥2,810	80	20	100
			Knowledge and Use of Language Conventions	≥10,660	≥2,810	81	19	100

Table 5.9 Inter-Rater Agreement, English Language Arts Items, Spring 2019

Course	Task Type	Question /Form	Trait	Total Reads	Read 2x	Inter-Rater Reliability %		
						EX	AD	EX + AD
English I	Research Simulation	902161 Form A (Seniors)	Reading Comprehension and Written Expression	≥80	≥10	100	0	100
			Knowledge and Use of Language Conventions	≥80	≥10	67	33	100
	Narrative Writing	906152 Form A (Seniors)	Written Expression	≥70	≥10	100	0	100
			Knowledge and Use of Language and Conventions	≥70	≥10	80	20	100
	Literary Analysis (AI)	902152 Form D	Reading Comprehension and Written Expression	≥29,410	≥6,940	86	14	100
			Knowledge and Use of Language and Conventions	≥29,410	≥6,940	85	15	100
	Narrative Writing	983215 Form E	Written Expression	≥23,690	≥4,870	79	20	99
			Knowledge and Use of Language and Conventions	≥23,690	≥4,870	77	23	100
	Research Simulation (AI)	914552 Forms D and E	Reading Comprehension and Written Expression	≥52,880	≥11,490	76	24	100
			Knowledge and Use of Language and Conventions	≥52,880	≥11,490	75	24	99* (na = 0)

Table 5.9 Inter-Rater Agreement, English Language Arts Items, Spring 2019 (continued)

Course	Task Type	Question /Form	Trait	Total Reads	Read 2x	Inter-Rater Reliability %		
						EX	AD	EX + AD
English II	Research Simulation	902331 Form A (Seniors)	Reading Comprehension and Written Expression	≥940	≥160	94	6	100
			Knowledge and Use of Language and Conventions	≥940	≥160	95	5	100
	Narrative Writing	902354 Form A (Seniors)	Written Expression	≥950	≥220	100	0	100
			Knowledge and Use of Language and Conventions	≥950	≥220	97	3	100
	Literary Analysis (AI)	906197 Form D	Reading Comprehension and Written Expression	≥25,430	≥5,770	79	21	100
			Knowledge and Use of Language and Conventions	≥25,430	≥5,770	80	20	100
	Narrative Writing (AI)	983642 Form E	Written Expression	≥21,670	≥4,650	84	16	100
			Knowledge and Use of Language and Conventions	≥21,670	≥4,650	82	18	100
	Research Simulation	983688 Forms D and E	Reading Comprehension and Written Expression	≥46,670	≥9,270	78	22	100
			Knowledge and Use of Language and Conventions	≥46,670	≥9,270	77	22	99* (na = 0)

\*Total Ex + AD does not add up to 100% due to rounding

Table 5.10 Inter-Rater Agreement, English Language Arts Items, Summer 2019

Course	Task Type	Question	Trait	Total Reads	Read 2x	Inter-Rater Reliability %		
						EX	AD	EX + AD
English I	Research Simulation (AI)	902161	Reading Comprehension and Written Expression	≥2,180	≥550	87	13	100
			Knowledge and Use of Language and Conventions	≥2,180	≥550	88	12	100
	Narrative Writing (AI)	906152	Written Expression	≥2,160	≥760	93	7	100
			Knowledge and Use of Language and Conventions	≥2,160	≥760	92	8	100
English II	Research Simulation (AI)	902331	Reading Comprehension and Written Expression	≥1,870	≥560	91	9	100
			Knowledge and Use of Language and Conventions	≥1,870	≥560	90	10	100
	Literary Analysis (AI)	906197	Reading Comprehension and Written Expression	≥1,880	≥570	88	12	100
			Knowledge and Use of Language and Conventions	≥1,880	≥570	88	12	100

Table 5.11 Inter-Rater Agreement, Mathematics Items, Fall 2018

Course	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
Algebra I	13	Part B	≥6,320	≥1,140	92	8	100
	15	Part A	≥6,180	≥1,330	100	0	100
		Part B	≥6,180	≥1,330	99	1	100
	28	Part A	≥6,020	≥1,470	99	0	99* (na = 0)
		Part B	≥6,020	≥1,470	99	1	100
		Part C	≥6,020	≥1,470	99	1	100
	29	Part A	≥5,900	≥1,620	99	1	100
		Part B	≥5,900	≥1,620	97	3	100
	43	N/A	≥6,010	≥1,550	96	4	100
	44	Part A	≥6,090	≥1,570	100	0	100
		Part B	≥6,090	≥1,570	98	2	100
	45	Part A	≥5,870	≥1,530	94	6	100
		Part B	≥5,870	≥1,530	97	3	100
Geometry	13	N/A	≥5,820	≥1,420	96	4	100
	15	N/A	≥5,600	≥1,390	96	4	100
	27	N/A	≥5,710	≥1,530	96	4	100
	28	Part A	≥5,710	≥1,310	98	2	100
		Part B	≥5,710	≥1,310	98	2	100
		Part C	≥5,710	≥1,310	98	1	99
	43	N/A	≥5,740	≥1,450	97	3	100
	44	Part A	≥5,690	≥1,450	95	4	99
		Part B	≥5,690	≥1,450	97	3	100

\*Total Ex + AD does not add up to 100% due to rounding

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

Table 5.12 Inter-Rater Agreement, Mathematics Items, Spring 2019

Course	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
Algebra I Form A (Seniors)	13	N/A	≥250	<10	NR	NR	NR
	15	Part A	≥260	≥10	100	0	100
		Part B	≥260	≥10	100	0	100
	28	Part C	≥250	<10	NR	NR	N
	29	Part A	≥230	<10	NR	NR	NR
		Part B	≥230	<10	NR	NR	NR
	43	N/A	≥270	≥80	98	2	100
	44	N/A	≥220	≥20	100	0	100
	45	Part A	≥220	<10	NR	NR	NR
Part B		≥220	<10	NR	NR	NR	
Algebra Form D	13	Part B	≥29,960	≥5,390	92	8	100
	15	N/A	≥28,910	≥6,480	95	5	100
	29	Part B	≥29,770	≥5,330	95	5	100
	43	N/A	≥29,750	≥6,050	96	4	100
	44	N/A	≥28,920	≥8,620	97	3	100
	45	N/A	≥28,410	≥7,090	95	4	99* (na = 0)
Algebra Forms D and E	28	Part A	≥52,800	≥11,120	98	2	100
		Part B	≥52,800	≥11,120	95	5	100
		Part C	≥52,800	≥11,120	93	7	100
Algebra Form E	13	N/A	≥23,670	≥4,850	88	12	100
	15	N/A	≥22,660	≥5,500	91	9	100
	29	Part A	≥22,960	≥5,150	98	2	100
		Part B	≥22,960	≥5,150	97	3	100
	43	N/A	≥23,070	≥4,170	95	5	100
	44	N/A	≥23,110	≥6,320	95	5	100
	45	N/A	≥22,980	≥5,360	97	3	100

Table 5.12 Inter-Rater Agreement, Mathematics Items, Spring 2019 (continued)

Course	Question	Part(s)**	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
Geometry Form A (Seniors)	13	N/A	≥400	≥70	100	0	100
	15	N/A	≥400	≥110	98	2	100
	27	N/A	≥390	≥100	100	0	100
	28	Part A	≥430	≥90	98	2	100
		Part B	≥430	≥90	100	0	100
		Part C	≥430	≥90	100	0	100
	43	Part A	≥410	≥90	100	0	100
		Part B	≥410	≥90	100	0	100
44	Part C	≥430	≥70	100	0	100	
Geometry Form D	13	N/A	≥20,160	≥3,890	92	8	100
	15	N/A	≥19,870	≥4,940	99	1	100
	25	N/A	≥20,250	≥4,790	95	5	100
	28	Part A	≥20,530	≥4,500	94	6	100
		Part B	≥20,530	≥4,500	96	4	100
	43	N/A	≥20,170	≥5,100	97	3	100
	44	N/A	≥19,750	≥4,930	98	2	100
45	N/A	≥20,130	≥4,570	95	5	100	
Geometry Form E	13	N/A	≥17,970	≥3,590	92	8	100
	15	N/A	≥17,570	≥4,280	99	1	100
	25	N/A	≥17,470	≥4,370	91	9	100
	28	Part A	≥17,670	≥2,860	97	3	100
		Part B	≥17,670	≥2,860	98	2	100
		Part C	≥17,670	≥2,860	98	1	99* (na = 0)
	43	Part B	≥18,590	≥3,390	97	3	100
	44	N/A	≥17,490	≥4,370	99	1	100
45	N/A	≥17,720	≥3,930	94	6	100	

\*Total Ex + AD does not add up to 100% due to rounding

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

Table 5.13 Inter-Rater Agreement, Mathematics Items, Summer 2019

Course	Question	Part(s)*	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
Algebra I	13	Part B	≥2,190	≥400	97	3	100
	15	Part A	≥2,120	≥460	100	0	100
		Part B	≥2,120	≥460	100	0	100
	28	Part A	≥2,110	≥530	100	0	100
		Part B	≥2,110	≥530	100	0	100
		Part C	≥2,110	≥530	100	0	100
	29	N/A	≥2,080	≥620	100	0	100
	43	N/A	≥2,090	≥530	99	1	100
	44	Part A	≥2,100	≥500	100	0	100
		Part B	≥2,100	≥500	100	0	100
	45	Part A	≥2,080	≥570	97	3	100
Part B		≥2,080	≥570	100	0	100	
Geometry	13	N/A	≥300	≥80	95	5	100
	15	N/A	≥290	≥70	100	0	100
	27	N/A	≥290	≥90	96	4	100
	28	Part A	≥290	≥70	100	0	100
		Part B	≥290	≥70	100	0	100
		Part C	≥290	≥70	100	0	100
	43	N/A	≥280	≥80	100	0	100
44	Part C	≥310	≥50	100	0	100	

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



Table 5.14 Inter-Rater Agreement, Spanish Mathematics Items, Fall 2018

Course	Question	Part(s)*	Total Reads	Read 2x**	Inter-Rater Reliability %		
					EX	AD	EX + AD
Algebra I	13	Part B	≥20	N/A	N/A	N/A	N/A
	15	Part A	≥20	N/A	N/A	N/A	N/A
		Part B	≥20	N/A	N/A	N/A	N/A
	28	Part A	≥20	N/A	N/A	N/A	N/A
		Part B	≥20	N/A	N/A	N/A	N/A
		Part C	≥20	N/A	N/A	N/A	N/A
	29	Part A	≥20	N/A	N/A	N/A	N/A
		Part B	≥20	N/A	N/A	N/A	N/A
	43	N/A	≥20	N/A	N/A	N/A	N/A
	44	Part A	≥20	N/A	N/A	N/A	N/A
		Part B	≥20	N/A	N/A	N/A	N/A
	45	Part A	≥20	N/A	N/A	N/A	N/A
		Part B	≥20	N/A	N/A	N/A	N/A
	Geometry	13	N/A	<10	N/A	N/A	N/A
15		N/A	<10	N/A	N/A	N/A	N/A
27		N/A	<10	N/A	N/A	N/A	N/A
28		Part A	<10	N/A	N/A	N/A	N/A
		Part B	<10	N/A	N/A	N/A	N/A
		Part C	<10	N/A	N/A	N/A	N/A
43		N/A	<10	N/A	N/A	N/A	N/A
44		Part A	<10	N/A	N/A	N/A	N/A
	Part B	<10	N/A	N/A	N/A	N/A	

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

\*\* Due to low numbers of Spanish Mathematics test takers in Fall 2018, all Spanish Mathematics responses were scored directly by expert scorers/supervisors and not routed for second reads. As a result, no inter-rater reliability percentages were generated.

Table 5.15 Inter-Rater Agreement, Spanish Mathematics Items, Spring 2019

Course	Question	Part(s)*	Total Reads	Read 2x**	Inter-Rater Reliability %		
					EX	AD	EX + AD
Algebra I Form A (Seniors)	13	N/A	<10	N/A	N/A	N/A	N/A
	15	Part A	<10	N/A	N/A	N/A	N/A
		Part B	<10	N/A	N/A	N/A	N/A
	28	Part C	<10	N/A	N/A	N/A	N/A
	29	Part A	<10	N/A	N/A	N/A	N/A
		Part B	<10	N/A	N/A	N/A	N/A
	43	N/A	<10	N/A	N/A	N/A	N/A
	44	N/A	<10	N/A	N/A	N/A	N/A
	45	Part A	<10	<10	NR	NR	NR
Part B		<10	<10	NR	NR	NR	
Algebra Form D	13	Part B	≥110	<10	NR	NR	NR
	15	N/A	≥100	N/A	N/A	N/A	N/A
	28	Part A	≥110	≥30	100	0	100
		Part B	≥110	≥30	94	6	100
		Part C	≥110	≥30	94	6	100
	29	Part B	≥100	N/A	N/A	N/A	N/A
	43	N/A	≥100	≥10	100	0	100
	44	N/A	≥110	≥40	100	0	100
	45	N/A	≥90	N/A	N/A	N/A	N/A

Course	Question	Part(s)*	Total Reads	Read 2x**	Inter-Rater Reliability %		
					EX	AD	EX + AD
Geometry Form A (Seniors)	13	N/A	<10	N/A	N/A	N/A	N/A
	15	N/A	<10	N/A	N/A	N/A	N/A
	27	N/A	<10	N/A	N/A	N/A	N/A
	28	Part A	<10	N/A	N/A	N/A	N/A
		Part B	<10	N/A	N/A	N/A	N/A
		Part C	<10	N/A	N/A	N/A	N/A
	43	Part A	<10	N/A	N/A	N/A	N/A
Part B		<10	N/A	N/A	N/A	N/A	
44	Part C	<10	N/A	N/A	N/A	N/A	
Geometry Form D	13	N/A	≥50	≥10	100	0	100
	15	N/A	≥30	N/A	N/A	N/A	N/A
	25	N/A	≥40	N/A	N/A	N/A	N/A
	28	Part A	≥40	N/A	N/A	N/A	N/A
		Part B	≥40	N/A	N/A	N/A	N/A
	43	N/A	≥40	≥10	100	0	100
	44	N/A	≥40	≥10	100	0	100
45	N/A	≥40	≥10	100	0	100	

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

\*\* Second Reads may be less than 10% of Total Reads or N/A for some items, because the smaller quantities of responses allowed scoring to be done directly by expert scorers/supervisors or via paired scoring between a supervisor and scorer. As a result, fewer were routed through the 10% read-behind process.

Table 5.16 Inter-Rater Agreement, Spanish Mathematics Items, Summer 2019

Course	Question	Part(s)*	Total Reads	Read 2x**	Inter-Rater Reliability %		
					EX	AD	EX + AD
Algebra I	13	Part B	<10	N/A	N/A	N/A	N/A
	15	Part A	<10	N/A	N/A	N/A	N/A
		Part B	<10	N/A	N/A	N/A	N/A
	28	Part A	<10	N/A	N/A	N/A	N/A
		Part B	<10	N/A	N/A	N/A	N/A
		Part C	<10	N/A	N/A	N/A	N/A
	29	N/A	<10	N/A	N/A	N/A	N/A
	43	N/A	<10	N/A	N/A	N/A	N/A
	44	Part A	<10	N/A	N/A	N/A	N/A
		Part B	<10	N/A	N/A	N/A	N/A
	45	Part A	<10	N/A	N/A	N/A	N/A
Part B		<10	N/A	N/A	N/A	N/A	
Geometry***	13	N/A	N/A	N/A	N/A	N/A	N/A
	15	N/A	N/A	N/A	N/A	N/A	N/A
	27	N/A	N/A	N/A	N/A	N/A	N/A
	28	Part A	N/A	N/A	N/A	N/A	N/A
		Part B	N/A	N/A	N/A	N/A	N/A
		Part C	N/A	N/A	N/A	N/A	N/A
	43	N/A	N/A	N/A	N/A	N/A	N/A
44	Part C	N/A	N/A	N/A	N/A	N/A	

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

\*\* Due to low numbers of Spanish Mathematics test takers in Summer 2019, all Spanish Algebra I responses were scored directly by expert scorers/supervisors and not routed for second reads. As a result, no inter-rater reliability percentages were generated.

\*\*\* There were no test takers for Spanish Geometry.

**Table 5.17 Inter-Rater Agreement, Mathematics Items, Field Test 2019**

Course	Question	Part(s)*	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	AD	EX + AD
Algebra I	986936	Part A	≥1,700	≥360	99	1	100
		Part B	≥1,700	≥360	98	2	100
	986937	Part A	≥1,730	≥330	97	3	100
		Part C	≥1,730	≥330	93	7	100
		Part D	≥1,730	≥330	95	5	100
	986938	N/A	≥1,720	≥380	96	4	100
	987007	Part B	≥1,760	≥330	96	4	100
		Part C	≥1,760	≥330	91	9	100
Geometry	986939	N/A	≥1,680	≥370	96	4	100
	986940	N/A	≥1,710	≥360	91	9	100
	987006	N/A	≥1,700	≥370	87	13	100
	987008	Part C	≥1,760	≥330	96	4	100

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

### 5.3 Technology-Enhanced Item Scoring Process

All technology-enhanced items, as well as EBSR, MPSR, and SA items, were processed through DRC's autoscoring engine and scored according to the assigned scoring rules established during content development by PARCC or DRC 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 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 rubric resided in one place along with the item image and other metadata. This scoring information included details that varied by item type. For example, for a drag-and-drop item, the information included which object is 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 report was generated that showed all student responses, their frequencies, and their received scores.

#### 5.4 Multiple-Choice and Multiple-Select Item Scoring Process

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

#### 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 5.2 demonstrate that the items were scored reliably. These efforts by DRC address multiple best practices of the testing industry but are particularly related to AERA, APA, & NCME (2014) Standards 4.18, 4.20, 6.8, and 6.9:

*Standard 4.18* Procedures for scoring and, if relevant, scoring criteria, should be presented by the test developer with sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical for extended-response items such as performance tasks, portfolios, and essays. (91)

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

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

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

## Chapter 6: Operational Data Analyses

---

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

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

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

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

For all 2018–2019 LEAP 2025 high school analyses, the Louisiana student population was used. In Section 6.3, the characteristics of calibration samples, such as subgroups, are discussed. Chapter 3 presents the test specifications. Information regarding reported data is discussed in detail in Chapter 7.

### 6.1 Classical Item Statistics

In this section, summary test statistics for each form and subject area of the LEAP 2025 high school tests are presented. These statistics are followed by item-level statistics for each subject area of the LEAP 2025 test. These statistics were produced using census data with first-time test takers. Students whose results were included in the item-level statistics summary needed to meet at least the following psychometric analysis criteria (note that the criteria used to filter data for item statistics analyses are slightly different than those used to produce students' performance statistics in this report):

- Student has total raw score in the data
- Student did not take administration error form
- Student did not take braille form
- Student did not take Spanish form
- Student's test score was not voided
- Student took the assessment for the first time (initial testers)
- Student finished all sessions
- Student's constructed-response items were scored

### Test-Level Statistics

Table 6.1 presents the number of items, score points, mean and standard deviation of the raw scores, and the average form difficulty for each subject for each administration. Form difficulty for a student was calculated by dividing the student's raw score by the total score points of the test.

As can be seen in the table, average form difficulty was similar in the fall and spring administrations. Average form difficulty in the summer administration was lower for all tests than in the fall and spring administrations, likely due to the fact that the summer form was a retest form. The average form difficulty for ELA ranged from 0.40 to 0.47 with the fall and spring administrations. The difficulty of the spring administration forms was 0.40 (Form D) and 0.45 (Form E) for English I and 0.40 (Form D) and 0.47 (Form E) for English II. The average form difficulty for mathematics ranged from 0.31 to 0.35 for the fall and spring administrations. The average form difficulty of the spring administration of mathematics was 0.32 (Form D) and 0.35 (Form E) for Algebra I and 0.31 (Form D) and 0.32 (Form E) for Geometry. In general, the 2019 LEAP 2025 High School tests were relatively difficult, and the mathematics tests were more difficult than the ELA tests.

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

Administration	Course	Form	Total Items*	Total Points	Mean Raw Score (Std. Dev.)	Average Form Difficulty (Std. Dev.)
Fall 2018	English I	B	34	94	41.64 (17.44)	0.45 (0.13)
	English II	B	34	94	40.46 (16.78)	0.44 (0.13)
	Algebra I	B	39	68	20.88 (9.73)	0.31 (0.18)
	Geometry	B	38	68	22.60 (13.22)	0.34 (0.18)
Spring 2019	English I	D	34	94	36.85 (17.69)	0.40 (0.13)
		E	33	90	40.43 (16.86)	0.45 (0.12)
	English II	D	34	94	37.18 (17.41)	0.40 (0.11)
		E	33	90	41.88 (16.45)	0.47 (0.11)
	Algebra I	D	39	68	21.36 (12.07)	0.32 (0.16)
		E	39	68	22.98 (12.40)	0.35 (0.16)
	Geometry	D	39	68	20.37 (12.52)	0.31 (0.16)
		E	39	68	21.43 (12.39)	0.32 (0.15)
Summer 2019	English I	A	33	90	27.47 (15.97)	0.32 (0.12)
	English II	B	34	94	26.87 (16.37)	0.30 (0.12)
	Algebra I	BR	39	68	17.62 (9.36)	0.26 (0.17)
	Geometry	BR	38	68	20.35 (18.93)	0.31 (0.14)

\*For English I and English II, each writing prompt component is counted as one item. The WE writing component is weighted in total points.

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 subject for each administration.

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



**Table 6.2 LEAP 2025 High School  $p$ -Values and Item-Total Correlation ( $R_{it}$ ) Descriptive Statistics**

Admin.	Course	Form	Total Items*	Item $p$ -Value				Average Item-Total Correlation			
				Mean	Std. Dev.	Min.	Max	Mean	Std. Dev.	Min.	Max
Fall 2018	English I	B	34	0.48	0.13	0.20	0.77	0.45	0.17	0.20	0.80
	English II	B	34	0.45	0.15	0.21	0.80	0.45	0.19	0.09	0.82
	Algebra I	B	39	0.35	0.15	0.05	0.72	0.33	0.17	0.05	0.62
	Geometry	B	38	0.39	0.18	0.08	0.84	0.48	0.13	0.16	0.73
Spring 2019	English I	D	34	0.44	0.13	0.25	0.79	0.47	0.17	0.22	0.81
		E	33	0.49	0.11	0.32	0.83	0.45	0.15	0.20	0.77
	English II	D	34	0.42	0.12	0.21	0.68	0.46	0.18	0.15	0.82
		E	33	0.49	0.12	0.30	0.72	0.45	0.15	0.24	0.78
	Algebra I	D	39	0.36	0.15	0.12	0.82	0.42	0.15	0.13	0.71
		E	39	0.39	0.16	0.10	0.89	0.42	0.15	0.17	0.70
	Geometry	D	39	0.35	0.14	0.06	0.69	0.46	0.15	0.15	0.76
		E	39	0.37	0.15	0.06	0.64	0.45	0.14	0.20	0.75
Summer 2019	English I	A	33	0.35	0.11	0.08	0.69	0.45	0.17	0.11	0.78
	English II	B	34	0.33	0.13	0.15	0.68	0.47	0.19	0.14	0.86
	Algebra I	BR	39	0.30	0.14	0.03	0.63	0.33	0.20	-0.13	0.69
	Geometry	BR	38	0.35	0.14	0.14	0.78	0.68	0.16	0.20	0.93

\*For English I and English II, each writing prompt component is counted as one item. The WE writing component is weighted in total points.

### Item-Level Statistics

Tables in Appendix G present the item statistics for each operational item included in the regular forms, organized by content area and administration. The tables include item number,  $p$ -value, item-total correlation ( $R_{it}$ ), omit rates, total N, adjusted N (adjusted N excludes omitted responses, responses that were not scored, or responses that received a non-score code), the percentage at each score point for polytomous items, and the percentage that chose each option for multiple-choice (MC) items. The summer administration population is not state representative, and the number of students was very small, so the interpretation of statistics in the summer administration should be done with caution.

#### *p*-Value

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

In terms of  $p$ -values, test scores tend to be more precise when their average  $p$ -values are between the mid-0.50s and the low 0.70s. However, it is important to select items based on content rather than on purely statistical criteria when building a criterion-referenced test. As shown in Table 6.2, the average  $p$ -values of the fall and spring administrations ranged from 0.35 to 0.49. In the spring administration, Form D tended to have lower  $p$ -values than Form E as Form D was administered when certain accommodations, such as text-to-speech, were required. The range of average  $p$ -values was lower in the summer administration, ranging from

0.30 to 0.35. The average  $p$ -values of the English I and English II forms were higher than the average  $p$ -values of the Algebra I forms.

It is important that one examines the range of  $p$ -values, not just the average  $p$ -value, to determine whether a test measures well. It is desirable for a test to measure well throughout the range of skills present in the test form. That is, it is important that the items measure the performance of both low-scoring and high-scoring students, not just students in the center of the distribution. Having a range of  $p$ -values also helps to prevent floor and/or ceiling effects so that the test does not have large numbers of students at the minimum or maximum possible scores. The fall and spring English forms have items with  $p$ -values ranging from 0.20 to 0.83 (see Appendix G) and the summer English forms have items with  $p$ -values ranging from 0.08 to 0.69. The  $p$ -values on the mathematics forms range from 0.05 to 0.89 (see Appendix G) for the fall and spring administrations and from 0.03 to 0.78 for the summer administration. Such a broad range of  $p$ -values, which indicates that the items measure well throughout the range of skill levels at a given grade, supports the accuracy of the LEAP 2025 high school test scores.

#### *Item-Total Correlations*

An item-total correlation is the correlation between an item score and the total test score, where the item score is not included in the total score. It indicates how well an item differentiates 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 on the LEAP 2025 High School forms had item-total correlations above this threshold. Any item with an item-total correlation below the 0.20 threshold was further analyzed to ensure that the item was correctly keyed.

#### *Omit Rates*

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

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

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

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

The results presented in the Tables in Appendix G show that the percentage of students who omitted most of the items on the fall and spring LEAP 2025 High School tests was less than 5, suggesting that most students were able to complete the test in the prescribed amount of time. There were a small number of Algebra I and Geometry items that exceeded the omit rate of 5%. This is likely due to the difficulty of the items, given that these items also have low  $p$ -values. Lu & Sireci (2007) report that the Education Testing Service has used an

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

These item level statistics are reviewed at the beginning of the operational analysis process to ensure that items are not flawed, and a careful review is given to determine that the answer key is correct.

An MC item is reviewed during the key check process if

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

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

- $p$ -value is less than 0.30 or more than .80,
- percentage of students who reached any possible score point is less than 3%,
- item-total correlation is less than 0.30, or
- percentage of students omitting or not reaching the item is 15% or greater.

### Item Response Theory (IRT)

Item parameters for items included in the LEAP 2025 High School tests were estimated using a marginal maximum-likelihood (MML) procedure and the 2-parameter logistic (2PL) model for MC items and the generalized partial credit (GPC) model (Muraki, 1992) for non-MC items. Under the 2PL model, the probability that a student with a trait or scale score of  $\theta$  will respond correctly to MC item  $j$  is

$$P_j(\theta) = \frac{1}{1 + \exp[-Da_j(\theta - b_j)]}$$

where  $D$  is 1.7,  $a_j$  is the item discrimination, and  $b_j$  is the item difficulty. Under the GPC model, the probability that a student with a trait or scale score of  $\theta$  will respond in category  $x$  to partial-credit item  $j$  is

$$P_{jx}(\theta) = \frac{\exp[\sum_{k=0}^x Da_j(\theta - b_j + d_{jk})]}{\sum_{h=0}^{m_i} \exp[\sum_{k=0}^h Da_j(\theta - b_j + d_{jk})]}$$

where  $d_{jk}$  is the relative difficulty of score category  $x$  of item  $j$ , and  $m_i$  is the maximum item score for item  $j$ .

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

## 6.2 Calibration Sample

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

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

Sample data was used for calibration with all subjects for the spring 2019 administration. Intact forms with established scoring tables were used for the fall and summer administrations, making calibration unnecessary. Since full census data was not available, the intention was to use data files that had at least 10,000 test takers for a form, or 20,000 test takers for a subject, scored to completion. The sample was evaluated using spring 2018 demographic information and student performance information to confirm that the sample was representative of the state's student population. Tables 6.3 and 6.4 show the representativeness of the calibration samples compared to the census data from the spring 2019 administration, including data for initial testers only. The census data in these tables included initial testers who received a scale score. With LEAP 2025 High School tests, psychometric analyses such as item calibration and item statistics are computed excluding re-testers since students taking the exam again are not representative of the general population. The calibration samples were representative of census data. The spring 2019 resampling strategy was successful in sampling a group that was representative of Louisiana students.

Table 6.3 Summary of Calibration and Census Data: Spring Administration Form D

Course	Group or Subgroup	Calibration Sample		Census Data Initial Testers Only		Census % - Calib %
		N	%	N	%	
English I	All Students	≥13,800	100.00%	≥25,020	100.00%	0.00%
	Gender					
	Male	≥7,010	50.80%	≥12,740	50.92%	0.12%
	Female	≥6,790	49.20%	≥12,280	49.08%	-0.12%
	Race/Ethnicity					
	Hispanic/Latino	≥900	6.54%	≥1,770	7.10%	0.56%
	American Indian or Alaska Native	≥100	0.75%	≥170	0.69%	-0.05%
	Asian	≥230	1.73%	≥450	1.82%	0.09%
	Black or African American	≥5,850	42.44%	≥10,980	43.90%	1.45%
	Native Hawaiian or Other Pacific	<10	NR	≥10	0.06%	NR
	White	≥6,430	46.60%	≥11,140	44.53%	-2.07%
	Two or More Races	≥260	1.91%	≥470	1.91%	0.00%
	Economic Status					
	Economically Disadvantaged	≥8,380	60.72%	≥15,310	61.20%	0.48%
	Not Economically Disadvantaged	≥4,220	30.58%	≥7,270	29.09%	-1.48%
English II	All Students	≥10,420	100.00%	≥21,840	100.00%	0.00%
	Gender					
	Male	≥5,170	49.66%	≥10,920	50.00%	0.34%
	Female	≥5,240	50.34%	≥10,920	50.00%	-0.34%
	Race/Ethnicity					
	Hispanic/Latino	≥520	5.07%	≥1,240	5.69%	0.62%
	American Indian or Alaska Native	≥80	0.80%	≥130	0.61%	-0.18%
	Asian	≥200	2.01%	≥400	1.84%	-0.17%
	Black or African American	≥4,090	39.28%	≥9,400	43.05%	3.77%
	Native Hawaiian or Other Pacific	<10	NR	≥20	0.09%	NR
	White	≥5,340	51.26%	≥10,260	46.99%	-4.26%
	Two or More Races	≥160	1.54%	≥370	1.73%	0.20%
	Economic Status					
	Economically Disadvantaged	≥5,950	57.09%	≥13,070	59.87%	2.78%
	Not Economically Disadvantaged	≥3,670	35.27%	≥7,400	33.88%	-1.39%

Table continues

Calibration and Census Data: Spring Form D Administration (continued)						
		Calibration Sample		Census Data Initial Testers Only		
Course	Group or Subgroup	N	%	N	%	Census % - Calib %
Algebra I	All Students	≥7,360	100.00%	≥25,630	100.00%	0.00%
	Gender					
	Male	≥3,750	50.94%	≥13,020	50.81%	-0.14%
	Female	≥3,610	49.06%	≥12,610	49.19%	0.14%
	Race Ethnicity					
	Hispanic/Latino	≥450	6.13%	≥1,890	7.39%	1.26%
	American Indian or Alaska Native	≥40	0.58%	≥180	0.70%	0.12%
	Asian	≥130	1.86%	≥410	1.63%	-0.23%
	Black or African American	≥3,330	45.23%	≥11,110	43.34%	-1.89%
	Native Hawaiian or Other Pacific	<10	NR	≥10	0.07%	NR
	White	≥3,250	44.14%	≥11,500	44.88%	0.74%
	Two or More Races	≥140	1.95%	≥500	1.99%	0.03%
	Economic Status					
	Economically Disadvantaged	≥4,470	60.73%	≥15,750	61.47%	0.74%
	Not Economically Disadvantaged	≥2,320	31.54%	≥7,460	29.10%	-2.44%
Geometry	All Students	≥6,800	100.00%	≥18,710	100.00%	0.00%
	Gender					
	Male	≥3,210	47.25%	≥8,890	47.52%	0.27%
	Female	≥3,580	52.75%	≥9,820	52.48%	-0.27%
	Race Ethnicity					
	Hispanic/Latino	≥350	5.22%	≥1,220	6.55%	1.33%
	American Indian or Alaska Native	≥20	0.43%	≥90	0.51%	0.08%
	Asian	≥150	2.23%	≥360	1.97%	-0.27%
	Black or African American	≥2,810	41.43%	≥7,800	41.71%	0.28%
	Native Hawaiian or Other Pacific	<10	NR	≥20	0.14%	NR
	White	≥3,330	49.06%	≥8,850	47.34%	-1.72%
	Two or More Races	≥100	1.51%	≥330	1.78%	0.27%
	Economic Status					
	Economically Disadvantaged	≥3,690	54.37%	≥10,660	56.98%	2.62%
	Not Economically Disadvantaged	≥2,740	40.40%	≥6,970	37.25%	-3.15%

Table 6.4 Summary of Calibration and Census Data: Spring Administration Form E

Calibration and Census Data: Spring Form E Administration						
Course		Calibration Sample		Census Data Initial Testers Only		Census % - Calib %
		N	%	N	%	
English I	All Students	≥7,250	100.00%	≥20,890	100.00%	0.00%
	Gender					
	Male	≥3,450	47.61%	≥10,040	48.06%	0.45%
	Female	≥3,800	52.39%	≥10,850	51.94%	-0.45%
	Race Ethnicity					
	Hispanic/Latino	≥470	6.55%	≥1,420	6.81%	0.26%
	American Indian or Alaska Native	≥50	0.76%	≥150	0.73%	-0.03%
	Asian	≥120	1.78%	≥370	1.80%	0.02%
	Black or African American	≥3,160	43.57%	≥8,780	42.03%	-1.54%
	Native Hawaiian or Other Pacific	<10	NR	≥10	0.09%	NR
	White	≥3,290	45.46%	≥9,740	46.62%	1.16%
	Two or More Races	≥130	1.81%	≥400	1.92%	0.12%
	Economic Status					
	Economically Disadvantaged	≥4,270	58.87%	≥12,210	58.45%	-0.42%
	Not Economically Disadvantaged	≥2,330	32.19%	≥6,640	31.81%	-0.38%
English II	All Students	≥8,410	100.00%	≥18,980	100.00%	0.00%
	Gender					
	Male	≥4,010	47.79%	≥9,100	47.97%	0.18%
	Female	≥4,390	52.21%	≥9,870	52.03%	-0.18%
	Race Ethnicity					
	Hispanic/Latino	≥480	5.79%	≥1,110	5.88%	0.09%
	American Indian or Alaska Native	≥80	0.96%	≥130	0.70%	-0.26%
	Asian	≥140	1.72%	≥370	1.96%	0.24%
	Black or African American	≥3,630	43.21%	≥7,680	40.46%	-2.75%
	Native Hawaiian or Other Pacific	<10	NR	≥10	0.09%	NR
	White	≥3,910	46.60%	≥9,300	49.00%	2.40%
	Two or More Races	≥140	1.68%	≥360	1.91%	0.23%
	Economic Status					
	Economically Disadvantaged	≥4,860	57.81%	≥10,820	57.01%	-0.80%
	Not Economically Disadvantaged	≥2,960	35.28%	≥6,910	36.45%	1.17%

Table continues

Calibration and Census Data: Spring Form E Administration (continued)						
		Calibration Sample		Census Data Initial Testers Only		
Course		N	%	N	%	Census % - Calib %
Algebra I	All Students	≥5,750	100.00%	≥20,920	100.00%	0.00%
	Gender					
	Male	≥2,770	48.24%	≥10,170	48.64%	0.41%
	Female	≥2,970	51.76%	≥10,740	51.36%	-0.41%
	Race Ethnicity					
	Hispanic/Latino	≥330	5.86%	≥1,200	5.76%	-0.10%
	American Indian or Alaska Native	≥20	0.50%	≥130	0.66%	0.16%
	Asian	≥110	1.96%	≥390	1.87%	-0.10%
	Black or African American	≥2,540	44.24%	≥8,620	41.25%	-2.99%
	Native Hawaiian or Other Pacific	<10	NR	≥10	0.05%	NR
	White	≥2,610	45.47%	≥10,120	48.41%	2.94%
	Two or More Races	≥110	1.95%	≥410	2.00%	0.06%
	Economic Status					
	Economically Disadvantaged	≥3,450	60.11%	≥12,080	57.75%	-2.36%
	Not Economically Disadvantaged	≥1,840	31.99%	≥6,760	32.34%	0.34%
Geometry	All Students	≥6,560	100.00%	≥16,700	100.00%	0.00%
	Gender					
	Male	≥3,020	46.13%	≥7,760	46.52%	0.39%
	Female	≥3,530	53.87%	≥8,930	53.48%	-0.39%
	Race Ethnicity					
	Hispanic/Latino	≥330	5.08%	≥880	5.28%	0.20%
	American Indian or Alaska Native	≥20	0.38%	≥90	0.59%	0.21%
	Asian	≥150	2.35%	≥360	2.17%	-0.17%
	Black or African American	≥2,760	42.16%	≥6,650	39.86%	-2.31%
	Native Hawaiian or Other Pacific	<10	NR	≥10	0.09%	NR
	White	≥3,170	48.46%	≥8,360	50.07%	1.61%
	Two or More Races	≥100	1.52%	≥320	1.95%	0.42%
	Economic Status					
	Economically Disadvantaged	≥3,520	53.70%	≥9,130	54.68%	0.98%
	Not Economically Disadvantaged	≥2,660	40.61%	≥6,550	39.25%	-1.36%



### 6.3 Calibration and Linking

Item calibration and linking for the LEAP 2025 high school assessments were performed based on item response theory (IRT). The calibration and linking methodology used for the spring 2019 LEAP 2025 High School administration closely followed most of the PARCC methods referenced in the PARCC document *Final Technical Report for 2015 Administration*. To maintain comparability to PARCC, the 2PL/GPC IRT model was applied to item calibration using the software IRTPRO (Cai et al., 2011). To avoid local independence between traits, the writing traits written expression (WE) and knowledge and use of language conventions (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 used to generate final scoring tables. All IRTPRO and STUIRT command files were prepared following PARCC examples. Descriptions of the PARCC calibration and equating approach can be found in the PARCC documents *Final Technical Report for 2015 Administration* (see <https://eric.ed.gov/?q=source%3a%22Partnership+for+Assessment+of+Readiness+for+College+and+Careers%22&id=ED599097>) and *Final Technical Report for 2016 Administration* (see <https://eric.ed.gov/?q=PARCC+Final+Technical+Report&id=ED599197>).

There were two CBT test forms per course for the spring administration. Table 6.5 summarizes the student count and item count by course.

The following steps were taken to place the 2018–2019 LEAP 2025 tests on the 2018 LEAP 2025 baseline scale:

- 1.1.1 Calibrate the LEAP 2025 High School tests.
- 1.1.2 Link the 2018–2019 LEAP 2025 High School tests to the 2018 LEAP 2025 baseline scale under the nonequivalent common item design.

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

#### Calibration of the 2018–2019 LEAP 2025 Tests

For the LEAP 2025 item calibration, the 2PL/GPC IRT model was applied to the Louisiana students' calibration samples using the software IRTPRO (Cai et al., 2011). Table 6.5 shows the number of students in the calibration samples and the number of items. Spring calibration included samples of students that were representative of the general population. For English I and II, reading items (RL/RI) in the writing prompts are not counted in the N-Items column because the calibration does not include reading item scores; it only includes WE item scores. A RL/RI score and a WE item score for the same writing prompt are the same. There were thirty-two English items and thirty-nine mathematics items in both administrations. Common items

were included in forms D and E, thus the total N-item per calibration for a course is less than the sum of the N-Items counts in each form.

**Table 6.5 Summary of Student Count and Item Count**

Course	Form	N-Students	N-Items	Total N-Items
English I	D	≥13,800	32	50
	E	≥7,250	32	
English II	D	≥10,420	32	50
	E	≥8,410	32	
Algebra I	D	≥7,360	39	62
	E	≥5,750	39	
Geometry	D	≥6,800	39	62
	E	≥6,560	39	

Figure 6.2 illustrates the common items, equating item sets, and unique items per form. There are separate equating item sets for forms D and E, and when combined, the set mirrors the form blueprint and is used to link to the 2018 LEAP 2025 scale. These across-year equating items were on forms D and E without overlap to mitigate item exposure as the LEAP 2025 assessment will use intact forms for future fall and summer administrations. The two forms were calibrated concurrently and the common items among forms allowed the forms to be on the same scale.

**Figure 6.2 LEAP 2025 HS Spring Administration Forms**

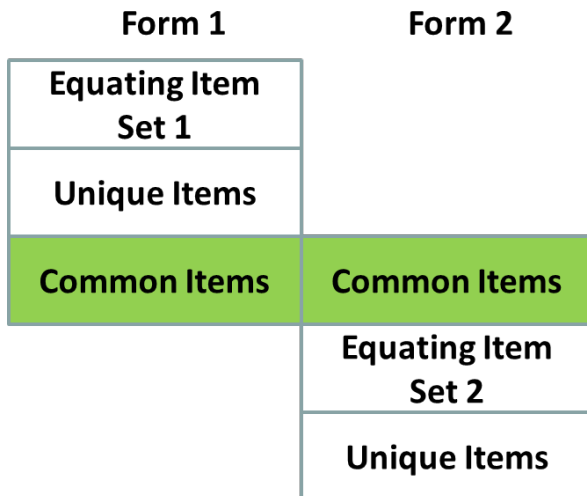
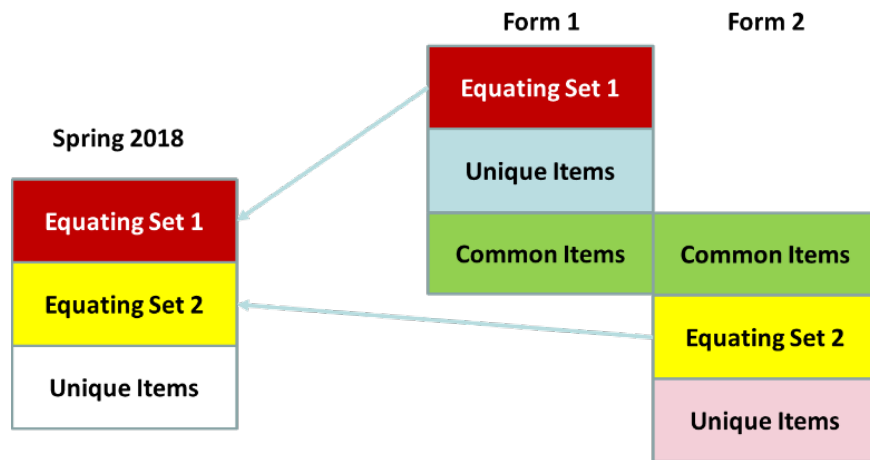


Figure 6.3 illustrates how the 2018–2019 LEAP 2025 forms were linked to the 2018 spring administration. Both Equating Set 1 and Equating Set 2 from forms 1 and 2 (Forms D and E) were used together to obtain one set of Stocking & Lord equating parameters.

Figure 6.3 LEAP 2025 HS Spring Administration Year-to-Year Equating



### Separate Calibration for ELA Prose Constructed-Response Tasks

For English I and II, the same sample repeated design was applied to the WE and WKL calibration to address the issue of local independence for ELA prose constructed response (PCR) tasks. For English I, for example, two datasets of responses were generated. One calibration dataset included two WE responses and the responses to the other items but excluded the WKL responses. The other calibration dataset included two WKL responses and the response of the other items but excluded the WE responses. Therefore, these datasets were the same except for WE and WKL items, 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 6.6 illustrates the calibration data structure for WE.

Table 6.6 Calibration Data Structure for ELA WE

Form	Other Items	Unique WE	Common Items	Common WE	Unique Form 2 Items	Unique WE
1	XXXXXXXX	X	XXXXXXXX	X		
2			XXXXXXXX	X	XXXXXXXX	XXXXXXXX

### IRT Item Fit

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

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

After convergence was achieved for each IRT data set, an item characteristic curve (ICC) for each item was plotted with empirical students’ performances from theta ability -4 to 4. No items were suppressed from

calibration due to poor fit. No items were removed from the anchor sets due to poor fit with the LEAP 2025 High School tests.

After calibration, the IRT model fit was evaluated by reviewing item chi-square values that were calculated using IRTPRO item parameters and item responses from students in the calibration sample. Adjusted fit values were calculated and flagged if they exceeded 0.35 (Pearson, 2017).

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:

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

Table 6.7 shows the adjusted item fit C values using the chi-square statistics and calibration sample sizes for the English and mathematics content. The average adjusted fit ranged from 0.07 to 0.10 for the spring administration. No items were excluded based on model fit statistics because the adjusted item fits for all items were lower than the criterion value of 0.45, as can be seen in the maximum values. The largest adjusted fit value was 0.28 for English I.

**Table 6.7 Summary of Adjusted Fit for Spring Administration**

Course	N-Items	Mean	Std. Dev.	Min.	Max.	N-Flagged Items
English I	50	0.10	0.06	0.03	0.28	0
English II	50	0.09	0.04	0.03	0.22	0
Algebra I	62	0.07	0.03	0.01	0.20	0
Geometry	62	0.07	0.04	0.01	0.21	0

### Linking 2018–2019 LEAP 2025 to PARCC Scale

This section explains the linking procedure used to place the LEAP 2025 High School spring 2019 administration onto the LEAP 2025 scale. The 2017–2018 fall and spring administrations were the first administrations of the LEAP 2025 High School tests, and the Stocking & Lord procedure (1983) was used to link the LEAP 2025 tests to the PARCC scale using intact PARCC items embedded in the test forms. This yielded item parameters on the PARCC scale. The post-equated Louisiana item parameters were based on only Louisiana students' responses; therefore, to distinguish these two sets of item parameters, item parameters based on only Louisiana students' responses will be called LEAP 2025 item parameters and the corresponding scale will be called the LEAP 2025 scale.

Three anchor sets were used in the spring 2019 LEAP 2025 High School assessments equating process. One anchor set was used to link to the LEAP 2025 scale. The use of multiple anchor sets will assist in establishing comparability between the LEAP 2025 scale and the PARCC scale for current and future administrations.

Anchor 1 items were intact PARCC items embedded in the spring 2019 LEAP 2025 forms. Anchor 2 items were items common to the spring 2018 and the spring 2019 LEAP 2025 High School forms. These parameters were generated using Louisiana student data and were used to place the spring administration onto the LEAP 2025 scale. Anchor 3 item parameters consisted of all Anchor 2 item parameters, and Anchor 1 item parameters were used for the remaining items. Table 6.8 summarizes the number and score points of the initial anchor item selection before equating. Table 6.8 also summarizes the number and score points of the final anchor

item selections. The difference between the initial number of anchor items and the final number of anchor items is the number of anchor items that were dropped.

**Table 6.8 The Number of Linking Items for Linking LEAP 2025 Spring 2019 to LEAP 2025 HS Spring 2018**

Course		Anchor 1		Anchor 2		Anchor 3	
		N. of Items	Score Points	N. of Items	Score Points	N. of Items	Score Points
English I	Initial	44	97	14	31	44	97
	Final	33	73	11	25	38	85
English II	Initial	41	91	12	27	41	91
	Final	31	69	9	18	31	68
Algebra I	Initial	39	67	16	25	42	74
	Final	28	52	15	24	37	66
Geometry	Initial	45	79	16	27	46	83
	Final	35	63	14	24	38	70

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

Table 6.9 presents the slope and intercept equating constants for each anchor set for each subject. The constants from Anchor 2 were used to bring the 2019 LEAP 2025 estimated parameters onto the LEAP 2025 scale.

**Table 6.9 Stocking & Lord Transformation Constants for Linking LEAP 2025 Spring 2019 to LEAP 2025 HS Spring 2018**

Course	Anchor 1		Anchor 2		Anchor 3	
	Slope	Intercept	Slope	Intercept	Slope	Intercept
English I	0.867933	0.161118	0.924291	0.089432	0.887791	0.069541
English II	0.882093	0.340173	0.911542	0.253819	0.886878	0.340465
Algebra I	1.006859	-0.043168	0.991766	0.105916	1.007307	0.027138
Geometry	0.961210	0.049833	0.962862	0.211570	0.959705	0.096150

Figures 6.3 through 6.6 show test characteristic curves (TCCs) for the anchor items, the corresponding LEAP 2025 spring 2019 estimated anchor items before equating (ANC) and after equating (EQ\_ANC), intact PARCC parameters (Pre\_EQ), and all 2019 LEAP 2025 estimated items (EQ\_ALL) for the LEAP 2025 High School Spring 2019 assessments after applying the Stocking & Lord equating procedure. The blue solid line illustrates the pre-equated anchor items, the red dotted line is the 2019 LEAP 2025 equated anchor items, the black solid line is the form construction target (2018 LEAP 2025), and the green dotted line and the brown solid line are the 2019 LEAP 2025 High School equated items from forms D and E, respectively. Anchor items for anchor sets 1, 2, and 3 are different as mentioned above. For all tests, the TCCs for anchor items and the corresponding 2019 estimated anchor items overlapped across most ability levels.

When anchor (blue) and pre-equated (black) TCCs are close to each other, it means the anchor value in the condition is similar to the TCC with parameters' pre-equated form. The anchor and pre-equated TCCs are closely overlapped with most tests with the Anchor 1 condition.

Pre-equated (black) and equated (green) TCCs overlapped closely with all tests with the Anchor 1 condition. Under the Anchor 2 condition, those TCCs were slightly farther from each other with English I and English II, indicating that equating results could differ if parameters from the fall 2017 administration of the LEAP 2025

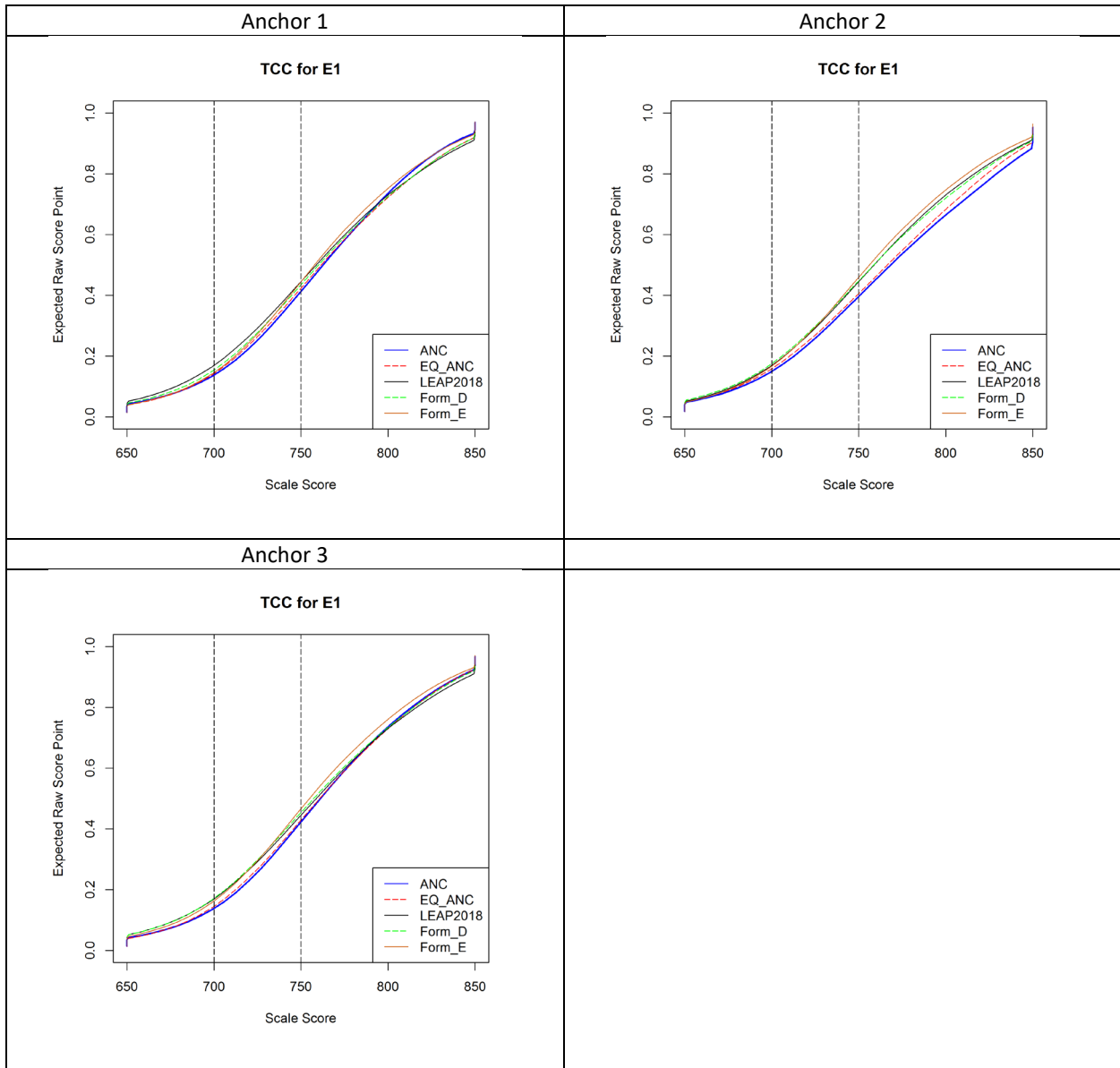
assessments were used. The number of anchor items with Anchor 2 conditions is small, thus it is likely that the parameter difference as well as the limited number of anchor items resulted in the TCCs difference. To clarify, the results from Anchor 2 were used as the final equating results for the spring 2019 LEAP 2025 High School tests.

Figures 6.7 to 6.10 present scatterplots of slope item parameters and difficulty item parameters and their correlation after linking 2019 LEAP 2025 to the LEAP 2025 scale.

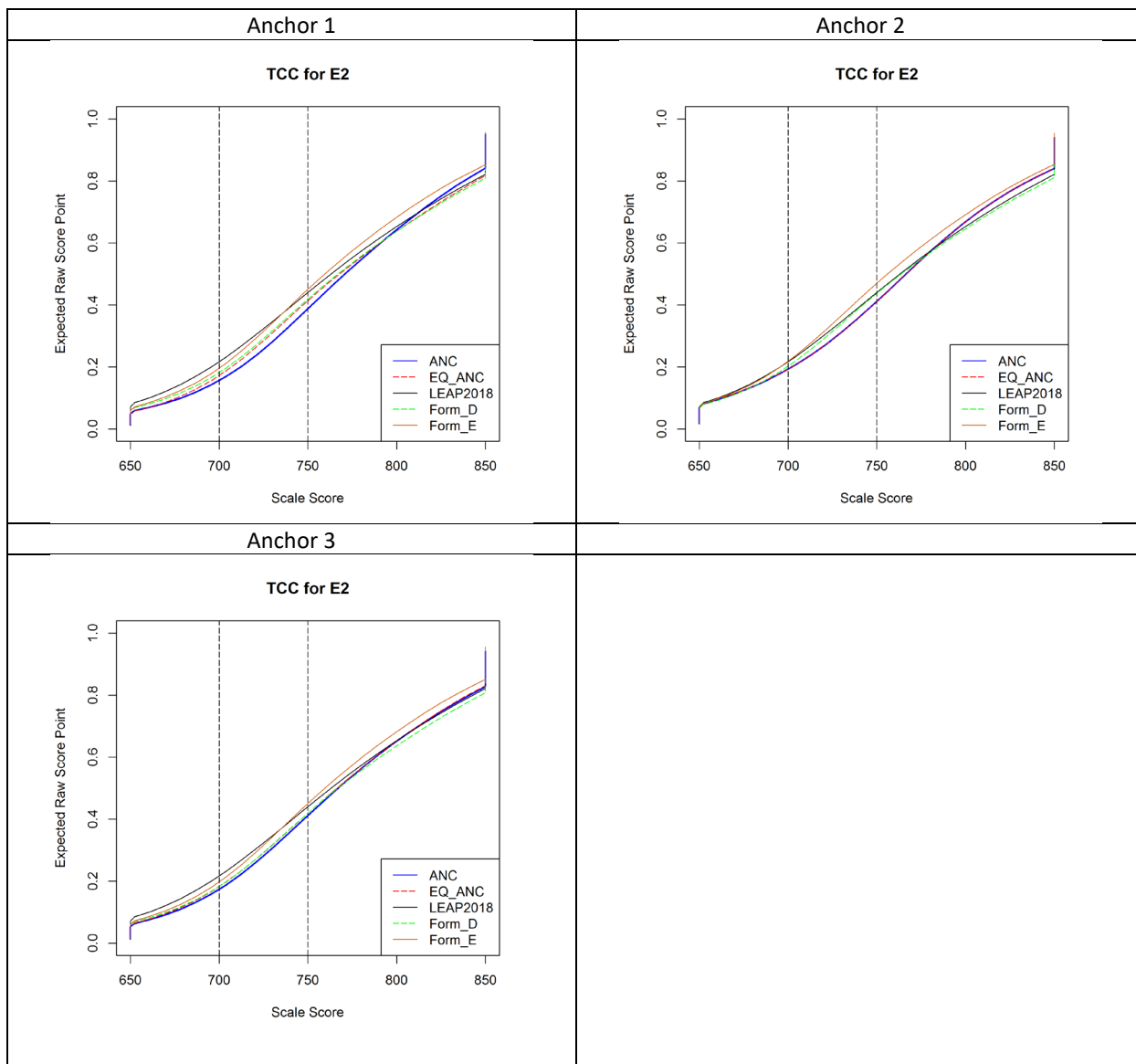
The parameters were mostly close to the identity line with Anchor 1, and the correlation ranged from 0.93 to 0.98. Most parameters were around the identity line with Anchor 2, as the anchor parameters were from Louisiana students, except the writing items in English I and English II. The correlation for Anchor 2 ranged from 0.95 to 0.98.

Compared to English I and English II, Algebra I and Geometry item discrimination parameters were a little scattered from the identity line. It is usual to find higher correlations for difficulty parameters than those for slope parameters.

**Figure 6.4 English I TCC between Anchor, 2019 LEAP 2025 Equated Anchor, LEAP 2025 Spring 2018, and Forms D and E with ALL 2019 LEAP 2025 Items**

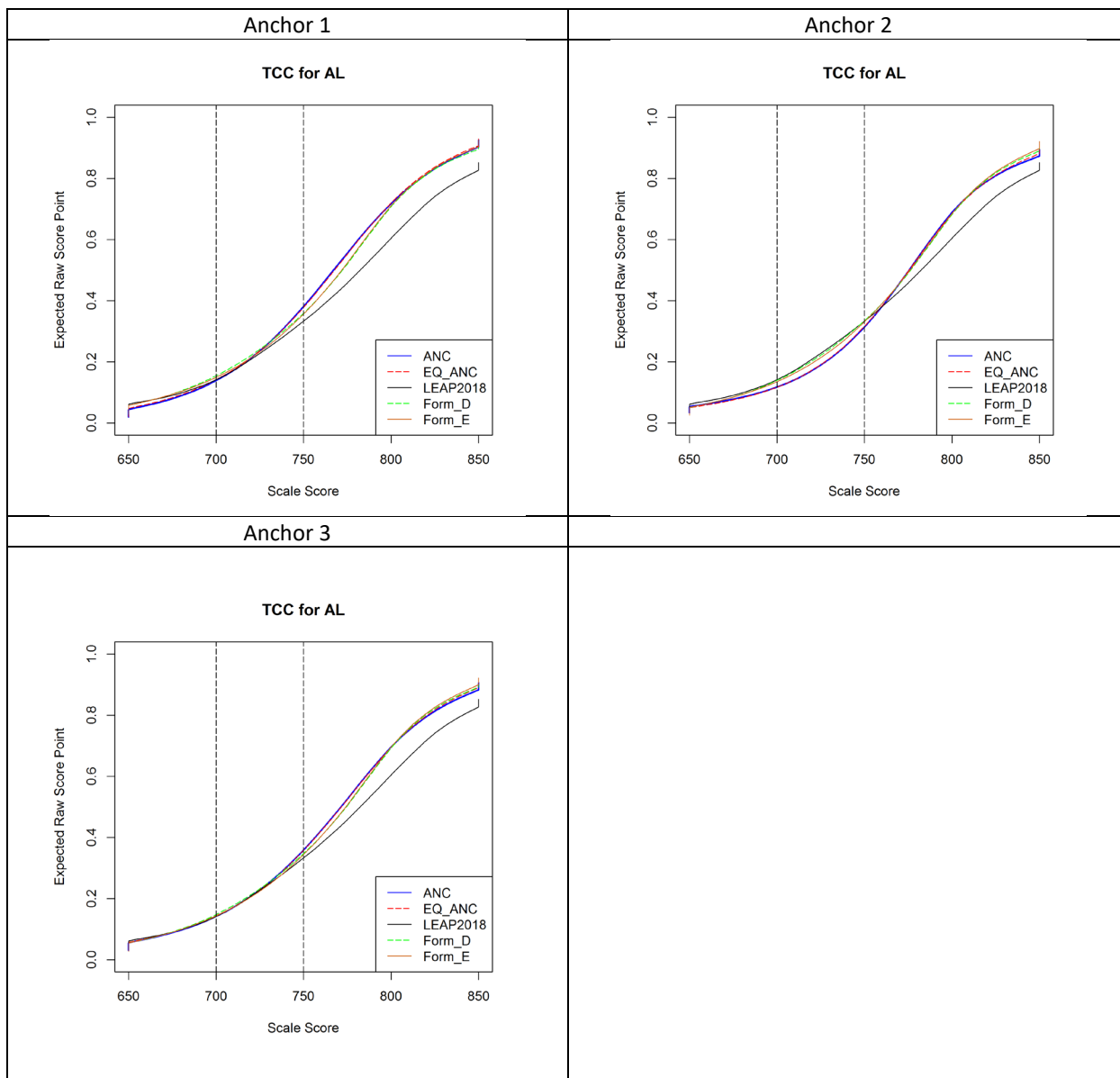


**Figure 6.5 English II TCC between Anchor, 2019 LEAP 2025 Equated Anchor, LEAP 2025 Spring 2018, and Forms D and E with ALL 2019 LEAP 2025 Items**

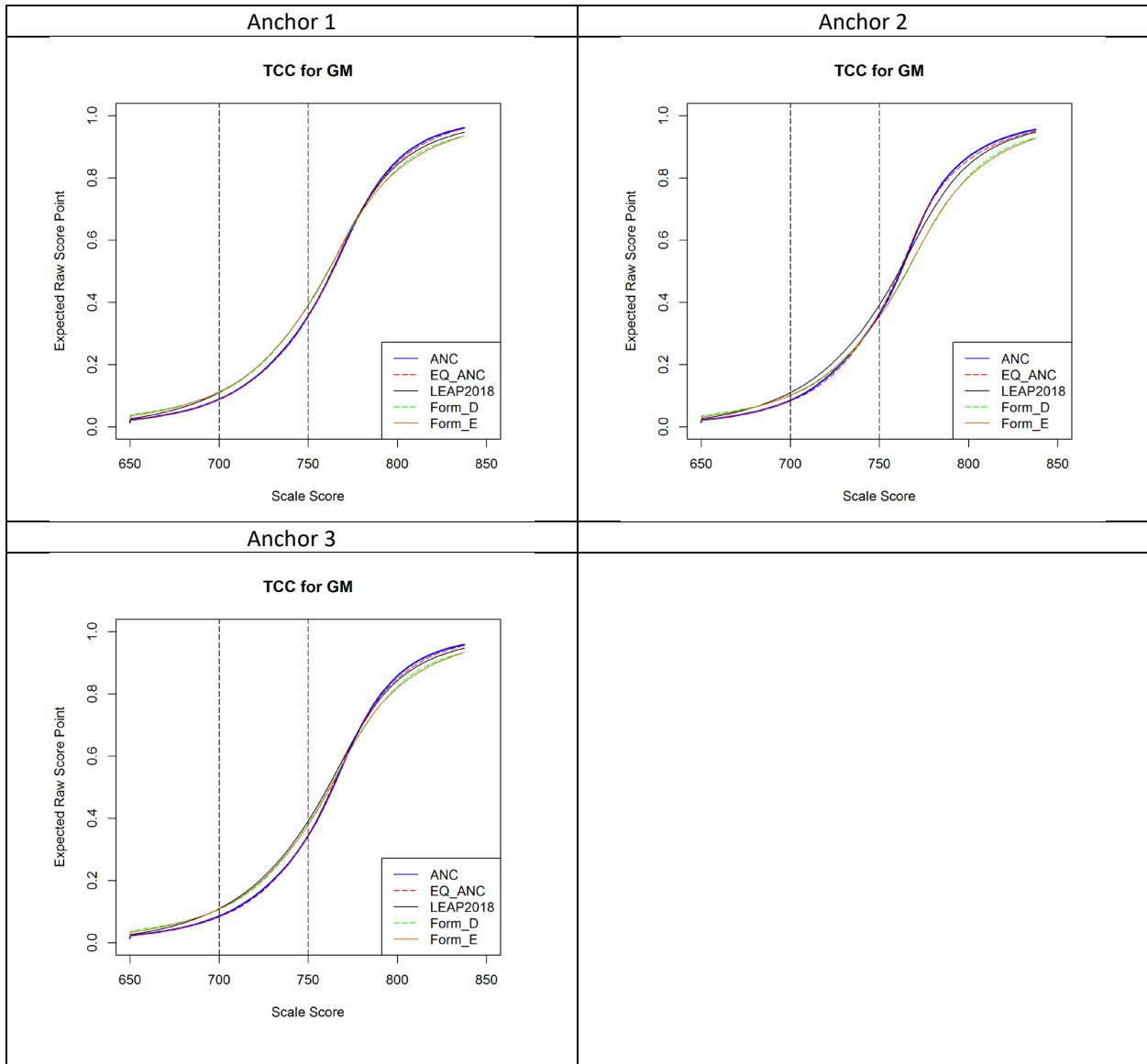




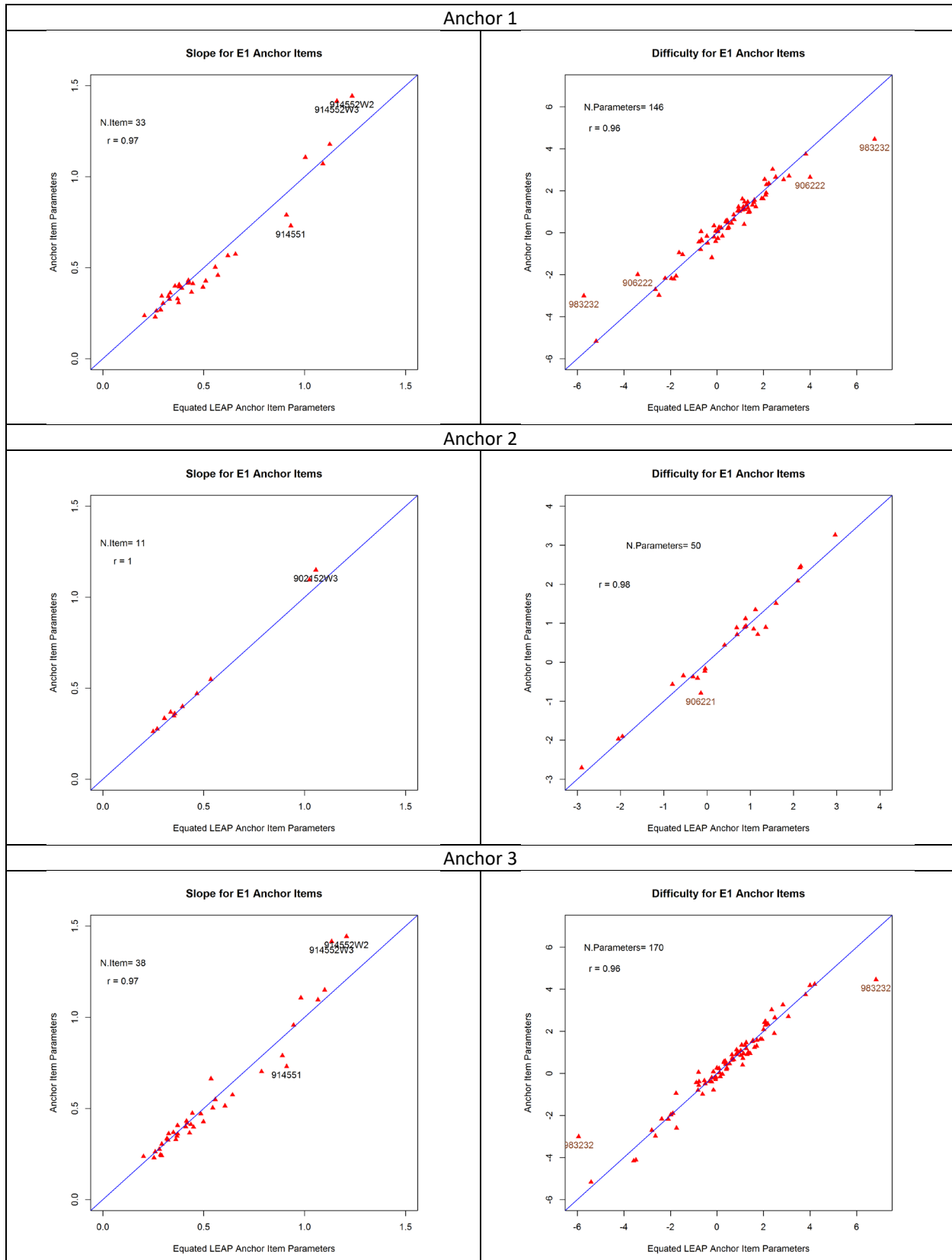
**Figure 6.6 Algebra I TCC between Anchor, 2019 LEAP 2025 Equated Anchor, LEAP 2025 Spring 2018, and Forms D and E with ALL 2019 LEAP 2025 Items**



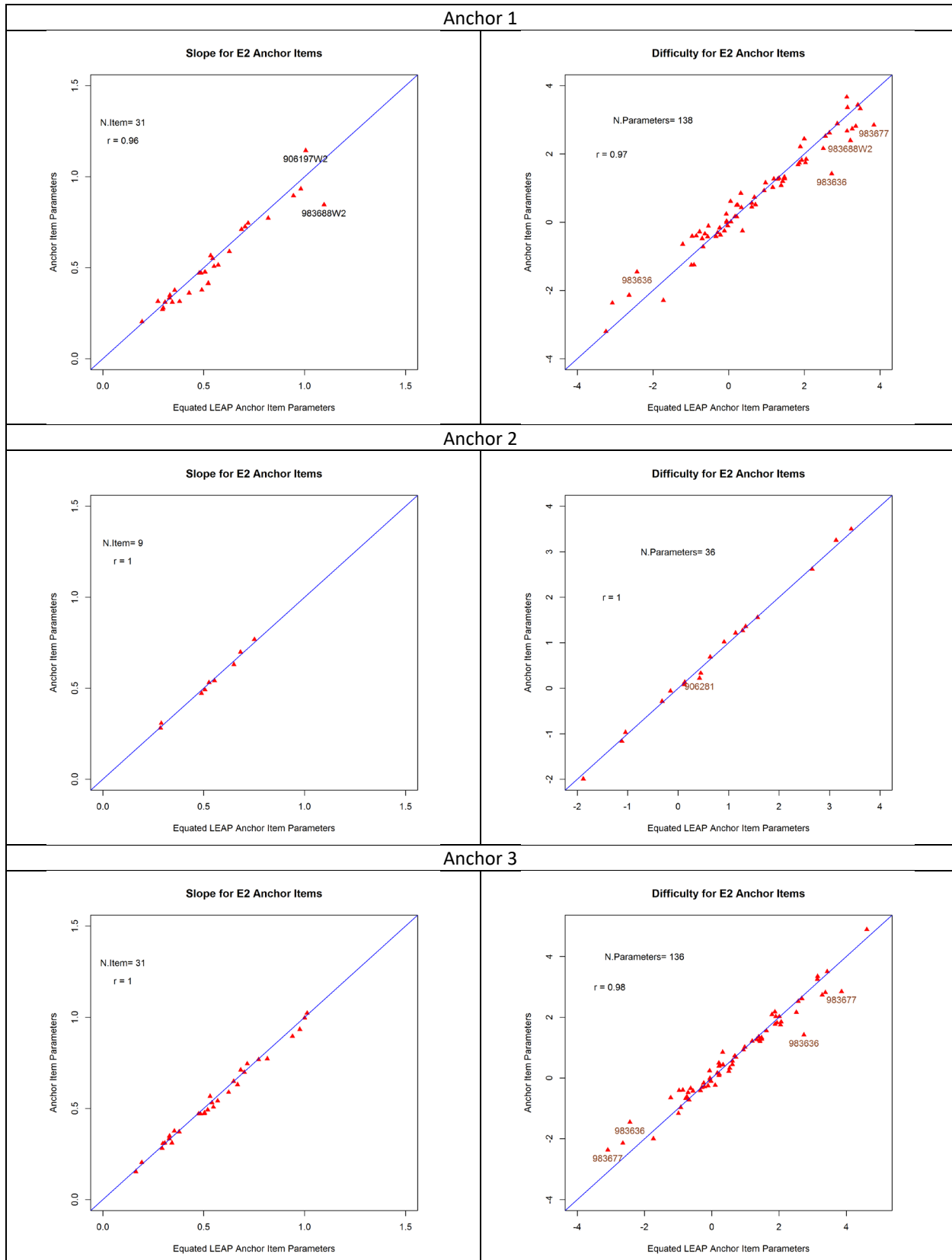
**Figure 6.7 Geometry TCC between Anchor, 2019 LEAP 2025 Equated Anchor, LEAP 2025 Spring 2018, and Forms D and E with ALL 2019 LEAP 2025 Items**



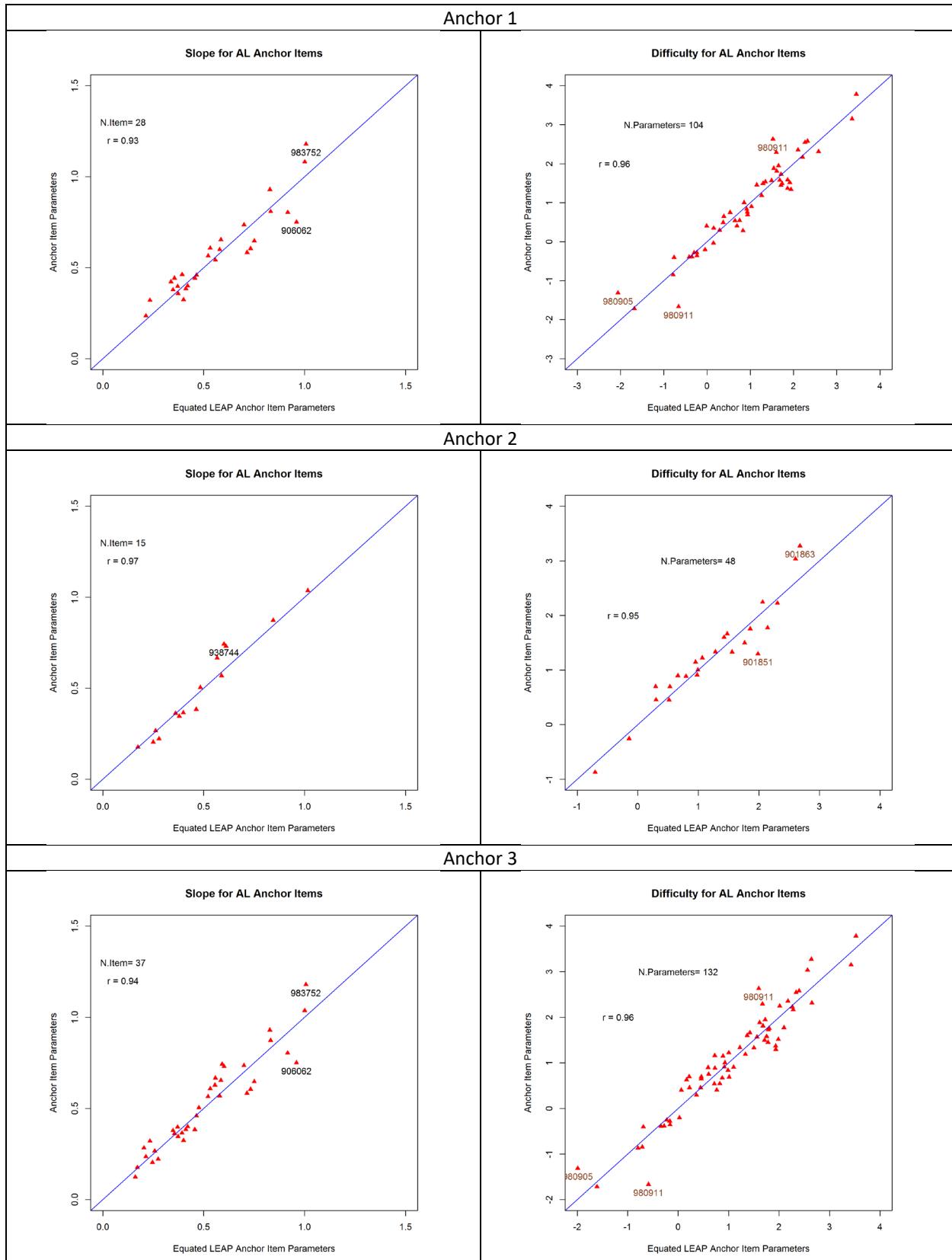
**Figure 6.8 English I Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters**



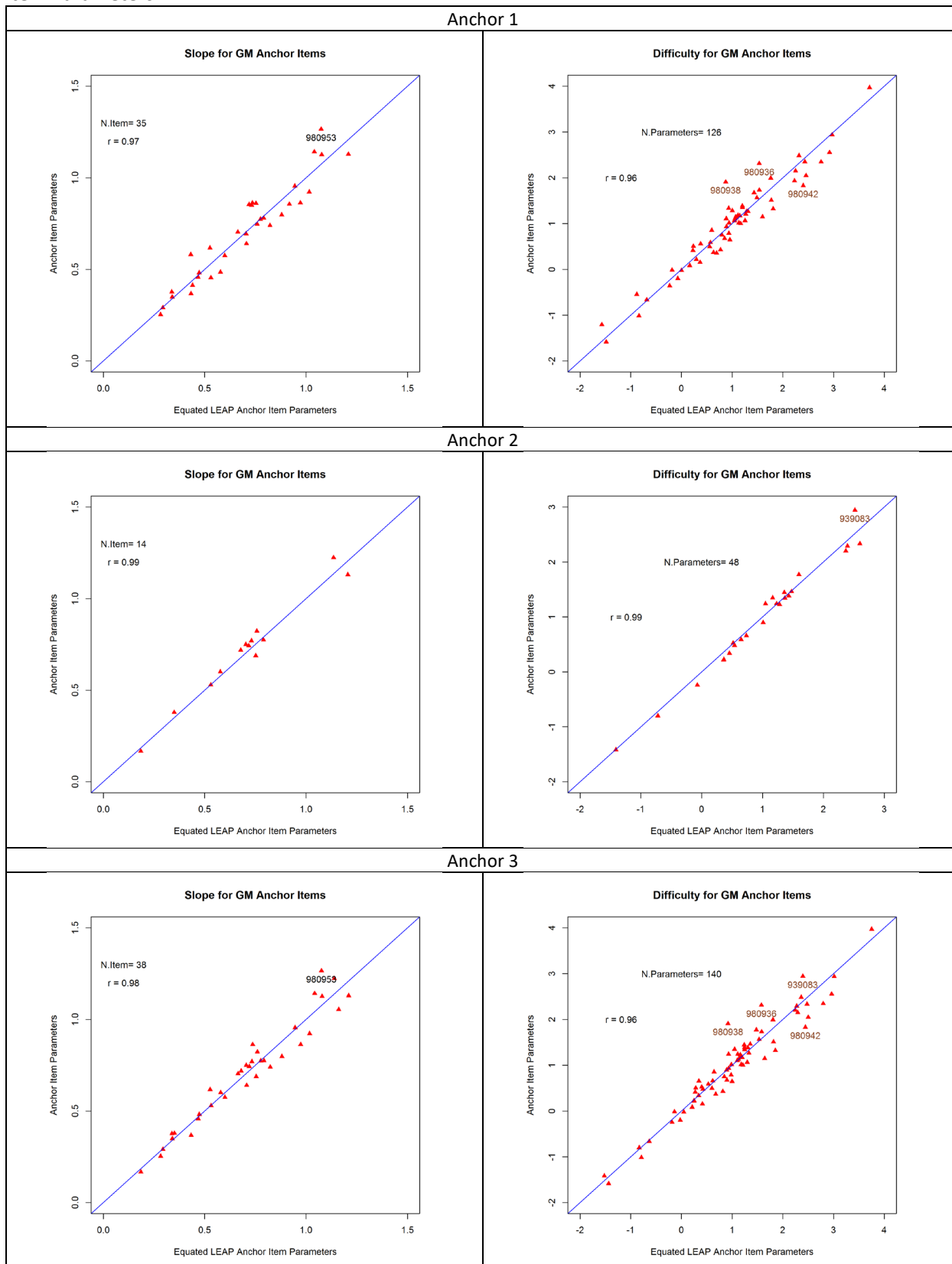
**Figure 6.9 English II Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters**



**Figure 6.10 Algebra Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters**



**Figure 6.11 Geometry Anchor Slope and Difficulty between Equated Anchor Item parameters and Anchor Item Parameters**



## Evaluation of Anchor Item Stability

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

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

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

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

- Exclude CR items from anchor set if categories were collapsed due to small sample size.
- Exclude items with content or parameter estimation issues.
- Run Robust Z method and remove flagged items from anchor set using the criterion value of  $|1.96|$ .
- Run STUIRT using the remaining items after Robust Z is applied, and flag items for further inspection if the WRMSD is greater than the values in Table 6.10. If the items are flagged, then they are 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 category and subcategory before and after dropping an anchor item.

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

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

where  $D$  is the difference between intact and estimated item parameters of an anchor item and  $Md$  is a median of differences between intact and estimated item parameters for all items. The critical value of  $\pm 1.96$  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

$$WRMSD = \sqrt{\sum_{Q=1}^{41} W_Q [ICC_Q(EST) - ICC_Q(INTACT)]^2},$$

where  $Q$  represents a quadrature point (i.e., node),  $W$  represents its weight given quadrature point  $Q$  from IRTPRO output,  $INTACT$  represents intact PARCC item parameters, and  $EST$  represents estimated item parameters corresponding to intact PARCC item parameters. Table 6.10 summarizes WRMSD flagging criteria for inspection and possible removal of linking items.

**Table 6.10 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
$\geq 8$	$\geq 7$	0.090	0.999

#### Lowest and Highest Obtainable Scale Scores

A maximum likelihood (MML) procedure cannot produce scale score estimates for students with perfect scores or scores below the level expected when students are guessing. In addition, although MML estimates are available for students with extreme scores other than zero or perfect, occasionally these estimates have standard errors of measurement that are very large, and differences between these extreme values have little meaning. Therefore, scores are established for these students based on a rational but necessary non-MML procedure. These values, which are set separately by course, are called the lowest obtainable scale score (LOSS) and the highest obtainable scale score (HOSS). All LEAP 2025 High School content areas in 2018 used the same LOSS and HOSS values established by PARCC. The LOSS value was 650, and the HOSS value was 850.

#### Category- and Subcategory-Level Subscores

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

Additionally, subcategory ratings are reported at the student level. English I and English II have three subcategories for reading (i.e., literary text, informational text, and vocabulary) and two subcategories for writing (i.e., written expression and knowledge and use of language conventions).

Algebra I and Geometry have four reporting categories. Category A, Major Content, is further reported at three subcategories in algebra I and two subcategories in Geometry. Category and subcategory performance are reported in one of three ratings: *Strong*, *Moderate*, or *Weak*.

Although the performance ratings are determined only by the items included within a category or subcategory, the level of knowledge and ability needed to achieve a performance rating is connected to the level of knowledge and ability required to reach the achievement levels in the overall test: 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.

The 2018–2019 LEAP 2025 High School reporting categories are summarized in Table 6.11.

**Table 6.11 LEAP 2025 High School Reporting Categories**

Course	Category	Subcategory
English I/II	<ol style="list-style-type: none"> <li>1. Reading</li> <li>2. Writing</li> </ol>	<ol style="list-style-type: none"> <li>1. Reading Informational Texts—RI</li> <li>2. Reading Literature—RL</li> <li>3. Reading Vocabulary—RV</li> <li>4. Written Expression—WE</li> <li>5. Written Knowledge of Language—WKL</li> </ol>
Algebra I	<ol style="list-style-type: none"> <li>1. A—Major Content</li> <li>2. B—Additional and Supporting Content</li> <li>3. C—Expressing Mathematical Reasoning</li> <li>4. D—Modeling and Application</li> </ol>	<ol style="list-style-type: none"> <li>1. A.1—Interpreting Functions</li> <li>2. A.2—Solving Algebraically</li> <li>3. A.3—Solving Graphically/Rate of Change</li> </ol>
Geometry	<ol style="list-style-type: none"> <li>1. A—Major Content</li> <li>2. B—Additional and Supporting Content</li> <li>3. C—Expressing Mathematical Reasoning</li> <li>4. D—Modeling and Application</li> </ol>	<ol style="list-style-type: none"> <li>1. A.1—Congruence Transformations/Similarity</li> <li>2. A.2—Similarity in Trigonometry/Modeling and Applying</li> </ol>

Reading and writing category scores were produced for English I and English II. The reading category score range was 10–90 and the writing category score range was 10–60. The method for scaling the reporting category scores followed the PARCC methodology (Pearson, 2017). For the reading category, two theta score points corresponding to English I and English II scale scores of 725 and 750 were used for scaling. Linear transformation constants mapping the two theta points to the scale score points of 40 and 50 were calculated for the reading category. After these transformation values were applied to item parameters belonging to the reading category, a scoring table was generated using the TCC inverse method. A similar approach was applied to scale the writing category, using two scale score points of 30 and 35. Two cut scores, 40 and 50 for reading and 30 and 35 for writing, were used to produce three performance-level ratings for each category (see Table 6.12 for cut scores for summatives, categories, and subcategories).

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

**Table 6.12 Cut Scores for Summative, Category, and Subcategory**

Performance Level	Summative Test	ELA Category		ELA Subcategory/ Mathematics Category/Subcategory*
		Reading	Writing	
1				
2	700	30	25	
3	725	40	30	$\theta_{L3}$
4	750	50	35	$\theta_{L4}$
5	Around 800			

\*Thetas are those from summative tests (i.e., 725 & 750).

Note: Yellow highlight shows cut scores for category and subcategory.

## 6.4 Comparability: Form Equating

The primary purpose of form equating is to establish score equivalency between two (or more) forms. Equivalency is established by first building the forms to be equated according to tight content specifications. Then the form scores are placed on the same scale (by equating), such that students performing on two scaled assessments at the same level of underlying achievement should receive the same scale score on both forms, 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 differences in item difficulty. LDOE and DRC strive to maintain equivalent samples or use near-census samples over the years, minimizing the potential differences caused by the different samples. Differences in the raw-to-scale-score relationship, therefore, can be primarily attributed to the differences in item difficulty.

In the spring of 2019, the forms used were post-equated forms linking to the LEAP 2025 scale. The equating was conducted using the test characteristic transformation function method in the common-item non-equivalent-groups design (Stocking & Lord, 1983). The fall 2018 and summer 2019 forms were intact forms.

Table 6.13 through Table 6.16 provide scale scores at selected percentiles that can be used to compare the distributional characteristics of the LEAP 2025 2018–2019 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 the proficiency (i.e., achievement) of the students in the samples or growth in student achievement across years; (2) unevenness in the respective distributions that combine with the number-correct-to-scale-score scoring method, leaving “gaps” in the scale; or (3) other sources of equating error. Other sources of equating error can include subtle content differences between forms, handscoring differences, or unusual student samples. Some equating errors will always be present between forms. This means that the forms will not measure identically, even under optimal testing conditions. In general, however, the test characteristic function equating techniques will “level” the equated forms through the raw-to-scale-score adjustment.

**Table 6.13 Comparisons of Scale Scores at Selected Percentiles—English I**

	<b>2017</b>	<b>2018</b>	<b>2018</b>	<b>2019</b>	<b>2019</b>	<b>2018</b>	<b>2019</b>
<b>Percentile</b>	<b>Fall</b>	<b>Fall</b>	<b>Spring</b>	<b>Spring Form D</b>	<b>Spring Form E</b>	<b>Summer</b>	<b>Summer</b>
99	818	821	824	820	824	756	750
95	795	796	799	796	802	736	734
90	784	784	788	785	790	727	725
85	776	774	778	777	782	721	719
80	771	767	773	769	776	715	715
75	765	760	767	764	770	710	713
70	762	755	762	759	765	708	709
65	757	749	757	753	759	704	707
60	753	742	754	748	756	701	704
55	750	737	749	743	753	699	702
50	745	731	745	738	748	693	699
45	742	726	740	732	744	691	696
40	737	720	737	729	739	688	693
35	734	712	731	723	734	684	693
30	728	707	728	717	729	684	689
25	723	700	722	713	726	681	686
20	717	694	716	706	718	677	682
15	711	688	707	699	712	673	677
10	702	681	697	686	705	669	677
5	693	666	685	674	691	658	665
1	671	650	660	650	669	650	650

**Table 6.14 Comparisons of Scale Scores at Selected Percentiles—English II**

	2017	2018	2018	2019	2019	2018	2019
Percentile	Fall	Fall	Spring	Spring Form D	Spring Form E	Summer	Summer
99	838	846	846	842	847	760	761
95	805	811	817	810	818	737	733
90	784	788	799	795	802	725	722
85	772	775	788	787	791	720	714
80	763	763	780	779	783	715	709
75	754	754	773	771	778	709	704
70	748	747	765	764	770	703	702
65	740	738	761	759	765	700	699
60	734	731	754	752	761	696	693
55	726	724	749	745	754	689	690
50	720	717	745	741	749	685	687
45	714	712	738	734	743	681	684
40	707	707	733	728	739	676	680
35	702	702	726	722	732	676	676
30	694	693	722	715	726	671	672
25	688	687	714	705	720	666	668
20	684	680	707	697	713	659	663
15	677	672	699	688	703	659	663
10	668	663	687	672	693	652	652
5	658	652	668	656	675	650	650
1	650	650	650	650	650	650	650

**Table 6.15 Comparisons of Scale Scores at Selected Percentiles—Algebra I**

	<b>2017</b>	<b>2018</b>	<b>2018</b>	<b>2019</b>	<b>2019</b>	<b>2018</b>	<b>2019</b>
<b>Percentile</b>	<b>Fall</b>	<b>Fall</b>	<b>Spring</b>	<b>Spring Form D</b>	<b>Spring Form E</b>	<b>Summer</b>	<b>Summer</b>
99	801	810	827	836	839	758	759
95	778	785	800	799	803	738	739
90	761	772	787	786	789	730	732
85	749	760	777	776	780	724	728
80	743	754	769	768	772	721	725
75	733	744	763	761	766	718	721
70	727	738	757	753	761	714	717
65	724	734	751	748	756	714	717
60	721	727	748	745	751	710	713
55	718	723	744	739	746	710	713
50	714	719	738	735	740	705	708
45	710	715	734	728	737	705	708
40	710	711	731	725	734	705	704
35	705	711	727	721	730	700	704
30	705	707	723	717	723	700	698
25	700	702	715	712	719	695	698
20	695	696	711	708	714	695	692
15	695	696	707	703	710	688	692
10	688	690	702	697	704	688	685
5	680	683	690	691	692	680	675
1	669	650	673	668	677	655	650

**Table 6.16 Comparisons of Scale Scores at Selected Percentiles—Geometry**

	2017	2018	2018	2019	2019	2018	2019
Percentile	Fall	Fall	Spring	Spring Form D	Spring Form E	Summer	Summer
99	797	799	796	801	801	820	816
95	779	781	779	783	784	760	785
90	768	771	771	774	774	724	738
85	761	764	764	767	768	719	723
80	755	760	758	761	763	716	720
75	749	754	754	755	758	712	717
70	744	751	749	750	753	712	717
65	740	746	746	746	750	709	714
60	736	742	742	742	746	709	710
55	732	736	738	738	742	705	710
50	727	731	734	734	738	705	706
45	724	729	731	731	733	705	706
40	722	724	727	729	731	701	701
35	716	721	724	726	728	701	701
30	712	715	721	723	725	696	696
25	709	711	718	719	722	696	696
20	705	707	715	716	719	696	690
15	701	702	707	712	711	691	690
10	696	697	702	707	706	685	684
5	691	692	697	694	701	678	676
1	678	677	677	675	686	670	666

Additional evidence of comparability can be found by reviewing the test characteristic curves (TCCs) for the spring 2018 and 2019 administrations of the LEAP 2025 assessments seen in Figure 6.12. For most content areas, the TCCs for the two years were similar across ability ranges. The English II content on Form E of spring 2019 was easier than on the spring 2018 and Form D of spring 2019 for high performing students. The Algebra I content spring 2019 forms was easier than on the spring 2018 form for high performing students. Overall, Geometry in spring 2019 was more difficult than in spring 2018. Note that this different difficulty is adjusted by reporting different scale scores given raw scores. A scale score of a difficult form is higher than that of an easy form given the same raw score.

Figure 6.13 shows SEMs for the Spring 2018 and 2019 LEAP 2025 HS assessments. For most content areas, the SEMs were similar across ability ranges, especially in the middle ability ranges.

Figure 6.12 TCCs Across Years: Spring Administrations

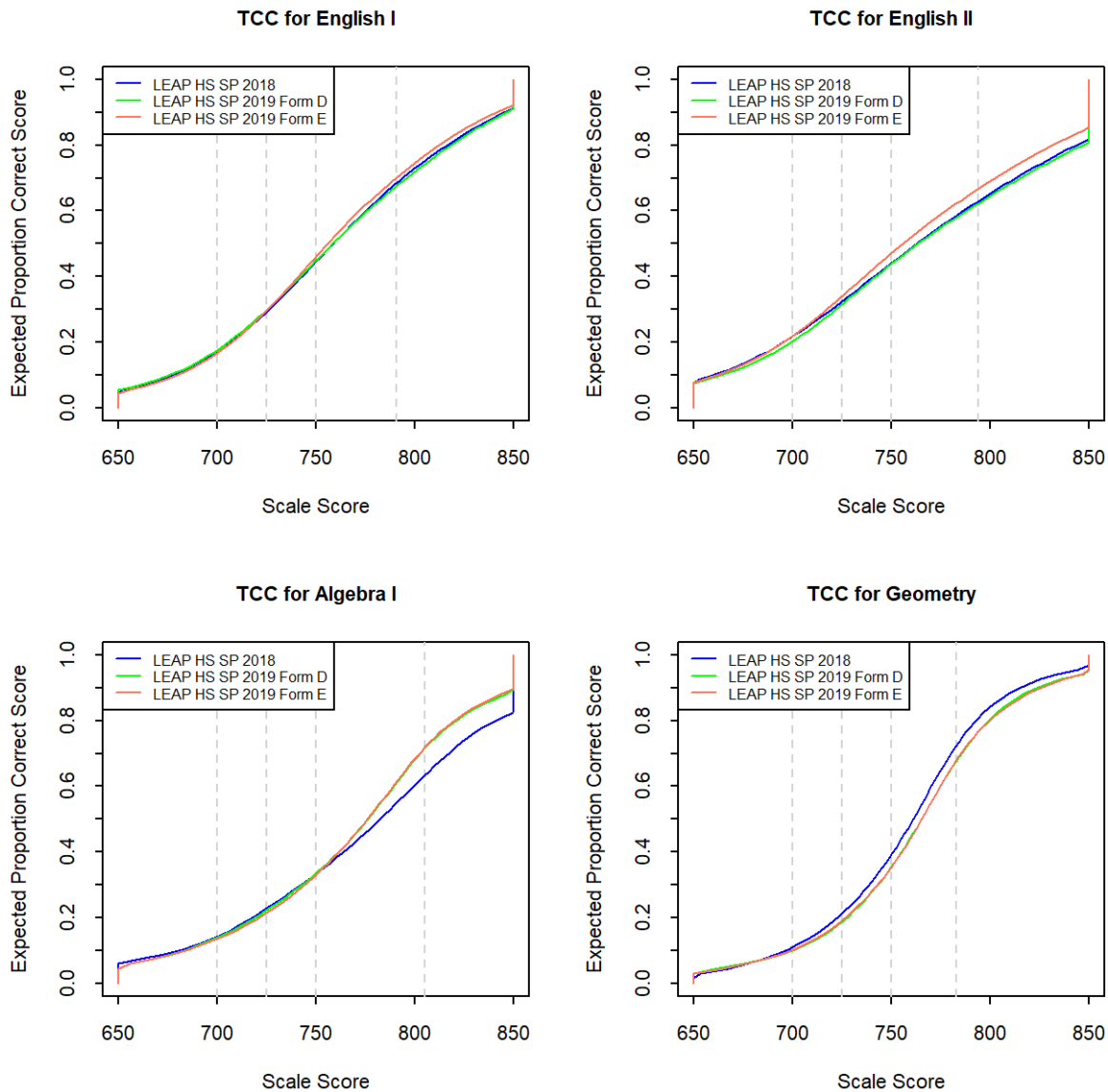
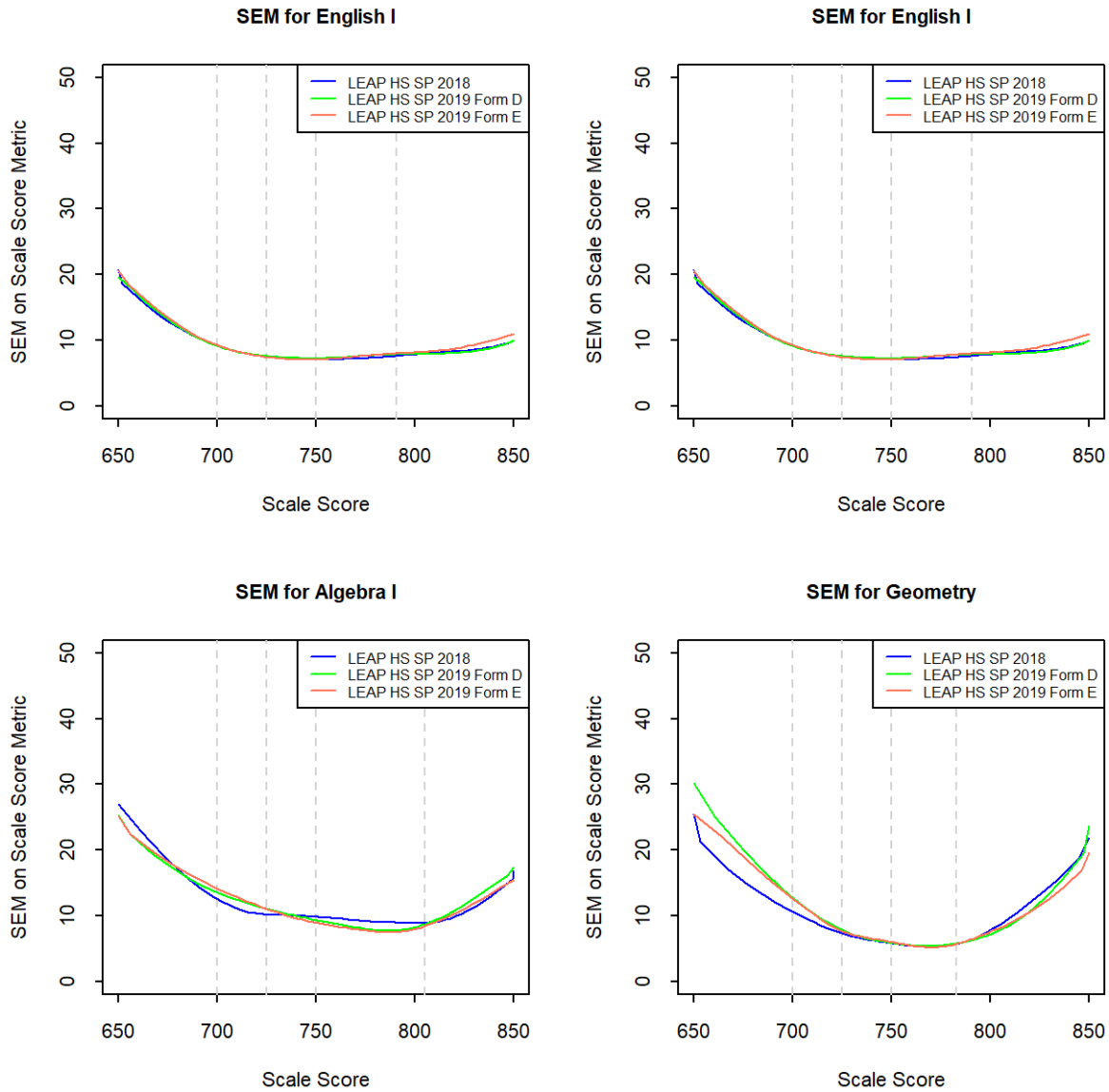


Figure 6.13 SEM Across Years: Spring Administrations



### 6.5 Summary

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

*Standard 1.8* The composition of any sample of test takers from which validity evidence is obtained should be described in as much detail as is practical and permissible, including major relevant socio-demographic and developmental characteristics. (25)



*Standard 4.14* For a test that has a time limit, test development research should examine the degree to which scores include a speed component and should evaluate the appropriateness of that component, given the domain the test is designed to measure. (90)

*Standard 5.2* The procedures for constructing scales used for reporting scores and the rationale for these procedures should be described clearly. (102)

*Standard 5.13* When claims of form-to-form score equivalence are based on equating procedures, detailed technical information should be provided on the method by which equating functions were established and on the accuracy of the equating functions. (105)

*Standard 5.15* In equating studies that employ an anchor test design, the characteristics of the anchor test and its similarity to the forms being equated should be presented, including both content specifications and empirically determined relationships among test scores. If anchor items are used in the equating study, the representativeness and psychometric characteristics of the anchor items should be presented. (105)

*Standard 7.2* The population for whom a test is intended and specifications for the test should be documented. If normative data are provided, the procedures used to gather the data should be explained; the norming population should be described in terms of relevant demographic variables; and the year(s) in which the data were collected should be reported. (126)

## Chapter 7: Test Results

---

This chapter of the technical report contains information on the results of the Spring LEAP 2025 High School administration of English I, English II, Algebra I, and Geometry. The scale score results and achievement level information are presented here. Presenting the results by achievement level translates the quantitative scale provided through scale scores into a qualitative description of student achievement. The levels are *Advanced*, *Mastery*, *Basic*, *Approaching Basic*, and *Unsatisfactory*.

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

This chapter also provides descriptions of the score reports, data structure, and interpretive guide for the LEAP 2025 administrations. The American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA, & NCME, 2014) *Standards for Educational & Psychological Testing* addressed in Chapter 7 are 5.1, 6.10, 7.0, and 12.18. Each standard is presented in the pertinent section of this chapter.

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

### 7.1 Student Participation

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

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

The number of students who attempted each test, the number of students whose results were reportable from each test, and the number of students whose results were included in the technical report sample for each test are summarized by grade in Tables 7.1–7.4. The “Attempted” category includes all the students who attempted at least one item on the assessment. The “Reportable” category includes students who finished all sections in the assessment, which includes students in private school and home-study programs. The “Technical Report Sample” category represents the sample of students included in the analyses for this report, and they are the students who finished all sections of the assessment and counted toward the state total score; students in private school and home-study programs were excluded from this sample.

**Table 7.1 Count of Students who Attempted, were Reportable, and Included in the Technical Report  
Sample: English I**

Administration	Group	Grade							
		6	7	8	9*	10	11	12	Total
Fall	Attempted	<10	<10	≥20	≥5,300	≥1,260	≥180	≥40	≥6,810
	Reportable	<10	<10	≥20	≥5,220	≥1,210	≥180	≥40	≥6,690
	Technical Report	<10	<10	≥20	≥5,220	≥1,210	≥180	≥40	≥6,680
Spring	Attempted	<10	<10	≥2,150	≥42,620	≥3,020	≥340	≥80	≥48,220
	Reportable	<10	<10	≥2,130	≥42,290	≥2,920	≥320	≥80	≥47,750
	Technical Report	<10	<10	≥2,080	≥41,920	≥2,920	≥320	≥80	≥47,330
Summer	Attempted	<10	<10	≥30	≥1,500	≥410	≥20	<10	≥1,970
	Reportable	<10	<10	≥30	≥1,460	≥390	≥20	<10	≥1,910
	Technical Report	<10	<10	≥10	≥1,450	≥390	≥20	<10	≥1,900

\* Grade 9 includes the grade that is coded as "T9."

**Table 7.2 Count of Students who Attempted, were Reportable, and Included in the Technical Report  
Sample: English II**

Administration	Group	Grade							
		6	7	8	9*	10	11	12	Total
Fall	Attempted	<10	<10	<10	≥510	≥6,790	≥1,890	≥550	≥9,760
	Reportable	<10	<10	<10	≥500	≥6,730	≥1,840	≥530	≥9,610
	Technical Report	<10	<10	<10	≥500	≥6,720	≥1,840	≥530	≥9,590
Spring	Attempted	<10	<10	<10	≥2,770	≥37,690	≥2,620	≥920	≥44,000
	Reportable	<10	<10	<10	≥2,730	≥37,510	≥2,530	≥900	≥43,690
	Technical Report	<10	<10	<10	≥2,710	≥37,220	≥2,520	≥900	≥43,360
Summer	Attempted	<10	<10	<10	≥70	≥820	≥560	≥270	≥1,730
	Reportable	<10	<10	<10	≥70	≥800	≥550	≥270	≥1,700
	Technical Report	<10	<10	<10	≥70	≥800	≥550	≥270	≥1,690

\* Grade 9 includes the grade that is coded as "T9."

**Table 7.3 Count of Students who Attempted, were Reportable, and included in the Technical Report  
Sample: Algebra I**

Administration	Group	Grade							
		6	7	8	9*	10	11	12	Total
Fall	Attempted	<10	<10	≥10	≥3,180	≥1,750	≥610	≥250	≥5,810
	Reportable	<10	<10	≥10	≥3,130	≥1,690	≥590	≥240	≥5,680
	Technical Report	<10	<10	≥10	≥3,130	≥1,690	≥590	≥240	≥5,670
Spring	Attempted	≥10	≥210	≥7,790	≥34,600	≥5,940	≥940	≥270	≥49,780
	Reportable	≥10	≥210	≥7,780	≥34,290	≥5,820	≥900	≥260	≥49,300
	Technical Report	≥10	≥210	≥7,700	≥33,950	≥5,810	≥900	≥260	≥48,870
Summer	Attempted	<10	<10	≥30	≥1,350	≥500	≥80	≥30	≥2,010
	Reportable	<10	<10	≥30	≥1,320	≥490	≥80	≥30	≥1,970
	Technical Report	<10	<10	≥20	≥1,320	≥490	≥80	≥30	≥1,950

\* Grade 9 includes the grade that is coded as "T9."

**Table 7.4 Count of Students who Attempted, were Reportable, and Included in the Technical Report  
Sample: Geometry**

Administration	Group	Grade							
		6	7	8	9*	10	11	12	Total
Fall	Attempted	<10	<10	<10	≥1,200	≥3,000	≥910	≥300	≥5,420
	Reportable	<10	<10	<10	≥1,200	≥2,970	≥890	≥290	≥5,370
	Tech Report	<10	<10	<10	≥1,200	≥2,960	≥890	≥290	≥5,350
Spring	Attempted	<10	<10	≥220	≥6,530	≥24,970	≥4,400	≥410	≥36,560
	Reportable	<10	<10	≥220	≥6,500	≥24,820	≥4,340	≥400	≥36,320
	Tech Report	<10	<10	≥220	≥6,460	≥24,600	≥4,300	≥400	≥36,000
Summer	Attempted	<10	<10	≥10	≥20	≥170	≥50	≥20	≥280
	Reportable	<10	<10	≥10	≥20	≥160	≥50	≥20	≥280
	Tech Report	<10	<10	≥10	≥20	≥160	≥50	≥20	≥270

\* Grade 9 includes the grade that is coded as "T9."

The counts and percentages of students in demographic groups by grade for the group of students comprising the technical report sample for the spring 2019 administration are summarized in Table 7.5 through Table 7.20. The same information regarding the technical report samples for the fall 2018 and summer 2019 administrations can be found in Appendix H.

**Table 7.5 Count of Students taking the Spring 2019 LEAP 2025 Administration: English I Form D**

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	≥1,040	≥22,830	≥1,790	≥170	<10	≥25,850
<b>Gender</b>								
Female	<10	<10	≥560	≥11,260	≥590	≥70	<10	≥12,480
Male	<10	<10	≥480	≥11,570	≥1,200	≥100	<10	≥13,360
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	≥110	≥1,410	≥210	≥40	<10	≥1,770
American Indian or Alaska Native	<10	<10	<10	≥150	≥20	<10	<10	≥170
Asian	<10	<10	≥70	≥350	≥10	<10	<10	≥450
Black or African American	<10	<10	≥380	≥10,120	≥980	≥100	<10	≥11,590
Native Hawaiian or Other Pacific	<10	<10	<10	≥10	<10	<10	<10	≥10
White	<10	<10	≥440	≥10,340	≥510	≥20	<10	≥11,320
Two or More Races	<10	<10	≥20	≥430	≥40	<10	<10	≥500
<b>Education Classification</b>								
Regular	<10	<10	≥770	≥18,570	≥1,220	≥120	<10	≥20,680
Special	<10	<10	≥10	≥2,990	≥570	≥50	<10	≥3,630
Gifted	<10	<10	≥250	≥1,260	<10	<10	<10	≥1,520
<b>Economic Status</b>								
Economically Disadvantaged	<10	<10	≥540	≥13,970	≥1,490	≥130	<10	≥16,140
Not Economically Disadvantaged	<10	<10	≥440	≥6,720	≥210	≥20	<10	≥7,410
<b>English Learner Status</b>								
Non-EL	<10	<10	≥1,030	≥22,200	≥1,610	≥130	<10	≥24,970
EL	<10	<10	≥10	≥630	≥180	≥40	<10	≥870
<b>Migrant Status</b>								
Nonmigrant	<10	<10	≥1,040	≥22,800	≥1,780	≥170	<10	≥25,810
Migrant	<10	<10	<10	≥20	≥10	<10	<10	≥40
<b>Section 504 Status</b>								
Non-Section 504	<10	<10	≥980	≥20,510	≥1,490	≥150	<10	≥23,140
Section 504	<10	<10	≥60	≥2,320	≥300	≥20	<10	≥2,700
<b>Homeless Status</b>								
Not Homeless	<10	<10	≥1,030	≥22,440	≥1,750	≥170	<10	≥25,390
Homeless	<10	<10	≥10	≥390	≥40	<10	<10	≥450
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	≥1,000	≥22,530	≥1,790	≥170	<10	≥25,510
Military Affiliated	<10	<10	≥30	≥290	<10	<10	<10	≥340
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	≥1,040	≥22,760	≥1,780	≥170	<10	≥25,760
Foster Care	<10	<10	<10	≥70	≥10	<10	<10	≥80

**Table 7.6 Reportable Percentage of Students taking the Spring 2019 LEAP 2025 Administration:  
English I Form D**

Group	Grade							Total
	6	7	8	9	10	11	12	
<b>All Students</b>	0.00	0.00	4.03	88.32	6.96	0.68	0.00	100
<b>Gender</b>								
Female	0.00	0.00	4.48	90.17	4.72	0.62	0.00	100
Male	0.00	0.00	3.61	86.59	9.05	0.75	0.00	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	6.41	79.26	11.97	2.36	0.00	100
American Indian or Alaska Native	0.00	0.00	0.56	85.96	13.48	0.00	0.00	100
Asian	0.00	0.00	16.30	79.07	3.52	1.10	0.00	100
Black or African American	0.00	0.00	3.29	87.33	8.48	0.90	0.00	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	92.86	7.14	0.00	0.00	100
White	0.00	0.00	3.94	91.27	4.56	0.23	0.00	100
Two or More Races	0.00	0.00	5.15	85.94	8.91	0.00	0.00	100
<b>Education Classification</b>								
Regular	0.00	0.00	3.73	89.77	5.91	0.59	0.00	100
Special	0.00	0.00	0.49	82.38	15.70	1.43	0.00	100
Gifted	0.00	0.00	16.62	82.85	0.39	0.13	0.00	100
<b>Economic Status</b>								
Economically Disadvantaged	0.00	0.00	3.39	86.53	9.24	0.85	0.00	100
Not Economically Disadvantaged	0.00	0.00	6.05	90.72	2.93	0.30	0.00	100
<b>English Learner Status</b>								
Non-EL	0.00	0.00	4.14	88.89	6.46	0.52	0.00	100
EL	0.00	0.00	1.14	72.29	21.21	5.36	0.00	100
<b>Migrant Status</b>								
Nonmigrant	0.00	0.00	4.04	88.36	6.92	0.68	0.00	100
Migrant	0.00	0.00	0.00	63.41	34.15	2.44	0.00	100
<b>Section 504 Status</b>								
Non-Section 504	0.00	0.00	4.23	88.62	6.48	0.67	0.00	100
Section 504	0.00	0.00	2.33	85.83	11.07	0.78	0.00	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	4.06	88.36	6.90	0.68	0.00	100
Homeless	0.00	0.00	2.63	86.18	10.31	0.88	0.00	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	3.94	88.34	7.02	0.69	0.00	100
Military Affiliated	0.00	0.00	10.79	86.88	2.33	0.00	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	4.05	88.33	6.94	0.68	0.00	100
Foster Care	0.00	0.00	0.00	87.21	11.63	1.16	0.00	100

Table 7.7 Count of Students taking the Spring 2019 LEAP 2025 Administration: English I Form E

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	≥1,040	≥19,070	≥1,120	≥140	<10	≥21,390
<b>Gender</b>								
Female	<10	<10	≥600	≥9,900	≥420	≥40	<10	≥10,980
Male	<10	<10	≥430	≥9,170	≥700	≥90	<10	≥10,400
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	≥110	≥1,100	≥140	≥40	<10	≥1,410
American Indian or Alaska Native	<10	<10	<10	≥140	≥10	<10	<10	≥150
Asian	<10	<10	≥70	≥280	≥10	<10	<10	≥370
Black or African American	<10	<10	≥330	≥8,140	≥600	≥60	<10	≥9,150
Native Hawaiian or Other Pacific	<10	<10	<10	≥10	<10	<10	<10	≥10
White	<10	<10	≥480	≥9,010	≥330	≥30	<10	≥9,860
Two or More Races	<10	<10	≥30	≥360	≥10	<10	<10	≥400
<b>Education Classification</b>								
Regular	<10	<10	≥790	≥17,390	≥1,000	≥130	<10	≥19,330
Special	<10	<10	<10	≥530	≥100	≥10	<10	≥650
Gifted	<10	<10	≥240	≥1,140	≥10	<10	<10	≥1,400
<b>Economic Status</b>								
Economically Disadvantaged	<10	<10	≥520	≥11,200	≥900	≥110	<10	≥12,740
Not Economically Disadvantaged	<10	<10	≥470	≥6,090	≥150	≥10	<10	≥6,740
<b>English Learner Status</b>								
Non-EL	<10	<10	≥1,030	≥18,710	≥1,000	≥100	<10	≥20,860
EL	<10	<10	<10	≥360	≥110	≥40	<10	≥520
<b>Migrant Status</b>								
Nonmigrant	<10	<10	≥1,040	≥19,060	≥1,120	≥140	<10	≥21,370
Migrant	<10	<10	<10	≥10	<10	<10	<10	≥20
<b>Section 504 Status</b>								
Non-Section 504	<10	<10	≥990	≥17,670	≥960	≥130	<10	≥19,770
Section 504	<10	<10	≥40	≥1,400	≥150	≥10	<10	≥1,610
<b>Homeless Status</b>								
Not Homeless	<10	<10	≥1,030	≥18,750	≥1,100	≥140	<10	≥21,030
Homeless	<10	<10	<10	≥320	≥20	<10	<10	≥350
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	≥1,000	≥18,770	≥1,120	≥140	<10	≥21,040
Military Affiliated	<10	<10	≥30	≥300	<10	<10	<10	≥350
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	≥1,040	≥19,020	≥1,120	≥140	<10	≥21,320
Foster Care	<10	<10	<10	≥50	<10	<10	<10	≥60

**Table 7.8 Reportable Percentage of Students taking the Spring 2019 LEAP 2025 Administration:  
English I Form E**

Group	Grade							Total
	6	7	8	9	10	11	12	
<b>All Students</b>	0.00	0.00	4.87	89.19	5.26	0.68	0.00	100
<b>Gender</b>								
Female	0.00	0.00	5.53	90.15	3.88	0.45	0.00	100
Male	0.00	0.00	4.17	88.18	6.73	0.92	0.00	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	8.01	78.51	10.50	2.98	0.00	100
American Indian or Alaska Native	0.00	0.00	1.27	92.36	6.37	0.00	0.00	100
Asian	0.00	0.00	20.11	75.93	2.91	1.06	0.00	100
Black or African American	0.00	0.00	3.68	89.01	6.61	0.70	0.00	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	88.89	5.56	5.56	0.00	100
White	0.00	0.00	4.86	91.39	3.43	0.32	0.00	100
Two or More Races	0.00	0.00	8.07	88.26	3.18	0.49	0.00	100
<b>Education Classification</b>								
Regular	0.00	0.00	4.12	89.99	5.21	0.68	0.00	100
Special	0.00	0.00	0.46	81.91	15.81	1.82	0.00	100
Gifted	0.00	0.00	17.27	81.58	1.07	0.07	0.00	100
<b>Economic Status</b>								
Economically Disadvantaged	0.00	0.00	4.10	87.92	7.07	0.90	0.00	100
Not Economically Disadvantaged	0.00	0.00	7.07	90.36	2.36	0.21	0.00	100
<b>English Learner Status</b>								
Non-EL	0.00	0.00	4.97	89.72	4.83	0.48	0.00	100
EL	0.00	0.00	0.95	68.43	22.31	8.32	0.00	100
<b>Migrant Status</b>								
Nonmigrant	0.00	0.00	4.87	89.20	5.25	0.68	0.00	100
Migrant	0.00	0.00	5.00	75.00	20.00	0.00	0.00	100
<b>Section 504 Status</b>								
Non-Section 504	0.00	0.00	5.05	89.39	4.90	0.66	0.00	100
Section 504	0.00	0.00	2.66	86.71	9.77	0.87	0.00	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	4.91	89.16	5.26	0.68	0.00	100
Homeless	0.00	0.00	2.24	91.32	5.60	0.84	0.00	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	4.77	89.23	5.32	0.68	0.00	100
Military Affiliated	0.00	0.00	10.83	86.89	1.71	0.57	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	4.88	89.19	5.25	0.68	0.00	100
Foster Care	0.00	0.00	0.00	90.32	9.68	0.00	0.00	100



Table 7.9 Count of Students taking the Spring 2019 LEAP 2025 Administration: English II Form D

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	<10	≥1,410	≥19,950	≥1,470	<10	≥22,840
<b>Gender</b>								
Female	<10	<10	<10	≥730	≥9,980	≥440	<10	≥11,170
Male	<10	<10	<10	≥680	≥9,960	≥1,020	<10	≥11,670
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	≥100	≥1,030	≥210	<10	≥1,350
American Indian or Alaska Native	<10	<10	<10	<10	≥120	<10	<10	≥130
Asian	<10	<10	<10	≥70	≥310	≥10	<10	≥400
Black or African American	<10	<10	<10	≥610	≥8,600	≥840	<10	≥10,050
Native Hawaiian or Other Pacific	<10	<10	<10	<10	≥10	<10	<10	≥10
White	<10	<10	<10	≥580	≥9,510	≥380	<10	≥10,480
Two or More Races	<10	<10	<10	≥30	≥340	≥10	<10	≥390
<b>Education Classification</b>								
Regular	<10	<10	<10	≥1,110	≥16,440	≥990	<10	≥18,540
Special	<10	<10	<10	≥120	≥2,300	≥470	<10	≥2,890
Gifted	<10	<10	<10	≥170	≥1,210	≥10	<10	≥1,400
<b>Economic Status</b>								
Economically Disadvantaged	<10	<10	<10	≥870	≥11,960	≥1,130	<10	≥13,970
Not Economically Disadvantaged	<10	<10	<10	≥470	≥6,880	≥200	<10	≥7,560
<b>English Learner Status</b>								
Non-EL	<10	<10	<10	≥1,390	≥19,580	≥1,280	<10	≥22,270
EL	<10	<10	<10	≥10	≥360	≥190	<10	≥570
<b>Migrant Status</b>								
Nonmigrant	<10	<10	<10	≥1,410	≥19,940	≥1,470	<10	≥22,830
Migrant	<10	<10	<10	<10	≥10	<10	<10	≥10
<b>Section 504 Status</b>								
Non-Section 504	<10	<10	<10	≥1,300	≥18,130	≥1,240	<10	≥20,670
Section 504	<10	<10	<10	≥110	≥1,820	≥230	<10	≥2,160
<b>Homeless Status</b>								
Not Homeless	<10	<10	<10	≥1,390	≥19,690	≥1,430	<10	≥22,520
Homeless	<10	<10	<10	≥20	≥250	≥40	<10	≥320
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	<10	≥1,390	≥19,730	≥1,470	<10	≥22,600
Military Affiliated	<10	<10	<10	≥20	≥220	<10	<10	≥240
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	<10	≥1,410	≥19,900	≥1,470	<10	≥22,780
Foster Care	<10	<10	<10	<10	≥40	<10	<10	≥50

**Table 7.10 Reportable Percentage of Students taking the Spring 2019 LEAP 2025 Administration: English II Form D**

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.00	0.00	0.00	6.20	87.35	6.46	0.00	100
<b>Gender</b>								
Female	0.00	0.00	0.00	6.55	89.43	4.02	0.00	100
Male	0.00	0.00	0.00	5.86	85.35	8.79	0.00	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.00	7.97	76.38	15.65	0.00	100
American Indian or Alaska Native	0.00	0.00	0.00	2.88	90.65	6.47	0.00	100
Asian	0.00	0.00	0.00	17.82	78.22	3.96	0.00	100
Black or African American	0.00	0.00	0.00	6.07	85.56	8.38	0.00	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	5.88	94.12	0.00	0.00	100
White	0.00	0.00	0.00	5.62	90.76	3.62	0.00	100
Two or More Races	0.00	0.00	0.00	8.18	87.72	4.09	0.00	100
<b>Education Classification</b>								
Regular	0.00	0.00	0.00	6.02	88.64	5.34	0.00	100
Special	0.00	0.00	0.00	4.22	79.54	16.25	0.00	100
Gifted	0.00	0.00	0.00	12.69	86.32	1.00	0.00	100
<b>Economic Status</b>								
Economically Disadvantaged	0.00	0.00	0.00	6.29	85.59	8.12	0.00	100
Not Economically Disadvantaged	0.00	0.00	0.00	6.21	91.06	2.72	0.00	100
<b>English Learner Status</b>								
Non-EL	0.00	0.00	0.00	6.28	87.95	5.77	0.00	100
EL	0.00	0.00	0.00	3.14	63.76	33.10	0.00	100
<b>Migrant Status</b>								
Nonmigrant	0.00	0.00	0.00	6.20	87.35	6.45	0.00	100
Migrant	0.00	0.00	0.00	0.00	83.33	16.67	0.00	100
<b>Section 504 Status</b>								
Non-Section 504	0.00	0.00	0.00	6.30	87.69	6.01	0.00	100
Section 504	0.00	0.00	0.00	5.21	84.08	10.71	0.00	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	0.00	6.19	87.44	6.37	0.00	100
Homeless	0.00	0.00	0.00	6.85	80.37	12.77	0.00	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	0.00	6.18	87.31	6.51	0.00	100
Military Affiliated	0.00	0.00	0.00	8.20	90.57	1.23	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	0.00	6.19	87.36	6.46	0.00	100
Foster Care	0.00	0.00	0.00	10.53	82.46	7.02	0.00	100

Table 7.11 Count of Students taking the Spring 2019 LEAP 2025 Administration: English II Form E

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	<10	≥1,290	≥17,260	≥1,040	<10	≥19,610
<b>Gender</b>								
Female	<10	<10	<10	≥680	≥8,970	≥380	<10	≥10,040
Male	<10	<10	<10	≥610	≥8,290	≥660	<10	≥9,570
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	≥120	≥900	≥150	<10	≥1,180
American Indian or Alaska Native	<10	<10	<10	<10	≥120	<10	<10	≥130
Asian	<10	<10	<10	≥70	≥280	≥10	<10	≥370
Black or African American	<10	<10	<10	≥460	≥7,050	≥530	<10	≥8,050
Native Hawaiian or Other Pacific	<10	<10	<10	<10	≥10	<10	<10	≥20
White	<10	<10	<10	≥580	≥8,550	≥310	<10	≥9,460
Two or More Races	<10	<10	<10	≥30	≥310	≥20	<10	≥370
<b>Education Classification</b>								
Regular	<10	<10	<10	≥1,040	≥15,640	≥950	<10	≥17,640
Special	<10	<10	<10	≥30	≥450	≥80	<10	≥560
Gifted	<10	<10	<10	≥210	≥1,170	<10	<10	≥1,400
<b>Economic Status</b>								
Economically Disadvantaged	<10	<10	<10	≥750	≥9,850	≥730	<10	≥11,340
Not Economically Disadvantaged	<10	<10	<10	≥470	≥6,410	≥150	<10	≥7,040
<b>English Learner Status</b>								
Non-EL	<10	<10	<10	≥1,270	≥17,000	≥900	<10	≥19,180
EL	<10	<10	<10	≥10	≥260	≥140	<10	≥420
<b>Migrant Status</b>								
Nonmigrant	<10	<10	<10	≥1,290	≥17,250	≥1,040	<10	≥19,590
Migrant	<10	<10	<10	<10	≥10	<10	<10	≥10
<b>Section 504 Status</b>								
Non-Section 504	<10	<10	<10	≥1,210	≥16,150	≥910	<10	≥18,280
Section 504	<10	<10	<10	≥80	≥1,110	≥130	<10	≥1,320
<b>Homeless Status</b>								
Not Homeless	<10	<10	<10	≥1,270	≥17,030	≥1,020	<10	≥19,340
Homeless	<10	<10	<10	≥20	≥230	≥10	<10	≥270
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	<10	≥1,260	≥17,000	≥1,030	<10	≥19,310
Military Affiliated	<10	<10	<10	≥20	≥260	<10	<10	≥290
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	<10	≥1,290	≥17,230	≥1,040	<10	≥19,560
Foster Care	<10	<10	<10	<10	≥30	<10	<10	≥40

**Table 7.12 Reportable Percentage of Students taking the Spring 2019 LEAP 2025 Administration: English II Form E**

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.00	0.00	0.00	6.61	88.05	5.33	0.01	100
<b>Gender</b>								
Female	0.00	0.00	0.00	6.82	89.37	3.80	0.00	100
Male	0.00	0.00	0.00	6.38	86.66	6.94	0.02	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.00	10.75	76.38	12.87	0.00	100
American Indian or Alaska Native	0.00	0.00	0.00	5.04	92.09	2.88	0.00	100
Asian	0.00	0.00	0.00	19.41	76.86	3.72	0.00	100
Black or African American	0.00	0.00	0.00	5.82	87.58	6.59	0.01	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	0.00	80.00	20.00	0.00	100
White	0.00	0.00	0.00	6.15	90.47	3.37	0.01	100
Two or More Races	0.00	0.00	0.00	10.08	84.08	5.84	0.00	100
<b>Education Classification</b>								
Regular	0.00	0.00	0.00	5.92	88.65	5.42	0.01	100
Special	0.00	0.00	0.00	6.01	79.51	14.49	0.00	100
Gifted	0.00	0.00	0.00	15.53	83.90	0.57	0.00	100
<b>Economic Status</b>								
Economically Disadvantaged	0.00	0.00	0.00	6.63	86.87	6.49	0.01	100
Not Economically Disadvantaged	0.00	0.00	0.00	6.72	91.05	2.23	0.00	100
<b>English Learner Status</b>								
Non-EL	0.00	0.00	0.00	6.66	88.64	4.69	0.01	100
EL	0.00	0.00	0.00	4.21	61.68	34.11	0.00	100
<b>Migrant Status</b>								
Nonmigrant	0.00	0.00	0.00	6.61	88.04	5.33	0.01	100
Migrant	0.00	0.00	0.00	0.00	93.33	6.67	0.00	100
<b>Section 504 Status</b>								
Non-Section 504	0.00	0.00	0.00	6.64	88.33	5.01	0.01	100
Section 504	0.00	0.00	0.00	6.10	84.10	9.80	0.00	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	0.00	6.59	88.08	5.32	0.01	100
Homeless	0.00	0.00	0.00	7.72	86.03	6.25	0.00	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	0.00	6.56	88.05	5.37	0.01	100
Military Affiliated	0.00	0.00	0.00	9.46	87.84	2.70	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	0.00	6.59	88.06	5.34	0.01	100
Foster Care	0.00	0.00	0.00	13.04	82.61	4.35	0.00	100

Table 7.13 Count of Students taking the Spring 2019 LEAP 2025 Administration: Algebra I Form D

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	≥100	≥3,880	≥18,600	≥3,650	≥590	<10	≥26,840
<b>Gender</b>								
Female	<10	≥40	≥2,090	≥9,200	≥1,530	≥230	<10	≥13,100
Male	<10	≥60	≥1,790	≥9,400	≥2,120	≥350	<10	≥13,740
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	≥220	≥1,280	≥310	≥80	<10	≥1,920
American Indian or Alaska Native	<10	<10	≥20	≥120	≥30	<10	<10	≥180
Asian	<10	≥20	≥140	≥220	≥10	<10	<10	≥410
Black or African American	<10	≥20	≥1,070	≥8,380	≥2,090	≥330	<10	≥11,920
Native Hawaiian or Other Pacific	<10	<10	<10	≥10	<10	<10	<10	≥10
White	<10	≥40	≥2,320	≥8,190	≥1,140	≥140	<10	≥11,850
Two or More Races	<10	<10	≥80	≥370	≥50	≥10	<10	≥530
<b>Education Classification</b>								
Regular	<10	≥60	≥3,010	≥15,540	≥2,510	≥370	<10	≥21,520
Special	<10	<10	≥50	≥2,390	≥1,090	≥210	<10	≥3,760
Gifted	<10	≥30	≥810	≥650	≥40	<10	<10	≥1,560
<b>Economic Status</b>								
Economically Disadvantaged	<10	≥30	≥1,560	≥11,840	≥2,900	≥450	<10	≥16,800
Not Economically Disadvantaged	<10	≥60	≥1,950	≥5,030	≥580	≥70	<10	≥7,710
<b>English Learner Status</b>								
Non-EL	<10	≥100	≥3,850	≥17,960	≥3,420	≥510	<10	≥25,860
EL	<10	<10	≥20	≥630	≥230	≥80	<10	≥970
<b>Migrant Status</b>								
Nonmigrant	<10	≥100	≥3,880	≥18,580	≥3,640	≥590	<10	≥26,810
Migrant	<10	<10	<10	≥10	<10	<10	<10	≥20
<b>Section 504 Status</b>								
Non-Section 504	<107	≥100	≥3,700	≥16,670	≥3,090	≥510	<10	≥24,100
Section 504	<10	<10	≥170	≥1,920	≥550	≥70	<10	≥2,740
<b>Homeless Status</b>								
Not Homeless	<10	≥100	≥3,850	≥18,280	≥3,590	≥570	<10	≥26,410
Homeless	<10	<10	≥20	≥320	≥60	≥10	<10	≥430
<b>Military Affiliation</b>								
Not Military Affiliated	<10	≥100	≥3,770	≥18,350	≥3,620	≥580	<10	≥26,450
Military Affiliated	<10	<10	≥100	≥240	≥30	<10	<10	≥380
<b>Foster Care Status</b>								
Not in Foster Care	<10	≥100	≥3,870	≥18,540	≥3,630	≥580	<10	≥26,750
Foster Care	<10	<10	<10	≥60	≥10	<10	<10	≥80

**Table 7.74 Reportable Percentage of Students taking the Spring 2019 LEAP 2025 Administration: Algebra I Form D**

Group	Grade							Total
	6	7	8	9	10	11	12	
<b>All Students</b>	0.03	0.39	14.46	69.30	13.62	2.20	0.00	100
<b>Gender</b>								
Female	0.02	0.33	15.95	70.23	11.68	1.80	0.00	100
Male	0.04	0.45	13.04	68.43	15.47	2.58	0.00	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.42	11.72	67.03	16.20	4.64	0.00	100
American Indian or Alaska Native	0.00	0.00	11.67	66.67	20.00	1.67	0.00	100
Asian	1.44	4.81	34.38	54.57	3.61	1.20	0.00	100
Black or African American	0.00	0.24	9.03	70.33	17.56	2.84	0.00	100
Native Hawaiian or Other Pacific	0.00	0.00	15.79	63.16	15.79	5.26	0.00	100
White	0.01	0.37	19.63	69.17	9.62	1.21	0.00	100
Two or More Races	0.00	0.75	16.32	70.17	10.88	1.88	0.00	100
<b>Education Classification</b>								
Regular	0.01	0.30	14.02	72.23	11.69	1.74	0.00	100
Special	0.00	0.05	1.36	63.77	29.21	5.61	0.00	100
Gifted	0.26	2.44	52.12	42.24	2.63	0.32	0.00	100
<b>Economic Status</b>								
Economically Disadvantaged	0.00	0.18	9.34	70.47	17.29	2.72	0.00	100
Not Economically Disadvantaged	0.00	0.87	25.33	65.31	7.53	0.96	0.00	100
<b>English Learner Status</b>								
Non-EL	0.03	0.40	14.90	69.47	13.23	1.97	0.00	100
EL	0.00	0.20	2.76	64.99	23.85	8.19	0.00	100
<b>Migrant Status</b>								
Nonmigrant	0.03	0.39	14.47	69.31	13.60	2.20	0.00	100
Migrant	0.00	0.00	7.14	60.71	32.14	0.00	0.00	100
<b>Section 504 Status</b>								
Non-Section 504	0.03	0.42	15.38	69.19	12.85	2.14	0.00	100
Section 504	0.00	0.15	6.42	70.29	20.40	2.74	0.00	100
<b>Homeless Status</b>								
Not Homeless	0.03	0.39	14.61	69.21	13.60	2.17	0.00	100
Homeless	0.00	0.23	5.34	75.17	15.08	4.18	0.00	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.03	0.39	14.28	69.39	13.70	2.21	0.00	100
Military Affiliated	0.00	0.77	26.80	63.14	7.99	1.29	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.03	0.39	14.49	69.30	13.60	2.20	0.00	100
Foster Care	0.00	0.00	5.68	71.59	20.45	2.27	0.00	100

Table 7.15 Count of Students taking the Spring 2019 LEAP 2025 Administration: Algebra I Form E

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	≥100	≥3,810	≥15,280	≥2,120	≥300	<10	≥21,640
<b>Gender</b>								
Female	<10	≥50	≥2,030	≥7,900	≥910	≥130	<10	≥11,040
Male	<10	≥50	≥1,780	≥7,380	≥1,200	≥160	<10	≥10,590
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	≥220	≥840	≥120	≥30	<10	≥1,230
American Indian or Alaska Native	<10	<10	≥10	≥100	≥20	<10	<10	≥140
Asian	<10	≥20	≥150	≥190	<10	<10	<10	≥380
Black or African American	<10	≥30	≥1,020	≥6,660	≥1,210	≥160	<10	≥9,090
Native Hawaiian or Other Pacific	<10	<10	<10	<10	<10	<10	<10	≥10
White	<10	≥40	≥2,320	≥7,180	≥690	≥90	<10	≥10,330
Two or More Races	<10	<10	≥80	≥290	≥50	<10	<10	≥430
<b>Education Classification</b>								
Regular	<10	≥60	≥2,980	≥14,260	≥1,990	≥280	<10	≥19,590
Special	<10	<10	≥20	≥350	≥100	≥10	<10	≥500
Gifted	<10	≥40	≥800	≥660	≥20	<10	<10	≥1,540
<b>Economic Status</b>								
Economically Disadvantaged	<10	≥30	≥1,510	≥9,330	≥1,650	≥220	<10	≥12,750
Not Economically Disadvantaged	<10	≥70	≥1,950	≥4,500	≥360	≥50	<10	≥6,950
<b>English Learner Status</b>								
Non-EL	<10	≥100	≥3,800	≥15,030	≥2,040	≥280	<10	≥21,270
EL	<10	<10	≥10	≥250	≥70	≥10	<10	≥360
<b>Migrant Status</b>								
Nonmigrant	<10	≥100	≥3,810	≥15,270	≥2,110	≥290	<10	≥21,620
Migrant	<10	<10	<10	≥10	<10	<10	<10	≥10
<b>Section 504 Status</b>								
Non-Section 504	<10	≥100	≥3,670	≥14,170	≥1,840	≥250	<10	≥20,070
Section 504	<10	<10	≥140	≥1,110	≥270	≥40	<10	≥1,570
<b>Homeless Status</b>								
Not Homeless	<10	≥100	≥3,790	≥15,000	≥2,070	≥290	<10	≥21,290
Homeless	<10	<10	≥20	≥280	≥40	<10	<10	≥350
<b>Military Affiliation</b>								
Not Military Affiliated	<10	≥100	≥3,700	≥15,070	≥2,100	≥300	<10	≥21,300
Military Affiliated	<10	<10	≥110	≥210	≥10	<10	<10	≥340
<b>Foster Care Status</b>								
Not in Foster Care	<10	≥100	≥3,810	≥15,240	≥2,100	≥300	<10	≥21,580
Foster Care	<10	<10	<10	≥30	≥10	<10	<10	≥50

**Table 7.16 Reportable Percentage of Students taking the Spring 2019 LEAP 2025 Administration: Algebra I Form E**

Group	Grade							Total
	6	7	8	9	10	11	12	
<b>All Students</b>	0.03	0.50	17.64	70.63	9.80	1.39	0.00	100
<b>Gender</b>								
Female	0.01	0.52	18.44	71.52	8.31	1.21	0.00	100
Male	0.06	0.49	16.82	69.71	11.35	1.58	0.00	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.65	18.30	68.10	10.36	2.59	0.00	100
American Indian or Alaska Native	0.00	0.00	8.33	71.53	18.06	2.08	0.00	100
Asian	1.56	5.47	39.06	51.04	2.34	0.52	0.00	100
Black or African American	0.00	0.37	11.24	73.25	13.38	1.76	0.00	100
Native Hawaiian or Other Pacific	0.00	0.00	20.00	80.00	0.00	0.00	0.00	100
White	0.01	0.40	22.49	69.48	6.68	0.95	0.00	100
Two or More Races	0.00	1.14	18.72	67.12	11.64	1.37	0.00	100
<b>Education Classification</b>								
Regular	0.01	0.33	15.25	72.81	10.17	1.44	0.00	100
Special	0.20	0.79	5.12	69.49	21.06	3.35	0.00	100
Gifted	0.32	2.60	52.24	43.28	1.43	0.13	0.00	100
<b>Economic Status</b>								
Economically Disadvantaged	0.00	0.24	11.87	73.17	12.98	1.74	0.00	100
Not Economically Disadvantaged	0.00	1.02	28.15	64.83	5.25	0.75	0.00	100
<b>English Learner Status</b>								
Non-EL	0.03	0.50	17.88	70.65	9.60	1.33	0.00	100
EL	0.00	0.54	4.07	69.38	21.14	4.88	0.00	100
<b>Migrant Status</b>								
Nonmigrant	0.03	0.50	17.66	70.63	9.80	1.38	0.00	100
Migrant	0.00	0.00	0.00	75.00	12.50	12.50	0.00	100
<b>Section 504 Status</b>								
Non-Section 504	0.03	0.53	18.32	70.61	9.21	1.29	0.00	100
Section 504	0.00	0.13	9.03	70.82	17.29	2.73	0.00	100
<b>Homeless Status</b>								
Not Homeless	0.03	0.51	17.83	70.48	9.75	1.39	0.00	100
Homeless	0.00	0.00	6.55	79.77	12.54	1.14	0.00	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.03	0.51	17.40	70.76	9.89	1.41	0.00	100
Military Affiliated	0.00	0.29	32.56	62.50	4.36	0.29	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.03	0.50	17.68	70.63	9.76	1.39	0.00	100
Foster Care	0.00	0.00	3.64	69.09	25.45	1.82	0.00	100



Table 7.17 Count of Students taking the Spring 2019 LEAP 2025 Administration: Geometry Form D

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	≥110	≥3,260	≥13,010	≥2,420	<10	≥18,820
<b>Gender</b>								
Female	<10	<10	≥50	≥1,790	≥6,840	≥1,140	<10	≥9,830
Male	<10	<10	≥50	≥1,470	≥6,170	≥1,270	<10	≥8,980
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	≥10	≥210	≥790	≥210	<10	≥1,220
American Indian or Alaska Native	<10	<10	<10	≥20	≥60	<10	<10	≥90
Asian	<10	<10	≥10	≥130	≥170	≥30	<10	≥360
Black or African American	<10	<10	≥20	≥840	≥5,580	≥1,430	<10	≥7,890
Native Hawaiian or Other Pacific	<10	<10	<10	<10	≥10	<10	<10	≥20
White	<10	<10	≥60	≥1,970	≥6,130	≥690	<10	≥8,860
Two or More Races	<10	<10	<10	≥60	≥230	≥30	<10	≥330
<b>Education Classification</b>								
Regular	<10	<10	≥50	≥2,570	≥11,310	≥1,970	<10	≥15,910
Special	<10	<10	<10	≥70	≥1,090	≥370	<10	≥1,530
Gifted	<10	<10	≥50	≥620	≥610	≥70	<10	≥1,360
<b>Economic Status</b>								
Economically Disadvantaged	<10	<10	≥20	≥1,310	≥7,760	≥1,750	<10	≥10,860
Not Economically Disadvantaged	<10	<10	≥80	≥1,720	≥4,630	≥580	<10	≥7,020
<b>English Learner Status</b>								
Non-EL	<10	<10	≥110	≥3,240	≥12,730	≥2,260	<10	≥18,360
EL	<10	<10	<10	≥20	≥270	≥150	<10	≥450
<b>Migrant Status</b>								
Nonmigrant	<10	<10	≥110	≥3,260	≥12,990	≥2,410	<10	≥18,790
Migrant	<10	<10	<10	<10	≥10	<10	<10	≥20
<b>Section 504 Status</b>								
Non-Section 504	<10	<10	≥110	≥3,120	≥11,870	≥2,190	<10	≥17,300
Section 504	<10	<10	<10	≥140	≥1,140	≥220	<10	≥1,510
<b>Homeless Status</b>								
Not Homeless	<10	<10	≥110	≥3,240	≥12,830	≥2,370	<10	≥18,560
Homeless	<10	<10	<10	≥20	≥180	≥40	<10	≥250
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	≥100	≥3,200	≥12,850	≥2,390	<10	≥18,560
Military Affiliated	<10	<10	<10	≥60	≥160	≥20	<10	≥250
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	≥110	≥3,260	≥12,980	≥2,410	<10	≥18,780
Foster Care	<10	<10	<10	<10	≥30	<10	<10	≥30

Table 7.18 Percentage of Students taking the Spring 2019 LEAP 2025 Administration: Geometry Form D

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.01	0.02	0.60	17.35	69.16	12.86	0.01	100
<b>Gender</b>								
Female	0.00	0.01	0.59	18.22	69.53	11.65	0.01	100
Male	0.01	0.02	0.61	16.40	68.75	14.19	0.01	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.90	17.49	64.52	17.09	0.00	100
American Indian or Alaska Native	0.00	0.00	0.00	20.83	70.83	8.33	0.00	100
Asian	0.27	0.82	4.62	37.23	48.37	8.70	0.00	100
Black or African American	0.00	0.00	0.29	10.76	70.81	18.13	0.01	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	18.52	59.26	22.22	0.00	100
White	0.00	0.00	0.69	22.23	69.20	7.87	0.01	100
Two or More Races	0.00	0.00	0.29	19.76	69.32	10.62	0.00	100
<b>Education Classification</b>								
Regular	0.00	0.01	0.35	16.16	71.09	12.38	0.01	100
Special	0.00	0.00	0.06	4.68	70.89	24.37	0.00	100
Gifted	0.07	0.15	4.17	45.39	44.66	5.56	0.00	100
<b>Economic Status</b>								
Economically Disadvantaged	0.00	0.00	0.24	12.11	71.51	16.13	0.02	100
Not Economically Disadvantaged	0.00	0.04	1.20	24.51	65.90	8.35	0.00	100
<b>English Learner Status</b>								
Non-EL	0.01	0.02	0.62	17.66	69.34	12.35	0.01	100
EL	0.00	0.00	0.00	4.86	61.59	33.55	0.00	100
<b>Migrant Status</b>								
Nonmigrant	0.01	0.02	0.60	17.35	69.15	12.87	0.01	100
Migrant	0.00	0.00	0.00	13.04	78.26	8.70	0.00	100
<b>Section 504 Status</b>								
Non-Section 504	0.01	0.02	0.64	18.03	68.61	12.70	0.01	100
Section 504	0.00	0.00	0.20	9.59	75.40	14.75	0.07	100
<b>Homeless Status</b>								
Not Homeless	0.01	0.02	0.61	17.46	69.11	12.79	0.01	100
Homeless	0.00	0.00	0.00	9.09	72.33	18.18	0.40	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.01	0.02	0.58	17.24	69.24	12.90	0.01	100
Military Affiliated	0.00	0.00	1.94	24.81	63.18	10.08	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.01	0.02	0.60	17.37	69.14	12.86	0.01	100
Foster Care	0.00	0.00	0.00	7.69	76.92	15.38	0.00	100

Table 7.19 Count of Students taking the Spring 2019 LEAP 2025 Administration: Geometry Form E

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	≥110	≥3,190	≥11,550	≥1,850	<10	≥16,710
<b>Gender</b>								
Female	<10	<10	≥60	≥1,780	≥6,150	≥930	<10	≥8,930
Male	<10	<10	≥50	≥1,410	≥5,390	≥910	<10	≥7,770
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	≥160	≥590	≥110	<10	≥880
American Indian or Alaska Native	<10	<10	<10	≥10	≥70	<10	<10	≥90
Asian	<10	<10	≥10	≥150	≥160	≥20	<10	≥360
Black or African American	<10	<10	≥20	≥850	≥4,700	≥1,100	<10	≥6,680
Native Hawaiian or Other Pacific	<10	<10	<10	<10	<10	<10	<10	≥10
White	<10	<10	≥60	≥1,920	≥5,780	≥570	<10	≥8,340
Two or More Races	<10	<10	<10	≥70	≥220	≥20	<10	≥320
<b>Education Classification</b>								
Regular	<10	<10	≥60	≥2,500	≥10,800	≥1,740	<10	≥15,110
Special	<10	<10	<10	≥20	≥200	≥40	<10	≥270
Gifted	<10	<10	≥50	≥660	≥530	≥60	<10	≥1,320
<b>Economic Status</b>								
Economically Disadvantaged	<10	<10	≥10	≥1,270	≥6,660	≥1,280	<10	≥9,230
Not Economically Disadvantaged	<10	<10	≥90	≥1,700	≥4,310	≥490	<10	≥6,600
<b>English Learner Status</b>								
Non-EL	<10	<10	≥110	≥3,170	≥11,430	≥1,780	<10	≥16,510
EL	<10	<10	<10	≥10	≥120	≥60	<10	≥190
<b>Migrant Status</b>								
Nonmigrant	<10	<10	≥110	≥3,190	≥11,540	≥1,850	<10	≥16,710
Migrant	<10	<10	<10	<10	<10	<10	<10	<10
<b>Section 504 Status</b>								
Non-Section 504	<10	<10	≥100	≥3,060	≥10,890	≥1,730	<10	≥15,810
Section 504	<10	<10	<10	≥120	≥650	≥110	<10	≥900
<b>Homeless Status</b>								
Not Homeless	<10	<10	≥110	≥3,160	≥11,410	≥1,810	<10	≥16,510
Homeless	<10	<10	<10	≥20	≥140	≥30	<10	≥200
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	≥100	≥3,120	≥11,370	≥1,830	<10	≥16,440
Military Affiliated	<10	<10	<10	≥60	≥170	≥10	<10	≥260
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	≥110	≥3,180	≥11,530	≥1,840	<10	≥16,690
Foster Care	<10	<10	<10	<10	≥10	<10	<10	≥20

Table 7.20 Percentage of Students taking the Spring Administration: Geometry Form E

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.01	0.02	0.68	19.10	69.10	11.09	0.00	100
<b>Gender</b>								
Female	0.00	0.03	0.68	19.91	68.88	10.49	0.00	100
Male	0.03	0.01	0.67	18.17	69.36	11.77	0.00	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.68	18.98	67.27	13.07	0.00	100
American Indian or Alaska Native	0.00	0.00	0.00	19.59	73.20	7.22	0.00	100
Asian	0.55	0.83	4.97	43.37	44.48	5.80	0.00	100
Black or African American	0.00	0.01	0.30	12.71	70.41	16.57	0.00	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	26.67	53.33	20.00	0.00	100
White	0.00	0.00	0.80	23.03	69.28	6.89	0.00	100
Two or More Races	0.00	0.00	0.61	22.49	69.60	7.29	0.00	100
<b>Education Classification</b>								
Regular	0.00	0.02	0.40	16.56	71.51	11.51	0.00	100
Special	0.00	0.00	0.36	8.00	75.27	16.36	0.00	100
Gifted	0.15	0.08	3.92	50.34	40.39	5.12	0.00	100
<b>Economic Status</b>								
Economically Disadvantaged	0.00	0.00	0.19	13.75	72.12	13.94	0.00	100
Not Economically Disadvantaged	0.00	0.06	1.36	25.83	65.31	7.43	0.00	100
<b>English Learner Status</b>								
Non-EL	0.01	0.02	0.68	19.25	69.20	10.83	0.00	100
EL	0.00	0.00	0.00	7.04	60.80	32.16	0.00	100
<b>Migrant Status</b>								
Nonmigrant	0.01	0.02	0.68	19.10	69.10	11.09	0.00	100
Migrant	0.00	0.00	0.00	16.67	83.33	0.00	0.00	100
<b>Section 504 Status</b>								
Non-Section 504	0.01	0.03	0.69	19.38	68.91	10.98	0.00	100
Section 504	0.00	0.00	0.44	14.21	72.48	12.87	0.00	100
<b>Homeless Status</b>								
Not Homeless	0.01	0.02	0.68	19.18	69.10	11.00	0.00	100
Homeless	0.00	0.00	0.00	12.81	69.46	17.73	0.00	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.01	0.02	0.66	19.00	69.14	11.16	0.00	100
Military Affiliated	0.00	0.00	1.50	25.09	66.67	6.74	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.01	0.02	0.68	19.10	69.11	11.07	0.00	100
Foster Care	0.00	0.00	0.00	17.39	60.87	21.74	0.00	100

Tables 7.21 through 7.24 summarize the mean scale scores, standard deviations, and the percentage of students in each achievement level for the 2018–2019 administration of the LEAP 2025 high school ELA and mathematics assessments. All three administrations are presented. Results from previous years are shown for comparison purposes.

**Table 7.21 Comparison of Percentage of Students in Each Achievement Level: English I**

	Year	Administration	Form	N	Scale Score		Percentage in Achievement Level				
					Mean	SD	1	2	3	4	5
All	2018	Fall	B	≥6,680	731.26	39.90	23.8	20.9	21.2	27.3	6.9
	2019	Spring	D	≥25,850	737.41	36.91	16.4	19.3	25.7	31.2	7.3
			E	≥21,390	747.16	33.38	7.4	17.4	27.9	38.4	8.8
			A*	≥80	714.24	30.15	38.6	21.7	24.1	15.7	0.0
2019	Summer	A	≥1,900	699.19	20.23	53.2	36.5	9.1	1.2	0.0	
First-Time Testers	2018	Fall	B	≥4,420	749.07	34.02	7.5	15.3	26.8	40.0	10.4
	2019	Spring	D	≥24,590	739.31	36.14	14.5	18.9	26.4	32.6	7.6
			E	≥20,530	748.58	32.55	6.2	16.8	28.3	39.6	9.1
			A*	≥50	720.14	29.66	32.2	18.6	28.8	20.3	0.0
2019	Summer	A	≥70	714.86	24.43	32.9	32.9	26.6	7.6	0.0	
Retesters	2018	Fall	B	≥1,910	691.99	21.90	62.4	30.7	6.4	0.5	0.0
	2019	Spring	D	≥700	684.00	20.16	77.4	20.7	1.8	0.0	0.0
			E	≥370	692.33	21.44	62.4	31.7	4.8	1.1	0.0
			A*	≥10	696.89	25.69	63.2	21.1	15.8	0.0	0.0
2019	Summer	A	≥1,760	697.78	19.12	55.1	36.8	7.6	0.5	0.0	
Previously Passed	2018	Fall	B	≥340	720.56	24.81	18.1	38.1	31.5	12.0	0.3
	2019	Spring	D	≥550	721.27	32.17	23.7	35.0	26.2	11.5	3.6
			E	≥470	729.70	34.60	15.9	30.8	30.3	17.2	5.9
			A*	<10	NR	NR	NR	NR	NR	NR	NR
2019	Summer	A	≥60	719.77	25.68	25.0	31.7	30.0	13.3	0.0	

Levels: 1 = *Unsatisfactory*, 2 = *Approaching Basic*, 3 = *Basic*, 4 = *Mastery*, 5 = *Advanced*

\* Senior form

**Table 7.22 Comparison of Percentage of Students in Each Achievement Level: English II**

	Year	Administration	Form	N	Scale Score		Percentage in Achievement Level				
					Mean	SD	1	2	3	4	5
All	2018	Fall	B	≥9,590	723.34	46.39	34.9	21.0	17.0	19.0	8.1
	2019	Spring	D	≥22,840	738.08	45.64	20.9	17.0	19.2	31.9	11.1
			E	≥19,610	748.12	41.91	13.1	15.4	22.7	34.1	14.7
			A*	≥890	687.50	25.58	73.7	19.2	4.8	2.1	0.1
2019	Summer	B	≥1,690	688.72	25.64	69.0	23.0	6.7	1.2	0.2	
First-Time Testers	2018	Fall	B	≥5,470	748.34	42.70	12.5	17.0	23.9	32.4	14.2
	2019	Spring	D	≥21,550	741.35	44.28	17.7	17.1	19.9	33.6	11.7
			E	≥18,720	750.77	40.45	10.7	15.1	23.4	35.5	15.3
			A*	≥90	703.64	33.59	47.9	27.7	16.0	8.5	0.0
2019	Summer	B	≥70	709.51	42.01	49.4	18.2	14.3	15.6	2.6	
Retesters	2018	Fall	B	≥3,900	688.63	24.51	67.0	25.8	6.7	0.5	0.0
	2019	Spring	D	≥970	674.06	22.27	85.7	11.7	2.2	0.3	0.0
			E	≥640	681.84	23.97	77.3	19.0	2.8	0.6	0.3
			A*	≥710	682.48	20.48	81.8	15.8	2.2	0.1	0.0
2019	Summer	B	≥1,600	687.56	24.16	70.1	23.1	6.3	0.5	0.1	
Previously Passed	2018	Fall	B	≥210	716.80	29.55	24.7	37.2	25.6	12.1	0.5
	2019	Spring	D	≥300	712.88	35.94	37.2	24.3	24.3	12.6	1.6
			E	≥240	720.69	34.28	28.3	28.3	24.2	16.8	2.5
			A*	≥80	710.82	32.33	36.0	38.2	13.5	11.2	1.1
2019	Summer	B	≥10	702.83	21.29	50.0	38.9	11.1	0.0	0.0	

Levels: 1 = *Unsatisfactory*, 2 = *Approaching Basic*, 3 = *Basic*, 4 = *Mastery*, 5 = *Advanced*

\* Senior form

**Table 7.23 Comparison of Percentage of Students in Each Achievement Level: Algebra I**

	Year	Administration	Form	N	Scale Score		Percentage in Achievement Level				
					Mean	SD	1	2	3	4	5
All	2018	Fall	B	≥5,670	724.88	32.88	20.8	36.8	19.7	21.1	1.6
	2019	Spring	D	≥26,940	738.04	34.94	12.4	24.8	28.2	30.5	4.1
			E	≥21,640	744.22	34.29	9.1	21.5	28.8	35.8	4.9
			AR*	≥260	711.23	23.51	29.5	48.1	14.0	8.0	0.4
2019	Summer	BR	≥1,950	708.44	20.36	32.8	46.2	18.3	2.7	0.0	
First-Time Testers	2018	Fall	B	≥3,130	740.33	32.73	9.4	25.3	26.2	36.3	2.9
	2019	Spring	D	≥25,270	739.47	34.84	11.4	23.9	28.7	31.7	4.3
			E	≥20,570	745.42	34.11	8.3	20.6	29.1	36.9	5.1
			AR*	≥70	725.11	29.60	19.2	32.9	21.9	24.7	1.4
2019	Summer	BR	≥70	722.90	25.20	19.0	32.9	30.4	17.7	0.0	
Retesters	2018	Fall	B	≥2,030	702.44	18.70	39.2	52.8	7.5	0.5	0.0
	2019	Spring	D	≥650	702.24	18.55	41.2	47.1	11.2	0.5	0.0
			E	≥270	704.62	18.05	39.1	49.3	10.5	1.1	0.0
			AR*	≥120	703.80	17.43	36.7	53.9	7.8	1.6	0.0
2019	Summer	BR	≥1,800	707.37	19.45	33.9	47.0	17.7	1.4	0.0	
Previously Passed	2018	Fall	B	≥510	719.58	23.71	17.2	43.9	28.5	10.4	0.0
	2019	Spring	D	≥1,020	725.31	30.45	19.0	34.0	27.2	18.3	1.5
			E	≥790	726.80	30.04	17.8	35.2	27.7	18.3	1.0
			AR*	≥60	710.27	19.04	27.0	54.0	17.5	1.6	0.0
2019	Summer	BR	≥70	719.17	26.96	20.8	41.7	18.1	19.4	0.0	

Levels: 1 = *Unsatisfactory*, 2 = *Approaching Basic*, 3 = *Basic*, 4 = *Mastery*, 5 = *Advanced*

\* Senior form

**Table 7.24 Comparison of Percentage of Students in Each Achievement Level: Geometry**

	Year	Administration	Form	N	Scale Score		Percentage in Achievement Level				
					Mean	SD	1	2	3	4	5
All	2018	Fall	B	≥5,350	733.72	28.31	11.8	29.4	28.5	26.5	3.9
	2019	Spring	D	≥18,870	737.04	26.74	5.2	29.6	33.7	26.2	5.3
			E	≥16,710	739.61	26.00	3.8	25.7	34.8	30.8	5.0
			AR*	≥400	710.38	20.62	29.7	49.4	16.2	4.2	0.5
2019	Summer	BR	≥270	710.67	28.59	33.5	51.6	7.6	2.2	5.1	
First-Time Testers	2018	Fall	B	≥4,710	737.54	27.26	8.4	26.4	31.0	29.8	4.4
	2019	Spring	D	≥18,490	737.41	26.71	5.0	29.2	33.8	26.6	5.4
			E	≥16,450	739.88	25.98	3.7	25.3	34.8	31.1	5.1
			AR*	≥270	711.90	21.36	27.6	51.1	15.1	5.5	0.7
2019	Summer	BR	≥70	729.61	44.24	24.3	40.0	8.6	7.1	20.0	
Retesters	2018	Fall	B	≥460	700.91	15.30	44.3	51.4	4.1	0.2	0.0
	2019	Spring	D	≥130	707.27	16.01	24.1	65.0	10.9	0.0	0.0
			E	≥70	710.87	15.47	17.3	65.3	14.7	2.7	0.0
			AR*	≥70	698.63	14.59	51.9	44.3	3.8	0.0	0.0
2019	Summer	BR	≥180	703.67	16.41	38.2	54.3	7.0	0.5	0.0	
Previously Passed	2018	Fall	B	≥180	718.47	20.45	17.4	49.5	24.5	8.7	0.0
	2019	Spring	D	≥240	725.51	22.13	9.0	43.0	32.4	15.2	0.4
			E	≥180	727.62	21.53	7.4	37.2	39.9	13.8	1.6
			AR*	≥50	720.66	16.31	6.0	48.0	42.0	4.0	0.0
2019	Summer	BR	≥10	709.32	16.44	21.1	68.4	10.5	0.0	0.0	

Levels: 1 = *Unsatisfactory*, 2 = *Approaching Basic*, 3 = *Basic*, 4 = *Mastery*, 5 = *Advanced*

\* Senior form

## 7.2 Reports

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

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

Standard 5.1 is related to Standard 6.10. It states:

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

Interpretations of test scores from each administration are disseminated in two ways: the individual score report and the *LEAP Interpretive Guide*.



In addition to providing interpretations of test results, the LDOE and DRC must ensure that those interpretations are understandable for the target audience. Standard 7.0 states:

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

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

The Individual Student-Level Report (ISR) is the primary means for sharing student test results with parents. As such, it is a stand-alone document from which parents can glean information that is relevant to understanding their children’s test scores. For more information about the test, parents are provided [Parent Guide to the LEAP 2025 Student Reports](#). In the 2018–2019 administration year, student reports for each school were posted by subject, then downloaded and printed from eDIRECT by the school systems and schools. eDIRECT is DRC’s secure online system that provides schools and districts access to student tests and reports.

### 7.3 Description of Each Type of Report

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

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

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

#### School Roster Report

A School Roster Report, which provides summary information about student performance on the LEAP 2025 high school ELA and mathematics assessments, is available to school systems and schools through eDIRECT. Total test scores and achievement level indicators are shown for the test of interest. Category and subcategory performance ratings are also reported for students. At the school level, the percentage of students at each achievement level and rating by category and subcategory are summarized. More details can be found in the [LEAP Interpretive Guide](#).

#### 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 parents. At the top of the page, overall student performance is reported by scale score and achievement level. In the middle of the page, category and subcategory performance indicators are reported. 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 will display a blank scale score for a given course.

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](#) was written to help Louisiana school system and school administrators, teachers, parents, and the general public 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.

## 7.4 Summary

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

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

*Standard 6.10* When test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the test covers, what scores represent, the precision/reliability of the scores, and how scores are intended to be used. (119)

*Standard 7.0* Information relating to tests should be clearly documented so that those who use tests can make informed decisions regarding which test to use for a specific purpose, how to administer the chosen test, and how to interpret test scores. (125)

*Standard 12.18* In educational settings, score reports should be accompanied by a clear presentation of information on how to interpret the scores, including the degree of measurement error associated with each score or classification level, and by supplementary information related to group summary scores. In addition, dates of test administration and relevant norming studies should be included in score reports. (200)

## Chapter 8: Performance-Level Setting

This chapter briefly describes the LEAP 2025 high school performance-level setting and presents the cut scores and achievement-level descriptors derived from the performance-level setting. Since the LDOE used PARCC cut scores for the LEAP 2025 high school 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 2017–2018 administrations, the LEAP 2025 High School assessments measured different content and constructs than did previous tests administered to Louisiana students. The new tests were built using the PARCC item bank and were fully aligned to the Louisiana Student Standards. The new tests were reported on new scales, and students were classified by achievement levels based on their knowledge and ability to perform different tasks in relation to the new test content and standards.

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

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

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

### 8.2 Cut Scores

This section presents the cut scores for each grade and content area of the LEAP 2025 High School assessments. Table 8.1 summarizes the cut scores.

**Table 8.1 LEAP 2025 High School Assessment Cut Scores**

Content	Approaching Basic	Basic	Mastery	Advanced
English I	700	725	750	791
English II	700	725	750	794
Algebra I	700	725	750	805
Geometry	700	725	750	783

### 8.3 Category Cut Scores

As stated in Chapter 6, section “Category- and Subcategory-Level Subscores,” student performance on ELA and mathematics reporting categories and subcategories was classified into one of three performance ratings: *Strong*, *Moderate*, and *Weak*. Detailed rules for calculating performance ratings for ELA and mathematics categories and subcategories can be found in that section.

### 8.4 Achievement-Level Definitions

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

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

Table 8.2 summarizes the LEAP 2025 High School scale-score ranges for each level of achievement.

**Table 8.2 Achievement-Level Scale-Score Ranges**

Achievement Level	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
English I	650–699	700–724	725–749	750–790	791–850
English II	650–699	700–724	725–749	750–793	794–850
Algebra I	650–699	700–724	725–749	750–804	805–850
Geometry	650–699	700–724	725–749	750–782	783–850

### 8.5 Summary

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

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

*Standard 5.21* When proposed score interpretations involve one or more cut scores, the rationale and procedures used for establishing cut scores should be documented clearly. (107)

*Standard 5.22* When cut scores defining pass-fail or proficiency levels are based on direct judgments about the adequacy of item or test performances, the judgmental process should be designed so that the participants providing the judgments can bring their knowledge and experience to bear in a reasonable way. (108)

## Chapter 9: Evidence of Construct-Related Validity

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

**Table 9.1 Evidence of Validity and the Corresponding Technical Report Chapter**

Source of Validity	Related Information	Related Chapter/Source
Evidence Based on Test Content	Item Development Process	Chapter 3 2018-2019 LEAP 2025 High School ELA and Mathematics Assessment Frameworks
	Test Blueprint and Item Alignment to Curriculum and Standards	Chapter 3 2018-2019 LEAP 2025 High School 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 2025 High School 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 Relations to Other Variables	Divergent Validity	Chapter 9
Evidence Based on the Consequences of Testing	Scale Score and Performance Level Information	Chapter 7
	Test Interpretive Guide	Chapter 4

In this chapter, DRC presents evidence of construct-related validity through studies of test reliability, convergent validity, and divergent validity. All analyses in this chapter are based on census data.

Chapter 9 of this report demonstrates adherence to the American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, &

NCME, 2014) Standards 1.13, 1.21, 2.0, 2.3, 2.13, 2.14, 2.16, and 2.19. Each standard is discussed in the pertinent section of this chapter.

## 9.1 Construct-Irrelevant Variance and Construct Underrepresentation

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

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

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

## 9.2 Reliability

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

The 2014 *Standards* indicates the following:

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

In accordance with the *Standards* in developing and maintaining tests of the highest quality, DRC has calculated the reliability of each LEAP 2025 test in a variety of ways: reliability of raw scores, overall standard error of measurement (SEM), IRT-based conditional SEM, and decision consistency of achievement-level classifications.

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

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

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

The total score reliabilities are discussed in Section 9.3 of this chapter. The category and subcategory reliabilities and SEMs are presented in Sections 9.11 and 9.4 and 9.11. The SEM of the total score is discussed in Section 9.4.

*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 9.3 and is reported in raw score units. The conditional SEM is discussed in Section 9.5 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 9.3 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 initial testers. Since the summer forms are primarily administered to students retesting, statistics for the summer form will not be reported. The summer form had been previously administered to a spring or fall population; therefore, the form's reliability information can be found in earlier technical reports. .

### 9.3 Test Reliability

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

Cronbach's coefficient alpha was computed using the formula

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



where  $n$  is the number of items on the test,  $\sigma_i^2$  is the variance of item  $i$ , and  $\sigma_x^2$  is the variance of the total test score.

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

The reliability coefficients for the fall and spring LEAP 2025 HS assessments are reported in Table 9.2. English I and English II have one writing component (RI or RL) that has the same score as another component (WE). The item score for the RI/RL writing component was excluded from the reliability computation. The reliability statistics ranged from 0.86 to 0.92 and from 0.90 to 0.91 for the fall and spring administrations, respectively. The two administrations had very similar reliability statistics. These results indicate acceptable reliability coefficients for the LEAP 2025 high school tests.

**Table 9.2 Reliability**

Administration	Course	Form	Number of Items	Number of Score Points	SEM	Cronbach's Alpha	N-Count
Fall 2018	English I	B	34	94	5.70	0.89	≥4,430
	English II	B	34	94	5.43	0.90	≥5,500
	Algebra I	B	39	68	3.68	0.86	≥3,130
	Geometry	B	38	68	3.82	0.92	≥4,730
Spring 2019	English I	D	34	94	5.55	0.90	≥45,850
		E	33	90	5.61	0.90	≥45,850
	English II	D	34	94	5.51	0.90	≥41,600
		E	33	90	5.60	0.90	≥41,600
	Algebra I	D	39	68	3.79	0.90	≥46,630
		E	39	68	3.73	0.91	≥46,630
	Geometry	D	39	68	3.64	0.91	≥35,680
		E	39	68	3.70	0.91	≥35,680

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

#### 9.4 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:

$$SEM = SD\sqrt{1 - R_{xx'}}, \quad (9.2)$$

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

The SEM at the test-form level was computed in raw score metric and is also presented in Table 9.2. With English I and English II, the raw score used to calculate the SD included the RI/RL component and weighting of WE.

### 9.5 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 accordance with Standard 2.14, which states:

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

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

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

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

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

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

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

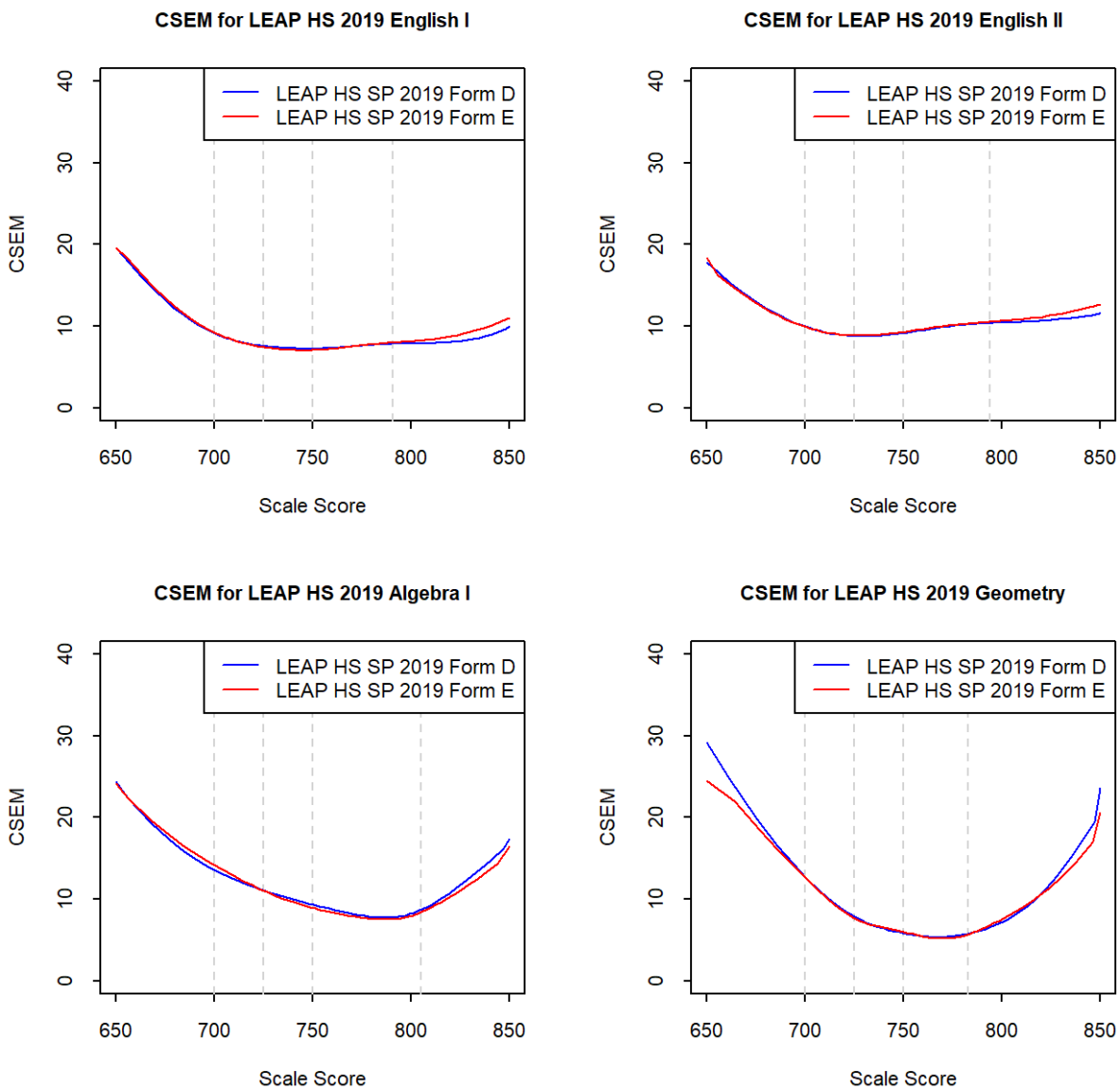
Note that the CSEMs vary in magnitude across the entire range of student ability estimates (i.e., scale scores) and are smaller in the middle of the score distribution and larger at the tails. This pattern is expected when IRT methods are used. The CSEMs at the four cut scores that define the performance 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**

Administration	Course	Form	<i>Approaching Basic</i>		<i>Basic</i>		<i>Mastery</i>		<i>Advanced</i>	
			Cut Score	CSEM	Cut Score	CSEM	Cut Score	CSEM	Cut Score	CSEM
Fall 2018	English I	B	700	11	725	9	750	9	791	10
	English II	B	700	12	725	11	750	11	794	12
	Algebra I	B	700	12	725	10	750	10	805	9
	Geometry	B	700	11	725	7	750	6	783	6
Spring 2019	English I	D	700	9	725	8	750	7	791	8
		E	700	9	725	7	750	7	791	8
	English II	D	700	10	725	9	750	9	794	10
		E	700	10	725	9	750	9	794	10
	Algebra I	D	700	14	725	11	750	9	805	9
		E	700	14	725	11	750	9	805	8
	Geometry	D	700	13	725	8	750	6	783	6
		E	700	13	725	8	750	6	783	6

Figure 9.1 displays the CSEM curves for each subject area. With fixed-form assessments, the estimates of measurement error tend to be higher at the low and high ends of the scale-score range, where few items measure the ability levels. Generally, there are few students with extreme scores, and these score levels cannot be estimated as accurately as levels toward the middle of the ability range. The middle of the ability range, where cut scores are located, shows lower measurement error than the low and high ends of the ability ranges. Figure 9.1 demonstrates that the tests are designed so that measurement error is minimized in the middle of the scale range, where most students are located.

Figure 9.1 CSEM Curves for LEAP High School 2019



## 9.6 Classification Accuracy and Consistency

### *Classification Accuracy*

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

### *Classification Consistency*

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

The Livingston-Lewis (1995) methodology was used to calculate classification accuracy statistics based on the spring 2019 LEAP 2025 results. The Livingston-Lewis procedure utilizes a beta-binomial model that requires two steps: (1) fitting proportion-correct true scores to a four-parameter beta distribution and (2) using the binomial distribution to estimate classification accuracy and consistency. All calculations for classification accuracy and consistency are based on census data.

Classification consistency and classification accuracy conditioned on achievement level (see Table 9.4 and 9.5) and on cut score (see Tables 9.6 and 9.7) are presented for the 2019 LEAP 2025 high school ELA and mathematics assessments in this section of the report. The magnitude of classification consistency and accuracy measures is influenced by several key features of a test's design, including the number of items, the location and number of cut scores, the score distribution, and the reliability and associated SEM. As seen in Table 9.4, classification accuracy conditioned on achievement level ranged from 0.00 to 0.85. As seen in Table 9.5, classification consistency conditioned on achievement level ranged from 0.21 to 0.79. For some mathematics tests, classification accuracy and consistency conditioned on the *Unsatisfactory* level were very low. A possible reason for these relatively low *Unsatisfactory* level values is the fact that there were not enough easy items to distinguish the *Unsatisfactory* level from the *Approaching Basic* performance level.

**Table 9.4 Classification Accuracy Conditioned on Level of Achievement**

Classification Accuracy							
Administration	Course	Form	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
Fall 2018	English I	B	0.67	0.67	0.70	0.82	0.68
	English II	B	0.71	0.62	0.64	0.76	0.74
	Algebra I	B	0.46	0.66	0.60	0.81	0.63
	Geometry	B	0.35	0.71	0.74	0.84	0.69
Spring 2019	English I	D	0.72	0.65	0.70	0.80	0.64
		E	0.70	0.66	0.72	0.80	0.65
	English II	D	0.77	0.65	0.62	0.76	0.70
		E	0.77	0.61	0.66	0.75	0.67
	Algebra I	D	0.00	0.78	0.59	0.84	0.76
		E	0.00	0.78	0.60	0.84	0.77
	Geometry	D	0.00	0.80	0.70	0.85	0.75
		E	0.00	0.80	0.69	0.85	0.75

**Table 9.5 Classification Consistency Conditioned on Level of Achievement**

Classification Consistency							
Administration	Course	Form	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
Fall 2018	English I	B	0.65	0.54	0.56	0.73	0.64
	English II	B	0.70	0.48	0.49	0.66	0.71
	Algebra I	B	0.44	0.49	0.47	0.75	0.60
	Geometry	B	0.39	0.55	0.62	0.78	0.68
Spring 2019	English I	D	0.69	0.51	0.56	0.71	0.60
		E	0.65	0.54	0.58	0.71	0.61
	English II	D	0.75	0.51	0.47	0.65	0.66
		E	0.74	0.49	0.50	0.63	0.64
	Algebra I	D	0.32	0.46	0.48	0.79	0.73
		E	0.32	0.46	0.48	0.79	0.74
	Geometry	D	0.21	0.57	0.60	0.77	0.72
		E	0.21	0.57	0.59	0.78	0.70

Perhaps the most important indices for accountability systems are those for the accuracy and consistency of classification decisions made at specific cut points. To evaluate decisions at specific cut points, the joint distribution of all the performance levels is collapsed into a dichotomized distribution around that specific cut point. As an example, for the LEAP 2025 assessments, a dichotomization at the cut point between the *Basic* and *Mastery* classifications was formed. The proportion of correct classifications below this particular cut point is equal to the sum of all the cells at the *Unsatisfactory*, *Approaching Basic*, and *Basic* levels, and the proportion of correct classifications above this 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 statistics and Table 9.7 shows the classification consistency estimates when

conditioned on LEAP 2025 High School cut scores. Table 9.6 shows that classification accuracy at achievement cut points ranged from 0.86 to 0.98. Table 9.7 shows that classification consistency at achievement cut points ranged from 0.81 to 0.98. Classification consistency and accuracy at achievement cut points tend to be higher values than those conditioned on performance level.

The classification accuracy statistics are at or above 0.86, while the classification consistency statistics are at or above 0.81. These results suggest that consistent and accurate performance-level classifications are being made for students in Louisiana based on the LEAP 2025 High School assessments.

**Table 9.6 Classification Accuracy at Achievement Cut Points**

Classification Accuracy						
Administration	Course	Form	<i>Unsatisfactory/ Approaching Basic</i>	<i>Approaching Basic/ Basic</i>	<i>Basic/ Mastery</i>	<i>Mastery/ Advanced</i>
Fall 2018	English I	B	0.96	0.93	0.90	0.95
	English II	B	0.95	0.91	0.90	0.94
	Algebra I	B	0.92	0.88	0.89	0.98
	Geometry	B	0.92	0.90	0.93	0.98
Spring 2019	English I	D	0.95	0.92	0.90	0.95
		E	0.95	0.92	0.90	0.95
	English II	D	0.95	0.92	0.90	0.93
		E	0.94	0.92	0.90	0.93
	Algebra I	D	0.90	0.86	0.92	0.98
		E	0.90	0.86	0.92	0.98
	Geometry	D	0.95	0.87	0.93	0.98
		E	0.95	0.87	0.93	0.98

**Table 9.7 Classification Consistency at Achievement Cut Points**

Classification Consistency						
Administration	Course	Form	<i>Unsatisfactory/ Approaching Basic</i>	<i>Approaching Basic/ Basic</i>	<i>Basic/ Mastery</i>	<i>Mastery/ Advanced</i>
Fall 2018	English I	B	0.95	0.90	0.86	0.92
	English II	B	0.92	0.87	0.86	0.92
	Algebra I	B	0.89	0.83	0.85	0.98
	Geometry	B	0.89	0.86	0.90	0.97
Spring 2019	English I	D	0.93	0.88	0.87	0.93
		E	0.93	0.88	0.86	0.93
	English II	D	0.93	0.89	0.87	0.90
		E	0.92	0.88	0.86	0.91
	Algebra I	D	0.86	0.81	0.88	0.98
		E	0.86	0.81	0.88	0.98
	Geometry	D	0.91	0.82	0.90	0.97
		E	0.91	0.82	0.90	0.97

## 9.7 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 geometry test is designed to measure a single overall construct—geometry achievement; therefore, the items comprising the LEAP 2025 Geometry test should measure only geometry, 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 the degree of the same general construct (Pearson, 2015, page 128). The reliability statistics shown in Table 9.2 are all above 0.89, indicating that items on the 2018 LEAP 2025 High School assessments are homogeneous. For a group of items to be homogeneous, the items must all measure the same construct (i.e., construct validity) or represent the same content domain (i.e., content validity). Because IRT models were used to calibrate test items and to report student scores, item fit is also relevant to construct validity. The extent to which test items function as the IRT model prescribes is relevant to the validation of test scores. As shown in Chapter 6, no items were flagged for poor model/data fit.

## 9.8 Principal Components Analysis

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

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

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

A PCA was conducted for each subject of the LEAP 2025 assessments. A large first principal component is evident in each analysis. It is common to have additional eigenvalues greater than 1.0, which may suggest the presence of other factors.

For the subjects of the LEAP 2025 assessments, the ratio of variance accounted for by the first factor to variance accounted for by the second factor is sufficiently large to indicate that the unidimensionality assumption holds. All the LEAP 2025 High School tests exhibit first principal components accounting for more than 20% of the test variance (



Table 9.8 through Table 9.11), except for the Algebra I spring 2019 administration. To further investigate the unidimensionality of the assessments, the ratio of the first eigenvalue to the second eigenvalue was found and is included in the row below the second component in each table. These ratios show that the first eigenvalue is at least four times as large as the second eigenvalue for all the LEAP 2025 assessments. This substantial difference in magnitude indicates that one factor appears to be dominant and that the LEAP 2025 High School tests are essentially unidimensional.

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

**Table 9.8 Principal Component Analysis: English I**

Administration	Form	Components	Eigenvalue	Percentage of Variance Explained	Cumulative Percentage of Variance Explained
Fall 2018	B	First Component	7.89	24.67	24.67
		Second Component	1.54	4.80	29.47
		Ratio (First/Second)	5.14	-	-
Spring 2019	D	First Component	8.37	26.15	26.15
		Second Component	1.38	4.31	30.47
		Ratio (First/Second)	6.06	-	-
Spring 2019	E	First Component	8.19	25.59	25.59
		Second Component	1.25	3.92	29.51
		Ratio (First/Second)	6.53	-	-

**Table 9.9 Principal Component Analysis: English II**

Administration	Form	Components	Eigenvalue	Percentage of Variance Explained	Cumulative Percentage of Variance Explained
Fall 2018	B	First Component	7.99	24.95	24.95
		Second Component	1.54	4.81	29.76
		Ratio (First/Second)	5.19	-	-
Spring 2019	D	First Component	8.28	25.86	25.86
		Second Component	1.38	4.32	30.18
		Ratio (First/Second)	5.98	-	-
Spring 2019	E	First Component	8.01	25.04	25.04
		Second Component	1.39	4.33	29.37
		Ratio (First/Second)	5.78	-	-

**Table 9.10 Principal Component Analysis: Algebra I**

Administration	Form	Components	Eigenvalue	Percentage of Variance Explained	Cumulative Percentage of Variance Explained
Fall 2018	B	First Component	6.80	17.44	17.44
		Second Component	1.43	3.66	21.09
		Ratio (First/Second)	4.77	-	-
Spring 2019	D	First Component	9.07	23.27	23.27
		Second Component	1.28	3.29	26.55
		Ratio (First/Second)	7.08	-	-
Spring 2019	E	First Component	9.22	23.64	23.64
		Second Component	1.28	3.28	26.92
		Ratio (First/Second)	7.20	-	-

**Table 9.11 Principal Component Analysis: Geometry**

Administration	Form	Components	Eigenvalue	Percentage of Variance Explained	Cumulative Percentage of Variance Explained
Fall 2018	B	First Component	10.73	28.25	28.25
		Second Component	1.54	4.05	32.29
		Ratio (First/Second)	6.98	-	-
Spring 2019	D	First Component	10.47	26.83	26.83
		Second Component	1.44	3.70	30.53
		Ratio (First/Second)	7.25	-	-
Spring 2019	E	First Component	10.03	25.73	25.73
		Second Component	1.39	3.57	29.29
		Ratio (First/Second)	7.21	-	-

## 9.9 Analyses by Reporting Categories and Subcategories

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

## 9.10 Correlations among Reporting Categories and Subcategories

This section reports the strength of the interrelationships among the reporting categories or subcategories by computing the correlation between them. Table 9.12 through Table 9.19 report the uncorrected Pearson product-moment (PPM) correlation coefficients, the PPM corrected for attenuation (CAPP). The PPM among the categories and subcategories is presented below the

diagonal portion of the matrix, the CAPPM is presented above the diagonal portion of the matrix, and the reliability coefficients used are shown in Table 9.12 through Table 9.19.

The uncorrected PPM in Table 9.12 through Table 9.19 should be interpreted in the context of the reliability coefficient. In general, lower PPM coefficients are expected between variables that are less reliable. In most cases, the PPM coefficients show that performance on one category or subcategory is moderately to strongly related to performance on another category or subcategory within the same grade and content area. The value of the correlation coefficients will be affected by the limited number of items measuring each category or subcategory. Therefore, caution should be used when comparing the PPM coefficients that measure the relationships between categories or subcategories to those that measure the relationships between content areas. A more modest relationship (i.e., smaller correlation coefficients) is expected to be reported between the categories or subcategories as a consequence of the lower number of items measuring each of the reporting categories. The PPM between two category subscores, for example, may be artificially low because of measurement error.

The CAPPM is reported along with the PPM as indicated by 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)

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

$$CAPPM = \frac{r_{xy}}{\sqrt{r_{xx}r_{yy}}}, \quad (9.5)$$

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

The English I and English II assessments show moderate relationships between the reading and writing categories, indicating that these two categories measure some different traits. Across all tables, the CAPPM indicates moderate or strong relationships between the subcategories. The CAPPM for reading vocabulary, written expression, and knowledge and use of language are moderate. In some cases, the CAPPM is greater than 1.0. "Disattenuated values greater than 1.00 indicate that measurement error is not randomly distributed" (Schumacker, 1996). The moderate or strong relationships suggested by the CAPPM in Table 9.12 through Table 9.19 are further evidence of the validity of the test construct. Since the overall content area is comprised of the category or subcategory subscores and the content area is expected to measure a single dimension, these subscores are expected to be moderately or highly related.

**Table 9.12 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Categories: English I**

Administration	Form	No.	Category	N Items	1	2
Fall 2018	B	1	Reading	30	.	0.81
		2	Writing	4	0.73	.
Spring 2019	D	1	Reading	30	.	0.82
		2	Writing	4	0.74	.
	E	1	Reading	29	.	0.84
		2	Writing	4	0.74	.

**Table 9.13 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Subcategories: English I**

Administration	Form	No.	Subcategory Uncorrected and Corrected Correlation Coefficients: English I						
			Subcategory	N Items	1	2	3	4	5
Fall 2018	B	1	Reading Literary Text	9	.	1.03	1.06	0.92	0.90
		2	Reading Informational Text	15	0.73	.	1.04	0.85	0.84
		3	Reading Vocabulary	6	0.56	0.61	.	0.73	0.73
		4	Written Expression	2	0.68	0.70	0.45	.	1.10
		5	Written Knowledge & Use of Language	2	0.67	0.69	0.45	0.95	.
Spring 2019	D	1	Reading Literary Text	9	.	0.96	0.91	0.94	0.90
		2	Reading Informational Text	14	0.65	.	0.94	0.86	0.83
		3	Reading Vocabulary	7	0.55	0.66	.	0.72	0.70
		4	Written Expression	2	0.66	0.70	0.52	.	1.11
		5	Written Knowledge & Use of Language	2	0.64	0.68	0.51	0.93	.
	E	1	Reading Literary Text	11	.	0.90	0.90	0.90	0.87
		2	Reading Informational Text	12	0.65	.	1.06	0.94	0.91
		3	Reading Vocabulary	6	0.55	0.66	.	0.82	0.81
		4	Written Expression	2	0.66	0.70	0.52	.	1.24
		5	Written Knowledge & Use of Language	2	0.64	0.68	0.51	0.93	.

**Table 9.14 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Categories: English II**

Administration	Form	No.	Category	N Items	1	2
Fall 2018	B	1	Reading	30	.	0.82
		2	Writing	4	0.74	.
Spring 2019	D	1	Reading	30	.	0.83
		2	Writing	4	0.75	.
	E	1	Reading	29	.	0.84
		2	Writing	4	0.75	.

**Table 9.15 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Subcategories: English II**

		Subcategory Uncorrected and Corrected Correlation Coefficients: English I							
Administration	Form	No.	Subcategory	N Items	1	2	3	4	5
Fall 2018	B	1	Reading Literary Text	7	.	1.11	1.06	1.04	1.03
		2	Reading Informational Text	14	0.72	.	1.00	0.86	0.85
		3	Reading Vocabulary	9	0.61	0.68	.	0.71	0.71
		4	Written Expression	2	0.71	0.71	0.51	.	1.10
		5	Written Knowledge & Use of Language	2	0.71	0.70	0.51	0.95	.
Spring 2019	D	1	Reading Literary Text	7	.	0.82	0.81	0.95	0.93
		2	Reading Informational Text	13	0.52	.	0.97	0.85	0.86
		3	Reading Vocabulary	10	0.49	0.69	.	0.73	0.75
		4	Written Expression	2	0.64	0.68	0.55	.	1.15
		5	Written Knowledge & Use of Language	2	0.63	0.68	0.56	0.96	.
	E	1	Reading Literary Text	10	.	0.73	0.76	0.87	0.85
		2	Reading Informational Text	11	0.52	.	1.05	0.90	0.91
		3	Reading Vocabulary	8	0.49	0.69	.	0.82	0.83
		4	Written Expression	2	0.64	0.68	0.55	.	1.25
		5	Written Knowledge & Use of Language	2	0.63	0.68	0.56	0.96	.

**Table 9.16 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Categories: Algebra I**

Administration	Form	No.	Category	N Items	1	2	3	4
Fall 2018	B	1	Major Content	21	.	1.02	0.90	0.98
		2	Additional & Supporting Con	10	0.64	.	0.98	0.93
		3	Expressing Mathematical Reasoning	3	0.54	0.54	.	0.90
		4	Modeling & Application	4	0.68	0.59	0.55	.
Spring 2019	D	1	Major Content	22	.	1.01	0.94	0.98
		2	Additional & Supporting Con	10	0.74	.	0.99	0.96
		3	Expressing Mathematical Reasoning	3	0.69	0.66	.	0.94
		4	Modeling & Application	4	0.75	0.67	0.66	.
	E	1	Major Content	22	.	1.00	0.94	0.95
		2	Additional & Supporting Con	10	0.74	.	0.96	0.92
		3	Expressing Mathematical Reasoning	3	0.69	0.66	.	0.91
		4	Modeling & Application	4	0.75	0.67	0.66	.

**Table 9.17 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Categories: Geometry**

Administration	Form	No.	Category	N Items	1	2	3	4
Fall 2018	B	1	Major Content	20	.	0.99	0.99	0.96
		2	Additional & Supporting Con	12	0.77	.	0.97	0.95
		3	Expressing Mathematical Reasoning	3	0.75	0.67	.	1.10
		4	Modeling & Application	3	0.75	0.67	0.76	.
Spring 2019	D	1	Major Content	19	.	0.99	1.01	0.95
		2	Additional & Supporting Con	13	0.75	.	1.02	1.01
		3	Expressing Mathematical Reasoning	3	0.76	0.69	.	1.09
		4	Modeling & Application	4	0.73	0.70	0.75	.
	E	1	Major Content	19	.	0.99	0.98	0.99
		2	Additional & Supporting Con	13	0.75	.	0.97	1.02
		3	Expressing Mathematical Reasoning	3	0.76	0.69	.	1.08
		4	Modeling & Application	4	0.73	0.70	0.75	.

**Table 9.18 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Subcategories: Algebra I**

Administration	Form	No.	Subcategory	N Items	1	2	3
Fall 2018	B	1	A1	5	.	1.35	1.15
		2	A2	6	0.51	.	1.32
		3	A3	6	0.49	0.46	.
Spring 2019	D	1	A1	6	.	0.38	1.21
		2	A2	7	0.21	.	0.22
		3	A3	6	0.67	0.11	.
	E	1	A1	7	.	0.39	1.27
		2	A2	6	0.21	.	0.24
		3	A3	6	0.67	0.11	.

**Table 9.19 Uncorrected Correlation Coefficient (below Diagonal) and Corrected Correlation Coefficient (above Diagonal) among Categories: Geometry**

Administration	Form	No.	Subcategory	N Items	1	2
Fall 2018	B	1	A1	13	.	1.00
		2	A2	7	0.74	.
Spring 2019	D	1	A1	11	.	0.95
		2	A2	8	0.69	.
	E	1	A1	11	.	1.03
		2	A2	8	0.69	.

### 9.11 Reliability of Reporting Categories, 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 categories or subcategories by subject using the census data. These statistics are presented in Tables 9.18 through 9.22. Reliability indices, such as Cronbach's coefficient alpha (and resulting SEM), are a function of the number of items on a test, the average covariance between item pairs, and the variance of a test's total score. In general, it is expected that the coefficient alpha would be lower for a category or subcategory assessed by a small number of items than for a category or subcategory assessed by a larger number of items.

### 9.12 Standard Error of Measurement of Reporting Categories or Subcategories

This chapter also reports the SEM associated with each of the categories and subcategories in Table 9.20 through Table 9.27. These SEMs are reported in the raw score metric.

**Table 9.20 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of English I Categories**

Administration	Form	Category	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
Fall 2018	B	Reading	30	64	30.81	11.91	4.50	0.86
		Writing	4	30	10.83	6.74	1.64	0.94
Spring 2019	D	Reading	30	64	28.88	12.06	4.44	0.86
		Writing	4	30	9.57	6.48	1.64	0.94
	E	Reading	30	64	28.88	12.06	4.52	0.86
		Writing	4	30	9.57	6.48	1.61	0.94



**Table 9.21 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of English I Subcategories**

Admin.	Form	Subcategory	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
Fall 2018	B	Reading Literary Text	9	20	7.86	4.07	2.45	0.64
		Reading Informational Text	15	32	16.33	6.71	3.11	0.79
		Reading Vocabulary	6	12	6.63	2.62	1.96	0.44
		Written Expression	2	24	8.04	5.15	1.99	0.85
		Knowledge & Use of Language	2	6	2.78	1.66	0.62	0.86
Spring 2019	D	Reading Literary Text	9	20	10.12	4.60	2.93	0.59
		Reading Informational Text	14	30	11.69	5.99	2.79	0.78
		Reading Vocabulary	7	14	7.06	3.25	2.00	0.62
		Written Expression	2	24	7.10	4.97	2.00	0.84
		Knowledge & Use of Language	2	6	2.47	1.59	0.61	0.85
	E	Reading Literary Text	11	22	10.12	4.60	2.44	0.72
		Reading Informational Text	12	26	11.69	5.99	3.09	0.73
		Reading Vocabulary	6	12	7.06	3.25	2.24	0.53
		Written Expression	2	24	7.10	4.97	2.50	0.75
		Knowledge & Use of Language	2	6	2.47	1.59	0.78	0.76

**Table 9.22 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of English II Categories**

Administration	Form	Category	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
Fall 2018	B	Reading	30	64	28.81	10.87	4.13	0.86
		Writing	4	30	11.65	7.08	1.62	0.95
Spring 2019	D	Reading	29	60	27.46	11.51	4.28	0.86
		Writing	4	30	11.42	6.91	2.16	0.90
	E	Reading	29	60	27.46	11.51	4.30	0.86
		Writing	4	30	11.42	6.91	2.05	0.91

**Table 9.23 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of English II Subcategories**

Administration	Form	Subcategory	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
Fall 2018	B	Reading Literary Text	7	16	6.60	3.12	2.10	0.55
		Reading Informational Text	14	30	12.10	5.56	2.63	0.78
		Reading Vocabulary	9	18	10.12	3.53	2.23	0.60
		Written Expression	2	24	8.82	5.39	1.98	0.87
		Knowledge & Use of Language	2	6	2.82	1.75	0.62	0.88
Spring 2019	D	Reading Literary Text	7	16	7.83	4.43	3.00	0.54
		Reading Informational Text	13	28	10.84	5.36	2.63	0.76
		Reading Vocabulary	10	20	8.79	3.82	2.20	0.67
		Written Expression	2	24	8.60	5.24	2.15	0.83
		Knowledge & Use of Language	2	6	2.83	1.72	0.70	0.84
	E	Reading Literary Text	10	20	7.83	4.43	2.43	0.70
		Reading Informational Text	11	24	10.84	5.36	2.77	0.73
		Reading Vocabulary	8	16	8.79	3.82	2.46	0.59
		Written Expression	2	24	8.60	5.24	2.56	0.76
		Knowledge & Use of Language	2	6	2.83	1.72	0.81	0.78

**Table 9.24 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Algebra I Categories**

Administration	Form	Category	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
Fall 2018	B	Major Content	21	28	11.89	4.64	2.60	0.68
		Additional & Supporting Content	11	14	4.68	2.49	1.62	0.58
		Expressing Mathematical Reasoning	3	11	1.00	1.39	0.96	0.53
		Modeling & Application	4	15	3.31	2.88	1.56	0.70
Spring 2019	D	Major Content	22	28	12.07	5.43	2.45	0.80
		Additional & Supporting Content	10	14	5.04	2.89	1.67	0.67
		Expressing Mathematical Reasoning	3	11	1.78	2.27	1.30	0.67
		Modeling & Application	4	15	3.15	3.20	1.66	0.73
	E	Major Content	22	28	12.07	5.43	2.47	0.79
		Additional & Supporting Content	10	14	5.04	2.89	1.62	0.69
		Expressing Mathematical Reasoning	3	11	1.78	2.27	1.27	0.69
		Modeling & Application	4	15	3.15	3.20	1.51	0.78

**Table 9.25 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Algebra I Subcategories**

Administration	Form	Subcategories	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
Fall 2018	B	A1	5	7	2.80	1.70	1.25	0.46
		A2	6	10	5.04	2.11	1.75	0.31
		A3	6	6	2.51	1.40	1.09	0.40
Spring 2019	D	A1	6	7	3.21	2.18	1.38	0.60
		A2	7	12	4.17	2.58	1.81	0.51
		A3	6	6	3.44	1.98	1.39	0.51
	E	A1	7	9	3.21	2.18	1.38	0.60
		A2	6	7	4.17	2.58	1.86	0.48
		A3	6	9	3.44	1.98	1.45	0.46

**Table 9.26 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Geometry Categories**

Administration	Form	Category	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
Fall 2018	B	Major Content	20	26	11.57	5.85	2.18	0.86
		Additional & Supporting Content	12	16	6.46	3.01	1.64	0.70
		Expressing Mathematical Reasoning	3	11	1.96	2.42	1.38	0.68
		Modeling & Application	3	15	2.61	3.38	1.85	0.70
Spring 2019	D	Major Content	19	26	10.76	5.43	2.19	0.84
		Additional & Supporting Content	13	16	6.18	3.13	1.77	0.68
		Expressing Mathematical Reasoning	3	11	2.05	2.54	1.45	0.67
		Modeling & Application	4	15	1.79	2.75	1.50	0.70
	E	Major Content	19	26	10.76	5.43	2.34	0.81
		Additional & Supporting Content	13	16	6.18	3.13	1.72	0.70
		Expressing Mathematical Reasoning	3	11	2.05	2.54	1.32	0.73
		Modeling & Application	4	15	1.79	2.75	1.59	0.67

**Table 9.27 Mean, Standard Deviation, and Standard Error of Measurement (SEM) of Geometry Subcategories**

Administration	Form	Subcategories	Number of Items	Number of Score Points	Mean Raw Score	Raw Score Std. Dev.	SEM	Cronbach's Alpha
Fall 2018	B	A1	13	17	7.87	3.90	1.77	0.79
		A2	7	9	3.70	2.35	1.28	0.70
Spring 2019	D	A1	11	17	7.66	3.71	1.88	0.74
		A2	8	9	3.10	2.17	1.18	0.71
	E	A1	11	17	7.66	3.71	1.85	0.75
		A2	8	9	3.10	2.17	1.38	0.60

### 9.13 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 High School assessments, correlations were computed between the English I, English II, Algebra I and Geometry total scores for students who took more than one subject test in 2019. These correlations are based on the census data, and the results are shown in Table 9.28 and Table 9.29 for the fall 2018 and spring 2019 administrations respectively. The correlation coefficients ranged from 0.54 to 0.88. The lowest correlation was observed between English II and algebra I in the fall 2018 administration, and the highest correlation was between English I and English II in the spring 2019 administration. Similar patterns were observed in both administrations. The correlation coefficients suggest that individual student scores across subjects are moderately related, indicating that these tests measure a similar knowledge base or general underlying ability while still measuring some different traits as planned.

**Table 9.28 Inter-Correlation of HS Content Area Scale Scores in Fall Administration**

	English I	English II	Algebra I	Geometry
English I	-	0.83	0.76	0.79
English II	0.83 (25)*	-	0.54	0.72
Algebra I	0.76 (916)	0.54 (234)	-	0.85
Geometry	0.79 (362)	0.72 (1,190)	0.85 (41)	-

\*The count of observations in the analysis is in parenthesis

**Table 9.29 Inter-Correlation of HS Content Area Scale Scores in Spring Administration**

	English I	English II	Algebra I	Geometry
English I	-	0.88	0.72	0.66
English II	0.88 (301)*	-	0.65	0.65
Algebra I	0.72 (32,531)	0.65 (3,788)	-	0.80
Geometry	0.66 (4,900)	0.65 (23,305)	0.80 (315)	-

\*The count of observations in the analysis is in parenthesis

## 9.14 Summary

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

*Standard 1.13* If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided. (26)

*Standard 1.21* When statistical adjustments, such as those for restriction of range or attenuation, are made, both adjusted and unadjusted coefficients, as well as the specific procedure used, and all statistics used in the adjustment, should be reported. Estimates of the construct-criterion relationship that remove the effects of measurement error on the test should be clearly reported as adjusted estimates. (29)

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

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

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

*Standard 2.14* When possible and appropriate, conditional standard errors of measurement should be reported at several score levels unless there is evidence that the standard error is constant across score levels. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score. (46)

*Standard 2.16* When a test or combination of measures is used to make classification decisions, estimates should be provided of the percentage of test takers who would be classified in the same way on two replications of the procedure. (46)

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

## Chapter 10: Fairness

---

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

This chapter demonstrates how the Leap 2025 assessments adhere to AERA, APA, & NCME Standards 3.1–3.6 and 3.16. These standards are from Chapter 3 of the *Standards*, which is titled “Fairness in Testing.” Each of these standards is presented in this chapter.

Standard 3.6 states:

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

Test scores of examinee subgroups that differ in meaning are an ongoing concern in any large-scale testing program. To lessen the possibility of differences in test score meaning, DRC follows several steps in the item development and item selection processes, as is explained in Section 10.1 of this chapter. In addition, LDOE assessment research and development experts and Louisiana educators conduct content and bias reviews on items during the selection process, as explained in Chapter 3. These practices adhere to Standard 3.3, which states,

Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test. (64)

The PARCC consortium 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 process, and any items flagged for DIF are further examined to determine possible bias. During the ELA and mathematics test development process, DRC content experts tried to avoid including 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 Standard 3.3.

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

*Standard 3.4* Test takers should receive comparable treatment during the test administration and scoring process. (65)

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

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

*Standard 3.1* Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (63)

*Standard 3.2* Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (64)

This chapter explains the steps taken by DRC to minimize words, phrases, and content that may be regarded as offensive by members of particular demographic subgroups. Section 3.2 of Chapter 3 discusses the content and bias review conducted for LEAP 2025. This review is also critical in fulfilling Standards 3.1 and 3.2. In addition to the Louisiana-developed items, the PARCC operational items used in the 2019 LEAP 2025 forms were critical to the forms construction process. Refer to the New Meridian website for the bias and sensitivity guidelines used and the processes and procedures followed by PARCC pertaining to these items (see <https://newmeridiancorp.org/>).

The DIF and reliability analyses in this section are based on the CIA data described in Chapter 6. The impact analyses (scale score mean and standard deviation) are based on the technical report sample described in Chapter 7.

## 10.1 Minimizing Bias through Careful Test Development

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

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

Second, item writers and test developers followed PARCC Fairness and Sensitivity Guidelines for reducing or eliminating bias. DRC test development staff reviewed all items and other testing materials with these guidelines in mind. Internal editorial reviews were conducted by at least three different people: a content editor who directly supervised the item writers, a style editor, and a content supervisor. The final test was again reviewed by people in these same roles and was also subjected to an independent review by 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 these conditions may indicate that an item is testing a construct irrelevant to what 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 10.1-10.4) 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 item development, test development and test construction processes to lessen the influence of these elements for large numbers of students. Unfortunately, these elements may continue to play a substantial role in some cases. To assess the extent to which items may be performing differently for various subgroups of interest, DIF analyses are conducted after each operational test administration.

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

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

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

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

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



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

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

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

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

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

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

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

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

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

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

### DIF Statistics for Demographic Groups

DIF analyses were conducted for groups defined by demographic characteristics. Tables 10.1 to Table 10.4 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.

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

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

*Section 504 Status:* Focal group is students who are classified as Section 504; reference group is all others.

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

*Military Affiliation:* Focus group is students who are affiliated with the military; reference group is all others.

*Foster Care Status:* Focus group is students who are in foster care; reference group is all others.

A negative SMD value implies that the focal group has a lower mean item score than the reference group, whereas a positive value implies that the focal group has a higher mean item score than the reference group, conditioned on the matching test score.

The minimum case count for the focal group was set at 200, and the minimum case count for the reference group was set at 400. The DIF analyses are not performed for subgroups of less than 200. In these cases, the statistical procedures do not have sufficient power to detect potential differences.

Tables 10.1 to Table 10.4 summarize the number of DIF flags for the spring 2019 administrations by subject 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 fall 2018 and summer 2019 administrations used intact forms. Therefore, DIF was not computed.

The DIF statistics for Form D of English I (see Table 10.1) can be considered an example. On the English I form, a total of three items were flagged for DIF for the female subgroup; two of the five items had moderate positive DIF and one of the five exhibited large negative DIF. For the ethnicity categories, four items were flagged for moderate negative DIF—one item each for Hispanic/Latino and Asian and two items for Black or African American. Finally, three items were flagged for moderate negative DIF for the English Learner group.

Table 10.1 Spring 2019 Administration DIF Statistics of English I: Number of Flagged Items

DIF Statistics				Count of Items at DIF Magnitude			
Form	Number of Items	Category	Group	Moderate		Large	
				B-	B+	C-	C+
D	34	Gender	Female	0	2	1	0
		Ethnicity	Hispanic/Latino	1	0	0	0
		Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
		Ethnicity	Asian	1	0	0	0
		Ethnicity	Black or African American	2	0	0	0
		Ethnicity	Two or More Races	0	0	0	0
		Education Classification	Special Education	0	0	0	0
		Education Classification	Gifted or Talented	0	0	0	0
		English Learner Status	EL	3	0	0	0
		Economic Status	Economically Disadvantaged	0	0	0	0
		Section 504 Status	Section 504	0	0	0	0
		Homeless Status	Homeless	0	0	0	0
		Military Affiliation	Affiliated	0	0	0	0
		Foster Care Status	Foster Care	NR	NR	NR	NR
E	33	Gender	Female	0	1	1	0
		Ethnicity	Hispanic/Latino	0	0	0	0
		Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
		Ethnicity	Asian	1	1	0	0
		Ethnicity	Black or African American	2	0	0	0
		Ethnicity	Two or More Races	0	0	0	0
		Education Classification	Special Education	0	0	0	0
		Education Classification	Gifted or Talented	0	0	0	0
		English Learner Status	EL	1	0	1	0
		Economic Status	Economically Disadvantaged	0	0	0	0
		Section 504 Status	Section 504	0	0	0	0
		Homeless Status	Homeless	0	0	0	0
		Military Affiliation	Affiliated	0	0	0	0
		Foster Care Status	Foster Care	NR	NR	NR	NR

Table 10.2 Spring 2019 Administration DIF Statistics of English II: Number of Flagged Items

DIF Statistics				Count of Items at DIF Magnitude			
Form	Number of Items	Category	Group	Moderate		Large	
				B-	B+	C-	C+
D	34	Gender	Female	2	2	0	0
		Ethnicity	Hispanic/Latino	1	0	0	0
		Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
		Ethnicity	Asian	0	0	0	0
		Ethnicity	Black or African American	0	0	0	0
		Ethnicity	Two or More Races	0	0	0	0
		Education Classification	Special Education	0	0	0	0
		Education Classification	Gifted or Talented	0	1	0	0
		English Learner Status	EL	2	0	0	0
		Economic Status	Economically Disadvantaged	0	0	0	0
		Section 504 Status	Section 504	0	0	0	0
		Homeless Status	Homeless	0	0	0	0
		Military Affiliation	Affiliated	0	0	0	0
		Foster Care Status	Foster Care	NR	NR	NR	NR
E	33	Gender	Female	2	2	0	0
		Ethnicity	Hispanic/Latino	0	0	0	0
		Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
		Ethnicity	Asian	1	1	0	0
		Ethnicity	Black or African American	1	0	0	0
		Ethnicity	Two or More Races	0	0	0	0
		Education Classification	Special Education	0	0	0	0
		Education Classification	Gifted or Talented	0	0	0	0
		English Learner Status	EL	1	1	0	0
		Economic Status	Economically Disadvantaged	0	0	0	0
		Section 504 Status	Section 504	0	0	0	0
		Homeless Status	Homeless	0	0	0	0
		Military Affiliation	Affiliated	0	0	0	0
		Foster Care Status	Foster Care	NR	NR	NR	NR

Table 10.3 Spring 2019 Administration DIF Statistics of Algebra I: Number of Flagged Items

DIF Statistics				Count of Items at DIF Magnitude			
Form	Number of Items	Category	Group	Moderate		Large	
				B-	B+	C-	C+
D	39	Gender	Female	0	0	0	0
		Ethnicity	Hispanic/Latino	0	0	1	0
		Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
		Ethnicity	Asian	0	2	0	0
		Ethnicity	Black or African American	0	0	1	0
		Ethnicity	Two or More Races	0	0	0	0
		Education Classification	Special Education	1	0	2	0
		Education Classification	Gifted or Talented	0	1	0	0
		English Learner Status	EL	1	0	1	0
		Economic Status	Economically Disadvantaged	1	0	0	0
		Section 504 Status	Section 504	0	0	0	0
		Homeless Status	Homeless	0	0	0	0
		Military Affiliation	Affiliated	0	1	0	0
		Foster Care Status	Foster Care	NR	NR	NR	NR
E	39	Gender	Female	0	0	0	0
		Ethnicity	Hispanic/Latino	0	0	0	0
		Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
		Ethnicity	Asian	0	2	0	0
		Ethnicity	Black or African American	1	0	0	0
		Ethnicity	Two or More Races	0	0	0	0
		Education Classification	Special Education	0	1	0	0
		Education Classification	Gifted or Talented	0	2	0	0
		English Learner Status	EL	1	0	0	0
		Economic Status	Economically Disadvantaged	0	0	0	0
		Section 504 Status	Section 504	0	0	0	0
		Homeless Status	Homeless	0	0	0	0
		Military Affiliation	Affiliated	0	1	0	0
		Foster Care Status	Foster Care	NR	NR	NR	NR

Table 10.4 Spring 2019 Administration DIF Statistics of Geometry: Number of Flagged Items

DIF Statistics				Count of Items at DIF Magnitude			
Form	Number of Items	Category	Group	Moderate		Large	
				B-	B+	C-	C+
D	39	Gender	Female	0	1	0	0
		Ethnicity	Hispanic/Latino	0	0	0	0
		Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
		Ethnicity	Asian	1	4	0	0
		Ethnicity	Black or African American	0	0	0	0
		Ethnicity	Two or More Races	0	0	0	0
		Education Classification	Special Education	0	0	0	0
		Education Classification	Gifted or Talented	0	0	0	1
		English Learner Status	EL	1	0	0	0
		Economic Status	Economically Disadvantaged	0	0	0	0
		Section 504 Status	Section 504	0	0	0	0
		Homeless Status	Homeless	NR	NR	NR	NR
		Military Affiliation	Affiliated	0	0	0	0
		Foster Care Status	Foster Care	NR	NR	NR	NR
E	39	Gender	Female	0	0	0	0
		Ethnicity	Hispanic/Latino	0	0	0	0
		Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
		Ethnicity	Asian	0	0	0	0
		Ethnicity	Black or African American	0	0	0	0
		Ethnicity	Two or More Races	0	0	0	0
		Education Classification	Special Education	0	0	0	0
		Education Classification	Gifted or Talented	0	0	0	0
		English Learner Status	EL	1	0	0	0
		Economic Status	Economically Disadvantaged	0	0	0	0
		Section 504 Status	Section 504	0	0	0	0
		Homeless Status	Homeless	NR	NR	NR	NR
		Military Affiliation	Affiliated	0	0	0	0
		Foster Care Status	Foster Care	NR	NR	NR	NR

### DIF Statistics for Test Language

All items on one CBT form of the mathematics test are transadapted from English into Spanish. Transadaptation takes into consideration linguistic and cultural differences and grade-level appropriate words. By accounting for these differences, the achievement of Spanish speakers can be measured in the same way as the achievement of English speakers. Please refer to Appendix E for more information about the transadaptation of Spanish mathematics forms. To help confirm that the test items can be measured similarly regardless of the language in which the items are published, a DIF set of analyses was performed. Two DIF analyses were performed using the 2019 LEAP 2025 mathematics operational items regardless of student count in the reference or focal group. Smaller counts for the groups needed to be tolerated since the overall count for those being administered the Spanish form was low.

For the first analysis, student responses for the shared operational items between 2018 and 2019 LEAP 2025 mathematics were combined. This approach increased the number of students who took the Spanish versions of the items. The Mantel-Haenszel (MH) and the Standardized Mean Difference (SMD) DIF procedures were performed on these common items. The second analysis focused on the items that were not shared between the 2018 and 2019 administrations. Although the MH and the SMD DIF procedures were performed on all 2019 LEAP 2025 operational items, the DIF flags were applied, where appropriate, to items that were not shared between 2018 and 2019.

For both analyses, DIF results were carefully reviewed whenever sample sizes were smaller than the required minimum sample size and when an item showed large (i.e., C) DIF. Table 10.5 summarizes how many items overall exhibited moderate or large DIF in mathematics.

**Table 10.5 2019 LEAP 2025 DIF Statistics: Number of Flagged Items, Mathematics**

DIF Statistics: Mathematics				Count of Items at DIF Magnitude			
				Moderate		Large	
Content Area	Number of Items	Category	Group	B-	B+	C-	C+
Algebra I	14	Test Language	Spanish		2	4	
Geometry	16	Test Language	Spanish		1	1	

### 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.

Table 10.6 through Table 10.13 present the number of students and test form reliability statistics (i.e., coefficient alpha; see Chapter 9). Scale score means, standard deviations, and effect sizes (i.e., Cohen's *d*) for the various subgroups of interest are reported by form in Table 10.14 through Table 10.21

### 10.4 Reliability

Tables 10.6–10.13 show the test form reliability coefficients and SEM by student gender, ethnicity, education classification, economic status, EL status, migrant status, Section 504 status, homeless status, military affiliation, and foster care status. The reliability coefficients for English I and II forms ranged from 0.83 to 0.92 and from 0.85 to 0.94 for the fall 2018 and spring 2019 administrations, respectively. For algebra I and geometry, the reliability coefficients ranged from 0.71 to 0.93 and from 0.80 to 0.95 for the fall 2018 and spring 2019 administrations, respectively. These analyses show that the test reliability is of acceptable magnitude for all the subgroups. Note that the reliability coefficients are NR for subgroups with fewer than 10 students.



Table 10.6 Fall 2018 Administration English I Reliability and SEM by Subgroup

Group	N Count	Cronbach's Alpha	SEM
<b>All Students</b>	≥4,430	0.89	5.70
<b>Gender</b>			
Female	≥2,180	0.89	5.66
Male	≥2,250	0.90	5.59
<b>Ethnicity</b>			
Hispanic/Latino	≥350	0.91	5.61
American Indian or Alaska Native	≥20	0.92	5.27
Asian	≥80	0.87	6.07
Black or African American	≥1,660	0.87	5.59
Native Hawaiian or Other Pacific	<10	NR	NR
White	≥2,200	0.87	5.78
Two or More Races	≥100	0.87	5.77
<b>Education Classification</b>			
Regular Education	≥3,750	0.88	5.69
Special Education	≥330	0.84	5.16
Gifted or Talented	≥350	0.84	5.67
<b>Economic Status</b>			
Economically Disadvantaged	<10	NR	NR
Not Economically Disadvantaged	≥2,640	0.89	5.73
<b>English Learner Status</b>			
Not English Learner	≥4,320	0.89	5.71
English Learner	≥110	0.82	5.14
<b>Migrant Status</b>			
Migrant	<10	NR	NR
Not Migrant	≥4,430	0.89	5.70
<b>Section 504 Status</b>			
Non-Section 504	≥4,080	0.89	5.70
Section 504	≥340	0.88	5.48
<b>Homeless Status</b>			
Not Homeless	≥4,360	0.86	6.59
Homeless	≥70	0.84	6.15
<b>Military Affiliation</b>			
Not Military Affiliated	≥4,260	0.86	6.62
Military Affiliated	≥170	0.83	6.01
<b>Foster Care Status</b>			
Not in Foster Care	≥4,400	0.86	6.60
Foster Care	≥70	0.84	6.15

Table 10.7 Spring 2019 Administration English I Reliability and SEM by Subgroup

Group	Form D			Form E		
	N Count	Cronbach's Alpha	SEM	N Count	Cronbach's Alpha	SEM
<b>All Students</b>	≥24,930	0.90	5.64	≥20,830	0.90	5.43
<b>Gender</b>						
Female	≥12,240	0.89	5.67	≥10,830	0.89	5.44
Male	≥12,680	0.90	5.52	≥9,990	0.90	5.37
<b>Ethnicity</b>						
Hispanic/Latino	≥1,760	0.92	5.38	≥1,410	0.92	5.27
American Indian or Alaska Native	≥170	0.88	5.60	≥150	0.86	5.55
Asian	≥450	0.91	5.68	≥370	0.90	5.33
Black or African American	≥10,920	0.87	5.53	≥8,730	0.88	5.38
Native Hawaiian or Other Pacific	≥10	0.94	5.53	≥10	0.92	4.95
White	≥11,120	0.89	5.76	≥9,730	0.88	5.45
Two or More Races	≥470	0.89	5.73	≥400	0.89	5.46
<b>Education Classification</b>						
Regular Education	≥20,160	0.88	5.69	≥18,880	0.89	5.44
Special Education	≥3,260	0.84	4.95	≥580	0.89	5.15
Gifted or Talented	≥1,500	0.85	5.73	≥1,370	0.87	5.20
<b>Economic Status</b>						
Economically Disadvantaged	≥15,240	0.89	5.57	≥12,170	0.88	5.42
Not Economically Disadvantaged	≥7,270	0.88	5.77	≥6,630	0.88	5.43
<b>English Learner Status</b>						
Not English Learner	≥24,000	0.90	5.67	≥20,230	0.89	5.45
English Learner	≥930	0.86	4.88	≥590	0.89	4.72
<b>Migrant Status</b>						
Migrant	≥30	0.86	5.85	≥10	0.94	5.37
Not Migrant	≥24,890	0.90	5.64	≥20,810	0.90	5.43
<b>Section 504 Status</b>						
Non-Section 504	≥22,430	0.90	5.65	≥19,330	0.90	5.44
Section 504	≥2,490	0.87	5.42	≥1,490	0.89	5.31
<b>Homeless Status</b>						
Not Homeless	≥24,500	0.90	5.64	≥20,490	0.90	5.43
Homeless	≥430	0.89	5.33	≥330	0.87	5.27
<b>Military Affiliation</b>						
Not Military Affiliated	≥24,590	0.90	5.64	≥20,490	0.90	5.43
Military Affiliated	≥330	0.88	5.80	≥340	0.88	5.47
<b>Foster Care Status</b>						
Not in Foster Care	≥24,850	0.90	5.64	≥20,770	0.90	5.43
Foster Care	≥430	0.89	5.33	≥330	0.87	5.27

Table 10.8 Fall 2018 Administration English II Reliability and SEM by Subgroup

Group	N Count	Cronbach's Alpha	SEM
<b>All Students</b>	≥5,500	0.90	5.43
<b>Gender</b>			
Female	≥2,690	0.89	5.39
Male	≥2,810	0.90	5.35
<b>Ethnicity</b>			
Hispanic/Latino	≥500	0.90	5.42
American Indian or Alaska Native	≥30	0.91	5.04
Asian	≥120	0.91	5.62
Black or African American	≥2,270	0.87	5.38
Native Hawaiian or Other Pacific	<10	NR	NR
White	≥2,460	0.89	5.43
Two or More Races	≥100	0.89	5.43
<b>Education Classification</b>			
Regular Education	≥4,810	0.89	5.41
Special Education	≥340	0.85	5.11
Gifted or Talented	≥340	0.87	5.53
<b>Economic Status</b>			
Economically Disadvantaged	<10	NR	NR
Not Economically Disadvantaged	≥3,950	0.89	5.43
<b>English Learner Status</b>			
Not English Learner	≥5,290	0.89	5.44
English Learner	≥200	0.82	5.01
<b>Migrant Status</b>			
Migrant	≥10	0.90	5.11
Not Migrant	≥5,490	0.90	5.43
<b>Section 504 Status</b>			
Non-Section 504	≥5,090	0.89	5.44
Section 504	≥410	0.89	5.16
<b>Homeless Status</b>			
Not Homeless	≥5,370	0.89	5.49
Homeless	≥120	0.87	5.36
<b>Military Affiliation</b>			
Not Military Affiliated	≥5,380	0.89	5.43
Military Affiliated	≥120	0.85	6.25
<b>Foster Care Status</b>			
Not in Foster Care	≥5,460	0.89	5.47
Foster Care	≥120	0.87	5.36

Table 10.9 Spring 2019 Administration English II Reliability and SEM by Subgroup

Group	Form D			Form E		
	N Count	Cronbach's Alpha	SEM	N Count	Cronbach's Alpha	SEM
<b>All Students</b>	≥21,770	0.90	5.55	≥18,920	0.90	5.32
<b>Gender</b>						
Female	≥10,880	0.89	5.50	≥9,840	0.89	5.27
Male	≥10,880	0.90	5.48	≥9,070	0.90	5.29
<b>Ethnicity</b>						
Hispanic/Latino	≥1,230	0.91	5.49	≥1,110	0.91	5.34
American Indian or Alaska Native	≥130	0.88	5.72	≥130	0.87	5.01
Asian	≥400	0.89	5.64	≥370	0.90	5.19
Black or African American	≥9,340	0.88	5.44	≥7,630	0.88	5.33
Native Hawaiian or Other Pacific	≥10	0.89	5.53	≥10	0.85	5.65
White	≥10,250	0.88	5.62	≥9,290	0.88	5.28
Two or More Races	≥370	0.89	5.51	≥360	0.89	5.34
<b>Education Classification</b>						
Regular Education	≥17,890	0.88	5.57	≥17,030	0.89	5.33
Special Education	≥2,480	0.85	4.94	≥490	0.90	5.16
Gifted or Talented	≥1,390	0.86	5.48	≥1,390	0.86	5.03
<b>Economic Status</b>						
Economically Disadvantaged	≥13,010	0.89	5.50	≥10,770	0.89	5.34
Not Economically Disadvantaged	≥7,390	0.88	5.57	≥6,910	0.87	5.25
<b>English Learner Status</b>						
Not English Learner	≥21,270	0.90	5.55	≥18,550	0.89	5.32
English Learner	≥490	0.85	5.03	≥370	0.87	4.88
<b>Migrant Status</b>						
Migrant	≥10	0.91	4.80	≥10	0.85	5.56
Not Migrant	≥21,750	0.90	5.55	≥18,910	0.90	5.32
<b>Section 504 Status</b>						
Non-Section 504	≥19,830	0.90	5.55	≥17,720	0.89	5.32
Section 504	≥1,930	0.88	5.39	≥1,200	0.90	5.27
<b>Homeless Status</b>						
Not Homeless	≥21,480	0.90	5.55	≥18,660	0.90	5.32
Homeless	≥280	0.89	5.34	≥260	0.88	5.27
<b>Military Affiliation</b>						
Not Military Affiliated	≥21,520	0.90	5.55	≥18,630	0.90	5.33
Military Affiliated	≥240	0.87	5.57	≥290	0.87	5.19
<b>Foster Care Status</b>						
Not in Foster Care	≥21,710	0.90	5.55	≥18,880	0.90	5.32
Foster Care	≥280	0.89	5.34	≥260	0.88	5.27

Table 10.10 Fall 2018 Administration Algebra I Reliability and SEM by Subgroup

Group	N Count	Cronbach's Alpha	SEM
<b>All Students</b>	≥3,130	0.86	3.68
<b>Gender</b>			
Female	≥1,500	0.85	3.68
Male	≥1,630	0.86	3.67
<b>Ethnicity</b>			
Hispanic/Latino	≥340	0.85	3.57
American Indian or Alaska Native	≥10	0.71	3.95
Asian	≥60	0.89	3.99
Black or African American	≥1,280	0.80	3.42
Native Hawaiian or Other Pacific	<10	NR	NR
White	≥1,350	0.85	3.79
Two or More Races	≥60	0.84	3.65
<b>Education Classification</b>			
Regular Education	≥2,740	0.85	3.66
Special Education	≥230	0.79	3.18
Gifted or Talented	≥150	0.85	3.97
<b>Economic Status</b>			
Economically Disadvantaged	<10	NR	NR
Not Economically Disadvantaged	≥1,950	0.86	3.73
<b>English Learner Status</b>			
Not English Learner	≥2,950	0.85	3.69
English Learner	≥180	0.80	3.16
<b>Migrant Status</b>			
Migrant	<10	NR	NR
Not Migrant	≥3,130	0.86	3.68
<b>Section 504 Status</b>			
Non-Section 504	≥2,860	0.86	3.69
Section 504	≥270	0.83	3.43
<b>Homeless Status</b>			
Not Homeless	≥3,060	0.90	3.15
Homeless	≥70	0.88	2.97
<b>Military Affiliation</b>			
Not Military Affiliated	≥3,110	0.89	3.15
Military Affiliated	≥10	0.88	3.16
<b>Foster Care Status</b>			
Not in Foster Care	≥3,130	0.90	3.15
Foster Care	≥70	0.88	2.97

Table 10.11 Spring 2019 Administration Algebra I Reliability and SEM by Subgroup

Group	Form D			Form E		
	N Count	Cronbach's Alpha	SEM	N Count	Cronbach's Alpha	SEM
<b>All Students</b>	≥25,510	0.90	3.74	≥20,850	0.91	3.78
<b>Gender</b>						
Female	≥12,560	0.90	3.77	≥10,710	0.90	3.79
Male	≥12,950	0.91	3.70	≥10,130	0.91	3.75
<b>Ethnicity</b>						
Hispanic/Latino	≥1,880	0.90	3.60	≥1,200	0.90	3.74
American Indian or Alaska Native	≥180	0.90	3.72	≥130	0.88	3.72
Asian	≥410	0.93	4.30	≥390	0.93	4.31
Black or African American	≥11,020	0.86	3.41	≥8,570	0.87	3.41
Native Hawaiian or Other Pacific	≥10	0.92	3.97	≥10	0.95	3.59
White	≥11,480	0.91	3.94	≥10,120	0.91	3.95
Two or More Races	≥500	0.91	3.82	≥410	0.92	3.82
<b>Education Classification</b>						
Regular Education	≥20,620	0.89	3.73	≥18,870	0.90	3.71
Special Education	≥3,370	0.80	3.01	≥470	0.90	3.47
Gifted or Talented	≥1,520	0.92	4.33	≥1,500	0.92	4.25
<b>Economic Status</b>						
Economically Disadvantaged	≥15,650	0.88	3.53	≥12,030	0.88	3.58
Not Economically Disadvantaged	≥7,450	0.91	4.06	≥6,750	0.91	4.03
<b>English Learner Status</b>						
Not English Learner	≥24,500	0.90	3.75	≥20,440	0.91	3.78
English Learner	≥1,010	0.87	3.19	≥400	0.87	3.23
<b>Migrant Status</b>						
Migrant	≥20	0.93	3.82	≥10	0.89	3.76
Not Migrant	≥25,490	0.90	3.74	≥20,830	0.91	3.78
<b>Section 504 Status</b>						
Non-Section 504	≥23,010	0.91	3.77	≥19,400	0.91	3.79
Section 504	≥2,500	0.86	3.39	≥1,440	0.89	3.49
<b>Homeless Status</b>						
Not Homeless	≥25,110	0.90	3.74	≥20,530	0.91	3.78
Homeless	≥400	0.85	3.33	≥310	0.83	3.33
<b>Military Affiliation</b>						
Not Military Affiliated	≥25,140	0.90	3.73	≥20,510	0.91	3.77
Military Affiliated	≥370	0.91	3.99	≥330	0.91	4.03
<b>Foster Care Status</b>						
Not in Foster Care	≥25,430	0.90	3.74	≥20,800	0.91	3.78
Foster Care	≥400	0.85	3.33	≥310	0.83	3.33

Table 10.12 Fall 2018 Administration Geometry Reliability and SEM by Subgroup

Group	N Count	Cronbach's Alpha	SEM
<b>All Students</b>	≥4,730	0.92	3.82
<b>Gender</b>			
Female	≥2,430	0.91	3.86
Male	≥2,300	0.92	3.77
<b>Ethnicity</b>			
Hispanic/Latino	≥460	0.92	3.71
American Indian or Alaska Native	≥20	0.93	4.03
Asian	≥140	0.93	4.33
Black or African American	≥1,980	0.89	3.35
Native Hawaiian or Other Pacific	<10	NR	NR
White	≥2,010	0.91	4.05
Two or More Races	≥90	0.89	4.06
<b>Education Classification</b>			
Regular Education	≥4,010	0.91	3.73
Special Education	≥240	0.91	3.07
Gifted or Talented	≥470	0.91	4.32
<b>Economic Status</b>			
Economically Disadvantaged	<10	NR	NR
Not Economically Disadvantaged	≥3,650	0.92	3.86
<b>English Learner Status</b>			
Not English Learner	≥4,550	0.92	3.84
English Learner	≥180	0.88	2.97
<b>Migrant Status</b>			
Migrant	<10	NR	NR
Not Migrant	≥4,730	0.92	3.82
<b>Section 504 Status</b>			
Non-Section 504	≥4,420	0.92	3.84
Section 504	≥310	0.93	3.37
<b>Homeless Status</b>			
Not Homeless	≥4,640	0.92	3.83
Homeless	≥80	0.86	3.03
<b>Military Affiliation</b>			
Not Military Affiliated	≥4,640	0.92	3.81
Military Affiliated	≥80	0.88	4.16
<b>Foster Care Status</b>			
Not in Foster Care	≥4,700	0.92	3.81
Foster Care	≥80	0.86	3.03

Table 10.13 Spring 2019 Administration Geometry Reliability and SEM by Subgroup

Group	Form D			Form E		
	N Count	Cronbach's Alpha	SEM	N Count	Cronbach's Alpha	SEM
<b>All Students</b>	≥18,640	0.91	3.66	≥16,630	0.91	3.68
<b>Gender</b>						
Female	≥9,780	0.91	3.69	≥8,890	0.91	3.70
Male	≥8,850	0.92	3.62	≥7,730	0.92	3.64
<b>Ethnicity</b>						
Hispanic/Latino	≥1,220	0.91	3.60	≥870	0.90	3.70
American Indian or Alaska Native	≥90	0.92	3.88	≥90	0.89	3.86
Asian	≥360	0.93	4.20	≥360	0.93	4.20
Black or African American	≥7,750	0.86	3.22	≥6,600	0.87	3.26
Native Hawaiian or Other Pacific	≥20	0.93	3.78	≥10	0.93	4.09
White	≥8,840	0.91	3.89	≥8,350	0.91	3.85
Two or More Races	≥330	0.91	3.78	≥320	0.91	3.76
<b>Education Classification</b>						
Regular Education	≥15,800	0.90	3.63	≥15,040	0.90	3.61
Special Education	≥1,480	0.83	2.86	≥270	0.89	3.30
Gifted or Talented	≥1,350	0.92	4.17	≥1,320	0.91	4.13
<b>Economic Status</b>						
Economically Disadvantaged	≥10,610	0.89	3.41	≥9,070	0.89	3.45
Not Economically Disadvantaged	≥6,960	0.92	3.93	≥6,540	0.91	3.91
<b>English Learner Status</b>						
Not English Learner	≥18,170	0.91	3.67	≥16,410	0.91	3.68
English Learner	≥470	0.87	3.17	≥210	0.91	3.24
<b>Migrant Status</b>						
Migrant	≥20	0.94	3.51	<10	NR	NR
Not Migrant	≥18,620	0.91	3.66	≥16,620	0.91	3.68
<b>Section 504 Status</b>						
Non-Section 504	≥17,160	0.91	3.68	≥15,750	0.91	3.69
Section 504	≥1,480	0.89	3.31	≥870	0.90	3.39
<b>Homeless Status</b>						
Not Homeless	≥18,400	0.91	3.66	≥16,430	0.91	3.68
Homeless	≥240	0.89	3.24	≥190	0.87	3.12
<b>Military Affiliation</b>						
Not Military Affiliated	≥18,390	0.91	3.66	≥16,360	0.91	3.67
Military Affiliated	≥250	0.90	3.85	≥260	0.91	3.87
<b>Foster Care Status</b>						
Not in Foster Care	≥18,600	0.91	3.66	≥16,600	0.91	3.68
Foster Care	≥240	0.89	3.24	≥190	0.87	3.12



## 10.5 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 and is given by the following formula:

$$d = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{(n_a + n_b) - 2}}},$$

where  $\bar{x}_a$  is the mean score of group A,  $\bar{x}_b$  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  $d = 0.34$  for two groups, then it may be interpreted that the SMD between the two groups is 0.34 of the pooled standard deviation. Cohen (1988) offered guidelines for interpreting the meaning of the  $d$  statistic:  $d = 0.20$  is a small ES,  $d = 0.50$  is a medium ES, and  $d = 0.80$  is a large ES.

Using Cohen's (1988) guidelines, certain trends become apparent in Tables 10.14–10.21. Results are NR for subgroups with fewer than 10 students. For all subjects across both the fall and spring administrations, small differences in test scores were seen between females and males, with females slightly outperforming males. Mean scale scores and ESs show that Asian and white students tend to outperform other ethnicity groups across subjects. For most ELA and mathematics tests, there were clear performance differences between regular education and special education students in Education Classification and Not English Learner and English Learner in EL status.

Table 10.14 Fall 2018 Administration Impact Analysis: English I

Group	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
<b>All Students</b>	≥6,680	731.26	39.90	
<b>Gender</b>				
Male	≥3,810	723.18	38.74	
Female	≥2,870	741.97	38.88	-0.48
<b>Ethnicity</b>				
White	≥2,740	748.94	38.17	
Hispanic/Latino	≥640	722.64	39.43	0.68
American Indian or Alaska Native	≥20	748.19	37.10	0.02
Asian	≥100	749.83	38.55	-0.02
Black or African American	≥3,030	715.86	34.50	0.91
Native Hawaiian or Other Pacific	<10	NR	NR	NR
Two or More Races	≥130	740.78	36.12	0.21
<b>Education Classification</b>				
Regular Education	≥5,360	734.13	37.63	
Special Education	≥950	697.13	27.31	1.02
Gifted or Talented	≥360	778.91	31.06	-1.20
<b>Economic Status</b>				
Not Economically Disadvantaged	≥3,130	743.53	37.75	
Economically Disadvantaged	<10	NR	NR	NR
<b>English Learner Status</b>				
Not English Learner	≥6,320	733.24	39.71	
English Learner	≥360	696.84	24.61	0.93
<b>Migrant Status</b>				
Nonmigrant	≥6,680	731.27	39.89	
Migrant	<10	NR	NR	NR
<b>Section 504 Status</b>				
Non-Section 504	≥5,970	733.29	40.14	
Section 504	≥710	714.27	33.28	0.48
<b>Homeless Status</b>				
Not Homeless	≥6,540	731.68	39.98	
Homeless	≥140	711.49	30.09	0.51
<b>Military Affiliation</b>				
Not Military Affiliated	≥6,490	730.51	39.90	
Military Affiliated	≥190	756.84	30.30	-0.66
<b>Foster Care Status</b>				
Not in Foster Care	≥6,640	731.21	39.94	
Foster Care	≥40	737.90	33.37	-0.17

Table 10.15 Spring 2019 Administration Impact Analysis: English I

Group	Form D				Form E			
	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
<b>All Students</b>	≥25,850	737.41	36.91		≥21,390	747.16	33.38	
<b>Gender</b>								
Male	≥13,360	730.54	36.82		≥10,400	741.61	33.38	
Female	≥12,480	744.76	35.57	-0.39	≥10,980	752.42	32.51	-0.33
<b>Ethnicity</b>								
White	≥11,320	750.10	34.97		≥9,860	758.19	30.80	
Hispanic/Latino	≥1,770	726.66	41.29	0.65	≥1,410	739.39	38.30	0.59
American Indian or Alaska Native	≥170	740.99	32.04	0.26	≥150	756.11	28.54	0.07
Asian	≥450	767.42	38.73	-0.49	≥370	772.05	33.86	-0.45
Black or African American	≥11,590	725.01	32.91	0.74	≥9,150	734.82	30.30	0.76
Native Hawaiian or Other Pacific	≥10	761.50	48.78	-0.33	≥10	743.78	34.38	0.47
Two or More Races	≥500	746.40	35.07	0.11	≥400	757.52	33.14	0.02
<b>Education Classification</b>								
Regular Education	≥20,680	740.44	33.73		≥19,330	745.51	31.89	
Special Education	≥3,630	702.81	29.16	1.14	≥650	721.63	33.56	0.75
Gifted or Talented	≥1,520	778.71	30.76	-1.14	≥1,400	781.91	30.00	-1.15
<b>Economic Status</b>								
Not Economically Disadvantaged	≥7,410	757.25	33.87		≥6,740	763.68	30.53	
Economically Disadvantaged	≥16,140	729.23	34.75	0.81	≥12,740	739.30	31.34	0.78
<b>English Learner Status</b>								
Not English Learner	≥24,970	738.76	36.39		≥20,860	748.21	32.78	
English Learner	≥870	698.83	30.37	1.10	≥520	705.74	30.23	1.30
<b>Migrant Status</b>								
Nonmigrant	≥25,810	737.44	36.91		≥21,370	747.16	33.37	
Migrant	≥40	719.68	32.36	0.48	≥20	741.50	43.52	0.17
<b>Section 504 Status</b>								
Non-Section 504	≥23,140	739.26	37.04		≥19,770	748.29	33.18	
Section 504	≥2,700	721.59	31.70	0.48	≥1,610	733.35	32.63	0.45
<b>Homeless Status</b>								
Not Homeless	≥25,390	737.68	36.91		≥21,030	747.46	33.36	
Homeless	≥450	722.56	34.03	0.41	≥350	729.25	29.26	0.55
<b>Military Affiliation</b>								
Not Military Affiliated	≥25,510	737.14	36.88		≥21,040	746.93	33.37	
Military Affiliated	≥340	757.75	33.49	-0.56	≥350	760.84	30.77	-0.42
<b>Foster Care Status</b>								
Not in Foster Care	≥25,760	737.47	36.90		≥21,320	747.19	33.38	
Foster Care	≥80	718.73	35.12	0.51	≥60	737.23	29.64	0.30

Table 10.16 Fall 2018 Administration Impact Analysis: English II

Group	Fall 2018			
	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
<b>All Students</b>	≥9,590	723.34	46.39	
<b>Gender</b>				
Male	≥5,550	714.79	43.98	
Female	≥4,040	735.11	47.05	-0.45
<b>Ethnicity</b>				
White	≥3,390	743.76	48.78	
Hispanic/Latino	≥940	712.30	44.00	0.66
American Indian or Alaska Native	≥50	730.40	42.69	0.27
Asian	≥170	752.94	58.19	-0.19
Black or African American	≥4,860	709.74	38.19	0.79
Native Hawaiian or Other Pacific	≥10	728.70	46.15	0.31
Two or More Races	≥150	735.15	48.17	0.18
<b>Education Classification</b>				
Regular Education	≥7,880	726.61	44.32	
Special Education	≥1,340	687.07	29.07	0.93
Gifted or Talented	≥360	785.67	44.99	-1.33
<b>Economic Status</b>				
Not Economically Disadvantaged	≥4,530	743.41	45.28	
Economically Disadvantaged	<10	NR	NR	NR
<b>English Learner Status</b>				
Not English Learner	≥8,930	725.92	46.46	
English Learner	≥660	688.94	28.27	0.81
<b>Migrant Status</b>				
Nonmigrant	≥9,580	723.31	46.37	
Migrant	≥10	741.63	54.21	-0.39
<b>Section 504 Status</b>				
Non-Section 504	≥8,520	725.85	46.94	
Section 504	≥1,070	703.52	36.12	0.49
<b>Homeless Status</b>				
Not Homeless	≥9,360	723.65	46.54	
Homeless	≥230	711.26	38.02	0.27
<b>Military Affiliation</b>				
Not Military Affiliated	≥9,470	722.69	46.07	
Military Affiliated	≥120	771.40	44.50	-1.06
<b>Foster Care Status</b>				
Not in Foster Care	≥9,540	723.19	46.31	
Foster Care	≥50	750.31	53.69	-0.59

Table 10.17 Spring 2019 Administration Impact Analysis: English II

Group	Form D				Form E			
	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
<b>All Students</b>	≥22,840	738.07	45.64		≥19,610	748.12	41.91	
<b>Gender</b>								
Male	≥11,670	728.66	45.32		≥9,570	741.04	42.83	
Female	≥11,170	747.92	43.86	-0.43	≥10,040	754.87	39.86	-0.33
<b>Ethnicity</b>								
White	≥10,480	753.12	42.92		≥9,460	761.94	38.35	
Hispanic/Latino	≥1,350	729.28	48.52	0.55	≥1,180	738.98	45.99	0.58
American Indian or Alaska Native	≥130	747.06	42.23	0.14	≥130	760.58	34.81	0.04
Asian	≥400	773.85	44.86	-0.48	≥370	776.17	42.50	-0.37
Black or African American	≥10,050	721.45	41.65	0.75	≥8,050	731.28	38.52	0.80
Native Hawaiian or Other Pacific	≥10	771.94	38.22	-0.44	≥20	755.25	38.91	0.17
Two or More Races	≥390	750.76	42.42	0.05	≥370	757.10	42.44	0.13
<b>Education Classification</b>								
Regular Education	≥18,540	741.52	41.95		≥17,640	745.82	39.98	
Special Education	≥2,890	692.93	35.02	1.18	≥560	713.21	42.10	0.81
Gifted or Talented	≥1,400	785.63	38.73	-1.06	≥1,400	791.15	36.45	-1.14
<b>Economic Status</b>								
Not Economically Disadvantaged	≥7,560	758.17	41.84		≥7,040	765.03	37.90	
Economically Disadvantaged	≥13,970	726.77	43.09	0.74	≥11,340	737.76	39.65	0.70
<b>English Learner Status</b>								
Not English Learner	≥22,270	739.30	45.25		≥19,180	749.24	41.37	
English Learner	≥570	690.33	33.71	1.09	≥420	698.07	34.78	1.24
<b>Migrant Status</b>								
Nonmigrant	≥22,830	738.08	45.64		≥19,590	748.13	41.92	
Migrant	≥10	729.42	44.42	0.19	≥10	738.20	34.78	0.24
<b>Section 504 Status</b>								
Non-Section 504	≥20,670	740.42	45.42		≥18,280	749.64	41.55	
Section 504	≥2,160	715.67	41.43	0.55	≥1,320	727.21	41.31	0.54
<b>Homeless Status</b>								
Not Homeless	≥22,520	738.40	45.61		≥19,340	748.38	41.92	
Homeless	≥320	714.98	41.93	0.51	≥270	729.98	37.17	0.44
<b>Military Affiliation</b>								
Not Military Affiliated	≥22,600	737.86	45.66		≥19,310	747.84	41.93	
Military Affiliated	≥240	758.00	38.64	-0.44	≥290	766.37	36.34	-0.44
<b>Foster Care Status</b>								
Not in Foster Care	≥22,780	738.14	45.63		≥19,560	748.14	41.93	
Foster Care	≥50	711.44	39.70	0.59	≥40	739.00	35.47	0.22

Table 10.18 Fall 2018 Administration Impact Analysis: Algebra I

Group	Fall 2018			
	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
<b>All Students</b>	≥5,670	724.88	32.88	
<b>Gender</b>				
Male	≥3,080	723.61	32.03	
Female	≥2,590	726.38	33.80	-0.08
<b>Ethnicity</b>				
White	≥1,890	740.91	34.26	
Hispanic/Latino	≥610	719.11	31.25	0.65
American Indian or Alaska Native	≥20	750.95	33.01	-0.29
Asian	≥80	754.43	39.86	-0.39
Black or African American	≥2,960	714.45	26.77	0.88
Native Hawaiian or Other Pacific	<10	NR	NR	NR
Two or More Races	≥100	735.39	33.30	0.16
<b>Education Classification</b>				
Regular Education	≥4,690	727.49	32.25	
Special Education	≥810	702.65	23.51	0.80
Gifted or Talented	≥170	757.07	33.38	-0.92
<b>Economic Status</b>				
Not Economically Disadvantaged	≥2,470	736.94	33.94	
Economically Disadvantaged	<10	NR	NR	NR
<b>English Learner Status</b>				
Not English Learner	≥5,250	726.37	33.04	
English Learner	≥420	706.25	23.91	0.62
<b>Migrant Status</b>				
Nonmigrant	≥5,670	724.85	32.88	
Migrant	<10	NR	NR	NR
<b>Section 504 Status</b>				
Non-Section 504	≥5,000	726.55	33.31	
Section 504	≥670	712.39	26.31	0.43
<b>Homeless Status</b>				
Not Homeless	≥5,530	725.16	33.00	
Homeless	≥140	714.54	26.01	0.32
<b>Military Affiliation</b>				
Not Military Affiliated	≥5,650	724.77	32.86	
Military Affiliated	≥20	746.46	30.09	-0.66
<b>Foster Care Status</b>				
Not in Foster Care	≥5,660	724.91	32.88	
Foster Care	≥10	708.40	28.73	0.50

Table 10.19 Spring 2019 Administration Impact Analysis: Algebra I

Group	Form D				Form E			
	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
<b>All Students</b>	≥26,840	738.15	34.92		≥21,640	744.22	34.29	
<b>Gender</b>								
Male	≥13,740	736.00	35.25		≥10,590	742.52	35.08	
Female	≥13,100	740.41	34.42	-0.13	≥11,040	745.86	33.42	-0.10
<b>Ethnicity</b>								
White	≥11,850	749.82	35.41		≥10,330	754.83	33.66	
Hispanic/Latino	≥1,920	732.33	34.54	0.50	≥1,230	741.72	33.69	0.39
American Indian or Alaska Native	≥180	742.16	33.53	0.22	≥140	744.25	29.75	0.31
Asian	≥410	772.46	41.03	-0.64	≥380	782.92	40.05	-0.83
Black or African American	≥11,920	726.04	29.04	0.73	≥9,090	730.68	28.89	0.77
Native Hawaiian or Other Pacific	≥10	741.42	39.35	0.24	≥10	740.60	50.80	0.42
Two or More Races	≥530	742.38	35.54	0.21	≥430	748.33	35.55	0.19
<b>Education Classification</b>								
Regular Education	≥21,520	740.05	32.32		≥19,590	741.87	32.52	
Special Education	≥3,760	709.99	24.63	0.96	≥500	728.02	32.69	0.43
Gifted or Talented	≥1,560	779.85	37.46	-1.22	≥1,540	779.51	36.19	-1.15
<b>Economic Status</b>								
Not Economically Disadvantaged	≥7,710	756.15	36.15		≥6,950	759.92	34.70	
Economically Disadvantaged	≥16,800	729.97	31.00	0.80	≥12,750	735.88	30.84	0.75
<b>English Learner Status</b>								
Not English Learner	≥25,860	739.00	34.86		≥21,270	744.67	34.22	
English Learner	≥970	715.64	28.31	0.67	≥360	718.30	27.74	0.77
<b>Migrant Status</b>								
Nonmigrant	≥26,810	738.15	34.91		≥21,620	744.22	34.29	
Migrant	≥20	736.71	38.84	0.04	≥10	745.56	31.38	-0.04
<b>Section 504 Status</b>								
Non-Section 504	≥24,100	739.75	35.19		≥20,070	745.32	34.19	
Section 504	≥2,740	724.08	28.86	0.45	≥1,570	730.21	32.41	0.44
<b>Homeless Status</b>								
Not Homeless	≥26,410	738.41	34.95		≥21,290	744.49	34.34	
Homeless	≥430	722.23	28.47	0.46	≥350	728.06	26.03	0.48
<b>Military Affiliation</b>								
Not Military Affiliated	≥26,450	737.91	34.83		≥21,300	743.93	34.23	
Military Affiliated	≥380	754.67	37.14	-0.48	≥340	762.32	33.25	-0.54
<b>Foster Care Status</b>								
Not in Foster Care	≥26,750	738.20	34.92		≥21,580	744.24	34.30	
Foster Care	≥80	722.90	29.17	0.44	≥50	737.40	29.63	0.20

Table 10.20 Fall 2018 Administration Impact Analysis: Geometry

Group	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
<b>All Students</b>	≥5,350	733.72	28.31	
<b>Gender</b>				
Male	≥2,670	732.15	28.93	
Female	≥2,680	735.28	27.60	-0.11
<b>Ethnicity</b>				
White	≥2,090	746.72	25.67	
Hispanic/Latino	≥530	727.42	29.54	0.73
American Indian or Alaska Native	≥20	737.48	30.96	0.36
Asian	≥140	754.90	31.44	-0.31
Black or African American	≥2,450	722.36	24.32	0.98
Native Hawaiian or Other Pacific	<10	NR	NR	NR
Two or More Races	≥90	742.92	24.61	0.15
<b>Education Classification</b>				
Regular Education	≥4,530	732.63	26.49	
Special Education	≥340	709.53	24.83	0.88
Gifted or Talented	≥480	760.92	26.53	-1.07
<b>Economic Status</b>				
Not Economically Disadvantaged	≥3,840	737.51	27.62	
Economically Disadvantaged	<10	NR	NR	NR
<b>English Learner Status</b>				
Not English Learner	≥5,110	734.97	27.99	
English Learner	≥240	707.83	21.72	0.98
<b>Migrant Status</b>				
Nonmigrant	≥5,350	733.69	28.30	
Migrant	<10	NR	NR	NR
<b>Section 504 Status</b>				
Non-Section 504	≥4,970	734.74	28.10	
Section 504	≥380	720.49	27.74	0.51
<b>Homeless Status</b>				
Not Homeless	≥5,250	734.05	28.33	
Homeless	≥100	717.08	21.46	0.60
<b>Military Affiliation</b>				
Not Military Affiliated	≥5,260	733.35	28.28	
Military Affiliated	≥90	755.52	20.88	-0.79
<b>Foster Care Status</b>				
Not in Foster Care	≥5,320	733.60	28.27	
Foster Care	≥30	754.03	28.05	-0.72



Table 10.21 Spring 2019 Administration Impact Analysis: Geometry

Group	Form D				Form E			
	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
<b>All Students</b>	≥18,820	737.09	26.75		≥16,710	739.61	26.00	
<b>Gender</b>								
Male	≥8,980	736.82	27.22		≥7,770	739.56	26.47	
Female	≥9,830	737.33	26.31	-0.02	≥8,930	739.66	25.58	0.00
<b>Ethnicity</b>								
White	≥8,860	747.08	25.59		≥8,340	748.09	24.79	
Hispanic/Latino	≥1,220	734.41	25.90	0.49	≥880	740.10	24.54	0.32
American Indian or Alaska Native	≥90	746.19	26.33	0.03	≥90	746.62	22.78	0.06
Asian	≥360	761.99	31.69	-0.58	≥360	764.38	30.23	-0.65
Black or African American	≥7,890	724.78	21.97	0.93	≥6,680	727.32	21.86	0.88
Native Hawaiian or Other Pacific	≥20	746.70	30.10	0.01	≥10	753.53	31.61	-0.22
Two or More Races	≥330	741.42	26.13	0.22	≥320	743.27	25.19	0.19
<b>Education Classification</b>								
Regular Education	≥15,910	736.92	25.14		≥15,110	737.58	24.74	
Special Education	≥1,530	715.10	20.40	0.88	≥270	726.05	24.80	0.47
Gifted or Talented	≥1,360	763.74	27.75	-1.06	≥1,320	765.61	25.61	-1.13
<b>Economic Status</b>								
Not Economically Disadvantaged	≥7,020	748.29	26.70		≥6,600	749.99	25.59	
Economically Disadvantaged	≥10,860	729.75	24.03	0.74	≥9,230	732.37	23.68	0.72
<b>English Learner Status</b>								
Not English Learner	≥18,360	737.49	26.71		≥16,510	739.83	25.95	
English Learner	≥450	720.60	22.58	0.63	≥190	721.70	23.67	0.70
<b>Migrant Status</b>								
Nonmigrant	≥18,790	737.09	26.74		≥16,710	739.61	26.00	
Migrant	≥20	737.04	31.98	0.00	<10	NR	NR	NR
<b>Section 504 Status</b>								
Non-Section 504	≥17,300	738.06	26.74		≥15,810	740.13	25.97	
Section 504	≥1,510	725.95	24.19	0.46	≥900	730.61	24.83	0.37
<b>Homeless Status</b>								
Not Homeless	≥18,560	737.25	26.76		≥16,510	739.80	25.99	
Homeless	≥250	724.73	22.62	0.47	≥200	724.29	22.03	0.60
<b>Military Affiliation</b>								
Not Military Affiliated	≥18,560	736.93	26.76		≥16,440	739.44	25.98	
Military Affiliated	≥250	748.49	23.25	-0.43	≥260	750.01	25.24	-0.41
<b>Foster Care Status</b>								
Not in Foster Care	≥18,780	737.11	26.75		≥16,690	739.63	25.99	
Foster Care	≥30	727.18	20.62	0.37	≥20	724.83	27.33	0.57

Additional data for scale score means are provided in Tables 10.22 and 10.25. These tables report the number of students, scale score means, and standard deviations for each Special Education Classification. Groups that have fewer than 10 students are not reported (NR) in the tables.

**Table 10.22 Special Education Classification Scale Score Means and Standard Deviations: English I**

Admin.	Form	Group	Yes			No		
			N	Mean	Std. Dev.	N	Mean	Std. Dev.
Fall 2018	B	Gifted	≥160	793.26	25.59	≥6,520	729.74	38.97
		Talented	≥200	767.60	30.35	≥6,480	730.12	39.63
		Autism	<50	NR	NR	≥6,650	731.40	39.90
		Deaf-Blindness	<50	NR	NR	≥6,680	731.26	39.90
		Developmental Delay	<50	NR	NR	≥6,680	731.26	39.90
		Emotional Disturbance	<50	NR	NR	≥6,650	731.42	39.89
		HI—Deaf	<50	NR	NR	≥6,680	731.29	39.90
		HI—Hard-of-Hearing	<50	NR	NR	≥6,680	731.30	39.90
		Mild Mental Disability	≥60	679.97	17.73	≥6,620	731.76	39.73
		Moderate Mental Disability	<50	NR	NR	≥6,680	731.26	39.90
		Orthopedic Impairment	<50	NR	NR	≥6,670	731.28	39.89
		Other Health Impairment	≥190	702.92	28.24	≥6,490	732.10	39.89
		Specific Learning Disability	≥540	694.36	25.61	≥6,140	734.56	39.28
		Speech or Language Impairment	<50	NR	NR	≥6,650	731.34	39.93
		Traumatic Brain Injury	<50	NR	NR	≥6,680	731.27	39.90
		Visual Impairment	<50	NR	NR	≥6,670	731.28	39.91
Other	<50	NR	NR	≥6,680	731.27	39.90		
Spring 2019	D	Gifted	≥740	790.92	25.17	≥25,110	735.83	36.02
		Talented	≥780	767.24	31.13	≥25,060	736.47	36.69
		Autism	≥140	718.65	34.87	≥25,710	737.51	36.90
		Deaf-Blindness	<50	NR	NR	≥25,850	737.41	36.91
		Developmental Delay	<50	NR	NR	≥25,850	737.41	36.91
		Emotional Disturbance	≥110	704.37	31.68	≥25,730	737.56	36.87
		HI—Deaf	<50	NR	NR	≥25,840	737.42	36.91
		HI—Hard-of-Hearing	<50	NR	NR	≥25,810	737.43	36.92
		Mild Mental Disability	≥160	684.55	20.71	≥25,680	737.76	36.75
		Moderate Mental Disability	<50	NR	NR	≥25,850	737.42	36.91
		Orthopedic Impairment	<50	NR	NR	≥25,810	737.43	36.91
		Other Health Impairment	≥740	704.50	28.76	≥25,100	738.39	36.68
		Specific Learning Disability	≥2,210	700.48	27.28	≥23,630	740.88	35.78

Admin.	Form	Group	Yes			No		
			N	Mean	Std. Dev.	N	Mean	Std. Dev.
		Speech or Language Impairment	≥100	717.54	31.41	≥25,750	737.49	36.91
		Traumatic Brain Injury	<50	NR	NR	≥25,840	737.42	36.91
		Visual Impairment	<50	NR	NR	≥25,810	737.42	36.91
		Other	<50	NR	NR	≥25,850	737.41	36.91
	E	Gifted	≥660	795.22	24.38	≥20,720	745.61	32.46
		Talented	≥730	769.73	29.48	≥20,650	746.36	33.23
		Autism	<50	NR	NR	≥21,340	747.17	33.39
		Deaf-Blindness	<50	NR	NR	≥21,390	747.16	33.38
		Developmental Delay	<50	NR	NR	≥21,390	747.16	33.38
		Emotional Disturbance	<50	NR	NR	≥21,350	747.21	33.35
		HI—Deaf	<50	NR	NR	≥21,380	747.16	33.38
		HI—Hard-of-Hearing	<50	NR	NR	≥21,370	747.16	33.38
		Mild Mental Disability	<50	NR	NR	≥21,370	747.20	33.35
		Moderate Mental Disability	<50	NR	NR	≥21,390	747.16	33.37
		Orthopedic Impairment	<50	NR	NR	≥21,370	747.15	33.37
		Other Health Impairment	≥160	724.71	34.59	≥21,220	747.33	33.31
		Specific Learning Disability	≥300	713.16	29.40	≥21,080	747.65	33.18
		Speech or Language Impairment	≥50	730.95	32.85	≥21,330	747.20	33.37
		Traumatic Brain Injury	<50	NR	NR	≥21,380	747.16	33.38
		Visual Impairment	<50	NR	NR	≥21,380	747.16	33.38
Other	<50	NR	NR	≥21,390	747.16	33.38		
Summer 2019	A	Gifted	<50	NR	NR	≥1,900	699.19	20.23
		Talented	<50	NR	NR	≥1,890	699.15	20.20
		Autism	<50	NR	NR	≥1,880	699.24	20.22
		Deaf-Blindness	<50	NR	NR	≥1,900	699.19	20.23
		Developmental Delay	<50	NR	NR	≥1,890	699.20	20.23
		Emotional Disturbance	<50	NR	NR	≥1,880	699.28	20.16
		HI—Deaf	<50	NR	NR	≥1,890	699.23	20.20
		HI—Hard-of-Hearing	<50	NR	NR	≥1,890	699.20	20.24
		Mild Mental Disability	<50	NR	NR	≥1,850	699.39	20.29
		Moderate Mental Disability	<50	NR	NR	≥1,900	699.20	20.23
		Orthopedic Impairment	<50	NR	NR	≥1,890	699.24	20.24
		Other Health Impairment	≥100	695.80	21.07	≥1,790	699.39	20.17

Admin.	Form	Group	Yes			No		
			N	Mean	Std. Dev.	N	Mean	Std. Dev.
		Specific Learning Disability	≥320	693.29	17.15	≥1,570	700.42	20.61
		Speech or Language Impairment	<50	NR	NR	≥1,880	699.22	20.25
		Traumatic Brain Injury	<50	NR	NR	≥1,900	699.19	20.23
		Visual Impairment	<50	NR	NR	≥1,890	699.20	20.23
		Other	<50	NR	NR	≥1,890	699.18	20.24

**Table 10.23 Special Education Classification Scale Score Means and Standard Deviations: English II**

Special Education Classification Scale Score Means and Standard Deviations: English II								
Admin.	Form	Group	Yes			No		
			N	Mean	Std. Dev.	N	Mean	Std. Dev.
Fall 2018	B	Gifted	≥140	810.91	33.69	≥9,450	722.04	45.29
		Talented	≥220	770.07	44.06	≥9,370	722.21	45.86
		Autism	≥50	694.73	37.97	≥9,540	723.50	46.39
		Deaf-Blindness	<50	NR	NR	≥9,590	723.34	46.39
		Developmental Delay	<50	NR	NR	≥9,590	723.34	46.39
		Emotional Disturbance	<50	NR	NR	≥9,560	723.45	46.43
		HI—Deaf	<50	NR	NR	≥9,590	723.35	46.40
		HI—Hard-of-Hearing	<50	NR	NR	≥9,570	723.43	46.40
		Mild Mental Disability	≥80	673.54	21.13	≥9,510	723.77	46.32
		Moderate Mental Disability	<50	NR	NR	≥9,590	723.36	46.39
		Orthopedic Impairment	<50	NR	NR	≥9,580	723.40	46.37
		Other Health Impairment	≥250	686.10	29.85	≥9,340	724.38	46.34
		Specific Learning Disability	≥790	685.97	25.92	≥8,800	726.70	46.35
		Speech or Language Impairment	≥50	702.12	43.61	≥9,540	723.46	46.38
		Traumatic Brain Injury	<50	NR	NR	≥9,590	723.35	46.39
		Visual Impairment	<50	NR	NR	≥9,580	723.34	46.40
Other	<50	NR	NR	≥9,590	723.38	46.39		
Spring 2019	D	Gifted	≥720	801.34	31.12	≥22,120	736.01	44.55
		Talented	≥680	768.93	39.06	≥22,160	737.13	45.50
		Autism	≥140	715.81	45.36	≥22,700	738.21	45.61
		Deaf-Blindness	<50	NR	NR	≥22,840	738.07	45.64
		Developmental Delay	<50	NR	NR	≥22,840	738.07	45.64
		Emotional Disturbance	≥90	695.13	35.01	≥22,750	738.25	45.59
		HI—Deaf	<50	NR	NR	≥22,820	738.11	45.63

Special Education Classification Scale Score Means and Standard Deviations: English II										
Admin.	Form	Group	Yes			No				
			N	Mean	Std. Dev.	N	Mean	Std. Dev.		
		HI—Hard-of-Hearing	<50	NR	NR	≥22,790	738.14	45.63		
		Mild Mental Disability	≥130	669.71	19.92	≥22,710	738.47	45.45		
		Moderate Mental Disability	<50	NR	NR	≥22,840	738.08	45.64		
		Orthopedic Impairment	<50	NR	NR	≥22,800	738.12	45.63		
		Other Health Impairment	≥570	696.47	35.62	≥22,260	739.15	45.37		
		Specific Learning Disability	≥1,700	688.87	31.22	≥21,140	742.04	44.29		
		Speech or Language Impairment	≥100	712.60	41.10	≥22,740	738.18	45.63		
		Traumatic Brain Injury	<50	NR	NR	≥22,830	738.08	45.64		
		Visual Impairment	<50	NR	NR	≥22,820	738.08	45.64		
		Other	<50	NR	NR	≥22,840	738.07	45.64		
	E	Gifted	≥730	806.54	29.61	≥18,870	745.85	40.66		
		Talented	≥670	774.33	35.80	≥18,940	747.19	41.81		
		Autism	<50	NR	NR	≥19,560	748.17	41.89		
		Deaf-Blindness	<50	NR	NR	≥19,610	748.12	41.91		
		Developmental Delay	<50	NR	NR	≥19,610	748.12	41.91		
		Emotional Disturbance	<50	NR	NR	≥19,570	748.18	41.89		
		HI—Deaf	<50	NR	NR	≥19,610	748.12	41.91		
		HI—Hard-of-Hearing	<50	NR	NR	≥19,600	748.12	41.91		
		Mild Mental Disability	<50	NR	NR	≥19,590	748.18	41.87		
		Moderate Mental Disability	<50	NR	NR	≥19,610	748.12	41.91		
		Orthopedic Impairment	<50	NR	NR	≥19,590	748.11	41.92		
		Other Health Impairment	≥120	719.79	43.43	≥19,480	748.31	41.84		
		Specific Learning Disability	≥250	699.97	33.16	≥19,350	748.76	41.64		
		Speech or Language Impairment	<50	NR	NR	≥19,560	748.14	41.91		
		Traumatic Brain Injury	<50	NR	NR	≥19,610	748.12	41.91		
		Visual Impairment	<50	NR	NR	≥19,600	748.12	41.91		
		Other	<50	NR	NR	≥19,610	748.12	41.91		
		Summer 2019	B	Gifted	<50	NR	NR	≥1,690	688.74	25.63
				Talented	<50	NR	NR	≥1,680	688.67	25.68
				Autism	<50	NR	NR	≥1,680	688.77	25.64
Deaf-Blindness	<50			NR	NR	≥1,690	688.72	25.64		
Developmental Delay	<50			NR	NR	≥1,690	688.72	25.64		

Special Education Classification Scale Score Means and Standard Deviations: English II								
Admin.	Form	Group	Yes			No		
			N	Mean	Std. Dev.	N	Mean	Std. Dev.
		Emotional Disturbance	<50	NR	NR	≥1,670	688.86	25.67
		HI—Deaf	<50	NR	NR	≥1,690	688.76	25.64
		HI—Hard-of-Hearing	<50	NR	NR	≥1,690	688.79	25.62
		Mild Mental Disability	<50	NR	NR	≥1,660	688.98	25.67
		Moderate Mental Disability	<50	NR	NR	≥1,690	688.72	25.64
		Orthopedic Impairment	<50	NR	NR	≥1,690	688.73	25.65
		Other Health Impairment	≥70	683.81	24.89	≥1,620	688.94	25.66
		Specific Learning Disability	≥220	679.31	22.78	≥1,470	690.14	25.76
		Speech or Language Impairment	<50	NR	NR	≥1,690	688.71	25.65
		Traumatic Brain Injury	<50	NR	NR	≥1,690	688.70	25.65
		Visual Impairment	<50	NR	NR	≥1,690	688.71	25.65
		Other	<50	NR	NR	≥1,690	688.70	25.64

**Table 10.24 Special Education Classification Scale Score Means and Standard Deviations: Algebra I**

Special Education Classification Scale Score Means and Standard Deviations: Algebra I								
Admin.	Form	Group	Yes			No		
			N	Mean	Std. Dev.	N	Mean	Std. Dev.
Fall 2018	B	Gifted	<50	NR	NR	≥5,630	724.41	32.48
		Talented	≥130	748.15	29.64	≥5,540	724.31	32.75
		Autism	<50	NR	NR	≥5,650	724.90	32.87
		Deaf-Blindness	<50	NR	NR	≥5,670	724.88	32.88
		Developmental Delay	<50	NR	NR	≥5,670	724.88	32.88
		Emotional Disturbance	<50	NR	NR	≥5,640	724.99	32.90
		HI—Deaf	<50	NR	NR	≥5,670	724.89	32.88
		HI—Hard-of-Hearing	<50	NR	NR	≥5,660	724.90	32.89
		Mild Mental Disability	≥60	693.32	19.03	≥5,610	725.23	32.83
		Moderate Mental Disability	<50	NR	NR	≥5,670	724.88	32.88
		Orthopedic Impairment	<50	NR	NR	≥5,670	724.89	32.88
		Other Health Impairment	≥160	704.07	23.19	≥5,510	725.49	32.92
		Specific Learning Disability	≥460	701.60	22.01	≥5,210	726.94	32.89

Special Education Classification Scale Score Means and Standard Deviations: Algebra I									
Admin.	Form	Group	Yes			No			
			N	Mean	Std. Dev.	N	Mean	Std. Dev.	
		Speech or Language Impairment	<50	NR	NR	≥5,650	724.97	32.87	
		Traumatic Brain Injury	<50	NR	NR	≥5,670	724.88	32.88	
		Visual Impairment	<50	NR	NR	≥5,670	724.89	32.87	
		Other	<50	NR	NR	≥5,670	724.92	32.88	
Spring 2019	D	Gifted	≥750	798.92	31.75	≥26,080	736.39	33.40	
		Talented	≥800	761.92	33.35	≥26,040	737.42	34.71	
		Autism	≥150	724.84	36.01	≥26,690	738.23	34.90	
		Deaf-Blindness	<50	NR	NR	≥26,840	738.15	34.92	
		Developmental Delay	<50	NR	NR	≥26,840	738.15	34.92	
		Emotional Disturbance	≥90	710.44	27.67	≥26,740	738.25	34.90	
		HI—Deaf	<50	NR	NR	≥26,830	738.16	34.92	
		HI—Hard-of-Hearing	<50	NR	NR	≥26,800	738.17	34.92	
		Mild Mental Disability	≥170	698.37	16.74	≥26,660	738.42	34.86	
		Moderate Mental Disability	<50	NR	NR	≥26,840	738.16	34.91	
		Orthopedic Impairment	<50	NR	NR	≥26,790	738.17	34.91	
		Other Health Impairment	≥780	709.55	24.20	≥26,060	739.01	34.83	
		Specific Learning Disability	≥2,270	708.35	21.76	≥24,560	740.92	34.62	
		Speech or Language Impairment	≥120	722.73	34.84	≥26,720	738.22	34.90	
	Traumatic Brain Injury	<50	NR	NR	≥26,830	738.16	34.92		
	Visual Impairment	<50	NR	NR	≥26,810	738.16	34.92		
	Other	<50	NR	NR	≥26,830	738.16	34.92		
		E	Gifted	≥730	798.29	29.06	≥20,900	742.31	32.87
			Talented	≥800	762.21	33.39	≥20,840	743.53	34.13
			Autism	≥50	744.36	35.21	≥21,590	744.22	34.29
			Deaf-Blindness	<50	NR	NR	≥21,640	744.22	34.29
			Developmental Delay	<50	NR	NR	≥21,640	744.22	34.29
			Emotional Disturbance	<50	NR	NR	≥21,620	744.24	34.28
			HI—Deaf	<50	NR	NR	≥21,640	744.22	34.29
	HI—Hard-of-Hearing		<50	NR	NR	≥21,630	744.22	34.28	
	Mild Mental Disability		<50	NR	NR	≥21,640	744.23	34.29	
	Moderate Mental Disability		<50	NR	NR	≥21,640	744.22	34.29	
	Orthopedic Impairment	<50	NR	NR	≥21,630	744.22	34.29		
	Other Health Impairment	≥140	725.37	30.28	≥21,490	744.35	34.28		

Special Education Classification Scale Score Means and Standard Deviations: Algebra I								
Admin.	Form	Group	Yes			No		
			N	Mean	Std. Dev.	N	Mean	Std. Dev.
		Specific Learning Disability	≥190	716.77	22.92	≥21,440	744.47	34.27
		Speech or Language Impairment	≥50	745.05	40.42	≥21,580	744.22	34.27
		Traumatic Brain Injury	<50	NR	NR	≥21,640	744.22	34.29
		Visual Impairment	<50	NR	NR	≥21,630	744.22	34.28
		Other	<50	NR	NR	≥21,640	744.22	34.29
Summer 2019	BR	Gifted	<50	NR	NR	≥1,950	708.44	20.36
		Talented	<50	NR	NR	≥1,930	708.40	20.37
		Autism	<50	NR	NR	≥1,940	708.45	20.41
		Deaf-Blindness	<50	NR	NR	≥1,950	708.44	20.36
		Developmental Delay	<50	NR	NR	≥1,950	708.44	20.36
		Emotional Disturbance	<50	NR	NR	≥1,940	708.44	20.40
		HI—Deaf	<50	NR	NR	≥1,950	708.43	20.36
		HI—Hard-of-Hearing	<50	NR	NR	≥1,950	708.45	20.36
		Mild Mental Disability	<50	NR	NR	≥1,930	708.57	20.32
		Moderate Mental Disability	<50	NR	NR	≥1,950	708.44	20.36
		Orthopedic Impairment	<50	NR	NR	≥1,950	708.49	20.34
		Other Health Impairment	≥90	702.81	19.76	≥1,850	708.73	20.36
		Specific Learning Disability	≥240	701.00	17.62	≥1,700	709.51	20.51
		Speech or Language Impairment	<50	NR	NR	≥1,940	708.46	20.34
		Traumatic Brain Injury	<50	NR	NR	≥1,950	708.44	20.37
		Visual Impairment	<50	NR	NR	≥1,950	708.42	20.38
Other	<50	NR	NR	≥1,950	708.45	20.35		



**Table 10.25 Special Education Classification Scale Score Means and Standard Deviations: Geometry**

Special Education Classification Scale Score Means and Standard Deviations: Geometry								
Admin.	Form	Group	Yes			No		
			N	Mean	Std. Dev.	N	Mean	Std. Dev.
Fall 2018	B	Gifted	≥170	777.87	21.34	≥5,180	732.21	27.29
		Talented	≥300	751.19	24.24	≥5,050	732.66	28.20
		Autism	<50	NR	NR	≥5,340	733.74	28.31
		Deaf-Blindness	<50	NR	NR	≥5,350	733.72	28.31
		Developmental Delay	<50	NR	NR	≥5,350	733.72	28.31
		Emotional Disturbance	<50	NR	NR	≥5,340	733.77	28.28
		HI—Deaf	<50	NR	NR	≥5,350	733.73	28.31
		HI—Hard-of-Hearing	<50	NR	NR	≥5,350	733.79	28.29
		Mild Mental Disability	<50	NR	NR	≥5,340	733.85	28.25
		Moderate Mental Disability	<50	NR	NR	≥5,350	733.72	28.31
		Orthopedic Impairment	<50	NR	NR	≥5,350	733.73	28.31
		Other Health Impairment	≥50	712.58	25.46	≥5,300	733.94	28.26
		Specific Learning Disability	≥190	706.70	21.21	≥5,160	734.72	28.05
		Speech or Language Impairment	<50	NR	NR	≥5,330	733.76	28.30
		Traumatic Brain Injury	<50	NR	NR	≥5,350	733.73	28.31
		Visual Impairment	<50	NR	NR	≥5,350	733.75	28.31
Other	<50	NR	NR	≥5,350	733.74	28.30		
Spring 2019	D	Gifted	≥700	777.10	23.02	≥18,120	735.54	25.66
		Talented	≥660	749.74	25.29	≥18,150	736.62	26.68
		Autism	≥80	726.24	24.73	≥18,740	737.13	26.75
		Deaf-Blindness	<50	NR	NR	≥18,820	737.09	26.75
		Developmental Delay	<50	NR	NR	≥18,820	737.09	26.75
		Emotional Disturbance	≥50	712.15	22.92	≥18,760	737.16	26.72
		HI—Deaf	<50	NR	NR	≥18,800	737.10	26.75
		HI—Hard-of-Hearing	<50	NR	NR	≥18,790	737.10	26.75
		Mild Mental Disability	≥60	705.05	15.95	≥18,760	737.19	26.71
		Moderate Mental Disability	<50	NR	NR	≥18,820	737.09	26.75
		Orthopedic Impairment	<50	NR	NR	≥18,790	737.11	26.74
		Other Health Impairment	≥290	717.11	22.32	≥18,520	737.40	26.69
		Specific Learning Disability	≥890	712.78	17.95	≥17,920	738.29	26.53

Special Education Classification Scale Score Means and Standard Deviations: Geometry								
Admin.	Form	Group	Yes			No		
			N	Mean	Std. Dev.	N	Mean	Std. Dev.
		Speech or Language Impairment	≥60	720.49	18.67	≥18,750	737.14	26.75
		Traumatic Brain Injury	<50	NR	NR	≥18,810	737.09	26.75
		Visual Impairment	<50	NR	NR	≥18,800	737.09	26.74
		Other	<50	NR	NR	≥18,820	737.09	26.75
	E	Gifted	≥690	778.07	21.44	≥16,020	737.95	24.87
		Talented	≥630	751.96	22.69	≥16,080	739.13	26.00
		Autism	<50	NR	NR	≥16,680	739.63	25.99
		Deaf-Blindness	<50	NR	NR	≥16,710	739.61	26.00
		Developmental Delay	<50	NR	NR	≥16,710	739.61	26.00
		Emotional Disturbance	<50	NR	NR	≥16,700	739.63	26.00
		HI—Deaf	<50	NR	NR	≥16,710	739.61	26.00
		HI—Hard-of-Hearing	<50	NR	NR	≥16,700	739.62	26.00
		Mild Mental Disability	<50	NR	NR	≥16,710	739.62	26.00
		Moderate Mental Disability	<50	NR	NR	≥16,710	739.61	26.00
		Orthopedic Impairment	<50	NR	NR	≥16,700	739.61	25.99
		Other Health Impairment	≥70	723.21	24.78	≥16,640	739.68	25.98
		Specific Learning Disability	≥90	719.59	19.99	≥16,620	739.73	25.98
		Speech or Language Impairment	<50	NR	NR	≥16,670	739.62	26.00
		Traumatic Brain Injury	<50	NR	NR	≥16,710	739.61	26.00
		Visual Impairment	<50	NR	NR	≥16,710	739.61	26.00
Other	<50	NR	NR	≥16,710	739.61	26.00		
Summer 2019	B	Gifted	<50	NR	NR	≥260	706.95	21.95
		Talented	<50	NR	NR	≥270	710.26	28.13
		Autism	<50	NR	NR	≥270	710.67	28.59
		Deaf-Blindness	<50	NR	NR	≥270	710.67	28.59
		Developmental Delay	<50	NR	NR	≥270	710.67	28.59
		Emotional Disturbance	<50	NR	NR	≥270	711.19	28.47
		HI—Deaf	<50	NR	NR	≥270	710.67	28.59
		HI—Hard-of-Hearing	<50	NR	NR	≥270	710.79	28.57
		Mild Mental Disability	<50	NR	NR	≥260	711.10	28.73
		Moderate Mental Disability	<50	NR	NR	≥270	710.67	28.59
		Orthopedic Impairment	<50	NR	NR	≥270	710.67	28.59
		Other Health Impairment	<50	NR	NR	≥260	710.81	28.86

Special Education Classification Scale Score Means and Standard Deviations: Geometry								
Admin.	Form	Group	Yes			No		
			N	Mean	Std. Dev.	N	Mean	Std. Dev.
		Specific Learning Disability	<50	NR	NR	≥250	711.97	29.11
		Speech or Language Impairment	<50	NR	NR	≥270	710.74	28.62
		Traumatic Brain Injury	<50	NR	NR	≥270	710.67	28.59
		Visual Impairment	<50	NR	NR	≥270	710.67	28.59
		Other	<50	NR	NR	≥270	710.67	28.64

## 10.6 Summary

In summary, the overall purpose of this chapter is to address fairness concerns that are relevant to the administration of LEAP 2025 assessments. The information in this chapter addresses multiple best practices of the testing industry and is particularly related to the following standards:

*Standard 3.1* Those responsible for test development, revision, and administration should design all steps of the testing process to promote valid score interpretations for intended score uses for the widest possible range of individuals and relevant subgroups in the intended population. (63)

*Standard 3.2* Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics. (64)

*Standard 3.3* Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test. (64)

*Standard 3.4* Test takers should receive comparable treatment during the test administration and scoring process. (65)

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

*Standard 3.6* Where credible evidence indicates that test scores may differ in meaning for relevant subgroups in the intended examinee population, test developers and/or users are responsible for examining the evidence for validity of score interpretations for intended uses for individuals from those subgroups. What constitutes a significant difference in subgroup scores and what actions are taken in response to such differences may be defined by applicable laws. (65)

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

## Appendix A—Text Complexity Placemat Template

Worksheet: Text Complexity Analysis		
Title	Author	Text Description

### Recommended Placement for Assessment: Grade 4



Qualitative Measures	Quantitative Measures												
<b>PURPOSE:</b> <b>TEXT STRUCTURE</b> Organization of Main Ideas: Text Features: Use of Images: <b>LANGUAGE FEATURES</b> Conventinality: Vocabulary: Sentence Structure: <b>KNOWLEDGE DEMANDS</b> Subject Matter Knowledge: Intertextuality:	<b>Common Core State Standards Appendix A Complexity Band Level</b> (if applicable): <b>Lexile or Other Quantitative Measure of the Text:</b>												
	<b>Considerations for Passage Selection</b>												
	Passage selection should be based on the ELA Content Specifications targets and the cognitive demands of the assessment tasks. <b>Potential Challenges This Text May Pose (check all that apply):</b> <table border="1" style="width: 100%;"> <tbody> <tr> <td style="width: 40px;"></td> <td>Accessibility</td> </tr> <tr> <td></td> <td>Sentence and text structures</td> </tr> <tr> <td></td> <td>Archaic language, slang, idioms, or other language challenges</td> </tr> <tr> <td></td> <td>Background knowledge</td> </tr> <tr> <td></td> <td>Bias and sensitivity issues</td> </tr> <tr> <td></td> <td>Word count</td> </tr> </tbody> </table>		Accessibility		Sentence and text structures		Archaic language, slang, idioms, or other language challenges		Background knowledge		Bias and sensitivity issues		Word count
	Accessibility												
	Sentence and text structures												
	Archaic language, slang, idioms, or other language challenges												
	Background knowledge												
	Bias and sensitivity issues												
	Word count												

Adapted from Smarter Balanced and the 2012 ELA SCASS work

## Appendix B—Item Content and Bias Review

---

### ***English Language Arts:***

Educators reviewed items for passage set quality and overall grade-level appropriateness. Item-specific characteristics reviewed included content alignment; cognitive complexity; bias, fairness, and sensitivity; and technical design. For content alignment, educators reviewed items to determine if the item was aligned to the content and skills indicated in the associated standard(s). With regard to cognitive complexity, educators reviewed the items for grade-level appropriateness and appropriate range of difficulty. Educators also assigned a depth-of-knowledge (DOK) level based on Webb’s DOK scale and analyzed each item for appropriate source of challenge, which indicates that the most difficult part of the item is indeed the skill defined in the standard the item is purported to measure. For bias, fairness, and sensitivity, educators reviewed items to ensure that barriers to successful performance on the test items were nonexistent or were removed via suggested revisions. For technical design the different parts of the items were analyzed to ensure that each functioned as it should.

### ***Mathematics:***

Educators reviewed items for content alignment, cognitive complexity, difficulty, bias, fairness, and sensitivity, and technical design. For content alignment, educators reviewed the items to determine if the item was aligned to the LSSM and/or the LEAP 2025 Evidence Statements and was appropriate for the grade level. With regard to cognitive complexity, educators assigned a depth-of-knowledge (DOK) level based on Webb’s DOK scale and analyzed each item for appropriate source of challenge to estimate a difficulty level (low, medium, high), to verify that the item only measured student mastery of the aligned standard and not some other irrelevant concept or skill. For bias, fairness, and sensitivity, educators reviewed items to ensure that barriers to successful performance on the test items were nonexistent or were removed via suggested revisions. Technical design referred to analyzing the different parts of the item to ensure that each functioned as it should.

### ***Summer/Fall 2018 Item Content and Bias Review Process***

The newly-developed items available for field-testing are aligned to the Louisiana Student Standards for ELA and Mathematics and/or LEAP 2025 Mathematics Evidence Statements as determined by committees of Louisiana educators during the Summer/Fall 2018 LEAP 2025 High School Item Content and Bias Review. The process used to train educators on their role at the Item Review is outlined below.

1. Committee members attended a general session and received an overview of the process, which included:
  - a. Background of the review process that took place prior to acquired items being field-tested
  - b. Overview of purpose of the review:
    - i. To confirm content alignment of each item to designated Louisiana Student Standard
    - ii. To confirm grade appropriateness of each item
    - iii. To confirm cognitive complexity of each item
    - iv. To confirm the correct key(s) or response for each item
    - v. To confirm items are free of issues of bias, fairness, or sensitivity that could impact student responses to item

- c. Overview of review process:
      - i. Committee members review each item individually and decide status of each item: accepted, accepted with revisions, or rejected.
      - ii. Group discusses and comes to consensus regarding status of each item.
      - iii. Items are either accepted or edited by group as needed.
      - iv. Items that are accepted or accepted with revisions are considered appropriate for future field-testing.
2. After the general session, committee members reported to their assigned committees, based on course.
3. Each group 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: item ID, passage, keys, primary, secondary, and tertiary standards (as applicable), and depth of knowledge.
  - b. The mathematics spreadsheet contained prepopulated columns of the following information for each item: ABBI ID, IDEAS ID, max points, Louisiana Student Standards for Math (LSSM)/LEAP evidence statement, secondary and tertiary standards (as applicable), estimated difficulty level (low, medium, high), cognitive complexity (depth of knowledge 1, 2, or 3), key, and item type.
4. The facilitator reviewed each column in the spreadsheet so committee members could better understand what information needed to be filled in.
5. The facilitator went over what each column represented to make sure all participants understood the information being presented.
6. The facilitator went over the columns to be filled in individually by each committee member as they reviewed an item (accept, accept with revisions, reject), and any comments they might have about the item.
7. The facilitator then reviewed the process to be used with each item.
  - a. For the first couple of items, the facilitator and committee members reviewed the item(s) together. They then came to group consensus as to the status of the items. Items were edited as necessary by the group.
  - b. When ready, committee members reviewed each item individually (facilitator indicated how many items to do before getting back together for group discussion of each item).
  - c. 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.
  - d. This process continued for all items taken to the item review.
  - e. Committee members were provided with an evaluation at the end of the meeting so that DRC and LDOE can make improvements for future item reviews.

## Universal Design




- Universal Design Principles make test items accessible for the widest range of the population possible. Some of these include:
  - Use simple, common words instead of low-frequency words when possible.
  - Avoid irregularly-spelled words, words with ambiguous or multiple meanings, and technical terms unless they are defined and integral to the meaning.
  - Ensure clarity of noun-pronoun relationships

## Bias and Sensitivity



- In order to have fairness in assessment, it is critical to ensure test materials are free of possible barriers to success among diverse groups of test takers
- Barriers can be reduced by ensuring items:
  - do not measure irrelevant knowledge or skill
  - do not anger, upset or distract test takers
  - treat all groups of people with respect

## Most Common Forms of Bias



- stereotypical representations
- geographical bias
- socioeconomic bias
- religious bias
- gender bias
- linguistic bias
- exclusion or underrepresentation of groups

Table B.1. LEAP 2025 2018 Item Content and Bias Review Totals

Content Area	Total Items Reviewed	Total Accepted Items at DOK 1	Total Accepted Items at DOK 2	Total Accepted Items at DOK 3	Accepted as Is	Accepted with Revision	Rejected
English I	36	0	11	25	8	28	0
English II	36	0	11	25	7	29	0
Algebra I	4	0	2	2	0	4	0
Geometry	4	0	3	1	1	3	0

Examples of the review sheet for ELA and mathematics that was completed by each reviewer follow.



### ELA Item Review Google Sheet

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Session	Session Sequence	Passage Title	ABBI ID	IDEAS ID	Item Type	Primary Standard	Secondary Standard	Tertiary Standard	Key	DOK	Content Alignment Standards Key Ideas Grade Appropriate	Cognitive Complexity DOK Source of Challenge	Technical Design Language Demands Correct Answers Distractions	Universal Design (Accessibility) Graphics Bias Sensibility	Status Accept Accept w/Revisions Reject	Comments
2												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
3												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
4												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
5												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
6												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
7												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
8												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
9												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
10												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
11												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
12												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
13												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
14												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
15												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
16												--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	

### Math Item Review Google Sheet

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Session	Session Sequence	ABBI ID	IDEAS ID	Item Type	Primary Standard	Secondary Standard	Key	Calc/No Calc	Content Alignment Standards	Cognitive Complexity DOK Level of Difficulty		Universal Design Language Demand Bias Grade Appropriateness Source of Challenge	Technical Design Correct Answer Distractions Graphics	Rubric	Status Accept Accept w/Revisions Reject	Comments
2											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
3											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
4											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
5											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
6											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
7											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
8											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
9											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
10											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
11											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
12											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
13											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
14											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
15											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	
16											--	--	--	<input type="checkbox"/>	<input type="checkbox"/>	--	

Table B.2. LEAP 2025 June 2018 Algebra I Content and Bias Review Committee Make-Up

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

Table B.3. LEAP 2025 June 2018 Geometry Content and Bias Review Committee Make-Up

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

Table B.4. LEAP 2025 October 2018 English I Content and Bias Review Committee Make-Up

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

Table B.5. LEAP 2025 October 2018 English II Content and Bias Review Committee Make-Up

<b>Member #</b>	<b>Gender</b>	<b>Race/Ethnicity</b>	<b>Background</b>
1	Female	White	Teacher (VI)
2	Female	Asian	Teacher (SPED)
3	Female	Native American/African American	Teacher
4	Male	White	Teacher
5	Female	White	Teacher
6	Female	White	Teacher
7	Female	African American	Administrator
8	Female	African American	Administrator
9	Male	White	Teacher
10	Female	White	Teacher

## Appendix C—Accommodated Print Form Creation

---

### ***Guidelines for Building Accommodated Print Forms***

Careful consideration is given to all items that are used for accommodated print (AP) forms and/or braille forms. Fairness for all populations, item integrity, and student-item interaction for technology-enhanced (TE) items are all factors when selecting the items that will appear on an AP form. TE items are modified so that students who interact with an item on an AP form will have a similar experience to students who interact with that same item in the online environment. This maintains both the rigor and the content being assessed. Some examples of the modification process are provided below.

- Drag-and-drop items in the online environment require a student to place the answer options in an interactive table. For the AP form, the student is presented with a table with the same information as the interactive table (column or row headers, any completed cells, and blank spaces) and the answer options are listed below the table (similar to the online form in which the options are listed either below or to the right of the table). The directions are modified to ask the student to write the correct answer in its corresponding box. Students are also able to circle the text and draw arrows to indicate where it should be placed or add labels to the answer choices and write only the label in the box, as long as the intended response is clear to the test administrator who will transcribe the answers into the online system.
- Matching items in the online environment require a student to select a checkbox in one or more columns for each of multiple rows. In the AP 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 a student to click on the selected text, which highlights the selected word, phrase, or sentence. 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 menu 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 require a student to type the 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 environment require 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 environment require a student to complete a graph by plotting points, adding *Xs* to create a line plot, or raising/lowering bars to create a bar graph or histogram. In the AP form, the student is provided with the same coordinate plane, number line, line plot, or bar graph as in the online item, including titles, axis labels, and keys, and is asked to complete the graph.

Displaying items similarly in both accommodated print forms and the online environment (and allowing students to interact with the items in a similar manner) maintains item integrity by assessing a similar

construct in a similar manner regardless of where a student encounters an item. This provides students who are unable to access the assessment online with an assessment at the same level of rigor as the online test.

AP forms are thoroughly reviewed by DRC and LDOE content experts 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.

## Appendix D—Item Alignment Review Process

---

The acquired items available for use on forms are aligned to the Louisiana Student Standards for ELA and Mathematics and/or LEAP 2025 Mathematics Evidence Statements. Their alignment was determined by committees of Louisiana educators during the June 2018 LEAP 2025 High School Item Alignment Review. The process used to train educators at the Item Alignment Review is outlined below.

1. Committee members attended a general session and received an overview of the process, which included the following:
  - a. Background on items and the review process that took place prior to acquired items being field-tested
  - b. Overview of the purpose of the alignment review
    - i. To confirm the content alignment of each item to a designated Louisiana State Standard
    - ii. To confirm the grade appropriateness of each item
    - iii. To confirm that items are free of issues of bias, fairness, or sensitivity that could impact student responses to each item
  - c. Overview of the process itself
    - i. Committee members will review each item individually and decide the status of each item. The status of each item could be “accepted with current alignment,” “accepted with realignment,” or “rejected.”
    - ii. The group will discuss the items and come to a consensus regarding the status of each item.
    - iii. Items that appropriately measure the intended standard and/or evidence statement and are free of issues of bias, fairness, or sensitivity that could impact student responses are accepted and considered appropriate for inclusion in LDOE item bank.
2. After the general session, committee members reported to their committees, which were assigned based on course.
3. Each group 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 with the following information for each item: item ID, passage, keys, primary, secondary, and tertiary standards (as applicable), and depth of knowledge.
  - b. The mathematics spreadsheet contained prepopulated columns with the following information for each item: ABBI ID, IDEAS ID, maximum points, Louisiana Student Standards for Mathematics (LSSM)/LEAP evidence statement, secondary and tertiary standards (as applicable), estimated difficulty level (low, medium, high), cognitive complexity (depth of knowledge 1, 2, or 3), key, and item type.
4. The facilitator reviewed each column in the spreadsheet.
  - a. The facilitator went over what each column represented to make sure all participants understood the information being presented.
  - b. 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 the place in the spreadsheet for 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.

5. The facilitator then reviewed the process to be used with each item.
  - a. For the first couple of items, the facilitator and committee members reviewed the item and its alignment(s) together. They then came to group consensus as to the status of the items.
  - b. When ready, committee members reviewed each item individually (the facilitator indicated how many items to do before getting back together for the group discussion of each item).
  - c. 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 to each item in a given set through group consensus.
  - d. This process continued for all items taken to the item alignment review.
  - e. Committee members were provided with an evaluation at the end of the meeting so that DRC and LDOE could make improvements to future item alignment reviews.

Table D.1. LEAP 2025 June 2018 Item Alignment Review Totals

Content Area	Total Items Reviewed	Total Accepted Items at DOK 1	Total Accepted Items at DOK 2	Total Accepted Items at DOK 3	Accepted with Current Alignment	Accepted with Realignment	Rejected
English I	206	0	157	45	165	37	4
English II	233	0	182	46	193	35	5
Algebra I	106	64	26	14	96	8	2
Geometry	106	25	59	16	96	4	6

Table D.2. LEAP 2025 June 2018 Algebra I Item Alignment Review Committee Makeup

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

Table D.3. LEAP 2025 June 2018 Geometry Item Alignment Review Committee Makeup

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

Table D.4. LEAP 2025 June 2018 English I Item Alignment and Passage Review Committee Makeup

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

Table D.5. LEAP 2025 June 2018 English II Item Alignment and Passage Review Committee Makeup

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



## Appendix E—Transadaptation Process for Spanish Mathematics Forms

---

For English Learners, the LDOE offers the mathematics assessments in Spanish for computer-based tests (CBTs) in order to mirror the English language forms and the text-to-speech (TTS) forms. The Spanish-language versions of the test were developed through transadaptation. Transadaptation takes into consideration the grade-level appropriateness of the words and sentence structures used and the linguistic and cultural differences that exist between speakers of two different languages. Accounting for these differences allows experts to ensure that a Spanish-language version of an item will measure the same construct as the English-language version of the item at the same level of rigor. The item is, therefore, expected to measure the achievement of English Learners in the same way that the English version of the item does for native speakers of English.

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

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

The Spanish Test Development team then reconciled any discrepancies and submitted the transadaptations to a senior Spanish Test Development team member for resolution. After approval by the senior Spanish Test Development team member, the item moved forward to be imported into DRC's item banking system.

Both previously transadapted items and newly transadapted items were imported into DRC's item banking system and formatted for online use. Each Spanish item was paired with the corresponding English item in the item bank, and the Spanish item was formatted. Graphics for the item were then finalized for review. The

finalized transadaptation was then compared to the Spanish version of the item in the DRC assessment system and the English version of the item, and all changes were verified.

DRC's Spanish Test Development team then used the final, approved communication assistance scripts in English to transadapt descriptions of graphics as necessary. These descriptions were used when preparing the TTS forms for review. Scripting the TTS forms and reviewing the finalized Spanish forms were conducted by native Spanish speakers at DRC prior to submitting the forms to the LDOE for a translation review by a third-party translation vendor. The vendor reviewed the transadapted forms and provided feedback to the LDOE and DRC. Experienced DRC Spanish Test Development team members and the translation vendor resolved any issues, and DRC made modifications as necessary. The forms were then approved by both DRC and the LDOE translation vendor.

## Appendix F—LEAP 2025 Spring 2019 Handscoring/AI Documentation

---

# LEAP 2025 SPRING 2019

## HANDSCORING/AI DOCUMENTATION

### LEAP 2025 GRADES 3-8

ELA, Math, Science, and Social Studies

### LEAP 2025 HIGH SCHOOL AND EOC

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

## Contents

Schedule, Locations, and Staffing .....	1
Training and Scoring Schedule .....	1
Scorer Degree Requirements .....	2
Training .....	3
Training Materials .....	4
EOC Biology, LEAP 2025 Biology, LEAP 205 U.S. History, and LEAP 2025 Grades 3-8 Science and Social Studies .....	4
EOC English III.....	4
LEAP 2025 Algebra I, Geometry, and Grades 3-8 Math (Items and Materials Developed by DRC).....	5
LEAP 2025 Algebra I, Geometry, English I, English II, and Grades 3-8 ELA and Math (Items and Materials Developed by PARCC) .....	6
Algebra I, Geometry, and Grades 3-8 Math Training Set Composition .....	7
English I, English II, and Grades 3-8 ELA Training Set Composition .....	8
English I Constructed Response (CR) Items and Associated Training Materials.....	10
English II Prose Constructed Response (CR) Items and Associated Training Materials.....	10
<i>Grades 3-8 ELA CR Items and Associated Training Materials</i> .....	11
Algebra I Items and Associated Training Materials.....	12
Geometry Items and Associated Training Materials.....	13
<i>Grade 3 Math Items and Associated Training Materials</i> .....	14
<i>Grade 4 Math Items and Associated Training Materials</i> .....	14
<i>Grade 5 Math Items and Associated Training Materials</i> .....	14
<i>Grade 6 Math Items and Associated Training Materials</i> .....	15
<i>Grade 7 Math Items and Associated Training Materials</i> .....	15
<i>Grade 8 Math Items and Associated Training Materials</i> .....	15
Qualifying .....	16
EOC Biology and EOC English III .....	16
EOC Biology .....	16
EOC English III.....	16

LEAP 2025 Constructed-Response and Extended-Response Items .....	17
LEAP 2025 English I, English II, and Grades 3-8 ELA.....	17
LEAP 2025 Algebra I, Geometry, and Grades 3-8 Math.....	17
LEAP 2025 U.S. History and Grades 3-8 Social Studies .....	18
LEAP 2025 Biology and Grades 3-8 Science .....	18
Spring 2019 Scoring Plan.....	19
LEAP 2025 High School and EOC.....	19
LEAP 2025 Grades 3-8 .....	19
Handscoring Rules.....	20
AI Scoring .....	20
Handscoring .....	20
Reader Monitoring Procedures.....	21
Team Leader Read-Behinds .....	21
Validity Sets and Inter-Rater Reliability .....	21
Calibration Sets .....	24
Handscoring Quality Control Reports .....	25
Scoring Summary Report Sample .....	25
Scoring Summary Report Sample with AI (Reader ID #3) .....	25
Reader Feedback Logs.....	26
Handling Unusual Responses.....	26
Nonscore Codes and Definitions.....	26
Nonscore Code Definitions .....	27
Nonscore Codes by Course .....	27
Alerts.....	27
Artificial Intelligence Scoring .....	28
AI Scoring – Measurement, Inc. ....	28
Model Building.....	29
Evaluation Metric.....	30
Scoring Responses with the AI Engine .....	31

Quality Control of the AI Engine .....	32
Identifying Responses for Human Review .....	32
Alert Detection System .....	33
Identification of Nonscorable Responses .....	34
Identifying Copied Text and Plagiarism with the AI Engine .....	34
AI Scoring – Pearson .....	37
The Intelligent Essay Assessor.....	37
How the Intelligent Essay Assessor was Trained .....	38
Quality Monitoring.....	40
Scoring (DRC) .....	41
Rescores .....	41
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 Data – EOC English III (Spring 2019) .....	48
Quadratic Weighted Kappa (QWK) and Inter-rater Reliability (IRR).....	48
Score Point Distribution (SPD) .....	48
AI Model Data – LEAP 2025 U.S. History ER (Spring 2019) .....	49
Quadratic Weighted Kappa (QWK), Inter-rater Reliability (IRR), and Score Point Distribution (SPD) 49	
AI Model Building – Social Studies Grades 5-8 ERs (Spring 2019) .....	50
Quadratic Weighted Kappa (QWK), Inter-rater Reliability (IRR), and Score Point Distribution (SPD) 50	
AI Model CR Performance – ELA Grades 6-8, English I, and English II (Spring 2019) .....	51
Spring 2019 LEAP 2025 and EOC Items – IRR and SPD from Previous Administrations .....	52
<i>Algebra I</i> .....	52
<i>Algebra I (continued)</i> .....	53
<i>Geometry</i> .....	54

<i>Math Grade 3</i> .....	56
<i>Math Grade 4</i> .....	57
Math Grade 5 .....	58
<i>Math Grade 6</i> .....	59
<i>Math Grade 7</i> .....	60
<i>Math Grade 8</i> .....	61
English I .....	62
English II .....	63
EOC English III (All Report Data Scored by DRC and MI [AI]) .....	64
<i>ELA Grade 3</i> .....	65
<i>ELA Grade 4</i> .....	65
<i>ELA Grade 5</i> .....	66
<i>ELA Grade 6</i> .....	66
<i>ELA Grade 7</i> .....	67
<i>ELA Grade 8</i> .....	67
Biology (EOC).....	68
Biology ERs and CRs (LEAP 2025) .....	69
Grade 3 Science.....	70
Grade 4 Science.....	70
Grade 5 Science.....	70
Grade 6 Science.....	71
Grade 7 Science.....	71
Grade 8 Science.....	71
<i>U.S. History ERs and CRs (LEAP 2025)</i> .....	72
<i>Social Studies Grade 3</i> .....	73
<i>Social Studies Grade 4</i> .....	73
<i>Social Studies Grade 5</i> .....	73
<i>Social Studies Grade 6</i> .....	74
<i>Social Studies Grade 7</i> .....	74



*Social Studies Grade 8* ..... 74

# Schedule, Locations, and Staffing

All reader training and handscoring for the spring 2019 administration of LEAP 2025 grades 3-8 and high school assessments and End-of-Course (EOC) high school assessments will take place at the DRC scoring center locations noted in the table below.

## Training and Scoring Schedule

DRC's reader training and scoring schedule is based on the spring testing windows of April 1, 2019 – May 3, 2019 (LEAP 2025 grades 3-8) and April 15, 2019 – May 17, 2019 (LEAP 2025 and EOC high school).

Reader training and scoring locations and the anticipated dates for each are noted below.

Grade/Content Area or Course	DRC Scoring Center Location	Anticipated Staffing	2019 Reader Training and Scoring Window
3 ELA	Columbus, OH	2 Scoring Directors, 5 Team Leaders, 55 Readers	May 9 – June 5
4 ELA	Madison, WI	2 Scoring Directors, 6 Team Leaders, 60 Readers	May 9 – June 5
5 ELA	Madison, WI	2 Scoring Directors, 4 Team Leaders, 45 Readers	April 4 – May 10
6 ELA	Plymouth, MN	1 Scoring Director, 1 Team Leader, 13 Readers	April 1 – May 10
7 ELA	Plymouth, MN	2 Scoring Directors, 4 Team Leaders, 30 Readers	April 1 – May 10
8 ELA	Atlanta, GA	2 Scoring Directors, 4 Team Leaders, 30 Readers	April 4 – May 10
3 Math	Woodbury, MN	2 Scoring Directors, 6 Team Leaders, 60 Readers	May 9 – June 5
4 Math	Plymouth, MN	2 Scoring Directors, 4 Team Leaders, 40 Readers	May 13 – June 5
5 Math	Sharonville, OH	2 Scoring Directors, 4 Team Leaders, 40 Readers	April 4 – May 10
6 Math	Lake Mary, FL	2 Scoring Directors, 5 Team Leaders, 50 Readers	April 4 – May 10
7 Math	Woodbury, MN	2 Scoring Directors, 5 Team Leaders, 50 Readers	April 4 – May 10
8 Math	Sharonville, OH	2 Scoring Directors, 5 Team Leaders, 50 Readers	April 4 – May 10
3 Science	Indianapolis, IN	2 Scoring Directors, 5 Team Leaders, 55 Readers	May 9 – June 5
4 Science	Indianapolis, IN	2 Scoring Directors, 9 Team Leaders, 90 Readers	May 9 – June 5
5 Science	Atlanta, GA	2 Scoring Directors, 6 Team Leaders, 55 Readers	April 4 – May 10
6 Science (CRs)	Indianapolis, IN	1 Scoring Director, 4 Team Leaders, 40 Readers	April 4 – May 10
6 Science (ERs)	Indianapolis, IN	1 Scoring Director, 3 Team Leaders, 25 Readers	April 1 – May 10
7 Science	Indianapolis, IN	2 Scoring Directors, 7 Team Leaders, 65 Readers	April 4 – May 10
8 Science (CRs)	Atlanta, GA	1 Scoring Director, 3 Team Leaders, 30 Readers	April 4 – May 10
8 Science (ERs)	Plymouth, MN	1 Scoring Director, 3 Team Leaders, 30 Readers	April 3 – May 10
3 SS	Atlanta, GA	1 Scoring Director, 2 Team Leaders, 20 Readers	May 13 – June 5
4 SS	Atlanta, GA	1 Scoring Director, 2 Team Leaders, 20 Readers	May 13 – June 5
5 & 6 SS (CRs only)	Atlanta, GA	1 Scoring Director, 4 Team Leaders, 40 Readers	April 1 – May 10
7 & 8 SS (CRs only)	Atlanta, GA	1 Scoring Director, 4 Team Leaders, 40 Readers	April 1 – May 10
5, 6, 7, & 8 SS (ERs only)	Plymouth, MN	1 Scoring Director, 4 Team Leaders, 40 Readers	April 3 – May 10

<b>Grade/Content Area or Course</b>	<b>DRC Scoring Center Location</b>	<b>Anticipated Staffing</b>	<b>2019 Reader Training and Scoring Window</b>
<b>LEAP 2025 Algebra I</b>	Sharonville, OH	2 Scoring Directors, 4 Team Leaders, 40 Readers	April 11 – May 22
<b>LEAP 2025 Geometry</b>	Sharonville, OH	2 Scoring Directors, 2 Team Leaders, 25 Readers	April 11 – May 22
<b>LEAP 2025 English I</b>	Plymouth, MN	1 Scoring Director, 1 Team Leader, 15 Readers	April 11 – May 22
<b>LEAP 2025 English II</b>	Plymouth, MN	1 Scoring Director, 1 Team Leader, 23 Readers	April 11 – May 22
<b>EOC English III</b>	Plymouth, MN	1 Scoring Director, 1 Team Leader, 5 Readers	April 11 – May 22
<b>EOC Biology</b>	Plymouth, MN	1 Scoring Director, 5 Readers	April 11 – May 22
<b>LEAP 2025 Biology</b>	Plymouth, MN	1 Scoring Director, 3 Team Leaders, 30 Readers	April 15 – May 22
<b>LEAP 2025 U.S. History</b>	Plymouth, MN	1 Scoring Director, 4 Team Leaders, 40 Readers	April 10 – May 22

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 content area or course 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 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 and English III

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

- LEAP 2025 grades 3-8 Science and Social Studies, as well as select items for grades 3-8 Math (noted as DRC Material Type on pages 14-15) originating with the spring 2018 DRC field test
- LEAP 2025 Biology and U.S. History, as well as select items for Algebra I and Geometry (noted as DRC Material Type on pages 12-13) originating with the spring 2018 DRC field test

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

- LEAP 2025 grades 3-8 ELA and Math items developed by PARCC
- LEAP 2025 Algebra I, Geometry, English I, and English II items developed by PARCC

The materials include:

- Passages, items/prompts, associated stimuli for applicable content areas/courses 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 2019 administration of the LEAP 2025 grades 3-8 and high school and EOC assessments.

## Training Materials

### *EOC Biology, LEAP 2025 Biology, LEAP 205 U.S. History, and LEAP 2025 Grades 3-8 Science and Social Studies*

Reader training for the EOC biology task is conducted using item-specific anchor sets, training sets, and qualifying sets provided by Pacific Metrics. The LEAP 2025 biology, U.S. history, and grades 3-8 science and social studies item-specific training materials were developed by DRC.

<b>Set Type*</b>	<b>Biology, U.S. History, and Grades 3-8 Science and Social Studies Training Materials</b>	<b>Annotated</b>
Anchor Set	Most item-specific anchor sets contain at least two responses per score point (with at least one example of each of the top scores).*	Yes
Training Sets	There are at least two training sets for each item <ul style="list-style-type: none"> <li>● 10 responses per training set</li> <li>● All numeric score points are represented*</li> </ul>	No
Qualifying Sets	There are two qualifying sets for each item <ul style="list-style-type: none"> <li>● 10 responses per qualifying set</li> <li>● All numeric score points are represented*</li> </ul>	No
*Examples of responses at the top score points or for all score-point combinations may not be present in some anchor, training, and 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 identify 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.

<b>Set Type</b>	<b>Content and Style Dimension Training Materials</b>	<b>Annotated</b>
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.		

EOC 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 a score point 1 for each element.

<b>Set Type</b>	<b>Conventions (F, U, M, S) Element Training Materials</b>	<b>Annotated</b>
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, and Grades 3-8 Math (Items and Materials Developed by DRC)*

Training materials for math items developed by DRC and field tested in spring of 2018 are made up of item-specific anchor sets, training sets, and qualifying sets developed by DRC.

<b>Set Type*</b>	<b>Algebra I, Geometry, and Grades 3-8 Math Training Materials</b>	<b>Annotated</b>
Anchor Set	Each item-specific anchor set contains at least two responses per score point (with at least one of each of the top score points).	Yes
Training Sets	There are two training sets for each item representing the range of responses <ul style="list-style-type: none"> <li>● 10 responses per training set</li> <li>● All numeric score points are represented*</li> </ul>	No
Qualifying Sets	There are three qualifying sets for each item <ul style="list-style-type: none"> <li>● 10 responses per qualifying set</li> <li>● All numeric score points are represented*</li> </ul>	No
*Examples of responses at the top score points may not be present in some anchor, training, and 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 identify examples of these scores during live scoring to supplement reader training.		

## *LEAP 2025 Algebra I, Geometry, English I, English II, and Grades 3-8 ELA and Math (Items and Materials Developed by PARCC)*

For all LEAP 2025 English I, English II, and grades 3-8 ELA items and a portion of the LEAP 2025 Algebra I, Geometry, and grades 3-8 math items (exceptions are referred to on page 5 in the *LEAP 2025 Algebra I, Geometry, and grades 3-8 Math [Items Developed by DRC]* section and specifically noted on pages 11-15), DRC will use the PARCC-approved training and qualifying materials provided by New Meridian. Training materials for each item can be grouped into one of two categories: “prototype” item materials or “abbreviated” item materials.

### **Prototype Item Materials**

Some PARCC 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 PARCC 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.<sup>1</sup> 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.

---

<sup>1</sup> Item associations were determined by PARCC and 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/course and task type.

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

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

<b>Set Type</b>	<b>Mathematics Prototype Item Training</b>	<b>Mathematics Abbreviated Item Training</b>	<b>Annotated</b>
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



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

<b>Set Type</b>	<b>English Prototype Item Training</b>	<b>English Abbreviated Item Training**</b>	<b>Annotated</b>
Anchor Set (for the RCWE and WE traits)	3 responses per score point <ul style="list-style-type: none"> <li>Anchor Sets for prototype RST and LAT item training include scores for the combined trait Reading Comprehension and Written Expression (RCWE).</li> <li>Anchor sets for prototype NWT item training include scores for Written Expression (WE).</li> </ul>	3 responses per score point <ul style="list-style-type: none"> <li>Anchor Sets for abbreviated RST and LAT item training include scores for the combined trait Reading Comprehension and Written Expression (RCWE).</li> <li>Anchor Sets for abbreviated NWT item training include scores for Written Expression (WE).</li> </ul>	Yes
Practice Set 1	5 responses representing the range of responses for <ul style="list-style-type: none"> <li>the RCWE trait (for LAT and RST items)</li> <li>the WE trait (for NWT items)</li> </ul>	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.

\*\*Some English abbreviated item training sets approved by PARCC were for items that have previously been field tested only. The abbreviated (FT) training materials that were provided to DRC for these ELA CRs consist of a full-length anchor set with some annotations and a five-response practice set (unannotated). The full range of score points may not be represented in some anchor sets or practice sets.

Set Type	English Prototype and Abbreviated Item Training	Annotated
Anchor Set (for the Knowledge of Language and Conventions trait)	<ul style="list-style-type: none"> <li>• There are 3 responses per score point in each set.</li> <li>• 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.</li> <li>• 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).</li> <li>• In addition, readers will be required to qualify on the Knowledge of Language and Conventions trait during each prototype item qualifying session.</li> </ul>	Yes

Some items selected for use on the spring 2019 administration were previously only field tested by PARCC. Consequently, the abbreviated training materials available for use with these items are abridged versions of typical abbreviated sets of materials. They consist of:

- Anchor Set (for ELA, some have annotations and some lack examples of the top scores)
- One Practice Set of 5 responses (scored but unannotated in the case of ELA)
- Approximately 10 validity responses

Since these materials are somewhat limited compared to typical abbreviated materials, DRC will bolster the training by using the PARCC-approved field test validity responses provided by New Meridian as additional practice responses. DRC Scoring Directors will then pull additional validity responses from operational Louisiana responses to use during the scoring window. The Scoring Directors will also find examples of higher-scoring responses that might be missing from the field test anchors. The validity and additional exemplar responses, along with the DRC Scoring Directors’ notes for all papers used during the training of the abbreviated (FT) items, will be submitted to LDOE for approval.

While the field test-only abbreviated item materials are somewhat limited compared to regular abbreviated materials (the main difference being a lack of formal written annotations and fewer practice responses), using the PARCC-approved validity responses provided by New Meridian as additional practice responses is intended to help fill that gap. It is important to note that readers still must qualify via standard qualification procedures on the prototype items for all items by first going through full training with the appropriate prototype Anchor Set, Practice Sets 1-4, and Qualifying Sets 1-3 (as well as the Conventions sets).

*English I Constructed Response (CR) Items and Associated Training Materials*

Question	Form	Task	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
7	D	LAT	902152	VH017536_2T	Abbreviated	VH037763_2T
20	D	RST	914552	GG431834057	Abbreviated	VH017542 2T
9	E	RST	914552	GG431834057	Abbreviated	VH017542 2T
14	E	NWT	983215	GG604245591	Abbreviated (FT)	6139
9	A (SR)	RST	902161	VH017542_2T	Prototype	N/A
14	A (SR)	NWT	906152	VH084830	Abbreviated	6139
9	C (AE)	RST	902194	VH017614_2T	Abbreviated	VH017542 2T
14	C (AE)	NWT	902203	6139	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.						
Abbreviated (FT) – Item has previously only been field tested by Pearson/PARCC. Abbreviated (FT) training materials for ELA consist of a full-length anchor set with some annotations and a five-response practice set (unannotated).						

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

Question	Form	Task	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
7	D	LAT	906197	HH428127697	Abbreviated	7233 2T
20	D	RST	983688	HH607742252	Abbreviated	7121 2T
9	E	RST	983688	HH607742252	Abbreviated	7121 2T
14	E	NWT	983642	HH432845949	Abbreviated	VF908613
9	A (SR)	RST	902331	VH004490	Abbreviated	7121_2T
14	A (SR)	NWT	902354	7064	Abbreviated	VF908613
7	C (AE)	LAT	906181	HH431436431	Abbreviated	7233 2T
20	C (AE)	RST	906190	HH433954866	Abbreviated	7121 2T
*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 CR Items and Associated Training Materials*

Grade	Question	Task	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
3	7	RST	915227	A1598	Abbreviated (FT)	VF906000
	12	NWT	913497	AA431426588	Abbreviated	VF910093
4	7	LAT	913567	VH170170	Abbreviated	VF925727
	20	RST	982233	VH060330	Abbreviated (FT)	VF653524
5	7	LAT	801310	VF821667	Abbreviated	VF882724
	20	RST	915510	VH198972	Abbreviated (FT)	2208
6	9	RST	913715	DD502035970	Abbreviated	3538
	14	NWT	913694	D1466	Abbreviated	VH000592
7	9	RST	915582	E1567	Abbreviated (FT)	VH014400
	14	NWT	913842	EE430133306	Abbreviated	4284
8	7	LAT	913958	F1460	Abbreviated	5271
	20	RST	982327	FF506834510	Abbreviated (FT)	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.						
Abbreviated (FT) – Item has previously only been field tested by Pearson/PARCC. Abbreviated (FT) training materials for ELA consist of a full-length anchor set with some annotations and a five-response practice set (unannotated).						

*Algebra I Items and Associated Training Materials*

Question	Form	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
13	D	901832	3031-M44083P	Abbreviated	3003_M43111
15	D	938741	MA10144 (DRC ID)	DRC	N/A
28	D	980927	VH251952	Abbreviated	VH046614
29	D	938735	MA10137 (DRC ID)	DRC	N/A
43	D	938744	MA10147 (DRC ID)	DRC	N/A
44	D	938737	MA10139 (DRC ID)	DRC	N/A
45	D	938769	MA10178 (DRC ID)	DRC	N/A
13	E	980924	M44463	Abbreviated	VH046614
15	E	980909	M43216	Abbreviated	VH046614
28	E	980927	VH251952	Abbreviated	VH046614
29	E	980911	2679-M43312	Abbreviated	3003-M43111
43	E	901851	M41726	Abbreviated	3003-M43111
44	E	938737	MA10139 (DRC ID)	DRC	N/A
45	E	980923	M000312	Abbreviated	3003-M43111
13	A (SR)	901836	M43318	Abbreviated	3003-M43111
15	A (SR)	901882	VH196970	Abbreviated	VH046614
28	A (SR)	901859	3003-M43111	Prototype	N/A
29	A (SR)	901814	M47147	Abbreviated	2407-M41752
43	A (SR)	938769	MA10178 (DRC ID)	DRC	N/A
44	A (SR)	901848	M47287	Abbreviated	M41686
45	A (SR)	901857	VH046479	Abbreviated	2407-M41752
13	B (AE)	901832	3031 M44083P	Abbreviated	3003_M43111
15	B (AE)	901882	VH196970	Abbreviated	VH046614
28	B (AE)	901687	2407_M41752_AT	Prototype	N/A
29	B (AE)	938737	MA10139 (DRC ID)	DRC	N/A
43	B (AE)	901851	M41726	Abbreviated	3003_M43111
44	B (AE)	901705	VF883359_AT	Abbreviated	VH046614
45	B (AE)	901857	VH046479	Abbreviated	2407_M41752
*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.					
DRC Material Type – Training materials built by DRC using 2018 field test responses. These materials consist of an annotated Anchor Set, two Practice Sets, and three Qualifying Sets specific to each CR.					

*Geometry Items and Associated Training Materials*

Question	Form	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
13	D	902012	M41169	Abbreviated	VF935309
15	D	980937	M43798	Abbreviated	2904-M43021
27	D	939083	MGM0141 (DRC ID)	DRC	N/A
28	D	980942	VH236248	Abbreviated	2904-M43021
43	D	939077	MGM0135 (DRC ID)	DRC	N/A
44	D	980938	M100106	Abbreviated	VF935309
45	D	980936	VH239429	Abbreviated	2904-M43021
13	E	902012	M41169	Abbreviated	VF935309
15	E	980937	M43798	Abbreviated	2904-M43021
25	E	980929	M1000516	Abbreviated (FT)	2904-M43021
28	E	902042	3020-M44058	Abbreviated	3042-M44133
43	E	980930	M1000518	Abbreviated (FT)	2904-M43021
44	E	980938	M100106	Abbreviated	VF935309
45	E	980936	VH239429	Abbreviated	2904-M43021
13	A (SR)	901939	M43794	Abbreviated	V935309
15	A (SR)	902046	M46668	Abbreviated	3042-M44133
27	A (SR)	902027	M43233	Abbreviated	VH001716
28	A (SR)	902036	2904-M43021	Prototype	N/A
43	A (SR)	902047	VH150404	Abbreviated	V935309
44	A (SR)	939101	MGM0160 (DRC ID)	DRC	N/A
13	B (AE)	902012	M41169	Abbreviated	VF935309
15	B (AE)	902046	M46668	Abbreviated	3042_M44133
27	B (AE)	902027	M43233	Abbreviated	VH001716
28	B (AE)	902042	3020-M44058	Abbreviated	3042-M44133
43	B (AE)	902062	VH150384	Abbreviated	VF613786
44	B (AE)	939101	MGM0160 (DRC ID)	DRC	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.					
DRC Material Type – Training materials built by DRC using 2018 field test responses. These materials consist of an annotated Anchor Set, two Practice Sets, and three Qualifying Sets specific to each CR.					
Abbreviated (FT) Material Type – Item has previously only been field tested by Pearson/PARCC. Abbreviated (FT) training materials consist of a full-length Anchor Set and a five-response Practice Set (both are annotated).					

### Grade 3 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
17	981736	VH054794	Abbreviated	VH093931
18	914048	M05158	Abbreviated	M00848
32	898001	N/A	DRC	N/A
33	981742	M300388PD	Abbreviated	M00848
48	914039	M02527	Abbreviated	M00848
49	981747	4127-M03599P	Abbreviated	M01883
*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.				
DRC Material Type – Training materials built by DRC using 2018 field test responses. These materials consist of an annotated Anchor Set, two Practice Sets, and three Qualifying Sets specific to each CR.				

### Grade 4 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
17	914084	4112-M03491P	Abbreviated	0081_M00445
18	914086	M04133	Abbreviated	M03436
32	981831	M400526	Abbreviated	M03436
33	899959	N/A	DRC	N/A
48	899955	N/A	DRC	N/A
49	981927	0318-M01475	Abbreviated (FT)	M03436
*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.				
DRC Material Type – Training materials built by DRC using 2018 field test responses. These materials consist of an annotated Anchor Set, two Practice Sets, and three Qualifying Sets specific to each CR.				
Abbreviated (FT) Material Type – Item has previously only been field tested by Pearson/PARCC. Abbreviated (FT) training materials consist of a full-length Anchor Set and a five response Practice Set (both are annotated).				

### Grade 5 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
17	914152	M03820	Abbreviated	M03555
18	914148	M03888	Abbreviated	VH141466
32	902410	N/A	DRC	N/A
33	902414	N/A	DRC	N/A
48	914195	0154-M00796	Abbreviated	VH084803
49	934015	N/A	DRC	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.				
DRC Material Type – Training materials built by DRC using 2018 field test responses. These materials consist of an annotated Anchor Set, two Practice Sets, and three Qualifying Sets specific to each CR.				

### Grade 6 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
30	981963	M25151	Abbreviated	VH122131
34	981961	VH082639	Abbreviated	VH122131
35	981954	VH139067	Abbreviated	M21577
36	981956	VH220482	Abbreviated	M21577
47	914231	1740-M23030	Abbreviated	VH122131
48	903511	N/A	DRC	N/A
49	914281	M25152	Abbreviated	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.				
DRC Material Type – Training materials built by DRC using 2018 field test responses. These materials consist of an annotated Anchor Set, two Practice Sets, and three Qualifying Sets specific to each CR.				

### Grade 7 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
31	914362	VH083535	Abbreviated	VF643181
34	982922	M25544	Abbreviated	M20598
36	868848	M25578	Abbreviated	M20598
37	900539	N/A	DRC	N/A
47	900520	N/A	DRC	N/A
48	914339	VH151385	Prototype	N/A
49	982929	M22009	Abbreviated	M22018
*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.				
DRC Material Type – Training materials built by DRC using 2018 field test responses. These materials consist of an annotated Anchor Set, two Practice Sets, and three Qualifying Sets specific to each CR.				

### Grade 8 Math Items and Associated Training Materials

Question	DRC Item ID	PARCC UIN	Material Type	Associated Prototype Item*
31	983010	VH097312	Abbreviated	M21063
34	982987	M800114	Abbreviated (FT)	M21063
35	982999	M22203	Abbreviated	M21063
36	870899	1282-M21381	Abbreviated	M20198
42	899312	N/A	DRC	N/A
46	914381	M25425	Abbreviated	M21063
48	899329	N/A	DRC	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.				
Abbreviated (FT) Material Type – Item has previously only been field tested by Pearson/PARCC. Abbreviated (FT) training materials consist of a full-length Anchor Set and a five response Practice Set (both are annotated).				
DRC Material Type – Training materials built by DRC using 2018 field test responses. These materials consist of an annotated Anchor Set, two Practice Sets, and three Qualifying Sets specific to each CR.				



# 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 the spring 2019 administration of the LEAP 2025 and EOC items as were established by the vendors who scored these items previously. Qualifying standards for LEAP 2025 biology and grades 3-8 science were approved by the LDOE in February 2019.

## EOC Biology and EOC English III

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

### *EOC Biology*

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

### *EOC English III*

Course	Qualifying Standard
EOC English III (Content and Style)	EOC 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.
EOC 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.

## LEAP 2025 Constructed-Response and Extended-Response Items

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

### LEAP 2025 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 scores).

### LEAP 2025 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, 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.

*LEAP 2025 U.S. History and Grades 3-8 Social Studies*

Course and Item Type	Qualifying Standard
<b>U.S. History and Grades 3-8 Social Studies</b> 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.
<b>U.S. History and Grades 5-8 Social Studies</b> 0-8 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.

*LEAP 2025 Biology and Grades 3-8 Science*

Course and Item Type	Qualifying Standard	
<b>Biology and Grades 3-8 Science</b> 0-2 point CRs	0-2 Rubric	Scorers must qualify with 80% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
<b>Biology and Grades 3-8 Science</b> Composite (multi-part) ER items*	0-1 Rubric	Scorers must qualify with 90% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
	0-2 Rubric	Scorers must qualify with 80% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
	0-3 Rubric	Scorers must qualify with 70% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
	0-4 Rubric	Scorers must qualify with 70% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
	0-5 Rubric	Scorers must qualify with 70% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
	0-6 Rubric	Scorers must qualify with 60% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
	0-7 Rubric	Scorers must qualify with 60% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
	0-8 Rubric	Scorers must qualify with 60% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
<b>Grades 3 and 4 Science</b> Comprehensive (single part) ER items	0-6 Rubric	Scorers must qualify with 60% exact agreement or higher on one or more of the qualifying sets in order to score student responses.
<b>Biology and Grades 5-8 Science</b> Comprehensive (single part) ER items	0-9 Rubric	Scorers must qualify with 60% exact agreement or higher on one or more of the qualifying sets in order to score student responses.

\*Qualifying Sets are made up of 10 responses comparable to the Anchor Set responses. For multi-part (composite) Biology and Grades 3-8 Science ERs, 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 6 and Part B with a top score of 3, a scorer would need to achieve 60% perfect agreement on Part A and 70% perfect agreement on Part B on one or more of the qualifying sets. A scorer may qualify on one part in the first qualifying set and the other part in the second qualifying set.

# Spring 2019 Scoring Plan

The charts below provide an overview of the Spring 2019 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	NWT_GG604245591 (Form E) RST_VH017614_2T (Form C – AE) NWT_6139 (Form C – AE) RST_VH017542_2T (Form A – SR) NWT_VH084830 (Form A –SR)	LAT_VH017536_2T (Form D) RST_GG431834057 (Forms D & E)	Pearson
LEAP 2025 English II	RST_HH607742252 (Forms D & E) LAT_HH431436431 (Form C – AE) RST_HH433954866 (Form C – AE) NWT_7064 (Form A – SR) RST_VH004490 (Form A – SR)	LAT_HH428127697 (Form D) NWT_HH432845949 (Form E)	Pearson
EOC English III	Writing Prompt (Form X – AE)	Writing Prompt (Form W)	Measurement Inc.
LEAP 2025 Algebra I	All CRs	N/A	
LEAP 2025 Geometry	All CRs	N/A	
LEAP 2025 Biology	All CRs and ERs	N/A	
EOC Biology	ERs (operational and AE form)	N/A	
LEAP 2025 U.S. History	All CRs, ER (AE form)	ER (operational)	Measurement Inc.
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 CRs	N/A	
ELA grade 4	Both CRs	N/A	
ELA grade 5	Both CRs	N/A	
ELA grade 6	N/A	Both CRs	Pearson
ELA grade 7	RST_E1567	NWT_EE430133306	Pearson
ELA grade 8	RST_FF506834510	LAT_F1460	Pearson
Math grades 3-8	All CRs	N/A	
Science grades 3-8	All CRs and ERs	N/A	
Social Studies grades 3 and 4	All CRs	N/A	
Social Studies grades 5-8	All CRs	All ERs	Measurement Inc.
*DRC's handscoring teams will provide a second read for at least ten percent of all AI-scored responses.			

# Handscoring Rules

## AI Scoring

For EOC English III 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 select CRs in LEAP 2025 English I, English II, and grades 6-8 ELA (see Spring 2019 Scoring Plan on page 19), 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. (For additional information about the nonscore, alert, and flagged response review process, please see the Handling Unusual Responses section starting on page 26.) 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 2019 Scoring Plan) 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 score associated with the first scorer 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 the Scoring Rules found in the '/Scoring Documentation' folder posted on the Reporting SFTP site.)

# 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. (When test numbers are low and smaller groups of 10 or fewer readers are used, these groups may be supervised directly by the Scoring Director.) If scoring patterns are apparent among individual scorers, Team Leaders or Scoring Directors 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, and English II, and grades 3-8 Math and ELA items from New Meridian, DRC will utilize the same validity responses that Pearson used. These validity responses were approved by PARCC and supplied by New Meridian.

DRC scoring supervisors will identify validity responses during live scoring for all newly operational LEAP 2025 items in Algebra I, Geometry, Biology, U.S. History, and grades 3-8 Math, Science, and Social Studies. DRC will post these validity responses and their scores to the secure LDOE SFTP site for LDOE content staff to review and approve. Validity responses previously identified and approved for EOC Biology and EOC English III will be reused.

The validity responses will be added to DRC’s image handscoring system prior to the beginning of scoring (with the exception of the previously noted LEAP 2025 validity responses, which will be identified and added during live 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 25) 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). MI’s PEG AI scoring system will provide the first scores (the scores of record) for EOC English III and the LEAP 2025 U.S. History and grades 5-8 Social Studies ERs. For LEAP 2025 English I and English II, and select CRs in grades 5-8 ELA (see Spring 2019 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 28. 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 the inter-rater and validity exact agreement rates at or above the percentages noted below. When a reader’s validity or inter-rater agreement falls 5% or more below these expectations, or if Perfect Agreement + Adjacent percentages fall below the rates noted, 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 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 EOC and LEAP 2025 items are shown below.

<b>Agreement Rate Expectations for Validity and Inter-Rater Reliability – LEAP 2025 and EOC</b>			
<b>Content Area/Course</b>	<b>Score Point Range</b>	<b>Perfect Agreement</b>	<b>Perfect Agreement + Adjacent</b>
<b>Grades 3-8 ELA, English I, English II</b>	0-3 or 0-4 Rubric, Multi-trait	65% (each trait)	96% (each trait)
<b>EOC English III</b> (Style and Content)	1-4 (each domain)	70%	95%
<b>EOC English III</b> (F, U, M, S)	0-1 (each domain)	80%	95%
<b>Algebra I, Geometry, Grades 3-8 Math</b>	0-1 Rubric	90%	95%
<b>Algebra I, Geometry, Grades 3-8 Math</b>	0-2 Rubric	80%	95%
<b>Algebra I, Geometry, Grades 3-8 Math</b>	0-3 Rubric	70%	95%
<b>Algebra I, Geometry, Grades 3-8 Math</b>	0-4 Rubric	65%	95%
<b>EOC Biology</b>	0-4 Rubric	80%	95%
<b>LEAP 2025 Biology and Grades 3-8 Science</b> CR items	0-2 Rubric	80%	95%
<b>LEAP 2025 Biology and Grades 3-8 Science</b> Composite (multi-part) ER items	0-1 Rubric	90%	100%
	0-2 Rubric	80%	95%
	0-3 Rubric	70%	95%
	0-4 Rubric	70%	95%
	0-5 Rubric	70%	95%
	0-6 Rubric	60%	93%
	0-7 Rubric	60%	93%
	0-8 Rubric	60%	90%
<b>LEAP 2025 Grades 3 and 4 Science</b> Comprehensive (single part) ER items	0-6 Rubric	60%	93%
<b>LEAP 2025 Biology and Grades 5-8 Science</b> Comprehensive (single part) ER items	0-9 Rubric	60%	90%
<b>LEAP 2025 U.S. History and Grades 3-8 Social Studies</b> CR items	0-2	80%	95%
<b>LEAP 2025 U.S. History and Grades 5-8 Social Studies</b> 0-8 point, 2-dimension ER items (Content 0-4; Claims 0-4)	0-4 (each trait)	70%	95%



Each reader will be expected to maintain an acceptable level of exact agreement on validity responses and on inter-rater reliability as described above. Additionally, readers will be expected 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" in the tables on the previous page.

## 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 or supplied by New Meridian (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 the set 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
<b>Current Handscore</b>	<b>11,834</b>	<b>87</b>	<b>12</b>	<b>1</b>	<b>58,111</b>	<b>47</b>	<b>32</b>	<b>12</b>	<b>6</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
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	28	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)

#### Grade 10 English II Q000

#### Conventions

Totals	Inter-rater Reliability				Score Point Distribution										
	2X	%EX	%AD	%NA	Total	%0	%1	%2	%3	%B	%F	%N	%R	%T	%U
<b>Current Handscore</b>	<b>636</b>	<b>87</b>	<b>13</b>	<b>0</b>	<b>2,440</b>	<b>62</b>	<b>26</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>3</b>
3	318	87	13	0	2,122	63	27	3	0	0	0	2	0	0	2
11775	18	86	14	0	18	72	28	0	0	0	0	0	0	0	0
13021	36	81	19	0	36	64	31	0	0	0	0	6	0	0	0
16132	76	83	17	0	76	64	33	0	0	0	3	0	0	0	0
<b>18887</b>	<b>107</b>	<b>91</b>	<b>9</b>	<b>0</b>	<b>107</b>	<b>16</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>2</b>	<b>34</b>	<b>3</b>	<b>2</b>	<b>39</b>

## 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 25). 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

Handscored 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 definitive

proof of plagiarism in a response but suspect it to be likely, the response is scored using the rubric and processed as an alert. All responses with a possible plagiarism alert are sent to LDOE for final determination. (For additional information on final alert processing, see *Alerts* section below).

The non-score codes and the 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	Response does not contain enough original writing to evaluate. There is an insufficient amount of original writing to score and/or the response is composed of copied text. (Insufficient also means copied text that may have slight changes but does not introduce original ideas/thoughts.)
N	Don't understand/know
R	Refusal to respond
T	Off-topic
U	Incoherent, unintelligible, or undecipherable

### Nonscore Codes by Course

Course	B	F	I	N	R	T	U
LEAP 2025 Algebra I, English I, English II, Geometry, 3-8 ELA, and 3-8 math	✓	✓	N/A	✓	✓	✓	✓
EOC English III	✓	✓	✓	N/A	N/A	N/A	✓
EOC Biology	✓	✓	N/A	N/A	✓	N/A	✓
LEAP 2025 Biology, 3-8 Science, U.S. History, and 3-8 Social Studies ERs and CRs	✓	✓	✓	✓	✓	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, content area/course, 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 for LDOE to review. 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 students’ responses to the writing prompt for EOC English III and the extended-response items (ER) for LEAP 2025 U.S. History and grades 5-8 Social Studies. Pearson’s Intelligent Essay Assessor (IEA) is used to score student responses to selected constructed-response (CR) items in grades 6-8 ELA, English I, and English II.

### AI Scoring – Measurement, Inc.

The items in the following table will be AI scored by MI during the Spring 2019 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 2019 test may be found in the Appendix.)

Course	Item Type	IDEAS ID	Model Built
EOC English III	Writing Prompt	851370	Fall 2016
LEAP 2025 U.S. History	ER	892955	Fall 2017
LEAP 2025 U.S. History	ER	894104	Fall 2017
LEAP 2025 Grade 5 Social Studies	ER	807773**	Fall 2016
LEAP 2025 Grade 6 Social Studies	ER	804889*	Fall 2016
LEAP 2025 Grade 7 Social Studies	ER	805627*	Fall 2016
LEAP 2025 Grade 8 Social Studies	ER	808905*	Fall 2016
<p>*In spring 2017, human scored targeted samples of ≈ 500 responses per item used to augment and retrain the original AI models built in 2016. These samples were intended to find high score points to add to the existing AI models for the purpose of retraining the models prior to operational scoring in spring 2017.  **Similarly, the original 2016 model for grade 5 ER 807773 will be augmented prior to operational scoring in spring 2019 using a targeted sample of spring 2019 responses.</p>			

## 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<sup>2</sup> 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.

---

<sup>2</sup> 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  $\kappa\omega(\text{PEG}, \text{R1})$  and quadratic weighted kappa between Reader 1 and Reader 2 is referred to as  $\kappa\omega(\text{R1}, \text{R2})$ .

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  $\kappa\omega(\text{R1}, \text{R2})$  will usually lead to a low  $\kappa\omega(\text{PEG}, \text{R1})$ . 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  $\kappa\omega(\text{PEG}, \text{R1})$ . In cases of four or more levels (e.g. a score ranging from 1-4, or broader) a  $\kappa\omega(\text{PEG}, \text{R1})$  of 0.7 has become a rule of thumb as a go-no-go metric. In these broader scales, a  $\kappa\omega(\text{PEG}, \text{R1})$  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  $\kappa\omega(\text{PEG}, \text{R1})$  and  $\kappa\omega(\text{R1}, \text{R2})$  can be used as an additional evaluation metric. MI defines that value as follows:

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

When  $\Delta\kappa$  is positive, PEG's scores are more in agreement with Reader 1 than Reader 1's scores are in agreement with Reader 2. When  $\Delta\kappa$  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  $\Delta\kappa$  maintains its weight as a relative value between the two kappa values. That is, a larger  $\Delta\kappa$  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  $\geq 0.7$ ), 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.

$\Delta\kappa$  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,  $\kappa\omega$  and  $\Delta\kappa$  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 tables in Section 2 of the "DRC – Streaming Scoring" document (see Appendix) also contain 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 51-53 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 nonscorable

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 pages 35-36 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 Nonscorable Responses*

PEG's nonscorable 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 nonscorable parameters which can be configured in this way. Each nonscorable 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 controls 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 predominantly 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 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, EOC English III responses having an insufficient amount of

original writing to score will receive a nonscore of “I.” For LEAP 2025 U.S. History and grades 5-8 Social Studies ERs, 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 also receive a score of “I” (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 (this process is explained in the next paragraph), 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 EOC English III, this is the prompt and any associated reading material provided with each test item. For the LEAP 2025 U.S. History and grades 5-8 Social Studies ERs, this includes the prompt and any associated 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 configurations are noted in 2a–4b on page 36.

2. EOC 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. LEAP 2025 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. LEAP 2025 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 scores and apply the appropriate scores or nonscores as necessary, depending on the content area scoring rules. For EOC English III and LEAP 2025 U.S. History and Social Studies, flagged responses composed entirely of text copied from item source material (or copied text combined with an insufficient amount of original student work) 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 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 2019 LEAP 2025 administration. AI scoring models for each of these items were previously built and used by Pearson during PARCC operational scoring. (Model-building data for all items included on the Spring 2019 test may be found in the Appendix.)

Course	Task Type	IDEAS ID	PARCC UIN	Model Built
English I	LAT	902152	VH017536_2T	2017
English I	RST	914552	GG431834057	2018
English II	LAT	906197	HH428127697	2017
English II	NWT	983642	HH432845949	2017
Grade 6 ELA	RST	913715	DD502035970	2017
Grade 6 ELA	NWT	913694	D1466	2017
Grade 7 ELA	NWT	913842	EE430133306	2017
Grade 8 ELA	LAT	913958	F1460	2017

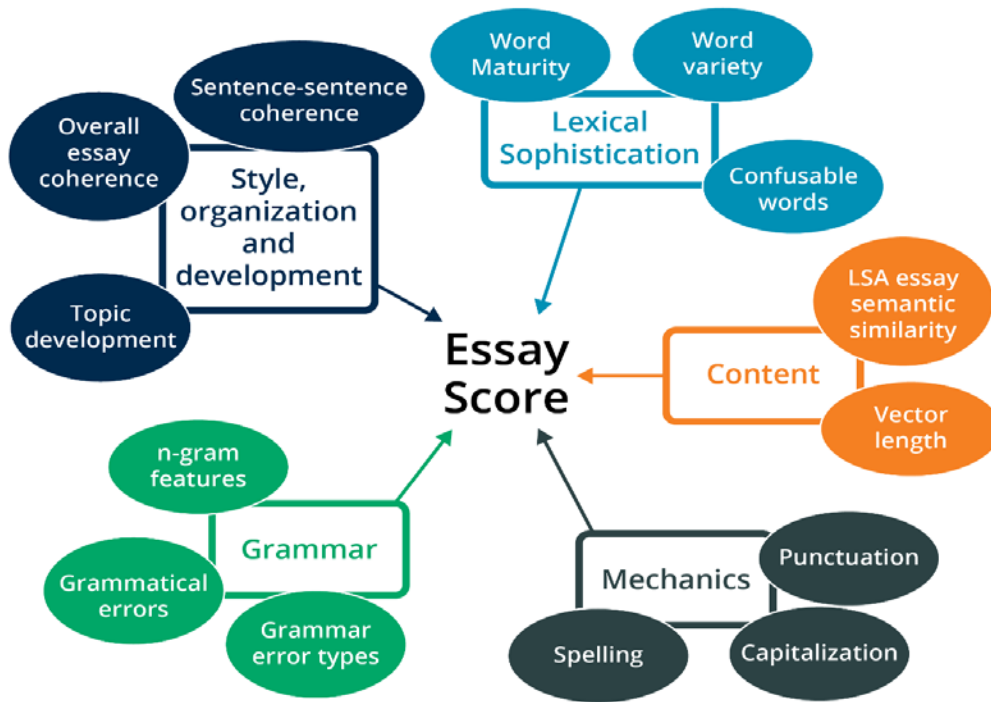
### *The Intelligent Essay Assessor*

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.

<b>Evaluation of Automated Scoring Systems</b>	
<b>Criterion</b>	<b>Threshold</b>
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.*

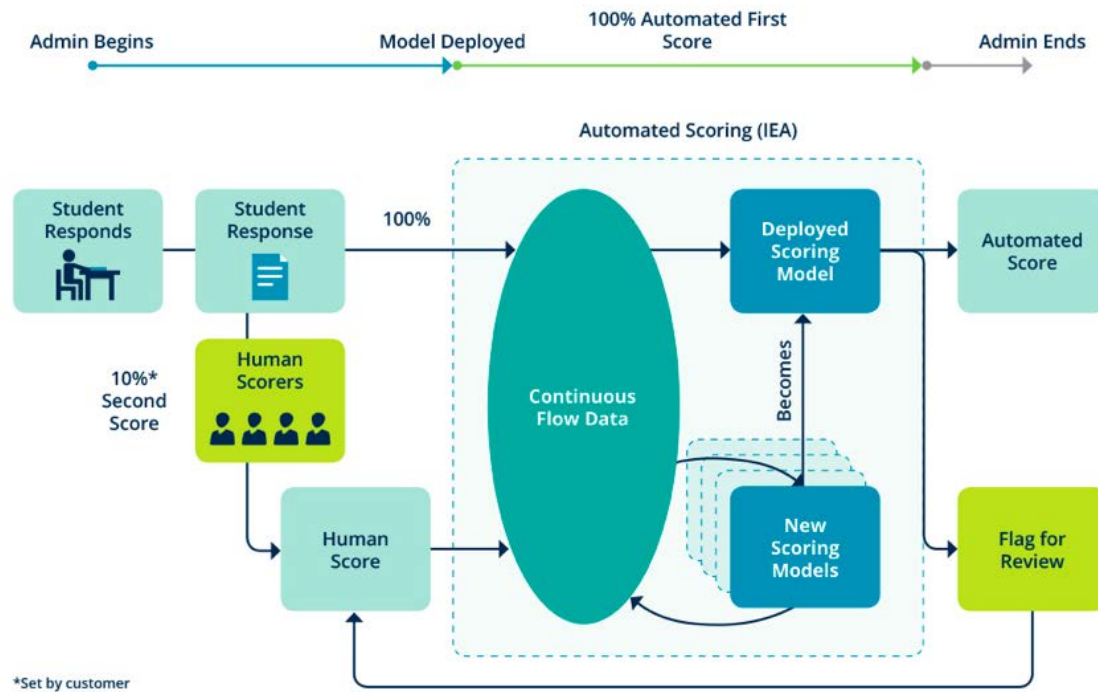
While the engine was being trained, scoring and psychometrics teams met daily to review progress, quality, and next steps. When IEA met or exceeded the performance criteria for a given constructed response item, it took over as the first scorer for that item.

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 on the following page depicts the entire Continuous Flow process.



## Continuous Flow Scoring



**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. The comparison is cumulative. Every response gets checked against every other response that has been received, as they come in, within that same administration and within that prompt. 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 2019 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, scoring supervisors 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 student responses.

## Rescores

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.

Please refer to [LEAP 2025 HS\\_EOC Processing Rules – Scoring.pdf](#) on the LDOE Reporting SFTP site at /<YYYY> - EOC LEAP 2025 HS <Spring>/Processing Rules - Final/ for a complete description of the rescore rules and process.

# 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 will 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	<i>Required.</i>

### 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	<i>Required.</i>

### 2.2.2 ConstructedResponseList

Name	Type	Min	Max	Description	Range
ConstructedResponse	ConstructedResponse	0	<i>unbounded</i>	List of individual CRs to be scored	<i>Required.</i> 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	<i>Required.</i>

### 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	<i>Required.</i> Will fall in the range 0-999. See <b>section 3</b> for more information
Description	string	0	1	Short description of status	<i>Required.</i> See <b>section 3</b> 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 Data – EOC English III (Spring 2019)

### Quadratic Weighted Kappa (QWK) and Inter-rater Reliability (IRR)

Course	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
EOC English III	851370	H-H*	0.69	61	36	3	0.77	62	37	1	0.27	76	24	0.44	72	28	0.39	77	23	0.49	82	18
		AI-H**		58	40	2		61	39	0		71	29		70	30		78	22		83	17

\*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	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%
EOC English III	851370	H	11	43	37	9	6	37	45	11	23	77	38	62	26	74	20	80
		AI	11	42	37	9	7	37	45	12	24	76	39	61	26	74	20	80

## AI Model Data – LEAP 2025 U.S. History ER (Spring 2019)

### 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	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

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 model was 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 2019)

### 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	807773	2599	0.89	H to H <sup>1</sup>	78	21	1	Human	62	25	12	2	0	0.88	H to H <sup>1</sup>	79	20	1	Human	67	23	9	1	0
		≈500		H to H <sup>3</sup>	92	7	1	Human	3	29	48	17	3		H to H <sup>3</sup>	91	8	1	Human	8	33	45	11	2
		15%		AI to H	77	23	1	AI	50	27	18	4	1		AI to H	77	23	1	AI	54	26	16	4	1
6	804889	2975	0.79	H to H <sup>1</sup>	67	32	1	Human	42	44	12	1	0	0.76	H to H <sup>1</sup>	68	31	1	Human	52	38	9	1	0
		≈500		H to H <sup>2</sup>	98	2	0	Human	7	28	50	14	1		H to H <sup>2</sup>	99	1	0	Human	14	47	32	6	1
		15%		AI to H	71	28	0	AI	38	43	16	2	1		AI to H	73	25	2	AI	52	35	11	1	0
7	805627	2610	0.83	H to H <sup>1</sup>	73	25	2	Human	45	41	12	2	0	0.83	H to H <sup>1</sup>	73	25	2	Human	57	31	11	2	0
		≈500		H to H <sup>2</sup>	98	1	0	Human	9	18	39	26	8		H to H <sup>2</sup>	98	1	1	Human	12	20	38	22	8
		15%		AI to H	71	29	1	AI	35	40	16	7	1		AI to H	74	25	2	AI	52	28	14	3	3
8	808905	2543	0.86	H to H	65	33	2	Human	30	36	25	7	2	0.86	H to H	64	34	2	Human	30	37	25	7	2
		≈500		H to H <sup>2</sup>	90	9	0	Human	1	6	34	35	24		H to H <sup>2</sup>	91	8	1	Human	1	7	35	34	23
		15%		AI to H	67	32	1	AI	25	33	24	13	5		AI to H	70	28	2	AI	21	37	26	12	4

H to H<sup>1</sup> – Human scored 2016 Field Test sample of ≈ 2500 responses per item.

H to H<sup>2</sup>, H to H<sup>3</sup> – Human scored targeted samples of ≈ 500 responses per item used to augment and retrain the original AI models from 2016. These samples come from spring operational responses and are intended to find high score points to add to the existing AI models for the purpose of retraining the models prior to operational scoring. H to H<sup>2</sup> augmentation sample was scored in spring 2017. H to H<sup>3</sup> augmentation sample is to be scored in spring 2019.

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.

## AI Model CR Performance – ELA Grades 6-8, English I, and English II (Spring 2019)

Prompt	Grade	Trait	IEA-Human Agreement					
			Exact	SP0	SP1	SP2	SP3	SP4
E06_N_D1466	6	1	Blue	Green	Blue	Blue	Blue	Blue
		2	Blue	Blue	Blue	Blue	Blue	Blue
E06_R_DD502035970	6	1	Blue	Green	Blue	Blue	Blue	Blue
		2	Blue	Blue	Blue	Blue	Blue	Blue
E07_N_EE430133306	7	1	Blue	Blue	Green	Blue	Blue	Blue
		2	Blue	Blue	Orange	Blue	Blue	Blue
E08_L_F1460	8	1	Blue	Blue	Blue	Blue	Blue	Blue
		2	Blue	Blue	Blue	Blue	Blue	Blue
E09_L_VH017536_2T	9	1	Blue	Blue	Blue	Blue	Blue	Blue
		2	Blue	Blue	Blue	Green	Blue	Blue
E09_R_GG431834057	9	1	Blue	Green	Blue	Blue	Blue	Blue
		2	Blue	Green	Blue	Blue	Blue	Blue
E10_L_HH428127697	10	1	Blue	Blue	Blue	Blue	Blue	Blue
		2	Blue	Green	Blue	Blue	Blue	Blue
E10_N_HH432845949	10	1	Blue	Blue	Blue	Blue	Blue	Blue
		2	Blue	Blue	Blue	Blue	Blue	Blue

- Trait 1 = Reading Comprehension and Written Expression or Written Expression
- Trait 2 = Conventions
- Blue indicates IEA-Human performance higher than Human-Human performance
- Green indicates IEA-Human performance is within 5.25% of Human-Human performance
- Orange indicates IEA-Human performance is more than 5.25% below Human-Human performance
- Source – Pearson

## Spring 2019 LEAP 2025 and EOC Items – IRR and SPD from Previous Administrations

### Algebra I

IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
901882	A, B	VH196970	Pearson Spring 2016	9,586	Part A	0,1	1,950	97.7%	99.9%	71.1%	12.5%				16.4%
				9,586	Part B	0,1,2	1,950	89.8%	97.1%	66.2%	6.7%	3.9%			23.1%
901882	A, B	VH196970	DRC Fall 2017, Op	8,522	Part A	0,1	1,940	99.0%	100.0%	94.0%	3.0%				4.0%
				8,522	Part B	0,1,2	1,940	99.0%	100.0%	94.0%	2.0%	1.0%			4.0%
901882	A, B	VH196970	DRC Spring 2018, Op	50,072	Part A	0,1	10,654	99.0%	100.0%	90.0%	8.0%				2.0%
				50,072	Part B	0,1,2	10,654	97.0%	100.0%	93.0%	3.0%	2.0%			2.0%
901882	A, B	VH196970	DRC Summer 2018, Op	1,625	Part A	0,1	372	99.0%	100.0%	97.0%	0.0%				3.0%
				1,625	Part B	0,1,2	372	99.0%	100.0%	96.0%	1.0%	0.0%			3.0%
901882	A, B	VH196970	DRC Fall 2018, Op	9,092	Part A	0,1	1,940	99.0%	100.0%	94.0%	3.0%				4.0%
				9,092	Part B	0,1,2	1,940	99.0%	100.0%	94.0%	2.0%	1.0%			4.0%
901836	A	M43318	DRC Fall 2017, Op	8,509	Overall	0,1,2,3	2,084	96.0%	100.0%	71.0%	12.0%	8.0%	3.0%		6.0%
901836	A	M43318	DRC Fall 2018, Op	9,062	Overall	0,1,2,3	2,084	96.0%	100.0%	71.0%	12.0%	8.0%	3.0%		6.0%
901814	A	M47147	DRC Fall 2017, Op	8,780	Part A	0,1,2	2,184	97.0%	100.0%	78.0%	8.0%	7.0%			8.0%
				8,780	Part B	0,1,2	2,184	99.0%	100.0%	88.0%	3.0%	1.0%			8.0%
901814	A	M47147	DRC Summer 2018, Op	1,637	Part A	0,1,2	412	97.0%	99.0%	88.0%	3.0%	1.0%			8.0%
				1,637	Part B	0,1,2	412	99.0%	100.0%	91.0%	1.0%	0.0%			8.0%
901859	A	3003-M43111	Pearson Spring 2016	253,395	Part C	0,1,2,3	48,917	92.1%	99.3%	43.7%	5.4%	14.9%	25.9%		10.1%
901859	A	3003-M43111	DRC Fall 2017, Op	8,485	Part C	0,1,2,3	2,504	98.0%	100.0%	73.0%	4.0%	7.0%	13.0%		2.0%
938769	A, D	MA10178	DRC Spring 2018, FT	1,579	Overall	0,1,2,3	324	94.0%	99.0%	65.0%	14.0%	13.0%	6.0%		2.0%
901848	A	M47287	Pearson Spring 2016	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%
901848	A	M47287	DRC Fall 2017, Op	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%
901848	A	M47287	DRC Summer 2018, Op	1,580	Overall	0,1,2,3,4	428	100.0%	100.0%	87.0%	1.0%	0.0%	0.0%	0.0%	12.0%
901857	A, B	VH046479	Pearson Spring 2017	78,418	Part A	0,1,2	13,963	88.2%	99.8%	51.2%	36.3%	2.9%			9.6%
				78,418	Part B	0,1	13,963	91.8%	99.7%	68.9%	19.0%			12.1%	
901857	A, B	VH046479	DRC Fall 2017, Op	8,686	Part A	0,1,2	2,258	94.0%	100.0%	77.0%	13.0%	1.0%			9.0%
				8,686	Part B	0,1	2,258	97.0%	100.0%	86.0%	5.0%			9.0%	
901857	A, B	VH046479	DRC Spring 2018, Op	8,686	Part A	0,1,2	2,258	94.0%	100.0%	77.0%	13.0%	1.0%			9.0%
				8,686	Part B	0,1	2,258	97.0%	100.0%	86.0%	5.0%			9.0%	
901857	A, B	VH046479	DRC Summer 2018, Op	49,959	Part A	0,1,2	11,927	88.0%	100.0%	57.0%	33.0%	4.0%			5.0%
				49,959	Part B	0,1	11,927	94.0%	100.0%	80.0%	14.0%			5.0%	
901857	A, B	VH046479	DRC Fall 2018, Op	1,623	Part A	0,1,2	396	92.0%	100.0%	80.0%	14.0%	0.0%			6.0%
				1,623	Part B	0,1	396	99.0%	100.0%	93.0%	1.0%			6.0%	

Form Key: Form A = Seniors only, Form B = Administrative Error (AE), Forms D and E = Operational

Algebra I (continued)

IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
901832	B, D	3031-M44083P	Pearson Spring 2016	95,907	Part B	0,1,2	18,835	91.3%	99.9%	29.9%	45.1%	11.9%			13.1%
901832	B, D	3031-M44083P	DRC Spring 2018, Op	55,162	Part B	0,1,2	10,236	91.0%	100.0%	32.0%	47.0%	21.0%			0.0%
901832	B, D	3031-M44083P	DRC Fall 2018, Op	6,329	Part B	0,1,2	1,140	92.0%	100.0%	51.0%	40.0%	9.0%			0.0%
938741	D	MA10144	DRC Spring 2018, FT	1,620	Overall	0,1,2,3	350	95.0%	99.0%	50.0%	10.0%	17.0%	11.0%		2.0%
980927	D, E	VH251952	Pearson Spring 2018	124,433	Part A	0,1,2	23,748	97.3%	99.6%	69.6%	15.0%	4.9%			10.6%
				124,433	Part B	0,1,2	23,748	95.4%	99.3%	72.0%	8.3%	5.7%			14.0%
				124,433	Part C	0,1,2	23,748	90.9%	98.8%	67.5%	11.5%	7.0%			13.9%
938735	D	MA10137	DRC Spring 2018, FT	1,655	Part B	0,1,2,3	316	94.0%	98.0%	79.0%	9.0%	7.0%	5.0%		0.0%
938744	D	MA10147	DRC Spring 2018, FT	1,606	Overall	0,1,2,3	344	90.0%	98.0%	67.0%	18.0%	4.0%	8.0%		1.0%
938737	B, D, E	MA10139	DRC Spring 2018, FT	1,582	Overall	0,1,2,3,4	382	94.0%	100.0%	71.0%	12.0%	4.0%	2.0%	5.0%	7.0%
938769	D	MA10178	DRC Spring 2018, FT	1,579	Overall	0,1,2,3	324	94.0%	99.0%	65.0%	14.0%	13.0%	6.0%		2.0%
980924	E	M44463	Pearson Spring 2017	77,183	Overall	0,1,2,3	14,754	87.6%	99.0%	36.9%	14.7%	30.4%	11.2%		6.8%
980909	E	M43216	Pearson Spring 2018	98,152	Overall	0,1,2,3	18,677	87.5%	99.3%	61.9%	13.9%	10.6%	3.7%		9.9%
980911	E	2679-M43312	Pearson 2015 FT	1,799	Part A	0,1,2	402	95.0%	99.5%	70.9%	12.4%	2.9%			13.9%
				1,799	Part B	0,1,2	402	94.5%	100.0%	19.2%	62.6%	3.3%			15.0%
901851	B, E	M41726	DRC Spring 2018, Op	52,490	Overall	0,1,2,3	11,918	92.0%	100.0%	57.0%	14.0%	15.0%	8.0%		6.0%
901851	B, E	M41726	DRC Fall 2018, Op	6,011	Overall	0,1,2,3	1,556	96.0%	100.0%	66.0%	11.0%	9.0%	4.0%		9.0%
980923	E	M000312	Pearson 2017 FT	1,593	Overall	0,1,2,3	264	89.0%	100.0%	65.1%	15.0%	7.6%	6.3%		6.1%
901687	B	2407-M41752	DRC Spring 2018, OP	53,117	Part A	0,1,2	11,413	98.0%	100.0%	74.0%	3.0%	19.0%			4.0%
				53,117	Part B	0,1,2	11,413	96.0%	100.0%	83.0%	7.0%	6.0%			4.0%
				53,117	Part C	0,1,2	11,413	98.0%	100.0%	89.0%	4.0%	3.0%			4.0%
901687	B	2407-M41752	DRC Spring 2018, OP	6,022	Part A	0,1,2	1,470	99.0%	100.0%	80.0%	2.0%	10.0%			7.0%
				6,022	Part B	0,1,2	1,470	99.0%	100.0%	87.0%	3.0%	2.0%			7.0%
				6,022	Part C	0,1,2	1,470	99.0%	100.0%	90.0%	1.0%	1.0%			7.0%
901705	B	VF883359	DRC Spring 2018, Op	53,281	Part A	0,1,2,3	11,808	98.0%	100.0%	89.0%	4.0%	1.0%	2.0%		5.0%
				53,281	Part B	0,1	11,808	93.0%	100.0%	84.0%	11.0%				5.0%
901705	B	VF883359	DRC Fall 2018, Op	6,097	Part A	0,1,2,3	1,570	100.0%	100.0%	87.0%	2.0%	1.0%			8.0%
				6,097	Part B	0,1	1,570	98.0%	100.0%	84.0%	7.0%				8.0%

Form Key: Form A = Seniors only, Form B = Administrative Error (AE), Forms D and E = Operational

Geometry

IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
902012	B, D, E	M41169	Pearson Spring 2016	90,471	Overall	0,1,2,3	16,723	87.1%	98.7%	46.2%	12.3%	14.6%	7.0%		19.9%
902012	B, D, E	M41169	DRC Spring 2018, Op	38,108	Overall	0,1,2,3	9,066	90.0%	100.0%	45.0%	15.0%	26.0%	9.0%		5.0%
902012	B, D, E	M41169	DRC Fall 2018, Op	5,823	Overall	0,1,2,3	1,424	96.0%	100.0%	47.0%	14.0%	23.0%	9.0%		7.0%
980937	D, E	M43798	Pearson Spring 2017	42,156	Overall	0,1,2,3	7,901	95.2%	99.5%	65.8%	14.1%	3.6%	1.2%		15.3%
939083	D	MGM0141	DRC Spring 2018, FT	1,592	Overall	0,1,2,3,4	354	95.0%	100.0%	70.0%	3.0%	6.0%	5.0%	11.0%	4.0%
980942	D	VH236248	Pearson 2016 FT	1,633	Part A	0,1,2,3	341	84.2%	98.2%	44.4%	25.3%	13.5%	6.1%		10.8%
				1,633	Part B	0,1,2,3	341	79.5%	97.7%	54.4%	16.8%	12.7%	3.4%		12.8%
939077	D	MGM0135	DRC Spring 2018, FT	1,595	Overall	0,1,2,3,4	356	95.0%	100.0%	70.0%	14.0%	7.0%	2.0%	2.0%	5.0%
980938	D, E	M100106	Pearson 2017 FT	1,635	Overall	0,1,2,3,4	314	93.0%	98.7%	73.8%	5.2%	5.7%	3.9%		11.4%
980936	D, E	VH239429	Pearson Spring 2017	42,154	Overall	0,1,2,3	8,173	84.3%	99.1%	71.6%	16.1%	3.6%	2.3%		6.4%
980929	E	M1000516	Pearson 2017 FT	1,612	Overall	0,1,2,3,4	314	87.9%	96.8%	63.1%	7.5%	6.8%	3.9%	6.8%	12.0%
902042	B, E	3020-M44058	Pearson Spring 2016	45,304	Part A	0,1,2,3	8,509	94.5%	99.7%	47.9%	29.7%	7.3%	4.1%		11.0%
				45,304	Part B	0,1	8,509	96.1%	99.8%	61.4%	21.9%				16.7%
				45,304	Part C	0,1,2	8,509	94.8%	97.7%	61.2%	4.7%	12.2%			21.9%
902042	B, E	3020-M44058	DRC Spring 2018, Op	38,085	Part A	0,1,2,3	8,517	96.0%	100.0%	55.0%	34.0%	5.0%	3.0%		4.0%
				38,085	Part B	0,1	8,517	97.0%	100.0%	78.0%	19.0%				4.0%
				38,085	Part C	0,1,2	8,517	97.0%	99.0%	79.0%	5.0%	13.0%			4.0%
902042	B, E	3020-M44058	DRC Fall 2018, Op	5,710	Part A	0,1,2,3	1,318	98.0%	100.0%	56.0%	30.0%	6.0%	2.0%		6.0%
				5,710	Part B	0,1	1,318	98.0%	100.0%	77.0%	17.0%				6.0%
				5,710	Part C	0,1,2	1,318	98.0%	99.0%	76.0%	5.0%	14.0%			6.0%
980930	E	M1000518	Pearson 2017 FT	1,500	Part B	0,1,2,3	298	95.3%	100.0%	60.4%	11.0%	11.9%	1.4%		15.2%
901939	A	M43794	DRC Fall 2017, Op	6,811	Overall	0,1,2,3	1,696	93.0%	100.0%	72.0%	10.0%	11.0%	2.0%		5.0%
901939	A	M43794	DRC Summer 2018, Op	450	Overall	0,1,2,3	162	98.0%	100.0%	74.0%	0.0%	2.0%	4.0%		19.0%

Form Key: Form A = Seniors only, Form B = Administrative Error (AE), Forms D and E = Operational

Geometry (continued)

IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
902046	A, B	M46668	Pearson Spring 2016	42,630	Overall	0,1,2,3	7,622	92.9%	98.9%	70.0%	8.5%	5.2%	0.5%		15.8%
902046	A, B	M46668	DRC Fall 2017, Op	6,821	Overall	0,1,2,3	1,880	97.0%	100.0%	78.0%	9.0%	3.0%	0.0%		9.0%
902046	A, B	M46668	DRC Spring 2018, Op	38,108	Overall	0,1,2,3	9,657	95.0%	100.0%	76.0%	10.0%	6.0%	1.0%		7.0%
902046	A, B	M46668	DRC Summer 2018, Op	423	Overall	0,1,2,3	148	99.0%	100.0%	74.0%	3.0%	3.0%	0.0%		19.0%
902046	A, B	M46668	DRC Fall 2018, Op	5,601	Overall	0,1,2,3	1,396	96.0%	100.0%	73.0%	9.0%	7.0%	1.0%		10.0%
902027	A, B	M43233	Pearson Spring 2017	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%
902027	A, B	M43233	DRC Spring 2018, Op	38,085	Overall	0,1,2,3,4	9,519	94.0%	100.0%	60.0%	13.0%	10.0%	5.0%	6.0%	7.0%
902027	A, B	M43233	DRC Summer 2018, Op	420	Overall	0,1,2,3,4	156	96.0%	100.0%	70.0%	3.0%	2.0%	1.0%	2.0%	22.0%
902027	A, B	M43233	DRC Fall 2018, Op	5,712	Overall	0,1,2,3,4	1,530	96.0%	100.0%	60.0%	10.0%	8.0%	5.0%	7.0%	9.0%
902036	B	2904-M43021	Pearson Spring 2016	42,708	Part A	0,1,2	8,216	95.9%	99.5%	53.9%	6.8%	33.2%			6.1%
				42,708	Part B	0,1,2	8,216	94.7%	99.2%	61.0%	6.5%	24.1%			8.5%
				42,708	Part C	0,1,2	8,216	94.9%	98.4%	75.2%	3.5%	4.2%			17.2%
902036	B	2904-M43021	DRC Fall 2017, Op	6,800	Part A	0,1,2	1,518	99.0%	100.0%	69.0%	9.0%	20.0%			2.0%
				6,800	Part B	0,1,2	1,518	96.0%	99.0%	68.0%	10.0%	19.0%			2.0%
				6,800	Part C	0,1,2	1,518	97.0%	99.0%	91.0%	3.0%	3.0%			2.0%
902036	B	2904-M43021	DRC Summer 2018, Op	433	Part A	0,1,2	110	100.0%	100.0%	84.0%	3.0%	6.0%			8.0%
				433	Part B	0,1,2	110	100.0%	100.0%	86.0%	1.0%	6.0%			8.0%
				433	Part C	0,1,2	110	100.0%	100.0%	89.0%	1.0%	2.0%			8.0%
902047	B	VH150404	Pearson Spring 2016	47,576	Part A	0,1,2	8,713	97.9%	99.6%	64.1%	6.6%	2.8%			26.5%
				47,576	Part B	0,1,2	8,713	92.1%	99.7%	52.6%	19.2%	7.0%			21.2%
902047	B	VH150404	DRC Fall 2017, Op	6,775	Part A	0,1,2	1,636	98.0%	100.0%	81.0%	10.0%	4.0%			5.0%
				6,775	Part B	0,1,2	1,636	96.0%	100.0%	74.0%	15.0%	6.0%			5.0%
902047	B	VH150404	DRC Summer 2018, Op	430	Part A	0,1,2	134	99.0%	100.0%	80.0%	2.0%	3.0%			14.0%
				430	Part B	0,1,2	134	100.0%	100.0%	77.0%	4.0%	5.0%			14.0%
939101	A, B	MGM0160	DRC Spring 2018, FT	1,665	Part C	0,1,2,3,4	336	80.0%	97.0%	73.0%	15.0%	8.0%	2.0%	1.0%	1.0%
902062	B	VH150384	Pearson Spring 2016	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%
902062	B	VH150384	DRC Spring 2018, Op	38,056	Overall	0,1,2,3,4	9,554	96.0%	100.0%	79.0%	9.0%	4.0%	1.0%	1.0%	7.0%
902062	B	VH150384	DRC Fall 2018, Op	5,747	Overall	0,1,2,3,4	1,452	97.0%	100.0%	76.0%	9.0%	4.0%	2.0%	1.0%	9.0%

Form Key: Form A = Seniors only, Form B = Administrative Error (AE), Forms D and E = Operational



Math Grade 3

IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
981736	Op	VH054794	Pearson Spring 2017	52,491	Part A	0,1,2	9,873	76.2%	98.9%	46.9%	33.0%	16.7%			3.4%
				52,491	Part B	0,1,2	9,885	82.8%	98.3%	35.4%	22.5%	37.8%			4.3%
914048	Op	M05158	Pearson Spring 2017	79,640	Overall	0,1,2,3	7,819	92.4%	99.3%	52.8%	18.5%	11.5%	15.7%		1.5%
914048	Op	M05158	DRC Spring 2018, Op	61,502	Overall	0,1,2,3	11,828	90.0%	100.0%	34.0%	30.0%	24.0%	6.0%		6.0%
898001	Op	N/A	DRC Spring 2018, FT	1,659	Part A	0,1,2	318	94.0%	100.0%	41.0%	21.0%	37.0%			1.0%
				1,659	Part B	0,1	318	98.0%	100.0%	95.0%	4.0%			1.0%	
981742	Op	M300388PD	Pearson 2017 FT	1,500	Part B	0,1,2	295	88.1%	98.3%	73.4%	7.3%	17.4%			1.9%
914039	Op	M02527	Pearson Spring 2017	7,113	Overall	0,1,2,3	699	92.7%	98.7%	37.9%	30.2%	23.4%	1.7%		6.8%
914039	Op	M02527	DRC Spring 2018, Op	61,394	Overall	0,1,2,3	11,578	88.0%	100.0%	18.0%	28.0%	45.0%	7.0%		1.0%
981747	Op	4127-M03599P	Pearson Spring 2018	102,233	Part B	0,1,2,3	20,403	90.8%	99.2%	48.2%	25.5%	8.8%	13.4%		4.1%
				102,233	Part C	0,1,2	20,403	92.4%	99.8%	32.9%	28.8%	33.1%			5.1%

Math Grade 4

IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
914084	Op	4112-M03491P	Pearson Spring 2017	383,723	Part C	0,1,2	37,737	94.9%	99.9%	65.0%	28.5%	2.5%			4.0%
914084	Op	4112-M03491P	DRC Spring 2018, Op Paper	5,830	Part C	0,1,2	1,238	96.0%	100.0%	67.0%	28.0%	3.0%			1.0%
914084	Op	4112-M03491P	DRC Spring 2018, Op Online	56,155	Part C	0,1,2	10,776	95.0%	100.0%	63.0%	28.0%	5.0%			4.0%
914086	Op	M04133	Pearson Spring 2017	107,359	Overall	0,1,2,3	10,670	91.1%	99.4%	53.1%	23.9%	7.4%	14.9%		0.7%
914086	Op	M04133	DRC Spring 2018, Op	61,742	Overall	0,1,2,3	11,702	95.0%	100.0%	54.0%	24.0%	7.0%	9.0%		5.0%
981831	Op	M400526	Pearson 2017 FT	1,500	Overall	0,1,2,3	288	85.8%	99.3%	47.2%	21.4%	22.1%	9.2%		0.1%
899959	Op	N/A	DRC Spring 2018, FT	1,622	Overall	0,1,2,3	302	82.0%	99.0%	34.0%	24.0%	11.0%	30.0%		0.0%
899955	Op	N/A	DRC Spring 2018, FT	1,651	Part A	0,1,2	306	88.0%	98.0%	39.0%	10.0%	49.0%			1.0%
				1,651	Part B	0,1	306	96.0%	100.0%	88.0%	11.0%				1.0%
981927	Op	0318-M01475	Pearson 2017 FT	1,500	Part A	0,1,2	300	98.7%	100.0%	54.8%	10.8%	33.6%			0.7%
				1,500	Part B	0,1,2	300	99.3%	100.0%	79.8%	3.3%	15.4%			1.6%
				1,500	Part C	0,1,2	300	93.7%	99.7%	64.2%	8.9%	24.5%			2.3%

Math Grade 5

IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
914152	Op	M03820	Pearson Spring 2017	216,578	Overall	0,1,2,3,4	43,004	76.0%	98.0%	26.0%	25.6%	22.4%	15.5%	9.3%	1.3%
914148	Op	M03888	Pearson Spring 2017	72,736	Overall	0,1,2,3	7,272	86.7%	98.9%	39.9%	27.5%	13.2%	18.7%		0.8%
914148	Op	M03888	DRC Spring 2018, Op	59,662	Overall	0,1,2,3	11,464	93.0%	99.0%	57.0%	22.0%	8.0%	12.0%		1.0%
902410	Op	N/A	DRC Spring 2018, FT	1,653	Part B	0,1,2	306	87.0%	100.0%	46.0%	20.0%	33.0%			1.0%
902414	Op	N/A	DRC Spring 2018, FT	1,651	Overall	0,1,2,3	318	87.0%	99.0%	63.0%	20.0%	7.0%			0.0%
914195	Op	0154-M00796	Pearson Spring 2017	92,904	Part B	0,1,2	9,282	95.9%	99.8%	80.4%	8.3%	6.4%			4.8%
914195	Op	0154-M00796	DRC Spring 2018, Op	61,037	Part B	0,1,2	11,260	91.0%	100.0%	75.0%	15.0%	10.0%			0.0%
934015	Op	N/A	DRC Spring 2018, FT	1,660	Part B	0,1	320	93.0%	100.0%	85.0%	15.0%				0.0%
				1,660	Part C	0,1,2,3,4	320	89.0%	98.0%	58.0%	19.0%	11.0%	4.0%	7.0%	0.0%

Math Grade 6

IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
981963	Op	M25151	Pearson Spring 2018	130,590	Overall	0,1,2,3,4	25,899	68.6%	96.5%	35.2%	23.2%	19.4%	13.4%	6.2%	2.6%
981961	Op	VH082639	Pearson 2015 FT	1,500	Part A	0,1,2	348	90.2%	100.0%	54.6%	26.9%	14.3%			4.2%
				1,500	Part B	0,1	348	90.8%	100.0%	53.9%	39.5%			6.6%	
981954	Op	VH139067	Pearson Spring 2017	111,824	Part A	0,1,2	21,162	93.0%	98.4%	78.8%	5.4%	11.5%			4.3%
				111,824	Part B	0,1,2,3,4	21,162	86.7%	98.3%	59.0%	15.5%	8.5%	3.8%	9.2%	4.0%
981956	Op	VH220482	Pearson Spring 2017	111,824	Part B	0,1,2	22,112	92.4%	99.3%	31.8%	15.7%	49.6%			2.8%
914231	Op	1740-M23030	Pearson Spring 2017	89,916	Overall	0,1,2,3	8,905	70.5%	96.2%	40.2%	18.0%	20.4%	19.0%		2.3%
914231	Op	1740-M23030	DRC Spring 2018, Op	58,067	Overall	0,1,2,3	11,448	74.0%	96.0%	43.0%	18.0%	19.0%	17.0%		2.0%
903511	Op	N/A	DRC Spring 2018, FT	1,652	Part B	0,1,2,3	310	85.0%	98.0%	76.0%	10.0%	10.0%	5.0%		0.0%
914281	Op	M25152	Pearson Spring 2017	112,484	Overall	0,1,2,3	11,247	89.0%	99.0%	53.8%	14.1%	12.3%	17.2%		2.6%
914281	Op	M25152	DRC Spring 2018, Op	57,609	Overall	0,1,2,3	11,534	91.0%	99.0%	63.0%	13.0%	8.0%	14.0%		2.0%

Math Grade 7

IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
914362	Op	VH083535	Pearson Spring 2016	100,577	Part A	0,1,2,3	19,892	90.4%	98.5%	74.7%	5.1%	4.9%	12.5%		2.8%
				100,577	Part B	0,1,2,3	19,892	90.3%	98.8%	70.8%	5.6%	5.6%	13.8%		4.3%
914362	Op	VH083535	DRC Spring 2018, Op	56,482	Part A	0,1,2,3	10,560	96.0%	100.0%	86.0%	3.0%	3.0%	7.0%		0.0%
				56,482	Part B	0,1,2,3	10,560	96.0%	100.0%	84.0%	3.0%	3.0%	9.0%		0.0%
982922	Op	M25544	Pearson 2015 FT	1,800	Overall	0,1,2,3	404	87.6%	99.0%	50.4%	13.7%	22.2%	7.0%		6.6%
868848	Op	M25578	Pearson Spring 2017	13,001	Overall	0,1,2,3	2,576	93.6%	99.0%	74.6%	5.4%	8.6%	1.4%		10.1%
900539	Op	N/A	DRC Spring 2018, FT	1,646	Part A	0,1,2	316	91.0%	99.0%	46.0%	37.0%	17.0%			0.0%
				1,646	Part B	0,1	316	97.0%	100.0%	62.0%	38.0%				0.0%
982929	Op	M22009	Pearson Spring 2018	124,808	Overall	0,1,2,3	24,757	83.2%	98.8%	45.6%	20.8%	20.4%	11.3%		2.0%
900520	Op	N/A	DRC Spring 2018, FT	1,624	Overall	0,1,2,3	348	97.0%	100.0%	77.0%	6.0%	4.0%	9.0%		3.0%
914339	Op	VH151385	Pearson Spring 2017	88,725	Part A	0,1,2	8,838	95.4%	99.4%	66.9%	7.8%	20.9%			4.4%
				88,725	Part B	0,1,2	8,838	95.5%	99.7%	77.2%	5.7%	9.7%			7.4%
914339	Op	VH151385	DRC Spring 2018, Op	56,454	Part A	0,1,2	10,887	98.0%	100.0%	73.0%	7.0%	19.0%			2.0%
				56,454	Part B	0,1,2	10,887	98.0%	100.0%	83.0%	6.0%	10.0%			2.0%

Math Grade 8

IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
983010	Op	VH097312	Pearson Spring 2018	28,653	Part A	0,1,2	5,561	95.7%	99.8%	63.6%	19.8%	8.2%			8.5%
				28,653	Part B	0,1,2,3,4	5,561	90.6%	98.9%	72.2%	9.5%	6.1%	1.6%	0.3%	10.4%
982987	Op	M800114	Pearson 2017 FT	1,500	Part A	0,1,2	300	93.3%	98.0%	73.6%	7.9%	15.3%			3.2%
				1,500	Part B	0,1,2	300	89.3%	98.7%	69.6%	12.2%	13.7%			4.5%
982999	Op	M22203	Pearson Spring 2017	69,637	Overall	0,1,2,3	13,500	84.2%	96.8%	54.6%	23.7%	8.5%	8.8%		4.4%
870899	Op	1282-M21381	Pearson Spring 2015	48,511	Part A	0,1,2	9,762	89.2%	97.5%	72.2%	9.3%	8.9%			9.7%
				48,511	Part B	0,1	9,762	91.0%	99.2%	65.5%	22.2%				12.2%
899312	Op	N/A	DRC Spring 2018, FT	1,648	Part B	0,1,2	318	85.0%	98.0%	27.0%	30.0%	43.0%			0.0%
914381	Op	M25425	Pearson Spring 2017	69,637	Overall	0,1,2,3,4	6,943	90.8%	99.1%	52.2%	12.5%	26.2%	1.9%	1.0%	6.2%
914381	Op	M25425	DRC Spring 2018, Op	49,280	Overall	0,1,2,3,4	10,088	94.0%	100.0%	59.0%	16.0%	20.0%	2.0%	0.0%	2.0%
899329	Op	N/A	DRC Spring 2018, FT	1,653	Part B	0,1	314	90.0%	100.0%	51.0%	49.0%				0.0%
				1,653	Part C	0,1	314	94.0%	100.0%	57.0%	43.0%				0.0%

English I

Task	IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
LAT	902152	D	VH017536_2T	Pearson Spr 17	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%
					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%		
LAT	902152	D	VH017536_2T	DRC Spr 18	51,374	RCWE	0,1,2,3,4	n/a	n/a	n/a	10,844	81.0%	100%	32.0%	44.0%	19.0%	3.0%	0%	2.0%
					51,374	Conv	0,1,2,3	n/a	n/a	n/a	10,844	82.0%	100%	26.0%	45.0%	23.0%	3.0%		
LAT	902152	D	VH017536_2T	DRC Fall 18	7,444	RCWE	0,1,2,3,4	n/a	n/a	n/a	1,870	86.0%	100%	44.0%	32.0%	16.0%	3.0%	0.0%	3.0%
					7,444	Conv	0,1,2,3	n/a	n/a	n/a	1,870	85.0%	100%	39.0%	34.0%	20.0%	4.0%		
RST	914552	D, E	GG431834057	Pearson Spr 18	66,624	RCWE	0,1,2,3,4	2,058	7,456	62,441	13,132	75.6%	99.6%	26.3%	28.9%	27.4%	11.3%	2.2%	4.0%
					66,624	Conv	0,1,2,3	2,058	7,456	62,441	13,132	76.1%	99.5%	25.2%	30.3%	27.4%	12.1%		
NWT	983215	E	GG604245591	Pearson 17 FT	1,696	Expr	0,1,2,3,4	1,430	155	0	299	74.9%	97.0%	24.7%	25.1%	26.3%	11.6%	5.0%	7.3%
					1,696	Conv	0,1,2,3	1,430	155	0	299	72.6%	100.0%	22.8%	22.8%	27.5%	14.7%		
RST	902161	A	VH017542_2T	Pearson Spr 17	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%
					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%		
RST	902161	A	VH017542_2T	DRC Fall 17*	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%
					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%		
RST	902161	A	VH017542_2T	DRC Spr 18	50,817	RCWE	0,1,2,3,4	n/a	n/a	n/a	10,136	81.0%	100%	17.0%	37.0%	32.0%	11.0%	1.0%	2.0%
					50,817	Conv	0,1,2,3	n/a	n/a	n/a	10,136	79.0%	100%	17.0%	36.0%	32.0%	13.0%		
RST	902161	A	VH017542_2T	DRC Fall 18	7,444	RCWE	0,1,2,3,4	n/a	n/a	n/a	1,870	84.0%	100%	30.0%	30.0%	24.0%	10.0%	1.0%	3.0%
					7,444	Conv	0,1,2,3	n/a	n/a	n/a	1,870	84.0%	100%	30.0%	29.0%	25.0%	12.0%		
NWT	906512	A	VH084830	Pearson Spr 17	61,936	Expr	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%
					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%		
NWT	906512	A	VH084830	DRC Fall 17*	5,047	Expr	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%
					5,047	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%		
RST	902194	C	VH017614_2T	Pearson Spr 17	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%
					3,179	Conv	0,1,2,3	3,012	317	0	620	79.4%	99.7%	47.8%	30.2%	11.7%	2.8%		
RST	902194	C	VH017614_2T	DRC Sum 18	1,546	RCWE	0,1,2,3,4	n/a	n/a	n/a	338	86.0%	100%	56.0%	32.0%	7.0%	1.0%	0%	4.0%
					1,546	Conv	0,1,2,3	n/a	n/a	n/a	338	82.0%	100%	57.0%	32.0%	7.0%	1.0%		
NWT	902203	C	6139	Pearson Spr 17	126,941	Expr	0,1,2,3,4	7,555	14,727	112,973	24,056	76.8%	99.6%	22.5%	33.7%	25.4%	9.3%	2.5%	6.6%
					126,941	Conv	0,1,2,3	7,555	14,727	112,973	24,056	76.4%	99.8%	26.8%	30.6%	25.9%	10.1%		
NWT	902203	C	6139	DRC Sum 18	1,510	WE	0,1,2,3,4	n/a	n/a	n/a	408	87.0%	100%	63.0%	27.0%	1.0%	0%	0%	9.0%
					1,510	Conv	0,1,2,3	n/a	n/a	n/a	408	93.0%	100%	73.0%	17.0%	1.0%	0%		

Form Key: Forms D and E = Operational, Form A = Seniors only, Form C = Administrative Error (AE)

\*Handscored by DRC in Fall of 2017

English II

Task	IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
LAT	906197	D	HH428127697	Pearson Spr 17	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%
LAT	906197	D	HH428127697	DRC Spr 18	48,997	RCWE	0,1,2,3,4	n/a	n/a	n/a	10,486	77.0%	99.0%	22.0%	37.0%	32.0%	6.0%	0.0%	3.0%
LAT	906197	D	HH428127697	DRC Fall 18	10,724	RCWE	0,1,2,3,4	n/a	n/a	n/a	2,872	80.0%	99.0%	38.0%	31.0%	21.0%	5.0%	1.0%	4.0%
RST	983688	D, E	HH607742252	Pearson 2017 FT	1,604	RCWE	0,1,2,3,4	1,487	162	0	312	78.2%	100.0%	28.1%	29.5%	20.2%	7.2%	2.2%	12.8%
NWT	983642	E	HH432845949	Pearson Spr 17	57,527	Expr	0,1,2,3,4	28,646	6,810	26,290	13,745	76.6%	99.6%	16.3%	23.2%	32.9%	15.6%	4.9%	7.1%
RST	902331	A	VH004490	Pearson Spr 17**	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%
RST	902331	A	VH004490	Pearson Spr 16**	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%
RST	902331	A	VH004490	DRC Fall 17*	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%
RST	902331	A	VH004490	DRC Spr 18	48,949	RCWE	0,1,2,3,4	n/a	n/a	n/a	10,460	79.0%	100.0%	15.0%	35.0%	34.0%	11.0%	2.0%	3.0%
RST	902331	A	VH004490	DRC Fall 18	10,714	RCWE	0,1,2,3,4	n/a	n/a	n/a	2,826	84.0%	100.0%	30.0%	33.0%	22.0%	9.0%	2.0%	3.0%
NWT	902354	A	7064	Pearson Spr 17	4,409	Expr	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%
NWT	902354	A	7064	DRC Fall 17*	9,721	Expr	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%
LAT	906181	C	HH431436431	Pearson Spr 17	57,534	RCWE	0,1,2,3,4	28,697	6,808	27,606	14,813	75.9%	98.9%	33.6%	37.4%	17.6%	5.6%	1.6%	4.2%
LAT	906181	C	HH431436431	DRC Sum18	2,632	RCWE	0,1,2,3,4	n/a	n/a	n/a	864	90.0%	100%	75.0%	18.0%	1.0%	0%	0%	6.0%
RST	906190	C	HH433954866	Pearson Spr 17	57,526	RCWE	0,1,2,3,4	26,197	5,981	27,528	13,108	76.8%	99.8%	28.7%	30.7%	21.0%	8.4%	1.9%	9.2%
RST	906190	C	HH433954866	DRC Sum18	2,440	RCWE	0,1,2,3,4	n/a	n/a	n/a	636	93.0%	99.0%	70.0%	19.0%	1.0%	0%	0%	9.0%

Form Key: Forms D and E = Operational, Form A =Seniors only, Form C = Administrative Error (AE)

\* 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.



EOC English III (All Report Data Scored by DRC and MI [AI])

IDEAS ID	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Trait	Score Points	Read 2x	Exact%	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	NS%
851370	Form W, Op	Op Fall 2017	8,382	Content	1, 2, 3, 4	3,672	91%	9%	0%		18%	34%	28%	11%	10%
			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%		
851368	Form X, AE	Op Fall 2016	7,783	Content	1, 2, 3, 4	2,204	83%	17%	0%		17%	39%	38%	5%	2%
			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 X, AE	Op Summer 2017	686	Content	1, 2, 3, 4	486	94%	6%	0%		56%	26%	2%	0%	15%
			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%		
851368	Form X, AE	Op Spring 2018	59,941	Content	1, 2, 3, 4	51,890	91%	9%	0%		14%	40%	38%	6%	2%
			59,941	Style	1, 2, 3, 4	51,890	93%	7%	0%		9%	32%	45%	13%	2%
			59,941	Formation (F)	0, 1	51,890	97%	3%		15%	83%				2%
			59,941	Usage (U)	0, 1	51,890	95%	5%		26%	73%				2%
			59,941	Mechanics (M)	0, 1	51,890	96%	4%		20%	78%				2%
59,941	Spelling (S)	0, 1	51,890	97%	3%		13%	86%					2%		
Form W (851370) will be AI scored by MI with human backreads by DRC. Form X (851368), the AE form, will be handscored by DRC supervisors.															

### ELA Grade 3

Task	IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
RST	915227	Op	A1598	Pearson 2016 FT	1,582	RCWE	0,1,2,3	n/a	n/a	n/a	339	69.3%	99.4%	52.7%	39.1%	7.1%	0.0%		1.1%
					1,582	Conventions	0,1,2,3	n/a	n/a	n/a	339	69.3%	98.2%	56.6%	32.7%	8.1%	1.5%		1.1%
NWT	913497	Op	AA431426588	Pearson Spring 17	118,416	Expression	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%
					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%
NWT	913497	Op	AA431426588	DRC Spring 18	62,260	Expression	0,1,2,3	n/a	n/a	n/a	13,242	80%	99%	31%	50%	13%	2%		4%
					62,260	Conventions	0,1,2,3	n/a	n/a	n/a	13,242	77%	99%	16%	58%	20%	2%		4%

### ELA Grade 4

Task	IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
LAT	913567	Op	VH170170	Pearson Spring 2017	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%
					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%
LAT	913567	Op	VH170170	DRC Spring 2018	62,127	RCWE	0,1,2,3,4	n/a	n/a	n/a	12,196	83%	100%	26%	36%	34%	3%	0%	1%
					62,127	Conventions	0,1,2,3	n/a	n/a	n/a	12,196	81%	100%	25%	36%	34%	4%		1%
RST	982233	Op	VH060330	Pearson 2017 FT	1,500	RCWE	0,1,2,3,4	1,468	150	0	300	77.7%	100.0%	26.0%	51.8%	17.5%	3.1%	0.0%	1.6%
					1,500	Conventions	0,1,2,3	1,468	150	0	300	78.0%	100.0%	19.7%	56.2%	19.4%	3.2%		1.6%

ELA Grade 5

Task	IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
LAT	801310	Op	VF821667	DRC Spring 2016	60,357	RCWE	0,1,2,3,4	n/a	n/a	n/a	14,914	77%	99%	45%	42%	11%	1%	0%	1%
					60,357	Conventions	0,1,2,3	n/a	n/a	n/a	14,914	75%	98%	24%	50%	22%	3%		1%
LAT	801310	Op	VF821667	Pearson Spring 2017	11,258	RCWE	0,1,2,3,4	11,045	1,127	0	2,231	86.8%	99.6%	79.7%	13.3%	1.2%	0.1%	0.0%	5.7%
					11,258	Conventions	0,1,2,3	11,045	1,127	0	2,231	81.6%	99.2%	64.7%	25.3%	3.9%	0.3%		5.7%
RST	915510	Op	VH198972	Pearson 2016 FT	1,561	RCWE	0,1,2,3,4	n/a	n/a	n/a	332	69.3%	100.0%	38.5%	40.9%	15.8%	3.7%	0.3%	0.8%
					1,561	Conventions	0,1,2,3	n/a	n/a	n/a	332	69.9%	98.8%	28.4%	42.9%	22.8%	5.1%		0.8%

ELA Grade 6

Task	IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
RST	913715	Op	DD502035970	Pearson Spring 2017	128,716	RCWE	0,1,2,3,4	36,320	13,240	93,042	29,065	72.5%	99.0%	32.1%	35.1%	24.5%	5.8%	1.0%	1.4%
					128,716	Conventions	0,1,2,3	36,320	13,240	93,042	29,065	71.3%	98.8%	32.2%	32.5%	25.6%	8.3%		1.4%
NWT	913694	Op	D1466	Pearson Spring 2017	127,628	Expression	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%
					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%
NWT	913694	Op	D1466	DRC Spring 2018	58,773	Expression	0,1,2,3,4	n/a	n/a	n/a	11,768	74%	99%	41%	24%	25%	6%	2%	0%
					58,773	Conventions	0,1,2,3	n/a	n/a	n/a	11,768	71%	99%	31%	38%	23%	6%		0%

ELA Grade 7

Task	IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
RST	915582	Op	E1567	Pearson Spring 2017	1,630	RCWE	0,1,2,3,4	n/a	n/a	n/a	345	75.7%	99.4%	30.8%	32.5%	22.9%	8.0%	2.8%	3.0%
					1,630	Conventions	0,1,2,3	n/a	n/a	n/a	345	76.2%	100.0%	30.5%	32.6%	23.2%	10.8%		3.0%
NWT	913842	Op	EE43013306	Pearson Spring 2017	128,845	Expression	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%
					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%
NWT	913842	Op	EE43013306	DRC Spring 2018	57,320	Expression	0,1,2,3,4	n/a	n/a	n/a	11,538	73%	99%	35%	13%	25%	18%	8%	0%
					57,320	Conventions	0,1,2,3	n/a	n/a	n/a	11,538	70%	99%	27%	23%	29%	20%		0%

ELA Grade 8

Task	IDEAS ID	Spring 2019 Form	PARCC UIN	Source of IRR and SPD Data	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 %
LAT	913958	Op	F1460	Pearson Spring 2017	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%
					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%
LAT	913958	Op	F1460	DRC Spring 2018	57,038	RCWE	0,1,2,3,4	n/a	n/a	n/a	12,090	73%	99%	18%	32%	35%	125%	1%	0%
					57,038	Conventions	0,1,2,3	n/a	n/a	n/a	12,090	76%	100%	14%	31%	39%	15%		0%
RST	982327	Op	FF506834510	Pearson 2017 FT	1,625	RCWE	0,1,2,3,4	1,496	165	0	317	74.8%	99.4%	42.8%	22.6%	16.7%	6.1%	2.5%	9.2%
					1,625	Conventions	0,1,2,3	1,496	165	0	317	74.1%	98.1%	35.4%	23.1%	23.4%	8.9%		9.2%

Biology (EOC)

IDEAS ID	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	NS%
845227	Form H, Op	DRC Spring 2017, OP	48,129	0-4	11,158	90%	8%	1%	42%	25%	18%	10%	3%	2%
845227	Form H, Op	DRC Fall 2017, OP	12,420	0-4	2,790	94%	6%	0%	47%	27%	13%	6%	3%	4%
845227	Form H, Op	DRC Summer 2018, OP	3,156	0-4	990	99%	1%	0%	60%	23%	4%	0%	0%	12%
845226	Form K, AE	DRC Fall 2016, OP	12,560	0-4	3,680	91%	9%	0%	45%	24%	11%	6%	3%	12%
845226	Form K, AE	DRC Summer 2017, OP	2,952	0-4	1,128	98%	2%	0%	61%	16%	2%	0%	0%	20%

Biology ERs and CRs (LEAP 2025)

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	SP 5 %	SP 6 %
965124	ER	B	Spring 2018 FT	4842	Part A (0-3)	4842	69%	29%	2%	9%	34%	36%	20%			
				4842	Part B (0-6)	4842	61%	26%	14%	36%	19%	21%	8%	10%	2%	3%
965129	CR	B, C	Spring 2018 FT	332	0-2	1566	81%	19%	1%	58%	29%	10%				
965237	CR	B, C	Spring 2018 FT	360	0-2	1607	96%	4%	0%	82%	14%	3%				
965295	CR	B, C	Spring 2018 FT	318	0-2	1622	76%	23%	1%	57%	33%	10%				
965286*	ER	A, C	Spring 2018 FT (re-scored Oct. 2018)	5,140	Part A (0-6)	5,140	82%	15%	3%	47%	13%	13%	15%	2%	1%	2%
				5,140	Part B (0-3)	5,140	84%	13%	3%	36%	30%	12%	16%			
965286	ER	A, C	Fall 2018 Op	7,446	Part A (0-6)	1,588	87%	10%	3%	55%	13%	13%	14%	2%	1%	1%
				7,446	Part B (0-3)	1,588	85%	14%	1%	41%	35%	11%	11%			
965190	CR	A	Spring 2018 FT	324	0-2	1626	84%	15%	1%	65%	20%	14%				
965190	CR	A	Fall 2018 Op	7,357	0-2	1,448	93%	7%	0%	78%	13%	7%				
965222	CR	A	Spring 2018 FT	350	0-2	1592	93%	7%	0%	64%	28%	4%				
965222	CR	A	Fall 2018 Op	7,279	0-2	1,516	92%	8%	0%	71%	25%	2%				
965279	CR	A	Spring 2018 FT	316	0-2	1625	75%	25%	1%	46%	31%	22%				
965279	CR	A	Fall 2018 Op	7,540	0-2	1,484	86%	14%	0%	57%	31%	11%				

Form Key: Form A = Administrative Error (AE), Forms B and C = Operational

All data from DRC 2018 field test handscoring. Nonscores excluded.  
 \*ER 965286 FT item was re-scored in October 2018 using updated rubric.

*Grade 3 Science*

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Score Points	Read 2x	Exact %	Adj %	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	SP 5 %	SP 6 %
957382	ER	Op	Spring 2018 FT	2,768	0-6	536	78%	18%	4%	63%	13%	15%	4%	4%	0%	0%
957435	CR	Op	Spring 2018 FT	1,660	0-2	320	87%	13%	0%	58%	33%	7%				
957418	CR	Op	Spring 2018 FT	1,661	0-2	322	88%	12%	0%	36%	61%	2%				
957409	CR	Op	Spring 2018 FT	1,675	0-2	350	87%	13%	0%	40%	40%	19%				

*Grade 4 Science*

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Score Points	Read 2x	Exact %	Adj %	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	SP 5 %	SP 6 %
957054	ER	Op	Spring 2018 FT	2,778	0-6	556	74%	23%	3%	6%	13%	40%	37%	3%	0%	0%
957144	CR	Op	Spring 2018 FT	1,668	0-2	326	88%	12%	0%	83%	14%	1%				
957090	CR	Op	Spring 2018 FT	1,665	0-2	330	79%	21%	0%	45%	49%	6%				
957099	CR	Op	Spring 2018 FT	1,657	0-2	314	96%	4%	0%	71%	25%	3%				

*Grade 5 Science*

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Score Points	Read 2x	Exact %	Adj %	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	SP 5 %	SP 6 %	SP 7 %	SP 8 %	SP 9 %
959503	ER	Op	Spring 2018 FT	4,992	0-9	4,992	67%	23%	10%	42%	12%	11%	9%	9%	6%	5%	3%	2%	1%
959557	CR	Op	Spring 2018 FT	1,667	0-2	346	89%	7%	4%	29%	51%	19%							
959548	CR	Op	Spring 2018 FT	1,658	0-2	324	96%	4%	1%	69%	12%	19%							
959530	CR	Op	Spring 2018 FT	1,690	0-2	382	98%	2%	0%	56%	7%	37%							

### Grade 6 Science

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	SP 5 %	SP 6 %
958421	ER	Op	Spring 2018 FT	4,988	Part A (0-3)	4,988	86%	8%	6%	68%	19%	0%	13%			
				4,988	Part B (0-3)	4,988	80%	19%	2%	58%	29%	11%	2%			
				4,988	Part C (0-3)	4,988	85%	12%	3%	62%	17%	19%	2%			
958378	CR	Op	Spring 2018 FT	1,652	0-2	314	86%	14%	0%	81%	14%	55				
958308	CR	Op	Spring 2018 FT	1,653	0-2	316	88%	11%	1%	68%	29%	3%				
958359	CR	Op	Spring 2018 FT	1,648	0-2	320	91%	95	0%	74%	20%	6%				

### Grade 7 Science

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	SP 5 %	SP 6 %
959635	ER	Op	Spring 2018 FT	4,952	Part A (0-3)	4,952	78%	16%	6%	71%	17%	10%	2%			
				4,952	Part B (0-4)	4,952	81%	15%	4%	71%	19%	8%	1%	0%		
				4,952	Part C (0-2)	4,952	96%	4%	0%	88%	10%	1%				
959748	CR	Op	Spring 2018 FT	1,646	0-2	312	82%	18%	0%	30%	50%	20%				
959657	CR	Op	Spring 2018 FT	1,651	0-2	332	92%	8%	0%	39%	42%	19%				
959715	CR	Op	Spring 2018 FT	1,647	0-2	336	92%	8%	0%	92%	6%	1%				

### Grade 8 Science

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %	SP 5 %	SP 6 %
959334	ER	Op	Spring 2018 FT	4,950	Part A (0-3)	4,950	62%	30%	8%	28%	28%	28%	15%			
				4,950	Part B (0-6)	4,950	49%	32%	20%	12%	13%	20%	21%	17%	10%	7%
959309	CR	Op	Spring 2018 FT	1,656	0-2	324	90%	10%	0%	87%	11%	1%				
959291	CR	Op	Spring 2018 FT	1,639	0-2	320	86%	13%	1%	42%	51%	6%				
959221	CR	Op	Spring 2018 FT	1,648	0-2	326	88%	12%	0%	76%	20%	3%				



U.S. History ERs and CRs (LEAP 2025)

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Trait	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %
894256	ER	B	Spring 2017 FT	5,000	Content	0-4	5,000	66%	32%	2%	28%	37%	24%	8%	3%
				5,000	Claims	0-4	5,000	64%	33%	3%	34%	33%	21%	8%	3%
892956	ER	B	Spring 2018 OP	16,722	Content	0-4	10142	93%	7%	0%	27%	30%	26%	11%	4%
				16,722	Claims	0-4	10142	93%	7%	0%	32%	27%	25%	10%	4%
892956	ER	B	Fall 2018 OP	12,519	Content	0-4	8,054	91%	9%	0%	39%	32%	19%	5%	2%
				12,519	Claims	0-4	8,054	92%	8%	0%	45%	27%	18%	5%	1%
894104	ER	A	Spring 2017 FT	5,000	Content	0-4	5,000	62%	33%	5%	31%	34%	22%	9%	4%
				5,000	Claims	0-4	5,000	61%	32%	7%	39%	28%	21%	9%	4%
894104	ER	A	Fall 2017 OP	7,649	Content	0-4	4028	90%	9%	0%	36%	34%	20%	6%	2%
				7,649	Claims	0-4	4028	89%	11%	0%	45%	30%	17%	6%	1%
894104	ER	A	Spring 2018 OP	14,069	Content	0-4	9990	96%	4%	0%	21%	34%	26%	11%	6%
				14,069	Claims	0-4	9990	95%	5%	0%	31%	33%	21%	10%	3%
894104	ER	A	Sum 2018 OP	215	Content	0-4	152	96%	4%	0%	75%	17%	6%	1%	0%
				215	Claims	0-4	152	99%	1%	0%	83%	12%	3%	1%	0%
892955	ER	F	Spring 2017 FT	5,000	Content	0-4	5,000	65%	32%	3%	34%	29%	25%	9%	3%
				5,000	Claims	0-4	5,000	64%	32%	4%	37%	26%	25%	10%	3%
892955	ER	F	Spring 2018 OP	10,506	Content	0-4	5,426	94%	6%	0%	16%	32%	31%	15%	3%
				10,506	Claims	0-4	5,426	93%	7%	0%	21%	28%	30%	15%	3%
894271	CR	F	Spring 2017 FT	1,658		0-2	316	66%	34%	1%	54%	37%	8%		
894271	CR	F	Spring 2018 FT	1,331		0-2	254	82%	18%	0%	29%	48%	23%		
957768	CR	F	Spring 2018 FT	1,557		0-2	294	86%	14%	0%	48%	27%	25%		
894225	CR	B	Spring 2017 FT	1,660		0-2	320	71%	29%	0%	44%	34%	22%		
894225	CR	B	Spring 2018 OP	39,705		0-2	7600	80%	19%	0%	55%	24%	21%		
894225	CR	B	Fall 2018 OP	9,205		0-2	1,694	88%	12%	0%	75%	15%	10%		
892994	CR	B	Spring 2017 FT	1,659		0-2	318	68%	31%	1%	13%	43%	44%		
892994	CR	B	Spring 2018 OP	39,867		0-2	7282	78%	22%	0%	22%	55%	23%		
892994	CR	B	Fall 2018 OP	9,375		0-2	1,728	80%	20%	0%	43%	39%	18%		
894188	CR	A	Fall 2017 OP	6,150		0-2	1,190	85%	15%	0%	53%	33%	14%		
894188	CR	A	Spring 2017 FT	1,655		0-2	310	74%	25%	1%	33%	38%	28%		
894188	CR	A	Spring 2018 FT	1,382		0-2	248	87%	13%	0%	55%	31%	13%		
894188	CR	A	Sum 2018 OP	154		0-2	30	73%	27%	0%	69%	28%	3%		
894149	CR	A	Spring 2017 FT	1,653		0-2	306	76%	21%	3%	44%	26%	30%		
894149	CR	A	Fall 2017 OP	6,056		0-2	1,132	87%	13%	0%	68%	18%	14%		
894149	CR	A	Spring 2018 FT	1,339		0-2	252	84%	16%	0%	64%	19%	17%		
894149	CR	A	Sum 2018 OP	145		0-2	26	92%	8%	0%	88%	10%	3%		
Form Key: Form A = Seniors only, Form B = Administrative Error (AE), Form F = Operational															
All data from DRC handscoring. Nonscores excluded.															
AI scoring models for ER items 892955 and 894104 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 2019 LEAP 2025 USH items (Op 2-point CRs and all AE form items) will be handscored.															

### Social Studies Grade 3

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Trait	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %
801184	CR	Op	Spring 2016 FT	1,281		0-2	248	78%	20%	2%	76%	15%	9%		
801184	CR	Op	Spring 2017 Op	62,961		0-2	11,436	89%	10%	1%	53%	18%	22%		
890683	CR	Op	Spring 2017 FT	1,654		0-2	308	81%	18%	2%	42%	38%	16%		

### Social Studies Grade 4

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Trait	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %
801539	CR	Op	Spring 2016 FT	1,654		0-2	308	71%	25%	3%	29%	37%	30%		
801539	CR	Op	Spring 2017 Op	62,340		0-2	11,406	82%	17%	1%	40%	36%	20%		
890820	CR	Op	Spring 2017 FT	1,654		0-2	308	85%	15%	0%	80%	17%	2%		

### Social Studies Grade 5

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Trait	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %
807773	ER	Op	Spring 2016 FT	5,668	Content	0-4	5,668	78%	21%	1%	62%	25%	12%	2%	0%
					Claims	0-4		79%	20%	1%	67%	23%	9%	1%	0%
890885	CR	Op	Spring 2017 FT	1,650		0-2	300	76%	23%	1%	54%	39%	6%		
890920	CR	Op	Spring 2017 FT	1,647		0-2	294	71%	29%	0%	63%	28%	9%		

### Social Studies Grade 6

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Trait	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %
804889	ER	Op	Spring 2016 FT	5,108	Content	0-4	5,108	67%	32%	1%	42%	44%	12%	1%	0%
					Claims	0-4		68%	31%	1%	52%	38%	9%	1%	0%
804889	ER	Op	Spring 2017 Op	71,724	Content	0-4	39,110	93%	6%	0%	56%	32%	10%	2%	0%
					Claims	0-4		93%	7%	0%	66%	24%	8%	1%	0%
804851	CR	Op	Spring 2016 FT	1,632		0-2	320	73%	28%	0%	46%	50%	5%		
804851	CR	Op	Spring 2017 Op	56,842		0-2	10,362	80%	20%	0%	41%	53%	6%		
949224	CR	Op	Spring 2018 FT	1,629		0-2	300	87%	13%	0%	45%	53%	2%		

### Social Studies Grade 7

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Trait	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %
805627	ER	Op	Spring 2016 FT	5,066	Content	0-4	5,066	73%	25%	2%	45%	41%	12%	2%	0%
					Claims	0-4		73%	25%	2%	57%	31%	11%	2%	0%
805627	ER	Op	Spring 2017 Op	68,833	Content	0-4	34,732	91%	9%	0%	48%	34%	13%	4%	1%
					Claims	0-4		91%	8%	0%	56%	28%	11%	4%	1%
891266	CR	Op	Spring 2017 FT	1,648		0-2	296	75%	25%	0%	43%	43%	14%		
805632	CR	Op	Spring 2016 FT	1,626		0-2	314	83%	17%	0%	42%	34%	24%		
805632	CR	Op	Spring 2017 Op	56,280		0-2	10,274	80%	19%	1%	47%	28%	25%		

### Social Studies Grade 8

IDEAS ID	Item Type	Spring 2019 Form	Source of IRR and SPD Data	Total Reads	Trait	Score Points	Read 2x	Exact %	Adj%	Non-Adj%	SP 0 %	SP 1 %	SP 2 %	SP 3 %	SP 4 %
808905	ER	Op	Spring 2016 FT	5,068	Content	0-4	5,068	65%	33%	2%	30%	36%	25%	7%	2%
					Claims	0-4		64%	34%	2%	30%	37%	25%	7%	2%
808905	ER	Op	Spring 2017 Op	65,286	Content	0-4	30,674	89%	11%	1%	32%	30%	25%	9%	3%
					Claims	0-4		88%	11%	1%	32%	29%	25%	9%	4%
808955	CR	Op	Spring 2016 FT	1,623		0-2	320	79%	21%	0%	39%	40%	21%		
808955	CR	Op	Spring 2017 Op	54,395		0-2	10,174	77%	22%	0%	32%	51%	17%		
892278	CR	Op	Spring 2017 FT	1,656		0-2	312	79%	20%	1%	43%	41%	15%		
892278	CR	Op	Spring 2018 Op	55,340		0-2	10,110	78%	21%	0%	43%	44%	13%		

## Appendix G—Classical Item Statistics

Table G.1 Operational Item Statistics—English I Fall Administration

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D
1	ESR	≥4,430	≥4,430	0.43	0.36	0.02	45.10	24.43	30.45		
2	TE	≥4,430	≥4,420	0.38	0.47	0.25	50.89	22.36	26.50		
3	ESR	≥4,430	≥4,430	0.61	0.31	0.05	29.91	18.73	51.32		
4	ESR	≥4,430	≥4,430	0.40	0.32	0.05	49.43	21.57	28.96		
5	ESR	≥4,430	≥4,430	0.77	0.35	0.02	19.04	8.83	72.10		
6	ESR	≥4,430	≥4,430	0.57	0.39	0.11	36.29	12.62	50.98		
7	CR	≥4,430	≥4,330	0.28	0.79	1.40	24.57	43.45	24.97	4.53	0.27
8	CR	≥4,430	≥4,330	0.28	0.79	1.40	24.57	43.45	24.97	4.53	0.27
9	CR	≥4,430	≥4,330	0.40	0.77	1.40	19.04	43.77	30.18	4.80	
10	ESR	≥4,430	≥4,420	0.33	0.36	0.18	59.54	13.73	26.55		
11	TE	≥4,430	≥4,420	0.70	0.44	0.36	11.99	36.20	51.45		
12	ESR	≥4,430	≥4,420	0.69	0.44	0.29	19.11	24.59	56.01		
13	ESR	≥4,430	≥4,420	0.47	0.36	0.25	47.42	10.84	41.49		
14	ESR	≥4,430	≥4,430	0.55	0.25	0.07	38.18	13.55	48.21		
15	ESR	≥4,430	≥4,430	0.42	0.39	0.09	54.00	8.77	37.14		
16	ESR	≥4,430	≥4,430	0.54	0.34	0.16	28.80	33.92	37.12		
17	TE	≥4,430	≥4,430	0.34	0.44	0.05	64.57	1.96	33.42		
18	ESR	≥4,430	≥4,430	0.46	0.49	0.05	37.30	34.01	28.65		
19	ESR	≥4,430	≥4,430	0.56	0.45	0.09	42.26	4.10	53.55		
20	ESR	≥4,430	≥4,430	0.73	0.44	0.05	16.90	19.97	63.08		
21	TE	≥4,430	≥4,430	0.38	0.44	0.05	43.11	37.05	19.79		
22	CR	≥4,430	≥4,320	0.41	0.80	1.65	11.72	31.96	35.54	16.16	2.14
23	CR	≥4,430	≥4,320	0.41	0.80	1.65	11.72	31.96	35.54	16.16	2.14
24	CR	≥4,430	≥4,320	0.55	0.79	1.65	11.90	30.29	36.56	18.77	
25	MS	≥4,430	≥4,430	0.54	0.34	0.02	26.36	39.66	33.96		
26	ESR	≥4,430	≥4,430	0.65	0.37	0.11	28.32	13.07	58.49		
27	ESR	≥4,430	≥4,430	0.50	0.34	0.11	38.13	24.58	37.18		
28	TE	≥4,430	≥4,430	0.52	0.28	0.05	44.07	7.50	48.38		
29	ESR	≥4,430	≥4,430	0.49	0.44	0.14	39.45	22.89	37.52		
30	TE	≥4,430	≥4,420	0.46	0.43	0.23	33.87	40.38	25.53		
31	ESR	≥4,430	≥4,420	0.20	0.20	0.34	66.06	28.22	5.39		
32	ESR	≥4,430	≥4,420	0.60	0.31	0.25	18.57	41.78	39.40		
33	ESR	≥4,430	≥4,420	0.38	0.46	0.23	53.23	16.59	29.95		
34	TE	≥4,430	≥4,420	0.43	0.50	0.29	41.92	29.19	28.60		

Table G.2 Operational Item Statistics—English II Fall Administration

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D
1	ESR	≥5,500	≥5,500	0.39	0.32	0.07	55.42	11.66	32.85		
2	TE	≥5,500	≥5,500	0.43	0.34	0.09	43.25	26.42	30.23		
3	ESR	≥5,500	≥5,490	0.23	0.09	0.22	67.71	18.54	13.53		
4	MS	≥5,500	≥5,490	0.39	0.49	0.24	52.79	16.18	30.80		
5	MS	≥5,500	≥5,490	0.21	0.29	0.18	68.31	22.08	9.42		
6	TE	≥5,500	≥5,490	0.38	0.33	0.24	27.67	67.44	4.65		
7	CR	≥5,500	≥5,310	0.35	0.81	2.54	17.41	35.14	34.63	8.43	0.89
8	CR	≥5,500	≥5,310	0.35	0.81	2.54	17.41	35.14	34.63	8.43	0.89
9	CR	≥5,500	≥5,310	0.46	0.80	2.54	17.78	34.50	34.17	10.04	
10	ESR	≥5,500	≥5,470	0.48	0.40	0.65	43.29	17.29	38.77		
11	ESR	≥5,500	≥5,450	0.53	0.33	0.89	40.71	10.97	47.43		
12	MS	≥5,500	≥5,440	0.28	0.33	1.05	49.41	44.47	5.07		
13	TE	≥5,500	≥5,400	0.27	0.45	1.92	56.75	30.07	11.26		
14	ESR	≥5,500	≥5,500	0.79	0.39	0.02	14.87	12.02	73.09		
15	ESR	≥5,500	≥5,500	0.31	0.28	0.09	51.15	35.03	13.73		
16	ESR	≥5,500	≥5,500	0.60	0.32	0.04	17.21	45.36	37.39		
17	TE	≥5,500	≥5,500	0.43	0.29	0.05	20.90	72.33	6.72		
18	MS	≥5,500	≥5,500	0.53	0.41	0.11	30.62	32.27	37.01		
19	ESR	≥5,500	≥5,500	0.80	0.43	0.02	14.31	12.35	73.32		
20	MS	≥5,500	≥5,490	0.39	0.47	0.15	34.12	54.42	11.31		
21	ESR	≥5,500	≥5,500	0.59	0.47	0.13	31.96	17.56	50.35		
22	CR	≥5,500	≥5,330	0.41	0.82	2.14	10.42	32.43	36.63	14.62	2.83
23	CR	≥5,500	≥5,330	0.41	0.82	2.14	10.42	32.43	36.63	14.62	2.83
24	CR	≥5,500	≥5,330	0.51	0.81	2.14	12.87	33.74	35.32	15.00	
25	ESR	≥5,500	≥5,500	0.76	0.34	0.02	17.45	12.26	70.27		
26	ESR	≥5,500	≥5,490	0.47	0.35	0.22	34.56	36.99	28.24		
27	ESR	≥5,500	≥5,500	0.60	0.49	0.09	23.33	33.63	42.95		
28	ESR	≥5,500	≥5,500	0.63	0.30	0.07	24.30	26.13	49.50		
29	MS	≥5,500	≥5,500	0.50	0.43	0.13	38.90	22.63	38.35		
30	MS	≥5,500	≥5,490	0.49	0.50	0.16	36.28	28.73	34.83		
31	MS	≥5,500	≥5,490	0.44	0.41	0.15	20.77	70.53	8.55		
32	TE	≥5,500	≥5,490	0.44	0.46	0.27	33.96	44.07	21.70		
33	MS	≥5,500	≥5,490	0.28	0.39	0.25	58.67	25.84	15.24		
34	TE	≥5,500	≥5,470	0.34	0.51	0.62	46.94	37.84	14.60		

Table G.3 Operational Item Statistics—Algebra I Fall Administration

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D	% at 5	% at 6
1	MC	≥3,130	≥3,120	0.52	0.30	0.22							
2	MC	≥3,130	≥3,130	0.51	0.45	0.19							
3	SA	≥3,130	≥3,070	0.26	0.51	2.04	72.26	25.70					
4	MC	≥3,130	≥3,120	0.21	0.24	0.35							
5	MC	≥3,130	≥3,120	0.35	0.05	0.38							
6	TE	≥3,130	≥3,120	0.16	0.22	0.22	69.64	27.55	2.58				
7	MPSR	≥3,130	≥3,110	0.30	0.40	0.73	45.85	47.96	5.45				
8	MC	≥3,130	≥3,120	0.53	0.34	0.35							
9	MC	≥3,130	≥3,130	0.30	0.25	0.19							
10	SA	≥3,130	≥3,130	0.66	0.62	0.06	11.00	12.31	14.89	24.08	37.66		
11	MC	≥3,130	≥3,130	0.47	0.36	0.13							
12	MC	≥3,130	≥3,120	0.37	0.13	0.26							
13	MC	≥3,130	≥3,130	0.30	0.14	0.10							
14	MC	≥3,130	≥3,120	0.51	0.39	0.22							
15	SA	≥3,130	≥3,130	0.49	0.61	0.13	35.14	31.28	33.45				
16	MC	≥3,130	≥3,110	0.38	0.12	0.61							
17	MC	≥3,130	≥3,120	0.34	0.24	0.38							
18	MC	≥3,130	≥3,130	0.35	0.21	0.10							
19	MS	≥3,130	≥3,130	0.36	0.45	0.13	63.81	36.07					
20	MC	≥3,130	≥3,120	0.38	0.09	0.29							
21	MPSR	≥3,130	≥3,130	0.28	0.18	0.10	52.84	38.81	8.26				
22	MC	≥3,130	≥3,120	0.35	0.18	0.22							
23	MC	≥3,130	≥3,130	0.72	0.40	0.06							
24	MC	≥3,130	≥3,130	0.36	0.30	0.10							
25	MC	≥3,130	≥3,120	0.44	0.16	0.26							
26	MPSR	≥3,130	≥3,130	0.30	0.13	0.03	49.04	41.96	8.96				
27	MC	≥3,130	≥3,130	0.34	0.40	0.16							
28	SA	≥3,130	≥3,060	0.16	0.24	2.39	82.17	15.43					
29	MPSR	≥3,130	≥3,130	0.43	0.26	0.16	31.89	49.84	18.11				
30	TE	≥3,130	≥3,090	0.34	0.51	1.21	47.23	35.65	15.91				
31	MC	≥3,130	≥3,120	0.42	0.22	0.35							
32	MC	≥3,130	≥3,120	0.41	0.09	0.35							
33	CR	≥3,130	≥2,800	0.24	0.61	6.70	53.54	14.41	14.32	7.27			
34	CR	≥3,130	≥3,130	0.54	0.60	0.10	14.45	23.44	47.77	14.25			
35	CR	≥3,130	≥2,990	0.05	0.45	2.93	86.29	5.90	2.49	0.92			
36	CR	≥3,130	≥2,860	0.10	0.54	5.84	70.50	3.19	8.71	3.41	3.00	1.24	1.21
37	CR	≥3,130	≥2,720	0.20	0.50	9.06	35.68	41.49	3.99	3.92	1.69		
38	CR	≥3,130	≥2,870	0.05	0.42	4.85	78.09	10.59	1.63	0.99	0.45		
39	CR	≥3,130	≥2,760	0.18	0.52	8.26	52.49	25.22	8.13	2.23			

Table G.4 Operational Item Statistics—Geometry Fall Administration

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D	% at 5	% at 6
1	MPSR	≥4,730	≥4,730	0.37	0.26	0.02	41.12	43.63	15.23				
2	TE	≥4,730	≥4,710	0.36	0.58	0.51	63.36	36.14					
3	MC	≥4,730	≥4,730	0.72	0.39	0.11							
4	MS	≥4,730	≥4,730	0.37	0.39	0.06	63.19	36.75					
5	TE	≥4,730	≥4,710	0.22	0.46	0.44	77.19	22.37					
6	TE	≥4,730	≥4,710	0.26	0.55	0.42	73.79	25.79					
7	MPSR	≥4,730	≥4,700	0.33	0.41	0.61	48.17	36.52	14.70				
8	MC	≥4,730	≥4,730	0.84	0.29	0.02							
9	SA	≥4,730	≥4,710	0.21	0.55	0.40	79.16	20.44					
10	TE	≥4,730	≥4,730	0.45	0.60	0.08	10.29	30.79	34.99	16.01	7.84		
11	SA	≥4,730	≥4,700	0.36	0.61	0.55	63.29	36.16					
12	CR	≥4,730	≥4,370	0.34	0.60	4.54	43.10	14.89	24.24	10.16			
13	CR	≥4,730	≥4,110	0.10	0.56	8.26	69.14	9.31	7.52	0.89			
14	SA	≥4,730	≥4,700	0.27	0.54	0.74	72.73	26.53					
15	MC	≥4,730	≥4,720	0.38	0.40	0.23							
16	SA	≥4,730	≥4,710	0.47	0.60	0.34	52.88	46.78					
17	MS	≥4,730	≥4,730	0.59	0.63	0.06	41.31	58.63					
18	MC	≥4,730	≥4,720	0.13	0.32	0.27							
19	SA	≥4,730	≥4,700	0.31	0.50	0.63	49.38	38.35	11.64				
20	MC	≥4,730	≥4,720	0.49	0.47	0.21							
21	MPSR	≥4,730	≥4,730	0.48	0.41	0.04	19.01	66.93	14.02				
22	MC	≥4,730	≥4,730	0.32	0.41	0.11							
23	TE	≥4,730	≥4,720	0.78	0.43	0.21	7.01	29.88	62.89				
24	CR	≥4,730	≥4,160	0.21	0.70	7.54	55.21	10.73	8.85	5.87	7.29		
25	TE	≥4,730	≥4,730	0.69	0.45	0.04	31.17	68.79					
26	SA	≥4,730	≥4,670	0.38	0.43	1.29	61.31	37.40					
27	MC	≥4,730	≥4,720	0.42	0.16	0.15							
28	MC	≥4,730	≥4,720	0.30	0.32	0.21							
29	MS	≥4,730	≥4,730	0.24	0.44	0.02	75.88	24.10					
30	MPSR	≥4,730	≥4,730	0.54	0.56	0.06	27.33	38.23	34.38				
31	MC	≥4,730	≥4,720	0.46	0.38	0.25							
32	SA	≥4,730	≥4,660	0.42	0.60	1.41	56.77	41.82					
33	TE	≥4,730	≥4,720	0.61	0.43	0.23	38.82	60.95					
34	TE	≥4,730	≥4,730	0.26	0.38	0.02	55.44	38.08	6.46				
35	MC	≥4,730	≥4,720	0.64	0.31	0.30							
36	CR	≥4,730	≥4,290	0.08	0.58	5.89	73.37	10.10	4.65	1.92	0.57		
37	CR	≥4,730	≥4,330	0.19	0.73	5.81	44.46	21.80	8.98	5.83	5.13	4.10	1.35
38	CR	≥4,730	≥4,220	0.24	0.70	6.97	50.01	4.29	8.17	13.71	2.79	6.34	3.86

Table G.5 Operational Item Statistics—English I Spring Administration Form D

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D
1	ESR	≥24,930	≥24,930	0.40	0.38	0.02	48.29	22.94	28.75		
2	TE	≥24,930	≥24,870	0.35	0.50	0.23	53.76	21.73	24.28		
3	ESR	≥24,930	≥24,900	0.59	0.32	0.10	32.08	18.75	49.06		
4	ESR	≥24,930	≥24,900	0.37	0.30	0.13	52.79	20.54	26.54		
5	ESR	≥24,930	≥24,900	0.73	0.40	0.12	21.64	10.18	68.06		
6	ESR	≥24,930	≥24,900	0.55	0.41	0.13	38.71	12.59	48.57		
7	CR	≥24,930	≥24,430	0.25	0.78	1.22	29.96	42.88	20.81	3.85	0.50
8	CR	≥24,930	≥24,430	0.25	0.78	1.22	29.96	42.88	20.81	3.85	0.50
9	CR	≥24,930	≥24,430	0.37	0.76	1.22	23.77	44.18	25.47	4.58	
10	ESR	≥24,930	≥24,860	0.43	0.26	0.28	54.40	5.02	40.30		
11	ESR	≥24,930	≥24,840	0.52	0.38	0.36	35.94	24.74	38.95		
12	ESR	≥24,930	≥24,820	0.56	0.51	0.43	36.84	14.80	47.93		
13	ESR	≥24,930	≥24,800	0.55	0.36	0.51	39.84	9.25	50.40		
14	ESR	≥24,930	≥24,920	0.43	0.34	0.04	53.49	7.03	39.44		
15	MS	≥24,930	≥24,880	0.37	0.54	0.20	48.51	28.90	22.40		
16	TE	≥24,930	≥24,890	0.42	0.36	0.17	38.80	38.86	22.17		
17	ESR	≥24,930	≥24,880	0.51	0.33	0.20	41.09	15.37	43.34		
18	ESR	≥24,930	≥24,880	0.45	0.47	0.19	50.61	7.83	41.36		
19	ESR	≥24,930	≥24,900	0.67	0.41	0.12	31.47	2.71	65.71		
20	MS	≥24,930	≥24,880	0.34	0.49	0.20	55.05	21.39	23.36		
21	TE	≥24,930	≥24,870	0.31	0.54	0.22	42.81	51.33	5.64		
22	CR	≥24,930	≥24,130	0.30	0.81	2.06	23.56	36.89	30.73	5.30	0.30
23	CR	≥24,930	≥24,130	0.30	0.81	2.06	23.56	36.89	30.73	5.30	0.30
24	CR	≥24,930	≥24,130	0.41	0.78	2.06	20.89	38.33	32.32	5.25	
25	ESR	≥24,930	≥24,920	0.79	0.35	0.04	16.99	8.46	74.51		
26	TE	≥24,930	≥24,910	0.46	0.22	0.10	48.36	12.07	39.47		
27	ESR	≥24,930	≥24,890	0.44	0.37	0.16	45.94	20.10	33.80		
28	ESR	≥24,930	≥24,900	0.52	0.36	0.14	40.74	13.47	45.65		
29	ESR	≥24,930	≥24,900	0.45	0.48	0.11	50.25	8.64	41.00		
30	ESR	≥24,930	≥24,900	0.39	0.38	0.11	42.09	37.53	20.26		
31	ESR	≥24,930	≥24,880	0.37	0.45	0.20	55.95	14.25	29.60		
32	ESR	≥24,930	≥24,870	0.33	0.34	0.23	55.17	22.81	21.79		
33	ESR	≥24,930	≥24,880	0.37	0.42	0.19	51.76	21.93	26.12		
34	TE	≥24,930	≥24,850	0.37	0.58	0.32	48.86	28.30	22.52		



**Table G.6 Operational Item Statistics—English I Spring Administration Form E**

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D
1	ESR	≥20,830	≥20,820	0.48	0.34	0.04	49.28	5.71	44.97		
2	MS	≥20,830	≥20,790	0.45	0.51	0.20	39.65	30.84	29.32		
3	TE	≥20,830	≥20,790	0.44	0.35	0.21	36.33	39.52	23.94		
4	ESR	≥20,830	≥20,770	0.57	0.26	0.29	37.19	10.56	51.96		
5	ESR	≥20,830	≥20,780	0.50	0.46	0.22	46.68	5.52	47.57		
6	ESR	≥20,830	≥20,780	0.74	0.30	0.22	25.34	1.98	72.46		
7	MS	≥20,830	≥20,750	0.41	0.48	0.37	47.93	22.56	29.13		
8	TE	≥20,830	≥20,740	0.36	0.50	0.42	35.52	57.38	6.69		
9	CR	≥20,830	≥20,430	0.36	0.77	1.41	13.43	38.04	38.92	7.16	0.54
10	CR	≥20,830	≥20,430	0.36	0.77	1.41	13.43	38.04	38.92	7.16	0.54
11	CR	≥20,830	≥20,430	0.48	0.73	1.41	11.49	39.11	40.10	7.38	
12	ESR	≥20,830	≥20,820	0.63	0.41	0.05	30.48	12.47	57.00		
13	TE	≥20,830	≥20,810	0.60	0.56	0.12	18.73	42.93	38.22		
14	MS	≥20,830	≥20,810	0.39	0.44	0.09	51.50	19.09	29.32		
15	TE	≥20,830	≥20,800	0.39	0.43	0.14	41.12	39.88	18.85		
16	CR	≥20,830	≥20,210	0.32	0.75	2.06	24.25	32.08	30.28	8.74	1.69
17	CR	≥20,830	≥20,210	0.44	0.73	2.06	19.89	35.45	31.36	10.33	
18	ESR	≥20,830	≥20,780	0.62	0.37	0.23	30.15	16.41	53.20		
19	ESR	≥20,830	≥20,810	0.44	0.44	0.11	53.22	6.33	40.35		
20	TE	≥20,830	≥20,810	0.64	0.51	0.12	32.04	8.50	59.34		
21	ESR	≥20,830	≥20,790	0.51	0.42	0.19	38.46	21.21	40.14		
22	TE	≥20,830	≥20,770	0.41	0.40	0.30	37.87	42.51	19.32		
23	MS	≥20,830	≥20,790	0.50	0.56	0.20	26.82	45.67	27.31		
24	ESR	≥20,830	≥20,820	0.83	0.29	0.07	12.76	8.24	78.93		
25	TE	≥20,830	≥20,810	0.46	0.20	0.12	47.59	12.90	39.40		
26	ESR	≥20,830	≥20,800	0.48	0.33	0.15	42.48	19.41	37.96		
27	ESR	≥20,830	≥20,810	0.56	0.32	0.11	37.37	12.22	50.30		
28	MS	≥20,830	≥20,800	0.46	0.33	0.16	34.66	38.16	27.02		
29	ESR	≥20,830	≥20,800	0.57	0.41	0.15	36.47	13.81	49.57		
30	ESR	≥20,830	≥20,790	0.44	0.34	0.18	44.27	23.04	32.51		
31	TE	≥20,830	≥20,780	0.50	0.33	0.22	45.87	7.15	46.75		
32	ESR	≥20,830	≥20,780	0.42	0.45	0.23	48.02	19.45	32.31		
33	TE	≥20,830	≥20,740	0.43	0.46	0.45	36.37	40.31	22.87		

Table G.7 Operational Item Statistics—English II Spring Administration Form D

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D
1	ESR	≥21,770	≥21,760	0.40	0.33	0.04	53.84	11.58	34.53		
2	TE	≥21,770	≥21,740	0.44	0.34	0.14	42.43	27.27	30.17		
3	ESR	≥21,770	≥21,730	0.23	0.15	0.17	67.10	19.45	13.28		
4	MS	≥21,770	≥21,730	0.38	0.49	0.18	53.49	16.27	30.05		
5	MS	≥21,770	≥21,730	0.21	0.29	0.16	66.84	23.17	9.83		
6	TE	≥21,770	≥21,720	0.39	0.35	0.20	27.48	67.51	4.81		
7	CR	≥21,770	≥21,000	0.34	0.79	2.41	17.42	34.29	36.43	7.75	0.61
8	CR	≥21,770	≥21,000	0.34	0.79	2.41	17.42	34.29	36.43	7.75	0.61
9	CR	≥21,770	≥21,000	0.46	0.78	2.41	17.76	33.16	36.76	8.82	
10	ESR	≥21,770	≥21,640	0.43	0.23	0.56	46.95	19.40	33.09		
11	ESR	≥21,770	≥21,630	0.51	0.38	0.64	41.21	15.53	42.63		
12	MS	≥21,770	≥21,590	0.40	0.38	0.81	48.98	21.56	28.65		
13	TE	≥21,770	≥21,480	0.30	0.47	1.31	50.50	37.22	10.97		
14	ESR	≥21,770	≥21,760	0.48	0.26	0.03	45.22	13.72	41.02		
15	TE	≥21,770	≥21,740	0.52	0.46	0.13	45.74	3.76	50.38		
16	MS	≥21,770	≥21,750	0.48	0.48	0.09	39.93	24.17	35.81		
17	ESR	≥21,770	≥21,720	0.37	0.41	0.19	40.07	45.99	13.75		
18	ESR	≥21,770	≥21,740	0.27	0.25	0.13	56.58	32.07	11.22		
19	ESR	≥21,770	≥21,730	0.68	0.38	0.17	18.52	27.83	53.49		
20	MS	≥21,770	≥21,730	0.28	0.45	0.18	54.74	34.42	10.66		
21	TE	≥21,770	≥21,730	0.63	0.46	0.14	31.78	10.83	57.25		
22	CR	≥21,770	≥20,950	0.34	0.82	2.57	21.13	28.42	39.26	7.10	0.36
23	CR	≥21,770	≥20,950	0.34	0.82	2.57	21.13	28.42	39.26	7.10	0.36
24	CR	≥21,770	≥20,950	0.45	0.82	2.57	20.68	29.13	38.03	8.42	
25	MS	≥21,770	≥21,750	0.55	0.49	0.07	27.96	33.12	38.85		
26	ESR	≥21,770	≥21,720	0.64	0.39	0.19	33.08	5.59	61.13		
27	ESR	≥21,770	≥21,720	0.46	0.36	0.19	49.38	9.40	41.04		
28	TE	≥21,770	≥21,730	0.42	0.36	0.18	40.64	34.67	24.50		
29	TE	≥21,770	≥21,730	0.34	0.48	0.17	42.06	46.89	10.88		
30	ESR	≥21,770	≥21,730	0.24	0.34	0.15	62.73	25.63	11.50		
31	ESR	≥21,770	≥21,730	0.40	0.46	0.17	49.99	19.51	30.32		
32	ESR	≥21,770	≥21,730	0.60	0.58	0.18	31.92	15.25	52.66		
33	TE	≥21,770	≥21,660	0.61	0.53	0.47	30.29	17.24	52.00		
34	ESR	≥21,770	≥21,710	0.38	0.40	0.27	33.71	55.75	10.27		

**Table G.8 Operational Item Statistics—English II Spring Administration Form E**

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D
1	ESR	≥18,920	≥18,920	0.50	0.24	0.03	42.82	13.35	43.80		
2	TE	≥18,920	≥18,890	0.57	0.43	0.15	41.25	3.62	54.98		
3	MS	≥18,920	≥18,900	0.52	0.43	0.12	35.70	24.43	39.76		
4	ESR	≥18,920	≥18,900	0.40	0.40	0.14	34.87	49.47	15.52		
5	ESR	≥18,920	≥18,900	0.30	0.24	0.10	52.08	35.23	12.59		
6	ESR	≥18,920	≥18,890	0.72	0.28	0.20	14.35	27.83	57.63		
7	MS	≥18,920	≥18,890	0.31	0.44	0.18	50.89	36.90	12.03		
8	TE	≥18,920	≥18,880	0.66	0.37	0.23	27.06	13.63	59.08		
9	CR	≥18,920	≥18,530	0.38	0.78	1.41	14.15	27.70	46.14	9.26	0.70
10	CR	≥18,920	≥18,530	0.38	0.78	1.41	14.15	27.70	46.14	9.26	0.70
11	CR	≥18,920	≥18,530	0.51	0.77	1.41	13.95	28.25	44.44	11.30	
12	ESR	≥18,920	≥18,910	0.61	0.46	0.08	30.74	17.40	51.78		
13	ESR	≥18,920	≥18,900	0.61	0.33	0.10	26.01	25.61	48.29		
14	MS	≥18,920	≥18,910	0.38	0.43	0.08	52.54	18.88	28.50		
15	ESR	≥18,920	≥18,900	0.55	0.35	0.13	40.64	9.46	49.78		
16	CR	≥18,920	≥18,320	0.45	0.72	2.45	8.07	25.01	43.97	16.22	3.53
17	CR	≥18,920	≥18,320	0.57	0.72	2.45	9.57	26.92	42.70	17.62	
18	ESR	≥18,920	≥18,880	0.52	0.37	0.22	44.09	7.82	47.87		
19	ESR	≥18,920	≥18,900	0.35	0.35	0.13	60.63	9.46	29.78		
20	ESR	≥18,920	≥18,900	0.46	0.31	0.12	39.06	28.77	32.05		
21	TE	≥18,920	≥18,900	0.30	0.34	0.14	46.95	46.21	6.70		
22	TE	≥18,920	≥18,840	0.72	0.60	0.43	17.62	20.16	61.78		
23	ESR	≥18,920	≥18,870	0.53	0.40	0.28	34.04	26.66	39.03		
24	MS	≥18,920	≥18,910	0.58	0.45	0.05	25.85	32.85	41.25		
25	ESR	≥18,920	≥18,890	0.70	0.36	0.16	27.40	4.73	67.71		
26	ESR	≥18,920	≥18,900	0.49	0.35	0.13	46.95	8.89	44.03		
27	TE	≥18,920	≥18,900	0.45	0.36	0.13	36.96	35.82	27.09		
28	MS	≥18,920	≥18,890	0.53	0.45	0.15	36.09	21.87	41.89		
29	MS	≥18,920	≥18,900	0.50	0.51	0.14	34.26	30.37	35.22		
30	MS	≥18,920	≥18,880	0.44	0.40	0.25	19.40	72.31	8.05		
31	TE	≥18,920	≥18,860	0.42	0.47	0.31	37.32	41.65	20.71		
32	MS	≥18,920	≥18,880	0.30	0.40	0.24	56.74	25.99	17.03		
33	TE	≥18,920	≥18,830	0.33	0.51	0.50	48.79	36.23	14.48		

Table G.9 Operational Item Statistics—Algebra I Spring Administration Form D

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D	% at 5	% at 6
1	MC	≥25,510	≥25,500	0.59	0.42	0.04							
2	MS	≥25,510	≥25,490	0.36	0.46	0.08	63.68	36.24					
3	SA	≥25,510	≥25,160	0.26	0.55	1.38	72.70	25.93					
4	MPSR	≥25,510	≥25,500	0.24	0.42	0.07	67.38	17.20	15.35				
5	MPSR	≥25,510	≥25,490	0.29	0.40	0.10	48.17	45.66	6.07				
6	MC	≥25,510	≥25,450	0.44	0.21	0.24							
7	MC	≥25,510	≥25,410	0.52	0.29	0.40							
8	MC	≥25,510	≥25,450	0.42	0.30	0.25							
9	MC	≥25,510	≥25,440	0.36	0.35	0.28							
10	SA	≥25,510	≥25,500	0.58	0.58	0.07	9.88	19.04	22.43	25.86	22.72		
11	MC	≥25,510	≥25,460	0.41	0.37	0.22							
12	CR	≥25,510	≥23,530	0.24	0.59	5.51	53.99	17.01	13.19	8.02			
13	MC	≥25,510	≥25,490	0.82	0.33	0.08							
14	MC	≥25,510	≥25,480	0.32	0.17	0.14							
15	TE	≥25,510	≥25,480	0.25	0.56	0.12	75.24	24.64					
16	TE	≥25,510	≥25,470	0.46	0.53	0.16	54.26	45.59					
17	MC	≥25,510	≥25,480	0.46	0.39	0.12							
18	MC	≥25,510	≥25,460	0.32	0.29	0.22							
19	MC	≥25,510	≥25,440	0.34	0.25	0.29							
20	MPSR	≥25,510	≥25,470	0.39	0.39	0.19	38.06	45.78	15.97				
21	MS	≥25,510	≥25,470	0.38	0.41	0.19	61.73	38.08					
22	SA	≥25,510	≥24,680	0.22	0.65	3.26	62.62	25.17	8.96				
23	MC	≥25,510	≥25,450	0.28	0.24	0.24							
24	MC	≥25,510	≥25,440	0.44	0.24	0.29							
25	MC	≥25,510	≥25,460	0.33	0.34	0.20							
26	MPSR	≥25,510	≥25,490	0.34	0.51	0.09	46.41	38.72	14.77				
27	MC	≥25,510	≥25,460	0.59	0.30	0.23							
28	SA	≥25,510	≥25,450	0.26	0.56	0.25	59.96	28.27	11.52				
29	MC	≥25,510	≥25,460	0.34	0.24	0.20							
30	MPSR	≥25,510	≥25,470	0.41	0.30	0.18	34.22	48.85	16.75				
31	TE	≥25,510	≥25,470	0.46	0.31	0.16	53.90	45.93					
32	MS	≥25,510	≥25,460	0.32	0.60	0.21	67.46	32.33					
33	MC	≥25,510	≥25,430	0.45	0.13	0.33							
34	CR	≥25,510	≥23,660	0.14	0.56	5.14	68.20	15.31	2.79	6.43			
35	CR	≥25,510	≥21,410	0.12	0.63	9.28	64.69	9.50	3.43	2.10	4.21		
36	CR	≥25,510	≥22,410	0.15	0.55	8.30	64.89	11.04	8.12	3.80			
37	CR	≥25,510	≥25,500	0.57	0.58	0.05	14.62	22.81	39.78	22.73			
38	CR	≥25,510	≥24,030	0.13	0.71	3.99	62.90	12.78	7.06	4.49	3.27	2.90	0.77
39	CR	≥25,510	≥25,390	0.14	0.58	0.48	64.78	21.88	6.09	3.71	3.06		

Table G.10 Operational Item Statistics—Algebra I Spring Administration Form E

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D	% at 5	% at 6
1	MC	≥20,850	≥20,830	0.64	0.42	0.05							
2	MS	≥20,850	≥20,830	0.41	0.44	0.05	59.06	40.88					
3	MC	≥20,850	≥20,820	0.35	0.21	0.11							
4	MC	≥20,850	≥20,830	0.25	0.32	0.08							
5	MPSR	≥20,850	≥20,830	0.31	0.40	0.08	44.20	48.50	7.22				
6	MPSR	≥20,850	≥20,820	0.39	0.40	0.14	38.88	44.77	16.21				
7	MC	≥20,850	≥20,800	0.55	0.31	0.23							
8	MC	≥20,850	≥20,800	0.43	0.31	0.23							
9	MC	≥20,850	≥20,800	0.35	0.29	0.21							
10	MPSR	≥20,850	≥20,840	0.54	0.59	0.02	6.63	22.25	31.48	25.99	13.63		
11	MC	≥20,850	≥20,800	0.57	0.30	0.23							
12	CR	≥20,850	≥19,920	0.39	0.63	3.13	36.73	18.47	27.80	12.56			
13	CR	≥20,850	≥18,240	0.17	0.62	9.02	60.54	13.76	8.74	4.48			
14	MC	≥20,850	≥20,830	0.89	0.25	0.07							
15	SA	≥20,850	≥19,910	0.29	0.58	4.50	67.37	28.13					
16	TE	≥20,850	≥20,810	0.28	0.56	0.16	71.63	28.21					
17	SA	≥20,850	≥20,800	0.52	0.58	0.20	32.15	31.45	36.21				
18	MC	≥20,850	≥20,820	0.50	0.37	0.12							
19	MS	≥20,850	≥20,780	0.10	0.43	0.30	89.80	9.90					
20	MC	≥20,850	≥20,790	0.37	0.24	0.28							
21	MPSR	≥20,850	≥20,830	0.42	0.46	0.07	34.84	45.49	19.60				
22	MS	≥20,850	≥20,810	0.40	0.40	0.19	60.02	39.78					
23	SA	≥20,850	≥20,280	0.45	0.52	2.71	53.75	43.54					
24	MC	≥20,850	≥20,780	0.28	0.21	0.30							
25	MC	≥20,850	≥20,790	0.46	0.27	0.28							
26	MC	≥20,850	≥20,800	0.31	0.48	0.21							
27	MC	≥20,850	≥20,800	0.47	0.17	0.21							
28	MC	≥20,850	≥20,810	0.63	0.31	0.19							
29	SA	≥20,850	≥20,810	0.27	0.53	0.15	58.08	28.85	12.92				
30	MC	≥20,850	≥20,820	0.46	0.18	0.12							
31	MPSR	≥20,850	≥20,830	0.50	0.39	0.08	22.85	54.27	22.80				
32	TE	≥20,850	≥20,610	0.38	0.56	1.15	43.63	35.84	19.39				
33	MC	≥20,850	≥20,800	0.46	0.20	0.21							
34	MC	≥20,850	≥20,810	0.34	0.44	0.19							
35	CR	≥20,850	≥19,330	0.26	0.62	5.25	57.09	9.94	16.00	9.72			
36	CR	≥20,850	≥17,870	0.16	0.65	8.71	60.38	11.77	4.64	2.79	6.15		
37	CR	≥20,850	≥18,760	0.15	0.62	7.12	64.87	14.44	6.00	4.68			
38	CR	≥20,850	≥19,870	0.15	0.70	3.31	58.79	14.54	8.47	5.31	4.17	3.35	0.68
39	CR	≥20,850	≥18,930	0.27	0.49	6.83	22.62	47.59	14.91	3.17	2.51		

Table G.11 Operational Item Statistics—Geometry Spring Administration Form D

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D	% at 5	% at 6
1	MPSR	≥18,640	≥18,630	0.35	0.25	0.04	42.27	44.67	13.02				
2	MC	≥18,640	≥18,620	0.45	0.26	0.12							
3	MC	≥18,640	≥18,620	0.69	0.37	0.12							
4	TE	≥18,640	≥18,620	0.50	0.60	0.10	49.83	50.07					
5	SA	≥18,640	≥18,330	0.38	0.55	1.66	60.93	37.41					
6	SA	≥18,640	≥18,350	0.31	0.52	1.56	67.99	30.45					
7	MPSR	≥18,640	≥18,560	0.41	0.17	0.43	33.38	50.70	15.49				
8	TE	≥18,640	≥18,620	0.36	0.57	0.14	63.67	36.19					
9	MC	≥18,640	≥18,620	0.60	0.32	0.13							
10	SA	≥18,640	≥18,630	0.39	0.64	0.07	21.84	28.40	28.43	14.24	7.02		
11	SA	≥18,640	≥18,510	0.20	0.57	0.72	79.55	19.73					
12	CR	≥18,640	≥17,300	0.33	0.61	4.42	44.18	15.36	24.21	9.03			
13	CR	≥18,640	≥16,280	0.06	0.53	8.63	75.49	9.48	2.04	0.31			
14	MC	≥18,640	≥18,630	0.49	0.32	0.06							
15	MC	≥18,640	≥18,620	0.39	0.38	0.13							
16	SA	≥18,640	≥18,580	0.43	0.57	0.34	56.99	42.67					
17	MC	≥18,640	≥18,620	0.35	0.31	0.13							
18	MS	≥18,640	≥18,630	0.33	0.57	0.06	67.28	32.66					
19	TE	≥18,640	≥18,640	0.37	0.29	0.03	34.86	55.49	9.62				
20	MC	≥18,640	≥18,620	0.44	0.32	0.11							
21	MPSR	≥18,640	≥18,620	0.54	0.34	0.11	22.96	46.61	30.32				
22	CR	≥18,640	≥16,860	0.20	0.71	6.23	63.67	5.23	6.68	5.28	9.56		
23	MPSR	≥18,640	≥18,630	0.49	0.32	0.06	25.52	49.99	24.42				
24	MC	≥18,640	≥18,620	0.34	0.47	0.11							
25	MC	≥18,640	≥18,620	0.46	0.40	0.11							
26	SA	≥18,640	≥18,450	0.35	0.47	1.06	64.00	34.94					
27	MC	≥18,640	≥18,620	0.40	0.15	0.10							
28	MC	≥18,640	≥18,630	0.30	0.30	0.08							
29	MS	≥18,640	≥18,630	0.22	0.44	0.09	77.83	22.08					
30	TE	≥18,640	≥18,620	0.26	0.28	0.13	55.41	37.36	7.09				
31	SA	≥18,640	≥18,510	0.41	0.58	0.69	58.66	40.64					
32	SA	≥18,640	≥18,390	0.38	0.60	1.34	60.80	37.87					
33	TE	≥18,640	≥18,620	0.47	0.50	0.13	52.48	47.39					
34	MPSR	≥18,640	≥18,630	0.51	0.59	0.09	35.50	26.28	38.13				
35	TE	≥18,640	≥18,540	0.19	0.47	0.53	81.00	18.47					
36	CR	≥18,640	≥16,450	0.09	0.59	7.35	69.22	11.40	4.78	1.71	1.15		
37	CR	≥18,640	≥16,000	0.11	0.60	9.17	72.89	3.59	3.49	5.87			
38	CR	≥18,640	≥16,930	0.10	0.60	6.15	72.60	9.81	6.60	1.81			
39	CR	≥18,640	≥17,480	0.16	0.76	3.94	57.76	13.01	8.12	5.74	5.81	2.23	1.10

Table G.12 Operational Item Statistics—Geometry Spring Administration Form E

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D	% at 5	% at 6
1	TE	≥16,630	≥16,620	0.29	0.39	0.01	49.70	42.10	8.18				
2	MC	≥16,630	≥16,610	0.47	0.25	0.10							
3	MS	≥16,630	≥16,600	0.36	0.57	0.18	63.84	35.97					
4	TE	≥16,630	≥16,600	0.55	0.58	0.13	45.14	54.72					
5	TE	≥16,630	≥16,550	0.23	0.45	0.46	77.12	22.42					
6	SA	≥16,630	≥16,360	0.33	0.51	1.63	65.49	32.88					
7	MPSR	≥16,630	≥16,530	0.32	0.36	0.57	49.32	37.00	13.11				
8	MC	≥16,630	≥16,600	0.48	0.20	0.13							
9	MC	≥16,630	≥16,610	0.64	0.29	0.11							
10	MPSR	≥16,630	≥16,620	0.41	0.46	0.02	11.15	32.83	37.53	16.94	1.53		
11	SA	≥16,630	≥16,510	0.22	0.56	0.68	77.54	21.78					
12	CR	≥16,630	≥15,420	0.34	0.60	4.41	42.49	15.56	24.36	10.32			
13	CR	≥16,630	≥14,520	0.06	0.52	8.82	75.32	9.48	2.10	0.44			
14	MC	≥16,630	≥16,620	0.51	0.32	0.07							
15	MC	≥16,630	≥16,590	0.39	0.36	0.25							
16	SA	≥16,630	≥16,550	0.52	0.39	0.48	47.48	52.04					
17	MS	≥16,630	≥16,610	0.60	0.59	0.11	39.88	60.01					
18	MS	≥16,630	≥16,610	0.36	0.55	0.07	64.39	35.54					
19	TE	≥16,630	≥16,620	0.37	0.30	0.03	35.66	54.44	9.87				
20	MC	≥16,630	≥16,600	0.46	0.44	0.15							
21	MPSR	≥16,630	≥16,620	0.48	0.42	0.04	18.41	67.85	13.71				
22	CR	≥16,630	≥14,330	0.16	0.67	9.54	60.78	8.74	7.49	4.83	4.38		
23	MPSR	≥16,630	≥16,610	0.52	0.33	0.10	22.78	49.66	27.46				
24	MC	≥16,630	≥16,600	0.38	0.47	0.13							
25	MC	≥16,630	≥16,610	0.58	0.50	0.10							
26	MC	≥16,630	≥16,600	0.43	0.32	0.17							
27	MC	≥16,630	≥16,610	0.49	0.25	0.13							
28	MC	≥16,630	≥16,590	0.30	0.35	0.22							
29	MC	≥16,630	≥16,600	0.34	0.30	0.17							
30	TE	≥16,630	≥16,610	0.26	0.29	0.11	55.18	36.86	7.85				
31	MC	≥16,630	≥16,610	0.38	0.31	0.11							
32	TE	≥16,630	≥16,610	0.33	0.53	0.13	66.60	33.27					
33	MC	≥16,630	≥16,600	0.49	0.26	0.13							
34	MPSR	≥16,630	≥16,610	0.55	0.59	0.10	32.17	25.98	41.75				
35	SA	≥16,630	≥16,340	0.15	0.59	1.70	83.84	14.46					
36	CR	≥16,630	≥14,210	0.12	0.59	9.49	71.07	3.98	3.89	6.55			
37	CR	≥16,630	≥15,050	0.11	0.58	6.75	70.18	11.43	7.32	1.59			
38	CR	≥16,630	≥15,680	0.22	0.75	3.93	41.86	21.84	11.24	5.53	6.18	5.41	2.23
39	CR	≥16,630	≥16,600	0.18	0.72	0.16	61.24	18.44	9.66	9.49	1.02		

Table G.13 Operational Item Statistics—English I Summer Administration

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D
1	ESR	≥90	≥90	0.47	0.34	0.00	43.75	18.75	37.50		
2	ESR	≥90	≥90	0.33	0.45	0.00	62.50	9.38	28.13		
3	ESR	≥90	≥90	0.51	0.38	0.00	32.29	33.33	34.38		
4	TE	≥90	≥90	0.23	0.55	0.00	76.04	1.04	22.92		
5	ESR	≥90	≥90	0.38	0.58	0.00	46.88	31.25	21.88		
6	ESR	≥90	≥90	0.41	0.44	2.08	55.21	5.21	37.50		
7	ESR	≥90	≥90	0.69	0.39	0.00	21.88	18.75	59.38		
8	TE	≥90	≥90	0.31	0.46	0.00	53.13	31.25	15.63		
9	CR	≥90	≥80	0.24	0.72	7.29	29.17	43.75	13.54	5.21	
10	CR	≥90	≥80	0.24	0.72	7.29	29.17	43.75	13.54	5.21	
11	CR	≥90	≥80	0.29	0.71	7.29	32.29	42.71	12.50	4.17	
12	ESR	≥90	≥90	0.43	0.55	2.08	42.71	27.08	28.13		
13	MS	≥90	≥90	0.32	0.64	2.08	51.04	30.21	16.67		
14	TE	≥90	≥90	0.32	0.20	0.00	47.92	40.63	11.46		
15	MS	≥90	≥90	0.38	0.20	1.04	47.92	27.08	23.96		
16	CR	≥90	≥80	0.16	0.78	13.54	47.92	20.83	11.46	2.08	1.04
17	CR	≥90	≥80	0.23	0.73	13.54	44.79	22.92	11.46	4.17	
18	ESR	≥90	≥90	0.38	0.19	1.04	43.75	35.42	19.79		
19	ESR	≥90	≥90	0.36	0.40	0.00	60.42	7.29	32.29		
20	ESR	≥90	≥90	0.46	0.35	0.00	46.88	13.54	39.58		
21	ESR	≥90	≥90	0.50	0.44	0.00	42.71	14.58	42.71		
22	TE	≥90	≥90	0.27	0.50	0.00	55.21	35.42	9.38		
23	TE	≥90	≥90	0.51	0.61	0.00	28.13	41.67	30.21		
24	ESR	≥90	≥90	0.30	0.48	0.00	64.58	10.42	25.00		
25	TE	≥90	≥90	0.31	0.26	0.00	45.83	45.83	8.33		
26	MS	≥90	≥90	0.38	0.34	0.00	33.33	57.29	9.38		
27	ESR	≥90	≥90	0.43	0.43	0.00	43.75	27.08	29.17		
28	ESR	≥90	≥90	0.08	0.11	0.00	89.58	4.17	6.25		
29	ESR	≥90	≥90	0.32	0.43	0.00	63.54	8.33	28.13		
30	ESR	≥90	≥90	0.41	0.47	0.00	53.13	12.50	34.38		
31	ESR	≥90	≥90	0.39	0.40	0.00	51.04	19.79	29.17		
32	ESR	≥90	≥90	0.27	0.19	0.00	60.42	25.00	14.58		
33	ESR	≥90	≥90	0.38	0.37	0.00	52.08	19.79	28.13		



**Table G.14 Operational Item Statistics—English II Summer Administration**

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D
1	ESR	≥80	≥80	0.34	0.43	0.00	58.82	15.29	25.88		
2	TE	≥80	≥80	0.38	0.39	1.18	44.71	32.94	21.18		
3	ESR	≥80	≥80	0.15	0.25	1.18	75.29	16.47	7.06		
4	MS	≥80	≥80	0.21	0.39	1.18	67.06	21.18	10.59		
5	MS	≥80	≥80	0.19	0.14	1.18	68.24	23.53	7.06		
6	TE	≥80	≥80	0.31	0.45	2.35	41.18	52.94	3.53		
7	CR	≥80	≥70	0.22	0.86	10.59	35.29	28.24	21.18	1.18	
8	CR	≥80	≥70	0.22	0.86	10.59	35.29	28.24	21.18	1.18	
9	CR	≥80	≥70	0.28	0.81	10.59	36.47	27.06	21.18	1.18	
10	ESR	≥80	≥80	0.37	0.34	2.35	55.29	12.94	29.41		
11	ESR	≥80	≥80	0.39	0.30	2.35	52.94	12.94	31.76		
12	MS	≥80	≥80	0.17	0.51	2.35	64.71	31.76	1.18		
13	TE	≥80	≥80	0.20	0.54	3.53	67.06	21.18	8.24		
14	ESR	≥80	≥80	0.65	0.27	0.00	27.06	15.29	57.65		
15	ESR	≥80	≥80	0.22	0.26	1.18	61.18	31.76	5.88		
16	ESR	≥80	≥80	0.54	0.43	1.18	21.18	48.24	29.41		
17	TE	≥80	≥80	0.37	0.45	2.35	29.41	64.71	3.53		
18	MS	≥80	≥80	0.34	0.42	2.35	49.41	30.59	17.65		
19	ESR	≥80	≥80	0.68	0.43	1.18	24.71	14.12	60.00		
20	MS	≥80	≥80	0.26	0.47	1.18	50.59	45.88	2.35		
21	ESR	≥80	≥80	0.43	0.52	1.18	48.24	15.29	35.29		
22	CR	≥80	≥70	0.22	0.82	9.41	37.65	24.71	18.82	4.71	
23	CR	≥80	≥70	0.22	0.82	9.41	37.65	24.71	18.82	4.71	
24	CR	≥80	≥70	0.27	0.79	9.41	37.65	29.41	16.47	2.35	
25	ESR	≥80	≥80	0.57	0.35	0.00	34.12	17.65	48.24		
26	ESR	≥80	≥80	0.42	0.24	1.18	43.53	28.24	27.06		
27	ESR	≥80	≥80	0.41	0.42	1.18	42.35	31.76	24.71		
28	ESR	≥80	≥80	0.51	0.34	1.18	36.47	24.71	37.65		
29	MS	≥80	≥80	0.30	0.42	1.18	61.18	15.29	22.35		
30	MS	≥80	≥80	0.31	0.61	1.18	57.65	21.18	20.00		
31	MS	≥80	≥80	0.30	0.47	1.18	44.71	48.24	5.88		
32	TE	≥80	≥80	0.29	0.32	3.53	52.94	31.76	11.76		
33	MS	≥80	≥80	0.30	0.37	1.18	54.12	30.59	14.12		
34	TE	≥80	≥80	0.24	0.33	3.53	55.29	35.29	5.88		

Table G.15 Operational Item Statistics—Algebra I Summer Administration

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D	% at 5	% at 6
1	MC	≥90	≥90	0.37	0.29	0.00							
2	MC	≥90	≥90	0.40	0.52	0.00							
3	SA	≥90	≥90	0.23	0.58	0.00	77.42	22.58					
4	MC	≥90	≥90	0.23	0.26	0.00							
5	MC	≥90	≥90	0.34	0.06	0.00							
6	TE	≥90	≥90	0.08	0.09	0.00	84.95	15.05					
7	MPSR	≥90	≥90	0.30	0.20	0.00	50.54	39.78	9.68				
8	MC	≥90	≥90	0.37	0.25	0.00							
9	MC	≥90	≥90	0.38	0.36	0.00							
10	SA	≥90	≥90	0.56	0.54	0.00	23.66	6.45	17.20	26.88	25.81		
11	MC	≥90	≥90	0.48	0.44	0.00							
12	MC	≥90	≥90	0.39	0.21	0.00							
13	MC	≥90	≥90	0.32	0.06	0.00							
14	MC	≥90	≥90	0.35	0.35	0.00							
15	SA	≥90	≥90	0.36	0.62	0.00	52.69	22.58	24.73				
16	MC	≥90	≥90	0.39	0.14	1.08							
17	MC	≥90	≥90	0.29	0.27	0.00							
18	MC	≥90	≥90	0.34	0.27	0.00							
19	MS	≥90	≥90	0.24	0.49	0.00	76.34	23.66					
20	MC	≥90	≥90	0.35	0.16	0.00							
21	MPSR	≥90	≥90	0.28	0.22	0.00	51.61	40.86	7.53				
22	MC	≥90	≥90	0.31	-0.13	0.00							
23	CR	≥90	≥70	0.04	0.38	11.83	70.97	5.38			2.15		
24	MC	≥90	≥90	0.63	0.40	0.00							
25	MC	≥90	≥90	0.26	0.37	0.00							
26	MC	≥90	≥90	0.46	0.23	0.00							
27	MPSR	≥90	≥90	0.30	0.37	0.00	54.84	30.11	15.05				
28	MC	≥90	≥90	0.26	0.47	0.00							
29	SA	≥90	≥90	0.13	0.45	2.15	84.95	12.90					
30	MPSR	≥90	≥90	0.40	-0.02	0.00	36.56	46.24	17.20				
31	TE	≥90	≥90	0.27	0.40	2.15	53.76	34.41	9.68				
32	MC	≥90	≥90	0.34	0.06	0.00							
33	MC	≥90	≥90	0.39	0.16	0.00							
34	CR	≥90	≥70	0.22	0.69	10.75	54.84	9.68	16.13	4.30			
35	CR	≥90	≥80	0.09	0.68	7.87	70.79	4.49	6.74	2.25	3.37	2.25	
36	CR	≥90	≥80	0.03	0.47	6.59	84.62	3.30	3.30				
37	CR	≥90	≥90	0.49	0.57	0.00	20.43	31.18	29.03	19.35			
38	CR	≥90	≥70	0.13	0.54	12.64	58.62	20.69	6.90				
39	CR	≥90	≥80	0.04	0.45	4.35	88.04	5.43	1.09	1.09			

Table G.16 Operational Item Statistics—Geometry Summer Administration

Item	Item Type	Total N	Adj. N	p-Value	Pbis	% Omit	% at 0	% at 1/A	% at 2/B	% at 3/C	% at 4/D	% at 5	% at 6
1	MPSR	≥70	≥70	0.40	0.63	0.00	40.26	38.96	20.78				
2	TE	≥70	≥70	0.25	0.88	2.60	72.73	24.68					
3	MC	≥70	≥70	0.59	0.47	1.30							
4	MC	≥70	≥70	0.36	0.69	1.30							
5	TE	≥70	≥70	0.19	0.82	3.90	77.92	18.18					
6	TE	≥70	≥70	0.25	0.77	2.60	72.73	24.68					
7	MPSR	≥70	≥70	0.32	0.69	2.60	51.95	28.57	16.88				
8	MC	≥70	≥70	0.78	0.34	1.30							
9	SA	≥70	≥70	0.21	0.93	2.60	76.62	20.78					
10	TE	≥70	≥70	0.43	0.85	2.60	14.29	40.26	16.88	10.39	15.58		
11	SA	≥70	≥70	0.35	0.72	2.60	63.64	33.77					
12	CR	≥70	≥60	0.32	0.76	14.29	42.86	11.69	15.58	11.69			
13	CR	≥70	≥60	0.14	0.79	14.29	63.64	9.09	9.09	2.60			
14	SA	≥70	≥70	0.30	0.82	0.00	70.13	29.87					
15	MC	≥70	≥70	0.28	0.50	1.30							
16	SA	≥70	≥70	0.33	0.79	1.30	66.23	32.47					
17	MS	≥70	≥70	0.39	0.72	1.30	59.74	38.96					
18	MC	≥70	≥70	0.22	0.57	3.90							
19	SA	≥70	≥70	0.35	0.70	1.30	49.35	29.87	19.48				
20	MC	≥70	≥70	0.38	0.67	1.30							
21	MPSR	≥70	≥70	0.47	0.65	1.30	25.97	51.95	20.78				
22	MC	≥70	≥70	0.35	0.68	2.60							
23	TE	≥70	≥70	0.66	0.54	1.30	18.18	29.87	50.65				
24	CR	≥70	≥60	0.20	0.83	15.58	55.84	6.49	3.90	1.30	11.69		
25	TE	≥70	≥70	0.55	0.51	0.00	45.45	54.55					
26	SA	≥70	≥70	0.26	0.61	1.30	72.73	25.97					
27	MC	≥70	≥70	0.48	0.20	2.60							
28	MC	≥70	≥70	0.25	0.51	1.30							
29	MS	≥70	≥70	0.20	0.75	1.30	79.22	19.48					
30	MPSR	≥70	≥70	0.38	0.64	1.30	55.84	11.69	31.17				
31	MC	≥70	≥70	0.43	0.48	1.30							
32	SA	≥70	≥70	0.25	0.86	2.60	72.73	24.68					
33	TE	≥70	≥70	0.45	0.62	3.90	53.25	42.86					
34	SA	≥70	≥70	0.29	0.68	3.90	57.14	22.08	16.88				
35	MC	≥70	≥70	0.50	0.51	1.30							
36	CR	≥70	≥60	0.16	0.83	12.99	62.34	7.79	3.90	2.60	7.79		
37	CR	≥70	≥60	0.22	0.90	11.84	50.00	13.16	5.26	3.95	1.32	9.21	5.26
38	CR	≥70	≥70	0.18	0.83	1.30	55.84	20.78	5.19	3.90	2.60	7.79	2.60

## Appendix H—Student Participation Counts

**Table H.1 Count of Students taking the Fall 2018 Administration: English I**

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	≥20	≥5,220	≥1,210	≥180	≥40	≥6,680
<b>Gender</b>								
Female	<10	<10	≥10	≥2,370	≥420	≥50	≥10	≥2,870
Male	<10	<10	≥10	≥2,850	≥790	≥120	≥30	≥3,810
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	≥400	≥160	≥60	<10	≥640
American Indian or Alaska Native	<10	<10	<10	≥20	<10	<10	<10	≥20
Asian	<10	<10	<10	≥80	≥10	<10	<10	≥100
Black or African American	<10	<10	<10	≥2,210	≥700	≥70	≥20	≥3,030
Native Hawaiian or Other Pacific	<10	<10	<10	<10	<10	<10	<10	<10
White	<10	<10	≥10	≥2,380	≥300	≥30	≥10	≥2,740
Two or More Races	<10	<10	<10	≥100	≥20	<10	<10	≥130
<b>Education Classification</b>								
Regular	<10	<10	≥10	≥4,280	≥870	≥150	≥30	≥5,360
Special	<10	<10	<10	≥590	≥320	≥30	<10	≥950
Gifted	<10	<10	<10	≥340	≥10	<10	<10	≥360
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	<10	<10	≥20	≥5,060	≥1,070	≥120	≥30	≥6,320
EL	<10	<10	<10	≥160	≥130	≥60	<10	≥360
<b>Migrant Status</b>								
Non-migrant	<10	<10	≥20	≥5,220	≥1,210	≥180	≥40	≥6,680
Migrant	<10	<10	<10	<10	<10	<10	<10	<10
<b>Section 504 Status</b>								
Non-section 504	<10	<10	≥20	≥4,720	≥1,030	≥150	≥30	≥5,970
Section 504	<10	<10	<10	≥500	≥170	≥20	<10	≥710
<b>Homeless Status</b>								
Not Homeless	<10	<10	≥20	≥5,130	≥1,180	≥170	≥40	≥6,540
Homeless	<10	<10	<10	≥90	≥30	<10	<10	≥140
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	≥20	≥5,050	≥1,200	≥180	≥40	≥6,490
Military Affiliated	<10	<10	<10	≥170	≥10	<10	<10	≥190
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	≥20	≥5,180	≥1,200	≥180	≥40	≥6,640
Foster Care	<10	<10	<10	≥30	<10	<10	<10	≥40

\* Economic status information is not available for the fall and summer administrations.

Table H.2 Percentage of Students taking the Fall 2018 Administration: English I

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.00	0.00	0.34	78.14	18.15	2.72	0.64	100
<b>Gender</b>								
Female	0.00	0.00	0.45	82.51	14.64	2.02	0.38	100
Male	0.00	0.00	0.26	74.85	20.80	3.25	0.84	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.16	63.41	25.58	10.08	0.78	100
American Indian or Alaska Native	0.00	0.00	0.00	85.19	14.81	0.00	0.00	100
Asian	0.00	0.00	0.98	83.33	10.78	4.90	0.00	100
Black or African American	0.00	0.00	0.13	73.04	23.30	2.61	0.92	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	75.00	25.00	0.00	0.00	100
White	0.00	0.00	0.58	86.95	10.97	1.13	0.36	100
Two or More Races	0.00	0.00	0.73	78.83	18.98	1.46	0.00	100
<b>Education Classification</b>								
Regular	0.00	0.00	0.34	79.84	16.28	2.81	0.73	100
Special	0.00	0.00	0.00	62.25	34.20	3.13	0.42	100
Gifted	0.00	0.00	1.38	95.04	3.31	0.28	0.00	100
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	0.00	0.00	0.36	80.04	17.06	1.91	0.62	100
EL	0.00	0.00	0.00	45.21	36.99	16.71	1.10	100
<b>Migrant Status</b>								
Non-migrant	0.00	0.00	0.34	78.15	18.14	2.72	0.64	100
Migrant	0.00	0.00	0.00	75.00	25.00	0.00	0.00	100
<b>Section 504 Status</b>								
Non-section 504	0.00	0.00	0.39	79.06	17.33	2.65	0.59	100
Section 504	0.00	0.00	0.00	70.53	25.00	3.35	1.12	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	0.35	78.33	18.02	2.67	0.63	100
Homeless	0.00	0.00	0.00	69.29	24.29	5.00	1.43	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	0.35	77.75	18.48	2.79	0.63	100
Military Affiliated	0.00	0.00	0.00	91.58	6.84	0.53	1.05	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	0.35	78.12	18.16	2.73	0.65	100
Foster Care	0.00	0.00	0.00	81.25	16.67	2.08	0.00	100

\* Economic status information is not available for the fall and summer administrations.

Table H.3 Count of Students taking the Fall 2018 Administration: English II

Group	Grade							Total
	6	7	8	9	10	11	12	
<b>All Students</b>	<10	<10	<10	≥500	≥6,720	≥1,840	≥530	≥9,590
<b>Gender</b>								
Female	<10	<10	<10	≥210	≥2,970	≥650	≥190	≥4,040
Male	<10	<10	<10	≥280	≥3,740	≥1,180	≥330	≥5,550
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	≥500	≥540	≥260	≥80	≥940
American Indian or Alaska Native	<10	<10	<10	<10	≥40	<10	<10	≥50
Asian	<10	<10	<10	≥20	≥100	≥30	≥10	≥170
Black or African American	<10	<10	<10	≥240	≥3,150	≥1,110	≥340	≥4,860
Native Hawaiian or Other Pacific	<10	<10	<10	<10	≥10	<10	<10	≥10
White	<10	<10	<10	≥160	≥2,740	≥400	≥80	≥3,390
Two or More Races	<10	<10	<10	≥10	≥110	≥10	<10	≥150
<b>Education Classification</b>								
Regular	<10	<10	<10	≥340	≥5,730	≥1,390	≥410	≥7,880
Special	<10	<10	<10	≥70	≥710	≥430	≥110	≥1,340
Gifted	<10	<10	<10	≥80	≥270	≥10	<10	≥360
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	<10	<10	<10	≥480	≥6,440	≥1,560	≥440	≥8,930
EL	<10	<10	<10	≥20	≥280	≥270	≥80	≥660
<b>Migrant Status</b>								
Non-migrant	<10	<10	<10	≥500	≥6,710	≥1,840	≥520	≥9,580
Migrant	<10	<10	<10	<10	≥10	<10	<10	≥10
<b>Section 504 Status</b>								
Non-section 504	<10	<10	<10	≥440	≥6,050	≥1,560	≥460	≥8,520
Section 504	<10	<10	<10	≥50	≥670	≥280	≥60	≥1,070
<b>Homeless Status</b>								
Not Homeless	<10	<10	<10	≥480	≥6,570	≥1,790	≥510	≥9,360
Homeless	<10	<10	<10	≥20	≥140	≥50	≥10	≥230
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	<10	≥490	≥6,600	≥1,840	≥520	≥9,470
Military Affiliated	<10	<10	<10	<10	≥110	<10	<10	≥120
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	<10	≥490	≥6,670	≥1,840	≥520	≥9,540
Foster Care	<10	<10	<10	<10	≥40	<10	<10	≥50

\* Economic status information is not available for the fall and summer administrations.

Table H.4 Percentage of Students taking the Fall 2018 Administration: English II

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.00	0.00	0.00	5.25	70.03	19.20	5.52	100
<b>Gender</b>								
Female	0.00	0.00	0.00	5.35	73.59	16.24	4.83	100
Male	0.00	0.00	0.00	5.18	67.44	21.35	6.03	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.00	5.60	57.34	28.19	8.87	100
American Indian or Alaska Native	0.00	0.00	0.00	3.64	74.55	16.36	5.45	100
Asian	0.00	0.00	0.00	14.12	62.35	17.65	5.88	100
Black or African American	0.00	0.00	0.00	5.06	64.90	22.91	7.13	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	0.00	100.00	0.00	0.00	100
White	0.00	0.00	0.00	4.92	80.83	11.90	2.36	100
Two or More Races	0.00	0.00	0.00	7.74	76.77	11.61	3.87	100
<b>Education Classification</b>								
Regular	0.00	0.00	0.00	4.43	72.73	17.64	5.20	100
Special	0.00	0.00	0.00	5.50	53.16	32.57	8.77	100
Gifted	0.00	0.00	0.00	21.95	73.71	3.79	0.54	100
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	0.00	0.00	0.00	5.37	72.11	17.56	4.96	100
EL	0.00	0.00	0.00	3.59	42.22	41.17	13.02	100
<b>Migrant Status</b>								
Non-migrant	0.00	0.00	0.00	5.25	70.04	19.21	5.50	100
Migrant	0.00	0.00	0.00	6.25	62.50	12.50	18.75	100
<b>Section 504 Status</b>								
Non-section 504	0.00	0.00	0.00	5.22	71.02	18.34	5.41	100
Section 504	0.00	0.00	0.00	5.47	62.15	25.97	6.40	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	0.00	5.16	70.21	19.11	5.52	100
Homeless	0.00	0.00	0.00	8.97	62.82	22.65	5.56	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	0.00	5.24	69.79	19.43	5.54	100
Military Affiliated	0.00	0.00	0.00	6.20	87.60	2.33	3.88	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	0.00	5.22	69.95	19.30	5.53	100
Foster Care	0.00	0.00	0.00	11.11	83.33	1.85	3.70	100

\* Economic status information is not available for the fall and summer administrations.

Table H.5 Count of Students taking the Fall 2018 Administration: Algebra I

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	≥10	≥3,130	≥1,690	≥590	≥240	≥5,670
<b>Gender</b>								
Female	<10	<10	<10	≥1,530	≥710	≥240	≥90	≥2,590
Male	<10	<10	≥10	≥1,600	≥970	≥340	≥140	≥3,080
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	≥290	≥220	≥70	≥20	≥610
American Indian or Alaska Native	<10	<10	<10	≥10	<10	<10	<10	≥20
Asian	<10	<10	<10	≥50	<10	<10	<10	≥80
Black or African American	<10	<10	<10	≥1,360	≥1,010	≥400	≥180	≥2,960
Native Hawaiian or Other Pacific	<10	<10	<10	<10	<10	<10	<10	<10
White	<10	<10	≥10	≥1,330	≥420	≥90	≥30	≥1,890
Two or More Races	<10	<10	<10	≥70	≥20	<10	<10	≥100
<b>Education Classification</b>								
Regular	<10	<10	≥10	≥2,690	≥1,350	≥420	≥190	≥4,690
Special	<10	<10	<10	≥290	≥300	≥150	≥40	≥810
Gifted	<10	<10	<10	≥140	≥20	<10	<10	≥170
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	<10	<10	≥10	≥3,000	≥1,490	≥520	≥210	≥5,250
EL	<10	<10	<1	≥120	≥190	≥70	≥20	≥420
<b>Migrant Status</b>								
Non-migrant	<10	<10	≥10	≥3,130	≥1,680	≥590	≥240	≥5,670
Migrant	<10	<10	<10	<10	<10	<10	<10	<10
<b>Section 504 Status</b>								
Non-section 504	<10	<10	≥10	≥2,820	≥1,450	≥490	≥210	≥5,000
Section 504	<10	<10	<10	≥310	≥230	≥90	≥20	≥670
<b>Homeless Status</b>								
Not Homeless	<10	<10	≥10	≥3,060	≥1,640	≥560	≥230	≥5,530
Homeless	<10	<10	<10	≥60	≥40	≥20	<10	≥140
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	≥10	≥3,110	≥1,680	≥590	≥230	≥5,650
Military Affiliated	<10	<10	<10	≥10	<10	<10	<10	≥20
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	≥10	≥3,130	≥1,680	≥590	≥240	≥5,660
Foster Care	<10	<10	<10	<10	<10	<10	<10	≥10

\* Economic status information is not available for the fall and summer administrations.



Table H.6 Percentage of Students taking the Fall 2018 Administration: Algebra I

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.00	0.00	0.32	55.20	29.79	10.42	4.26	100
<b>Gender</b>								
Female	0.00	0.00	0.31	58.87	27.47	9.58	3.77	100
Male	0.00	0.00	0.32	52.11	31.75	11.14	4.68	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.16	47.89	35.88	12.66	3.41	100
American Indian or Alaska Native	0.00	0.00	0.00	65.00	30.00	0.00	5.00	100
Asian	0.00	0.00	1.23	70.37	11.11	8.64	8.64	100
Black or African American	0.00	0.00	0.20	46.01	34.12	13.58	6.08	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	25.00	75.00	0.00	0.00	100
White	0.00	0.00	0.53	70.41	22.31	5.12	1.64	100
Two or More Races	0.00	0.00	0.00	70.59	19.61	7.84	1.96	100
<b>Education Classification</b>								
Regular	0.00	0.00	0.36	57.47	28.93	9.12	4.11	100
Special	0.00	0.00	0.00	36.67	38.15	19.63	5.56	100
Gifted	0.00	0.00	0.56	79.78	14.61	2.81	2.25	100
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	0.00	0.00	0.34	57.23	28.39	9.91	4.13	100
EL	0.00	0.00	0.00	29.93	47.27	16.86	5.94	100
<b>Migrant Status</b>								
Non-migrant	0.00	0.00	0.32	55.22	29.78	10.44	4.25	100
Migrant	0.00	0.00	0.00	42.86	42.86	0.00	14.29	100
<b>Section 504 Status</b>								
Non-section 504	0.00	0.00	0.36	56.34	29.13	9.90	4.27	100
Section 504	0.00	0.00	0.00	46.72	34.78	14.33	4.18	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	0.33	55.50	29.71	10.22	4.25	100
Homeless	0.00	0.00	0.00	44.30	32.89	18.12	4.70	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	0.32	55.19	29.80	10.46	4.23	100
Military Affiliated	0.00	0.00	0.00	57.14	28.57	3.57	10.71	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	0.32	55.21	29.79	10.43	4.25	100
Foster Care	0.00	0.00	0.00	50.00	30.00	10.00	10.00	100

\* Economic status information is not available for the fall and summer administrations.

Table H.7 Count of Students taking the Fall 2018 Administration: Geometry

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	<10	≥1,200	≥2,960	≥890	≥290	≥5,350
<b>Gender</b>								
Female	<10	<10	<10	≥640	≥1,510	≥420	≥100	≥2,680
Male	<10	<10	<10	≥550	≥1,450	≥470	≥190	≥2,670
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	≥100	≥250	≥140	≥30	≥530
American Indian or Alaska Native	<10	<10	<10	<10	≥10	<10	<10	≥20
Asian	<10	<10	<10	≥60	≥50	≥10	<10	≥140
Black or African American	<10	<10	<10	≥330	≥1,420	≥490	≥200	≥2,450
Native Hawaiian or Other Pacific	<10	<10	<10	<10	<10	<10	<10	<10
White	<10	<10	<10	≥670	≥1,150	≥220	≥40	≥2,090
Two or More Races	<10	<10	<10	≥20	≥50	≥10	<10	≥90
<b>Education Classification</b>								
Regular	<10	<10	<10	≥900	≥2,610	≥770	≥230	≥4,530
Special	<10	<10	<10	≥20	≥170	≥90	≥50	≥340
Gifted	<10	<10	<10	≥270	≥170	≥20	<10	≥480
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	<10	<10	<10	≥1,190	≥2,870	≥770	≥260	≥5,110
EL	<10	<10	<10	<10	≥90	≥120	≥20	≥240
<b>Migrant Status</b>								
Non-migrant	<10	<10	<10	≥1,200	≥2,960	≥890	≥290	≥5,350
Migrant	<10	<10	<10	<10	<10	<10	<10	<10
<b>Section 504 Status</b>								
Non-section 504	<10	<10	<10	≥1,150	≥2,750	≥800	≥260	≥4,970
Section 504	<10	<10	<10	≥40	≥210	≥90	20	≥380
<b>Homeless Status</b>								
Not Homeless	<10	<10	<10	≥1,190	≥2,900	≥860	≥280	≥5,250
Homeless	<10	<10	<10	<10	≥60	≥20	<10	≥100
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	<10	≥1,180	≥2,900	≥890	≥290	≥5,260
Military Affiliated	<10	<10	<10	≥20	≥60	<10	<10	≥90
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	<10	≥1,190	≥2,940	≥890	≥290	≥5,320
Foster Care	<10	<10	<10	≥10	≥20	<10	<10	≥30

\* Economic status information is not available for the fall and summer administrations.

Table H.8 Percentage of Students taking the Fall 2018 Administration: Geometry

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.00	0.00	0.00	22.47	55.36	16.68	5.49	100
<b>Gender</b>								
Female	0.00	0.00	0.00	24.10	56.38	15.66	3.87	100
Male	0.00	0.00	0.00	20.82	54.34	17.72	7.12	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.00	18.62	48.23	26.44	6.70	100
American Indian or Alaska Native	0.00	0.00	0.00	11.11	70.37	14.81	3.70	100
Asian	0.00	0.00	0.00	44.14	40.00	13.10	2.76	100
Black or African American	0.00	0.00	0.00	13.72	57.86	20.11	8.31	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	0.00	100.00	0.00	0.00	100
White	0.00	0.00	0.00	32.06	55.10	10.73	2.10	100
Two or More Races	0.00	0.00	0.00	29.47	54.74	10.53	5.26	100
<b>Education Classification</b>								
Regular	0.00	0.00	0.00	19.85	57.74	17.18	5.23	100
Special	0.00	0.00	0.00	7.35	50.88	26.47	15.29	100
Gifted	0.00	0.00	0.00	57.53	36.29	5.15	1.03	100
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	0.00	0.00	0.00	23.44	56.28	15.10	5.18	100
EL	0.00	0.00	0.00	2.43	36.44	49.39	11.74	100
<b>Migrant Status</b>								
Non-migrant	0.00	0.00	0.00	22.49	55.38	16.68	5.45	100
Migrant	0.00	0.00	0.00	0.00	40.00	20.00	40.00	100
<b>Section 504 Status</b>								
Non-section 504	0.00	0.00	0.00	23.23	55.25	16.13	5.38	100
Section 504	0.00	0.00	0.00	12.57	56.81	23.82	6.81	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	0.00	22.79	55.23	16.50	5.48	100
Homeless	0.00	0.00	0.00	5.83	62.14	26.21	5.83	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	0.00	22.41	55.11	16.93	5.54	100
Military Affiliated	0.00	0.00	0.00	25.56	70.00	2.22	2.22	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	0.00	22.40	55.33	16.77	5.50	100
Foster Care	0.00	0.00	0.00	33.33	60.61	3.03	3.03	100

\* Economic status information is not available for the fall and summer administrations.

Table H.9 Count of Students taking the Summer 2019 Administration: English I

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	≥10	≥1,450	≥390	≥20	<10	≥1,900
<b>Gender</b>								
Female	<10	<10	≥10	≥460	≥120	<10	<10	≥600
Male	<10	<10	<10	≥990	≥270	≥10	<10	≥1,290
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	≥120	≥60	<10	<10	≥180
American Indian or Alaska Native	<10	<10	<10	<10	<10	<10	<10	<10
Asian	<10	<10	<10	≥10	<10	<10	<10	≥10
Black or African American	<10	<10	≥10	≥990	≥250	≥20	<10	≥1,280
Native Hawaiian or Other Pacific	<10	<10	<10	<10	<10	<10	<10	<10
White	<10	<10	<10	≥300	≥70	<10	<10	≥380
Two or More Races	<10	<10	<10	≥10	<10	<10	<10	≥10
<b>Education Classification</b>								
Regular	<10	<10	≥10	≥1,040	≥260	≥10	<10	≥1,340
Special	<10	<10	<10	≥400	≥130	<10	<10	≥540
Gifted	<10	<10	<10	<10	<10	<10	<10	≥10
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	<10	<10	≥10	≥1,330	≥330	≥20	<10	≥1,710
EL	<10	<10	<10	≥120	≥50	<10	<10	≥180
<b>Migrant Status</b>								
Non-migrant	<10	<10	≥10	≥1,450	≥390	≥20	<10	≥1,890
Migrant	<10	<10	<10	<10	<10	<10	<10	<10
<b>Section 504 Status</b>								
Non-section 504	<10	<10	≥10	≥1,180	≥310	≥10	<10	≥1,540
Section 504	<10	<10	<10	≥270	≥70	<10	<10	≥350
<b>Homeless Status</b>								
Not Homeless	<10	<10	≥10	≥1,410	≥380	≥20	<10	≥1,850
Homeless	<10	<10	<10	≥40	<10	<10	<10	≥40
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	≥10	≥1,450	≥390	≥20	<10	≥1,890
Military Affiliated	<10	<10	<10	<10	<10	<10	<10	<10
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	≥10	≥1,450	≥390	≥20	<10	≥1,890
Foster Care	<10	<10	<10	<10	<10	<10	<10	<10

\* Economic status information is not available for the fall and summer administrations.

Table H.10 Percentage of Students taking the Summer 2019 Administration: English I

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.00	0.00	1.00	76.70	20.73	1.26	0.32	100
<b>Gender</b>								
Female	0.00	0.00	1.98	76.73	20.30	0.83	0.17	100
Male	0.00	0.00	0.54	76.68	20.93	1.47	0.39	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.00	66.14	32.28	1.06	0.53	100
American Indian or Alaska Native	0.00	0.00	0.00	33.33	66.67	0.00	0.00	100
Asian	0.00	0.00	0.00	84.62	15.38	0.00	0.00	100
Black or African American	0.00	0.00	1.01	77.59	19.53	1.56	0.31	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100
White	0.00	0.00	1.55	79.12	18.56	0.52	0.26	100
Two or More Races	0.00	0.00	0.00	78.95	21.05	0.00	0.00	100
<b>Education Classification</b>								
Regular	0.00	0.00	1.34	77.47	19.33	1.41	0.45	100
Special	0.00	0.00	0.18	74.91	23.99	0.92	0.00	100
Gifted	0.00	0.00	0.00	70.00	30.00	0.00	0.00	100
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	0.00	0.00	1.11	77.82	19.62	1.22	0.23	100
EL	0.00	0.00	0.00	66.12	31.15	1.64	1.09	100
<b>Migrant Status</b>								
Non-migrant	0.00	0.00	1.00	76.74	20.68	1.27	0.32	100
Migrant	0.00	0.00	0.00	60.00	40.00	0.00	0.00	100
<b>Section 504 Status</b>								
Non-section 504	0.00	0.00	1.17	76.89	20.65	1.10	0.19	100
Section 504	0.00	0.00	0.28	75.84	21.07	1.97	0.84	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	1.03	76.46	20.95	1.24	0.32	100
Homeless	0.00	0.00	0.00	85.71	12.24	2.04	0.00	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	1.00	76.83	20.63	1.21	0.32	100
Military Affiliated	0.00	0.00	0.00	33.33	50.00	16.67	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	1.00	76.69	20.73	1.27	0.32	100
Foster Care	0.00	0.00	0.00	80.00	20.00	0.00	0.00	100

\* Economic status information is not available for the fall and summer administrations.

Table H.11 Count of Students taking the Summer 2019 Administration: English II

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	<10	≥70	≥800	≥550	≥270	≥1,690
<b>Gender</b>								
Female	<10	<10	<10	≥20	≥260	≥150	≥80	≥520
Male	<10	<10	<10	≥50	≥540	≥390	≥180	≥1,170
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	<10	≥70	≥60	≥60	≥200
American Indian or Alaska Native	<10	<10	<10	<10	<10	<10	<10	<10
Asian	<10	<10	<10	<10	<10	<10	<10	≥10
Black or African American	<10	<10	<10	≥50	≥570	≥380	≥170	≥1,180
Native Hawaiian or Other Pacific	<10	<10	<10	<10	<10	<10	<10	<10
White	<10	<10	<10	≥10	≥140	≥90	≥20	≥260
Two or More Races	<10	<10	<10	<10	<10	<10	<10	≥10
<b>Education Classification</b>								
Regular	<10	<10	<10	≥40	≥600	≥400	≥260	≥1,310
Special	<10	<10	<10	≥20	≥190	≥140	≥10	≥360
Gifted	<10	<10	<10	<10	<10	<10	<10	≥10
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	<10	<10	<10	≥60	≥730	≥480	≥190	≥1,470
EL	<10	<10	<10	<10	≥60	≥70	≥80	≥210
<b>Migrant Status</b>								
Non-migrant	<10	<10	<10	≥70	≥800	≥550	≥270	≥1,690
Migrant	<10	<10	<10	<10	<10	<10	<10	<10
<b>Section 504 Status</b>								
Non-section 504	<10	<10	<10	≥50	≥650	≥430	≥200	≥1,350
Section 504	<10	<10	<10	≥10	≥140	≥110	≥60	≥330
<b>Homeless Status</b>								
Not Homeless	<10	<10	<10	≥60	≥780	≥530	≥260	≥1,650
Homeless	<10	<10	<10	<10	≥20	≥10	<10	≥40
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	<10	≥60	≥790	≥550	≥270	≥1,680
Military Affiliated	<10	<10	<10	<10	<10	<10	<10	<10
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	<10	≥60	≥790	≥550	≥270	≥1,680
Foster Care	<10	<10	<10	<10	<10	<10	<10	≥10

\* Economic status information is not available for the fall and summer administrations.

Table H.12 Percentage of Students taking the Summer 2019 Administration: English II

Group	Grade							Total
	6	7	8	9	10	11	12	
<b>All Students</b>	0.00	0.00	0.00	4.13	47.26	32.57	16.05	100
<b>Gender</b>								
Female	0.00	0.00	0.00	3.81	49.71	30.10	16.38	100
Male	0.00	0.00	0.00	4.27	46.15	33.68	15.90	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.00	1.93	35.75	29.47	32.85	100
American Indian or Alaska Native	0.00	0.00	0.00	0.00	50.00	25.00	25.00	100
Asian	0.00	0.00	0.00	0.00	27.78	33.33	38.89	100
Black or African American	0.00	0.00	0.00	4.38	48.27	32.69	14.66	100
Native Hawaiian or Other Pacific								
White	0.00	0.00	0.00	5.20	52.42	34.20	8.18	100
Two or More Races	0.00	0.00	0.00	0.00	60.00	40.00	0.00	100
<b>Education Classification</b>								
Regular	0.00	0.00	0.00	3.42	45.86	30.95	19.77	100
Special	0.00	0.00	0.00	6.50	52.03	38.75	2.71	100
Gifted	0.00	0.00	0.00	9.09	54.55	18.18	18.18	100
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	0.00	0.00	0.00	4.54	49.80	32.66	13.01	100
EL	0.00	0.00	0.00	1.37	30.14	31.96	36.53	100
<b>Migrant Status</b>								
Non-migrant	0.00	0.00	0.00	4.13	47.28	32.59	16.00	100
Migrant	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100
<b>Section 504 Status</b>								
Non-section 504	0.00	0.00	0.00	4.28	48.38	32.08	15.27	100
Section 504	0.00	0.00	0.00	3.54	42.77	34.51	19.17	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	0.00	4.12	47.27	32.61	16.00	100
Homeless	0.00	0.00	0.00	4.44	46.67	31.11	17.78	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	0.00	4.09	47.07	32.72	16.12	100
Military Affiliated	0.00	0.00	0.00	12.50	87.50	0.00	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	0.00	4.09	47.00	32.76	16.14	100
Foster Care	0.00	0.00	0.00	10.00	90.00	0.00	0.00	100

\* Economic status information is not available for the fall and summer administrations.

Table H.13 Count of Students taking the Summer 2019 Administration: Algebra I

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	<10	<10	≥20	≥1,320	≥490	≥80	≥30	≥1,950
<b>Gender</b>								
Female	<10	<10	≥10	≥570	≥210	≥30	≥10	≥850
Male	<10	<10	≥10	≥740	≥270	≥40	≥10	≥1,100
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	≥90	≥30	≥10	<10	≥140
American Indian or Alaska Native	<10	<10	<10	<10	<10	<10	<10	≥10
Asian	<10	<10	<10	<10	<10	<10	<10	<10
Black or African American	<10	<10	≥20	≥890	≥340	≥60	≥20	≥1,340
Native Hawaiian or Other Pacific	<10	<10	<10	<10	<10	<10	<10	<10
White	<10	<10	<10	≥290	≥100	<10	<10	≥410
Two or More Races	<10	<10	<10	≥20	<10	<10	<10	≥30
<b>Education Classification</b>								
Regular	<10	<10	≥20	≥1,050	≥350	≥60	≥20	≥1,520
Special	<10	<10	<10	≥260	≥130	≥10	<10	≥410
Gifted	<10	<10	<10	≥10	<10	<10	<10	≥10
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	<10	<10	≥20	≥1,260	≥460	≥60	≥20	≥1,840
EL	<10	<10	<10	≥60	≥30	≥10	<10	≥110
<b>Migrant Status</b>								
Non-migrant	<10	<10	≥20	≥1,320	≥490	≥80	≥30	≥1,950
Migrant	<10	<10	<10	<10	<10	<10	<10	<10
<b>Section 504 Status</b>								
Non-section 504	<10	<10	≥20	≥1,080	≥410	≥70	≥20	≥1,610
Section 504	<10	<10	<10	≥240	≥70	<10	≥10	≥330
<b>Homeless Status</b>								
Not Homeless	<10	<10	≥20	≥1,290	≥490	≥70	≥30	≥1,910
Homeless	<10	<10	<10	≥30	<10	<10	<10	≥40
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	≥20	≥1,310	≥490	≥70	≥30	≥1,930
Military Affiliated	<10	<10	<10	<10	<10	<10	<10	≥10
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	≥20	≥1,310	≥490	≥80	≥30	≥1,940
Foster Care	<10	<10	<10	<10	<10	<10	<10	<10

\* Economic status information is not available for the fall and summer administrations.



Table H.14 Percentage of Students taking the Summer 2019 Administration: Algebra I

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.00	0.00	1.33	67.67	25.32	4.14	1.53	100
<b>Gender</b>								
Female	0.00	0.00	1.40	67.72	25.50	4.09	1.29	100
Male	0.00	0.00	1.27	67.64	25.18	4.18	1.73	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.00	64.08	23.94	9.15	2.82	100
American Indian or Alaska Native	0.00	0.00	0.00	70.00	30.00	0.00	0.00	100
Asian	0.00	0.00	0.00	71.43	14.29	14.29	0.00	100
Black or African American	0.00	0.00	1.64	66.47	25.65	4.46	1.78	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100
White	0.00	0.00	0.96	71.33	25.54	1.69	0.48	100
Two or More Races	0.00	0.00	0.00	82.86	17.14	0.00	0.00	100
<b>Education Classification</b>								
Regular	0.00	0.00	1.64	69.01	23.57	4.01	1.77	100
Special	0.00	0.00	0.00	62.95	31.96	4.36	0.73	100
Gifted	0.00	0.00	5.26	63.16	21.05	10.53	0.00	100
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	0.00	0.00	1.41	68.40	25.14	3.69	1.36	100
EL	0.00	0.00	0.00	55.75	28.32	11.50	4.42	100
<b>Migrant Status</b>								
Non-migrant	0.00	0.00	1.33	67.69	25.29	4.15	1.54	100
Migrant	0.00	0.00	0.00	50.00	50.00	0.00	0.00	100
<b>Section 504 Status</b>								
Non-section 504	0.00	0.00	1.61	66.83	25.87	4.46	1.24	100
Section 504	0.00	0.00	0.00	71.68	22.71	2.65	2.95	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.00	1.36	67.47	25.69	3.92	1.57	100
Homeless	0.00	0.00	0.00	77.50	7.50	15.00	0.00	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.00	1.34	67.77	25.27	4.07	1.55	100
Military Affiliated	0.00	0.00	0.00	56.25	31.25	12.50	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.00	1.34	67.69	25.27	4.16	1.54	100
Foster Care	0.00	0.00	0.00	62.50	37.50	0.00	0.00	100

\* Economic status information is not available for the fall and summer administrations.

Table H.15 Count of Students taking the Summer 2019 Administration: Geometry

Group	Grade							Total
	6	7	8	9	10	11	12	
<b>All Students</b>	<10	<10	≥10	≥20	≥160	≥50	≥20	≥270
<b>Gender</b>								
Female	<10	<10	<10	≥10	≥70	≥20	<10	≥120
Male	<10	<10	≥10	≥10	≥80	≥20	≥10	≥140
<b>Ethnicity</b>								
Hispanic/Latino	<10	<10	<10	<10	<10	<10	<10	≥10
American Indian or Alaska Native	<10	<10	<10	<10	<10	<10	<10	<10
Asian	<10	<10	<10	<10	<10	<10	<10	≥10
Black or African American	<10	<10	<10	≥10	≥120	≥40	≥20	≥210
Native Hawaiian or Other Pacific	<10	<10	<10	<10	<10	<10	<10	<10
White	<10	<10	<10	<10	≥20	<10	<10	≥30
Two or More Races	<10	<10	<10	<10	<10	<10	<10	<10
<b>Education Classification</b>								
Regular	<10	<10	<10	≥10	≥130	≥40	≥20	≥210
Special	<10	<10	<10	<10	≥30	<10	<10	≥40
Gifted	<10	<10	≥10	<10	<10	<10	<10	≥10
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	<10	<10	≥10	≥20	≥160	≥40	≥20	≥260
EL	<10	<10	<10	<10	<10	<10	<10	<10
<b>Migrant Status</b>								
Non-migrant	<10	<10	≥10	≥20	≥160	≥50	≥20	≥270
Migrant	<10	<10	<10	<10	<10	<10	<10	<10
<b>Section 504 Status</b>								
Non-section 504	<10	<10	≥10	≥10	≥140	≥40	≥10	≥230
Section 504	<10	<10	<10	<10	≥10	≥10	<10	≥30
<b>Homeless Status</b>								
Not Homeless	<10	<10	≥10	≥20	≥160	≥40	≥20	≥270
Homeless	<10	<10	<10	<10	<10	<10	<10	<10
<b>Military Affiliation</b>								
Not Military Affiliated	<10	<10	≥10	≥20	≥160	50	≥20	≥270
Military Affiliated	<10	<10	<10	<10	<10	<10	<10	<10
<b>Foster Care Status</b>								
Not in Foster Care	<10	<10	≥10	≥20	≥150	≥50	≥20	≥260
Foster Care	<10	<10	<10	<10	<10	<10	<10	<10

\* Economic status information is not available for the fall and summer administrations.

Table H.16 Percentage of Students taking the Summer 2019 Administration: Geometry

Group	Grade							
	6	7	8	9	10	11	12	Total
<b>All Students</b>	0.00	0.36	5.09	8.36	59.27	18.55	8.36	100
<b>Gender</b>								
Female	0.00	0.00	3.17	7.94	61.90	19.84	7.14	100
Male	0.00	0.67	6.71	8.72	57.05	17.45	9.40	100
<b>Ethnicity</b>								
Hispanic/Latino	0.00	0.00	0.00	6.25	50.00	31.25	12.50	100
American Indian or Alaska Native								
Asian	0.00	9.09	81.82	9.09	0.00	0.00	0.00	100
Black or African American	0.00	0.00	0.47	8.49	60.85	20.28	9.91	100
Native Hawaiian or Other Pacific	0.00	0.00	0.00	0.00	100.00	0.00	0.00	100
White	0.00	0.00	11.43	8.57	71.43	8.57	0.00	100
Two or More Races								
<b>Education Classification</b>								
Regular	0.00	0.00	1.83	8.26	60.09	20.18	9.63	100
Special	0.00	0.00	0.00	6.98	72.09	16.28	4.65	100
Gifted	0.00	7.14	71.43	14.29	7.14	0.00	0.00	100
<b>Economic Status*</b>								
Economically Disadvantaged	—	—	—	—	—	—	—	—
Not Economically Disadvantaged	—	—	—	—	—	—	—	—
<b>English Learner Status</b>								
Non-EL	0.00	0.38	5.26	8.65	60.53	17.67	7.52	100
EL	0.00	0.00	0.00	0.00	22.22	44.44	33.33	100
<b>Migrant Status</b>								
Non-migrant	0.00	0.36	5.09	8.36	59.27	18.55	8.36	100
Migrant								
<b>Section 504 Status</b>								
Non-section 504	0.00	0.42	5.86	7.95	61.92	16.74	7.11	100
Section 504	0.00	0.00	0.00	11.11	41.67	30.56	16.67	100
<b>Homeless Status</b>								
Not Homeless	0.00	0.37	5.19	8.52	59.26	18.15	8.52	100
Homeless	0.00	0.00	0.00	0.00	60.00	40.00	0.00	100
<b>Military Affiliation</b>								
Not Military Affiliated	0.00	0.37	5.15	8.46	59.19	18.38	8.46	100
Military Affiliated	0.00	0.00	0.00	0.00	66.67	33.33	0.00	100
<b>Foster Care Status</b>								
Not in Foster Care	0.00	0.37	5.20	8.18	58.74	18.96	8.55	100
Foster Care	0.00	0.00	0.00	16.67	83.33	0.00	0.00	100

\* Economic status information is not available for the fall and summer administrations.

## 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](http://researchnetwork.pearson.com/wpcontent/uploads/Bulletin21_Evidence_Based_Standard_Setting.pdf)
- Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO for Windows [Computer software]. Lincolnwood, IL: Scientific Software International.
- Camilli, G., & Shepard, A. L. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage Publication.
- Center for Assessment. (2017, June). *LEAP 2017: English language arts -grade 6 summary – comparability with PARCC performance standards* (Memorandum). Dove, NH.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.
- Data Recognition Corporation. (2016). *Interpretive guide: Grades 3–8 ELA and math*. Maple Grove, MN: Author.
- Dorans, N. J., & Schmitt, M. P. (1991). *Constructed response and differential item functioning: A pragmatic approach* (Research Report No. RR-91-47). Princeton, NJ: Educational Testing Service.
- Educational Testing Service, Pearson, & Measured Progress. (2016). *Final technical report for 2015 administration*. PARCC. Retrieved from <https://eric.ed.gov/?q=source%3a%22Partnership+for+Assessment+of+Readiness+for+College+and+Careers%22&id=ED599097>
- Green, D. R. (1975). *Procedures for assessing bias in achievement tests*. Paper presented at the National Institute of Education Conference on Test Bias, Annapolis, MD.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Hingham, MA: Kluwer-Nijhoff Publishing.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel Procedure. In H. Wainer & H. I. Braun (Eds.), *Test Validity*, (pp. 129–145). Hillsdale, NJ: Erlbaum.

- Kim, S., & Kolen, M. (2004). STUIRT: A computer program for scale transformation under unidimensional item response 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*(2), 179–197.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*(4), 719–748.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159–176.
- Pearson. (2015). *Performance level setting technical report*. PARCC. Retrieved from <https://eric.ed.gov/?q=source%3a%22Partnership+for+Assessment+of+Readiness+for+College+and+Careers%22&pg=2&id=ED599257>
- Pearson. (2017). *PARCC: Final technical report for 2016 administration*. PARCC. Retrieved from <https://eric.ed.gov/?q=PARCC+Final+Technical+Report&id=ED599197>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.
- Schumacker, R. E. (1996). Disattenuating correlation coefficients. *Rasch Measurement Transactions, 10*(1), 479.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*(2), 201–210.
- Thompson, S., & Thurlow, M. (2002). *Universally designed assessments: Better tests for everyone!* (Policy Directions No. 14). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from <http://www.cehd.umn.edu/NCEO/OnlinePUBs/Policy14.htm>
- Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30*(3), 233–251.