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

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

April 2020

DATA RECOGNITION
DRC
CORPORATION

# 2019 Technical Summary

The tests used in Louisiana are carefully constructed to fairly assess the progress of Louisiana students. This document provides an overview of the process and summarizes some of the key psychometric information 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.

## INTRODUCTION

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 math students in grades 3 through 8, except those qualifying for the LEAP Alternate Assessment Level 1 (LAA 1), took the LEAP 2025 assessments.

Beginning in 2017-2018 the Louisiana Department of Education (LDOE) transitioned to LEAP 2025 high school ELA and mathematics assessments 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. The implementation (use) of LEAP 2025assessments provides a coherent and consistent measure of student performance and growth from grades three through high school. Students in grades 9-12 enrolled in English I, English II, Algebra I, and Geometry took the corresponding LEAP 2025 high school assessments.

The Board of Elementary and Secondary Education (BESE) and the LDOE are committed to ensuring that every student is on track to be successful in postsecondary education and the workforce through their comprehensive plan, Louisiana Believes. The LEAP 2025 supports this vision by measuring the full range of student performance, including the performance of high- and low-performing students and providing information for educators and parents about student readiness and whether students are "on track" for college and careers.

## TEST CONTENT DEVELOPMENT

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. All mathematics assessments were translated into Spanish forms.

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.

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 Partnership for Assessment of Readiness for College and Careers (PARCC) flagship blueprint that was used for the PARCC 2018–2019 test administrations. The specific course test blueprints for the 2018–2019 LEAP 2025 ELA assessments for high school were designed with the goal for all students to read, understand, and express understanding of complex, grade-level texts. The specific course test blueprints for the 2018–2019 LEAP 2025 mathematics assessments for high school 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 ELA and mathematics assessments for high school provide questions that have been reviewed by Louisiana educators to ensure their alignment to the Louisiana Student Standards and appropriateness for Louisiana students; measure the full range of student performance, including the performance of high- and low-performing students; and inform educators and parents about student readiness in ELA and mathematics and whether students are "on track" for college and 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 2016 Mathematics and ELA as follows:

The assessments shall be

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

## 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.

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

The reliability coefficients for the LEAP 2025 HS assessments are reported in Table 1. English I and English II have one writing component (RI or RL) that is the same score of 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. Reliability statistics were computed using only initial testers; therefore, it is not reported for the summer administration due to few initial testers.

**Table 1 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 |

## CONSTRUCT-RELATED VALIDITY

In addition to content validity addressed in the Test Content Development and Reliability sections, additional evidence of validity, especially construct-related validity, is demonstrated through studies of convergent and divergent 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 Algebra I test is designed to measure a single overall construct—Algebra I achievement; therefore, the items comprising the Algebra I LEAP 2025 should measure only Algebra I, not language or reading.

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.

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.

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 the content and the 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.

To present evidence of construct-related validity, the 2019 Louisiana LEAP High School 2025 Technical Report describes in detail the following validity studies:

- Decision Accuracy,
- Decision Consistency,
- Principal Components Analysis,
- Correlations among Claims, Subclaims, and Subcategories
- Reliability of Claims, Subclaims, and Subcategories, and
- Divergent (Discriminant) Validity.

## 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

### 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.

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

### 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.

## 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).

## 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.

## 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 2, ELA Reporting Categories and Subcategories. Mathematics has four subcategories, as described in Table 3, Overview of LEAP 2025 Mathematics Task Types and Reporting Categories. Subcategory performance is reported in one of three ratings: *Strong*, *Moderate*, or *Weak*.

**Table 2 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. |

**Table 3 Overview of LEAP 2025 Mathematics Task Types and Reporting Categories**

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

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.

### Use of the Category- and Subcategory-Level Subscores

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.

## EQUATING OF TEST FORMS

A statistical process called equating is needed to convert the scale of the form administered in the current administration to the scale of the forms in previous administrations. This is to ensure that scores from different administrations have the same meaning. Detailed technical information describing this process can be found in the full technical report. This process places the form scores on the same scale, such that students performing on an assessment at the same level of (underlying) achievement should receive the same scaled score.

All forms for a given course should provide comparable scores, and the passing standards across different administrations should be equivalent. Therefore, a form-equating procedure is conducted every year to establish score equivalency across forms. The form-equating process ensures that students are not given

an unfair advantage or disadvantage, despite whether a particular form students take is "easier" or "harder" than a form taken by other students.

## Measurement Model

LEAP 2025 high school item calibration and linking were performed based on item response theory (IRT). Calibration and linking methodology used for the LEAP 2025 High School administrations closely followed most of the PARCC methods referenced in the PARCC document *Final Technical Report for 2015 Administration*.

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

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

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

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

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

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

The software (IRTPRO (Cai, Thissen, & du Toit, 20122) 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. The program implements marginal maximum likelihood (MML) estimation techniques for items and MLE estimation of theta.

## Methodology

LEAP 2025 high school item calibration and linking were performed based on item response theory (IRT). Calibration and linking methodology used for the 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 written knowledge and use of language (WKL) were separately calibrated using the sparse matrix method.

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

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

1 Calibrate the LEAP 2025 High School tests.

2 Link the 2018-2019 LEAP 2025 High School tests to the 2018 LEAP 2025 baseline scale under the non-equivalent common item design.

## RESULTS SUMMARY

Table 4 through Table 7 summarize the mean scale score and standard deviation as well as the percentage of students in each achievement level based on the state population for the 2018-2019 administration of the LEAP 2025 high school ELA and mathematics assessments. All three administrations are presented.

**Table 4 Comparison of Percentage of Students in Each Achievement Level: English I**

| | Year | Administration | Form | | Scale Score | | Percentage in Achievement Level | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | *N* | 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

**Table 5 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

**Table 6 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

**Table 7 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

## REFERENCES


Beimers, J. N., Way, W. D., McClarty, K. L., & Miles, J. A. (2012, January). Evidence based standard setting: Establishing cut scores by integrating research evidence with expert content judgments. Austin, TX: Pearson. Retrieved from http://researchnetwork.pearson.com/wpcontent/uploads/Bulletin21_Evidence_Based_Standard_Setting.pdf

Cai, L., Thissen, D., & du Toit, S. H. C. (2011). IRTPRO for Windows [Computer software]. Lincolnwood, IL: Scientific Software International.

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.

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

Kim, S., & Kolen, M. (2004). *STUIRT: A computer program for scale transformation under unidimensional item repose theory models* (Version 1.0) [Computer software]. Iowa City, IA: University of Iowa.

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

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, 201–210.