2016 LEAP ELA and Mathematics Operational Technical Report

Submitted to the
Louisiana Department of Education

February 2017

DATA RECOGNITION
DRC
CORPORATION

This online-only document was published at a cost of $30,468. This document was published for the Louisiana Department of Education, P.O. Box 94064, Baton Rouge, LA 708042-9064, by Data Recognition Corporation, 13490 Bass Lake Road, Maple Grove, MN 55311. This material was printed in accordance with the standards for printing by State Agencies established pursuant to R.S. 43:31.

TABLE OF CONTENTS

Table of Contents i
Executive Summary 1
E.1 Background 1
E.2 Administration 1
E.3 Student Performance 2
E.4 Validity and Test Scores 3
Chapter 1: Introduction 4
1.1 Background of the Louisiana Education Assessment Program 4
1.2 Purpose of the LEAP 4
1.3 Design of the LEAP 5
1.4 Overview of This Report 5
Chapter 2: The Uses of Test Scores 7
2.1 Uses of Test Scores 7
2.2 Test-Level Scores 8
2.2.1 Scaled Scores 8
2.2.2 Levels of Achievement 8
2.2.3 Use of Test-Level Scores 8
2.3 Claim-Level and Subclaim-Level Subscores 9
2.3.1 Use of the Claim-Level and Subclaim-Level Subscores 9
Chapter 3: Test Content Development 10
3.1 Test Specifications 10
3.1.1 Review of Louisiana Content Standards and PARCC Claims, Subclaims, and Evidence Statements 12
3.1.2 Defining the Specific Test Blueprint 13
3.2 Item Development 14
3.2.1 Considerations of Test Fairness in Item Development 15
3.2.2 Item Writing 15
3.2.3 Pilot Tests 16
3.2.4 Item Reviews 16
3.3 Field Test Selection and Administration 17
3.4 Operational Test Selection 17
3.4.1 ELA Item/Passage Selection Process and Criteria 17
3.4.2 Mathematics Selection Process and Criteria 19
3.4.3 Item-Selection Options for Special Cases 22
3.4.4 Psychometric Review 22
3.5 Universal Design 23
3.6 Accommodations and Designated Supports 23
3.7 Standards and Content Specifications 24
3.8 Summary 25
Chapter 4: Test Administration 58
4.1 Training of Districts 58
4.2 Ancillary Materials 58
4.2.1 Return Material Forms and Guidelines 64
4.2.2 Security Checklists 64

Copyright © 2017 by Louisiana Department of Education.

4.2.3 Interpretive Guides ... 67
4.3 Test Security Measures ... 67
4.4 Test Administration ... 67
4.4.1 Time ... 67
4.4.2 Accommodations ... 68
4.5 Summary ... 77
Chapter 5: Constructed-Response and Technology-Enhanced Scoring ... 78
5.1 Constructed-Response Scoring Process ... 78
5.1.1 Selection of Scoring Evaluators ... 78
5.1.2 Handscoring Training Process ... 80
5.1.3 Monitoring the Scoring Process ... 84
5.1.4 Security ... 86
5.2 Technology-Enhanced Item Scoring Process ... 86
5.3 Multiple-Choice and Multi-Select Item Scoring Process ... 87
5.4 Inter-Rater Reliability ... 87
5.5 Summary ... 93
Chapter 6: Operational Data Analyses ... 94
6.1 Classical Item Statistics ... 94
6.1.1 Test-Level Statistics ... 94
6.1.2 Item-Level Statistics ... 96
6.2 Item Response Theory ... 131
6.3 Calibration Sample ... 132
6.4 Calibration and Linking ... 137
6.4.1 Calibration of LEAP 2016 Tests ... 138
6.4.2 Linking LEAP 2016 PBT to LEAP 2016 CBT for Grades 6 to 8 ... 142
6.4.3 Linking LEAP 2016 to new PARCC 2016 baseline scale ... 153
6.4.4 Linking LEAP 2015 to New PARCC 2016 Baseline Scale ... 173
6.5 Summary ... 176
Chapter 7: Test Results ... 177
7.1 Student Participation ... 177
7.2 Current Administration Data ... 183
7.3 Reports ... 185
7.3.1 Description of Each Type of Report ... 186
7.4 Data Structures ... 187
7.5 Interpreting Test Results ... 187
7.6 Summary ... 187
Chapter 8: Performance Level Setting ... 189
8.1 PARCC Performance Level Setting Process for English Language Arts and Mathematics ... 189
8.2 Cut Scores ... 189
8.2.1 Claim Cut Scores ... 190
8.3 Achievement-Level Descriptors ... 190
8.4 Summary ... 191
Chapter 9: Evidence of Construct-Related Validity ... 192
9.1 Construct-Irrelevant Variance and Construct Underrepresentation ... 192
9.2 Reliability ... 192

Copyright © 2017 by Louisiana Department of Education.

9.2.1 Test Reliability 194
9.2.2 Standard Error of Measurement (SEM) 195
9.2.3 Conditional Standard Error of Measurement (CSEM) 196
9.2.4 Classification Accuracy and Consistency 201
9.2.5 Convergent Validity 206
9.3 Principal Components Analysis 207
9.4 Analyses by Claims and Subclaims 209
9.4.1 Correlations among Claims and Subclaims 210
9.4.2 Reliability of Claims or Subclaims 214
9.4.3 Standard Error of Measurement of Claims or Subclaims 215
9.5 Divergent (Discriminant) Validity 219
9.6 Summary 220
Chapter 10: Fairness 222
10.1 Minimizing Bias through Careful Test Development 223
10.2 Evaluating Bias through Differential Item Functioning (DIF) Statistics 224
10.3 Evaluating Bias through Impact Analysis 234
10.3.1 Reliability 234
10.3.2 Effect Size 245
10.4 Mode Effect Study 264
10.4.1 Sampling Using PSM Method 264
10.5 Summary 266
Chapter 11: References 268

Copyright © 2017 by Louisiana Department of Education.

EXECUTIVE SUMMARY

This report is a technical summary of the 2016 administration of the Louisiana Educational Assessment Program (LEAP). The LEAP is a summative assessment in English Language Arts (ELA) and mathematics administered in grades 3 through 8. These tests are designed to measure students' knowledge of 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) item bank with items on pre-established item response theory (IRT) scales (less than 50%) as well as items from DRC's College and Career Ready (CCR) item bank for ELA and mathematics. This section provides a summary of the 2016 Operational Technical Report.

E.1 Background

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

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

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

As a result of the legislative agreement reached during summer 2015 and to maintain comparability with the 2015 assessments, the LEAP ELA and mathematics assessments at grades 3 through 8 for the 2015–2016 school year consisted of a combination of items taken from the PARCC assessments (less than 50%), with the remaining percentage of items coming from the College and Career Readiness Item Bank belonging to Data Recognition Corporation.

The information that follows describes the 2016 LEAP ELA and mathematics assessments and provides information about how to read and interpret the data on the 2016 assessment reports. For more information, see the introduction of any of the LEAP ELA or mathematics assessment guides, found at http://www.louisianabelieves.com/resources/library/assessment-guidance.

E.2 Administration

In spring 2016, Louisiana administered summative assessments in ELA and mathematics to students in grades 3 through 8. The LEAP was administered in both computer-based and paper-based format. The assessments were available on the computer from April 11, 2016, to May 6, 2016. The paper-

Copyright © 2017 by Louisiana Department of Education.

based tests were administered April 25-April 29, 2016. Test administration is discussed in Chapter 4 of this report.

Approximately 69 districts and 35 charter schools administered ELA and mathematics LEAP tests in grades 3 through 8. Table E.1 shows participation rates based on census data. For the purposes of this report, participation rate is defined as the percentage of students who earned a valid scaled score given the total number of students who were expected to take an online test or receive a test book. The Accountable column shows the total number of students who were expected to take an online test or receive a consumable test booklet. The Percent Reportable column shows the percentage of students who received a scaled score on the LEAP. Further analysis of participation rates is provided in Chapter 7 of this report. The results presented in Table E.1 and Chapter 7 are presented as evidence of reliability and validity of the scores from the LEAP assessments and should not be used for state accountability purposes.

Table E.1 Participation Rates: All Students

Grade	Accountable in ELA	Percent Reportable in ELA	Accountable in Mathematics	Percent Reportable in Mathematics*
3	≥58,180	98.20%	≥58,380	98.90%
4	≥54,960	98.27%	≥55,160	98.93%
5	≥53,610	98.08%	≥53,780	98.75%
6	≥53,690	98.17%	≥53,860	98.91%
7	≥52,130	97.85%	≥52,270	98.66%
8	≥52,640	97.61%	≥52,810	98.52%

*Algebra I students in Grade 8 had the option of taking Algebra EOC instead of LEAP mathematics test

E.3 Student Performance

Tables E.2 and E.3 present the percentage of students classified in each of the 2016 achievement levels.

Table E.2 Percentage of Students Classified in 2016 Achievement Levels Using 2016 Census Data: English Language Arts

Grade	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
3	17.4	20.6	26.0	33.6	2.4
4	9.0	20.5	28.4	34.9	7.2
5	9.0	21.3	31.1	35.1	3.5
6	6.9	19.5	34.8	33.7	5.0
7	10.6	18.9	29.1	30.0	11.4
8	7.4	15.7	28.2	41.0	7.7

Copyright © 2017 by Louisiana Department of Education.

Table E.3: Percentage of Students Classified in 2016 Achievement Levels Using 2016 Census Data: Mathematics

Grade	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
3	9.3	17.0	27.7	36.9	9.1
4	7.3	22.8	29.4	34.4	6.1
5	10.4	25.3	31.0	27.3	6.0
6	10.5	29.7	32.2	23.6	3.9
7	9.2	31.3	33.7	23.2	2.6
8	16.5	27.5	26.2	27.8	2.0

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 test scores. Chapter 2 discusses the uses of the LEAP tests scores. Chapter 3 discusses the test development process used to create the LEAP, which is important to the content-related validity of the LEAP scores. Chapter 4 presents information on test administration. Chapter 5 discusses the scoring of constructed-response items, as well as 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 2016 administration and gives an overview of the score reports that were delivered to the districts electronically for distribution to schools and parents. Chapter 8 highlights the standard-setting procedures implemented by PARCC since PARCC standards were used for the LEAP. Chapter 9 discusses reliability and construct-related validity. In addition, the assumption that the content-area LEAP tests are unidimensional is evaluated. For example, the grade-level mathematics portion of the LEAP test should measure one primary dimension (mathematics). Chapter 10 gives an overview of the statistical and development processes used to ensure fairness of the LEAP for all examinees. Some analyses in this document are based on the calibration sample, while others are based on census data. The sources of data used for particular analyses are indicated throughout the technical report.

Copyright © 2017 by Louisiana Department of Education.

CHAPTER 1: INTRODUCTION

The LEAP is designed to measure students' knowledge of ELA and mathematics. This report provides a technical overview of the ELA and mathematics assessments of the 2016 LEAP and presents evidence for the validity of the 2016 LEAP scores.

This chapter of the technical report describes the background, history, purpose, and design of the LEAP and provides an overview of the major sections for the current report.

1.1 Background of the Louisiana Education Assessment Program

In July 2010, the Board of Elementary and Secondary Education (BESE) adopted the Common Core State Standards (CCSS) in ELA and mathematics. To ease the transition to assessments based on the CCSS, transitional assessments for ELA and mathematics were administered in spring 2013. The LEAP transitional tests administered in 2012–2013 were revised. The prior LEAP tests at grades 4 and 8 measured skills presented in grade clusters, and *i*LEAP tests at grades 3, 5, 6, and 7 were based on grade level expectations. In the LEAP ELA transitional tests, the writing prompt was replaced by a new type of prompt that asked students to read one or two passages and then write a composition that included evidence from the text(s) to support the writer's ideas. The LEAP mathematics transitional tests administered in 2012–2013 changed to include only items that measured content common to the grade 4 or grade 8 GLEs and the CCSS.

In 2013–2014, further changes were made to the tests to continue the transition to the CCSS. The most substantial changes were to the ELA and mathematics tests, which measured only the content of the CCSS.

In 2013–2014, the ELA tests continued to use the text-based writing prompt. The Using Information Resources session, which focused on skimming and scanning skills to locate selected information, was replaced by the Research to Build Knowledge session, which required students to demonstrate a thorough understanding of the sources provided. A 4-point extended-response item was added to the Reading and Responding session of the grade 4 LEAP test. The Proofreading session of the LEAP test was replaced by the Language session, which used a format similar to the previous LEAP session but focused on more grade-specific skills.

In 2013–2014, the mathematics tests for LEAP were adjusted so that they were aligned to the CCSS for mathematics, and reporting was based on CCSS domains or combined domains.

In 2014-2015, the tests administered for ELA and mathematics were intact PARCC forms.

Test forms administered in spring 2016 were constructed by DRC. This report includes those activities associated with the 2016 operational administration identified as performed by DRC.

1.2 Purpose of the LEAP

The BESE and the Louisiana Department of Education (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.

Copyright © 2017 by Louisiana Department of Education.

1.3 Design of the LEAP

A paper-based test (PBT) was administered in grades 3 through 8 for both ELA and mathematics, and a computer-based test (CBT) was administered in both subjects in all grades except for grade 3. Additionally, the mathematics form was translated to Spanish in all grades. Braille and large print test forms were constructed for each grade and content area to enable students who are visually impaired to participate in LEAP testing.

1.4 Overview of This Report

This technical report documents the major activities of the testing cycle and provides comprehensive details that confirm that the processes and procedures applied in the LEAP 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. An overview of major activities documented within this report is provided below.

Use of Test Scores (Chapter 2)

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

Item and Test Development (Chapter 3)

Chapter 3 of the technical report provides a summary of the test development activities that occurred to create the spring 2016 operational test forms.

Test Administration (Chapter 4)

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

Scoring Constructed-Response Items (Chapter 5)

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

Operational Data Analyses (Chapter 6)

Chapter 6 of the technical report includes a detailed description of the operational analyses of the 2016 LEAP, which include the following major parts: the classical item analysis; calibration, scaling, and linking using item response theory (IRT) models; and student scoring. This chapter also describes the demographics of the calibration samples and compares it to state census data. It reports the results of the classical item analysis, as well as the results of the calibration, scaling, and linking.

Copyright © 2017 by Louisiana Department of Education.

Test Results and Reporting (Chapter 7)

Chapter 7 of the technical report contains information on the results of the Spring 2016 LEAP administration. Detailed summary statistics based on scaled scores and achievement-level information are also provided. Finally, this chapter presents information on the score reports sent to districts.

Standard Setting (Chapter 8)

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

Reliability and Validity Evidence (Chapter 9)

Chapter 9 of the technical report provides evidence of reliability and validity of the LEAP scores. This chapter provides detailed results of the reliability of the tests, as well as information on the decision consistency of the cut scores. It also provides evidence of construct validity for the LEAP scores.

Fairness (Chapter 10)

Chapter 10 of the technical report discusses fairness and how the LEAP tests are constructed to be fair to all Louisiana students. This chapter summarizes the results of the differential item functioning (DIF) analysis. It also discusses the results of an impact analysis to determine whether large differences exist between demographic groups in Louisiana. Results of the administration mode study are also summarized.

Copyright © 2017 by Louisiana Department of Education.

CHAPTER 2: THE USES OF TEST SCORES

Validity is the central component of the Louisiana Educational Assessment Program (LEAP) testing program. The following excerpt is from the Standards for Educational and Psychological Testing (hereafter the Standards) from the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME), 2014):

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

As stated by the *Standards*, the validity of a testing program hinges on the use of the test scores. Validity evidence that supports the uses of the LEAP test scores is provided in this technical report. This chapter examines some possible uses of the LEAP test scores.

Chapters 3 through 10 of this technical report provide additional evidence for these uses, as well as technical support for some of the interpretations and uses of test scores. The information in Chapters 3 through 10 also provides a firm foundation of evidence that the LEAP tests measure what they are intended to measure. However, this technical report cannot anticipate all possible interpretations and uses of the LEAP scores. It is recommended that policy and program evaluation studies, in accordance with the *Standards*, be conducted to support some of the uses of the LEAP scores that are anticipated.

2.1 Uses of Test Scores

To understand whether a test score is being used properly, it is necessary to understand the purpose of the test. The intended uses of the LEAP scores include the following:

- evaluating student's overall achievement in the subject matter
- identifying Louisiana students' strengths and weaknesses
- communicating expectations for all students
- evaluating school-, district-, and/or state-level programs
- informing stakeholders (teachers, school administrators, district administrators, Louisiana Department of Education (LDOE) staff members, parents, and the public) on the status of the progress toward meeting the state's academic achievement standards

This technical report refers to the use of the test-level scores (scaled scores and achievement levels); claim-level scores and achievement level classifications; and subclaim-level scores and achievement level classifications.

Copyright © 2017 by Louisiana Department of Education.

2.2 Test-Level Scores

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

The ELA and mathematics test forms were developed by DRC Test Development staff using the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium's item bank with items on pre-established IRT scales for ELA and mathematics, as well as DRC-developed items that were reviewed and approved for use by LDOE content experts. Braille and large-print forms were developed by DRC.

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

2.2.1 Scaled Scores

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

Scaled scores are derived from raw scores (number correct). 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. Scaled scores are more meaningful than raw scores, however, because they maintain their meaning year-to-year, thus allowing comparisons of different test forms across the entire range of the ability scale.

2.2.2 Levels of Achievement

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

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

2.2.3 Use of Test-Level Scores

The LEAP scaled scores and achievement levels provide summary evidence of student achievement in ELA or mathematics. Classroom teachers may use these scores as evidence of student achievement in these content areas. At the aggregate level, district and school administrators may use this information for activities such as planning curriculum. The results presented in this

Copyright © 2017 by Louisiana Department of Education.

technical report provide evidence that the scaled scores are a valid and reliable indicator of student performance in ELA and mathematics.

2.3 Claim-Level and Subclaim-Level Subscores

The Claim-level subscores indicate student performance in terms of performance rating for each Claim in ELA: Reading and Writing. A student's performance on the ELA Claims is reported in one of three ratings: *Weak*, *Moderate*, or *Strong*.

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

Although the performance ratings are determined only by the items included within a claim or subclaim, the level of knowledge and ability needed to achieve a performance rating is connected to the level of knowledge and ability required by the subject-level achievement tests: a *Weak* rating is comparable to the knowledge and ability required below the *Basic* achievement level; a *Moderate* rating requires similar knowledge and ability as the *Basic* achievement level; and a *Strong* rating requires similar knowledge and ability of at least the *Mastery* achievement level.

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

The purpose of reporting Claim- or subclaim-level subscores on LEAP tests is to show for each student the relationship between the overall achievement being measured and the skills in each of the areas defined by the Claims in ELA and subclaims in ELA and mathematics. Teachers may use these subscores for individual students as indicators of strengths and weaknesses, but they are best corroborated by other evidence, such as grades, teacher feedback, and scores on other tests. Chapter 3 of this technical report provides evidence of content validity that supports the use of the claim- or subclaim-level subscores. Chapter 9 of this technical report provides evidence of construct validity that further supports the use of these subscores.

District and school administrators may compare their aggregate results with the state means to better understand their strengths and weaknesses within a content area. Caution should be exercised when comparing claim- or subclaim-level subscores between students or across years, because different items will comprise these subscores and these items may vary in difficulty.

Copyright © 2017 by Louisiana Department of Education.

CHAPTER 3: TEST CONTENT DEVELOPMENT

Content-related validity in achievement tests is evidenced by a correspondence between test content and a specification of the content domain. Content-related validity can be demonstrated through consistent adherence to test blueprints, through a high-quality test development process that includes review of items for accessibility to English Language Learners and students with disabilities, and through alignment studies performed by independent groups. This section provides a detailed discussion of the test development cycle. In particular, it shows how the LEAP 2016 followed rigorous procedures to construct tests that reflect the full range of content that the LEAP 2016 was expected to cover.

This chapter is particularly relevant to Testing Standards 4.0, 4.1, and 4.7. It also addresses Standards 3.1, 3.2, 3.9., 4.12, and 7.4, which are discussed in pertinent sections of this chapter. Standards 4.0, 4.1, and 4.7 are from Chapter 4 of the AERA, APA, and NCME (2014) *Standards for Educational and Psychological Testing*, "Test Design and Development." Each of these Standards and the way each Standard is addressed are presented in 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)

The purpose of this chapter is to document the test development process used for the LEAP. It describes steps taken to create the LEAP tests, from the development of test specifications to the creation of operational forms.

3.1 Test Specifications

Standard 4.1 states the following:

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

The purpose of the test is discussed in Chapter 2. The LEAP 2016 domains are generally defined as the knowledge and skills that are identified within the 2015–2016 Louisiana Content Standards for ELA and mathematics. This framework, in turn, is based on prior consensus among the LDOE, Louisiana educators, and experienced subject-matter experts regarding what is important for teachers to teach and students to learn, as further described by the Partnership for Assessment of Readiness for College and Careers (PARCC) evidence statements. Following this content structure was one aspect that made LEAP scores comparable to other states in the nation that also followed the content structure. Concurrent with the development of LEAP 2016, Louisiana was in the process of revising its ELA and mathematics content standards. These revised standards will guide LEAP test development for future years.

Copyright © 2017 by Louisiana Department of Education.

Evidence of validity based on test content includes information about the test specifications, test design, and test blueprint. Test development involves creating a design framework from the statement of the construct to be measured.

The LEAP 2016 test specifications consisted of a test blueprint and a test design for each grade level/content area. In order to construct the assessments following the LDOE-approved test blueprints and test designs, LDOE and DRC collaborated to use two banks of items that were aligned to the standards. The blueprints and test designs were closely aligned to the PARCC Operational blueprint that had been used the previous school year in Louisiana and that were created by LDOE in collaboration with DRC. The English Language Arts and mathematics LEAP Assessments for grades 3–8 were developed based on the requirements of RFP # 678PUR-LEAP 2016 Math and ELA as follows:

The assessments shall be

- aligned to Louisiana ELA and mathematics content 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 that report student performance using achievement levels that are comparable to a significant number of other states having similarly high expectations for student learning and comparable to Louisiana's 2014–2015 grades 3 to 8 ELA and mathematics assessments;
- developed and reviewed with LDOE assessment staff involvement;
- not designed to be computer adaptive;
- used in assessing students' readiness to successfully transition to postsecondary education and the workplace; and
- administered through a separate administration contract in multiple modalities.

Additionally,

- less than 50 percent of the items included in the selected assessments shall be based upon a blueprint or intellectual property developed by the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium, or any other federally funded consortium of states; and
- no item included in the selected assessments shall be based upon a blueprint or intellectual property developed by a consortium of states predominantly funded by organizations primarily dedicated to political advocacy.

The products of the above requirements are dual-mode assessments—paper-based tests (PBT) and computer-based tests (CBT)—composed of PARCC and DRC test items aligned to the Louisiana content standards. The contract with PARCC provided for the use of items and related passages for one complete PARCC operational test form for each content area and grade based on PARCC's 2016 test blueprints that can be administered in a dual-mode testing environment (PBT or CBT). These PARCC items/passages and the DRC Career- and College-Ready (CCR) items became the available item pool used for 2016 forms construction. For ELA and mathematics, the LEAP 2016 test blueprints were finalized in December 2015. DRC and LDOE content experts carefully scrutinized each final blueprint to ensure optimal content coverage and prudent use of time and resources. In general, the blueprints represent content sampling proportions that reflect intended emphasis in instruction and mastery at each grade level and which are comparable to PARCC 2016 test blueprints. The test specifications provide the numbers of items by strand, assessment focus, and item type, and they demonstrate the desired proportions within test delivery and available item

Copyright © 2017 by Louisiana Department of Education.

pool constraints. The test designs for ELA and mathematics were finalized in March 2016 by the LDOE and DRC. All assessments were fixed forms.

3.1.1 Review of Louisiana Content Standards and PARCC Claims, Subclaims, and Evidence Statements

For the 2016 assessments, the Louisiana program required

- all items used to construct test forms be aligned to the 2015-2016 Louisiana Content Standards for ELA and mathematics;
- test forms be constructed to yield valid and reliable test results;
- tests that provide for the reporting of student performance using achievement levels that are comparable to a significant number of other states having similarly high expectations for student learning;
- tests that are comparable to Louisiana's 2014–2015 ELA and mathematics assessments.

The above implies that the 2016 assessments should be modeled after PARCC 2016 assessments, and the criteria used to produce the PARCC assessments should be used for the LEAP 2016 ELA and mathematics assessments, even though less than 50 percent of the 2016 assessment may be composed of PARCC items. The critical criteria used to develop 2016 test blueprints became the 2014–2015 Louisiana reporting categories and the 2015-2016 Louisiana standards, PARCC claims, subclaims, evidence statements, and scoring rubrics.

For ELA and mathematics, this meant using the same reporting categories that were used in 2015:

English Language Arts

Reading (Major Claim) and Reading Subclaims:
- Reading Informational Texts (RI)
- Reading Literary Texts (RL)
- Reading Vocabulary (RV)

Writing (Major Claim) and Writing Subclaims:
- Written Expression (WE)
- Writing Knowledge and Use of Language Conventions (WKL)

Mathematics

Major Content (Subclaim A)	solve problems involving the major grade-level content
Additional & Supporting Content (Subclaim B)	solve problems involving additional and supporting grade-level content
Expressing Mathematical Reasoning (Subclaim C)	express mathematical reasoning by constructing mathematical arguments and critiques
Modeling and Application (Subclaim D)	solve real-world problems engaging particularly in the modeling practice

The PARCC items are aligned to the Common Core State Standards (CCSS), which were also the Louisiana Academic Standards for 2015–2016. The DRC Career-and College-Ready (CCR) items

Copyright © 2017 by Louisiana Department of Education.

were also aligned to the CCSS, and during forms construction, PARCC criteria was consulted when selecting DRC CCR items.

3.1.2 Defining the Specific Test Blueprint

The 2014–2015 Louisiana PARCC ELA and Math assessments were composed of separate Performance-Based Assessments (PBA) and End-of-Year Assessments (EOY) with the combined scores from both assessments used to report summative assessment results. Since all constructed-response items (CR) were on the PBA tests and all writing scores and mathematics subclaim C and D scores were also derived from PBA test items, it became obvious that the source for these points would have to come from PARCC PBA items, with the remainder from PARCC EOY and/or DRC items as long as the total PARCC item contribution was less than 50 percent of the total items used for each form. As a result, the specific content area and grade-level test blueprints were designed based on five primary factors: (1) the content requirements of the Louisiana Content Standards, (2) the reporting needs of the assessments, (3) LDOE's expressed goal of shorter assessments compared to 2014–2015, (4) the use of specific PARCC versus DRC items to address content and reporting needs of the assessments, and (5) the legal or contractual limitation on the use of PARCC items.

3.1.2.1. English Language Arts Test Blueprints

To address the goal of reducing test administrations and duration of testing, the 2016 ELA tests kept a similar structure to the 2015 assessments: performance-based tasks followed by discrete passage sets. However, instead of administering each part in separate testing windows as was done in 2015, the number of tasks and sessions was reduced, and the test was administered during one PBT or CBT testing window. PARCC made a similar adjustment in 2016 to its test blueprints, moving from separate PBA/EOY tests to one combined end-of-grade summative test. Only two of the three types of performance tasks—Research Simulation Task and Literary Analysis Task or Narrative Writing Task—were included on each of the Louisiana grade-level tests, but all three types were represented across grades 3–8. This allows Louisiana to rotate the tasks given for each grade from year to year and encourages teachers to focus equally on all three writing types. Since the choice of Literary Analysis Task or Narrative Writing Task would be made during the forms construction process, alternative blueprints—one with a Research Simulation Task and a Literary Analysis Task and the other with a Research Simulation Task and a Narrative Writing Task—were created for each grade. During forms construction, the Narrative Writing Task was selected for grades 4 and 8, and the Literary Analysis Task was selected for grades 3, 5, 6, and 7, all based on item performance and the quality of the available passage sets for each task. The remainder of the test forms was composed of the DRC CCR reading literary and informational passage sets.

The session testing times shown in the ELA test blueprints (Tables 3.1 to 3.4) are based on PARCC testing times proportioned to be comparable based on the passage and item counts on ELA 2016 PARCC versus LEAP tests. In the grade 4 and grade 8 blueprints, the passage set that comes after the Narrative Writing Task in session two is designed to balance the reading load between the Literary Analysis Task and the Narrative Writing Task and to provide consistent timing in sessions two and three.

3.1.2.2. Mathematics Test Blueprints

To address the primary goal of reducing test administrations and duration of testing, the 2016 mathematics tests kept a similar structure to the 2015 assessments: test sessions composed of the four mathematics subclaims using the three mathematics task types (Table 3.5). In addition, instead

Copyright © 2017 by Louisiana Department of Education.

of administering each part (PBA or EOY) in separate testing windows as was done in 2015, the number of tasks and sessions was reduced and the test was administered during one PBT or CBT testing window. PARCC made a similar adjustment in 2016 of their test blueprints from separate PBA/EOY tests into one combined end-of-grade summative test. Since the selection of tasks aligned to subclaims is largely determined by the available item pools, Type II and Type III tasks aligned to subclaims C and D are available only from PARCC, while the Type I tasks aligned to subclaims A and B may come from a combination of the DRC and PARCC item pools, except all necessary two-point Type I items are only available from PARCC. The resulting LEAP 2016 Mathematics Test Blueprints are shown in Tables 3.6 to 3.11.

3.1.2.3. Final Mathematics Test Design

Unlike ELA, whose test blueprints were organized by test sessions 1–3, the mathematics test blueprints were organized by reporting categories, so it was necessary to define the general structure of the test forms into test sessions. The design goal was to have balanced test sessions with a variety of task types and equivalent testing times. For all grade 3–5 forms, the use of calculators is prohibited, except for those students with a calculator accommodation. For session one of grades 6–8, the use of calculators is prohibited, except for those students with a calculator accommodation; calculators are allowed for sessions two and three. The general test structures (Tables 3.16 to 3.21) guided test-form sequencing and design.

The session testing times shown in the general test structures are based on PARCC testing times proportioned to be comparable based on the task counts on mathematics PARCC 2016 versus LEAP 2016 tests.

PARCC's calculator guidelines, outlined in their evidence statement tables, provided the basis for calculator designation of tasks/items. In cases where DRC items had differing calculator designations, items were evaluated for calculator neutrality. If items were considered neutral, the items could be used in either no calculator or calculator sessions. If not, then the items were not used. In some cases, only PARCC items could be used for certain standards because of the differing calculator designations. All DRC mathematics items were developed to align to CCSS and not to PARCC evidence statements.

3.2 Item Development

Item development is 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 LEAP 2016 ELA and mathematics items come from two sources, PARCC and DRC, and were leased to LDOE for use in the Spring 2016 administrations. Louisiana law at the time specifies that less than 50% of the items in each form may be from a national item bank.

All items and passages that make up the research simulation tasks, literary analysis tasks, and narrative writing tasks are from PARCC and comprise less than 50% of the total ELA items used. The PARCC composition of the 2016 ELA test forms is shown in Table 3.12.

Copyright © 2017 by Louisiana Department of Education.

All items and passages that make up the Reading Literary or Informational passage sets are from DRC and comprise more than 50% of the total ELA items used. The DRC composition of the 2016 ELA test forms is shown in Table 3.13.

All items composing the mathematics Subclaims C and D (Type II and Type III tasks) and the two-point Type I tasks are from PARCC. Some of the one-point Type I tasks are also from PARCC. The PARCC composition of test forms is less than 50% of the total mathematics item used and is shown in Table 3.14.

All DRC items used are one-point Type I tasks aligned to Subclaims A and B and comprise more than 50% of the mathematics items used. The DRC composition of test forms is shown in Table 3.15.

The PARCC items selected for use on the 2016 forms were used for equating back to the 2014 LEAP tests. The 2014 LEAP tests also consisted of items from PARCC. Refer to PARCC's website for their processes, procedures, and timelines for item development, review, field testing, and item selection for operational use. DRC's item development process for those items selected for use on the 2016 forms included the following steps: a small-scale pilot of the mathematics items in spring 2014, a pilot test in fall 2014, and a field test in fall 2015. Items/tasks administered for the 2016 LEAP operational test complied with content specifications and with the item and task specifications that were consistent with PARCC criteria. Operational forms were selected based on LEAP test blueprint specifications supported by statistical data from PARCC operational testing.

3.2.1 Considerations of Test Fairness in Item Development

Standard 3.2 is particularly relevant to fairness in item development:

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

Bias and sensitivity guidelines used during the development of the PARCC and DRC items help ensure that the assessments are fair for all groups of test takers, despite differences in characteristics including, but not limited to, disability status, ethnic group, gender, regional background, native language, race, religion, sexual orientation, and socioeconomic status. DRC strongly relied on the Bias and Sensitivity Guidelines in the development of the assessments, particularly in item writing and review. Items had to comply with the Bias and Sensitivity Guidelines in order to be included in the assessments.

3.2.2 Item Writing

The DRC CCR item development occurred from 2013 to 2015. DRC worked with qualified item writers throughout the test development cycle to develop items. The item writers were trained on DRC's CCR content specifications, item and task specifications, and stimulus specifications (ELA). In addition, DRC test development experts held regular meetings to provide direction and feedback to the item writers. An item development plan was used to specify for item-writing teams the number and distribution of items to be written. Pools of items were written specifically to support a variety of approaches for operational use.

Copyright © 2017 by Louisiana Department of Education.

3.2.3 Pilot Tests

The DRC pilot test administration in spring 2014 was designed to collect data on the statistical quality of the mathematics items. One of the main goals was to try out a variety of new technology-enabled items to determine the best use when assessing the mathematics standards. The content specifications and item and task specifications were adjusted after the pilot test, prior to development of new items and tasks for the field test.

3.2.4 Item Reviews

PARCC conducted external bias and fairness reviews independent of DRC's reviews and as part of their ongoing item development practices. Refer to PARCC's website for information on PARCC's item review processes and procedures. The external DRC bias review committee comprised representatives from various backgrounds whose purpose was to screen the items for racial, socioeconomic, gender, and other sensitivity issues. This follows the *Standards*.

Standard 3.1states 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)

External reviewers, under DRC guidance, reviewed all CCR items and item stimuli for accessibility, bias/sensitivity, and content. (Item stimuli include the reading passages used on the ELA assessments and the figures and graphics used on the mathematics assessments.) Prior to the 2015 field test, twelve ELA panelists, ten mathematics panelists, and ten bias, fairness, and sensitivity panelists reviewed items for accessibility and bias/sensitivity. During the accessibility reviews, panelists identified issues that could negatively affect a student's ability to access stimuli, items, or performance tasks or that also negatively affected students' ability to provide valid evidence about an assessment target. During the bias and sensitivity review, experts identified content in stimuli or items that could negatively affect a student's ability to produce a correct response because of his or her background. The content review focused on developmental appropriateness and alignment of stimuli and items to the content specifications. Panelists in the content review also checked the accuracy of the content, answer keys, and scoring materials. Items flagged for accessibility, bias/sensitivity, and/or content concerns were either revised to address the issues identified by the panelists or removed from the item pool. The final and approved selections of DRC's external panel and internal content experts became DRC's item bank and was used to select items for the LEAP ELA and mathematics tests. Selected items were reviewed again for accessibility, bias/sensitivity, and content by LDOE content specialists during the forms construction process.

Copyright © 2017 by Louisiana Department of Education.

3.3 Field Test Selection and Administration

Approved items were included in DRC's field test administration from October to June 2015. The major purposes of the field test were to administer and calibrate a sufficiently large number of items to ensure a successful operational launch of summative assessments; to obtain classical statistics to inform item selection; to evaluate the protocols for the test administration and computer delivery system (technology infrastructure); and to implement targeted test accommodations and elements of universal design. A total of 5,311 items across all grade levels (K–12) from the ELA and mathematics field tests met these varied purposes. This pool and the PARCC pool of operational items were used to develop the LEAP administered in spring 2016.

3.4 Operational Test Selection

Operational item selections for 2016 were performed from December 2015 through March 2016 by LDOE and DRC. The PARCC and DRC CCR item pool was used to select fixed LEAP ELA and mathematics forms.

3.4.1 ELA Item/Passage Selection Process and Criteria

The item/passage-selection process used for forms construction was a content-focused, collaborative process between the LDOE and DRC ELA content specialists, followed by a psychometric evaluation of each selection. Item selection was greatly limited for the PARCC component of each form since PARCC provided only one set of passages and items for each type of performance task for each grade. Since PARCC items are the only links to the PARCC scale used for equating purposes, the critical psychometric consideration, other than individual item performance, is the degree to which the PARCC selection reflects the 2015 and 2016 targets. Although the PARCC pool was limited, when possible, very difficult (IRT B > 2.0) and/or not discriminating (IRT A <0.3) items were avoided.

Item Selection Guidelines

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

The data from the DRC CCR pilot test items consisted of classical statistics. This data was used only to flag potential item-performance issues, not for passage/item selection. Passage/item selection, therefore, was based mainly on alignment to Louisiana standards and test blueprint requirements as well as passage type, diversity of topics and authors, quality, reading load,

Copyright © 2017 by Louisiana Department of Education.

distribution of item types, and overall standards coverage. The goal was to select passages and items that are similar to PARCC's stand-alone passage sets.

3.4.1.1. ELA Content Review and Forms Development

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

Guidelines for Placing Tasks and Passage Sets into Forms

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

3.4.1.2. Review of the ELA Items and Forms

DRC and LDOE content-area assessment specialists also check to see that the items are in compliance with the guidelines provided by the Louisiana Department of Education, including alignment to the content standards. At every stage of the test development process, the alignment of the item to the content standard must be reviewed and verified, since establishing content validity is one of the most important aspects in the legal defensibility of a test. As a result, it is essential that an item selected for a form link directly to the content standard which it measures. The content-area assessment specialists also verify all items against their designated content codes and metadata, both to evaluate the correctness of the coding and to ensure that the given item measures what it purports to measure.

In addition, the content-area assessment specialists review each item for item quality, making sure that the test items are in compliance with industry guidelines for clarity, style, accuracy, and appropriateness for Louisiana students. While there are many published guidelines for reviewing assessment items, the list below serves to summarize the major considerations the content-area assessment specialists follow when reviewing items to make sure they conform to item quality standards for good, reliable, fair test questions.

Copyright © 2017 by Louisiana Department of Education.

Guidelines for Reviewing Items Selected for Forms

A good item should

- (if multiple choice [MC]) have only one clear, correct answer and contain answer choices that are reasonably parallel in length and structure;
- (if multiple select [MS]) have only the indicated number of correct answers and contain answer choices that are reasonably parallel in length and structure;
- have a correctly assigned content code (item map);
- measure one main idea or standard, unless it is a complex item, such as a prose constructed-response item (PCR);
- measure the objective or content standard(s) it is designed to measure;
- be at the appropriate level of rigor;
- be simple, direct, and free of ambiguity;
- make use of vocabulary and sentence structure that is appropriate to the grade level of the student being tested;
- be based on content that is accurate and current;
- (when appropriate) contain stimulus material that is clear and concise and provides all of the information that is needed;
- (when appropriate) contain graphics that are clearly labeled;
- contain answer choices that are plausible and reasonable in terms of the requirements of the question, as well as 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 subject area;
- be free of gender, ethnic, cultural, socioeconomic, and regional bias.

Paper-based test forms (PBT) were developed for grade 3, and paper- and computer-based test forms (CBT) were developed for grades 4–8. The dual-mode forms were essentially identical except for a small quantity (two to three items) of one-point technology-enhanced interactive items (TEIs) in each CBT form. Items used on PBT forms as replacements for the TEIs were selected-response items addressing the same content standards as the TEIs and, when possible, of similar rigor.

3.4.2 Mathematics Selection Process and Criteria

The item selection process used for forms construction was a content-focused, collaborative process between the LDOE and DRC mathematics content specialists, followed by a psychometric evaluation of each selection. Item selection was greatly limited for the subclaims C and D PARCC component of each form since PARCC provided only one set of Type II and Type III tasks each for most grades. For equating purposes, PARCC items are the only links to the PARCC scale. Therefore, in addition to individual PARCC item performance, a critical psychometric consideration was the degree to which the PARCC selection reflected the 2015 and 2016 targets. Although the PARCC pool was limited, when possible, very difficult (IRT B > 2.0) and/or not discriminating

Copyright © 2017 by Louisiana Department of Education.

(IRT A < 0.3) items were avoided. In one case, PARCC agreed to provide the replacement for an extremely difficult item.

Item-Selection Guidelines

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

The DRC items included classical statistics. This data was used only to flag for potential item-performance issues, not for item selection. Item selection was based on alignment to Louisiana standards and test blueprint requirements, calculator/non-calculator requirements, quality, and distribution of item types. The goal was to select items that are similar to PARCC Type I tasks and related mathematics evidence statements.

3.4.2.1. Mathematics Content Review and Forms Development

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

Guidelines for Placing Items into Forms

- Forms should include adequate content coverage, as required by the detailed test blueprint.
- No item in a form should “clue” another item on that same form. If it is necessary to use items with clueing issues due to item pool limitations, then these items should be positioned in separated testing sessions.
- Clang association should be avoided. Clang is when a distractor can be associated with a stem word by sound (e.g., rhyming, alliteration) rather than meaning.
- Forms should be ethnically diverse, both in terms of artwork and in terms of names.
- Forms should target an equal representation of genders, both in terms of artwork and names.
- Forms should include a wide range of topics and a variety of questions.
- Correct answer distributions should follow best practice (no more than three keys of the same answer option in a row).
- Forms *will not* contain any items that have been released to the public.

Copyright © 2017 by Louisiana Department of Education.

3.4.2.2. Review of the Mathematics Items and Forms

DRC and LDOE content-area assessment specialists also check to see that the items are in compliance with the guidelines provided by the LDOE, including alignment to the content standard. At every stage of the test development process, the alignment of the item to the content standard must be reviewed and verified, since establishing content validity is one of the most important aspects in the legal defensibility of a test. As a result, it is essential that an item selected for a form link directly to the content standard which it measures. The content-area assessment specialists also verify all items against their designated content codes and metadata, both to evaluate the correctness of the coding and to ensure that the given item measures what it purports to measure.

In addition, the content-area assessment specialists review each item for item quality, making sure that the test items are in compliance with industry guidelines for clarity, style, accuracy, and appropriateness for Louisiana students. While there are many published guidelines for reviewing assessment items, the list below serves to summarize the major considerations the content-area assessment specialists follow when reviewing items to make sure they conform to item quality standards for good, reliable, fair test questions.

Guidelines for Reviewing Items Selected for Forms

A good item should

- (if multiple choice [MC]) have only one clear, correct answer and contain answer choices that are reasonably parallel in length and structure;
- (if multiple select [MS]) have multiple correct answers, as identified by the item key, and contain answer choices that are reasonably parallel in length and structure;
- have a correctly assigned content code (item map);
- measure one main idea or standard, unless it is a complex task, such as Type II or Type III task;
- measure the objective or content standard(s) it is designed to measure;
- be at the appropriate level of rigor;
- be simple, direct, and free of ambiguity;
- make use of vocabulary and sentence structure that is appropriate to the grade level of the student being tested;
- be based on content that is accurate and current;
- (when appropriate) contain stimulus material that is clear and concise and provides all of the information that is needed;
- (when appropriate) contain graphics that are clearly labeled;
- contain answer choices that are plausible and reasonable in terms of the requirements of the question, as well as 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 subject area;
- be free of gender, ethnic, cultural, socioeconomic, and regional bias.

Copyright © 2017 by Louisiana Department of Education.

Paper-based test forms (PBT) were developed for grade 3, and paper- and computer-based test forms (CBT) were developed for grades 4–8. The dual-mode forms are essentially identical except for a small quantity (one to two items) of one-point technology-enhanced interactive items (TEIs) in each CBT form. Items used on PBT forms as replacements for the TEIs are selected-response items addressing the same content standards as the TEIs and, when possible, are of similar rigor. CBT short-answer items (SA) were reformatted as gridded-response items (GR) for use on PBT forms.

3.4.3 Item-Selection Options for Special Cases

It may not be possible to comply with all of the psychometric criteria for item/form difficulty due to item pool limitations. In these cases, critical psychometric guidelines are followed while allowing some tolerance on less critical item-selection guidelines. The tolerance of meeting target characteristics, the relative exposure of previously used operational items, and other considerations (such as content coverage) are possible areas affected in such cases.

3.4.4 Psychometric Review

The psychometric evaluation of each selection was centered on reviewing the PARCC items with operational item parameters, because the DRC items lacked item parameters and had only classical statistics based on pilot-testing data. As stated previously, the review of the DRC component was focused on assessing content coverage and conformity to test blueprint requirements.

3.4.4.1. Selecting Targets

The 2015 intact form served as a target for selecting the PARCC component of each form. The item pool PARCC provided for 2016 forms selection also served as another target. The two targets, however, overlapped considerably because the 2016 item pool PARCC provided was composed of the 2015 intact form with a few replacement items for the spring 2015 test items PARCC released in the fall of 2015. The rationale for the choice of the targets was that the PARCC component of each new form should closely resemble the total test in terms of psychometric characteristics.

3.4.4.2. Selecting Anchors

Anchor sets used in the common item non-equivalent group design underwent considerable scrutiny due to the generally accepted guideline that the anchor set should mirror the total test in terms of content and item characteristics. One of the critical psychometric considerations for an anchor set, other than individual item performance, is the extent to which the test characteristic curve (TCC) of the anchor set aligns with that of the total (or reference) test. That is, the expected proportion of the maximum raw score based on the TCC of the anchor set and the TCC of the reference test form should be closely overlapped. Another consideration related to the first is that the test information function (TIF) of the anchor set and that of the target test should be aligned, accounting for any difference in the number of items (score points). To compensate for the reduction in the length of the 2016 forms, this guideline was modified in such a way that the TIF for the anchor set may be taller than the TIF for the reference test (i.e., the anchor can have proportionately more highly discriminating items, especially around the performance cuts). Upon receiving an anchor set, DRC psychometricians generated graphs to compare the TCC and TIF of the anchor set to those of the target/reference forms to assess the level of alignment or lack thereof. When the alignment was not adequate and alternative choices were available, DRC content-area assessment specialists updated the anchor set for another round of psychometric review.

Copyright © 2017 by Louisiana Department of Education.

3.4.4.3. Matching the Anchor Items to the Target

Figures 3.1 to 3.6 for ELA and Figures 3.7 to 3.12 for mathematics show the TCC and TIF of the final anchor sets compared to those of the target forms. The two line graphs on the left display the TCC and TIF of the anchor set, the 2015 PARCC intact form, and the 2016 item pool provided by PARCC without taking into account the differences in the number of items (score points). The corresponding graphs on the right are scaled to allow for comparisons of the anchor set and the target forms of different test lengths. The TCCs on the lower right-hand corner show that the expected proportion of the maximum raw score of the anchor set and the target forms overlap closely.

3.5 Universal Design

Grade-level assessments that follow universal design guidelines allow participation of the widest possible range of students, resulting in more valid inferences about students' performance. Such assessments may reduce the need for accommodations by reducing or eliminating access barriers associated with the tests themselves. Table 3.22 presents the elements of universal design (Thompson and Thurlow, 2002). The elements of universal design are relevant to both item development and form construction. This section addresses how the elements of universal design were addressed in the construction of the spring 2016 test forms in compliance with AERA, APA, and NCME (2014) Standard 3.1, which states the following:

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

Universal design requires that grade-level assessments measure the performance of students with a wide range of abilities and skill repertoires, ensuring that students with diverse learning needs receive opportunities to demonstrate competence on the same content. To accommodate the greatest number of students administered the LEAP tests, the assessments include simple, clear, and intuitive instructions and procedures; maximum readability and comprehensibility; and maximum legibility. All of these design components are addressed primarily through the web formatting of the test forms. The page specifications define how directions and test items are placed on the pages, the location and appearance of headers and footers, spacing between an item stem and answer choices, and other page elements to ensure a consistent, legible appearance of online test forms. Written instructions at the beginning of each test session are clearly and simply stated, and the wording of such instructions is standardized as much as possible across content areas and grade levels to ensure clarity and consistency, while being comparable to PARCC.

3.6 Accommodations and Designated Supports

AERA, APA, and 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)

Copyright © 2017 by Louisiana Department of Education.

Students with disabilities or students who are English Language Learners may be provided test administration accommodation based on their Individualized Education Plan (IEP). More information on accommodations can be found in Section 4.4.2 of Chapter 4. Accommodation code definitions can be found in the *Test Administration Manual*.

Braille and large print versions were constructed for each grade/content area to enable students with visual impairments to participate in the LEAP testing. Braille and large print forms for ELA and mathematics are based on the standard-print forms. Specific recommendations on how to transcribe items into braille were provided by the braille publisher to produce the braille version of the LEAP and test administrator's notes that accompany the braille forms. The goal is to maximize the number of items on the braille form, and it was possible to transcribe all items into braille.

Additional access and accommodations features were available for print or online testing, such as blank scratch paper, bookmarks, calculators (calculator section only), color overlays, contrasting colors/reverse colors, directions in native languages, general directions clarified and/or read aloud, general masking, highlighters, line guides, magnifiers/variable zoom, measurement tools, noise buffer/headphones, strikethrough, write on test, mathematics text to speech, mathematics human reader, and a Spanish translation of the mathematics tests.

3.7 Standards and Content Specifications

AERA, APA, and 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. Indeed, the purpose of the item and task specifications is to define the characteristics of the items and tasks that will provide the evidence to support one or more claims. To do this, the item and task specifications delineate the types of evidence, or targets, that should be elicited for each reporting category within a grade level. Then, they provide explicit guidance on how to write items in order to elicit the desired evidence. To address LEAP 2016 comparability goals with PARCC 2016, PARCC claims, subclaims, and evidence statements were used as item and task specifications, along with guidance provided by the *Louisiana State Standards for ELA and Mathematics*.

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

These item specifications were developed for each grade level and standard in order to delineate the expectations of knowledge and skill to be included on test questions in each grade. In addition, the ELA and mathematics item and stimulus specifications provide guidance on determining the grade-appropriateness of task and stimulus materials (the materials that a student must refer to in working on a test question). The stimulus examples also provide information on characteristics of stimuli or activities to avoid because they are not important to the knowledge, skill, or process being measured. This is important because it underscores DRC's efforts to select items that are accessible

Copyright © 2017 by Louisiana Department of Education.

to the widest range of students possible; in other words, LEAP 2016 items were selected according to the principle of universal design.

Table 3.23 provides the distribution of ELA points on the LEAP 2016 by subclaim.

Table 3.24 provides the distribution of ELA points on the LEAP 2016 by claim.

Table 3.25 provides the distribution of ELA items and points on the LEAP 2016 by session and item type.

Tables 3.26 provide the distribution of mathematics points on the LEAP 2016 by subclaim.

Table 3.27 provides the distribution of mathematics tasks and points on the LEAP 2016 by task type.

3.8 Summary

In summary, the overall purpose of this chapter is to explicate the procedures used in the development of the LEAP 2016 grade-level assessments. The efforts by the LDOE and DRC in developing the LEAP are in alignment with multiple best practices of the test industry but, in particular, support the following AERA, APA, and 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)

Copyright © 2017 by Louisiana Department of Education.

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

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

Copyright © 2017 by Louisiana Department of Education.

Table 3.1 Grade 3 English Language Arts Test Blueprint

Grade 3 English Language Arts Test Blueprint and Test Design													
Section	**Task/ Item Set**	**Number of Passages**	**Claims/ Subclaims**	**Number of 2-point PARCC EBSR Items**	**Number of Points from 2-point PARCC EBSR Items**	**Number of 1-point DRC CCR Items (MC, MS, EBSR)**	**Number of Points from 1-point DRC CCR Items (MC, MS, EBSR)**	**Number of PARCC PCR Items**	**Number of Points from PARCC PCR Items**	**Total Items**	**Total Points**	**Standards**	**Testing Time (minutes)**
1	Research Simulation Task	2	Reading: Reading Informational Text/Reading Vocabulary*	6	12	0	0	1	3	6	15	RI.1-3, 5-10, RI.4, L.4, L.5	75
			Writing: Written Expression	0	0	0	0		9	1	9	W.1-2, 7-8, 10, L.3	
			Writing: Knowledge of Language and Conventions	0	0	0	0		3		3	L.1, 2**	
	Totals	2		6	12	0	0	1	15	7	27		
2	Literary Analysis Task	2	Reading: Reading Literary Text/Reading Vocabulary*	5	10	0	0	1	3	5	13	RL.1-3, 5-10 RL.4, L.4, L.5	75
			Writing: Written Expression	0	0	0	0		9	1	9	W.1-2, 10, L.3	
			Writing: Knowledge of Language and Conventions	0	0	0	0		3		3	L.1, 2**	
	Totals	2		5	10	0	0	1	15	6	25		
3	Reading Literary Texts	at least 3	Reading: Reading Literary Text/Reading Vocabulary*	0	0	at least 5	14	0	0	14	14	RL.1-3, 5-10, RL.4, L.4, L.5	60
	Reading Informational Texts		Reading: Reading Informational Text/Reading Vocabulary*	0	0	at least 5		0	0			RI.1-3, 5, 7-10, RI.4, L.4, L.5	
	Totals			0	0	14	14	0	0	14	14		
Grade 3 Totals		**at least 7**	Reading: Reading Literary Text/Reading Vocabulary*	5	10	at least 5	14	2	3	10+	18+	42	210
			Reading: Reading Informational Text/Reading Vocabulary*	6	12	at least 5			3	11+	20+		
			Writing: Written Expression	0	0	0	0		18	2	18	24	
			Writing: Knowledge of Language and Conventions	0	0	0	0		6		6		
			Total	11	22	14	14	2	30	27	66	66	

*There must be at least eight points of reading vocabulary items on the test. **Students advancing through the grades are expected to meet each year's grade-specific standards and retain or further develop skills and understandings mastered in the preceding grades.

Copyright © 2017 by Louisiana Department of Education.

Table 3.2 Grade 4 English Language Arts Test Blueprint

Grade 4 English Language Arts Test Blueprint and Test Design													
Section	**Task/ Item Set**	**Number of Passages**	**Claims/ Subclaims**	**Number of 2-point PARCC EBSR Items**	**Number of Points from 2-point PARCC EBSR Items**	**Number of 1-point DRC CCR Items (MC, MS, EBSR)**	**Number of Points from 1-point DRC CCR Items (MC, MS, EBSR)**	**Number of PARCC PCR Items**	**Number of Points from PARCC PCR Items**	**Total Items**	**Total Points**	**Standards**	**Testing Time (minutes)**
1	Research Simulation Task	3	Reading: Reading Informational Text/Reading Vocabulary*	7	14	0	0	1	4	7	18	RI.1-3, 5-10, RI.4, L.4, L.5	90
			Writing: Written Expression	0	0	0	0		12	1	12	W.1-2, 4, 7-10	
			Writing: Knowledge of Language and Conventions	0	0	0	0		3		3	L.1, 2**	
	Totals	3		7	14	0	0	1	19	8	33		
2	Narrative Writing Task	1	Reading: Reading Literary Text/Reading Vocabulary*	4	8	0	0	1	0	4	8	RL.1-3, 5-10 RL.4, L.4, L.5	75
			Writing: Written Expression	0	0	0	0		9	1	9	W.3, 4, 10, L.3	
			Writing: Knowledge of Language and Conventions	0	0	0	0		3		3	L.1, 2**	
	Reading Literary or Informational Texts	1	Reading: Reading Literary or Informational Text/Reading Vocabulary*	0	0	6	6	0	0	6	6	RL/RI.1-10; L.4, L.5	
	Totals	2		4	8	6	6	1	12	11	26		
3	Reading Literary Texts	at least 3	Reading: Reading Literary Text/Reading Vocabulary*	0	0	at least 4	15	0	0	15	15	RL.1-3, 5-10, RL.4, L.4, L.5	75
	Reading Informational Texts		Reading: Reading Informational Text/Reading Vocabulary*	0	0	at least 4		0	0			RI.1-3, 5, 7-10, RI.4, L.4, L.5	
	Totals			0	0	15	15	0	0	15	15		
Grade 4 Totals		**at least 8**	Reading: Reading Literary Text/Reading Vocabulary*	4	8	at least 5	15	2	0	9+	13+	47	240
			Reading: Reading Informational Text/Reading Vocabulary*	7	14	at least 5			4	12+	23+		
			Writing: Written Expression	0	0	0	0		21	2	21	27	
			Writing: Knowledge of Language and Conventions	0	0	0	0		6		6		
			Total	11	22	21	21	2	31	34	74	74	

*There must be at least eight points of reading vocabulary items on the test. **Students advancing through the grades are expected to meet each year's grade-specific standards and retain or further develop skills and understandings mastered in the preceding grades.

Copyright © 2017 by Louisiana Department of Education.

Table 3.3 Grades 5–7 English Language Arts Test Blueprints

Grades 5–7 English Language Arts Test Blueprint and Test Design													
Section	**Task/ Item Set**	**Number of Passages**	**Claims/ Subclaims**	**Number of 2-point PARCC EBSR Items**	**Number of Points from 2-point PARCC EBSR Items**	**Number of 1-point DRC CCR Items (MC, MS, EBSR)**	**Number of Points from 1-point DRC CCR Items (MC, MS, EBSR)**	**Number of PARCC PCR Items**	**Number of Points from PARCC PCR Items**	**Total Items**	**Total Points**	**Standards**	**Testing Time (minutes)**
1	Research Simulation Task	3	Reading: Reading Informational Text/ Reading Vocabulary*	7	14	0	0	1	4	7	18	RI.1-3, 5-10, RI.4, L.4, L.5	90
			Writing: Written Expression	0	0	0	0		12	1	12	W.1-2, 4, 7-10	
			Writing: Knowledge of Language and Conventions	0	0	0	0		3		3	L.1, 2**	
	Totals	3		7	14	0	0	1	19	8	33		
2	Literary Analysis Task	2	Reading: Reading Literary Text/Reading Vocabulary*	5	10	0	0	1	4	5	14	RL.1-3, 5-10 RL.4, L.4, L.5	75
			Writing: Written Expression	0	0	0	0		12	1	12	W.1-2, 4, 9, 10, L.3	
			Writing: Knowledge of Language and Conventions	0	0	0	0		3		3	L.1, 2**	
	Totals	2		5	10	0	0	1	19	6	29		
3	Reading Literary Texts	at least 3	Reading: Reading Literary Text/Reading Vocabulary*	0	0	at least 5	15	0	0	15	15	RL.1-3, 5-10, RL.4, L.4, L.5	75
	Reading Informational Texts		Reading: Reading Informational Text/Reading Vocabulary*	0	0	at least 5		0	0			RI.1-3, 5-10, RI.4, L.4, L.5	
	Totals			0	0	15	15	0	0	15	15		
Grades 5–7 Totals		**at least 8**	Reading: Reading Literary Text/Reading Vocabulary*	5	10	at least 5	15	2	4	10+	19+	47	240
			Reading: Reading Informational Text/Reading Vocabulary*	7	14	at least 5			4	12+	23+		
			Writing: Written Expression	0	0	0	0		24	2	24	30	
			Writing: Knowledge of Language and Conventions	0	0	0	0		6		6		
			Total	12	24	15	15	2	38	29	77	77	

*There must be at least eight points of reading vocabulary items on the test. **Students advancing through the grades are expected to meet each year's grade-specific standards and retain or further develop skills and understandings mastered in the preceding grades.

Copyright © 2017 by Louisiana Department of Education.

Table 3.4 Grade 8 English Language Arts Test Blueprint

Grade 8 English Language Arts Test Blueprint and Test Design													
Section	**Task/ Item Set**	**Number of Passages**	**Claims/ Subclaims**	**Number of 2-point PARCC EBSR Items**	**Number of Points from 2-point PARCC EBSR Items**	**Number of 1-point DRC CCR Items (MC, MS, EBSR)**	**Number of Points from 1-point DRC CCR Items (MC, MS, EBSR)**	**Number of PARCC PCR Items**	**Number of Points from PARCC PCR Items**	**Total Items**	**Total Points**	**Standards**	**Testing Time (minutes)**
1	Research Simulation Task	3	Reading: Reading Informational Text/Reading Vocabulary*	7	14	0	0	1	4	7	18	RI.1-3, 5-10, RI.4, L.4, L.5	90
			Writing: Written Expression	0	0	0	0		12	1	12	W.1-2, 4, 7-10	
			Writing: Knowledge of Language and Conventions	0	0	0	0		3		3	L.1, 2**	
	Totals	3		7	14	0	0	1	19	8	33		
2	Narrative Writing Task	1	Reading: Reading Literary Text/Reading Vocabulary*	4	8	0	0	1	0	4	8	RL.1-3, 5-10 RL.4, L.4, L.5	75
			Writing: Written Expression	0	0	0	0		12	1	12	W.3, 4, 10	
			Writing: Knowledge of Language and Conventions	0	0	0	0		3		3	L.1, 2**	
	Reading Literary or Informational Texts	1	Reading: Reading Literary or Informational Text/Reading Vocabulary	0	0	6	6	0	0	6	6	RL/RI.1-10; L.4, L.5	
	Totals	2		4	8	6	6	1	15	11	29		
3	Reading Literary Texts	at least 3	Reading: Reading Literary Text/Reading Vocabulary*	0	0	at least 5	15	0	0	15	15	RL.1-3, 5-10, RL.4, L.4, L.5	75
	Reading Informational Texts		Reading: Reading Informational Text/Reading Vocabulary*	0	0	at least 5		0	0			RI.1-3, 5-10, RI.4, L.4, L.5	
	Totals			0	0	15	15	0	0	15	15		
Grade 8 Totals		**at least 8**	Reading: Reading Literary Text/Reading Vocabulary*	4	8	at least 5	21	2	0	9+	13+	47	240
			Reading: Reading Informational Text/Reading Vocabulary*	7	14	at least 5			4	12+	23+		
			Writing: Written Expression	0	0	0	0		24	2	24	30	
			Writing: Knowledge of Language and Conventions	0	0	0	0		6		6		
			Total	11	22	21	21	2	34	34	77	77	

*There must be at least eight points of reading vocabulary items on the test. **Students advancing through the grades are expected to meet each year's grade-specific standards and retain or further develop skills and understandings mastered in the preceding grades.

Copyright © 2017 by Louisiana Department of Education.

Table 3.5 Overview of LEAP Mathematics Task Types and Reporting Categories

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

Copyright © 2017 by Louisiana Department of Education.

Table 3.6 Grade 3 Mathematics Test Blueprint

Reporting Category	Category Description	Task Types						Accessible Content
		Type I		Type II		Type III		
		Tasks	Points	Tasks	Points	Tasks	Points	
Subclaim A	Major Content	29	30					Louisiana Content Standards: 3.OA.A, B, C, D 3.NF.A 3.MD.A, C
Subclaim B	Additional and Supporting Content	10	10					Louisiana Content Standards: 3.NBT.A 3.MD.B, D 3.G.A
Subclaim C	Expressing Mathematical Reasoning			3	10			PARCC Evidence Statements: 3.C.1-1, -2, -3 3.C.2 3.C.3-1, -2 3.C.4-1, -2, -3, -4, -5, -6 3.C.5-1, -2 3.C.6-1, -2
Subclaim D	Modeling and Application					3	12	PARCC Evidence Statements: 3.D.1 3.D.2
	TOTAL	**39**	**40**	**3**	**10**	**3**	**12**	
		TOTAL TASKS		**45**	**TOTAL POINTS**		**62**	

Copyright © 2017 by Louisiana Department of Education.

Table 3.7 Grade 4 Mathematics Test Blueprint

Reporting Category	Category Description	Task Types						Accessible Content
		Type I		Type II		Type III		
		Tasks	Points	Tasks	Points	Tasks	Points	
Subclaim A	Major Content	28	30					Louisiana Content Standards: 4.OA.A 4.NBT.A, B 4.NF.A, B, C
Subclaim B	Additional and Supporting Content	10	10					Louisiana Content Standards: 4.OA.B, C 4.MD.A, B, C 4.G.A
Subclaim C	Expressing Mathematical Reasoning			3	10			PARCC Evidence Statements: 4.C.1-1, -2 4.C.2 4.C.3 4.C.4-1, -2, -3, -4, -5 4.C.5-1, -2, -3, -4, -5, -6 4.C.6-1, -2, -3 4.C.7-1, -2, -3, -4
Subclaim D	Modeling and Application					3	12	PARCC Evidence Statements: 4.D.1 4.D.2
	TOTAL	**38**	**40**	**3**	**10**	**3**	**12**	
		TOTAL TASKS		**44**	**TOTAL POINTS**		**62**	

Copyright © 2017 by Louisiana Department of Education.

Table 3.8 Grade 5 Mathematics Test Blueprint

Reporting Category	Category Description	Task Types						Accessible Content
		Type I		Type II		Type III		
		Tasks	Points	Tasks	Points	Tasks	Points	
Subclaim A	Major Content	28	30					Louisiana Content Standards: 5.NBT.A, B 5.NF.A, B 5.MD.C
Subclaim B	Additional and Supporting Content	10	10					Louisiana Content Standards: 5.OA.A, B 5.MD.A, B 5.G.A, B
Subclaim C	Expressing Mathematical Reasoning			3	10			PARCC Evidence Statements: 5.C.1-1, -2, -3 5.C.2-1, -2, -3, -4 5.C.3 5.C.4-1, -2, -3, -4 5.C.5-1, -2, -3 5.C.6 5.C.7-1, -2, -3, -4 5.C.8-2
Subclaim D	Modeling and Application					3	12	PARCC Evidence Statements: 5.D.1 5.D.2
	TOTAL	**38**	**40**	**3**	**10**	**3**	**12**	
		TOTAL TASKS		**44**	**TOTAL POINTS**		**62**	

Copyright © 2017 by Louisiana Department of Education.

Table 3.9 Grade 6 Mathematics Test Blueprint

Reporting Category	Category Description	Task Types						Accessible Content
		Type I		Type II		Type III		
		Tasks	Points	Tasks	Points	Tasks	Points	
Subclaim A	Major Content	28	30					Louisiana Content Standards: 6.RP.A 6.NA.A, C 6.EE.A, B, C
Subclaim B	Additional and Supporting Content	9	10					Louisiana Content Standards: 6.NS.B 6.G.A 6.SP.A, B
Subclaim C	Expressing Mathematical Reasoning			4	14			PARCC Evidence Statements: 6.C.1-1 6.C.2 6.C.3 6.C.4 6.C.5 6.C.6 6.C.7 6.C.8-1, -2 6.C.9
Subclaim D	Modeling and Application					3	12	PARCC Evidence Statements: 6.D.1 6.D.2 6.D.3
	TOTAL	**37**	**40**	**4**	**14**	**3**	**12**	
		TOTAL TASKS		**44**	**TOTAL POINTS**		**66**	

Copyright © 2017 by Louisiana Department of Education.

Table 3.10 Grade 7 Mathematics Test Blueprint

Reporting Category	Category Description	Task Types						Accessible Content
		Type I		Type II		Type III		
		Tasks	Points	Tasks	Points	Tasks	Points	
Subclaim A	Major Content	28	30					Louisiana Content Standards: 7.RP.A 7.NS.A 7.EE.A, B
Subclaim B	Additional and Supporting Content	8	10					Louisiana Content Standards: 7.G.A, B 7.SP.A, B, C
Subclaim C	Expressing Mathematical Reasoning			4	14			PARCC Evidence Statements: 7.C.1-1, -2 7.C.2 7.C.3 7.C.4 7.C.5 7.C.6-1 7.C.7-1, -2, -3, -4 7.C.8
Subclaim D	Modeling and Application					3	12	PARCC Evidence Statements: 7.D.1 7.D.2 7.D.3 7.D.4
	TOTAL	**36**	**40**	**4**	**14**	**3**	**12**	
		TOTAL TASKS		**43**	**TOTAL POINTS**		**66**	

Copyright © 2017 by Louisiana Department of Education.

Table 3.11 Grade 8 Mathematics Test Blueprint

Reporting Category	Category Description	Task Types						Accessible Content
		Type I		Type II		Type III		
		Tasks	Points	Tasks	Points	Tasks	Points	
Subclaim A	Major Content	28	30					Louisiana Content Standards: 8.EE.A, B, C 8.F.A 8.G.A, B
Subclaim B	Additional and Supporting Content	10	10					Louisiana Content Standards: 8.NS.A 8.F.B 8.G.C 8.SP.A
Subclaim C	Expressing Mathematical Reasoning			4	14			PARCC Evidence Statements: 8.C.1-1, -2 8.C.2 8.C.3-1, -2, -3 8.C.4-1 8.C.5-1, -2, -3 8.C.6
Subclaim D	Modeling and Application					3	12	PARCC Evidence Statements: 8.D.1 8.D.2 8.D.3 8.D.4
	TOTAL	**38**	**40**	**4**	**14**	**3**	**12**	
		TOTAL TASKS		**45**	**TOTAL POINTS**		**66**	

Copyright © 2017 by Louisiana Department of Education.

Table 3.12 PARCC Items Used in LEAP 2016 ELA Forms

Grade	Items	Percent Total Items
3	13	48.15
4	13	38.24
5	14	48.28
6	14	48.28
7	14	48.28
8	13	38.24

Table 3.13 DRC Items Used in LEAP 2016 ELA Forms

Grade	Items	Percent Total Items
3	14	51.85
4	21	61.76
5	15	51.72
6	15	51.72
7	15	51.72
8	21	61.76

Table 3.14 PARCC Items Used in LEAP 2016 Math Forms

Grade	Items	Percent Total Items
3	22	48.89
4	21	47.73
5	21	47.73
6	21	47.73
7	21	48.84
8	21	46.67

Table 3.15 DRC Items Used in LEAP 2016 Math Forms

Grade	Items	Percent Total Items
3	23	51.11
4	23	52.27
5	23	52.27
6	23	52.27
7	22	51.16
8	24	53.33

Copyright © 2017 by Louisiana Department of Education.

Table 3.16 General Mathematics Test Structure—Grade 3

Reporting Category	Test Session						TOTAL	
	Session 1 No Calculator		Session 2 No Calculator		Session 3 No Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Subclaim A	10	11	8	8	11	11	29	30
Subclaim B	3	3	6	6	1	1	10	10
Subclaim C	1	4	1	3	1	3	3	10
Subclaim D	1	3	1	3	1	6	3	12
TOTAL	**15**	**21**	**16**	**20**	**14**	**21**	**45**	**62**
Test Duration (minutes)	**75**		**75**		**75**		**225**	

Table 3.17 General Mathematics Test Structure—Grade 4

Reporting Category	Test Session						TOTAL	
	Session 1 No Calculator		Session 2 No Calculator		Session 3 No Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Subclaim A	10	10	9	10	9	10	28	30
Subclaim B	4	4	4	4	2	2	10	10
Subclaim C	1	4	1	3	1	3	3	10
Subclaim D	1	3	1	3	1	6	3	12
TOTAL	**16**	**21**	**15**	**20**	**13**	**21**	**44**	**62**
Test Duration (minutes)	**75**		**75**		**75**		**225**	

Table 3.18 General Mathematics Test Structure—Grade 5

Reporting Category	Test Session						TOTAL	
	Session 1 No Calculator		Session 2 No Calculator		Session 3 No Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Subclaim A	11	12	10	10	7	8	28	30
Subclaim B	2	2	4	4	4	4	10	10
Subclaim C	1	4	1	3	1	3	3	10
Subclaim D	1	3	1	3	1	6	3	12
TOTAL	**15**	**21**	**16**	**20**	**13**	**21**	**44**	**62**
Test Duration (minutes)	**75**		**75**		**75**		**225**	

Copyright © 2017 by Louisiana Department of Education.

Table 3.19 General Mathematics Test Structure—Grade 6

Reporting Category	Test Session						TOTAL	
	Session 1 No Calculator		Session 2 Calculator		Session 3 Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Subclaim A	14	14	7	9	7	7	28	30
Subclaim B	6	6	1	1	2	3	9	10
Subclaim C			2	7	2	7	4	14
Subclaim D			2	6	1	6	3	12
TOTAL	**20**	**20**	**12**	**23**	**12**	**23**	**44**	**66**
Test Duration (minutes)	**75**		**75**		**75**		**225**	

Table 3.20 General Mathematics Test Structure—Grade 7

Reporting Category	Test Session						TOTAL	
	Session 1 No Calculator		Session 2 Calculator		Session 3 Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Subclaim A	20	20	4	5	4	5	28	30
Subclaim B			4	5	4	5	8	10
Subclaim C			2	7	2	7	4	14
Subclaim D			2	6	1	6	3	12
TOTAL	**20**	**20**	**12**	**23**	**11**	**23**	**43**	**66**
Test Duration (minutes)	**75**		**75**		**75**		**225**	

Table 3.21 General Mathematics Test Structure—Grade 8

Reporting Category	Test Session						TOTAL	
	Session 1 No Calculator		Session 2 Calculator		Session 3 Calculator			
	Tasks	Points	Tasks	Points	Tasks	Points	Tasks	Points
Subclaim A	13	14	9	9	6	7	28	30
Subclaim B	6	6	1	1	3	3	10	10
Subclaim C			2	7	2	7	4	14
Subclaim D			2	6	1	6	3	12
TOTAL	**19**	**20**	**14**	**23**	**12**	**23**	**45**	**66**
Test Duration (minutes)	**75**		**75**		**75**		**225**	

Copyright © 2017 by Louisiana Department of Education.

Table 3.22 Elements of Universal Design

Element	Explanation
Inclusive Assessment Population	Tests designed for state, district, 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 brailled).
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.

Table 3.23 Distribution of Points by Subclaim—ELA

	Grade					
Subclaims	**3**	**4**	**5**	**6**	**7**	**8**
Reading - RI	15	22	19	20	18	23
Reading - RL	19	17	20	19	18	16
Reading - RV	8	8	8	8	10	8
Writing - WE	18	21	24	24	24	24
Writing - WKL	6	6	6	6	6	6
Total Points	**66**	**74**	**77**	**77**	**76**	**77**

Table 3.24 Distribution of Points by Claim—ELA

	Grade					
Claims	**3**	**4**	**5**	**6**	**7**	**8**
Reading	42	47	47	47	46	47
Writing	24	27	30	30	30	30
Total Points	**66**	**74**	**77**	**77**	**76**	**77**

Copyright © 2017 by Louisiana Department of Education.

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

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

These reporting categories are the same as the reporting categories on the Spring 2015 ELA student reports and provide parents and educators 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 districts are helping students achieve higher expectations.

Copyright © 2017 by Louisiana Department of Education.

Table 3.25 Distribution of ELA Items and Points by Session and Item Type

	Sub	Gr	Session	EBSR		MC		MS		PCR		Total Pts
				Items	Pts	Items	Pts	Items	Pts	Items	Pts	
Paper -Pencil	ELA	3	1. Research Simulation	6	12					1	15	66
			2. Literary Analysis	5	10					1	15	
			3. Reading Literary & Informational Texts	4	4	7	7	3	3			
	ELA	4	1. Research Simulation	7	14					1	19	74
			2. Narrative Writing & Reading Texts	6	10	3	3	1	1	1	12	
			3. Reading Literary & Informational Texts	3	3	9	9	3	3			
	ELA	5	1. Research Simulation	7	14					1	19	77
			2. Literary Analysis	5	10					1	19	
			3. Reading Literary & Informational Texts	4	4	9	9	2	2			
	ELA	6	1. Research Simulation	7	14					1	19	77
			2. Literary Analysis	5	10					1	19	
			3. Reading Literary & Informational Texts	3	3	11	11	1	1			
	ELA	7	1. Research Simulation	7	14					1	19	77
			2. Literary Analysis	5	10					1	19	
			3. Reading Literary & Informational Texts	3	3	10	10	2	2			
	ELA	8	1. Research Simulation	7	14					1	19	77
			2. Narrative Writing & Reading Texts	5	9	4	4	1	1	1	15	
			3. Reading Literary & Informational Texts	2	2	11	11	2	2			
Online	ELA	4	1. Research Simulation	7	14					1	19	74
			2. Narrative Writing & Reading Texts	6	10	2	2	1	1	1	12	
			3. Reading Literary & Informational Texts	2	2	8	8	3	3			
	ELA	5	1. Research Simulation	7	14					1	19	77
			2. Literary Analysis	5	10					1	19	
			3. Reading Literary & Informational Texts	3	3	8	8	2	2			
	ELA	6	1. Research Simulation	7	14					1	19	77
			2. Literary Analysis	5	10					1	19	
			3. Reading Literary & Informational Texts	3	3	9	9	1	1			
	ELA	7	1. Research Simulation	7	14					1	19	77
			2. Literary Analysis	5	10					1	19	
			3. Reading Literary & Informational Texts	3	3	8	8	2	2			
	ELA	8	1. Research Simulation	7	14					1	19	77
			2. Narrative Writing & Reading Texts	5	9	4	4	1	1	1	15	
			3. Reading Literary & Informational Texts	2	2	8	8	2	2			

Copyright © 2017 by Louisiana Department of Education.

Table 3.26 Distribution of Points by Subclaim—Mathematics

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

* one grade 7 Sublaim B item was omitted during scoring

Each item on the LEAP assessment is referred to as a task and is identified by one of three types: Type I, Type II, and Type III. As shown in the table below, each of the three task types is aligned to one of four reporting categories (also called subclaims): Major Content (subclaim A), Additional and Supporting Content (subclaim B), Expressing Mathematical Reasoning (subclaim C), and Modeling and Application (subclaim D). Each task type is designed to align with at least one of the Standards for Mathematical Practice (MP).

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

These reporting categories are the same as the reporting categories on the Spring 2015 mathematics student reports and will provide parents and educators 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 districts are helping students achieve higher expectations.

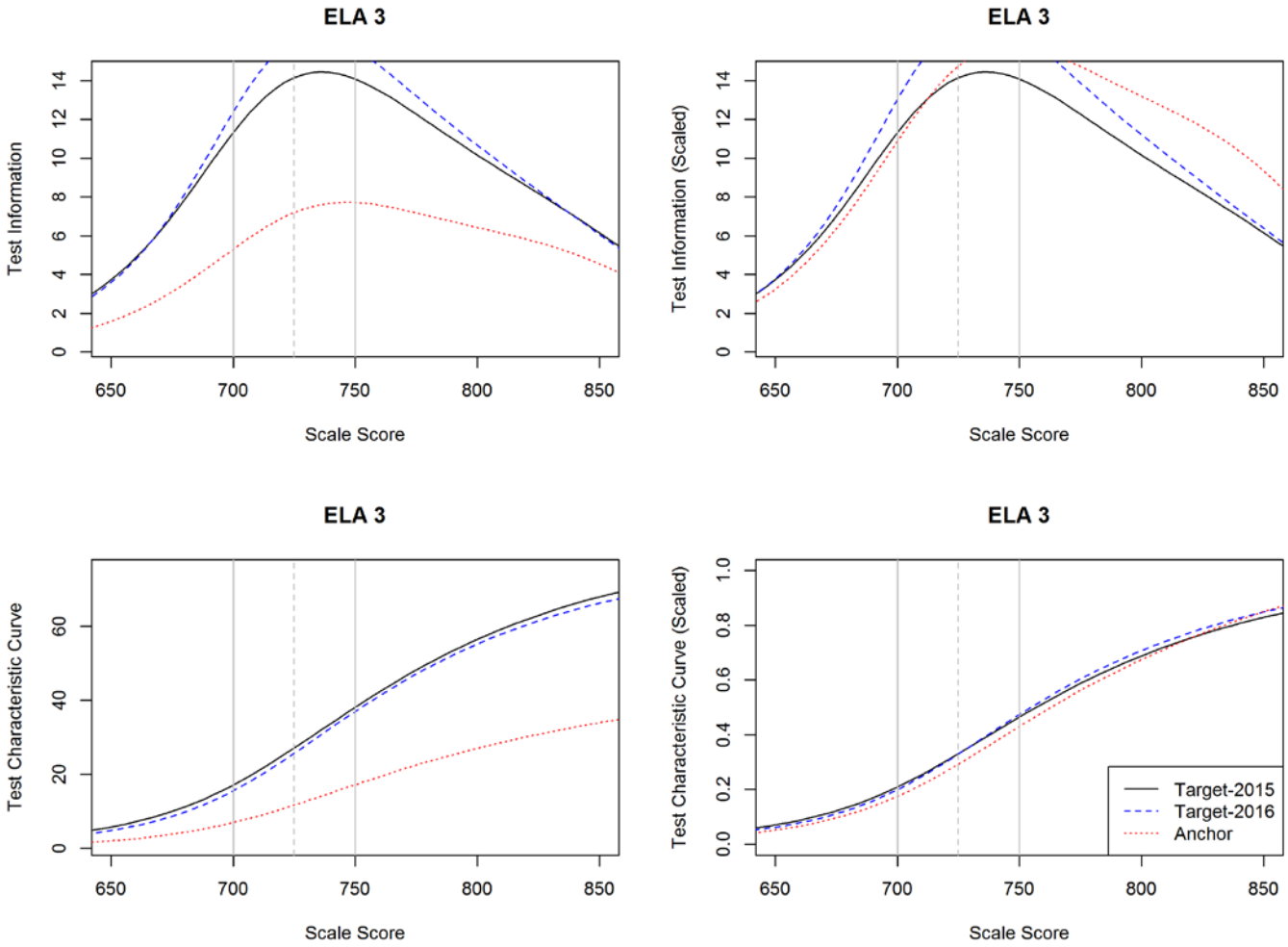
Copyright © 2017 by Louisiana Department of Education.

Table 3.27 Distribution of Mathematics Tasks and Points by Task Type

	Subject	Grade	Type I			Type II			Type III			Total Points
			(1 pt) Tasks	(2 pt) Tasks	Points	(3 pt) Tasks	(4 pt) Tasks	Points	(3 pt) Tasks	(6 pt) Tasks	Points	
Paper-Pencil	Math	3	38	1	40	2	1	10	2	1	12	62
	Math	4	36	2	40	2	1	10	2	1	12	62
	Math	5	36	2	40	2	1	10	2	1	12	62
	Math	6	34	3	40	2	2	14	2	1	12	66
	Math	7	32	4	40	2	2	14	2	1	12	66
	Math	8	36	2	40	2	2	14	2	1	12	66
Online	Math	4	36	2	40	2	1	10	2	1	12	62
	Math	5	36	2	40	2	1	10	2	1	12	62
	Math	6	34	3	40	2	2	14	2	1	12	66
	Math	7	32	4	40	2	2	14	2	1	12	66
	Math	8	36	2	40	2	2	14	2	1	12	66

Copyright © 2017 by Louisiana Department of Education.

Figure 3.1 2016 ELA Anchor Evaluation—Grade 3

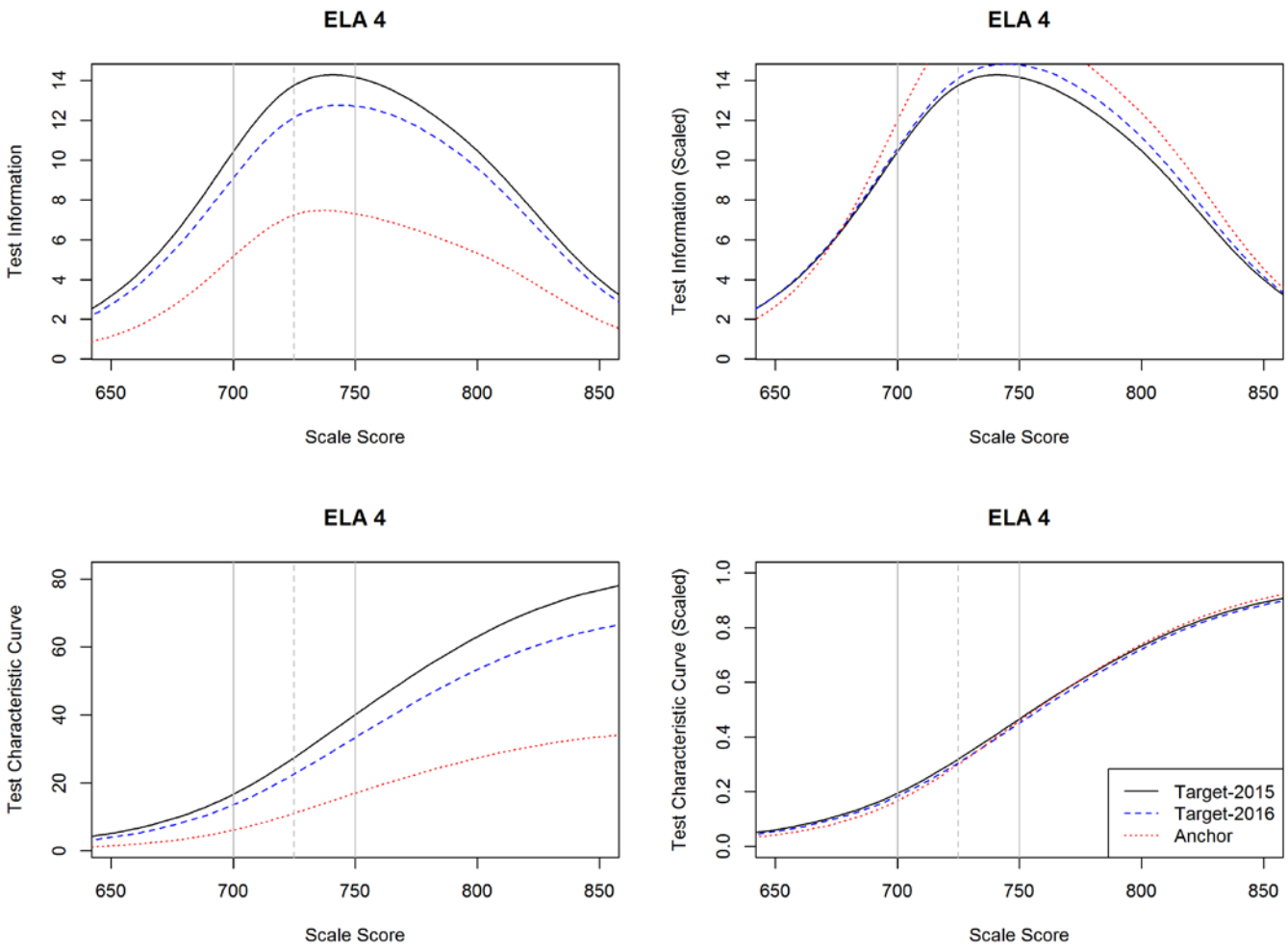


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.2 2016 ELA Anchor Evaluation—Grade 4

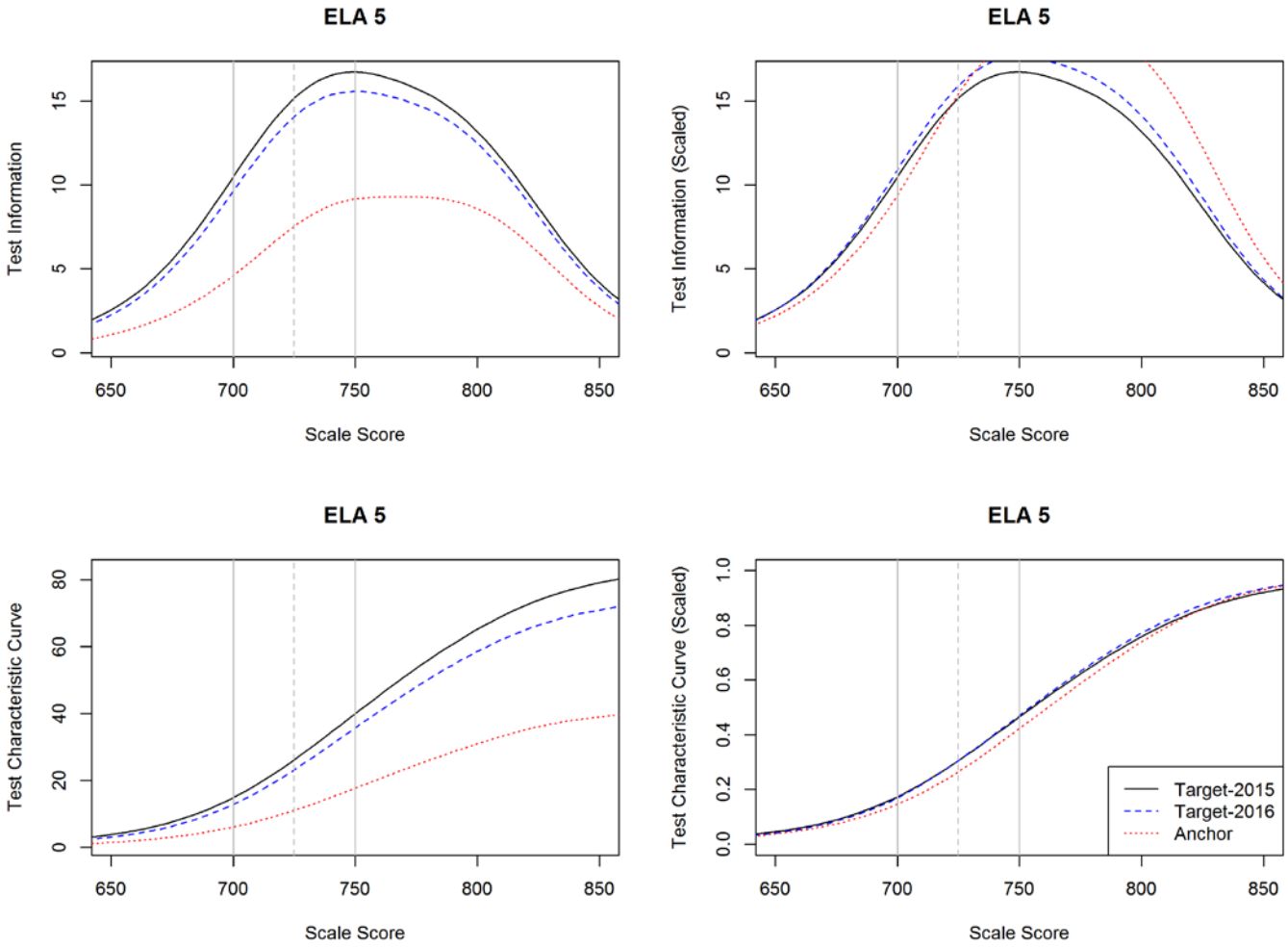


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.3 2016 ELA Anchor Evaluation—Grade 5

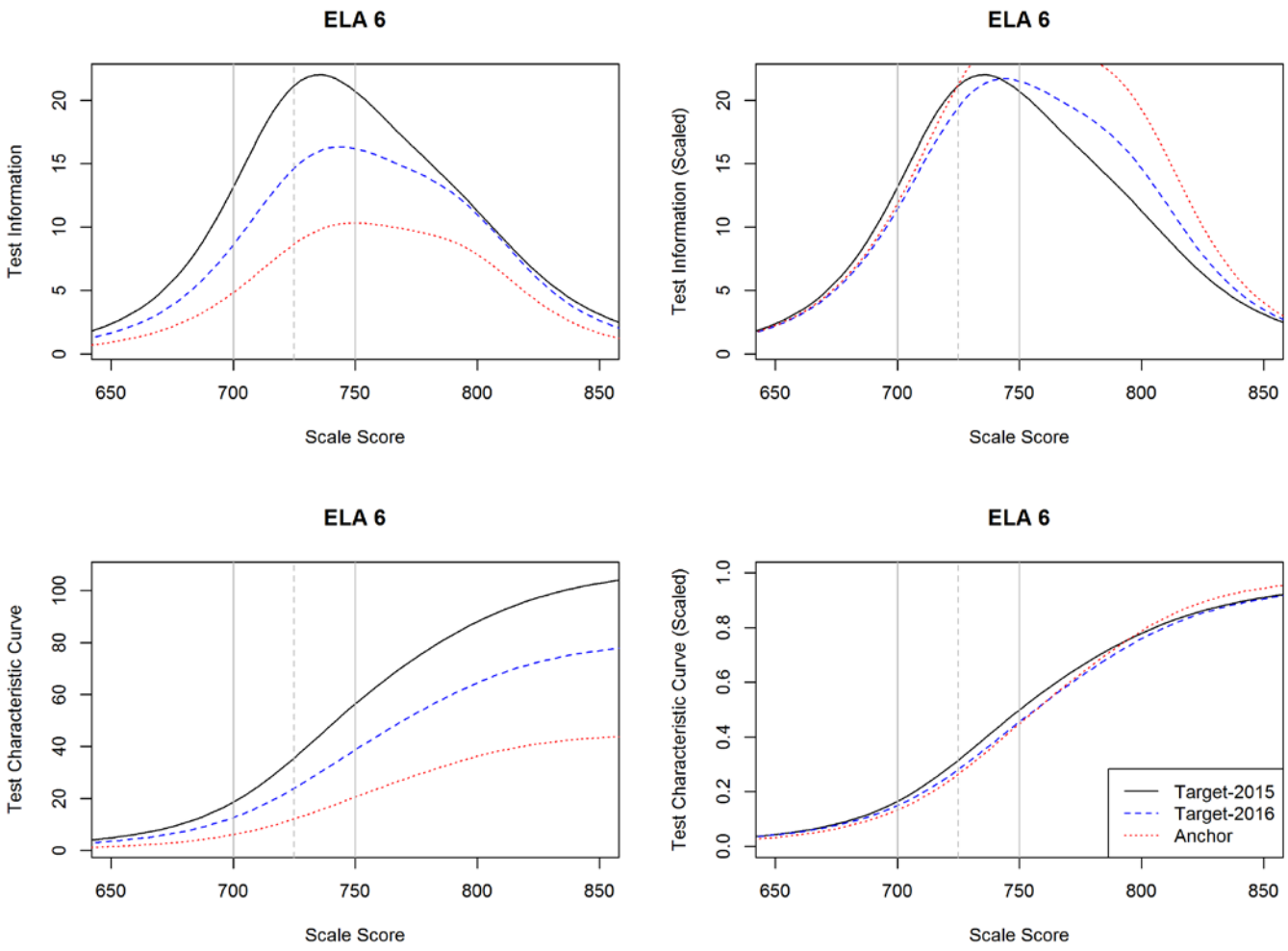


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.4 2016 ELA Anchor Evaluation—Grade 6

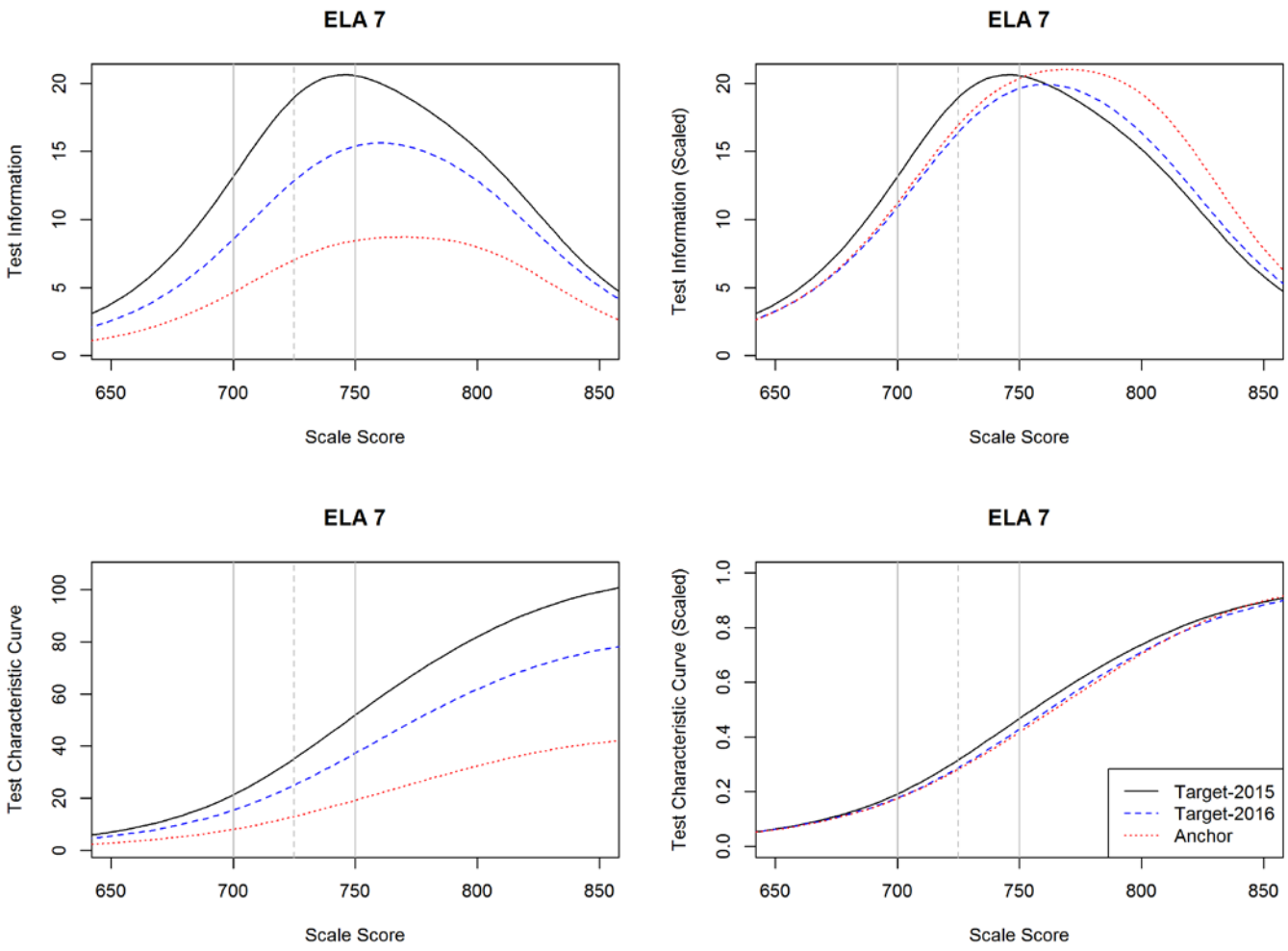


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.5 2016 ELA Anchor Evaluation—Grade 7

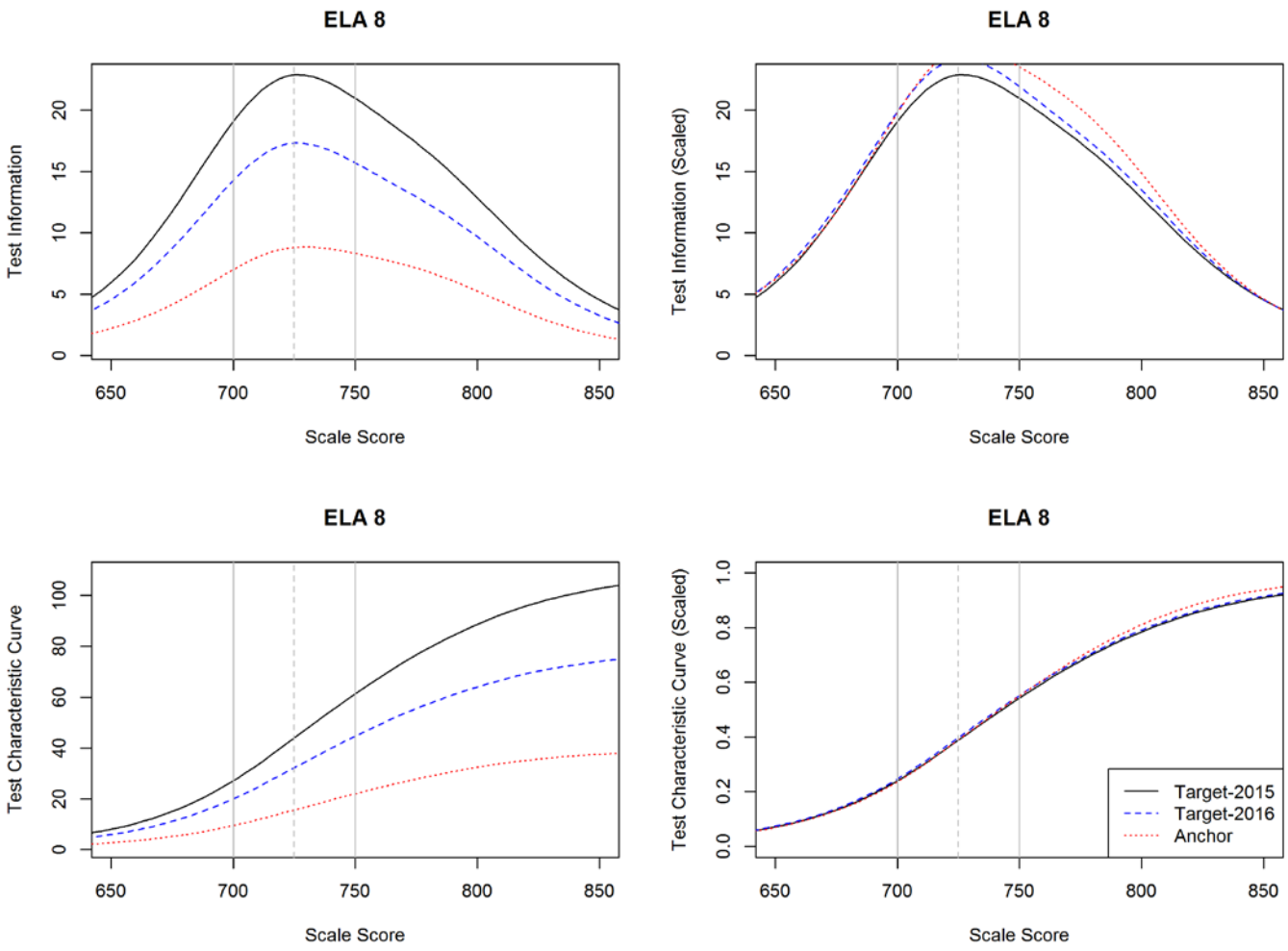


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.6 2016 ELA Anchor Evaluation—Grade 8

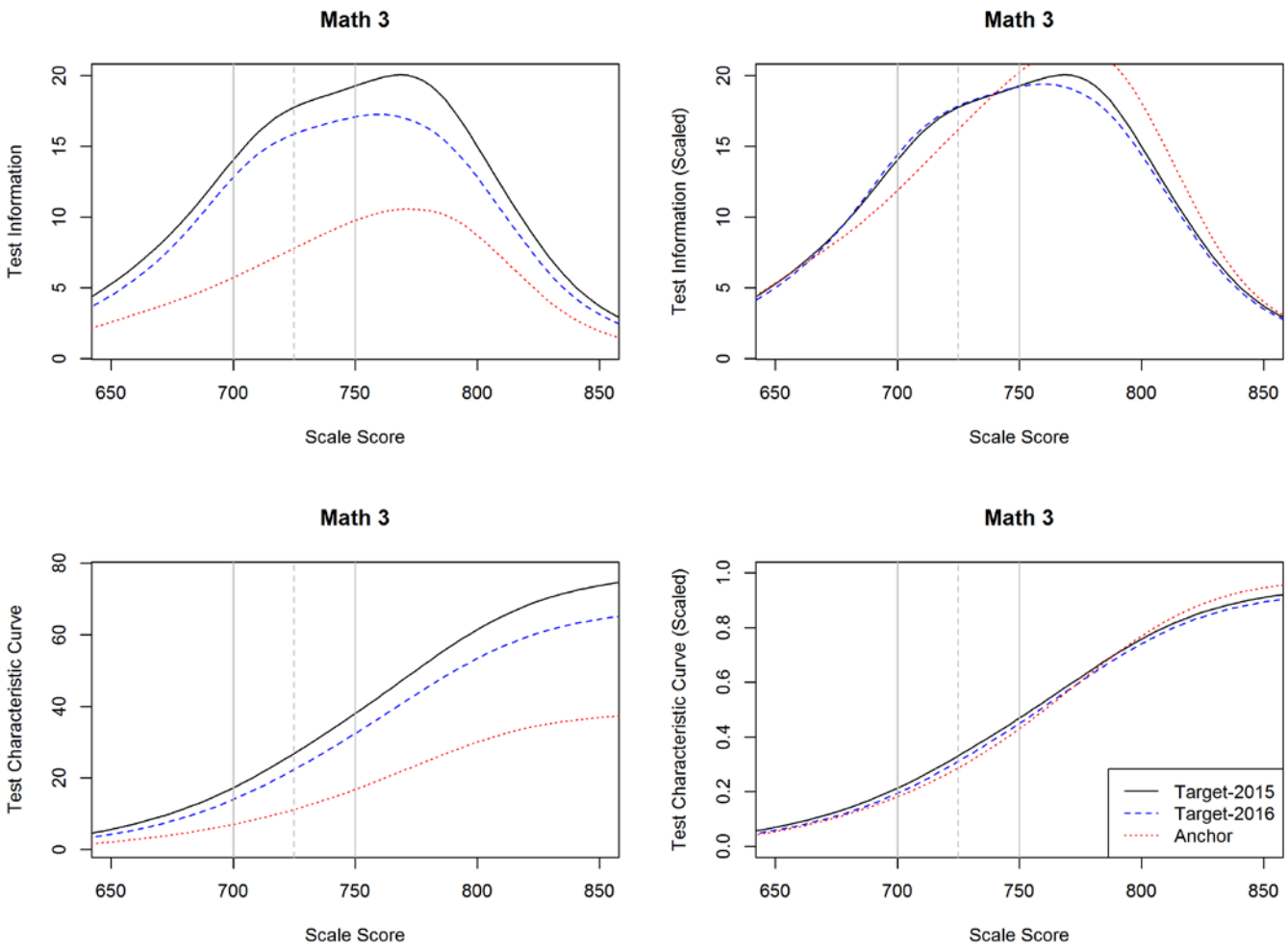


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.7 2016 Mathematics Anchor Evaluation—Grade 3

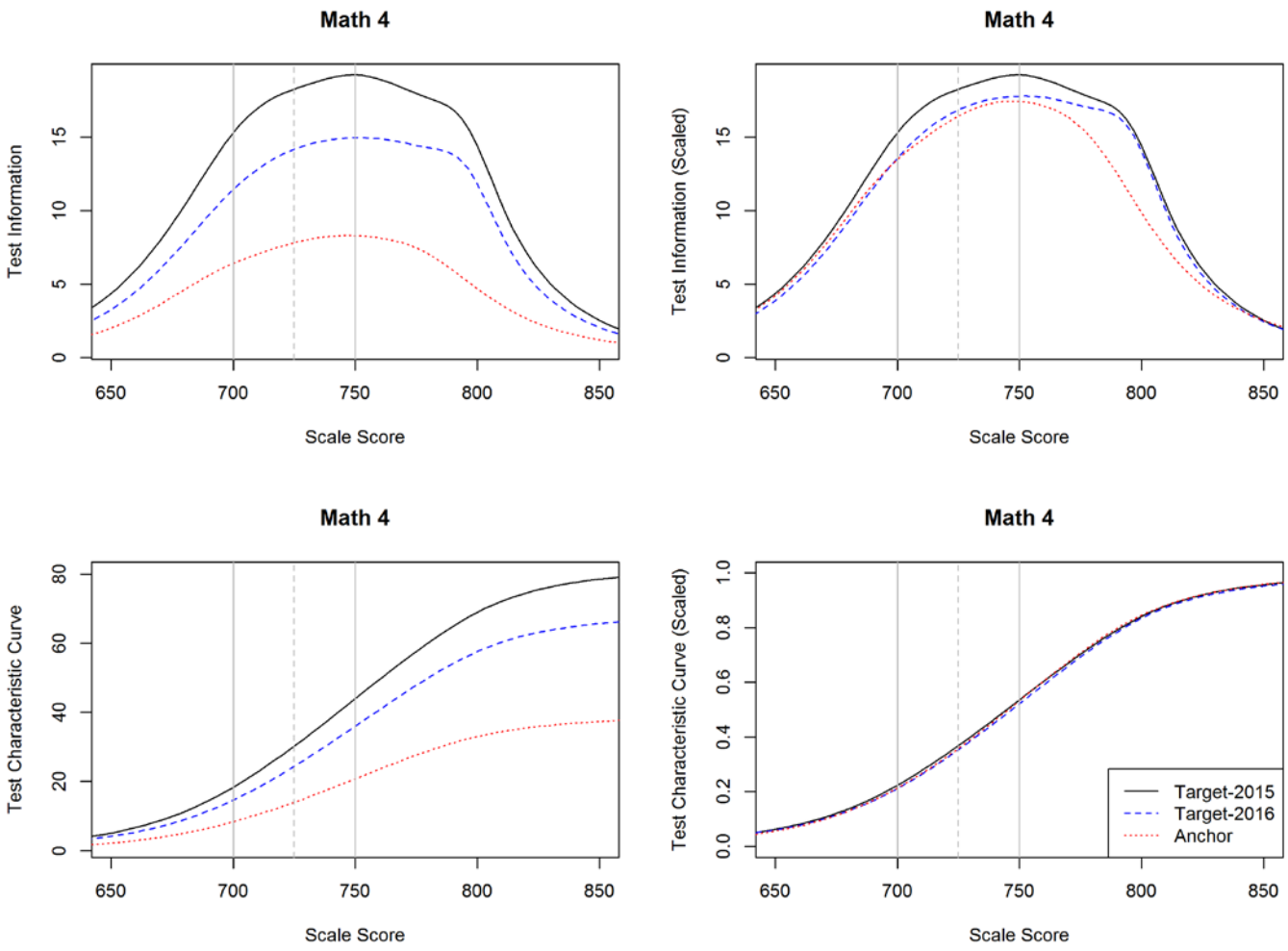


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.8 2016 Mathematics Anchor Evaluation—Grade 4

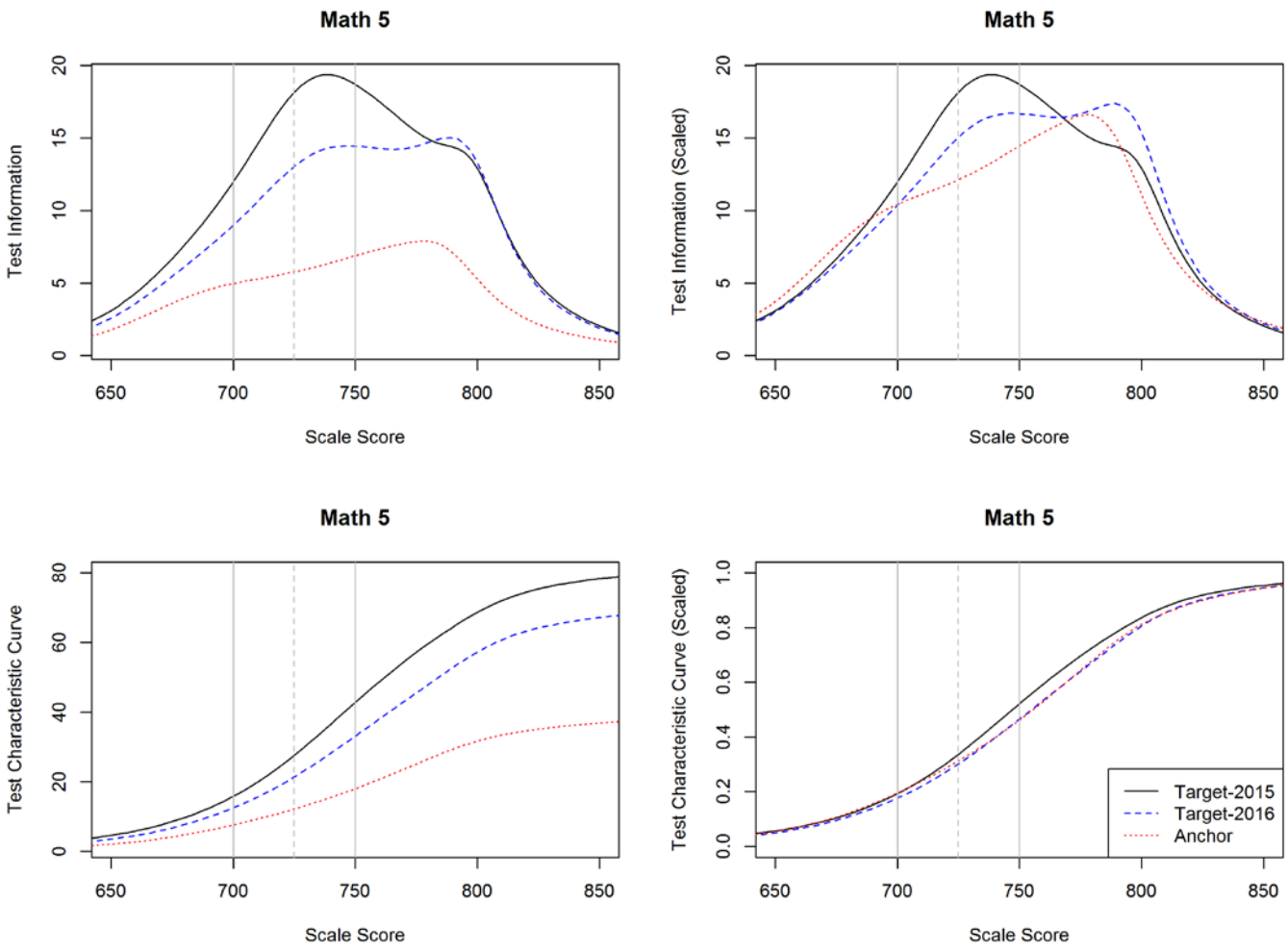


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.9 2016 Mathematics Anchor Evaluation—Grade 5

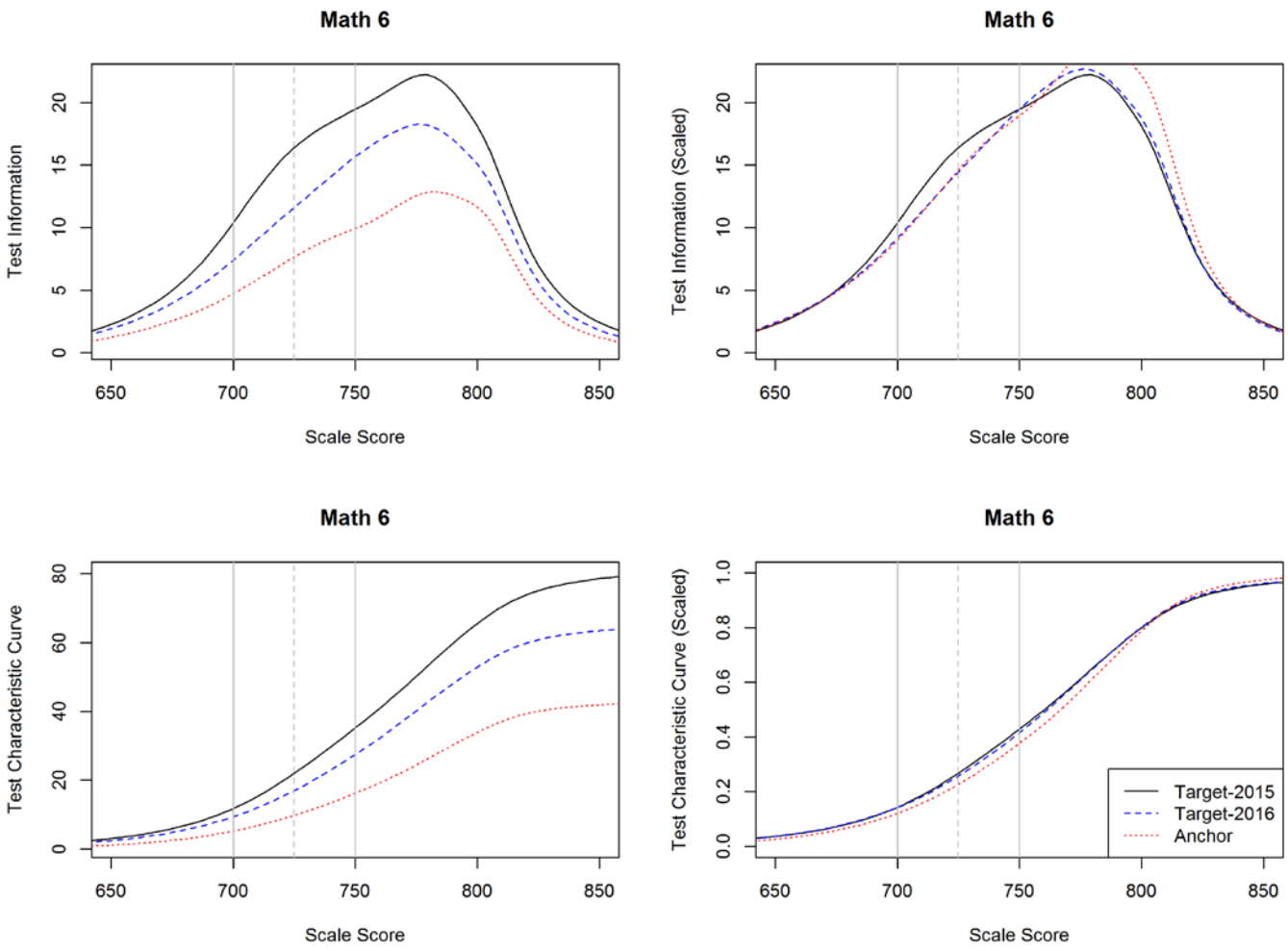


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.10 2016 Mathematics Anchor Evaluation—Grade 6

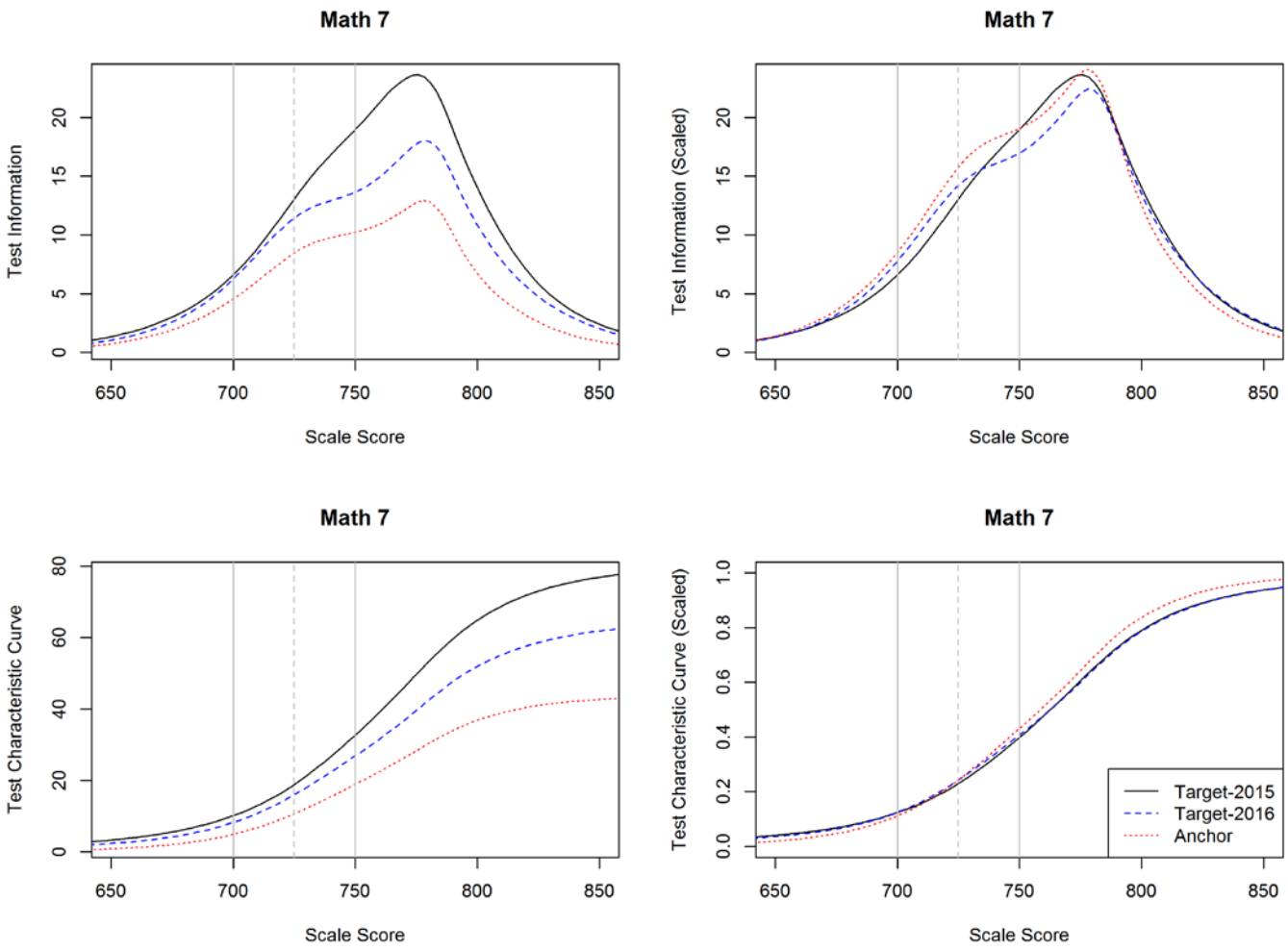


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.11 2016 Mathematics Anchor Evaluation—Grade 7

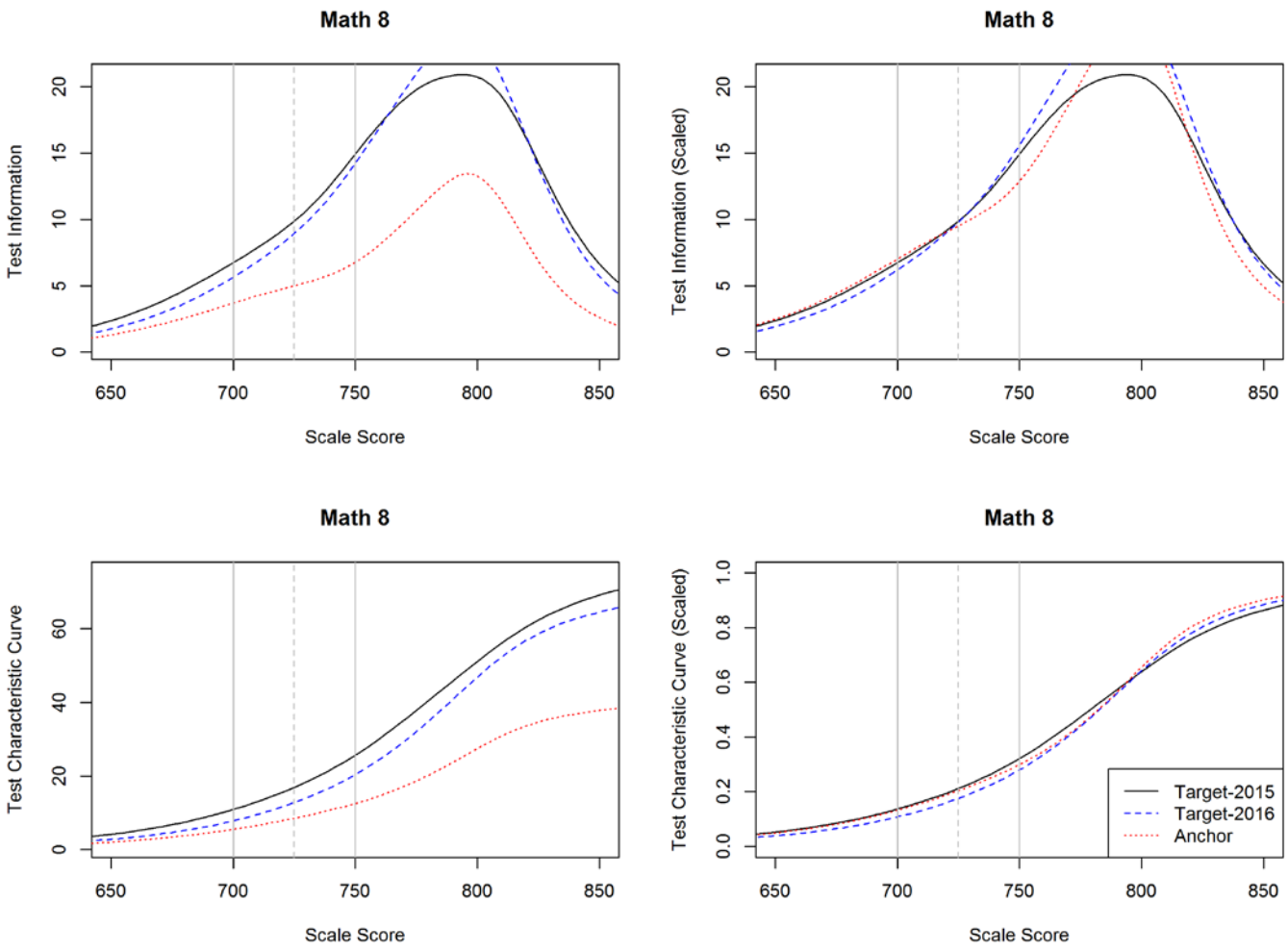


NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

Figure 3.12 2016 Mathematics Anchor Evaluation—Grade 8



NOTE:

- Target-2015 is the 2015 PARCC intact test form.
- Target-2016 is the 2016 item pool from PARCC.
- Anchor is the subset of Target-2016 selected for the 2016 LEAP test form.

Copyright © 2017 by Louisiana Department of Education.

CHAPTER 4: TEST ADMINISTRATION

Chapter 4 of the technical report describes processes and activities implemented and information disseminated to help ensure standardized test administration procedures and, thus, uniform test administration conditions for students. According to the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (2014) *Standards for Educational and Psychological Testing*, "The usefulness and interpretability of test scores require that a test be administered and scored according to the developer's instructions" (111). This chapter examines how test administration procedures implemented for the Louisiana Education Assessment Program (LEAP) 2016 strengthen and support the intended score interpretations and reduce construct-irrelevant variance that could threaten the validity of score interpretations.

Chapter 4 demonstrates adherence to AERA, APA, and NCME (2014) Standards 4.15, 4.16, 6.1, 6.2, 6.3, 6.4, 6.6, and 6.7. Each standard will be explicated within the relevant section of this chapter.

4.1 Training of Districts

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

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

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

4.2 Ancillary Materials

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

For the spring 2016 test administration, DRC produced four administration manuals:

Copyright © 2017 by Louisiana Department of Education.

1. LEAP/*i* LEAP *Grades 3–5 Paper-Based Test Administration Manual*
2. LEAP/*i* LEAP *Grades 6–8 Paper-Based Test Administration Manual*
3. *LEAP Computer-Based Test Administration Manual Grades 3-5*
4. *LEAP Computer-Based Test Administration Manual Grades 6-8*

DRC also produced a district *Test Coordinators Manual*. LDOE curriculum and assessment staff review, provide feedback, and give final approval for these manuals. The *Test Coordinators Manual* is inclusive of grades 3–8 ELA/mathematics/Science (paper). It provides detailed instructions for district and school test coordinators' responsibilities for distributing and collecting test materials for the following programs and for returning them to DRC for scoring.

Test Coordinators Manual Table of Contents

1. Key Dates Spring 2016
2. Spring 2016 Alerts
3. Oath of Security and Confidentiality Statement
4. General Information
5. Test Security
   - 5.1. Key Definitions
   - 5.2. Violations of Test Security
   - 5.3. Erasure Analysis
   - 5.4. Voiding Student Tests
6. Testing Guidelines
   - 6.1. Testing Eligibility
   - 6.2. Testing Conditions
   - 6.3. Testing in Class-sized Groups
   - 6.4. Test Schedule
   - 6.5. Extended Time for Testing
   - 6.6. Extended Breaks
   - 6.7. Makeup Testing
   - 6.8. Test Administration Resources
7. District Test Coordinator
   - 7.1. Conduct Training Session
   - 7.2. Receive Test Materials
   - 7.3. Large-print and Braille Test Materials
   - 7.4. Accommodated Materials
   - 7.5. Verify and Distribute Test Materials to School Test Coordinators
   - 7.6. Request Additional Test Materials and Bar-code Labels
   - 7.7. Collect Materials from Schools After Testing
   - 7.8. Used and Unused Answer Documents and Consumable Test Booklets (Defined)
   - 7.9. Unscorable Documents and Unscorable Document Labels
8. Directions for Returning Test Materials to DRC in May
   - 8.1. Pickup 1, 2, 3: Science and ELA/Math
   - 8.2. Final Checklist for Returning Test Materials to DRC
9. School Test Coordinator
   - 9.1. Receive and Verify Test Materials
   - 9.2. Conduct Test Administration and Security Training Session
   - 9.3. Supervise Application of Bar-code Labels and Coding of Answer Documents and Consumable Test Booklets

Copyright © 2017 by Louisiana Department of Education.

9.4. Soiled, Damaged, and Other Unscorable Answer Documents and Consumable Test Booklets
9.5. Verify and Distribute Materials to Test Administrators
9.6. Supervise Test Administration
9.7. Collect Test Materials
9.8. Used and Unused Answer Documents and Consumable Test Booklets (Defined)
9.9. Coding Responsibilities of Principals—Before Testing
9.10. Coding Responsibilities of Principals—After Testing
10. Directions for Returning Test Materials to the DTC
10.1. Pickup 1, 2, 3: Science and ELA/Math
11. Void Notification
12. Index

The test administration manuals are specific to grades and content areas—grades 3–5 or 6–8 online or paper and content areas ELA/Math/Science (paper) or ELA/Math (online). They provide detailed instructions for administering the LEAP assessments. The manuals include instructions for test security, test administrator responsibilities, test preparation, administration of tests (online or paper), and post-test procedures. Information included in the test administration manuals is listed below.

Paper Administration Table of Contents

1. Spring 2016 Notes and Reminders
2. Test Administrator Oath of Security and Confidentiality Statement
3. Overview
4. Test Security
4.1. Key Definitions
4.2. Violations of Test Security
4.3. Erasure Analysis
4.4. Reporting Testing Irregularities and Security Breaches
4.5. Voiding Student Tests
5. Test Administrator Responsibilities
6. Test Administration Checklists
6.1. Before Testing
6.2. During Testing
6.3. After Testing (Daily)
6.4. After Testing (Last Day)
7. Test Administrators' Frequently Asked Questions
8. Test Materials
8.1. Receipt of Test Materials
9. Testing Guidelines
9.1. Testing Eligibility
9.2. Test Schedule
9.3. Extended Time for Testing
9.4. Testing Times for Grades X–X
9.5. Makeup Testing
9.6. Testing Conditions
10. Special Populations and Accommodations

Copyright © 2017 by Louisiana Department of Education.

10.1. IDEA Special Education Students
10.2. Students with One or More Disabilities According to Section 504
10.3. Gifted and Talented Special Education Students
10.4. Test Accommodations for Special Education and Section 504 Students
10.5. Special Considerations for Deaf and Hard of Hearing Students
10.6. Limited English Proficient (LEP) Students
11. Hand-coded Consumable Test Booklets and Answer Documents
12. Students Absent from Testing
13. Consumable Test Booklet and Answer Document Coding
13.1. Coding the Demographic Section
13.2. Sample Grade X English Language Arts Consumable Test Booklet
13.3. Sample Grade X Science Answer Document
14. General Instructions for English Language Arts and mathematics
14.1. Student Marking/Erasing on Consumable Test Booklet
14.2. Reading Directions to Students
14.3. Special Instructions
15. Directions for Administering LEAP: English Language Arts and mathematics
16. General Instructions for Science
16.1. Student Marking/Erasing on Answer Document
16.2. Reading Directions to Students
16.3. Special Instructions
17. Directions for Administering LEAP/i LEAP: Science
18. Post-test Procedures
18.1. Test Administrator Oath of Security and Confidentiality Statement
18.2. Science Test Booklets
18.3. Used and Unused Answer Documents and Consumable Test Booklets (Defined)
18.4. Transferring Student Responses
18.5. Returning Test Materials to the School Test Coordinator
19. Index

Online Administration Table of Contents

1. Spring 2016 Notes and Reminders
2. Test Administrator Oath of Security and Confidentiality Statement
3. Overview
4. Test Security
4.1. Key Definitions
4.2. Violations of Test Security
4.3. Reporting Testing Irregularities and Security Breaches
4.4. Voiding Student Tests
5. Test Administrator Responsibilities
6. Test Administration Checklists
6.1. Before Testing
6.2. During Testing
6.3. After Testing (Daily)
6.4. After Testing (Last Day)
7. Test Administrators' Frequently Asked Questions
8. Test Materials
8.1. Receipt of Test Materials

Copyright © 2017 by Louisiana Department of Education.

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

The *Standards* contain multiple references relevant to test administration. Information in the LEAP test administration manuals addresses these in the following manner.

Directions for test administration found in the manual addresses Standard 4.15, which states

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

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

Furthermore, the standardized procedures addressed in the test administration manual need to be followed, as the *Standards* state in Standard 6.1, “ Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instructions from the test user” (114). To ensure the usefulness and interpretability of test scores and to minimize sources of construct-irrelevant variance, it was essential that the LEAP was administered according to the prescribed test administration manual. It should be noted that adhering to the test schedule is also a critical component. The test administration manuals included instructions for scheduling the test within the state testing window of April 11–May 6, 2016. The test administration manual also contained the schedule for timing each test session. The test timing schedule is presented in Table 4.1.

Copyright © 2017 by Louisiana Department of Education.

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

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

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

Test administration manuals outline the steps that teachers should take to prepare classroom environment testing for administering the LEAP online test. These 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 or removed or are out of the students' view.

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

The test administration manuals present instructions for post-test activities to ensure that online tests are submitted and printed test materials are handled properly to maintain the integrity of student information and test scores. Detailed instructions guide test examiners in submitting all online test records. For students who were administered a large print or braille version of the LEAP, examiners are instructed to transcribe students' responses from the large print test or Braille test book into the online testing system (INSIGHT) exactly as they responded in the large print or Braille test book.

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

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

Copyright © 2017 by Louisiana Department of Education.

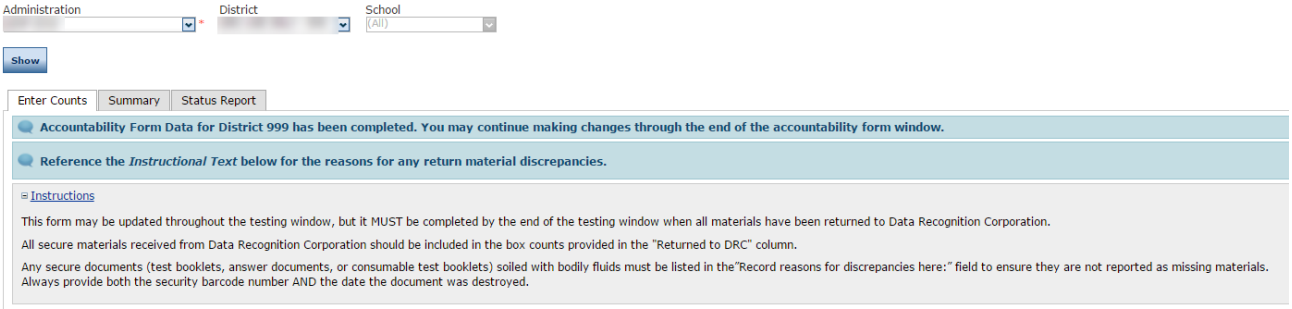
4.2.1 Return Material Forms and Guidelines

The *Test Coordinators Manual* instructs test coordinators regarding procedures for organizing and packing materials and returning them to DRC for secure inventory purposes. LDOE curriculum and assessment staff have opportunities to review, provide feedback, and give final approval. The purpose of the instructions is to ensure that secure test materials are properly accounted for and organized appropriately for return shipment.

4.2.2 Security Checklists

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

Figure 4.1 Sample Accountability Form



Copyright © 2017 by Louisiana Department of Education.

Accountability Form for [redacted]

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

Accountability Form for [redacted]

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

Record reasons for discrepancies here:

Enter Counts | Summary | Status Report

Instructions

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

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

Copyright © 2017 by Louisiana Department of Education.

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

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

Record reasons for discrepancies here:

Print

Enter Counts | Summary | Status Report

Instructions

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

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

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

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

Overall Status for District	
District	Status
	Completed

Export to Excel

Copyright © 2017 by Louisiana Department of Education.

4.2.3 Interpretive Guides

Essential to making valid interpretations of test scores is an understanding of what the test scores mean and how to interpret score reports. The *Interpretive Guide* is written for Louisiana teachers and administrators who receive the LEAP score reports from the 2016 administration. More detail about the guide can be found in Chapter 7.

4.3 Test Security Measures

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

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

4.4 Test Administration

The 2016 test was administered to students within the state testing window of April 11–May 6, 2016. The paper testing window was May 2–6, 2016. Each session within each content area of the LEAP was required to be administered in one block of time.

4.4.1 Time

Each section of each content area test was timed to provide sufficient time for students to attempt all items. The test administration manuals provided examiners with timing guidelines for the assessments. For the LEAP's custom sessions, test administrators were instructed to allow students to complete the assessment if they were making adequate progress. The timing schedule of the LEAP is presented in Table 4.1.

Copyright © 2017 by Louisiana Department of Education.

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

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

4.4.2 Accommodations

Accommodations are allowed on the LEAP. Accommodations have been split into three areas: Universal Tools, Designated Supports, and Accommodations.

- Universal Tools are available to all students taking an assessment.
- Designated Supports are available to students when deemed appropriate by a team of educators.
- Accommodations must appear in a student's IEP/504 Plan.

Accommodations may be used with students who qualify under Individual with Disabilities Act (IDEA) and have an IEP or Section 504 of the Americans with Disabilities Act and have a Section 504 plan, or who are identified as an English Language Learner (ELL) students. 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, and 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 *Test Administration Manual* contains the list of Universal Tools, Designated Supports, and Accommodations permissible for the LEAP assessments.

Copyright © 2017 by Louisiana Department of Education.

Braille and large print forms are provided for students who are visually challenged.

Tables 4.2 to 4.5 summarize the numbers of reportable students receiving accommodations by accommodation type for the 2016 LEAP. Accommodations are grouped into four sections: special education accommodation, Limited English Proficient (LEP) status accommodation, 504 status accommodation, and online accommodation. The analyses are based on census data and include only those students who received accommodations and received a scaled score on the ELA or mathematics LEAP.

Copyright © 2017 by Louisiana Department of Education.

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

Special Education Accommodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Frequency	Percent	Frequency	Percent
03	No Accommodation	≥2500	4.38%	≥2490	4.36%
03	Braille	<50	NR	<50	NR
03	Large Print	<50	NR	<50	NR
03	Answers Recorded	≥540	0.96%	≥530	0.94%
03	Extended Time	≥4650	8.14%	≥4620	8.09%
03	Transferred Answers	≥210	0.37%	≥190	0.34%
03	Individual/Small Group Administration	≥4710	8.25%	≥4670	8.19%
03	Tests Read Aloud	≥3520	6.17%	≥4050	7.10%
04	No Accommodation	≥2530	4.73%	≥2510	4.69%
04	Braille	<50	NR	<50	NR
04	Large Print	<50	NR	<50	NR
04	Answers Recorded	≥490	0.92%	≥480	0.91%
04	Extended Time	≥4770	8.92%	≥4750	8.89%
04	Transferred Answers	≥210	0.40%	≥220	0.41%
04	Individual/Small Group Administration	≥4750	8.89%	≥4730	8.85%
04	Tests Read Aloud	≥3650	6.83%	≥4250	7.95%
05	No Accommodation	≥2560	5.04%	≥2490	4.91%
05	Braille	<50	NR	<50	NR
05	Large Print	≥50	0.10%	≥50	0.10%
05	Answers Recorded	≥420	0.84%	≥410	0.82%
05	Extended Time	≥4850	9.53%	≥4830	9.49%
05	Transferred Answers	≥230	0.47%	≥220	0.45%
05	Individual/Small Group Administration	≥4840	9.52%	≥4830	9.50%
05	Tests Read Aloud	≥3720	7.33%	≥4300	8.45%
06	No Accommodation	≥2240	4.70%	≥2190	4.59%
06	Braille	<50	NR	<50	NR
06	Large Print	<50	NR	<50	NR
06	Answers Recorded	≥230	0.49%	≥230	0.48%
06	Extended Time	≥4560	9.53%	≥4470	9.35%
06	Transferred Answers	≥210	0.44%	≥200	0.43%
06	Individual/Small Group Administration	≥4440	9.29%	≥4350	9.10%
06	Tests Read Aloud	≥3530	7.39%	≥3920	8.21%

Copyright © 2017 by Louisiana Department of Education.

Special Education Accommodation Type (continued)					
		English Language Arts		Mathematics	
Grade	Accommodation	Frequency	Percent	Frequency	Percent
07	No Accommodation	≥2200	4.93%	≥2210	4.98%
07	Braille	<50	NR	<50	NR
07	Large Print	<50	NR	<50	NR
07	Answers Recorded	≥120	0.29%	≥110	0.26%
07	Extended Time	≥3840	8.60%	≥3820	8.57%
07	Transferred Answers	≥160	0.37%	≥140	0.33%
07	Individual/Small Group Administration	≥3670	8.23%	≥3630	8.16%
07	Tests Read Aloud	≥2900	6.50%	≥3240	7.28%
08	No Accommodation	≥2210	4.84%	≥1390	3.29%
08	Braille	<50	NR	<50	NR
08	Large Print	<50	NR	<50	NR
08	Answers Recorded	≥70	0.17%	≥70	0.17%
08	Extended Time	≥3860	8.44%	≥3830	9.02%
08	Transferred Answers	≥140	0.31%	≥130	0.31%
08	Individual/Small Group Administration	≥3690	8.06%	≥3660	8.63%
08	Tests Read Aloud	≥2860	6.25%	≥3270	7.71%

Copyright © 2017 by Louisiana Department of Education.

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

LEP Accommodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Frequency	Percent	Frequency	Percent
03	No Accommodation	≥430	0.76%	≥440	0.78%
03	Extended Time	≥1670	2.93%	≥1650	2.90%
03	Individual/Small Group Administration	≥1510	2.65%	≥1490	2.62%
03	English/Native Language Word-to-Word Dictionary	≥330	0.59%	≥340	0.60%
03	Directions Read Aloud/Clarified in Native Language	≥200	0.35%	≥200	0.36%
04	No Accommodation	≥260	0.50%	≥260	0.49%
04	Extended Time	≥1400	2.63%	≥1420	2.66%
04	Individual/Small Group Administration	≥1280	2.40%	≥1300	2.43%
04	English/Native Language Word-to-Word Dictionary	≥430	0.81%	≥450	0.85%
04	Directions Read Aloud/Clarified in Native Language	≥200	0.38%	≥190	0.37%
05	No Accommodation	≥180	0.36%	≥170	0.34%
05	Extended Time	≥1060	2.09%	≥1050	2.07%
05	Individual/Small Group Administration	≥920	1.82%	≥920	1.82%
05	English/Native Language Word-to-Word Dictionary	≥380	0.76%	≥370	0.74%
05	Directions Read Aloud/Clarified in Native Language	≥140	0.28%	≥120	0.24%
06	No Accommodation	≥110	0.23%	≥100	0.23%
06	Extended Time	≥770	1.63%	≥780	1.65%
06	Individual/Small Group Administration	≥520	1.10%	≥550	1.15%
06	English/Native Language Word-to-Word Dictionary	≥510	1.07%	≥490	1.04%
06	Directions Read Aloud/Clarified in Native Language	≥140	0.30%	≥90	0.20%
07	No Accommodation	≥100	0.24%	≥100	0.24%
07	Extended Time	≥780	1.75%	≥760	1.72%
07	Individual/Small Group Administration	≥470	1.06%	≥480	1.1%
07	English/Native Language Word-to-Word Dictionary	≥570	1.28%	≥550	1.24%
07	Directions Read Aloud/Clarified in Native Language	≥90	0.21%	≥60	0.14%

Copyright © 2017 by Louisiana Department of Education.

LEP Accommodation Type (continued)					
		English Language Arts		Mathematics	
Grade	Accommodation	Frequency	Percent	Frequency	Percent
08	No Accommodation	≥80	0.19%	≥70	0.18%
08	Extended Time	≥750	1.66%	≥790	1.87%
08	Individual/Small Group Administration	≥520	1.14%	≥580	1.37%
08	English/Native Language Word-to-Word Dictionary	≥610	1.34%	≥640	1.53%
08	Directions Read Aloud/Clarified in Native Language	≥140	0.32%	≥150	0.36%

Copyright © 2017 by Louisiana Department of Education.

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

Section 504 Status Accommodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Frequency	Percent	Frequency	Percent
03	No Accommodation	≥300	0.54%	≥270	0.48%
03	Large Print	<50	NR	<50	NR
03	Answers Recorded	≥70	0.14%	≥70	0.13%
03	Extended Time	≥3810	6.67%	≥3790	6.64%
03	Transferred Answers	≥50	0.09%	≥50	0.10%
03	Individual/Small Group Administration	≥3520	6.17%	≥3510	6.14%
03	Tests Read Aloud	≥1780	3.13%	≥2420	4.25%
04	No Accommodation	≥290	0.55%	≥280	0.53%
04	Large Print	<50	NR	<50	NR
04	Answers Recorded	≥110	0.22%	≥110	0.21%
04	Extended Time	≥4500	8.41%	≥4540	8.48%
04	Transferred Answers	≥80	0.15%	≥80	0.15%
04	Individual/Small Group Administration	≥4080	7.64%	≥4100	7.66%
04	Tests Read Aloud	≥2110	3.94%	≥2930	5.47%
05	No Accommodation	≥260	0.51%	≥220	0.44%
05	Braille	<50	NR	<50	NR
05	Large Print	<50	NR	<50	NR
05	Answers Recorded	≥90	0.19%	≥90	0.19%
05	Extended Time	≥4810	9.45%	≥4800	9.44%
05	Transferred Answers	≥90	0.18%	≥90	0.19%
05	Individual/Small Group Administration	≥4230	8.33%	≥4220	8.30%
05	Tests Read Aloud	≥2120	4.17%	≥2970	5.85%
06	No Accommodation	≥260	0.55%	≥240	0.51%
06	Large Print	<50	NR	<50	NR
06	Answers Recorded	≥60	0.13%	≥50	0.12%
06	Extended Time	≥4450	9.31%	≥4390	9.17%
06	Transferred Answers	≥60	0.13%	≥60	0.14%
06	Individual/Small Group Administration	≥3710	7.77%	≥3680	7.70%
06	Tests Read Aloud	≥1990	4.17%	≥2690	5.64%
07	No Accommodation	≥230	0.53%	≥210	0.48%
07	Braille	<50	NR	<50	NR
07	Large Print	<50	NR	<50	NR
07	Answers Recorded	≥50	0.13%	≥50	0.12%
07	Extended Time	≥3780	8.47%	≥3760	8.46%
07	Transferred Answers	≥70	0.16%	≥60	0.15%
07	Individual/Small Group Administration	≥3040	6.81%	≥3000	6.74%
07	Tests Read Aloud	≥1570	3.53%	≥2110	4.74%

Copyright © 2017 by Louisiana Department of Education.

Section 504 Status Accommodation Type (continued)					
		English Language Arts		Mathematics	
Grade	Accommodation	Frequency	Percent	Frequency	Percent
08	No Accommodation	≥220	0.48%	≥180	0.44%
08	Braille	<50	NR	<50	NR
08	Large Print	<50	NR	<50	NR
08	Answers Recorded	<50	NR	<50	NR
08	Extended Time	≥3570	7.80%	≥3520	8.29%
08	Transferred Answers	≥60	0.14%	≥50	0.12%
08	Individual/Small Group Administration	≥2850	6.23%	≥2830	6.68%
08	Tests Read Aloud	≥1430	3.13%	≥1950	4.61%

Copyright © 2017 by Louisiana Department of Education.

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

Online Accomodation Type					
		English Language Arts		Mathematics	
Grade	Accommodation	Frequency	Percent	Frequency	Percent
04	Text-to-Speech	<50	NR	≥60	13.16%
04	Human Read Aloud	<50	NR	<50	NR
04	Answers Recorded	<50	NR	<50	NR
04	Extended Time	≥70	17.11%	≥70	17.11%
04	Individual/Small Group Administration	≥80	18.42%	≥80	18.20%
05	Text-to-Speech	≥150	9.27%	≥280	16.73%
05	Human Read Aloud	≥90	5.32%	≥140	8.81%
05	Native Language Word-to-Word Dictionary	<50	NR	<50	NR
05	Directions in Native Language	<50	NR	<50	NR
05	Transferred Answers	<50	NR	<50	NR
05	Answers Recorded	<50	NR	<50	NR
05	Extended Time	≥330	19.67%	≥330	19.62%
05	Individual/Small Group Administration	≥290	17.37%	≥290	17.32%
06	Text-to-Speech	≥390	8.11%	≥520	10.97%
06	Human Read Aloud	≥70	1.62%	≥100	2.21%
06	Native Language Word-to-Word Dictionary	<50	NR	<50	NR
06	Directions in Native Language	<50	NR	<50	NR
06	Transferred Answers	<50	NR	<50	NR
06	Answers Recorded	<50	NR	<50	NR
06	Extended Time	≥870	18.25%	≥870	18.17%
06	Individual/Small Group Administration	≥600	12.47%	≥590	12.38%
07	Text-to-Speech	≥380	6.10%	≥490	7.88%
07	Human Read Aloud	≥120	2.03%	≥160	2.57%
07	Native Language Word-to-Word Dictionary	<50	NR	<50	NR
07	Directions in Native Language	<50	NR	<50	NR
07	Transferred Answers	<50	NR	<50	NR
07	Answers Recorded	<50	NR	<50	NR
07	Extended Time	≥950	15.05%	≥940	15.11%
07	Individual/Small Group Administration	≥650	10.44%	≥650	10.49%
08	Text-to-Speech	≥280	5.21%	≥370	8.33%
08	Human Read Aloud	≥90	1.70%	≥120	2.88%
08	Native Language Word-to-Word Dictionary	<50	NR	<50	NR
08	Directions in Native Language	<50	NR	<50	NR
08	Transferred Answers	<50	NR	<50	NR
08	Answers Recorded	<50	NR	<50	NR
08	Extended Time	≥710	13.12%	≥700	15.97%
08	Individual/Small Group Administration	≥460	8.52%	≥450	10.34%

Copyright © 2017 by Louisiana Department of Education.

4.5 Summary

In summary, the overall purpose of each of the test administration workshops and the ancillary materials is to keep districts informed about policies and procedures related to testing in general and the LEAP program in particular. The information imparted is clearly related to standardizing the administration of the LEAP, maintaining the security of the assessment, allowing access to the assessments for special populations by clearly delineating appropriate accommodations, and by providing guidance on appropriate interpretations of the test results. These communication and training efforts by the LDOE and the ancillary information developed by DRC address multiple best practices of the testing industry but in particular are related to the following *Standards*:

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

Copyright © 2017 by Louisiana Department of Education.

CHAPTER 5: CONSTRUCTED-RESPONSE AND TECHNOLOGY-ENHANCED SCORING

In this chapter, the scoring process used for 2016 LEAP ELA and mathematics is described, with a particular focus on the handscoring process of constructed-response items and the automated scoring of technology-enhanced items. At the end of this section, the results of the inter-rater reliability for the handscoring of the LEAP constructed-response items are presented.

Chapter 5 adheres to the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (2014) Standards 4.18, 4.20, 6.8, and 6.9. Each of these standards are 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 ELA and mathematics constructed-response items and technology-enhanced items. The scoring criteria used for each item are not presented in this chapter in order to preserve the integrity of the items for future use.

5.1 Constructed-Response Scoring Process

Constructed-response items were scored by human raters who were trained by DRC.

5.1.1 Selection of Scoring Evaluators

Standard 4.20 states the following:

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

Sections 5.1.1 and 5.1.2 explain how scorers are selected and trained for the LEAP ELA and mathematics handscoring process. Section 5.1.3 describes how the scorers are monitored throughout the handscoring process.

Copyright © 2017 by Louisiana Department of Education.

DRC strives to develop a highly qualified, experienced core of evaluators so that the integrity of all projects is appropriately maintained.

Recruitment and the Interview Process
All readers hired by DRC to score 2016 LEAP ELA and Mathematic responses had at least a four-year college degree in an appropriate field, such as a bachelor's degree in a STEM field or a liberal arts degree in English language arts.

DRC has a human resources director dedicated solely to recruiting and retaining the handscoring staff. Applications for reader positions are screened by the handscoring project manager, the human resources director, or recruiting staff to create a large pool of potential readers. In the screening process, preference is given to candidates with previous experience scoring large-scale assessments and with degrees emphasizing the appropriate content areas. At the personal interview, reader candidates are asked to demonstrate their own proficiency in writing by responding to a DRC writing topic and their proficiency in mathematics by solving word problems with correct work shown. All of this results in a highly qualified 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.

ELA and mathematics scorers hired by DRC meet the degree requirements as specified by the Partnership for Assessment of Readiness for College and Careers (PARCC) and conveyed by Pearson described in the following two paragraphs:

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

Copyright © 2017 by Louisiana Department of Education.

5.1.2 Handscoring Training Process

Standard 6.9 specifies:

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

Training Material Development

DRC scoring supervisors trained scorers using 2016 PARCC-approved training materials provided by PARCC, including the following:

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

Training and Qualifying Procedures

Handscoring involves training and qualifying team leaders and evaluators, monitoring scoring accuracy and production, and ensuring security of both the test materials and the scoring facilities. An explanation of the training and qualification procedures follows.

DRC used the PARCC-approved Math and ELA training and qualifying materials. These materials for each item were grouped into two categories: “prototype” items and “abbreviated” items.

Prototype Items

Some items included in the Louisiana forms were prototype items, meaning they had full sets of associated training materials, including anchor sets, practice sets, and qualifying sets. DRC started the training 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 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 training materials for an associated prototype item that was similar to the abbreviated one they would be scoring

Copyright © 2017 by Louisiana Department of Education.

on the Louisiana form.[1] Readers who did not qualify on the prototype item were not allowed to continue the training.

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

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

Table 5.1 Mathematics Training Set Composition

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

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

Copyright © 2017 by Louisiana Department of Education.

Table 5.2 ELA Training Set Composition

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

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

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

Copyright © 2017 by Louisiana Department of Education.

Qualifying Standards

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

Scorers demonstrated their ability to apply the scoring criteria by qualifying (i.e., scoring with acceptable agreement with true scores on qualifying sets). After each qualifying set was scored, the DRC scoring director responsible for training the item 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 actual student work.

Table 5.3 Mathematics Qualifying Standards

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

Table 5.4 Mathematics Qualifying Standards (Composite Items)*

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

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

Table 5.5 ELA Qualifying Standards

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

ELA readers were required to meet all three of the qualifications listed in the table. 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 non-adjacent. In other words, no more than 2 of the 60 applied scores could be non-adjacent (3 sets x 10 responses/set x 2 traits = 60 applied score).

Copyright © 2017 by Louisiana Department of Education.

5.1.3 Monitoring the Scoring Process

Standard 6.8 states:

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

Section 5.1.3 explains the monitoring procedures that DRC uses to ensure that handscoring evaluators follow established scoring criteria while items are being scored. Detailed scoring rubrics, which specify the criteria for scoring, are available for all constructed-response items.

Reader Monitoring Procedures

Throughout the handscoring process, DRC project managers, scoring directors, and team leaders reviewed the statistics that were generated on a daily basis. DRC used one team leader for every 10–12 readers, the same ratio that Pearson used for PARCC. If scoring patterns were apparent among individual scorers, team leaders dealt with issues of this sort 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 based on regular read-behinds and the continuous monitoring of inter-rater reliability and score point distributions, DRC also conducted validity scoring using PARCC-approved validity responses supplied by PARCC. Validity responses were inserted among the live student responses. DRC used the same validity responses that Pearson used, and these included both paper-based test (PBT) and computer-based test (CBT) responses.

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

Approximately 10% of all 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 this 10% of responses that received second reads. 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 re-dealt and rescored.

Copyright © 2017 by Louisiana Department of Education.

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.

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

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

Agreement Rate Requirements for Validity and Inter-Rater Reliability			
Subject	**Score Point Range**	**Perfect Agreement**	**Perfect Agreement + Adjacent**
Math	0–1	90%	95%
Math	0–2	80%	95%
Math	0–3	70%	95%
Math	0–4	65%	95%
ELA	Multi-trait 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 displayed under "Perfect Agreement" in the table above. Additionally, readers were required to maintain an acceptably low rate of nonadjacent agreement. To monitor this, each reader's percentages of exact and adjacent agreement rates, and required each reader to maintain the levels displayed under "Perfect Agreement + Adjacent" on the table above were summed.

Calibration Sets

PARCC provided DRC with PARCC-approved calibration sets. DRC used these sets to perform calibration across the entire scorer population for an item if trends were detected (e.g., low agreement between certain score points, if a certain type of response was missing from initial training). These calibrations were designed to help refocus scorers on how to properly use the scoring guidelines. They were selected to help illustrate particular points and familiarize scorers with the types of responses commonly seen during operational scoring. After the scorers took a calibration set, the scoring director reviewed it from the front of the room, using rubric language and the anchor responses to explain the reasoning behind each response's score.

Copyright © 2017 by Louisiana Department of Education.

Reports and Reader Feedback

Reader performance and intervention information were recorded in reader feedback logs. These Logs tracked information about actions taken with individual readers to ensure scoring consistency in regard to reliability, score point distribution, and validity performance.

5.1.4 Security

Each DRC scoring center is a secure facility. All employees are issued photo identification badges and are required to wear them in plain view at all times. Access to scoring centers is limited to badge-wearing staff and to visitors accompanied by authorized staff. All readers are made aware that no scoring materials may leave the scoring center and must sign legally binding confidentiality agreements before work begins. DRC retains these agreements for the duration of the contract. To prevent the unauthorized duplication of secure materials, cell phone/camera use within the scoring rooms is strictly forbidden. Readers only have access to the student responses they are qualified to score. Each scorer is assigned a unique username and password to access the DRC imaging system and must qualify before viewing any live student responses. DRC maintains full control of who may access the system and which item each scorer may score. No demographic data is available to scorers at any time.

5.2 Technology-Enhanced Item Scoring Process

All technology-enhanced items were processed through DRC's autoscoring engine and scored according to the assigned scoring rules. DRC ensured that all rubrics and scoring rules were verified for accuracy before scoring any technology-enhanced items. DRC established an adjudication process for technology-enhanced items and any gridded responses to verify that correct answers were identified. DRC's technology-enhanced scoring quality process included the following procedures:

- A scoring rubric was created for each technology-enhanced item. It was as simple as describing the one and only correct answer for dichotomously scored items (scored as either right or wrong). If partial credit was possible, the rubric described in detail the type of response that could receive credit for each score point.
- The information from the scoring rubric was entered into the scoring system within the item banking system so that the truth resided in one place, along with the item image and other metadata. This scoring information designated specific information that varied by item type. For example, for a drag-and-drop item, the information included which objects are to be placed in which 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.

Copyright © 2017 by Louisiana Department of Education.

- 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. Scoring was then rerun. This checking and modification process continued until no other issues were found.
- As a final check, a final report was run that showed all student responses, along with their frequencies and received scores.

In cases of braille, large print, or PBT non-accommodated form administration, student responses to PBT technology-enhanced-equivalent items were transcribed (entered) into the online system by a test administrator.

5.3 Multiple-Choice and Multi-Select Item Scoring Process

Responses to multiple-choice and multi-select items were captured during the CBT administration. In the case of braille, large print, or PBT non-accommodated form administration, student responses to these items were transcribed into the online system by a test administrator.

5.4 Inter-Rater Reliability

Approximately 10% of the responses in ELA and mathematics were scored independently by a second reader. The statistics for the inter-rater reliability were calculated for all items at all grades. To determine the reliability of scoring, the percentage of perfect agreement and adjacent agreement between the two readers was examined.

A total of 51 items were scored by human readers across all grades and content areas. Inter-rater reliability as well as the total number of item reads for each of the handscored items is shown in Table 5.6 for ELA items, Table 5.7 for mathematic items, and Table 5.8 for Spanish mathematics items.

As shown in Table 5.6, raters demonstrated at or above 98% perfect and adjacent agreement for all ELA handscored items. As shown in Table 5.7, raters demonstrated at or above 97% perfect and adjacent agreement for mathematics items. As shown in 5.8, raters demonstrated perfect agreement for Spanish mathematics items.

Copyright © 2017 by Louisiana Department of Education.

Table 5.6 Inter-Rater Agreement, English Language Arts Tasks

Grade	Task Type	Question	Trait	Total Reads	Read 2x	IRR %		
						EX	Adj	EX + Adj
3	Research Simulation	7	Reading Comprehension and Written Expression	63,727	12400	73	26	99
			Knowledge of Language and Conventions	63,727	12400	70	28	98
3	Literary Analysis	13	Reading Comprehension and Written Expression	63,992	12930	73	27	100
			Knowledge of Language and Conventions	63,992	12930	71	29	100
4	Research Simulation	8	Reading Comprehension and Written Expression	61,702	14620	73	26	99
			Knowledge of Language and Conventions	61,702	14620	72	26	98
4	Narrative Writing	13	Written Expression	60,346	11912	75	25	99
			Knowledge of Language and Conventions	60,346	11912	73	27	100
5	Research Simulation	8	Reading Comprehension and Written Expression	58,738	11696	82	18	100
			Knowledge of Language and Conventions	58,738	11696	80	20	100
5	Literary Analysis	14	Reading Comprehension and Written Expression	60,357	14914	77	22	99
			Knowledge of Language and Conventions	60,357	14914	75	23	98
6	Research Simulation	8	Reading Comprehension and Written Expression	59,674	13356	73	26	99
			Knowledge of Language and Conventions	59,674	13356	73	26	99
6	Literary Analysis	14	Reading Comprehension and Written Expression	59,546	10370	73	26	99
			Knowledge of Language and Conventions	59,546	13070	75	25	100
7	Research Simulation	8	Reading Comprehension and Written Expression	57,981	13136	73	25	2
			Knowledge of Language and Conventions	57,981	13136	73	26	2
7	Literary Analysis	14	Reading Comprehension and Written Expression	56,847	10868	73	26	2
			Knowledge of Language and Conventions	56,847	10868	74	25	1
8	Research Simulation	8	Reading Comprehension and Written Expression	57,482	11216	71	27	2
			Knowledge of Language and Conventions	57,482	11216	71	27	1
8	Narrative	13	Written Expression	57,222	10754	74	25	1
			Knowledge of Language and Conventions	57,222	10754	76	23	1

Copyright © 2017 by Louisiana Department of Education.

Table 5.7 Inter-Rater Agreement, Mathematics Items

Grade	Question	Part(s)	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	Adj	EX + Adj
3	14	N/A	63,617	11836	96	3	99
3	15	Part A	63,659	11920	97	3	100
3	30	Part C	63,658	11918	87	13	100
3	31	N/A	63,633	11868	88	11	99
3	44	Part C	63,817	12236	89	8	97
3	45	N/A	63,625	11852	90	10	100
4	15	Part A	59,958	11302	87	12	99
		Part B	59,958	11302	92	8	100
4	16	N/A	59,885	11170	90	10	100
4	30	N/A	59,860	11102	90	10	100
4	31	Part B	59,378	11060	94	6	100
4	31	Part A	524	128	100	0	100
		Part B	524	128	94	6	100
4	43	N/A	60,232	11850	96	4	100
4	44	Part B	59,499	11302	98	1	99
		Part C	59,499	11302	96	3	99
4	44	Part A	509	98	100	0	100
		Part B	509	98	98	2	100
		Part C	509	98	98	2	100
5	14	Part B	56,321	10446	92	8	100
		Part D	56,321	10446	88	12	100
5	14	Part A	1875	366	100	0	100
		Part B	1875	366	91	9	100
		Part C	1875	366	100	0	100
		Part D	1875	366	91	9	100
5	15	N/A	58,251	10952	96	4	100
5	30	N/A	58,228	10874	88	12	100
5	31	N/A	58,184	10798	84	14	98
5	43	Part A	58,139	10696	98	2	100
		Part B	58,139	10696	91	8	99
5	44	Part C	56,275	10354	86	13	99
5	44	Part A	1874	362	100	0	100
		Part B	1874	362	100	0	100
		Part C	1874	362	86	14	100

Copyright © 2017 by Louisiana Department of Education.

Grade	Question	Part(s)	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	Adj	EX + Adj
6	29	N/A	58,706	11512	81	17	2
6	30	N/A	58,380	10926	78	20	2
6	31	N/A	58,339	10870	90	8	1
6	32	N/A	58,298	10842	94	5	1
6	42	Part A	58,308	10722	91	9	0
		Part B	58,308	10722	94	6	0
6	43	N/A	58,402	11004	93	6	1
6	44	Part B	53,033	9720	99	0	0
		Part C	53,033	9720	100	0	0
6	44	Part A	5276	996	100	0	0
		Part B	5276	996	99	1	0
		Part C	5276	996	99	1	0
7	29	N/A	56,445	10616	94	6	0
7	30	N/A	56,501	10706	94	6	0
7	31	N/A	56,551	10978	98	2	0
7	32	N/A	56,420	10776	98	2	0
7	40	Part A	56,397	10412	94	6	0
		Part B	56,397	10412	90	10	0
7	41	Part A	56,705	11202	96	4	0
		Part B	56,705	11202	92	8	0
7	42	Part C	49,802	9700	95	5	0
		Part D	49,802	9700	92	8	0
7	42	Part A	6909	1364	100	0	0
		Part B	6909	1364	100	0	0
		Part C	6909	1364	93	7	0
		Part D	6909	1364	88	12	1
8	30	N/A	52,349	10152	88	11	1
8	31	N/A	52,294	10168	90	10	0
8	32	N/A	52,207	9936	97	3	0
8	33	N/A	52,230	10012	86	12	1
8	43	N/A	52,277	9978	98	2	0
8	44	Part A	52,214	9816	97	3	0
		Part B	52,214	9816	95	5	0
8	45	Part A	52,219	9910	98	2	0
		Part B	52,219	9910	96	4	0

Copyright © 2017 by Louisiana Department of Education.

Table 5.8 Inter-Rater Agreement, Spanish Mathematics Items

Grade	Question	Part(s)	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	Adj	EX + Adj
3	14	N/A	70	16	100	0	100
3	15	Part A	74	24	100	0	100
3	30	Part C	75	26	100	0	100
3	31	N/A	76	28	100	0	100
3	44	Part C	70	16	100	0	100
3	45	N/A	74	24	100	0	100
4	15	Part A	102	28	100	0	100
		Part B	102	28	100	0	100
4	16	N/A	98	20	100	0	100
4	30	N/A	104	32	94	6	100
4	31	Part B	97	18	100	0	100
4	31	Part A	2	2	100	0	100
		Part B	2	2	100	0	100
4	43	N/A	104	30	100	0	100
4	44	Part B	99	22	100	0	100
		Part C	99	22	100	0	100
5	14	Part B	96	28	100	0	100
		Part D	96	28	100	0	100
5	14	Part A	3	2	100	0	100
		Part B	3	2	100	0	100
		Part C	3	2	100	0	100
		Part D	3	2	100	0	100
5	15	N/A	94	22	100	0	100
5	30	N/A	96	26	100	0	100
5	31	N/A	97	26	100	0	100
5	43	Part A	94	20	100	0	100
		Part B	94	20	100	0	100
5	44	Part C	94	24	100	0	100
5	44	Part A	4	2	100	0	100
		Part B	4	2	100	0	100
		Part C	4	2	100	0	100

Copyright © 2017 by Louisiana Department of Education.

Grade	Question	Part(s)	Total Reads	Read 2x	Inter-Rater Reliability %		
					EX	Adj	EX + Adj
6	29	N/A	141	30	100	0	0
6	30	N/A	145	38	100	0	0
6	31	N/A	148	44	100	0	0
6	32	N/A	142	30	100	0	0
6	42	Part A	152	50	100	0	0
		Part B	152	50	100	0	0
6	43	N/A	142	32	100	0	0
6	44	Part B	139	30	100	0	0
		Part C	139	30	100	0	0
6	44	Part A	5	2	100	0	0
		Part B	5	2	100	0	0
		Part C	5	2	100	0	0
7	29	N/A	135	30	93	7	0
7	30	N/A	132	26	92	8	0
7	31	N/A	134	30	100	0	0
7	32	N/A	133	28	100	0	0
7	40	Part A	136	32	94	6	0
		Part B	136	32	94	6	0
7	41	Part A	133	28	100	0	0
		Part B	133	28	86	14	0
7	42	Part Ch	137	42	90	10	0
		Part Dh	137	42	100	0	0
7	42	Part A	7	6	100	0	0
		Part B	7	6	100	0	0
		Part C	7	6	100	0	0
		Part D	7	6	100	0	0
8	30	N/A	205	50	96	4	0
8	31	N/A	205	52	92	8	0
8	32	N/A	200	42	100	0	0
8	33	N/A	202	44	100	0	0
8	43	N/A	202	46	100	0	0
8	44	Part A	202	48	92	8	0
		Part B	202	48	96	4	0
8	45	Part A	199	42	100	0	0
		Part B	199	42	100	0	0

Copyright © 2017 by Louisiana Department of Education.

5.5 Summary

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

> ***Standard 4.18*** Procedures for scoring and, if relevant, scoring criteria, should be presented by the test developer with sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical for extended-response items such as performance tasks, portfolios, and essays. (91)
>
> ***Standard 4.20*** The process for selecting, training, qualifying, and monitoring scorers should be specified by the test developer. The training materials, such as the scoring rubrics and examples of test takers' responses that illustrate the levels on the rubric score scale, and the procedures for training scorers should result in a degree of accuracy and agreement among scorers that allows the scores to be interpreted as originally intended by the test developer. Specifications should also describe processes for assessing scorer consistency and potential drift over time in raters' scoring. (92)
>
> ***Standard 6.8*** Those responsible for test scoring should establish scoring protocols. Test scoring that involves human judgment should include rubrics, procedures, and criteria for scoring. When scoring of complex responses is done by computer, the accuracy of the algorithm and processes should be documented. (118)
>
> ***Standard 6.9*** Those responsible for test scoring should establish and document quality control processes and criteria. Adequate training should be provided. The quality of scoring should be monitored and documented. Any systematic source of scoring errors should be documented and corrected. (118)

Copyright © 2017 by Louisiana Department of Education.

CHAPTER 6: OPERATIONAL DATA ANALYSES

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

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

Chapter 6 demonstrates adherence in the LEAP program to American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (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 LEAP 2016 analyses, the Louisiana student population was used. In Section 6.3, the characteristics of calibration samples, such as subgroups, are discussed. Chapter 3 presents the test specifications. Information regarding reported data is discussed in detail in Chapter 7.

6.1 Classical Item Statistics

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

6.1.1 Test-Level Statistics

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

As can be seen in the table, for ELA, average form difficulty ranged from 0.33 to 0.53. For mathematics, average form difficulty ranged from 0.32 to 0.53. In general, LEAP 2016 tests were relatively difficult tests across all subjects and grades. For ELA, the grade 5 test was the most difficult, with 0.33 average form difficulty, and the grade 8 test

Copyright © 2017 by Louisiana Department of Education.

was the easiest, with 0.53 average form difficulty. For mathematics, the grade 8 test was the most difficult, with 0.33 average form difficulty, and the grade 4 test was the easiest, with 0.52/0.53 average form difficulty.

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

Content	Summarized Grade	Mode	Total Items	Total Points	Mean Raw Score (Std. Dev.)	Average Form Difficulty (Std. Dev.)
ELA	3	PBT	31	66	24.98 (11.81)	0.38 (0.16)
	4	CBT	37	74	29.61 (11.92)	0.40 (0.15)
	4	PBT	37	74	30.51 (12.81)	0.42 (0.13)
	5	CBT	33	77	25.25 (12.92)	0.33 (0.16)
	5	PBT	33	77	25.28 (11.99)	0.33 (0.18)
	6	CBT	33	77	33.41 (14.67)	0.44 (0.14)
	6	PBT	33	77	32.75 (13.51)	0.43 (0.15)
	7	CBT	32	76	32.56 (14.06)	0.43 (0.13)
	7	PBT	32	76	32.21 (13.15)	0.43 (0.14)
	8	CBT	37	77	40.43 (14.69)	0.53 (0.16)
	8	PBT	37	77	40.69 (13.80)	0.53 (0.15)
Mathematics	3	PBT	45	62	30.53 (11.93)	0.50 (0.26)
	4	CBT	44	62	32.55 (12.29)	0.53 (0.21)
	4	PBT	44	62	31.80 (12.62)	0.52 (0.21)
	5	CBT	44	62	27.64 (11.68)	0.45 (0.22)
	5	PBT	44	62	27.61 (11.66)	0.45 (0.21)
	6	CBT	44	66	26.20 (11.87)	0.40 (0.25)
	6	PBT	44	66	24.86 (11.35)	0.38 (0.24)
	7	CBT	42	65	23.67 (11.84)	0.37 (0.22)
	7	PBT	42	65	22.19 (11.62)	0.35 (0.21)
	8	CBT	45	66	21.36 (9.06)	0.33 (0.24)
	8	PBT	45	66	20.78 (10.07)	0.32 (0.23)

Copyright © 2017 by Louisiana Department of Education.

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

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

Table 6.2 LEAP Means, Standard Deviations for Raw Scores, p-values, Item-Total Correlation (Rit)

				Item P-value				Average Total Correlation			
Content	Grade	Mode	N of Items	Mean	Std. Dev.	Min.	Max	Mean	Std. Dev.	Min.	Max
ELA	3	PBT	29	0.46	0.16	0.24	0.82	0.41	0.11	0.15	0.57
	4	CBT	36	0.44	0.16	0.09	0.85	0.37	0.13	0.08	0.61
	4	PBT	36	0.44	0.15	0.10	0.78	0.38	0.13	0.14	0.63
	5	CBT	31	0.41	0.16	0.13	0.74	0.41	0.12	0.12	0.62
	5	PBT	31	0.43	0.17	0.13	0.75	0.39	0.11	0.18	0.58
	6	CBT	31	0.51	0.18	0.20	0.85	0.43	0.14	0.06	0.69
	6	PBT	31	0.52	0.17	0.22	0.86	0.41	0.13	0.04	0.63
	7	CBT	30	0.47	0.16	0.23	0.82	0.38	0.13	0.23	0.69
	7	PBT	30	0.48	0.16	0.21	0.83	0.37	0.13	0.22	0.66
	8	CBT	36	0.54	0.21	0.08	0.85	0.41	0.14	0.04	0.70
	8	PBT	36	0.56	0.19	0.15	0.85	0.41	0.12	0.15	0.67
Mathematics	3	PBT	45	0.57	0.25	0.06	0.97	0.41	0.12	0.15	0.74
	4	CBT	44	0.59	0.21	0.20	0.95	0.42	0.14	0.08	0.67
	4	PBT	44	0.57	0.20	0.17	0.92	0.43	0.12	0.12	0.70
	5	CBT	44	0.51	0.22	0.06	0.95	0.40	0.15	0.10	0.66
	5	PBT	44	0.52	0.21	0.06	0.93	0.40	0.13	0.13	0.65
	6	CBT	44	0.48	0.24	0.06	0.91	0.43	0.14	0.09	0.68
	6	PBT	44	0.46	0.23	0.06	0.90	0.41	0.14	0.06	0.63
	7	CBT	42	0.41	0.21	0.05	0.79	0.41	0.13	0.16	0.69
	7	PBT	42	0.38	0.20	0.04	0.78	0.40	0.13	0.17	0.71
	8	CBT	45	0.41	0.23	0.02	0.90	0.33	0.10	0.11	0.61
	8	PBT	45	0.40	0.22	0.03	0.91	0.36	0.11	0.13	0.65

6.1.2 Item-Level Statistics

Tables 6.3 through 6.13 present the item statistics for each item included in regular test forms by grade for ELA. Tables 6.14 through 6.24 show the item statistics for each item included in regular test forms by grade for mathematics. The tables include administration mode, item number, *p*-value, item-total correlation (R_{it}), omit rates, total

Copyright © 2017 by Louisiana Department of Education.

N, and adjusted N count (adjusted N excludes omits), and the percentage at each score point, if applicable, for each item by grade and content area.

p-value

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

In terms of *p*-values, test scores tend to be more precise when their average *p*-values are in the mid 0.50s to low 0.70s. However, in building a criterion-referenced test, it is important to select items on the basis of content rather than on purely statistical criteria. As shown in Table 6.2, the average *p*-values associated with the ELA forms range from 0.33 (grade 5, online and paper forms) to 0.53 (grade 8, online and paper forms). The average *p*-values associated with the mathematics forms (Table 6.2) range from 0.32 (grade 8, paper form) to 0.53 (grade 4, online form).

It is important that one examines the range of *p*-values, not just the average *p*-value, to determine whether a test measures well. It is desirable for the test to measure well throughout the range of skills present at a given grade. That is, it is important that the items measure the performance of both low-scoring and high-scoring students, as well as students in the center of the distribution. Having a range of *p*-values also helps to prevent floor and/or ceiling effects so that the test does not have large numbers of students at the minimum or maximum possible scores. The ELA forms have items with *p*-values ranging from the 0.08 to the 0.86 (see Tables 6.3 through 6.13) across all grade levels. The *p*-values on the mathematics forms range from the 0.02 to 0.97 (see Tables 6.14 through 6.24). Such a broad range of *p*-values, which indicates the items measure well throughout the range of skills at a given grade, supports the accuracy of the LEAP test scores.

Item-Total Correlations

An item-total correlation is the correlation between an item and the total test score, where the item score is not included in the total score. It indicates how well an item differentiates between low- and high-achieving students. In general, items with correlations below 0.20 are said to be poorly discriminating. The majority of the items in the LEAP had item-test 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. As a rule of thumb, an item is said to have a high omit rate if more than 5% of students failed to respond to the item.

Copyright © 2017 by Louisiana Department of Education.

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

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

The results in this section will show that, overall, student test scores are not adversely affected by the rate at which they complete the test. In general, students have ample time to complete all sections of the test.

Again, the results presented in Tables 6.3 through 6.24 show that the omit rates for the items on any of the LEAP regular forms were less than 5%, suggesting that the majority of students were able to complete the test in the prescribed amount of time.

Copyright © 2017 by Louisiana Department of Education.

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

ELA Grade 3 Paper-Based Test Administration										
Item	Item Type	Total N	Adj. N	*p*-value	Pbis Rit	Omit Rate	% at 0	% at 1	% at 2	% at 3
1	EBSR	≥57,160	≥56,970	0.67	0.41	0.32	21.81	21.23	56.64	
2	EBSR	≥57,160	≥56,820	0.40	0.36	0.58	52.94	13.48	33.00	
3	EBSR	≥57,160	≥56,840	0.50	0.48	0.56	44.90	10.08	44.46	
4	EBSR	≥57,160	≥56,780	0.51	0.28	0.65	43.92	9.97	45.46	
5	EBSR	≥57,160	≥56,740	0.64	0.49	0.73	29.33	13.58	56.36	
6	EBSR	≥57,160	≥56,700	0.67	0.47	0.79	26.79	12.16	60.25	
7	CR	≥57,160	≥56,480	0.24	0.53	0.86	42.24	44.31	11.22	1.06
7	CR	≥57,160	≥56,480	0.31	0.57	0.86	31.23	46.57	19.19	1.84
8	EBSR	≥57,160	≥56,880	0.39	0.49	0.49	43.85	33.58	22.08	
9	EBSR	≥57,160	≥56,820	0.37	0.40	0.58	52.93	18.76	27.73	
10	EBSR	≥57,160	≥56,710	0.67	0.48	0.78	23.09	19.64	56.50	
11	EBSR	≥57,160	≥56,710	0.25	0.15	0.79	66.14	16.33	16.74	
12	EBSR	≥57,160	≥56,660	0.47	0.52	0.86	26.51	52.55	20.08	
13	CR	≥57,160	≥56,450	0.25	0.55	0.75	38.79	46.98	11.98	1.02
13	CR	≥57,160	≥56,450	0.27	0.54	0.75	37.50	44.95	14.85	1.47
14	EBSR	≥57,160	≥56,840	0.46	0.35	0.55	53.81	45.65		
15	EBSR	≥57,160	≥56,750	0.37	0.46	0.70	62.73	36.57		
16	MC	≥57,160	≥56,800	0.73	0.29	0.58				
17	MS	≥57,160	≥56,090	0.35	0.36	1.86	63.53	34.61		
18	MC	≥57,160	≥56,180	0.59	0.45	1.53				
19	MC	≥57,160	≥56,880	0.48	0.27	0.42				
20	MC	≥57,160	≥56,710	0.45	0.35	0.67				
21	MC	≥57,160	≥56,810	0.58	0.38	0.55				
22	EBSR	≥57,160	≥56,670	0.32	0.45	0.86	67.74	31.40		
23	EBSR	≥57,160	≥56,200	0.28	0.43	1.67	70.49	27.84		
24	MS	≥57,160	≥56,390	0.36	0.44	1.35	63.32	35.34		
25	MC	≥57,160	≥56,420	0.82	0.38	1.20				
26	MS	≥57,160	≥55,390	0.31	0.21	3.09	67.35	29.56		
27	MC	≥57,160	≥55,570	0.52	0.26	2.15				

Copyright © 2017 by Louisiana Department of Education.

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

ELA Grade 4 Computer-Based Test Administration											
Item	Item Type	Total N	Adj. N	*p*-value	Pbis Rit	Omit Rate	% at 0	% at 1	% at 2	% at 3	% at 4
1	EBSR	≥450	≥450	0.39	0.40	0.00	52.85	16.01	31.14		
2	EBSR	≥450	≥450	0.25	0.23	0.00	63.16	22.81	14.04		
3	EBSR	≥450	≥450	0.65	0.44	0.00	16.89	36.18	46.93		
4	EBSR	≥450	≥450	0.34	0.41	0.00	46.93	38.16	14.91		
5	EBSR	≥450	≥450	0.15	0.19	0.00	81.80	6.36	11.84		
6	EBSR	≥450	≥450	0.52	0.31	0.22	34.43	26.32	39.04		
7	EBSR	≥450	≥450	0.36	0.50	0.22	42.76	42.98	14.04		
8	CR	≥450	≥450	0.28	0.49	0.00	15.79	60.09	20.18	2.85	0.22
8	CR	≥450	≥450	0.38	0.53	0.00	19.52	51.54	23.03	5.04	
9	EBSR	≥450	≥450	0.56	0.44	0.00	39.25	8.99	51.75		
10	EBSR	≥450	≥450	0.71	0.51	0.00	25.22	7.02	67.76		
11	EBSR	≥450	≥450	0.29	0.25	0.00	67.11	7.46	25.44		
12	EBSR	≥450	≥450	0.38	0.27	0.00	52.63	17.76	29.61		
13	CR	≥450	≥450	0.42	0.59	0.22	14.91	49.78	28.73	6.14	
13	CR	≥450	≥450	0.41	0.61	0.22	13.60	53.07	28.73	4.17	
14	TE	≥450	≥450	0.47	0.38	0.00	52.85	47.15			
15	EBSR	≥450	≥450	0.34	0.32	0.00	66.01	33.99			
16	MS	≥450	≥450	0.28	0.49	0.00	72.15	27.85			
17	EBSR	≥450	≥450	0.09	0.08	0.22	91.01	8.77			
18	MC	≥450	≥450	0.85	0.41	0.00					
19	MC	≥450	≥450	0.48	0.27	0.00					
20	MC	≥450	≥450	0.55	0.22	0.00					
21	MC	≥450	≥450	0.65	0.38	0.00					
22	EBSR	≥450	≥450	0.15	0.26	0.00	85.31	14.69			
23	MS	≥450	≥450	0.57	0.29	0.00	42.54	57.46			
24	MC	≥450	≥450	0.54	0.22	0.00					
25	MC	≥450	≥450	0.42	0.32	0.22					
26	TE	≥450	≥450	0.63	0.50	0.22	36.84	62.94			
27	MS	≥450	≥450	0.38	0.46	0.00	61.62	38.38			
28	MC	≥450	≥450	0.36	0.09	0.00					
29	EBSR	≥450	≥450	0.41	0.54	0.00	59.43	40.57			
30	MC	≥450	≥450	0.61	0.37	0.00					
31	TE	≥450	≥450	0.52	0.51	0.00	48.25	51.75			
32	MS	≥450	≥450	0.62	0.46	0.00	38.38	61.62			
33	MC	≥450	≥450	0.47	0.38	0.00					
34	MC	≥450	≥450	0.47	0.28	0.00					

Copyright © 2017 by Louisiana Department of Education.

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

ELA Grade 4 Paper-Based Test Administration										
Item	Item Type	Total N	Adj. N	*p*-value	Pbis Rit	Omit Rate	% at 0	% at 1	% at 2	% at 3
1	EBSR	≥53,520	≥53,380	0.38	0.39	0.27	54.97	13.74	31.02	
2	EBSR	≥53,520	≥53,250	0.28	0.27	0.50	64.19	15.57	19.75	
3	EBSR	≥53,520	≥53,340	0.66	0.44	0.34	17.47	33.40	48.79	
4	EBSR	≥53,520	≥53,310	0.32	0.43	0.39	50.32	34.72	14.57	
5	EBSR	≥53,520	≥53,250	0.15	0.22	0.51	82.96	4.10	12.43	
6	EBSR	≥53,520	≥53,210	0.56	0.32	0.58	32.41	23.29	43.72	
7	EBSR	≥53,520	≥53,110	0.37	0.41	0.76	39.92	44.26	15.06	
8	CR	≥53,520	≥52,650	0.32	0.60	0.74	14.80	45.61	33.98	3.73
8	CR	≥53,520	≥52,650	0.43	0.60	0.74	14.09	44.65	35.34	4.29
9	EBSR	≥53,520	≥53,380	0.55	0.45	0.25	40.92	7.88	50.94	
10	EBSR	≥53,520	≥53,280	0.68	0.54	0.45	28.46	6.56	64.52	
11	EBSR	≥53,520	≥53,290	0.32	0.24	0.44	63.85	7.41	28.30	
12	EBSR	≥53,520	≥53,210	0.42	0.25	0.57	49.66	15.20	34.57	
13	CR	≥53,520	≥53,030	0.45	0.62	0.55	15.08	41.98	34.45	7.58
13	CR	≥53,520	≥53,030	0.47	0.63	0.55	10.03	47.24	34.24	7.57
14	MC	≥53,520	≥51,620	0.53	0.37	3.52				
15	EBSR	≥53,520	≥51,130	0.31	0.35	4.47	66.26	29.26		
16	MS	≥53,520	≥51,370	0.32	0.49	4.02	65.62	30.36		
17	EBSR	≥53,520	≥51,320	0.10	0.14	4.11	85.98	9.90		
18	MC	≥53,520	≥51,400	0.78	0.47	3.93				
19	MC	≥53,520	≥51,340	0.41	0.24	4.03				
20	MC	≥53,520	≥53,340	0.53	0.36	0.32				
21	MC	≥53,520	≥53,350	0.57	0.29	0.30				
22	EBSR	≥53,520	≥53,290	0.14	0.25	0.43	85.42	14.14		
23	MS	≥53,520	≥53,120	0.50	0.35	0.75	49.37	49.88		
24	MC	≥53,520	≥53,330	0.52	0.28	0.35				
25	MC	≥53,520	≥53,320	0.38	0.33	0.34				
26	MC	≥53,520	≥53,120	0.55	0.34	0.72				
27	MS	≥53,520	≥53,130	0.39	0.40	0.72	61.00	38.28		
28	MC	≥53,520	≥53,040	0.39	0.16	0.80				
29	EBSR	≥53,520	≥52,980	0.42	0.54	1.01	57.14	41.86		
30	MC	≥53,520	≥53,350	0.57	0.40	0.30				
31	EBSR	≥53,520	≥53,050	0.67	0.54	0.89	32.36	66.75		
32	MS	≥53,520	≥52,700	0.60	0.46	1.54	39.82	58.65		
33	MC	≥53,520	≥53,150	0.50	0.35	0.53				
34	MC	≥53,520	≥53,180	0.43	0.31	0.60				

Copyright © 2017 by Louisiana Department of Education.

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

ELA Grade 5 Computer-Based Test Administration											
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
1	EBSR	≥1,690	≥1,690	0.65	0.34	0.18	31.48	7.21	61.13		
2	EBSR	≥1,690	≥1,680	0.31	0.33	0.30	51.03	35.62	13.05		
3	EBSR	≥1,690	≥1,680	0.50	0.49	0.35	19.14	60.90	19.61		
4	EBSR	≥1,690	≥1,680	0.74	0.38	0.41	17.72	17.13	64.74		
5	EBSR	≥1,690	≥1,680	0.44	0.57	0.41	46.54	17.60	35.44		
6	EBSR	≥1,690	≥1,680	0.59	0.55	0.47	16.42	48.38	34.73		
7	EBSR	≥1,690	≥1,680	0.31	0.33	0.41	60.07	17.37	22.15		
8	CR	≥1,690	≥1,680	0.22	0.62	0.53	39.16	36.56	18.90	4.13	0.65
8	CR	≥1,690	≥1,680	0.35	0.60	0.53	31.66	37.63	24.34	5.79	
9	EBSR	≥1,690	≥1,690	0.36	0.27	0.00	57.18	12.99	29.83		
10	EBSR	≥1,690	≥1,690	0.23	0.31	0.00	64.32	24.69	10.99		
11	EBSR	≥1,690	≥1,690	0.56	0.44	0.00	36.44	14.41	49.14		
12	EBSR	≥1,690	≥1,690	0.36	0.29	0.06	52.16	24.28	23.51		
13	EBSR	≥1,690	≥1,690	0.26	0.43	0.06	66.21	15.00	18.72		
14	CR	≥1,690	≥1,680	0.18	0.56	0.06	44.24	42.17	11.75	1.36	
14	CR	≥1,690	≥1,680	0.36	0.57	0.06	24.99	45.89	24.81	3.84	
15	MC	≥1,690	≥1,690	0.22	0.12	0.00					
16	MS	≥1,690	≥1,690	0.13	0.36	0.00	86.59	13.41			
17	EBSR	≥1,690	≥1,690	0.74	0.40	0.00	26.34	73.66			
18	MC	≥1,690	≥1,690	0.42	0.19	0.06					
19	TE	≥1,690	≥1,690	0.38	0.48	0.00	62.49	37.51			
20	MC	≥1,690	≥1,690	0.66	0.43	0.00					
21	MC	≥1,690	≥1,690	0.41	0.26	0.00					
22	EBSR	≥1,690	≥1,690	0.25	0.35	0.00	74.66	25.34			
23	MC	≥1,690	≥1,690	0.50	0.31	0.00					
24	MC	≥1,690	≥1,690	0.46	0.40	0.00					
25	TE	≥1,690	≥1,680	0.42	0.36	0.41	57.59	42.00			
26	EBSR	≥1,690	≥1,690	0.37	0.55	0.06	62.91	37.03			
27	MS	≥1,690	≥1,690	0.36	0.44	0.06	64.15	35.79			
28	MC	≥1,690	≥1,690	0.38	0.43	0.06					
29	MC	≥1,690	≥1,690	0.46	0.43	0.12					

Copyright © 2017 by Louisiana Department of Education.

Table 6.7 Item Statistics English Language Arts Grade 5 Paper-Based Test Administration

ELA Grade 5 Paper-Based Test Administration											
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
1	EBSR	≥50,890	≥50,790	0.70	0.34	0.20	26.92	6.03	66.85		
2	EBSR	≥50,890	≥50,710	0.34	0.35	0.34	44.99	41.58	13.09		
3	EBSR	≥50,890	≥50,770	0.54	0.48	0.23	17.02	58.00	24.74		
4	EBSR	≥50,890	≥50,720	0.75	0.38	0.32	17.06	15.73	66.88		
5	EBSR	≥50,890	≥50,730	0.49	0.58	0.31	43.17	14.59	41.93		
6	EBSR	≥50,890	≥50,700	0.65	0.57	0.36	15.12	40.06	44.46		
7	EBSR	≥50,890	≥50,580	0.35	0.31	0.61	54.58	20.38	24.42		
8	CR	≥50,890	≥50,500	0.17	0.51	0.56	47.42	36.84	13.19	1.67	0.12
8	CR	≥50,890	≥50,500	0.29	0.48	0.56	36.00	41.98	18.70	2.57	
9	EBSR	≥50,890	≥50,710	0.41	0.31	0.34	52.18	12.46	35.02		
10	EBSR	≥50,890	≥50,680	0.25	0.29	0.40	62.15	24.32	13.13		
11	EBSR	≥50,890	≥50,710	0.60	0.45	0.35	33.39	12.82	53.44		
12	EBSR	≥50,890	≥50,730	0.38	0.29	0.32	48.13	26.40	25.15		
13	EBSR	≥50,890	≥50,590	0.33	0.44	0.58	59.67	14.53	25.22		
14	CR	≥50,890	≥50,430	0.17	0.53	0.47	44.75	42.10	11.13	1.10	0.02
14	CR	≥50,890	≥50,430	0.35	0.54	0.47	23.73	49.82	22.68	2.86	
15	MC	≥50,890	≥50,660	0.27	0.18	0.43					
16	MS	≥50,890	≥50,630	0.13	0.37	0.51	86.44	13.05			
17	EBSR	≥50,890	≥50,620	0.72	0.40	0.53	27.86	71.61			
18	MC	≥50,890	≥50,560	0.35	0.21	0.59					
19	EBSR	≥50,890	≥50,700	0.33	0.24	0.36	66.87	32.76			
20	MC	≥50,890	≥50,640	0.70	0.41	0.32					
21	MC	≥50,890	≥50,610	0.41	0.29	0.51					
22	EBSR	≥50,890	≥50,680	0.27	0.37	0.40	72.56	27.04			
23	MC	≥50,890	≥50,540	0.53	0.34	0.39					
24	MC	≥50,890	≥50,220	0.45	0.38	0.61					
25	MC	≥50,890	≥50,690	0.62	0.23	0.34					
26	EBSR	≥50,890	≥50,500	0.41	0.54	0.76	58.07	41.17			
27	MS	≥50,890	≥50,600	0.38	0.41	0.56	61.47	37.97			
28	MC	≥50,890	≥50,530	0.39	0.40	0.54					

Copyright © 2017 by Louisiana Department of Education.

Table 6.8 Item Statistics English Language Arts Grade 6 Computer-Based Administration

ELA Grade 6 Computer-Based Test Administration											
Item	Item Type	Total N	Adj. N	*p*-value	Pbis Rit	Omit Rate	% at 0	% at 1	% at 2	% at 3	% at 4
1	EBSR	≥4,810	≥4,800	0.59	0.52	0.17	26.54	28.60	44.69		
2	EBSR	≥4,810	≥4,800	0.59	0.43	0.21	36.81	7.21	55.77		
3	EBSR	≥4,810	≥4,800	0.44	0.35	0.19	36.33	38.54	24.94		
4	EBSR	≥4,810	≥4,790	0.56	0.53	0.37	32.84	21.26	45.52		
5	EBSR	≥4,810	≥4,790	0.26	0.20	0.39	67.16	13.39	19.06		
6	EBSR	≥4,810	≥4,790	0.57	0.48	0.33	33.90	17.67	48.10		
7	EBSR	≥4,810	≥4,790	0.40	0.32	0.35	43.69	31.32	24.63		
8	CR	≥4,810	≥4,730	0.38	0.68	0.62	15.86	31.53	36.50	13.26	1.25
8	CR	≥4,810	≥4,730	0.51	0.69	0.62	15.51	31.07	37.37	14.45	
9	EBSR	≥4,810	≥4,800	0.61	0.44	0.12	21.97	33.51	44.40		
10	EBSR	≥4,810	≥4,800	0.37	0.36	0.15	51.15	23.84	24.86		
11	EBSR	≥4,810	≥4,800	0.34	0.38	0.12	53.21	24.53	22.14		
12	EBSR	≥4,810	≥4,800	0.58	0.45	0.12	29.85	23.99	46.04		
13	EBSR	≥4,810	≥4,800	0.46	0.41	0.12	43.42	20.70	35.75		
14	CR	≥4,810	≥4,780	0.31	0.68	0.31	22.64	38.87	30.18	7.34	0.44
14	CR	≥4,810	≥4,780	0.42	0.67	0.31	18.42	43.86	31.20	5.99	
15	TE	≥4,810	≥4,770	0.53	0.32	0.79	47.02	52.19			
16	MC	≥4,790	≥4,790	0.85	0.41	0.08					
17	EBSR	≥4,810	≥4,790	0.70	0.44	0.31	29.60	70.09			
18	MC	≥4,790	≥4,790	0.65	0.33	0.10					
19	MC	≥4,790	≥4,790	0.81	0.40	0.13					
20	MC	≥4,790	≥4,790	0.54	0.39	0.08					
21	MC	≥4,790	≥4,790	0.81	0.44	0.06					
22	TE	≥4,810	≥4,770	0.20	0.47	0.69	79.51	19.81			
23	EBSR	≥4,810	≥4,790	0.27	0.22	0.33	72.73	26.94			
24	MC	≥4,790	≥4,790	0.75	0.51	0.17					
25	MS	≥4,810	≥4,790	0.23	0.06	0.42	76.95	22.64			
26	MC	≥4,790	≥4,780	0.56	0.50	0.21					
27	EBSR	≥4,810	≥4,790	0.27	0.41	0.44	72.52	27.04			
28	MC	≥4,790	≥4,780	0.52	0.35	0.21					
29	MC	≥4,790	≥4,780	0.57	0.44	0.21					

Copyright © 2017 by Louisiana Department of Education.

Table 6.9 Item Statistics English Language Arts Grade 6 Paper-Based Test Administration

ELA Grade 6 Paper-Based Test Administration											
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
1	EBSR	≥47,840	≥47,730	0.57	0.53	0.24	29.96	25.97	43.83		
2	EBSR	≥47,840	≥47,680	0.63	0.41	0.34	33.19	7.20	59.28		
3	EBSR	≥47,840	≥47,690	0.43	0.32	0.31	37.03	38.94	23.72		
4	EBSR	≥47,840	≥47,640	0.58	0.48	0.43	31.22	22.13	46.22		
5	EBSR	≥47,840	≥47,570	0.26	0.19	0.58	66.34	13.91	19.17		
6	EBSR	≥47,840	≥47,560	0.59	0.46	0.58	30.27	21.15	48.00		
7	EBSR	≥47,840	≥47,490	0.42	0.32	0.74	40.30	34.06	24.89		
8	CR	≥47,840	≥46,780	0.36	0.63	1.09	15.69	33.27	40.48	8.10	0.24
8	CR	≥47,840	≥46,780	0.48	0.63	1.09	15.56	33.08	40.66	8.48	
9	EBSR	≥47,840	≥47,710	0.58	0.44	0.28	25.06	34.54	40.12		
10	EBSR	≥47,840	≥47,700	0.35	0.33	0.31	52.26	25.03	22.40		
11	EBSR	≥47,840	≥47,680	0.31	0.38	0.34	58.86	20.17	20.63		
12	EBSR	≥47,840	≥47,710	0.54	0.44	0.28	33.70	24.20	41.82		
13	EBSR	≥47,840	≥47,630	0.46	0.40	0.45	42.31	23.59	33.66		
14	CR	≥47,840	≥47,420	0.28	0.63	0.60	23.05	43.44	28.96	3.55	0.11
14	CR	≥47,840	≥47,420	0.40	0.63	0.60	15.53	50.16	30.34	3.08	
15	MC	≥47,840	≥47,660	0.55	0.36	0.36					
16	MC	≥47,840	≥47,720	0.86	0.42	0.23					
17	EBSR	≥47,840	≥47,680	0.69	0.40	0.35	30.85	68.80			
18	MC	≥47,840	≥47,390	0.60	0.31	0.93					
19	MC	≥47,840	≥47,620	0.81	0.40	0.44					
20	MC	≥47,840	≥47,420	0.58	0.40	0.81					
21	MC	≥47,840	≥47,590	0.79	0.43	0.52					
22	MC	≥47,840	≥47,600	0.68	0.43	0.48					
23	EBSR	≥47,840	≥47,640	0.26	0.18	0.43	74.14	25.43			
24	MC	≥47,840	≥47,610	0.75	0.51	0.46					
25	MS	≥47,840	≥47,360	0.22	0.04	1.01	77.35	21.64			
26	MC	≥47,840	≥47,440	0.57	0.50	0.72					
27	EBSR	≥47,840	≥47,630	0.28	0.40	0.45	71.55	28.01			
28	MC	≥47,840	≥47,560	0.52	0.33	0.57					
29	MC	≥47,840	≥47,600	0.58	0.44	0.48					

Copyright © 2017 by Louisiana Department of Education.

Table 6.10 Item Statistics English Language Arts Grade 7 Computer-Based Test Administration

ELA Grade 7 Computer-Based Test Administration											
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
1	EBSR	≥6,310	≥6,300	0.73	0.29	0.06	17.54	19.67	62.73		
2	EBSR	≥6,310	≥6,300	0.45	0.28	0.14	51.16	6.70	41.99		
3	EBSR	≥6,310	≥6,300	0.24	0.29	0.08	71.38	9.17	19.37		
4	EBSR	≥6,310	≥6,300	0.37	0.32	0.08	53.98	18.52	27.42		
5	EBSR	≥6,310	≥6,300	0.29	0.27	0.11	59.32	23.81	16.76		
6	EBSR	≥6,310	≥6,300	0.33	0.36	0.14	61.16	11.07	27.63		
7	EBSR	≥6,310	≥6,300	0.43	0.39	0.14	46.41	20.35	33.09		
8	CR	≥6,310	≥6,270	0.38	0.66	0.41	17.71	30.81	36.21	12.81	1.87
8	CR	≥6,310	≥6,270	0.53	0.64	0.41	13.78	30.46	37.87	17.30	
9	EBSR	≥6,310	≥6,300	0.75	0.26	0.11	23.29	2.68	73.93		
10	EBSR	≥6,310	≥6,300	0.59	0.44	0.11	36.70	9.11	54.08		
11	EBSR	≥6,310	≥6,300	0.55	0.30	0.11	40.46	8.90	50.53		
12	EBSR	≥6,310	≥6,300	0.42	0.33	0.11	53.78	7.98	38.13		
13	EBSR	≥6,310	≥6,300	0.23	0.31	0.11	73.07	7.67	19.15		
14	CR	≥6,310	≥6,280	0.36	0.68	0.38	15.97	37.76	34.41	9.65	1.79
14	CR	≥6,310	≥6,280	0.50	0.69	0.38	14.03	35.31	37.87	12.36	
15	MS	≥6,310	≥6,290	0.25	0.23	0.27	75.21	24.52			
16	EBSR	≥6,310	≥6,290	0.55	0.44	0.25	45.27	54.47			
18	MC	≥6,290	≥6,290	0.62	0.29	0.08					
19	MC	≥6,290	≥6,290	0.46	0.30	0.08					
20	MS	≥6,310	≥6,290	0.38	0.41	0.29	61.49	38.22			
21	TE	≥6,310	≥6,240	0.74	0.33	1.09	25.93	72.98			
22	MC	≥6,290	≥6,290	0.82	0.35	0.08					
23	MC	≥6,290	≥6,290	0.59	0.25	0.06					
24	EBSR	≥6,310	≥6,290	0.54	0.47	0.25	45.92	53.83			
25	TE	≥6,310	≥6,250	0.56	0.39	0.93	43.81	55.25			
26	EBSR	≥6,310	≥6,290	0.32	0.37	0.25	67.73	32.01			
27	MC	≥6,290	≥6,290	0.52	0.39	0.08					
28	MC	≥6,290	≥6,290	0.35	0.33	0.08					
29	MC	≥6,290	≥6,280	0.42	0.35	0.14					

Copyright © 2017 by Louisiana Department of Education.

Table 6.11 Item Statistics English Language Arts Grade 7 Paper-Based Test Administration

ELA Grade 7 Paper-Based Test Administration											
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
1	EBSR	≥44,650	≥44,540	0.69	0.32	0.23	20.82	19.46	59.49		
2	EBSR	≥44,650	≥44,470	0.49	0.32	0.40	48.79	5.00	45.81		
3	EBSR	≥44,650	≥44,410	0.21	0.22	0.54	74.82	8.50	16.14		
4	EBSR	≥44,650	≥44,480	0.36	0.31	0.37	53.72	19.23	26.68		
5	EBSR	≥44,650	≥44,400	0.29	0.27	0.56	59.11	23.42	16.92		
6	EBSR	≥44,650	≥44,320	0.32	0.32	0.73	62.15	11.03	26.09		
7	EBSR	≥44,650	≥44,290	0.47	0.37	0.80	42.05	20.38	36.77		
8	CR	≥44,650	≥43,970	0.37	0.62	1.22	14.42	34.11	39.50	9.98	0.49
8	CR	≥44,650	≥43,970	0.51	0.62	1.22	11.62	33.54	42.00	11.33	
9	EBSR	≥44,650	≥44,520	0.79	0.25	0.29	19.49	2.81	77.41		
10	EBSR	≥44,650	≥44,440	0.61	0.42	0.46	34.28	8.99	56.27		
11	EBSR	≥44,650	≥44,480	0.57	0.32	0.37	38.39	9.29	51.95		
12	EBSR	≥44,650	≥44,430	0.42	0.31	0.49	53.30	8.03	38.18		
13	EBSR	≥44,650	≥44,380	0.21	0.27	0.60	75.04	7.59	16.77		
14	CR	≥44,650	≥44,230	0.34	0.66	0.79	16.64	38.58	35.19	8.15	0.50
14	CR	≥44,650	≥44,230	0.50	0.66	0.79	12.99	34.26	40.87	10.95	
15	MS	≥44,650	≥44,460	0.23	0.22	0.42	76.81	22.77			
16	EBSR	≥44,650	≥44,450	0.56	0.45	0.45	44.20	55.36			
18	MC	≥44,650	≥44,290	0.59	0.31	0.63					
19	MC	≥44,650	≥44,270	0.45	0.30	0.66					
20	MS	≥44,650	≥44,370	0.40	0.42	0.62	59.91	39.46			
21	MC	≥44,650	≥44,410	0.70	0.31	0.51					
22	MC	≥44,650	≥44,340	0.83	0.32	0.67					
23	MC	≥44,650	≥44,400	0.59	0.23	0.52					
24	EBSR	≥44,650	≥44,230	0.55	0.46	0.93	44.13	54.93			
25	MC	≥44,650	≥44,350	0.61	0.39	0.63					
26	EBSR	≥44,650	≥44,430	0.32	0.36	0.48	67.17	32.34			
27	MC	≥44,650	≥44,410	0.58	0.40	0.50					
28	MC	≥44,650	≥44,400	0.35	0.27	0.52					
29	MC	≥44,650	≥44,450	0.46	0.34	0.42					

Copyright © 2017 by Louisiana Department of Education.

Table 6.12 Item Statistics English Language Arts Grade 8 Computer-Based Test Administration

ELA Grade 8 Computer-Based Test Administration											
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
1	EBSR	≥5,450	≥5,450	0.36	0.32	0.09	41.35	45.00	13.56		
2	EBSR	≥5,450	≥5,450	0.47	0.33	0.09	49.54	6.41	43.95		
3	EBSR	≥5,450	≥5,450	0.46	0.32	0.07	47.42	12.50	40.01		
4	EBSR	≥5,450	≥5,440	0.67	0.47	0.15	26.98	11.05	61.82		
5	EBSR	≥5,450	≥5,450	0.74	0.41	0.11	23.94	3.28	72.67		
6	EBSR	≥5,450	≥5,440	0.72	0.49	0.20	24.67	5.66	69.46		
7	EBSR	≥5,450	≥5,440	0.26	0.28	0.15	71.77	5.21	22.87		
8	CR	≥5,450	≥5,430	0.47	0.67	0.29	8.28	25.29	41.26	19.12	5.57
8	CR	≥5,450	≥5,430	0.58	0.68	0.29	9.29	27.11	41.90	21.22	
9	EBSR	≥5,450	≥5,450	0.64	0.44	0.04	3.57	63.84	32.55		
10	EBSR	≥5,450	≥5,450	0.62	0.47	0.09	24.10	26.89	48.92		
11	EBSR	≥5,450	≥5,450	0.24	0.24	0.05	68.66	14.44	16.84		
12	EBSR	≥5,450	≥5,450	0.65	0.39	0.07	29.36	11.25	59.31		
13	CR	≥5,450	≥5,400	0.52	0.69	0.82	8.27	18.93	37.99	22.65	11.16
13	CR	≥5,450	≥5,400	0.60	0.70	0.82	10.34	25.16	38.32	25.18	
14	MC	≥5,450	≥5,430	0.52	0.13	0.44					
15	MC	≥5,450	≥5,430	0.54	0.40	0.40					
16	MC	≥5,450	≥5,420	0.82	0.49	0.48					
17	EBSR	≥5,450	≥5,420	0.51	0.44	0.49	48.48	51.03			
18	MS	≥5,450	≥5,420	0.21	0.36	0.64	78.08	21.28			
19	MC	≥5,450	≥5,410	0.42	0.20	0.64					
20	MS	≥5,450	≥5,440	0.69	0.36	0.20	30.92	68.88			
21	TE	≥5,450	≥5,430	0.46	0.37	0.37	54.03	45.60			
22	MC	≥5,450	≥5,440	0.85	0.42	0.11					
23	MC	≥5,440	≥5,440	0.79	0.47	0.04					
24	EBSR	≥5,450	≥5,440	0.72	0.53	0.20	27.84	71.96			
25	MC	≥5,440	≥5,440	0.85	0.45	0.04					
26	MC	≥5,440	≥5,440	0.62	0.32	0.07					
27	MC	≥5,440	≥5,440	0.30	0.26	0.06					
28	MS	≥5,450	≥5,440	0.22	0.39	0.24	77.60	22.16			
29	TE	≥5,450	≥5,440	0.67	0.52	0.29	32.94	66.77			
30	TE	≥5,450	≥5,440	0.08	0.04	0.27	91.29	8.43			
31	EBSR	≥5,450	≥5,440	0.17	0.34	0.26	82.72	17.03			
32	MC	≥5,440	≥5,440	0.67	0.37	0.09					
33	MC	≥5,440	≥5,430	0.57	0.35	0.17					
34	MC	≥5,440	≥5,430	0.83	0.46	0.13					

Copyright © 2017 by Louisiana Department of Education.

Table 6.13 Item Statistics English Language Arts Grade 8 Paper-Based Test Administration

ELA Grade 8 Paper-Based Test Administration											
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**
1	EBSR	≥45,780	≥45,650	0.35	0.35	0.27	41.93	45.55	12.24		
2	EBSR	≥45,780	≥45,620	0.49	0.30	0.34	46.86	7.46	45.34		
3	EBSR	≥45,780	≥45,610	0.45	0.34	0.37	48.79	11.04	39.80		
4	EBSR	≥45,780	≥45,650	0.64	0.45	0.28	30.80	9.44	59.47		
5	EBSR	≥45,780	≥45,620	0.75	0.41	0.35	23.65	2.46	73.54		
6	EBSR	≥45,780	≥45,590	0.72	0.46	0.41	24.64	5.62	69.33		
7	EBSR	≥45,780	≥45,530	0.27	0.27	0.53	69.88	5.37	24.21		
8	CR	≥45,780	≥45,360	0.47	0.61	0.78	6.05	23.53	48.26	19.39	1.86
8	CR	≥45,780	≥45,360	0.62	0.62	0.78	5.91	24.01	48.15	21.03	
9	EBSR	≥45,780	≥45,680	0.62	0.42	0.22	4.15	66.77	28.87		
10	EBSR	≥45,780	≥45,620	0.60	0.50	0.34	26.40	26.62	46.63		
11	EBSR	≥45,780	≥45,640	0.26	0.27	0.29	66.44	14.09	19.18		
12	EBSR	≥45,780	≥45,510	0.65	0.38	0.57	29.11	10.41	59.90		
13	CR	≥45,780	≥45,090	0.51	0.66	1.35	7.38	16.34	44.31	25.50	4.96
13	CR	≥45,780	≥45,090	0.65	0.67	1.35	7.75	17.44	44.89	28.41	
14	MC	≥45,780	≥43,590	0.54	0.15	4.76					
15	MC	≥45,780	≥43,600	0.57	0.43	4.74					
16	MC	≥45,780	≥43,570	0.77	0.51	4.80					
17	EBSR	≥45,780	≥43,450	0.50	0.46	5.09	47.26	47.65			
18	MS	≥45,780	≥43,190	0.20	0.35	5.66	75.52	18.82			
19	MC	≥45,780	≥43,070	0.38	0.15	5.72					
20	MS	≥45,780	≥45,570	0.69	0.35	0.45	31.25	68.29			
21	MC	≥45,780	≥45,560	0.76	0.47	0.47					
22	MC	≥45,780	≥45,550	0.84	0.43	0.47					
23	MC	≥45,780	≥45,520	0.77	0.47	0.54					
24	EBSR	≥45,780	≥45,560	0.72	0.53	0.47	27.48	72.05			
25	MC	≥45,780	≥45,600	0.85	0.44	0.35					
26	MC	≥45,780	≥45,530	0.61	0.29	0.50					
27	MC	≥45,780	≥45,470	0.30	0.24	0.64					
28	MS	≥45,780	≥45,530	0.23	0.40	0.55	76.65	22.81			
29	MC	≥45,780	≥45,530	0.45	0.43	0.48					
30	MC	≥45,780	≥45,560	0.66	0.26	0.44					
31	EBSR	≥45,780	≥45,570	0.15	0.33	0.45	84.37	15.18			
32	MC	≥45,780	≥45,550	0.66	0.39	0.46					
33	MC	≥45,780	≥45,560	0.58	0.37	0.42					

Copyright © 2017 by Louisiana Department of Education.

Table 6.14 Item Statistics Mathematics Grade 3 Paper-Based Test Administration

Mathematics Grade 3 Paper-Based Test Administration													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
1	ESR	≥57,130	≥56,690	0.79	0.40	0.78	20.47	78.75					
2	SA,GR	≥57,130	≥56,380	0.79	0.34	1.32	20.40	78.28					
3	MC	≥57,130	≥56,510	0.45	0.52	1.03							
4	MC	≥57,130	≥56,780	0.86	0.30	0.57							
5	MC	≥57,130	≥56,830	0.72	0.45	0.48							
6	MC	≥57,130	≥56,910	0.91	0.27	0.35							
7	MC	≥57,130	≥56,660	0.27	0.34	0.71							
8	MC	≥57,130	≥56,670	0.95	0.15	0.69							
9	MPSR	≥57,130	≥56,980	0.52	0.58	0.27	26.84	42.74	30.14				
10	MC	≥57,130	≥56,710	0.56	0.45	0.60							
11	MC	≥57,130	≥56,570	0.25	0.38	0.65							
12	SA,GR	≥57,130	≥56,650	0.58	0.59	0.84	41.36	57.80					
13	MC	≥57,130	≥56,710	0.28	0.25	0.54							
14	CR	≥57,130	≥55,770	0.14	0.53	2.15	70.75	19.48	1.24	6.15			
15	CR	≥57,130	≥56,520	0.06	0.42	0.97	87.23	3.90	4.14	3.20	0.46		
16	MC	≥57,130	≥56,990	0.89	0.24	0.21							
17	MC	≥57,130	≥56,860	0.74	0.41	0.43							
18	MC	≥57,130	≥56,020	0.78	0.41	1.91							
19	MC	≥57,130	≥56,860	0.66	0.39	0.40							
20	ESR	≥57,130	≥56,030	0.41	0.42	1.93	58.11	39.96					
21	MC	≥57,130	≥56,430	0.95	0.27	0.94							
22	MC	≥57,130	≥55,420	0.85	0.30	0.66							
23	SA,GR	≥57,130	≥56,330	0.44	0.60	1.41	55.26	43.33					
24	MC	≥57,130	≥56,380	0.54	0.44	1.27							
25	SA,GR	≥57,130	≥56,550	0.55	0.35	1.02	44.35	54.62					
26	MC	≥57,130	≥56,610	0.67	0.44	0.61							
27	MC	≥57,130	≥56,790	0.75	0.50	0.52							
28	MC	≥57,130	≥56,380	0.51	0.45	1.14							
29	MC	≥57,130	≥56,660	0.17	0.30	0.69							
30	ESR	≥57,130	≥57,010	0.54	0.65	0.21	23.64	20.91	23.63	31.61			
31	CR	≥57,130	≥52,450	0.28	0.60	7.96	47.91	18.56	16.56	8.78			
32	MC	≥57,130	≥56,800	0.73	0.35	0.55							
33	MC	≥57,130	≥56,840	0.61	0.36	0.39							
34	MC	≥57,130	≥56,650	0.45	0.46	0.78							
35	MC	≥57,130	≥56,660	0.78	0.40	0.78							
36	MC	≥57,130	≥56,500	0.47	0.25	1.06							
37	MC	≥57,130	≥56,720	0.73	0.44	0.65							
38	SA,GR	≥57,130	≥56,320	0.64	0.55	1.43	35.69	62.88					
39	ESR	≥57,130	≥56,320	0.25	0.47	1.42	74.10	24.48					

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 3 Paper-Based Test Administration (continued)													
Item	**Item Type**	**Total N**	**Adj. N**	**p-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
40	MC	≥57,130	≥56,750	0.55	0.42	0.63							
41	MC	≥57,130	≥56,570	0.84	0.42	0.85							
42	MC	≥57,130	≥56,780	0.97	0.19	0.40							
43	MS	≥57,130	≥55,440	0.12	0.39	2.96	85.37	11.68					
44	CR	≥57,130	≥56,870	0.37	0.74	0.45	32.29	16.80	12.78	7.04	8.45	7.39	14.79
45	CR	≥57,130	≥56,250	0.34	0.44	1.39	35.15	33.98	20.73	8.58			

Copyright © 2017 by Louisiana Department of Education.

Table 6.15 Item Statistics Mathematics Grade 4 Computer-Based Test Administration

Mathematics Grade 4 Computer-Based Test Administration												
Item	Item Type	Total N	Adj. N	*p*-value	Pbis Rit	Omit Rate	% at 0	% at 1	% at 2	% at 3	% at 4	% at 5
1	MC	≥450	≥450	0.89	0.32	0.00						
2	SA,GR	≥450	≥450	0.73	0.25	0.00	27.41	72.59				
3	MC	≥450	≥450	0.66	0.38	0.00						
4	MS	≥450	≥450	0.36	0.51	0.00	63.82	36.18				
5	MC	≥450	≥450	0.59	0.42	0.00						
6	MC	≥450	≥450	0.38	0.54	0.00						
7	MC	≥450	≥450	0.79	0.45	0.00						
8	MC	≥450	≥450	0.63	0.52	0.00						
9	MS	≥450	≥450	0.65	0.42	0.00	35.09	64.91				
10	TE	≥450	≥450	0.79	0.33	0.00	21.27	78.73				
11	MC	≥450	≥450	0.79	0.39	0.00						
12	MC	≥450	≥450	0.92	0.32	0.00						
13	MC	≥450	≥450	0.71	0.43	0.00						
14	MC	≥450	≥450	0.38	0.35	0.00						
15	CR	≥450	≥450	0.37	0.66	0.22	34.65	17.11	25.00	10.96	12.06	
16	CR	≥450	≥440	0.20	0.58	1.97	61.62	18.42	12.06	5.70		
17	MC	≥450	≥450	0.72	0.34	0.22						
18	ESR	≥450	≥450	0.62	0.60	0.00	19.74	35.96	44.30			
19	MC	≥450	≥450	0.39	0.37	0.22						
20	MC	≥450	≥450	0.45	0.53	0.00						
21	MC	≥450	≥450	0.72	0.43	0.22						
22	MC	≥450	≥450	0.63	0.45	0.00						
23	MC	≥450	≥450	0.46	0.49	0.00						
24	MC	≥450	≥450	0.52	0.33	0.00						
25	MC	≥450	≥450	0.67	0.49	0.22						
26	MC	≥450	≥450	0.78	0.31	0.00						
27	MC	≥450	≥450	0.32	0.45	0.00						
28	MC	≥450	≥450	0.42	0.39	0.00						
29	MS	≥450	≥450	0.76	0.37	0.00	24.12	75.88				
30	CR	≥450	≥450	0.28	0.49	1.10	39.04	38.38	19.74	1.54		
31	MPSR	≥450	≥450	0.42	0.67	0.00	28.95	31.14	23.90	16.01		
32	SA,GR	≥450	≥450	0.93	0.17	0.00	6.80	93.20				
33	MC	≥450	≥450	0.37	0.41	0.00						
34	MC	≥450	≥450	0.83	0.27	0.00						
35	MC	≥450	≥450	0.93	0.26	0.00						
36	MPSR	≥450	≥450	0.49	0.57	0.00	41.01	19.96	39.04			
37	MC	≥450	≥450	0.30	0.08	0.00						
38	MC	≥450	≥450	0.69	0.35	0.00						
39	SA,GR	≥450	≥450	0.61	0.55	0.00	38.60	61.40				

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 4 Computer-Based Test Administration (continued)												
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**
40	MC	≥450	≥450	0.95	0.12	0.00						
41	MC	≥450	≥450	0.66	0.51	0.22						
42	MC	≥450	≥450	0.31	0.23	0.00						
43	CR	≥450	≥450	0.24	0.59	0.22	63.60	11.18	14.04	10.53		

Copyright © 2017 by Louisiana Department of Education.

Table 6.16 Item Statistics Mathematics Grade 4 Paper-Based Test Administration

Mathematics Grade 4 Paper-Based Test Administration													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
1	MC	≥53,530	≥53,280	0.86	0.35	0.41							
2	SA,GR	≥53,530	≥52,690	0.68	0.35	1.57	31.32	67.11					
3	MC	≥53,530	≥53,320	0.62	0.39	0.36							
4	MS	≥53,530	≥53,210	0.31	0.50	0.60	68.11	31.29					
5	MC	≥53,530	≥52,600	0.55	0.41	0.73							
6	MC	≥53,530	≥53,250	0.54	0.56	0.50							
7	MC	≥53,530	≥53,070	0.79	0.43	0.67							
8	MC	≥53,530	≥52,680	0.63	0.51	1.53							
9	MS	≥53,530	≥53,170	0.62	0.33	0.66	37.45	61.88					
10	MC	≥53,530	≥53,060	0.83	0.37	0.52							
11	MC	≥53,530	≥53,160	0.80	0.45	0.66							
12	MC	≥53,530	≥53,340	0.78	0.41	0.29							
13	MC	≥53,530	≥53,320	0.78	0.42	0.36							
14	MC	≥53,530	≥53,000	0.41	0.38	0.98							
15	CR	≥53,530	≥53,070	0.40	0.65	0.79	32.81	15.51	24.15	12.30	14.38		
16	CR	≥53,530	≥51,680	0.17	0.58	3.32	65.73	16.96	10.54	3.32			
17	MC	≥53,530	≥53,120	0.71	0.35	0.75							
18	ESR	≥53,530	≥53,270	0.63	0.59	0.49	19.20	35.60	44.71				
19	MC	≥53,530	≥52,750	0.36	0.34	0.44							
20	MC	≥53,530	≥53,300	0.42	0.53	0.39							
21	MC	≥53,530	≥52,700	0.74	0.41	1.54							
22	MC	≥53,530	≥53,290	0.56	0.40	0.39							
23	MC	≥53,530	≥53,140	0.43	0.53	0.71							
24	MC	≥53,530	≥53,210	0.44	0.32	0.57							
25	MC	≥53,530	≥53,130	0.65	0.49	0.34							
26	MC	≥53,530	≥53,290	0.79	0.32	0.40							
27	MC	≥53,530	≥53,310	0.32	0.42	0.36							
28	MC	≥53,530	≥53,210	0.43	0.34	0.51							
29	MS	≥53,530	≥53,190	0.80	0.39	0.64	20.10	79.26					
30	CR	≥53,530	≥52,950	0.25	0.57	0.95	44.56	35.28	18.90	0.18			
31	MPSR	≥53,530	≥53,280	0.39	0.63	0.46	28.64	33.25	29.12	8.52			
32	SA,GR	≥53,530	≥53,020	0.87	0.35	0.96	13.07	85.97					
33	MC	≥53,530	≥53,310	0.39	0.41	0.38							
34	MC	≥53,530	≥53,000	0.80	0.31	0.98							
35	MC	≥53,530	≥53,150	0.81	0.29	0.42							
36	MPSR	≥53,530	≥53,090	0.61	0.57	0.81	29.55	19.05	50.59				
37	MC	≥53,530	≥53,000	0.26	0.12	0.96							
38	MC	≥53,530	≥53,330	0.66	0.33	0.34							

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 4 Paper-Based Test Administration (continued)													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
39	SA,GR	≥53,530	≥52,560	0.56	0.57	1.81	43.36	54.83					
40	MC	≥53,530	≥53,150	0.92	0.27	0.70							
41	MC	≥53,530	≥53,270	0.61	0.57	0.41							
42	MC	≥53,530	≥53,030	0.30	0.24	0.79							
43	CR	≥53,530	≥52,640	0.24	0.58	1.29	61.23	14.12	12.14	10.85			
44	MPSR	≥53,530	≥53,390	0.47	0.70	0.27	12.42	20.17	10.46	10.81	32.14	4.12	9.60

Copyright © 2017 by Louisiana Department of Education.

Table 6.17 Item Statistics Mathematics Grade 5 Computer-Based Test Administration

Mathematics Grade 5 Computer-Based Test Administration													
Item	Item Type	Total N	Adj. N	p-value	Pbis Rit	Omit Rate	% at 0	% at 1	% at 2	% at 3	% at 4	% at 5	% at 6
1	MC	≥1,690	≥1,690	0.90	0.29	0.06							
2	MC	≥1,690	≥1,690	0.86	0.35	0.06							
3	MC	≥1,690	≥1,680	0.50	0.56	0.18							
4	SA,GR	≥1,690	≥1,690	0.21	0.51	0.12	78.72	21.16					
5	ESR	≥1,690	≥1,680	0.49	0.58	0.18	32.98	36.52	30.32				
6	MC	≥1,690	≥1,690	0.61	0.34	0.12							
7	MC	≥1,690	≥1,690	0.78	0.37	0.12							
8	MC	≥1,690	≥1,680	0.72	0.49	0.18							
9	MC	≥1,690	≥1,690	0.66	0.21	0.06							
10	MC	≥1,690	≥1,680	0.40	0.21	0.18							
11	TE	≥1,690	≥1,690	0.09	0.24	0.06	90.78	9.16					
12	MC	≥1,690	≥1,680	0.63	0.45	0.18							
13	MC	≥1,690	≥1,690	0.41	0.49	0.12							
14	MPSR	≥1,690	≥1,680	0.25	0.63	0.12	56.62	13.83	12.17	7.33	9.87		
15	CR	≥1,690	≥1,660	0.06	0.41	1.00	90.01	3.72	1.42	3.43			
16	MC	≥1,690	≥1,690	0.87	0.18	0.00							
17	MS	≥1,690	≥1,690	0.29	0.35	0.06	71.22	28.72					
18	MC	≥1,690	≥1,690	0.53	0.32	0.00							
19	MC	≥1,690	≥1,690	0.60	0.40	0.00							
20	MC	≥1,690	≥1,690	0.56	0.24	0.12							
21	MC	≥1,690	≥1,690	0.71	0.44	0.06							
22	MC	≥1,690	≥1,690	0.71	0.28	0.00							
23	MC	≥1,690	≥1,680	0.72	0.35	0.18							
24	MC	≥1,690	≥1,690	0.72	0.31	0.12							
25	SA,GR	≥1,690	≥1,690	0.50	0.18	0.00	49.94	50.06					
26	MC	≥1,690	≥1,690	0.61	0.10	0.00							
27	MC	≥1,690	≥1,690	0.28	0.50	0.12							
28	MC	≥1,690	≥1,680	0.60	0.43	0.18							
29	MC	≥1,690	≥1,690	0.46	0.52	0.00							
30	CR	≥1,690	≥1,680	0.38	0.64	0.06	37.41	24.29	24.17	13.77			
31	CR	≥1,690	≥1,670	0.22	0.62	0.41	64.01	16.84	6.50	11.82			
32	SA,GR	≥1,690	≥1,690	0.64	0.28	0.00	35.58	64.42					
33	MC	≥1,690	≥1,690	0.95	0.24	0.00							
34	MC	≥1,690	≥1,690	0.31	0.29	0.00							
35	MC	≥1,690	≥1,690	0.43	0.34	0.12							
36	MC	≥1,690	≥1,690	0.41	0.55	0.00							
37	MC	≥1,690	≥1,680	0.48	0.57	0.18							
38	MC	≥1,690	≥1,690	0.29	0.23	0.06							

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 5 Computer-Based Test Administration (continued)													
Item	**Item Type**	**Total N**	**Adj. N**	**p-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
39	ESR	≥1,690	≥1,690	0.45	0.61	0.00	34.81	40.84	24.35				
40	MC	≥1,690	≥1,690	0.50	0.52	0.06							
41	MC	≥1,690	≥1,690	0.80	0.40	0.00							
42	MC	≥1,690	≥1,690	0.55	0.49	0.06							
43	CR	≥1,690	≥1,680	0.18	0.47	0.12	62.59	25.65	6.80	4.73			
44	CR	≥1,690	≥1,690	0.33	0.66	0.06	20.09	21.87	22.58	16.49	12.17	6.03	0.71

Copyright © 2017 by Louisiana Department of Education.

Table 6.18 Item Statistics Mathematics Grade 5 Paper-Based Test Administration

Mathematics Grade 5 Paper-Based Test Administration													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
1	MC	≥50,880	≥50,680	0.88	0.34	0.38							
2	MC	≥50,880	≥50,730	0.86	0.34	0.24							
3	MC	≥50,880	≥50,660	0.47	0.54	0.41							
4	SA,GR	≥50,880	≥50,000	0.24	0.49	1.73	74.66	23.62					
5	ESR	≥50,880	≥50,590	0.49	0.57	0.58	31.23	38.32	29.88				
6	MC	≥50,880	≥50,590	0.62	0.36	0.48							
7	MC	≥50,880	≥50,510	0.75	0.36	0.70							
8	MC	≥50,880	≥50,520	0.70	0.52	0.69							
9	MC	≥50,880	≥50,450	0.64	0.25	0.81							
10	MC	≥50,880	≥50,520	0.41	0.22	0.67							
11	MC	≥50,880	≥50,360	0.25	0.18	0.72							
12	MC	≥50,880	≥50,630	0.60	0.46	0.46							
13	MC	≥50,880	≥50,600	0.43	0.46	0.47							
14	MPSR	≥50,880	≥50,560	0.26	0.59	0.59	54.34	15.11	12.55	8.07	9.30		
15	CR	≥50,880	≥48,780	0.06	0.41	3.85	86.30	4.54	1.19	3.83			
16	MC	≥50,880	≥50,760	0.89	0.15	0.22							
17	MS	≥50,880	≥50,700	0.30	0.34	0.36	70.18	29.45					
18	MC	≥50,880	≥50,560	0.55	0.33	0.49							
19	MC	≥50,880	≥50,670	0.59	0.42	0.38							
20	MC	≥50,880	≥50,600	0.57	0.31	0.50							
21	MC	≥50,880	≥50,450	0.71	0.42	0.78							
22	MC	≥50,880	≥50,590	0.70	0.26	0.54							
23	MC	≥50,880	≥50,590	0.72	0.42	0.54							
24	MC	≥50,880	≥50,540	0.72	0.34	0.48							
25	SA,GR	≥50,880	≥49,860	0.48	0.24	2.00	51.13	46.87					
26	MC	≥50,880	≥50,620	0.61	0.13	0.45							
27	MC	≥50,880	≥50,480	0.24	0.47	0.75							
28	MC	≥50,880	≥50,650	0.60	0.43	0.43							
29	MC	≥50,880	≥50,580	0.44	0.50	0.53							
30	CR	≥50,880	≥50,590	0.47	0.63	0.48	27.60	22.80	29.33	19.68			
31	CR	≥50,880	≥50,130	0.27	0.60	1.26	54.30	22.70	6.50	15.04			
32	SA,GR	≥50,880	≥50,040	0.61	0.33	1.65	37.97	60.38					
33	MC	≥50,880	≥50,600	0.93	0.27	0.43							
34	MC	≥50,880	≥48,530	0.32	0.23	4.60							
35	MC	≥50,880	≥50,640	0.49	0.33	0.45							
36	MC	≥50,880	≥50,480	0.40	0.53	0.75							
37	MC	≥50,880	≥50,460	0.45	0.52	0.78							
38	MC	≥50,880	≥50,490	0.29	0.22	0.39							

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 5 Paper-Based Test Administration (continued)													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
39	ESR	≥50,880	≥50,630	0.45	0.60	0.50	33.75	42.08	23.67				
40	MC	≥50,880	≥50,650	0.48	0.51	0.43							
41	MC	≥50,880	≥50,740	0.77	0.43	0.27							
42	MC	≥50,880	≥50,590	0.52	0.48	0.54							
43	CR	≥50,880	≥50,270	0.17	0.41	1.07	62.44	26.86	5.86	3.62			
44	CR	≥50,880	≥50,730	0.30	0.65	0.30	23.58	23.84	23.85	14.72	8.65	4.04	1.00

Copyright © 2017 by Louisiana Department of Education.

Table 6.19 Item Statistics Mathematics Grade 6 Computer-Based Test Administration

Mathematics Grade 6 Computer-Based Test Administration													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
1	MC	≥4,800	≥4,800	0.45	0.56	0.02							
2	MC	≥4,800	≥4,800	0.83	0.31	0.08							
3	SA,GR	≥4,800	≥4,780	0.44	0.64	0.33	56.25	43.41					
4	MC	≥4,800	≥4,790	0.91	0.29	0.10							
5	MC	≥4,800	≥4,800	0.54	0.60	0.06							
6	MC	≥4,800	≥4,790	0.55	0.38	0.25							
7	MC	≥4,800	≥4,800	0.37	0.41	0.08							
8	MC	≥4,800	≥4,800	0.52	0.49	0.08							
9	MC	≥4,800	≥4,800	0.33	0.44	0.08							
10	MC	≥4,800	≥4,800	0.77	0.38	0.04							
11	MC	≥4,800	≥4,790	0.62	0.33	0.12							
12	TE	≥4,800	≥4,790	0.16	0.42	0.15	83.43	16.42					
13	MC	≥4,800	≥4,790	0.59	0.40	0.10							
14	MC	≥4,800	≥4,790	0.21	0.36	0.12							
15	SA,GR	≥4,800	≥4,780	0.56	0.37	0.37	44.18	55.44					
16	MC	≥4,800	≥4,790	0.50	0.40	0.29							
17	MC	≥4,800	≥4,790	0.48	0.09	0.15							
18	MC	≥4,800	≥4,790	0.46	0.26	0.21							
19	MC	≥4,800	≥4,790	0.86	0.37	0.21							
20	MC	≥4,800	≥4,790	0.67	0.46	0.12							
21	MC	≥4,790	≥4,790	0.57	0.36	0.02							
22	MS	≥4,800	≥4,790	0.06	0.13	0.25	93.63	6.12					
23	MPSR	≥4,800	≥4,760	0.23	0.57	0.89	69.39	14.40	15.32				
24	MC	≥4,790	≥4,780	0.16	0.22	0.21							
25	MC	≥4,790	≥4,790	0.85	0.37	0.02							
26	MC	≥4,790	≥4,790	0.77	0.37	0.15							
27	MPSR	≥4,800	≥4,790	0.73	0.54	0.23	12.76	27.45	59.56				
28	MC	≥4,790	≥4,790	0.69	0.41	0.10							
29	CR	≥4,800	≥4,760	0.44	0.63	0.60	33.01	10.57	13.32	32.13	10.11		
30	CR	≥4,800	≥4,710	0.47	0.62	1.27	29.82	23.79	19.06	25.54			
31	CR	≥4,800	≥4,700	0.11	0.47	1.52	81.69	5.14	6.08	5.04			
32	CR	≥4,800	≥4,670	0.12	0.54	2.08	79.17	4.60	10.47	3.14			
33	MC	≥4,780	≥4,780	0.65	0.42	0.08							
34	MPSR	≥4,800	≥4,780	0.22	0.51	0.35	64.93	25.45	9.26				
35	MC	≥4,780	≥4,780	0.68	0.36	0.19							
36	MS	≥4,800	≥4,780	0.34	0.56	0.40	65.35	34.26					
37	MC	≥4,780	≥4,770	0.30	0.13	0.21							
38	MC	≥4,780	≥4,780	0.72	0.46	0.06							

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 6 Computer-Based Test Administration (continued)													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
39	MS	≥4,800	≥4,780	0.40	0.60	0.46	59.65	39.90					
40	MC	≥4,780	≥4,780	0.35	0.45	0.04							
41	MC	≥4,780	≥4,780	0.74	0.38	0.08							
42	CR	≥4,800	≥4,760	0.33	0.68	0.71	17.90	46.29	24.79	6.85	3.37		
43	CR	≥4,800	≥4,700	0.09	0.49	1.62	83.16	6.95	3.33	4.47			
44	CR	≥4,800	≥4,770	0.07	0.55	0.54	63.43	33.84	0.94	0.85	0.17	0.12	0.10

Copyright © 2017 by Louisiana Department of Education.

Table 6.20 Item Statistics Mathematics Grade 6 Paper-Based Test Administration

Mathematics Grade 6 Paper-Based Test Administration													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
1	MC	≥47,860	≥47,680	0.46	0.57	0.37							
2	MC	≥47,860	≥47,680	0.78	0.34	0.29							
3	SA,GR	≥47,860	≥45,490	0.33	0.60	4.95	63.89	31.16					
4	MC	≥47,860	≥47,710	0.90	0.29	0.27							
5	MC	≥47,860	≥47,610	0.48	0.59	0.47							
6	MC	≥47,860	≥47,550	0.53	0.34	0.56							
7	MC	≥47,860	≥47,470	0.33	0.34	0.78							
8	MC	≥47,860	≥47,540	0.48	0.46	0.63							
9	MC	≥47,860	≥47,650	0.31	0.44	0.40							
10	MC	≥47,860	≥47,680	0.75	0.39	0.34							
11	MC	≥47,860	≥47,400	0.61	0.29	0.66							
12	MC	≥47,860	≥47,570	0.33	0.44	0.56							
13	MC	≥47,860	≥47,550	0.58	0.36	0.62							
14	MC	≥47,860	≥47,560	0.19	0.30	0.58							
15	SA,GR	≥47,860	≥45,800	0.52	0.37	4.31	45.51	50.17					
16	MC	≥47,860	≥47,590	0.50	0.41	0.42							
17	MC	≥47,860	≥47,460	0.52	0.06	0.78							
18	MC	≥47,860	≥47,440	0.44	0.18	0.81							
19	MC	≥47,860	≥47,510	0.86	0.35	0.71							
20	MC	≥47,860	≥47,500	0.61	0.47	0.72							
21	MC	≥47,860	≥47,710	0.55	0.34	0.29							
22	MS	≥47,860	≥47,220	0.06	0.15	1.33	92.92	5.75					
23	MPSR	≥47,860	≥46,630	0.17	0.51	2.56	75.32	11.20	10.92				
24	MC	≥47,860	≥47,360	0.14	0.19	1.01							
25	MC	≥47,860	≥47,570	0.85	0.36	0.58							
26	MC	≥47,860	≥47,610	0.73	0.39	0.44							
27	MPSR	≥47,860	≥47,690	0.68	0.56	0.35	16.96	29.03	53.67				
28	MC	≥47,860	≥46,930	0.69	0.40	1.90							
29	CR	≥47,860	≥47,310	0.45	0.62	1.00	32.01	7.52	14.65	36.74	7.94		
30	CR	≥47,860	≥46,740	0.38	0.58	2.07	38.18	23.77	19.04	16.68			
31	CR	≥47,860	≥46,630	0.15	0.50	2.28	75.85	8.12	5.81	7.64			
32	CR	≥47,860	≥45,710	0.11	0.55	4.28	80.06	2.91	9.27	3.27			
33	MC	≥47,860	≥47,670	0.62	0.43	0.36							
34	MPSR	≥47,860	≥47,650	0.18	0.46	0.44	69.71	24.37	5.48				
35	MC	≥47,860	≥47,500	0.66	0.30	0.64							
36	MS	≥47,860	≥47,500	0.30	0.55	0.76	69.24	30.00					
37	MC	≥47,860	≥47,200	0.29	0.11	1.35							
38	MC	≥47,860	≥47,160	0.69	0.45	1.41							

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 6 Paper-Based Test Administration (continued)													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
39	MS	≥47,860	≥47,290	0.38	0.59	1.20	60.96	37.84					
40	MC	≥47,860	≥47,490	0.32	0.37	0.69							
41	MC	≥47,860	≥47,570	0.72	0.35	0.54							
42	CR	≥47,860	≥47,280	0.32	0.63	1.11	16.17	50.67	23.94	5.51	2.49		
43	CR	≥47,860	≥46,160	0.07	0.45	3.16	86.50	4.00	2.71	3.23			
44	CR	≥47,860	≥47,580	0.06	0.53	0.57	65.17	32.84	0.56	0.50	0.12	0.11	0.12

Copyright © 2017 by Louisiana Department of Education.

Table 6.21 Item Statistics Mathematics Grade 7 Computer-Based Test Administration

Mathematics Grade 7 Computer-Based Test Administration													
Item	Item Type	Total N	Adj. N	*p*-value	Pbis Rit	Omit Rate	% at 0	% at 1	% at 2	% at 3	% at 4	% at 5	% at 6
1	MC	≥6,270	≥6,260	0.47	0.41	0.08							
2	MC	≥6,270	≥6,260	0.50	0.42	0.06							
3	MC	≥6,270	≥6,260	0.79	0.39	0.03							
4	MC	≥6,270	≥6,260	0.30	0.17	0.16							
5	MC	≥6,270	≥6,260	0.72	0.44	0.11							
6	MC	≥6,270	≥6,260	0.48	0.35	0.05							
7	SA,GR	≥6,270	≥6,250	0.61	0.51	0.24	38.71	61.06					
8	MC	≥6,270	≥6,260	0.25	0.22	0.16							
9	MC	≥6,270	≥6,260	0.59	0.44	0.06							
10	SA,GR	≥6,270	≥6,260	0.40	0.47	0.14	59.91	39.95					
11	MC	≥6,270	≥6,260	0.60	0.42	0.08							
12	TE	≥6,270	≥6,260	0.23	0.37	0.11	77.38	22.51					
13	MS	≥6,270	≥6,260	0.07	0.40	0.14	93.03	6.82					
14	MC	≥6,270	≥6,260	0.50	0.40	0.14							
15	MS	≥6,270	≥6,260	0.18	0.46	0.18	82.24	17.58					
16	MC	≥6,270	≥6,260	0.63	0.29	0.16							
17	MC	≥6,270	≥6,250	0.55	0.45	0.22							
18	MC	≥6,270	≥6,240	0.34	0.30	0.33							
19	MC	≥6,270	≥6,250	0.29	0.27	0.24							
20	SA,GR	≥6,270	≥6,240	0.12	0.31	0.49	87.60	11.91					
21	MC	≥6,260	≥6,250	0.72	0.41	0.14							
22	MC	≥6,260	≥6,260	0.68	0.33	0.06							
23	MPSR	≥6,270	≥6,220	0.31	0.60	0.80	52.83	30.69	15.69				
24	MS	≥6,270	≥6,260	0.51	0.34	0.18	49.10	50.73					
25	MC	≥6,260	≥6,250	0.38	0.30	0.24							
26	MC	≥6,260	≥6,260	0.39	0.16	0.08							
27	MC	≥6,260	≥6,250	0.49	0.40	0.14							
28	MPSR	≥6,270	≥6,250	0.29	0.51	0.33	53.79	34.13	11.75				
29	CR	≥6,270	≥6,160	0.15	0.62	1.45	73.06	13.06	5.42	6.73			
30	CR	≥6,270	≥6,150	0.21	0.58	1.24	41.10	38.63	12.35	5.21	0.88		
31	CR	≥6,270	≥6,060	0.05	0.31	2.65	87.69	5.72	1.53	1.72			
32	CR	≥6,270	≥6,030	0.07	0.53	3.09	81.73	10.86	1.20	2.42			
33	MC	≥6,250	≥6,250	0.68	0.27	0.05							
34	MC	≥6,250	≥6,240	0.63	0.46	0.13							
35	MS	≥6,270	≥6,250	0.11	0.34	0.33	88.24	11.43					
37	TE	≥6,270	≥6,240	0.31	0.56	0.45	68.39	31.17					
38	MC	≥6,250	≥6,240	0.37	0.22	0.14							
39	MPSR	≥6,270	≥6,210	0.23	0.58	0.99	72.12	9.09	17.81				
40	CR	≥6,270	≥6,220	0.71	0.51	0.56	11.13	12.51	29.03	46.58			

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 7 Computer-Based Test Administration (continued)													
Item	**Item Type**	**Total N**	**Adj. N**	**p-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
41	CR	≥6,270	≥6,100	0.11	0.58	1.96	72.52	11.88	8.26	3.92	0.77		
42	MPSR	≥6,270	≥6,220	0.50	0.69	0.73	17.04	12.77	11.06	11.35	16.74	21.55	8.75
43	ESR	≥6,270	≥6,170	0.55	0.62	1.53	32.44	24.58	41.45				

Copyright © 2017 by Louisiana Department of Education.

Table 6.22 Item Statistics Mathematics Grade 7 Paper-Based Test Administration

Mathematics Grade 7 Paper-Based Test Administration													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
1	MC	≥44,570	≥44,240	0.44	0.38	0.68							
2	MC	≥44,570	≥44,240	0.47	0.40	0.68							
3	MC	≥44,570	≥44,390	0.78	0.39	0.38							
4	MC	≥44,570	≥44,070	0.27	0.18	1.07							
5	MC	≥44,570	≥44,320	0.65	0.44	0.53							
6	MC	≥44,570	≥44,350	0.45	0.40	0.43							
7	SA,GR	≥44,570	≥43,240	0.56	0.51	2.97	43.01	54.02					
8	MC	≥44,570	≥44,180	0.26	0.18	0.83							
9	MC	≥44,570	≥44,340	0.55	0.44	0.47							
10	SA,GR	≥44,570	≥43,610	0.38	0.43	2.14	60.56	37.31					
11	MC	≥44,570	≥44,310	0.55	0.42	0.56							
12	MS	≥44,570	≥44,160	0.20	0.39	0.90	79.66	19.44					
13	MS	≥44,570	≥44,090	0.06	0.35	1.07	93.36	5.56					
14	MC	≥44,570	≥43,620	0.47	0.39	0.66							
15	MS	≥44,570	≥44,070	0.16	0.43	1.11	83.51	15.38					
16	MC	≥44,570	≥44,230	0.59	0.30	0.69							
17	MC	≥44,570	≥44,230	0.52	0.45	0.66							
18	MC	≥44,570	≥44,250	0.35	0.24	0.63							
19	MC	≥44,570	≥44,090	0.28	0.23	0.86							
20	SA,GR	≥44,570	≥43,390	0.11	0.24	2.63	86.97	10.40					
21	MC	≥44,570	≥44,160	0.69	0.41	0.89							
22	MC	≥44,570	≥44,400	0.63	0.34	0.36							
23	MPSR	≥44,570	≥43,730	0.30	0.61	1.87	53.86	30.47	13.80				
24	MS	≥44,570	≥44,320	0.40	0.36	0.56	60.07	39.38					
25	MC	≥44,570	≥44,200	0.35	0.29	0.82							
26	MC	≥44,570	≥44,150	0.38	0.17	0.88							
27	MC	≥44,570	≥44,280	0.48	0.41	0.61							
28	MPSR	≥44,570	≥44,230	0.27	0.49	0.75	55.84	33.27	10.14				
29	CR	≥44,570	≥43,380	0.12	0.59	2.31	76.51	11.60	4.16	5.08			
30	CR	≥44,570	≥43,130	0.18	0.58	2.81	50.98	30.12	9.39	5.30	0.98		
31	CR	≥44,570	≥42,410	0.04	0.31	4.28	86.80	5.60	1.40	1.37			
32	CR	≥44,570	≥42,330	0.06	0.52	4.55	82.38	10.11	0.83	1.66			
33	MC	≥44,570	≥44,230	0.62	0.26	0.69							
34	MC	≥44,570	≥44,220	0.57	0.49	0.75							
35	MS	≥44,570	≥44,160	0.10	0.33	0.91	89.10	9.99					
37	MC	≥44,570	≥44,160	0.43	0.36	0.86							
38	MC	≥44,570	≥44,210	0.36	0.18	0.74							
39	MPSR	≥44,570	≥43,660	0.19	0.56	2.04	75.34	8.73	13.89				

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 7 Paper-Based Test Administration (continued)													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
40	CR	≥44,570	≥44,080	0.71	0.51	1.00	11.11	12.87	26.73	48.20			
41	CR	≥44,570	≥42,410	0.12	0.60	4.19	70.33	10.01	9.53	4.17	1.14		
42	MPSR	≥44,570	≥44,130	0.47	0.71	0.96	20.58	13.83	9.60	10.59	14.80	20.73	8.88
43	ESR	≥44,570	≥43,870	0.52	0.63	1.57	33.35	27.40	37.69				

Copyright © 2017 by Louisiana Department of Education.

Table 6.23 Item Statistics Mathematics Grade 8 Computer-Based Test Administration

Mathematics Grade 8 Computer-Based Test Administration													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
1	MC	≥4,440	≥4,430	0.73	0.21	0.07							
2	MC	≥4,440	≥4,430	0.60	0.31	0.11							
3	MC	≥4,440	≥4,430	0.33	0.31	0.07							
4	MC	≥4,440	≥4,430	0.48	0.40	0.16							
5	MC	≥4,440	≥4,430	0.79	0.29	0.16							
6	ESR	≥4,440	≥4,430	0.61	0.37	0.05	18.06	42.03	39.86				
7	MC	≥4,440	≥4,430	0.44	0.26	0.05							
8	MC	≥4,440	≥4,430	0.28	0.22	0.09							
9	MS	≥4,440	≥4,430	0.07	0.30	0.11	93.33	6.55					
10	MC	≥4,440	≥4,430	0.68	0.38	0.09							
11	MC	≥4,440	≥4,430	0.51	0.11	0.09							
12	MC	≥4,440	≥4,430	0.34	0.24	0.14							
13	MS	≥4,440	≥4,430	0.37	0.40	0.11	62.86	37.03					
14	MC	≥4,440	≥4,430	0.45	0.35	0.11							
15	MC	≥4,440	≥4,430	0.64	0.41	0.09							
16	MC	≥4,440	≥4,430	0.49	0.21	0.11							
17	MS	≥4,440	≥4,430	0.15	0.25	0.09	84.89	15.02					
18	SA,GR	≥4,440	≥4,410	0.52	0.27	0.50	47.91	51.60					
19	MC	≥4,440	≥4,430	0.62	0.38	0.20							
20	MC	≥4,430	≥4,430	0.78	0.31	0.05							
21	MC	≥4,430	≥4,420	0.24	0.27	0.25							
22	MC	≥4,430	≥4,430	0.59	0.34	0.11							
23	MC	≥4,430	≥4,430	0.46	0.35	0.09							
24	MC	≥4,430	≥4,430	0.58	0.31	0.09							
25	MC	≥4,430	≥4,430	0.36	0.41	0.00							
26	MC	≥4,430	≥4,430	0.21	0.27	0.09							
27	MC	≥4,430	≥4,430	0.51	0.26	0.07							
28	TE	≥4,440	≥4,420	0.27	0.39	0.25	72.39	27.36					
29	MC	≥4,430	≥4,430	0.54	0.47	0.07							
30	CR	≥4,440	≥4,300	0.21	0.57	2.18	49.41	29.55	8.27	3.20	6.58		
31	CR	≥4,440	≥4,230	0.17	0.61	3.67	64.37	18.63	8.20	4.10			
32	CR	≥4,440	≥4,270	0.02	0.25	2.88	92.27	3.13	0.50	0.36			
33	CR	≥4,440	≥4,240	0.09	0.37	3.31	77.82	9.75	6.85	1.24			
34	MC	≥4,420	≥4,420	0.90	0.27	0.05							
35	MC	≥4,420	≥4,410	0.39	0.25	0.34							
36	MC	≥4,420	≥4,420	0.42	0.27	0.07							
37	MC	≥4,420	≥4,420	0.78	0.31	0.07							
38	MS	≥4,440	≥4,420	0.26	0.45	0.34	73.54	26.13					

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 8 Computer-Based Test Administration (continued)													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
39	MS	≥4,440	≥4,420	0.20	0.50	0.36	79.37	20.27					
40	MC	≥4,420	≥4,420	0.66	0.31	0.14							
41	MPSR	≥4,440	≥4,420	0.13	0.18	0.32	79.44	15.29	4.95				
42	MC	≥4,420	≥4,420	0.53	0.36	0.14							
43	CR	≥4,440	≥4,330	0.03	0.37	1.82	89.37	7.03	1.13	0.09			
44	CR	≥4,440	≥4,350	0.18	0.32	1.37	57.66	18.45	16.06	2.95	3.00		
45	CR	≥4,440	≥4,300	0.06	0.45	2.41	78.13	11.67	4.23	0.90	0.77	0.52	0.68

Copyright © 2017 by Louisiana Department of Education.

Table 6.24 Item Statistics Mathematics Grade 8 Paper-Based Test Administration

Mathematics Grade 8 Paper-Based Test Administration													
Item	**Item Type**	**Total N**	**Adj. N**	***p*-value**	**Pbis Rit**	**Omit Rate**	**% at 0**	**% at 1**	**% at 2**	**% at 3**	**% at 4**	**% at 5**	**% at 6**
1	MC	≥42,490	≥42,300	0.70	0.27	0.37							
2	MC	≥42,490	≥42,320	0.57	0.34	0.39							
3	MC	≥42,490	≥42,270	0.30	0.34	0.46							
4	MC	≥42,490	≥42,180	0.46	0.44	0.66							
5	MC	≥42,490	≥42,290	0.73	0.33	0.43							
6	ESR	≥42,490	≥42,350	0.58	0.42	0.33	21.38	41.43	36.85				
7	MC	≥42,490	≥42,250	0.44	0.29	0.53							
8	MC	≥42,490	≥42,180	0.28	0.27	0.61							
9	MS	≥42,490	≥42,030	0.06	0.33	1.08	92.58	6.34					
10	MC	≥42,490	≥42,200	0.63	0.41	0.67							
11	MC	≥42,490	≥42,310	0.48	0.13	0.39							
12	MC	≥42,490	≥42,220	0.34	0.25	0.57							
13	MS	≥42,490	≥42,190	0.35	0.42	0.70	64.73	34.57					
14	MC	≥42,490	≥42,190	0.43	0.35	0.66							
15	MC	≥42,490	≥42,290	0.59	0.42	0.39							
16	MC	≥42,490	≥42,280	0.42	0.22	0.39							
17	MS	≥42,490	≥42,100	0.16	0.22	0.92	83.42	15.66					
18	SA,GR	≥42,490	≥40,980	0.46	0.35	3.56	52.12	44.32					
19	MC	≥42,490	≥42,150	0.59	0.39	0.71							
20	MC	≥42,490	≥42,310	0.74	0.33	0.37							
21	MC	≥42,490	≥42,160	0.24	0.34	0.71							
22	MC	≥42,490	≥42,190	0.57	0.34	0.68							
23	MC	≥42,490	≥42,190	0.43	0.39	0.55							
24	MC	≥42,490	≥42,070	0.54	0.34	0.96							
25	MC	≥42,490	≥42,080	0.34	0.43	0.94							
26	MC	≥42,490	≥42,150	0.21	0.25	0.77							
27	MC	≥42,490	≥42,080	0.51	0.30	0.90							
28	SA,GR	≥42,490	≥40,060	0.18	0.41	5.72	77.72	16.55					
29	MC	≥42,490	≥42,200	0.53	0.49	0.61							
30	CR	≥42,490	≥40,200	0.24	0.63	4.84	46.21	27.45	8.41	4.24	8.29		
31	CR	≥42,490	≥40,010	0.18	0.65	5.13	64.10	15.63	8.49	5.97			
32	CR	≥42,490	≥40,500	0.04	0.40	4.22	87.67	5.63	1.04	0.99			
33	CR	≥42,490	≥40,370	0.12	0.50	4.49	72.28	13.31	7.42	2.00			
34	MC	≥42,490	≥42,200	0.91	0.24	0.65							
35	MC	≥42,490	≥41,760	0.40	0.28	1.71							
36	MC	≥42,490	≥42,200	0.40	0.32	0.66							
37	MC	≥42,490	≥42,240	0.76	0.34	0.56							
38	MS	≥42,490	≥41,730	0.26	0.46	1.79	72.71	25.50					

Copyright © 2017 by Louisiana Department of Education.

Mathematics Grade 8 Paper-Based Test Administration (continued)													
Item	Item Type	Total N	Adj. N	*p*-value	Pbis Rit	Omit Rate	% at 0	% at 1	% at 2	% at 3	% at 4	% at 5	% at 6
39	MS	≥42,490	≥41,740	0.20	0.52	1.77	78.99	19.24					
40	MC	≥42,490	≥42,150	0.69	0.34	0.72							
41	MPSR	≥42,490	≥42,060	0.12	0.23	1.00	79.07	15.19	4.75				
42	MC	≥42,490	≥42,110	0.51	0.43	0.87							
43	CR	≥42,490	≥39,980	0.03	0.40	5.39	87.34	5.61	0.97	0.18			
44	CR	≥42,490	≥40,800	0.21	0.32	3.61	49.04	23.63	17.09	3.46	2.81		
45	CR	≥42,490	≥40,730	0.07	0.53	3.68	77.29	10.20	4.15	1.18	0.70	0.91	1.44

6.2 Item Response Theory

Item parameters for items contained in ELA and mathematics 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 θ 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 θ will respond in category *x* to partial-credit item *j* is

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

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

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

PARSCALE (Muraki and Bock, 2003) was used for the IRT calibrations. PARSCALE is a multipurpose program that implements a variety of IRT models associated with mixed-item formats and associated statistics. PARSCALE has been used to calibrate large data sets such as those of Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced assessments. The program implements marginal maximum likelihood (MML) estimation techniques for items and MLE estimation of theta.

Copyright © 2017 by Louisiana Department of Education.

6.3 Calibration Sample

This section describes the calibration sample in adherence to Standard 1.8 of the *Standards*. Standard 1.8 states the following:

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

The calibration sample included only students who had answered at least 25 percent of the items. All student data available at the time of calibration was used, resulting in a near-census data file. Tables 6.25 and 6.26 show the representativeness of the calibration samples compared to the census data. This table demonstrates that the calibration sample was representative of the state.

Table 6.25 Summary of Calibration and Census Data: English Language Arts

Calibration and Census Data: English Language Arts						
		Calibration Sample		Census Data		
Grade		N	%	N	%	(Calib % - Census %)
3	All Students	≥57,200	100.00%	≥57,270	100.00%	0.00%
	Gender					
	Male	≥29,280	51.12%	≥29,20	51.12%	0.00%
	Female	≥27,980	48.85%	≥27,980	48.85%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥3,800	6.65%	≥3,800	6.65%	0.00%
	American Indian or Alaska Native	≥350	0.63%	≥350	0.63%	0.00%
	Asian	≥800	1.41%	≥800	1.41%	0.00%
	Black or African American	≥25,240	44.08%	≥25,240	44.08%	0.00%
	Native Hawaiian or Other Pacific	≥50	0.09%	≥50	0.09%	0.00%
	White	≥25,380	44.33%	≥25,380	44.33%	0.00%
	Two or more races	≥1,550	2.72%	≥1,550	2.72%	0.00%
4	All Students	≥54,130	100.00%	≥54,150	100.00%	0.00%
	Gender					
	Male	≥27,840	51.41%	≥27,840	51.41%	0.00%
	Female	≥26,280	48.53%	≥26,280	48.53%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥3,350	6.19%	≥3,350	6.19%	0.00%
	American Indian or Alaska Native	≥340	0.63%	≥340	0.63%	0.00%
	Asian	≥810	1.51%	≥810	1.51%	0.00%
	Black or African American	≥24,340	44.96%	≥24,340	44.96%	0.00%
	Native Hawaiian or Other Pacific	≥50	0.09%	≥50	0.09%	0.00%
	White	≥23,890	44.12%	≥23,890	44.12%	0.00%
	Two or more races	≥1,290	2.40%	≥1,290	2.40%	0.00%

Copyright © 2017 by Louisiana Department of Education.

Calibration and Census Data: English Language Arts (continued)						
		Calibration Sample		Census Data		
Grade		N	%	N	%	(Calib % - Census %)
5	All Students	≥52,690	100.00%	≥52,740	100.00%	0.00%
	Gender					
	Male	≥27,180	51.53%	≥27,180	51.53%	0.00%
	Female	≥25,550	48.44%	≥25,550	48.44%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥3,160	5.99%	≥3,160	5.99%	0.00%
	American Indian or Alaska Native	≥360	0.70%	≥360	0.70%	0.00%
	Asian	≥800	1.52%	≥800	1.52%	0.00%
	Black or African American	≥23,390	44.36%	≥23,390	44.36%	0.00%
	Native Hawaiian or Other Pacific	≥40	0.08%	≥40	0.08%	0.00%
	White	≥23,850	45.23%	≥23,850	45.23%	0.00%
	Two or more races	≥1,070	2.04%	≥1,070	2.04%	0.00%
6	All Students	≥52,760	100.00%	≥52,810	100.00%	0.00%
	Gender					
	Male	≥26,980	51.08%	≥26,980	51.08%	0.00%
	Female	≥25,820	48.89%	≥25,820	48.89%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥2,950	5.59%	≥2,950	5.59%	0.00%
	American Indian or Alaska Native	≥410	0.78%	≥410	0.78%	0.00%
	Asian	≥830	1.58%	≥830	1.58%	0.00%
	Black or African American	≥23,850	45.16%	≥23,850	45.16%	0.00%
	Native Hawaiian or Other Pacific	≥40	0.09%	≥40	0.09%	0.00%
	White	≥23,740	44.97%	≥23,740	44.97%	0.00%
	Two or more races	≥930	1.77%	≥930	1.77%	0.00%
7	All Students	≥51,000	100.00%	≥51,110	100.00%	0.00%
	Gender					
	Male	≥26,360	51.58%	≥26,360	51.58%	0.00%
	Female	≥24,730	48.38%	≥24,730	48.38%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥2,750	5.38%	≥2,750	5.38%	0.00%
	American Indian or Alaska Native	≥360	0.71%	≥360	0.71%	0.00%
	Asian	≥800	1.57%	≥800	1.57%	0.00%
	Black or African American	≥22,700	44.42%	≥22,700	44.42%	0.00%
	Native Hawaiian or Other Pacific	≥40	0.09%	≥40	0.09%	0.00%
	White	≥23,550	46.09%	≥23,550	46.09%	0.00%
	Two or more races	≥840	1.65%	≥840	1.65%	0.00%

Copyright © 2017 by Louisiana Department of Education.

Calibration and Census Data: English Language Arts (continued)						
		Calibration Sample		Census Data		
Grade		N	%	N	%	(Calib % - Census %)
8	All Students	≥51,490	100.00%	≥51,490	100.00%	0.00%
	Gender					
	Male	≥26,340	51.16%	≥26,340	51.16%	0.00%
	Female	≥25,140	48.82%	≥25,140	48.82%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥2,690	5.24%	≥2,690	5.24%	0.00%
	American Indian or Alaska Native	≥360	0.72%	≥360	0.72%	0.00%
	Asian	≥770	1.50%	≥770	1.50%	0.00%
	Black or African American	≥23,280	45.21%	≥23,280	45.21%	0.00%
	Native Hawaiian or Other Pacific	≥40	0.08%	≥40	0.08%	0.00%
	White	≥23,520	45.69%	≥23,520	45.69%	0.00%
	Two or more races	≥760	1.49%	≥760	1.49%	0.00%

Copyright © 2017 by Louisiana Department of Education.

Table 6.26 Summary of Calibration and Census Data: Mathematics

Calibration and Census Data: Mathematics						
		Calibration Sample		Census Data		
Grade		N	%	N	%	(Calib % - Census %)
3	All Students	≥57,190	100.00%	≥57,250	100.00%	0.00%
	Gender					
	Male	≥29,260	51.11%	≥29,260	51.11%	0.00%
	Female	≥27,960	48.84%	≥27,960	48.84%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥3,810	6.66%	≥3,810	6.66%	0.00%
	American Indian or Alaska Native	≥350	0.63%	≥350	0.63%	0.00%
	Asian	≥800	1.41%	≥800	1.41%	0.00%
	Black or African American	≥25,240	44.09%	≥25,240	44.09%	0.00%
	Native Hawaiian or Other Pacific	≥50	0.09%	≥50	0.09%	0.00%
	White	≥25,380	44.33%	≥25,380	44.33%	0.00%
	Two or more races	≥1,550	2.71%	≥1,550	2.71%	0.00%
4	All Students	≥54,060	100.00%	≥54,160	100.00%	0.00%
	Gender					
	Male	≥27,850	51.42%	≥27,850	51.42%	0.00%
	Female	≥26,280	48.52%	≥26,280	48.52%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥3,360	6.21%	≥3,360	6.21%	0.00%
	American Indian or Alaska Native	≥340	0.63%	≥340	0.63%	0.00%
	Asian	≥810	1.51%	≥810	1.51%	0.00%
	Black or African American	≥24,350	44.95%	≥24,350	44.95%	0.00%
	Native Hawaiian or Other Pacific	≥50	0.09%	≥50	0.09%	0.00%
	White	≥23,890	44.11%	≥23,890	44.11%	0.00%
	Two or more races	≥1,300	2.40%	≥1,300	2.40%	0.00%
5	All Students	≥52,650	100.00%	≥52,740	100.00%	0.00%
	Gender					
	Male	≥27,170	51.53%	≥27,170	51.53%	0.00%
	Female	≥25,550	48.45%	≥25,550	48.45%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥3,160	5.99%	≥3,160	5.99%	0.00%
	American Indian or Alaska Native	≥360	0.70%	≥360	0.70%	0.00%
	Asian	≥800	1.52%	≥800	1.52%	0.00%
	Black or African American	≥23,400	44.38%	≥23,400	44.38%	0.00%
	Native Hawaiian or Other Pacific	≥40	0.09%	≥40	0.09%	0.00%
	White	≥23,840	45.21%	≥23,840	45.21%	0.00%
	Two or more races	≥1,070	2.03%	≥1,070	2.03%	0.00%

Copyright © 2017 by Louisiana Department of Education.

Calibration and Census Data: Mathematics (continued)						
		Calibration Sample		Census Data		
Grade		N	%	N	%	(Calib % - Census %)
6	All Students	≥52,710	100.00%	≥52,820	100.00%	0.00%
	Gender					
	Male	≥26,980	51.07%	≥26,980	51.07%	0.00%
	Female	≥25,820	48.88%	≥25,820	48.88%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥2,970	5.64%	≥2,970	5.64%	0.00%
	American Indian or Alaska Native	≥410	0.78%	≥410	0.78%	0.00%
	Asian	≥830	1.58%	≥830	1.58%	0.00%
	Black or African American	≥23,840	45.14%	≥23,840	45.14%	0.00%
	Native Hawaiian or Other Pacific	≥40	0.09%	≥40	0.09%	0.00%
	White	≥23,720	44.92%	≥23,720	44.92%	0.00%
	Two or more races	≥930	1.76%	≥930	1.76%	0.00%
7	All Students	≥50,880	100.00%	≥50,990	100.00%	0.00%
	Gender					
	Male	≥26,300	51.58%	≥26,300	51.58%	0.00%
	Female	≥24,670	48.39%	≥24,670	48.39%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥2,740	5.39%	≥2,740	5.39%	0.00%
	American Indian or Alaska Native	≥360	0.72%	≥360	0.72%	0.00%
	Asian	≥790	1.56%	≥790	1.56%	0.00%
	Black or African American	≥22,660	44.46%	≥22,660	44.46%	0.00%
	Native Hawaiian or Other Pacific	≥40	0.09%	≥40	0.09%	0.00%
	White	≥23,470	46.03%	≥23,470	46.03%	0.00%
	Two or more races	≥840	1.66%	≥840	1.66%	0.00%
8	All Students	≥47,000	100.00%	≥47,190	100.00%	0.00%
	Gender					
	Male	≥24,360	51.63%	≥24,360	51.63%	0.00%
	Female	≥22,800	48.32%	≥22,800	48.32%	0.00%
	Race Ethnicity					
	Hispanic/Latino	≥2,500	5.31%	≥2,500	5.31%	0.00%
	American Indian or Alaska Native	≥340	0.73%	≥340	0.73%	0.00%
	Asian	≥570	1.21%	≥570	1.21%	0.00%
	Black or African American	≥22,470	47.62%	≥22,470	47.62%	0.00%
	Native Hawaiian or Other Pacific	≥30	0.08%	≥30	0.08%	0.00%
	White	≥20,510	43.47%	≥20,510	43.47%	0.00%
	Two or more races	≥680	1.46%	≥680	1.46%	0.00%

Copyright © 2017 by Louisiana Department of Education.

6.4 Calibration and Linking

All LEAP 2016 item calibration and linking were performed based on item response theory (IRT).

Calibration and linking methodology used for the Spring 2016 LEAP administration closely followed most of the PARCC methods referenced in the PARCC document "Final Technical Report for 2015 Administration." To maintain comparability to PARCC, the 2PL/GPC IRT model was applied to item calibration using the software PARSCALE (Muraki and Bock, 2003). To avoid local independence between Writing traits, the Writing traits Written Expression (WE) and Writing Knowledge and Use of Language Conventions (WKL) were separately calibrated using sparse matrix or separate calibration.

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

Descriptions of the PARCC calibration and equating approach can be found in the PARCC documents "Final Technical Report for 2015 Administration" and "Linking Chapter_PARCC OPM_2016."

There were two test forms, computer-based test (CBT) and paper-based test (PBT), for each LEAP 2016 ELA and mathematics area, except for grade 3. There was only PBT for grade 3 ELA and mathematics. In general, a school took the same test mode for ELA and mathematics. For LEAP 2016 calibration, CBT and PBT were combined and calibrated together for grades 4 and 5 due to their small sample sizes, but CBT and PBT were separately calibrated for grades 6 to 8. Table 6.27 summarizes the count of students administered by test mode, as well as the number of items on the test form.

LEAP 2016 tests were placed on the PARCC 2016 base scale following this order:

1. Calibrate LEAP 2016 tests
2. Equate LEAP 2016 PBT to LEAP 2016 CBT tests
3. Link LEAP 2016 CBT to LEAP 2016 PARCC

CBT and PBT were separately calibrated for ELA and mathematics grades 6 to 8, following the PARCC approach. PBT was equated to CBT using all common items between CBT and PBT by the Stocking and Lord procedure. Most items for CBT and PBT were common except for two or three unique items from each form.

LEAP 2016 CBT forms were linked to the PARCC scale using all PARCC items embedded in LEAP 2016 CBT as anchors by the Stocking and Lord procedure.

Copyright © 2017 by Louisiana Department of Education.

PARCC established a new baseline scale using 2016 spring tests, and linked the 2015 PARCC scale to the new 2016 baseline scale. LEAP 2016 scores and item parameters from PARCC 2015 tests were also updated using the transformation constants provided by PARCC.

6.4.1 Calibration of LEAP 2016 Tests

For LEAP 2016 item calibration, the 2PL/GPC IRT model was applied to Louisiana students' calibration sample using the software PARSCALE (Muraki and Bock, 2003). Table 6.26 shows the number of students in the calibration samples and number of items by mode. There was no CBT form for grade 3, and few students took CBT in grades 4 and 5. There were more students for ELA grade 8 compared to those for mathematics grade 8 because high-performing students were allowed to take EOC Algebra I instead of the Grade 8 mathematics test. CBT and PBT consist of the same number of items, 29 to 36 for ELA tests and 43 to 45 items for mathematics tests.

There was only one grade 3 PBT form; all PBT items were calibrated together for each subject area. Grades 4 and 5 had one PBT form and one CBT form; however, both had small sample sizes. Therefore, PBT and CBT forms were combined and calibrated together for grades 4 and 5. For the other grades, each mode form was separately calibrated.

Table 6.27 Summary of Student Count and Item Count by Test Mode

		N			Percent		N-Items	
Content	Grade	All	CBT	PBT	CBT	PBT	CBT	PBT
ELA	3	≥57,200	*	≥57,200	*	100.00	*	29
	4	≥54,130	≥450	≥53,680	0.84	99.16	36	36
	5	≥52,690	≥1690	≥51,000	3.21	96.79	31	31
	6	≥52,760	≥4790	≥47,960	9.09	90.91	31	31
	7	≥51,000	≥6290	≥44,710	12.35	87.65	31	31
	8	≥51,490	≥5450	≥46,040	10.59	89.41	36	36
Mathematics	3	≥57,190	*	≥57,190	*	100.00	*	45
	4	≥54,060	≥450	≥53,610	0.84	99.16	44	44
	5	≥52,650	≥1690	≥50,960	3.21	96.79	44	44
	6	≥52,710	≥4800	≥47,910	9.11	90.89	44	44
	7	≥50,880	≥6270	≥44,610	12.33	87.67	43	43
	8	≥47,000	≥4430	≥42,560	9.44	90.56	45	45

*Grade 3 did not have a CBT form.

6.4.1.1. Separate Calibration for ELA Prose Constructed-Response (PCR) Tasks

To address the issue of local independence for ELA PCR tasks, two different calibration methods were applied. Each ELA test consists of two PCR tasks; there were two Written Expression (WE) traits and two Knowledge and Use of Language Conventions (WKL)

Copyright © 2017 by Louisiana Department of Education.

traits for each test. As can be seen in Table 6.27 for grades 3 to 5, a single calibration was performed by randomly splitting the students into two groups. Almost half of the dataset included response of other items and two WE responses, and the other calibration dataset included the same responses of other items and two WKL responses. Because these two sets of item responses were calibrated together, there is only one unique set of item parameters for each item. PARCC took this sparse matrix approach for all grades because its calibration sample size was large enough for all grades.

For grades 6 to 8, calibration was performed by mode; there were two calibration forms for each grade, as can be seen in Table 6.28. The same approach used for grades 4 and 5 was used to calibrate WE and WKL items for grades 6 to 8 ELA PBT form because there were sufficient samples. However, because sample size was not large enough for grades 6 to 8 CBT form, a different calibration approach was taken. For each grade CBT form, two sets of responses were generated. One calibration dataset includes two WE responses and responses of the other items excluding WKL responses, and the other calibration dataset includes two WKL responses and the same responses of the other items excluding WE responses. In short, responses of the other items, except for WE and WKL, were the same for both forms, and each dataset included either WE or WKL responses.

Table 6.28 Calibration Data Structure for ELA WE and WKL Traits with Sparse Matrix

Group	Other Items	WE	WKL
I	XXXXXXXX	XX	
II	XXXXXXXX		XX

Table 6.29 Calibration Data Structure for ELA WE and WKL Traits with Repeated Sample Procedure

Group	Other Items	WE	WKL
All Students	XXXXXXXX	XX	
All Students	XXXXXXXX		XX

6.4.1.2. Equating between ELA CBT Two Writing Traits

For each dataset, WE and WKL were separately calibrated for ELA grades 6 to 8. After separate calibration, Stocking and Lord procedure was applied using all items, except for WE and WKL, as anchor items by considering the form with WKL responses as the reference form and the form with WE responses as the focal form. Since most items were anchor items and the calibration samples were the same, transformation constants values were expected to be similar to identity constants (slope=1 and intercept=0). Table 6.30 shows the number of anchor items and transformation constants after Stocking and Lord equating.

Copyright © 2017 by Louisiana Department of Education.

Table 6.30 Transformation Constants for ELA CBT PCR Tasks

Grade	N of Anchor Items	Slope	Intercept
6	27	1.008160	-0.002750
7	26	1.002044	-0.002306
8	32	1.003291	-0.003663

As expected, the item parameters for both forms were very similar, and their transformation values were close to the identity values. After applying the Stocking and Lord transformation, all item parameters of the form with WKL were kept and WE item parameters from the other form were added to the form with WKL. Ultimately, there was one unique set of item parameters for each CBT item. Tables 6.31 and 6.32 present a summary of estimated item parameters for ELA and mathematics.

Table 6.31 Summary of ELA Item Parameter Estimates

			a parameter				b parameter			
Grade	Mode	No. of Items	Mean	Std. Dev.	Min.	Max	Mean	Std. Dev.	Min.	Max
3	PBT	45	0.70	0.19	0.32	1.18	-0.07	1.47	-3.94	3.69
4	CBT/ PBT	45	0.67	0.19	0.18	1.06	-0.04	1.41	-2.45	5.34
5	CBT / PBT	45	0.62	0.24	0.17	1.16	0.27	1.39	-3.80	3.41
6	CBT	44	0.70	0.25	0.14	1.34	0.56	1.51	-2.00	4.89
6	PBT	44	0.65	0.25	0.08	1.17	0.71	1.57	-1.95	4.78
7	CBT	43	0.69	0.24	0.23	1.25	0.62	1.16	-1.32	3.76
7	PBT	43	0.65	0.23	0.23	1.23	0.78	1.16	-1.17	3.75
8	CBT	45	0.52	0.19	0.15	1.02	1.09	1.64	-2.06	5.23
8	PBT	45	0.57	0.20	0.19	1.04	1.06	1.40	-2.15	4.95

Table 6.32 Summary of Mathematics Item Parameter Estimates

			a parameter				b parameter			
Grade	Mode	No. of Items	Mean	Std. Dev.	Min.	Max	Mean	Std. Dev.	Min.	Max
3	PBT	29	0.53	0.19	0.13	0.81	0.54	1.87	-4.18	6.54
4	CBT/ PBT	39	0.55	0.25	0.18	1.11	0.48	2.12	-4.06	8.00
5	CBT/ PBT	33	0.52	0.20	0.24	1.03	0.61	1.64	-4.61	3.66
6	CBT	31	0.60	0.28	0.14	1.15	0.20	1.54	-3.30	5.67
6	PBT	31	0.57	0.27	0.08	1.16	0.34	1.93	-3.48	9.06
7	CBT	31	0.51	0.23	0.22	0.96	0.17	2.47	-8.35	5.28
7	PBT	31	0.49	0.22	0.22	0.89	0.21	2.56	-8.18	5.89
8	CBT	36	0.61	0.29	0.19	1.11	-0.06	2.30	-4.84	7.21
8	PBT	36	0.60	0.27	0.21	1.12	-0.22	2.20	-5.37	6.47

All LEAP 2016 items successfully calibrated and converged.

Copyright © 2017 by Louisiana Department of Education.

6.4.1.3. IRT Item Fit

The usefulness of IRT models is dependent on the extent to which they effectively reflect the data. Hambleton, Swaminathan, and Rogers (1991) explain, "The advantages of item response models can be obtained only when the fit between the model and the test data of interest is satisfactory. A poorly fitting IRT model will not yield invariant item and ability parameters" (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, the IRT model fit was evaluated by reviewing item chi-square values from PARSCALE (Muraki and Bock, 2003) and calculating adjusted fit values and flagging them if > 0.45 (PARCC 2015 TR).

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

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

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

Tables 6.33 and 6.34 show adjusted item fit using PARSCALE (Muraki and Bock, 2003) chi-square values and calibration sample size for ELA and mathematics. Average adjusted fit for ELA ranged from 0.10 to 0.14, and mathematics ranged from 0.08 to 0.14. Items were not excluded based on model fit statistics because the adjusted item fits for all items were lower than the criterion value of 0.45, as can be seen in the maximum values for both ELA and mathematics. The largest adjusted fit value was 0.42 for mathematics grade 7.

Copyright © 2017 by Louisiana Department of Education.

Table 6.33 Summary of Adjusted Fit for ELA

Grade	Mode	No. Items	Mean	Std. Dev.	Min.	Max	No. Flagged Items	No. Not Available
3	PBT	29	0.10	0.05	0.04	0.23	0	0
4	CBT/PBT	39	0.10	0.06	0.04	0.26	0	0
5	CBT/PBT	33	0.11	0.04	0.05	0.21	0	0
6	CBT	31	0.14	0.04	0.07	0.26	0	0
6	PBT	31	0.10	0.05	0.05	0.25	0	0
7	CBT	31	0.14	0.06	0.07	0.37	0	0
7	PBT	31	0.11	0.06	0.03	0.33	0	0
8	CBT	36	0.13	0.04	0.06	0.22	0	0
8	PBT	36	0.10	0.05	0.03	0.20	0	0

Table 6.34 Summary of Adjusted Fit for Mathematics

Grade	Mode	No. Items	Mean	Std. Dev.	Min.	Max	No. Flagged Items	No. Not Available
3	PBT	45	0.08	0.03	0.04	0.20	0	0
4	CBT/PBT	45	0.09	0.04	0.04	0.23	0	0
5	CBT/PBT	45	0.09	0.04	0.04	0.18	0	0
6	CBT	44	0.12	0.04	0.08	0.22	0	0
6	PBT	44	0.08	0.03	0.04	0.15	0	0
7	CBT	43	0.14	0.06	0.05	0.42	0	0
7	PBT	43	0.11	0.06	0.04	0.41	0	0
8	CBT	45	0.12	0.03	0.08	0.21	0	0
8	PBT	45	0.08	0.03	0.03	0.16	0	0

6.4.2 Linking LEAP 2016 PBT to LEAP 2016 CBT for Grades 6 to 8

For grades 6 to 8, there was one CBT form and one PBT form, and they were separately calibrated following the PARCC procedure. These two forms consisted of common items, but there was a maximum of two or three unique items for each form. PBT forms were equated to CBT forms using all common items by Stocking and Lord equating method. Table 6.35 presents the number of anchor items and Stocking and Lord transformation Constants for ELA and mathematics grades 6 to 8.

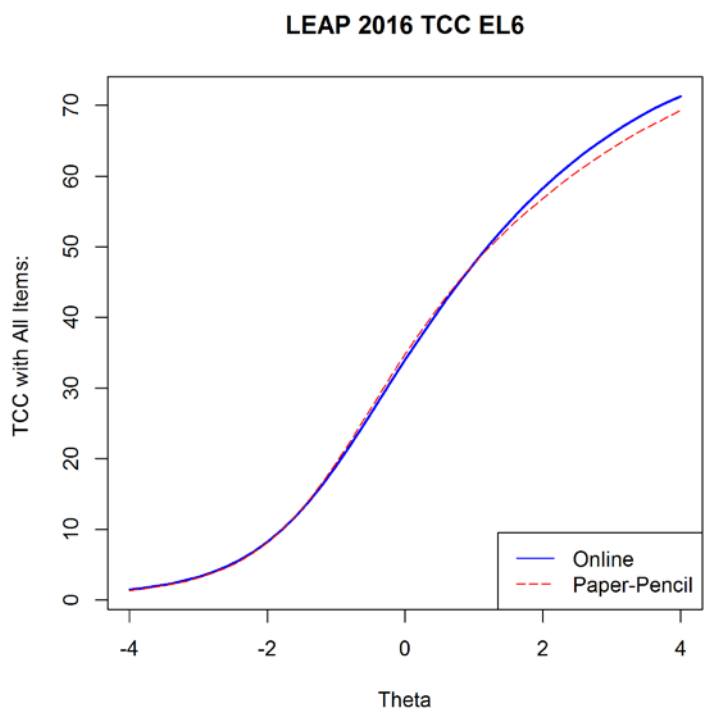
Table 6.35 Stocking and Lord Transformation Constants for Linking LEAP PBT to LEAP CBT

Content	Grade	No. of Anchor Items	Slope	Intercept
ELA	6	29	0.909901	-0.082811
	7	28	0.921696	-0.020650
	8	33	0.949990	-0.031919
Mathematics	6	42	0.958261	-0.130624
	7	40	0.987628	-0.134063
	8	44	1.132877	-0.067171

Copyright © 2017 by Louisiana Department of Education.

Figures 6.1 and 6.2 show test characteristic curves (TCC) for PBT and CBT anchor items for ELA and mathematics after Stocking and Lord equating procedure for grades 6 to 8. The *x*-axis represents the ability range and the *y*-axis represents estimated observed scores given ability. The blue solid line is for CBT (online) and the red dotted line is for PBT (Paper-Pencil). For ELA, the TCCs for CBT and PBT were overlapped for most of the ability levels across all three grades. There were some differences at the high ability range for grades 6 and 7. For mathematics, the TCCs for CBT and PBT were overlapped across all ability ranges.

Figure 6.1 ELA PBT to CBT Anchor Item Test Characteristic Curves



Copyright © 2017 by Louisiana Department of Education.

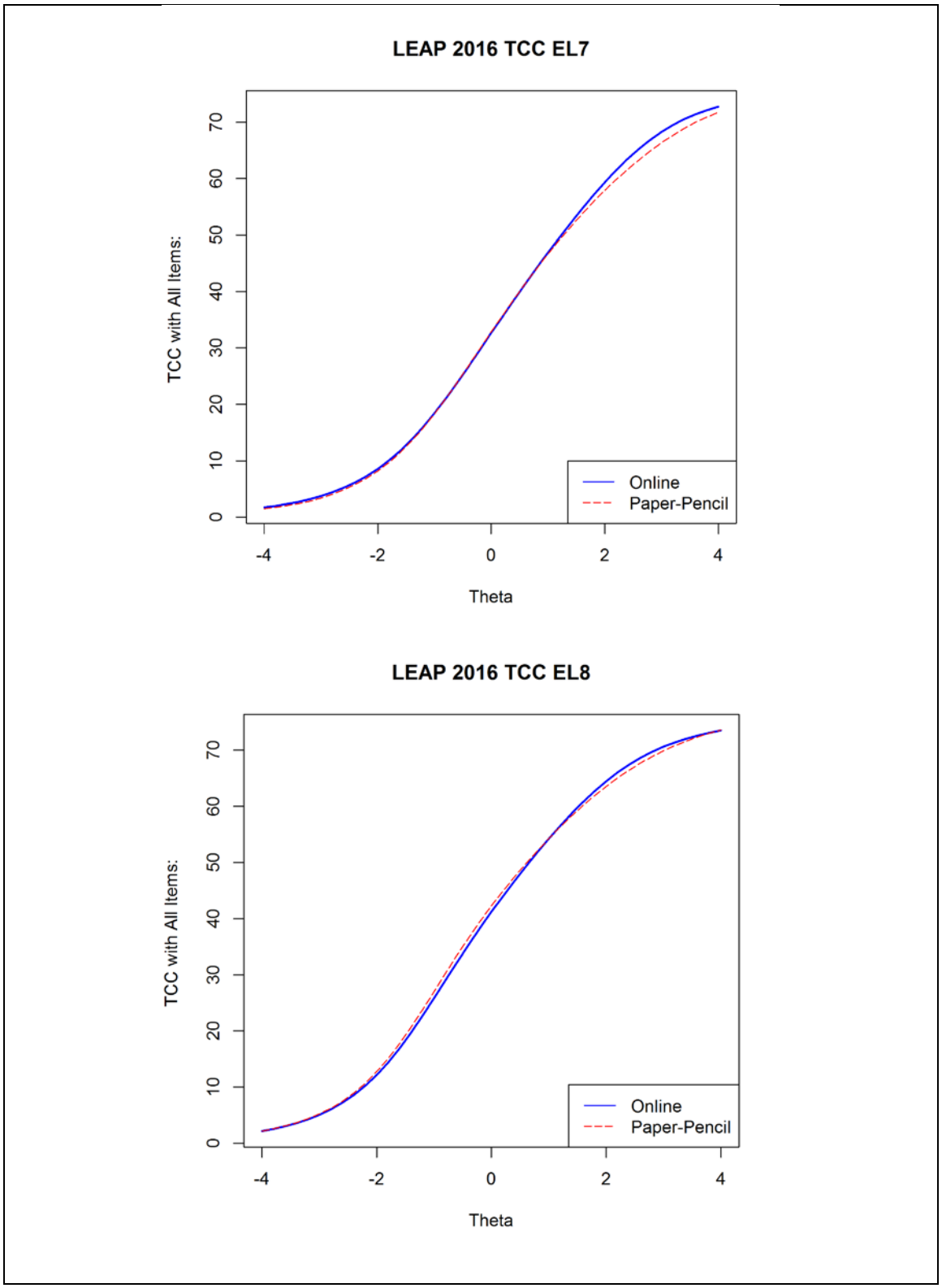
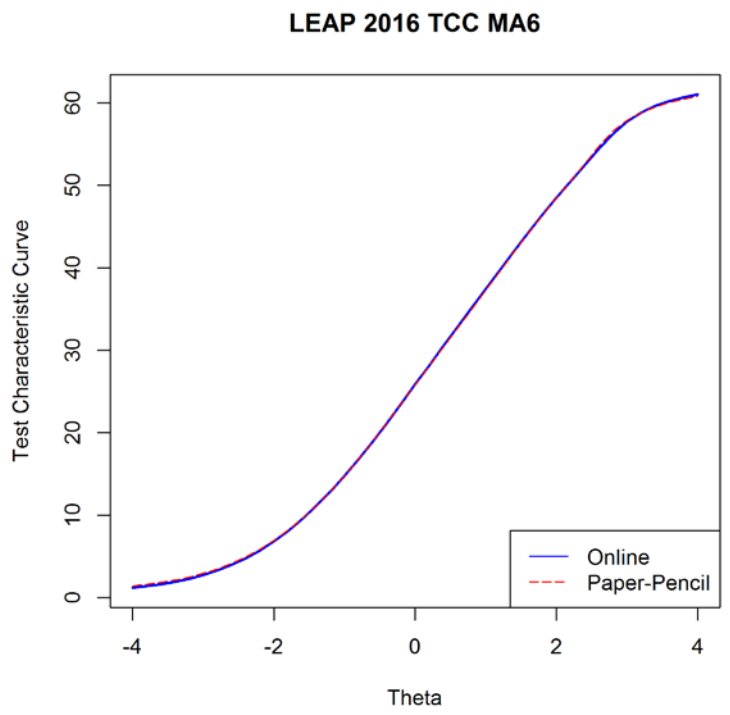
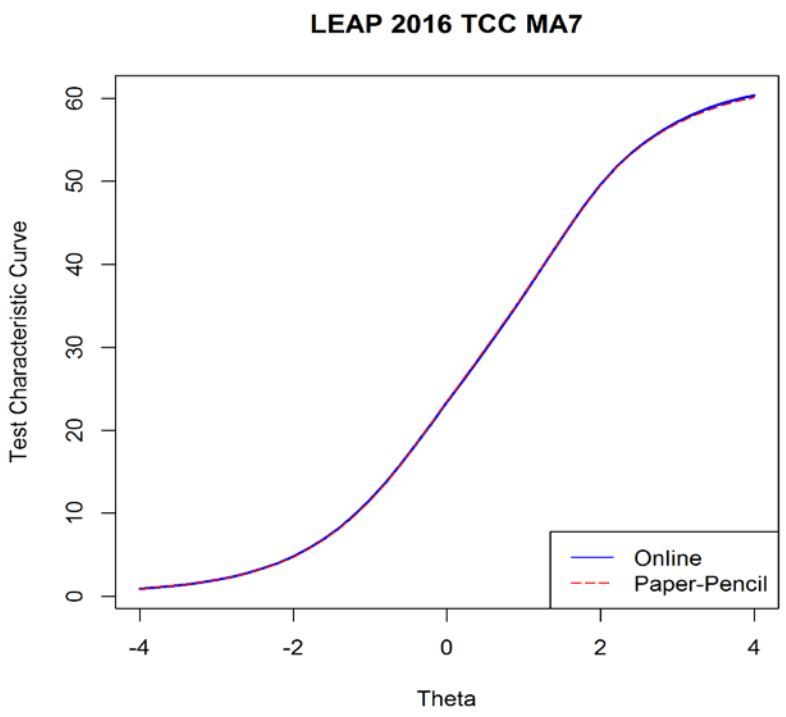


Figure 6.2 Mathematics PBT to CBT Anchor Item Test Characteristic Curves

Copyright © 2017 by Louisiana Department of Education.





Copyright © 2017 by Louisiana Department of Education.

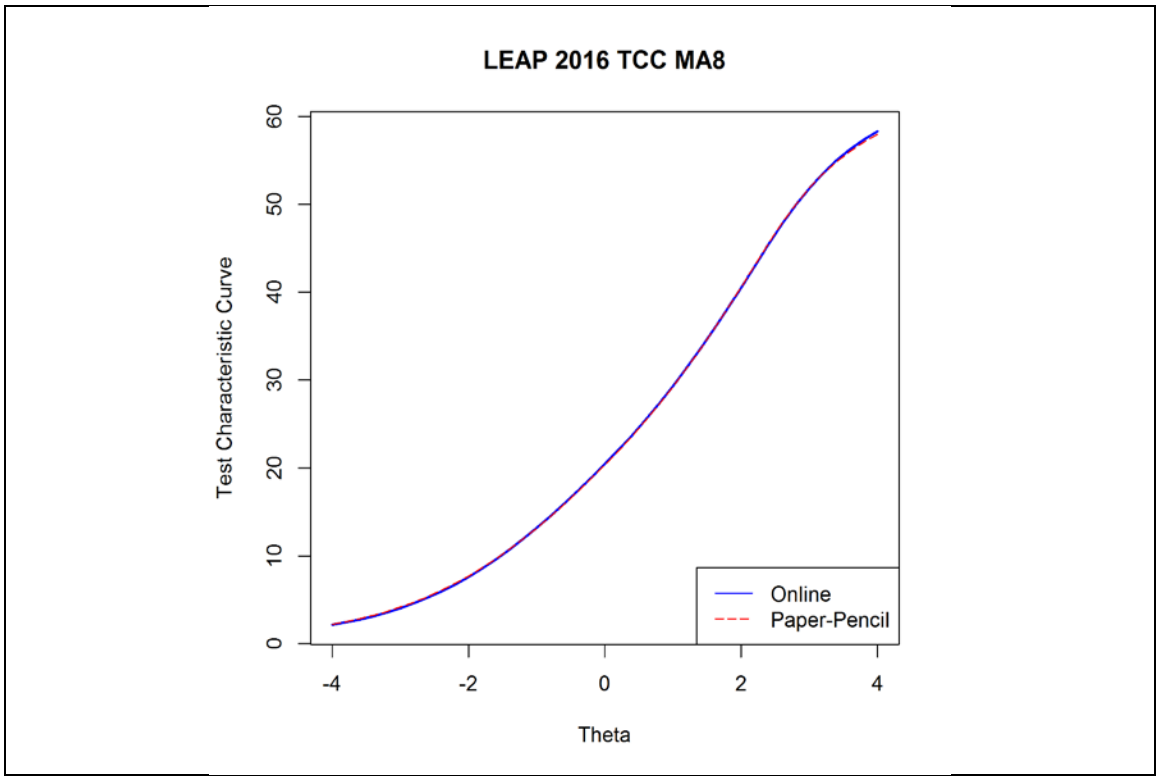
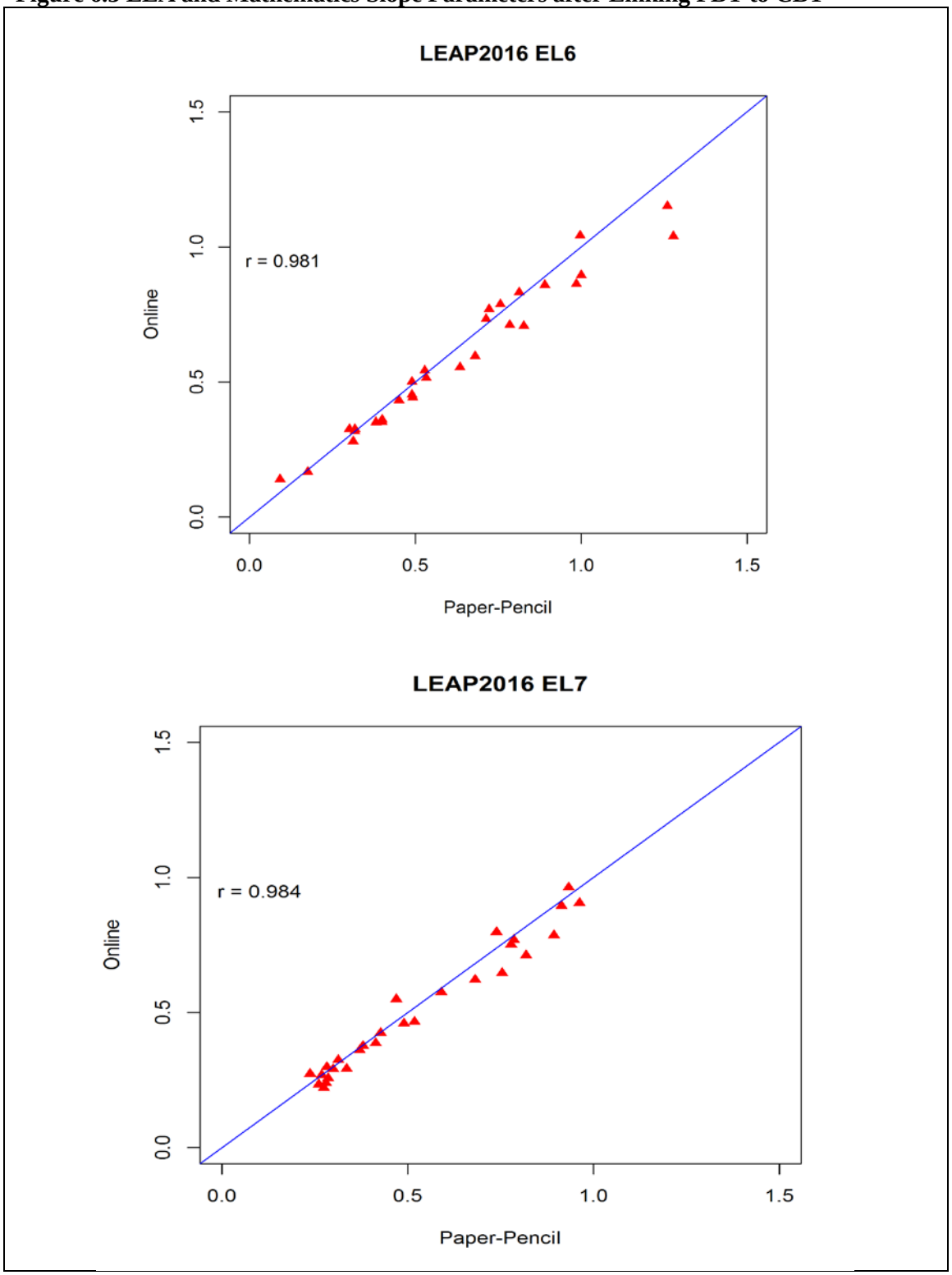


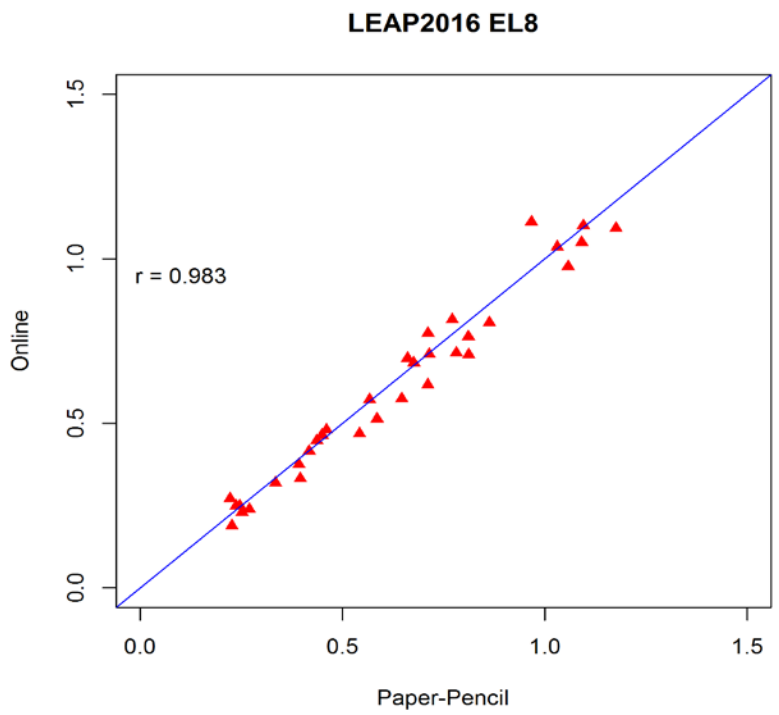
Figure 6.3 presents scatter plots of slope item parameters for CBT and PBT and their correlation after linking PBT to CBT for ELA and mathematics grades 6 to 8. The *x*-axis shows slope item parameters for PBT, and the *y*-axis presents slope item parameters for CBT. As can be seen in the figures, most item slope parameters were around the identity line. Since the slope item parameters for CBT and PBT were similar, all six correlations were above 0.96. Figure 6.4 presents scatter plots of difficulty parameters for CBT and PBT and their correlation after linking PBT to CBT. The *x*-axis shows difficulty item parameters for PBT, and the *y*-axis presents difficulty item parameters for CBT. The plots have more difficulty item parameters than slope parameters because there are more difficulty item parameters than slope parameters. Like slope parameters, most item difficulty parameters were around the identity line, and their correlations were well above 0.97.

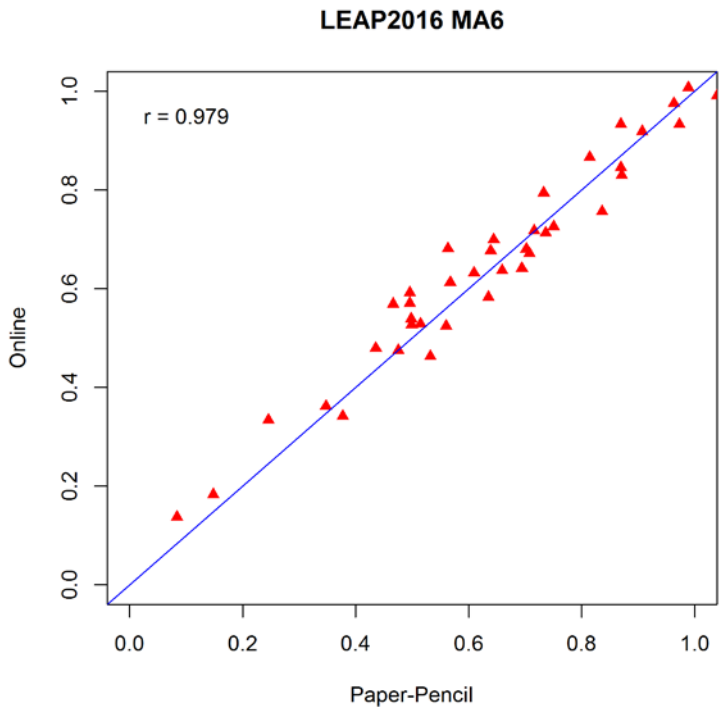
Copyright © 2017 by Louisiana Department of Education.

Figure 6.3 ELA and Mathematics Slope Parameters after Linking PBT to CBT

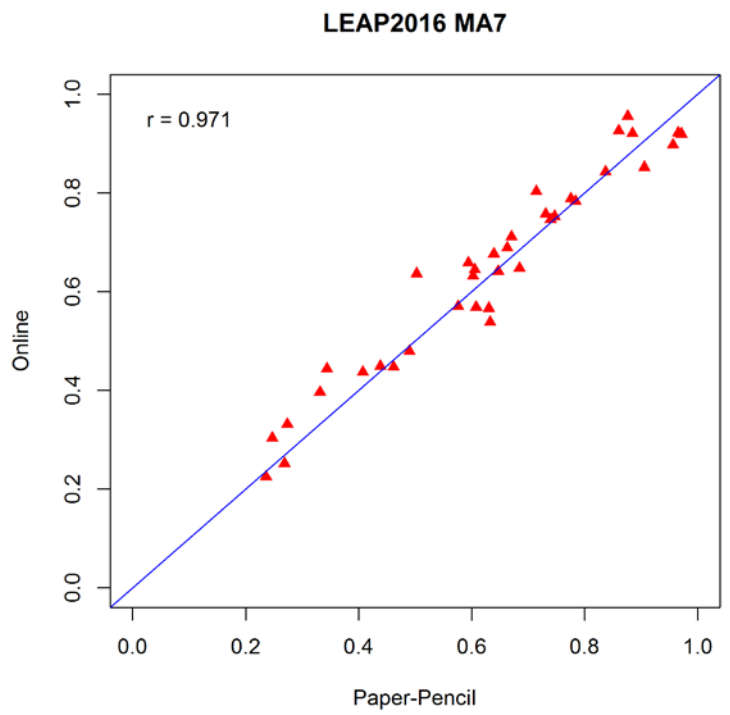


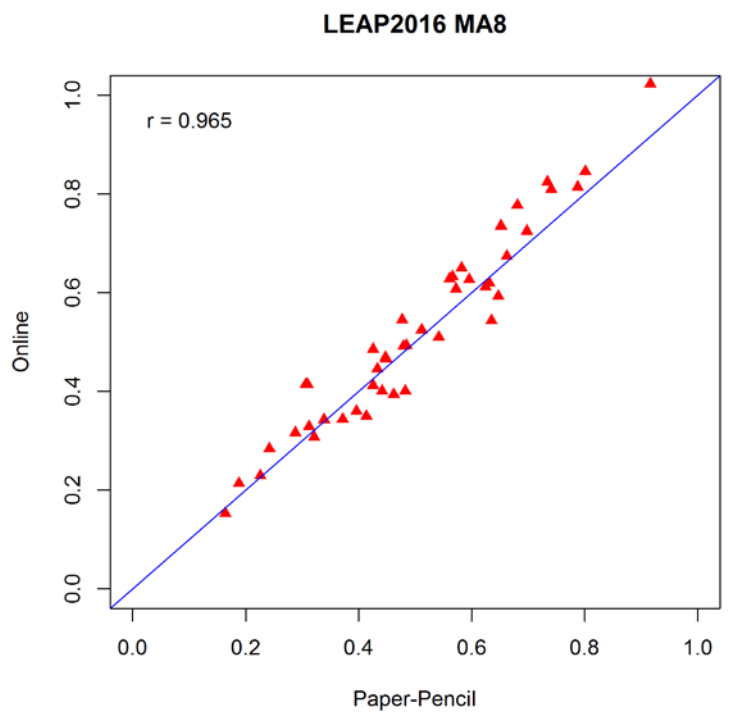
Copyright © 2017 by Louisiana Department of Education.





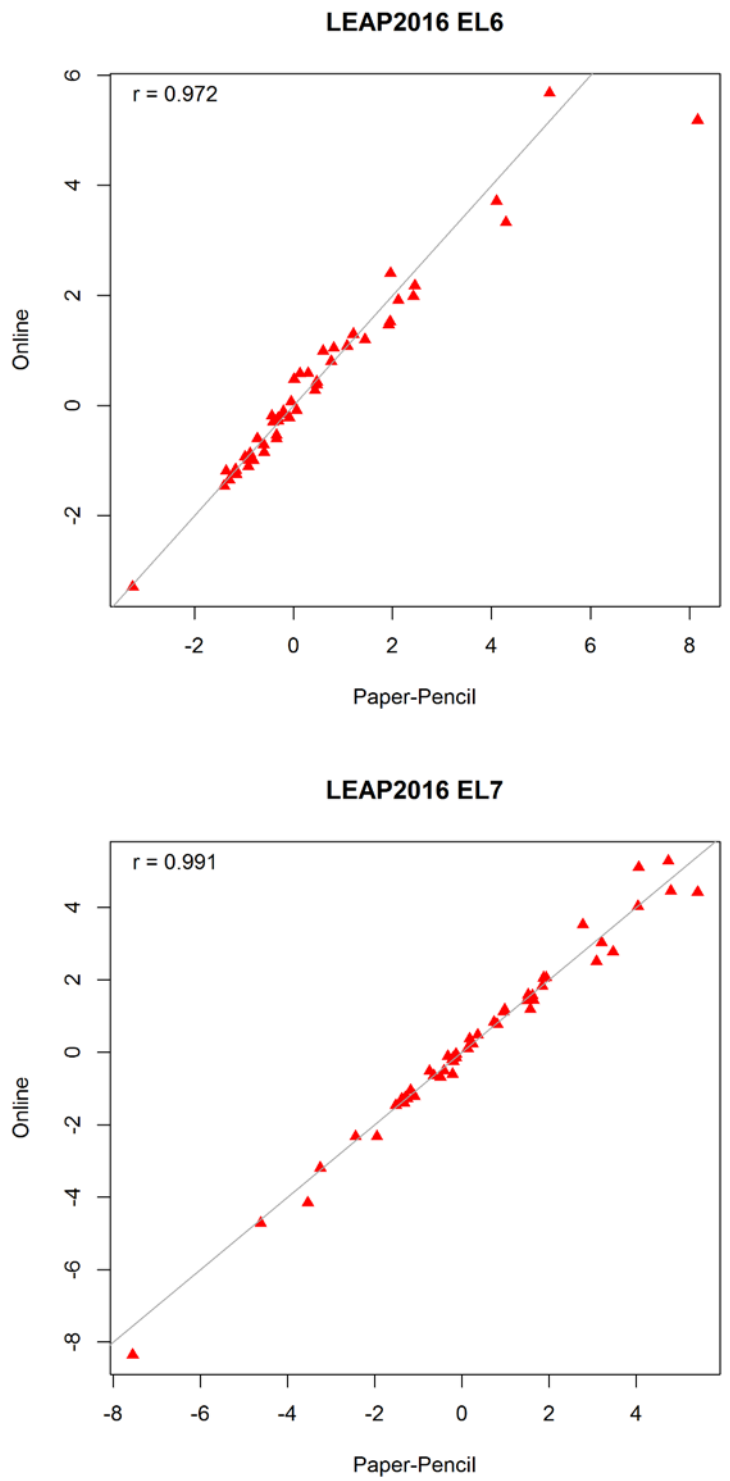
Copyright © 2017 by Louisiana Department of Education.



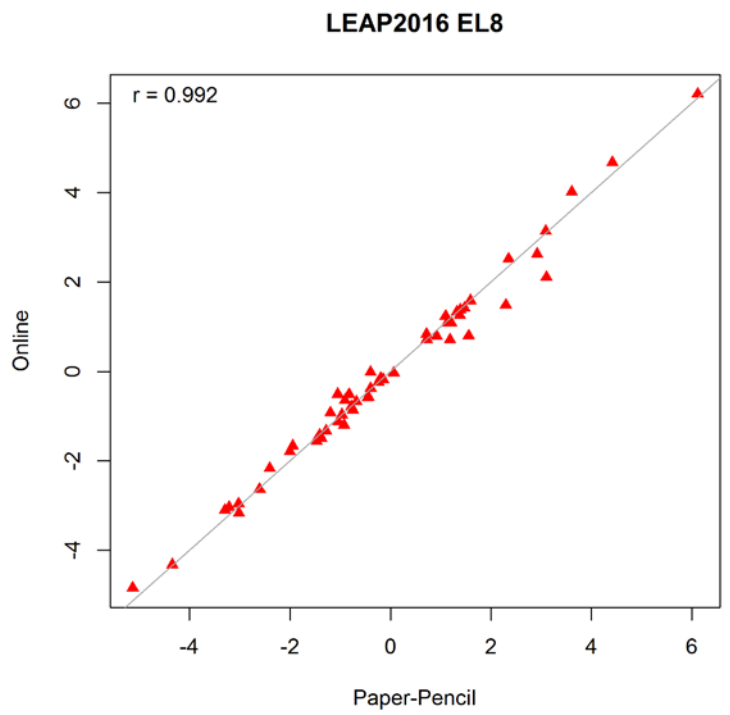


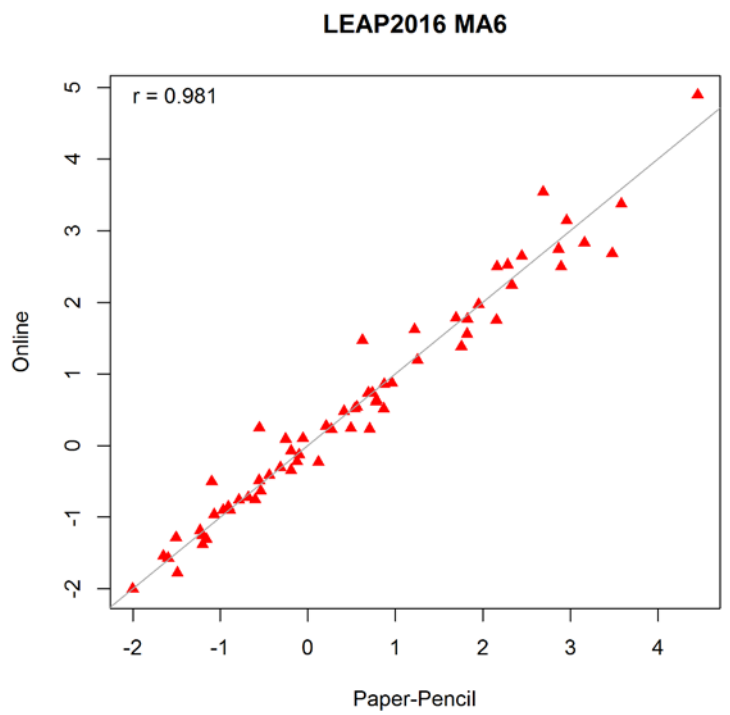
Copyright © 2017 by Louisiana Department of Education.

Figure 6.4 ELA and Mathematics Difficulty Parameters after Linking PBT to CBT

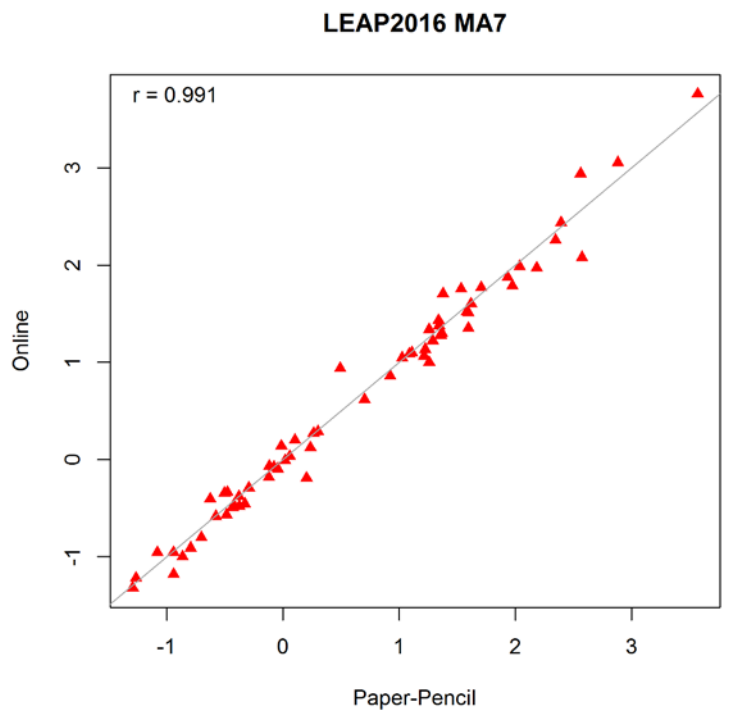


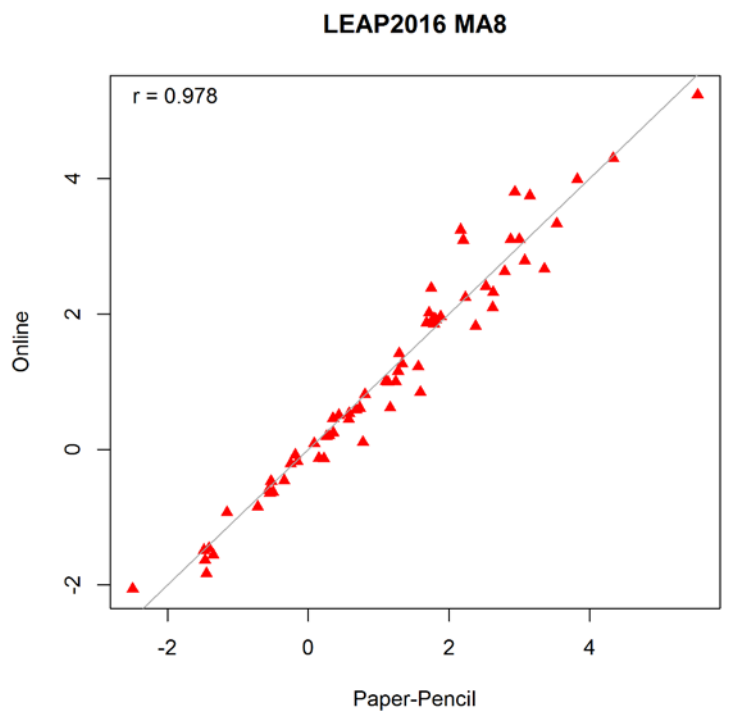
Copyright © 2017 by Louisiana Department of Education.





Copyright © 2017 by Louisiana Department of Education.





Copyright © 2017 by Louisiana Department of Education.

6.4.3 Linking LEAP 2016 to new PARCC 2016 baseline scale

After LEAP 2016 PBT was linked to LEAP 2016 CBT, LEAP 2016 CBT forms were linked to the PARCC scale using all PARCC items embedded in LEAP 2016 CBT as anchors by the Stocking and Lord procedure. Table 6.36 summarizes the number of anchor items and the Stocking and Lord transformation constants for ELA and mathematics grades 3 to 8. The difference between the initial number of anchor items and the final number of anchor items is the number of anchor items dropped.

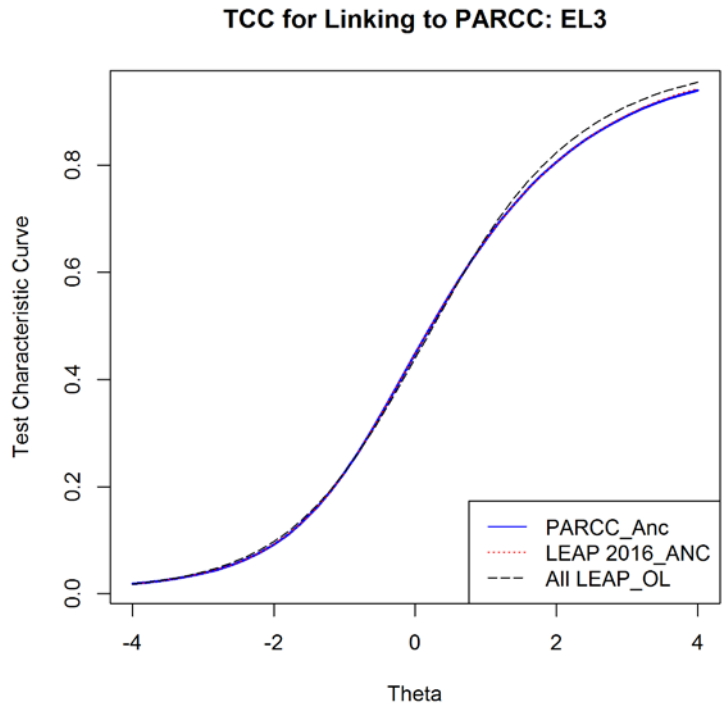
Table 6.36 Stocking and Lord Transformation Constants for Linking LEAP CBT 2016 to PARCC 2016

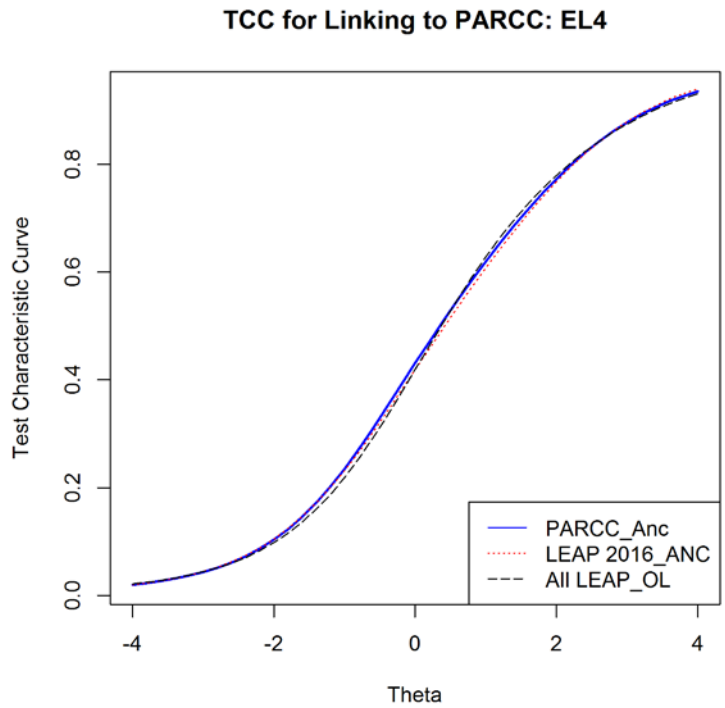
Content	Grade	Initial No. of Anchor Items	Final No. of Anchor Items	Slope	Intercept
ELA	3	15	13	0.87098	0.04207
	4	14	14	0.90251	0.06935
	5	15	13	0.91933	0.08710
	6	16	14	0.95472	0.19037
	7	16	14	0.98862	0.65841
	8	15	13	0.92659	0.22089
Mathematics	3	22	18	0.98872	0.07671
	4	21	16	1.01880	0.08576
	5	21	18	1.02084	-0.09682
	6	21	19	0.97344	-0.01105
	7	21	21	1.07433	0.08781
	8	21	17	0.80755	0.19705

Figures 6.5 and 6.6 show test characteristic curves for PARCC 2016 anchor items, corresponding LEAP 2016 anchor items (LEAP 2016 ANC), and all LEAP 2016 items (All LEAP_OL) for ELA and mathematics after the Stocking and Lord equating procedure. The blue solid line is for PARCC 2016 anchor items, the red dotted line is for LEAP 2016 anchor items, and the black dashed line is for all LEAP items. For ELA, the three TCCs were overlapped over most ability levels across all six grades. For ELA grade 6, the TCC for all LEAP items showed some differences compared to the other two TCCs. For mathematics, the three TCCs were also overlapped over most ability levels across all six grades. There were some differences at the extreme ranges, such as low ability or high ability.

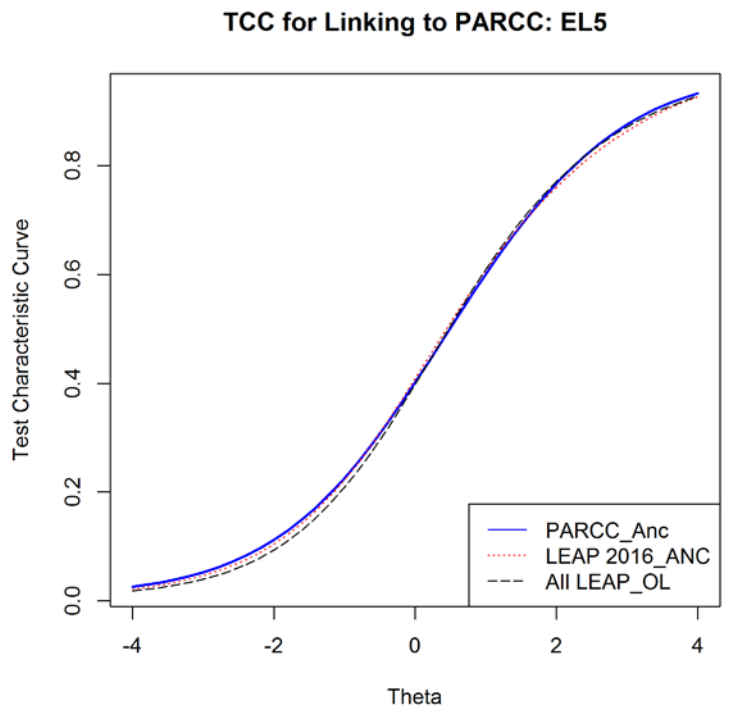
Copyright © 2017 by Louisiana Department of Education.

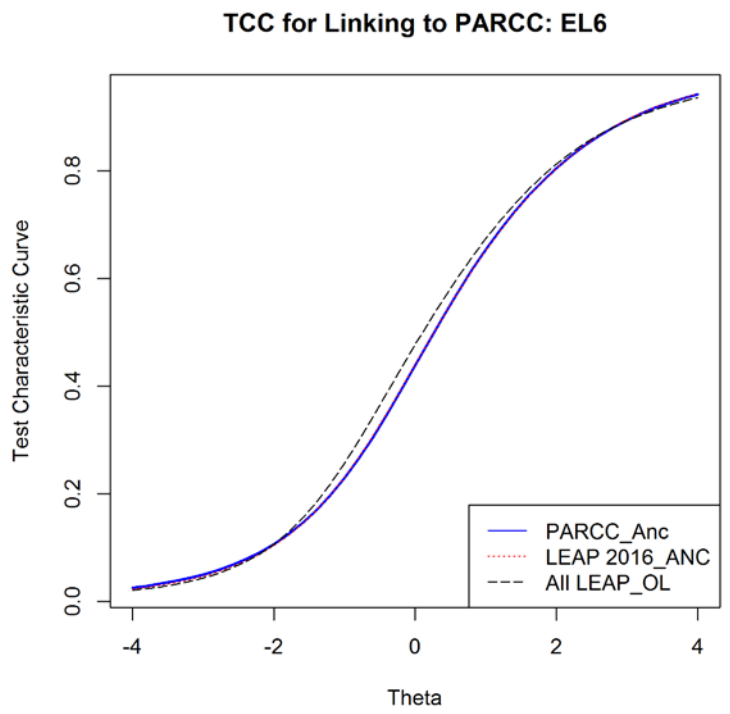
Figure 6.5 ELA TCC between PARCC 2016 Anchor, LEAP 2016 Anchor, and ALL LEAP Items



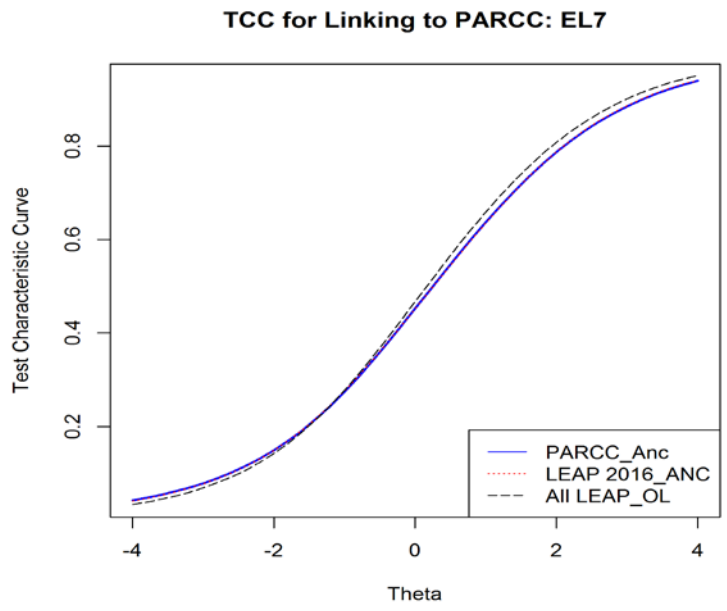


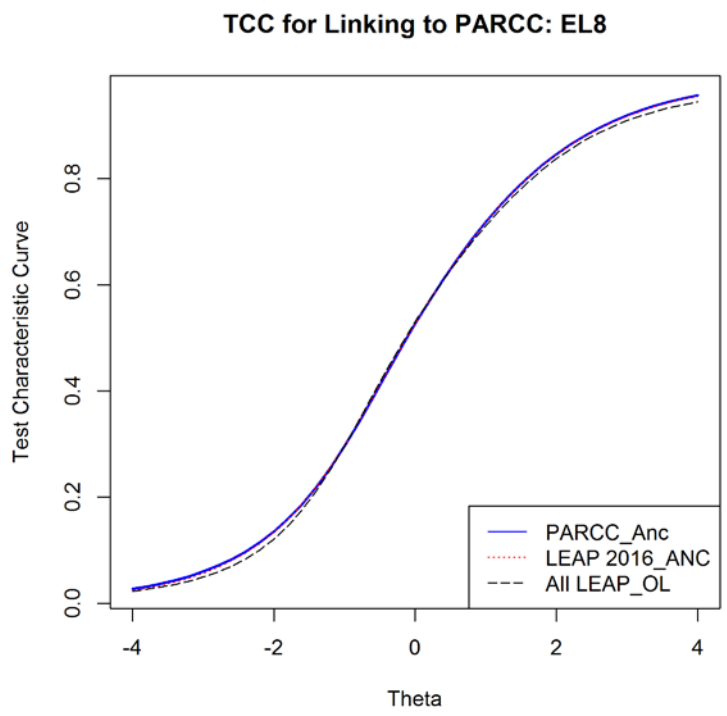
Copyright © 2017 by Louisiana Department of Education.





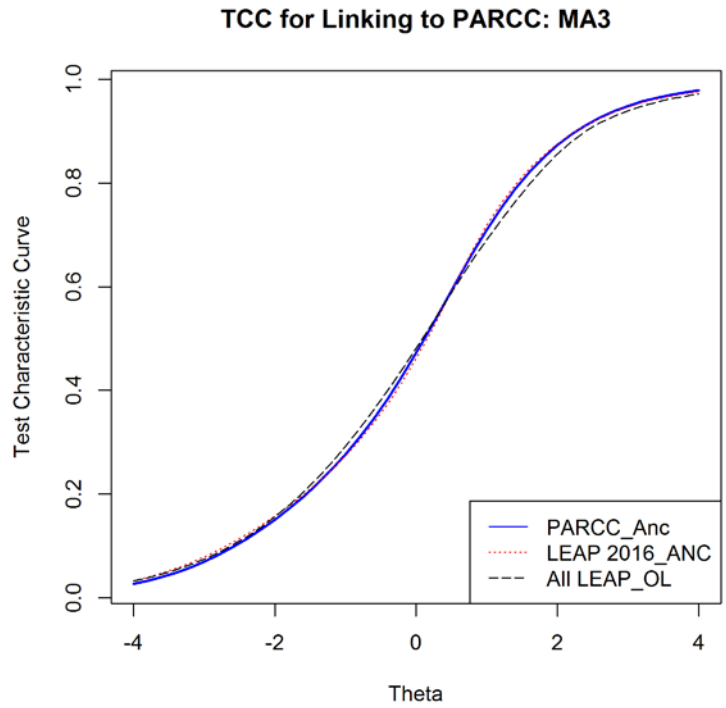
Copyright © 2017 by Louisiana Department of Education.

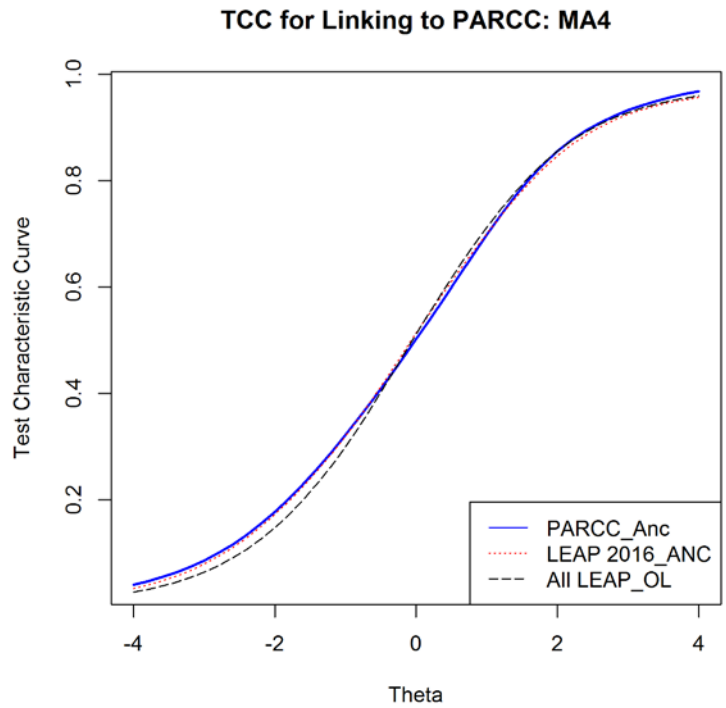




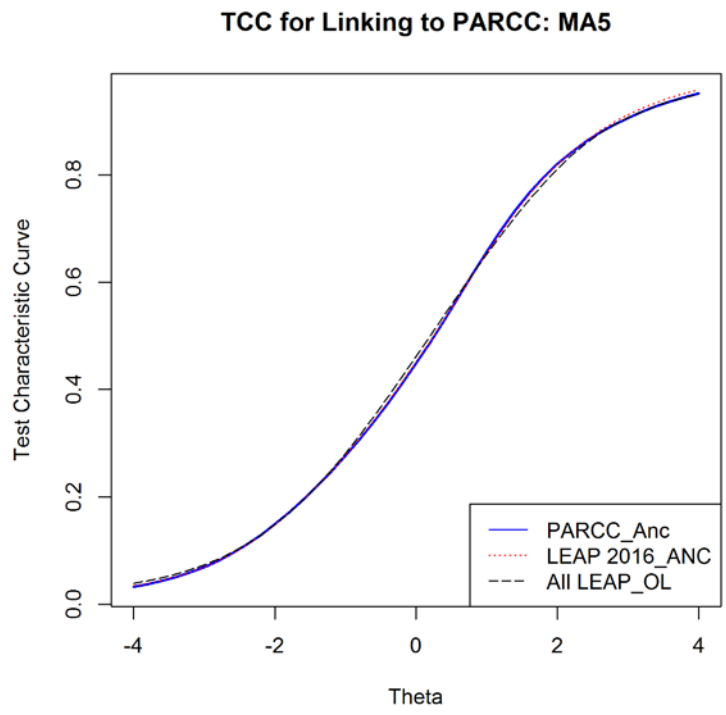
Copyright © 2017 by Louisiana Department of Education.

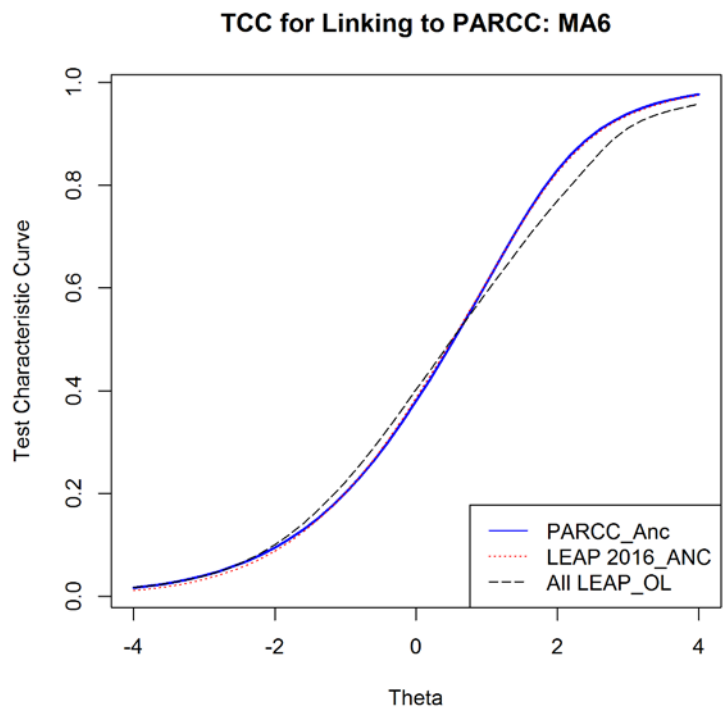
Figure 6.6 Mathematics TCC between PARCC 2016 Anchor, LEAP 2016 Anchor, and ALL LEAP Items



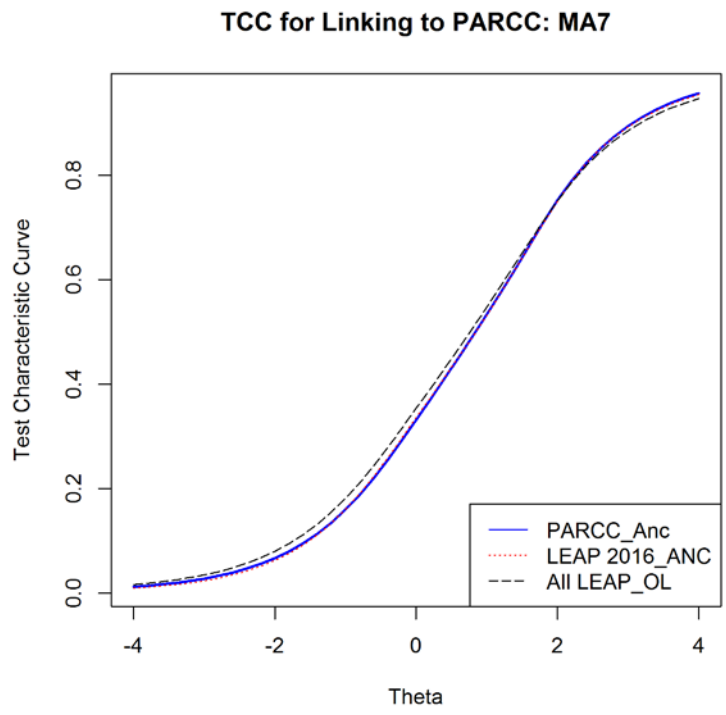


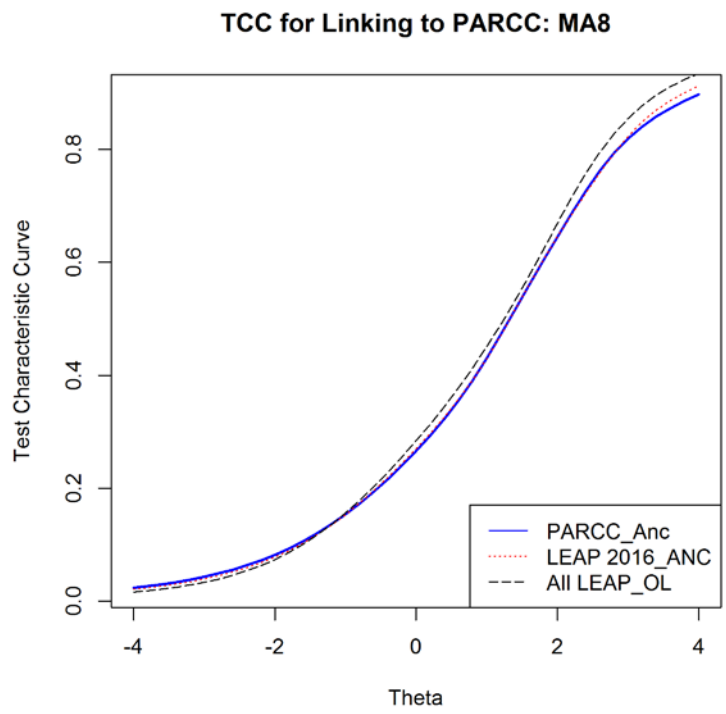
Copyright © 2017 by Louisiana Department of Education.





Copyright © 2017 by Louisiana Department of Education.



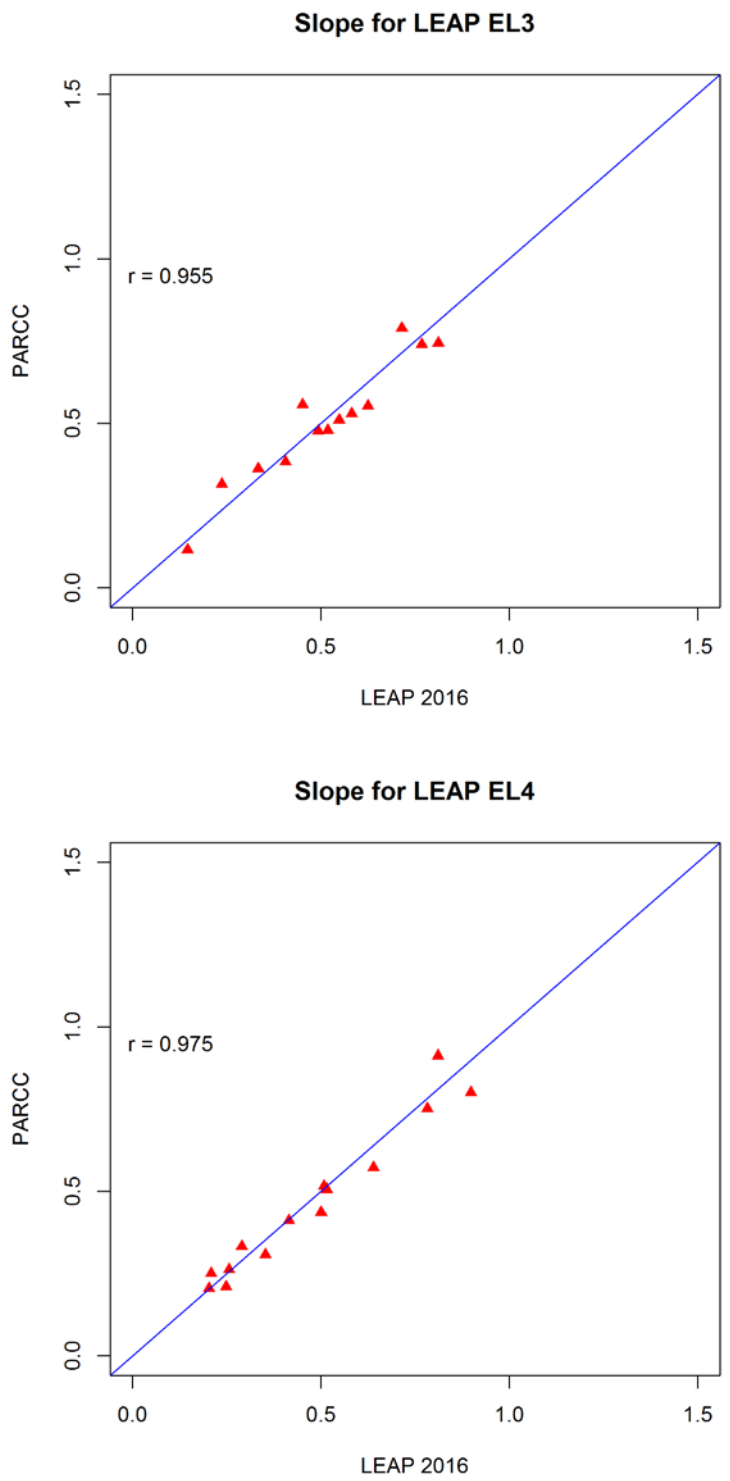


Figures 6.7 and 6.8 present scatter plots of ELA and mathematics slope item parameters and their correlation after linking LEAP 2016 to the PARCC 2016 scale. As can be seen in the ELA plots, most item slope parameters were around the identity line except for two items in ELA grade 5. The correlation between PARCC 2016 and LEAP 2016 anchors

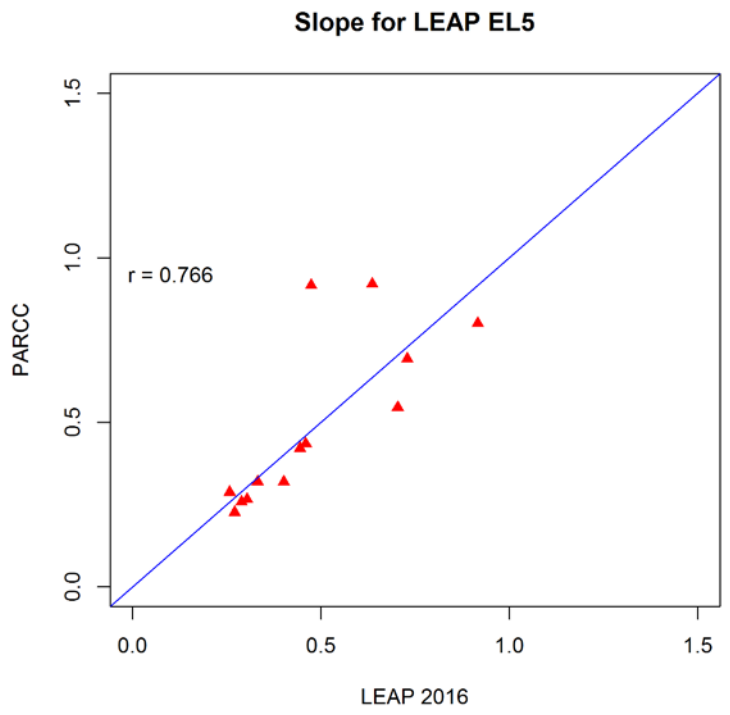
Copyright © 2017 by Louisiana Department of Education.

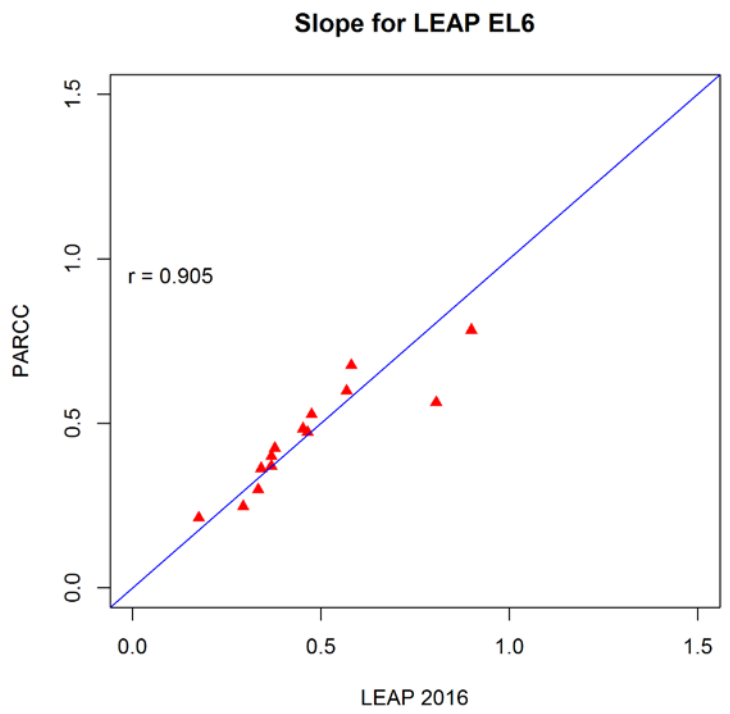
was 0.766 for ELA grade 5. For the other grades, correlations ranged from 0.905 to 0.985. For mathematics, most item slope parameters were around the identity line, and the correlations ranged from 0.795 to 0.933.

Figure 6.7 ELA Slope Parameters after Linking LEAP 2016 to PARCC Scale

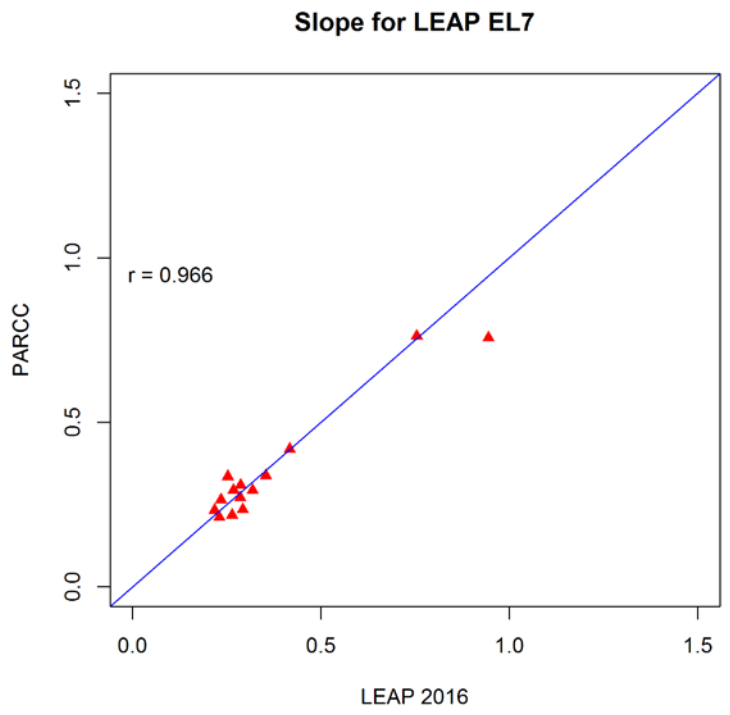


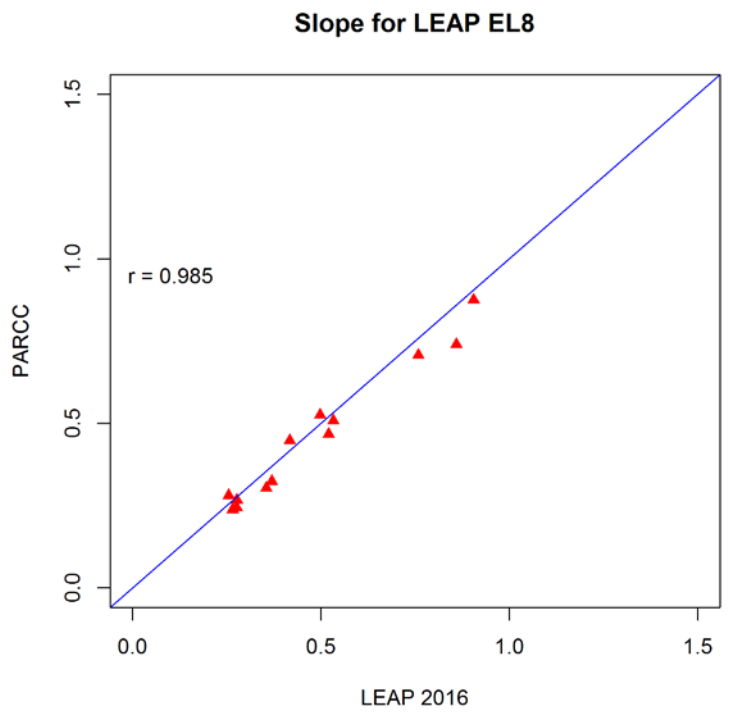
Copyright © 2017 by Louisiana Department of Education.





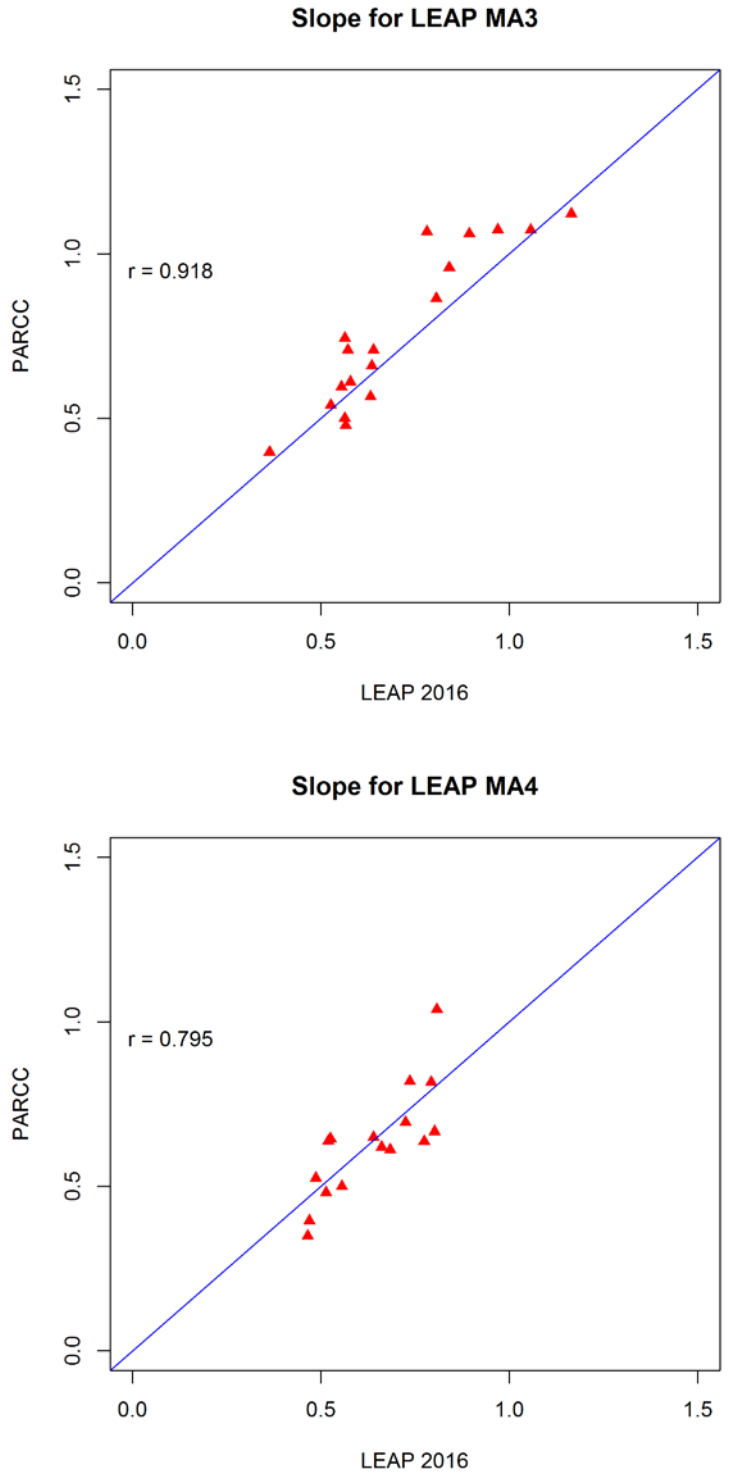
Copyright © 2017 by Louisiana Department of Education.



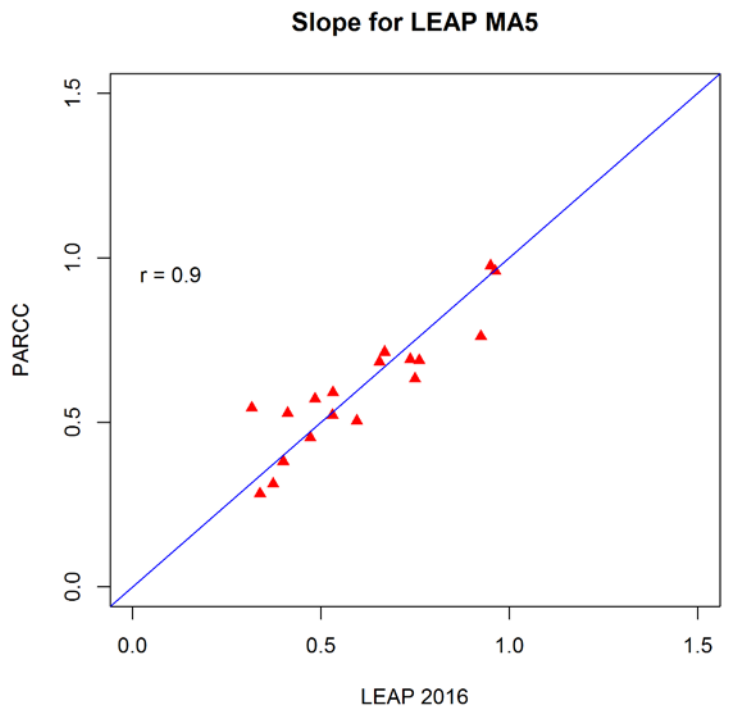


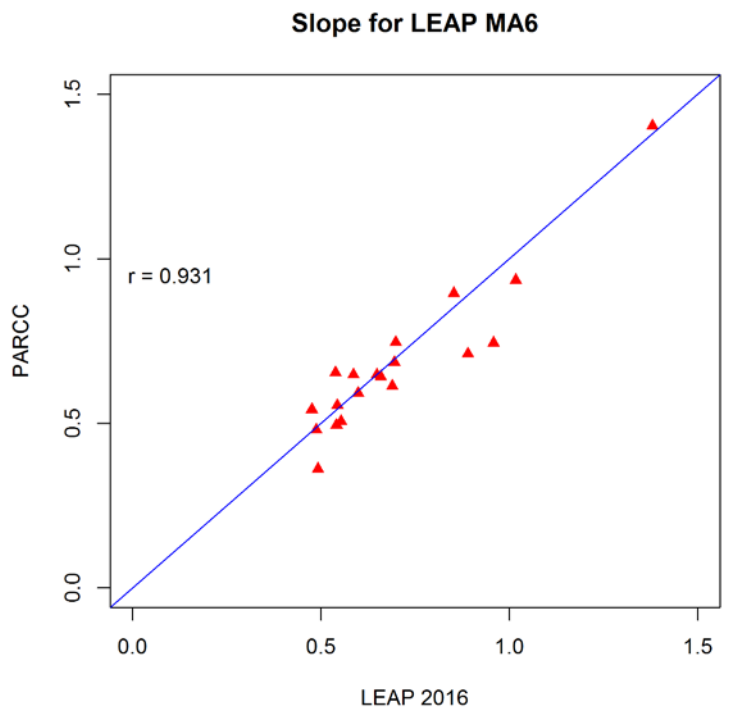
Copyright © 2017 by Louisiana Department of Education.

Figure 6.8 Mathematics Slope Parameters after Linking LEAP 2016 to PARCC Scale



Copyright © 2017 by Louisiana Department of Education.





Copyright © 2017 by Louisiana Department of Education.

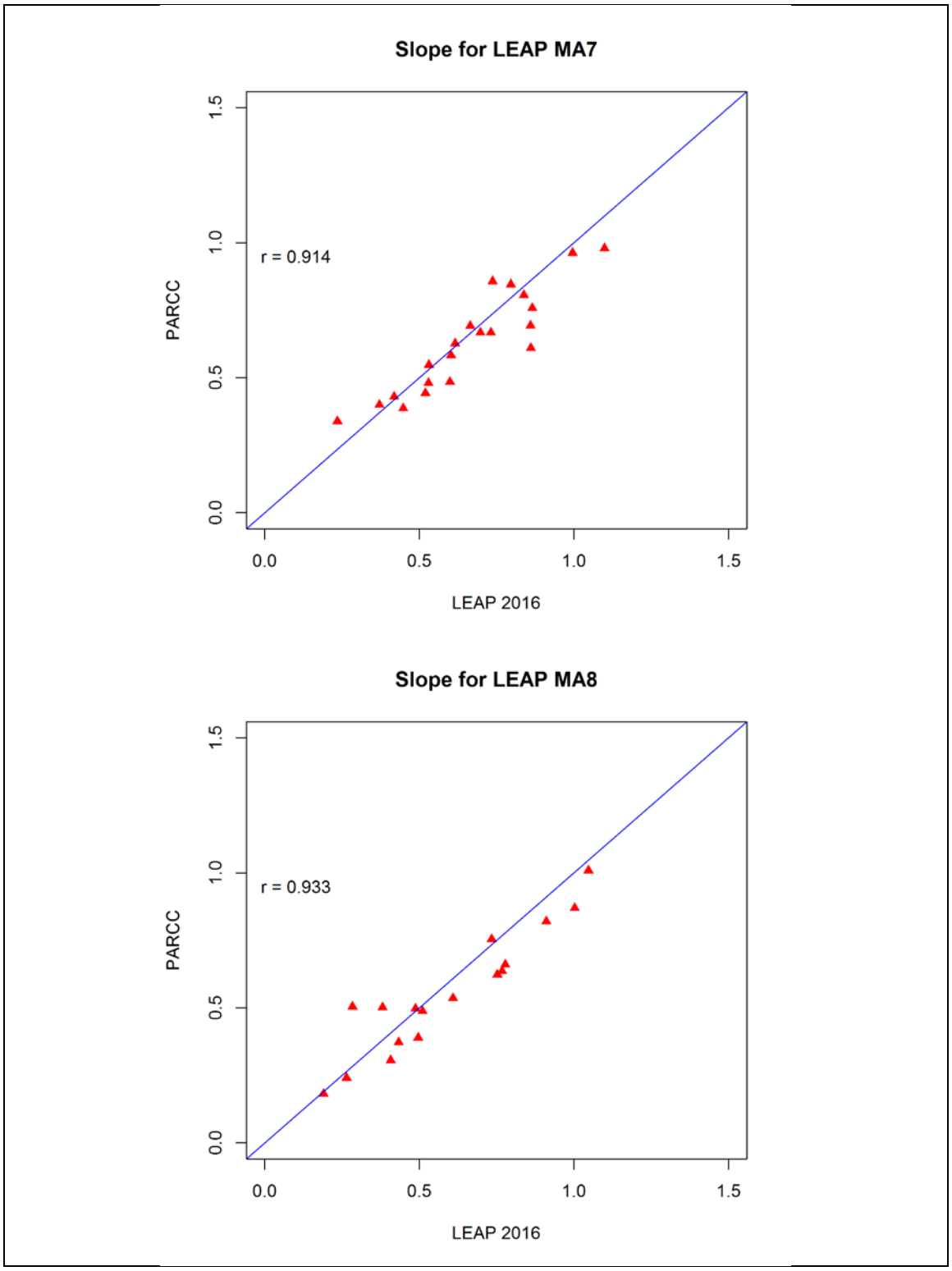
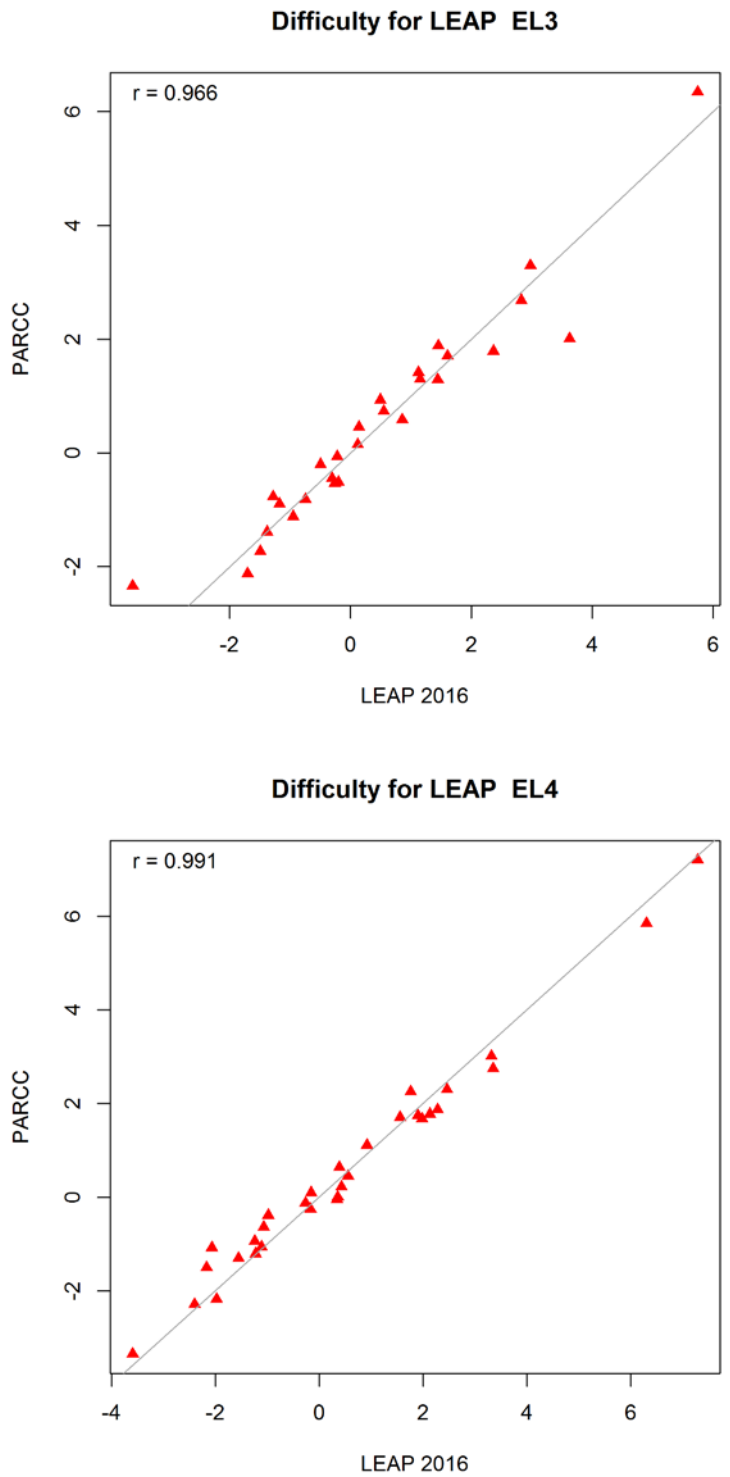


Figure 6.9 and 6.10 present scatter plots of the difficulty item parameters for ELA and mathematics and their correlation after linking LEAP 2016 to the PARCC 2016 scale. For ELA, most item difficulty parameters were around the identity line, and the correlations ranged from 0.966 to 0.991. Compared to ELA item difficulty parameters, mathematics

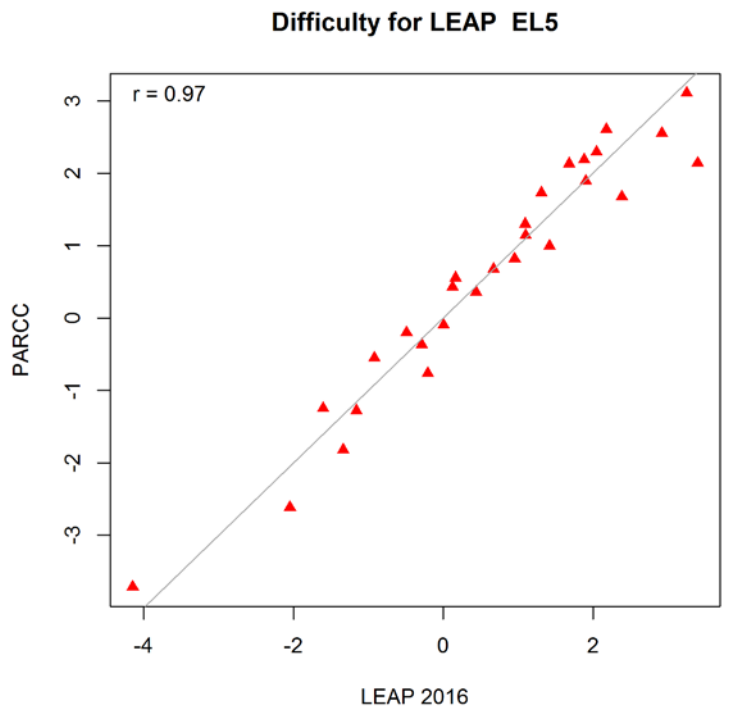
Copyright © 2017 by Louisiana Department of Education.

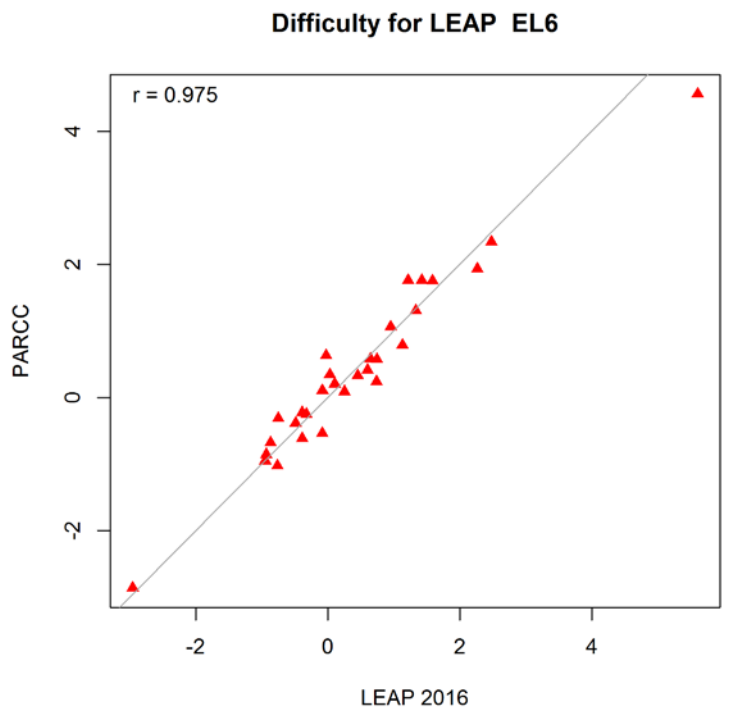
item difficulty parameters were slightly scattered from the identity line for most grades. Correlations ranged from 0.906 to 0.977 across grades.

Figure 6.9 ELA Difficulty Parameters after Linking LEAP 2016 to PARCC Scale

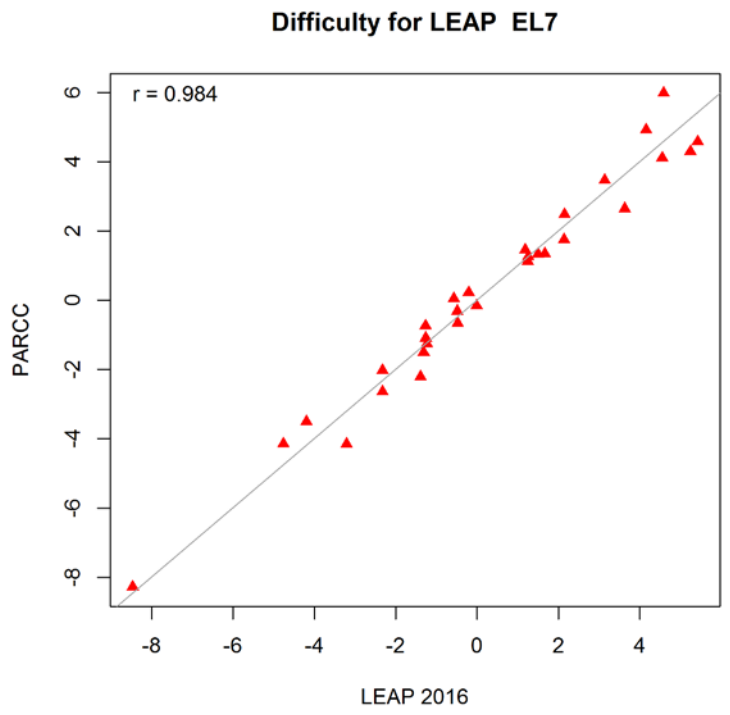


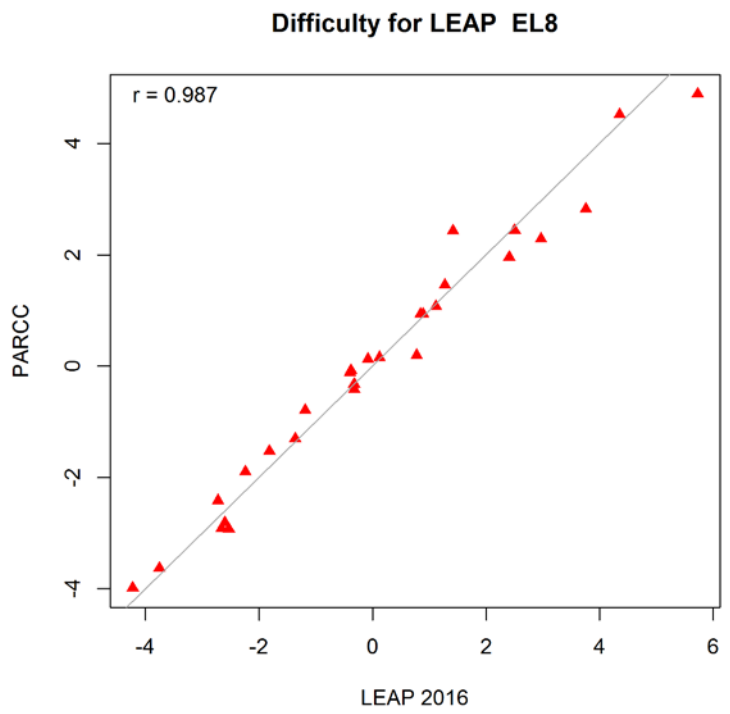
Copyright © 2017 by Louisiana Department of Education.





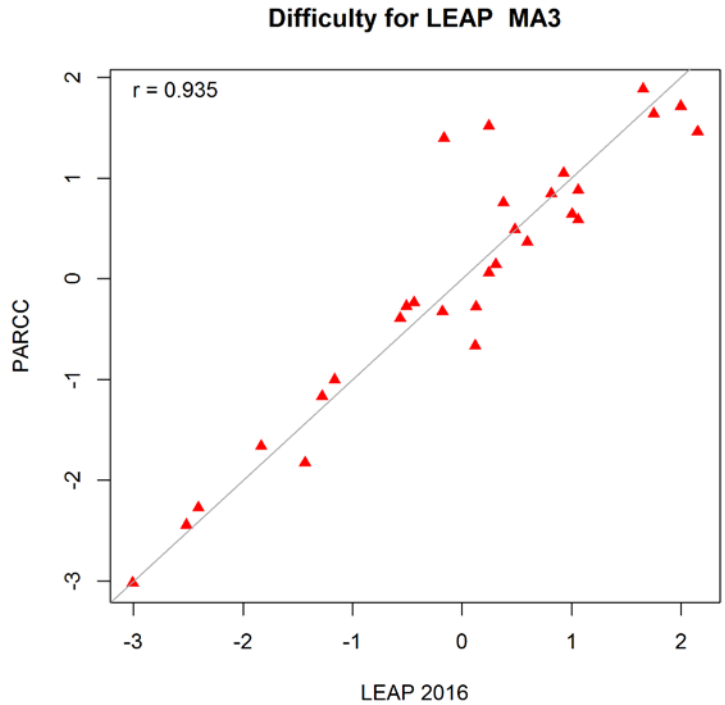
Copyright © 2017 by Louisiana Department of Education.

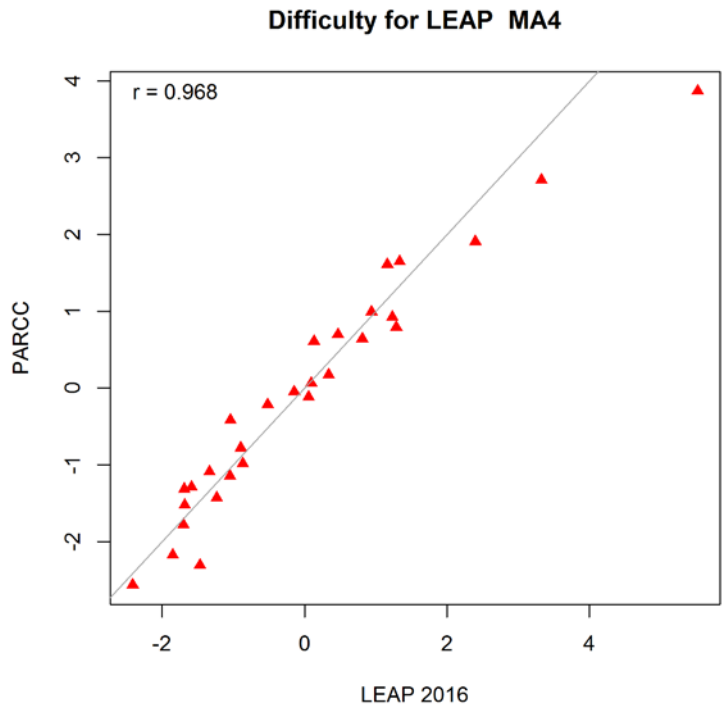




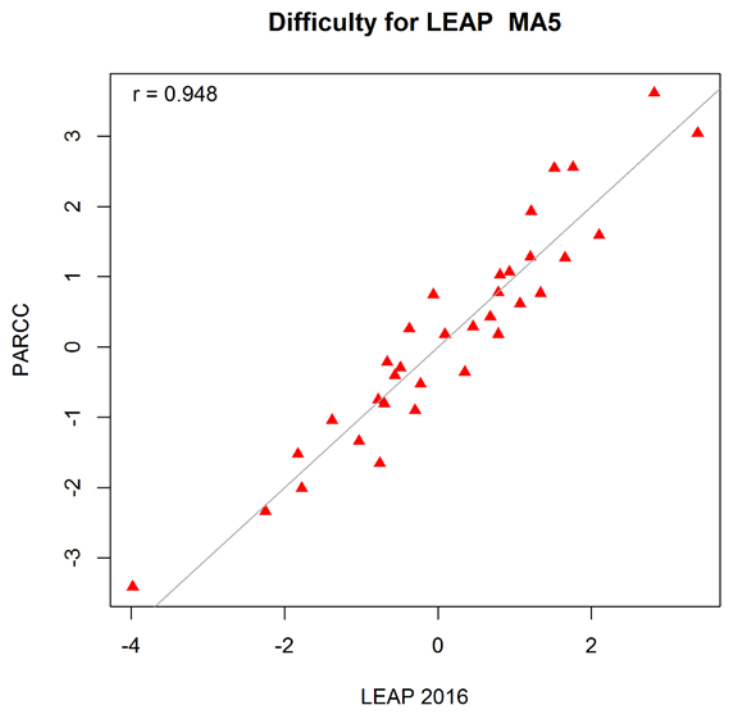
Copyright © 2017 by Louisiana Department of Education.

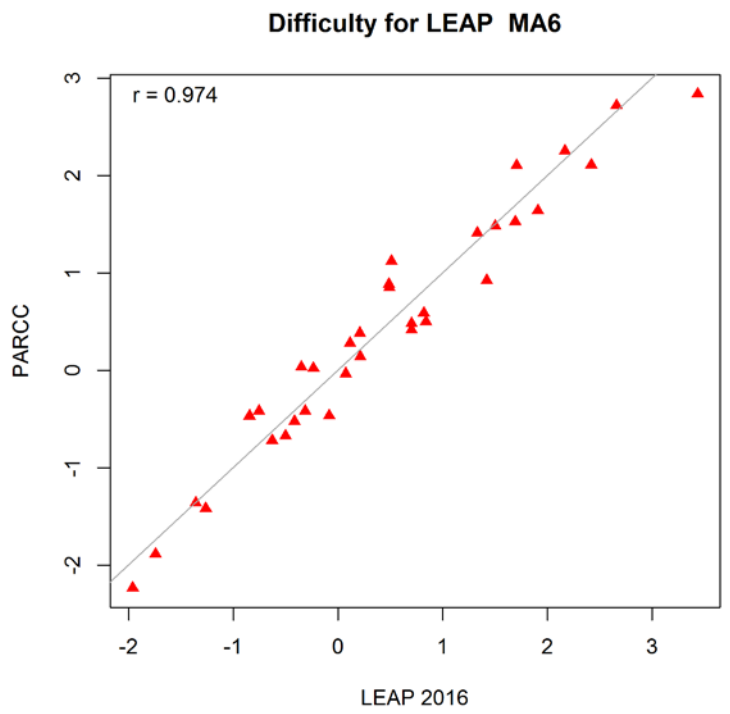
Figure 6.10 Mathematics Difficulty Parameters after Linking LEAP 2016 to PARCC Scale



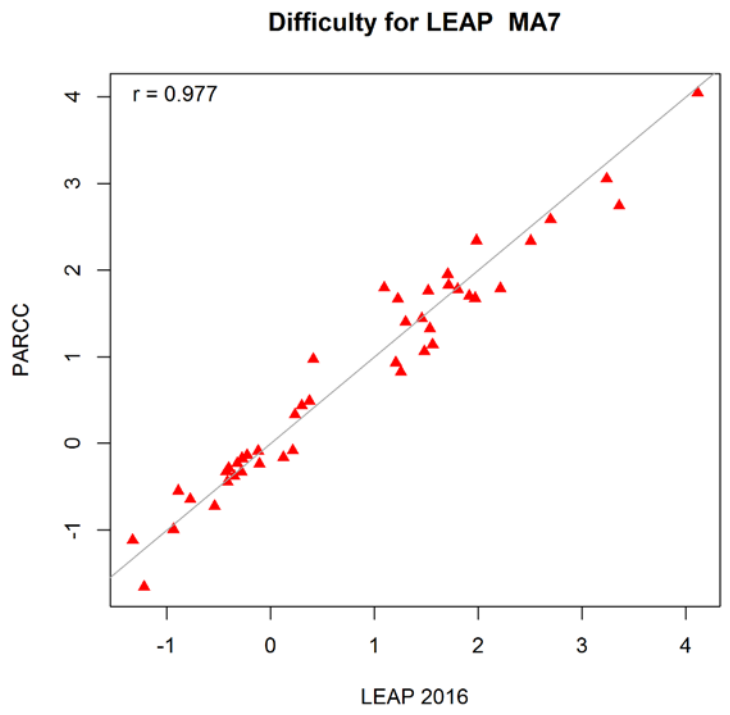


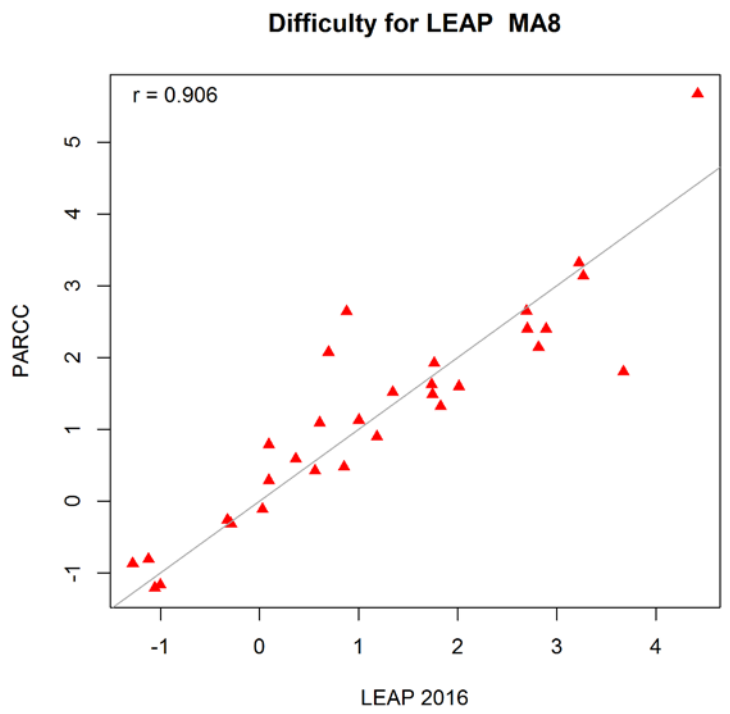
Copyright © 2017 by Louisiana Department of Education.





Copyright © 2017 by Louisiana Department of Education.





Copyright © 2017 by Louisiana Department of Education.

6.4.3.1. Evaluation of Anchor Item Stability

Standard 5.15 requires information about the anchors, stating the following:

> ***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)

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

Outliers of anchor items were reviewed with the weighted root mean square difference (WRMSD) method in addition to content perspective, such as number of items and score points for each claim/subclaim, and if approved by the LDOE, dropped from anchor sets and considered non-common anchor items during equating. All intact 2016 PARCC items were initially used as anchor items to link to PARCC 2016, and the quality of the anchor items was checked. Following are evaluation rules that were applied to check the quality of anchor items and the anchor set:

- Exclude CR items from anchor set if categories were collapsed due to small sample size.
- Exclude items with content or parameter estimation issues.
- Run STUIRT and flag items for further inspection if the weighted root mean square difference (WRMSD) was greater than the values in Table 6.37. If so, then the items were removed from the anchor set and the item characteristic curve (ICC) was reviewed together with WRMSD.
- Flag outliers using the plots of slope and difficulty item parameters with their correlations (Kolen and Brennan, 2014).
- Check score points and number of items by claim/subclaim before and after dropping an anchor item.

WRMSD values were calculated to compare with the ICCs using intact and estimated item parameters. For linking to PARCC 2016, PARCC 2016 item parameters were intact, and their corresponding LEAP 2016 calibrated item parameters were estimated item parameters.

WRMSD is defined as

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

Copyright © 2017 by Louisiana Department of Education.

where *Q* represents a quadrature point (or node), *W* represents its weight given quadrature point *Q* from PARCALE PH2 output, *INTACT* represents intact PARCC item parameters, and '*EST*' represents estimated item parameters corresponding to intact PARCC item parameters. Table 6.37 summarizes WRMSD flagging criteria for inspection and possible removal of linking items (from PARCC 2016 linking chapter).

Table 6.37 PARCC WRMSD Flagging Criteria

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

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

6.4.4 Linking LEAP 2015 to New PARCC 2016 Baseline Scale

Although PARCC had created a scale using the PARCC 2015 administration, they updated the scale using the PARCC 2016 administration. The 2015 PARCC scale was linked to the new 2016 baseline scale by the Stocking and Lord (1983) procedure. LEAP 2015 scores and item parameters from the PARCC 2016 tests were also updated using the transformation constants provided by PARCC. Table 6.38 shows the transformation constants, slope, and intercept, for this purpose.

Copyright © 2017 by Louisiana Department of Education.

Table 6.38 Transformation Constants from PARCC 2015 Scale to the PARCC 2016 Scale

Content	Grade	Slope	Intercept
ELA	3	0.9864	-0.0012
	4	0.9759	-0.0047
	5	1.0004	0.0363
	6	1.0145	0.0191
	7	0.9991	-0.0747
	8	1.0062	-0.1229
Mathematics	3	0.9918	-0.184
	4	0.988	-0.123
	5	0.9983	-0.1863
	6	1.0182	-0.0899
	7	1.0229	-0.0673
	8	1.0185	-0.0278

Lowest and Highest Obtainable Scaled Scores

A maximum likelihood procedure cannot produce scaled score estimates for students with perfect scores or scores below the level expected by guessing. In addition, although maximum likelihood 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 necessarily non-maximum likelihood procedure. These values, which are set separately by grade, are called the lowest obtainable scale score (LOSS) and the highest obtainable scale score (HOSS). All grades and subjects in LEAP 2016 used the same LOSS and HOSS values. The LOSS value is 650 and the HOSS value is 850.

6.4.4.1. Claim-Level/Subclaim-Level Scores

In addition to total test scaled scores, the claim-level scores were also computed for ELA and subclaim-level scores were computed for ELA and mathematics. They are used to classify students in one of the three performance levels on a claim and subclaim: *Weak Performance*, *Moderate Performance*, or *Strong Performance*.

Raw-score-to-scale-score scoring tables for each total test, claim, and subclaim were generated using the item parameters placed on 2016 PARCC scale. A raw-score-to-scale-score table for claim or subclaim level includes only the items that have their claim or subclaim reporting category. For all scoring tables, the TCC inverse method was applied to generate raw-to-scaled-score-tables following PARCC approach. LEAP 2016 Reporting Categories were the following:

Copyright © 2017 by Louisiana Department of Education.

English Language Arts

Objective 2 (Claim)

1. Reading
2. Writing

Objective 1 (Subclaim)

1. Reading—RI
2. Reading—RL
3. Reading—RV
4. Writing—WE
5. Writing—WKL

Mathematics

Objective 1 (Subclaim)

1. Math Subclaim A
2. Math Subclaim B
3. Math Subclaim C
6. Math Subclaim D

Reading and writing claim scores were reported for ELA assessments only. The reading claim score range is 10–90 and the writing claim score range is 10–60. Scaling claim followed PARCC methodology as described in "PARCC Spring 2016 Subclaim Performance Reporting Rules." For reading claim, two theta score points corresponding to ELA scaled scores of 700 and 750 were used for scaling. Linear transformation constants mapping the two theta points to scaled score points of 30 and 50 were calculated. After these transformation values were applied to item parameters belonging to reading claim, a scoring table was generated using the TCC inverse method. A similar approach was applied to scale writing claim using two scaled score points of 30 and 35. There are three performance levels for each claim. Two cut scores, 40 and 50 for reading and 30 and 35 for writing were used to produce three performance levels in each claim (See Table 6.38 for cut scores for summative, claim and subclaim).

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

Table 6.39 Cut scores for Summative, Claim, and Subclaim

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

*Subclaim thetas are those of summative tests, 725 and 750.

* Yellow highlight shows cut scores for claim and subclaim

Copyright © 2017 by Louisiana Department of Education.

6.5 Summary

In summary, the overall purpose of the operational data analysis is to ensure that the test items, as well as the overall test, are functioning appropriately. It also helps maintain the test scale across the years 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, in particular, are related to the following *Standards*:

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

Copyright © 2017 by Louisiana Department of Education.

CHAPTER 7: TEST RESULTS

This chapter of the technical report contains information on the results of the Spring 2016 LEAP administration of ELA and mathematics. The scaled score results are presented here. Achievement-level information is also provided. Presenting the results by achievement level translates the quantitative scale provided through scaled scores into a qualitative description of student achievement: *Advanced, Mastery, Basic, Approaching Basic,* and *Unsatisfactory.*

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

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

Results presented below are based on census data. The results presented here may differ slightly from the official state summary report of all student populations due to ongoing resolution of test materials and student information. The results in the tables in this chapter are presented as evidence of reliability and validity of the scores from the LEAP 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 tests:

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

For the purposes of this report, participation rate is defined as the percentage of students who received a valid scaled score given the total number of students who were expected to take the online test or receive a test book. These participation rates are summarized in Table 7.1. Both the percentage of students classified as reportable and the number of students classified as accountable are reported.. Reportable students include all students

Copyright © 2017 by Louisiana Department of Education.

with a valid scaled score. The Accountable column shows the total number of students who were expected to take the online test or received a test book. These include students who should have received a LEAP scaled score but who did not take the test and could not be assigned a scaled score.

Table 7.1.: Participation Rates

Participation Rates by Grade and Subgroup					
Grade	**Group**	**Accountable in ELA**	**Percent Reportable in ELA**	**Accountable in Mathematics**	**Percent Reportable in Mathematics**
3	**All Students**	≥58,180	98.20%	≥58,380	98.90%
	Gender				
	Female	≥28,350	98.43%	≥28,460	99.05%
	Male	≥29,710	98.29%	≥29,810	98.99%
	Ethnicity				
	Hispanic/Latino	≥3,800	97.79%	≥3,820	98.40%
	American Indian or Alaska Native	≥360	98.62%	≥360	99.45%
	Asian	≥800	99.25%	≥800	99.50%
	Black or African American	≥25,710	98.12%	≥25,820	98.92%
	Native Hawaiian or Other Pacific	≥50	100.00%	≥50	100.00%
	White	≥25,750	98.49%	≥25,810	99.02%
	Two or More Races	≥1,570	98.67%	≥1,580	98.99%
	Education Classification				
	Regular	≥52,000	98.25%	≥52,180	98.96%
	Special	≥6,180	97.80%	≥6,200	98.39%
	Economic Status				
	Economically Disadvantaged	≥41,700	99.04%	≥41,870	99.05%
	Not Economically Disadvantaged	≥15,000	99.13%	≥15,010	99.25%
	LEP Status				
	Non-LEP	≥55,920	98.22%	≥56,110	98.92%
	LEP	≥2,250	97.79%	≥2,270	98.33%
	Migrant Status				
	Nonmigrant	≥58,000	98.20%	≥58,210	98.90%
	Migrant	≥170	97.11%	≥170	97.71%
	504 Status				
	Non-504	≥53,270	98.18%	≥53,460	98.91%
	504	≥4,900	98.45%	≥4,920	98.76%

Copyright © 2017 by Louisiana Department of Education.

Participation Rates by Grade and Subgroup (continued)					
Grade	**Group**	**Accountable in ELA**	**Percent Reportable in ELA**	**Accountable in Mathematics**	**Percent Reportable in Mathematics**
4	**All Students**	≥54,960	98.27%	≥55,160	98.93%
	Gender				
	Female	≥26,650	98.35%	≥26,750	98.96%
	Male	≥28,240	98.29%	≥28,340	98.95%
	Ethnicity				
	Hispanic/Latino	≥3,340	97.40%	≥3,370	98.28%
	American Indian or Alaska Native	≥340	98.83%	≥340	99.42%
	Asian	≥810	99.38%	≥810	99.63%
	Black or African American	≥24,800	98.11%	≥24,920	98.86%
	Native Hawaiian or Other Pacific	≥50	96.08%	≥50	100.00%
	White	≥24,190	98.63%	≥24,250	99.09%
	Two or More Races	≥1,300	99.08%	≥1,310	99.62%
	Education Classification				
	Regular	≥49,130	98.30%	≥49,310	98.95%
	Special	≥5,820	98.03%	≥5,850	98.75%
	Economic Status				
	Economically Disadvantaged	≥39,110	99.06%	≥39,290	99.11%
	Not Economically Disadvantaged	≥14,610	98.73%	≥14,620	98.78%
	LEP Status				
	Non-LEP	≥53,150	98.34%	≥53,340	98.98%
	LEP	≥1,810	96.25%	≥1,820	97.43%
	Migrant Status				
	Nonmigrant	≥54,830	98.26%	≥55,030	98.93%
	Migrant	≥130	100.00%	≥130	100.00%
	504 Status				
	Non-504	≥49,060	98.32%	≥49,240	98.99%
	504	≥5,890	97.86%	≥5,910	98.43%

Copyright © 2017 by Louisiana Department of Education.

Participation Rates by Grade and Subgroup (continued)					
Grade	**Group**	**Accountable in ELA**	**Percent Reportable in ELA**	**Accountable in Mathematics**	**Percent Reportable in Mathematics**
5	**All Students**	≥53,610	98.08%	≥53,780	98.75%
	Gender				
	Female	≥25,980	98.10%	≥26,060	98.73%
	Male	≥27,600	98.13%	≥27,680	98.78%
	Ethnicity				
	Hispanic/Latino	≥3,150	97.05%	≥3,180	98.18%
	American Indian or Alaska Native	≥370	99.19%	≥370	99.46%
	Asian	≥790	99.50%	≥790	99.62%
	Black or African American	≥23,820	98.20%	≥23,900	98.98%
	Native Hawaiian or Other Pacific	≥40	95.45%	≥40	100.00%
	White	≥24,270	98.16%	≥24,320	98.57%
	Two or More Races	≥1,080	99.17%	≥1,080	99.26%
	Education Classification				
	Regular	≥47,800	98.10%	≥47,950	98.75%
	Special	≥5,810	97.99%	≥5,830	98.71%
	Economic Status				
	Economically Disadvantaged	≥37,700	98.76%	≥37,820	98.86%
	Not Economically Disadvantaged	≥14,730	98.63%	≥14,750	98.67%
	LEP Status				
	Non-LEP	≥52,240	98.17%	≥52,380	98.81%
	LEP	≥1,370	94.99%	≥1,390	96.35%
	Migrant Status				
	Nonmigrant	≥53,490	98.08%	≥53,650	98.74%
	Migrant	≥120	99.19%	≥120	100.00%
	504 Status				
	Non-504	≥47,410	98.12%	≥47,560	98.78%
	504	≥6,200	97.81%	≥6,210	98.49%

Copyright © 2017 by Louisiana Department of Education.

Participation Rates by Grade and Subgroup (continued)					
Grade	**Group**	**Accountable in ELA**	**Percent Reportable in ELA**	**Accountable in Mathematics**	**Percent Reportable in Mathematics**
6	**All Students**	≥53,690	98.17%	≥53,860	98.91%
	Gender				
	Female	≥26,210	98.33%	≥26,280	99.01%
	Male	≥27,420	98.17%	≥27,520	98.88%
	Ethnicity				
	Hispanic/Latino	≥2,960	97.37%	≥2,980	98.49%
	American Indian or Alaska Native	≥410	98.56%	≥420	99.52%
	Asian	≥830	99.28%	≥830	99.88%
	Black or African American	≥24,300	98.06%	≥24,400	98.87%
	Native Hawaiian or Other Pacific	≥40	95.74%	≥40	100.00%
	White	≥24,070	98.55%	≥24,130	99.04%
	Two or More Races	≥940	98.94%	≥940	99.37%
	Education Classification				
	Regular	≥48,100	98.18%	≥48,250	98.92%
	Special	≥5,590	98.05%	≥5,610	98.82%
	Economic Status				
	Economically Disadvantaged	≥37,650	99.03%	≥37,790	99.04%
	Not Economically Disadvantaged	≥14,750	98.96%	≥14,770	98.98%
	LEP Status				
	Non-LEP	≥52,560	98.25%	≥52,710	98.96%
	LEP	≥1,130	94.35%	≥1,150	96.53%
	Migrant Status				
	Nonmigrant	≥53,570	98.17%	≥53,740	98.91%
	Migrant	≥120	98.35%	≥120	100.00%
	504 Status				
	Non-504	≥47,570	98.10%	≥47,720	98.89%
	504	≥6,120	98.64%	≥6,130	99.09%

Copyright © 2017 by Louisiana Department of Education.

Participation Rates by Grade and Subgroup (continued)					
Grade	**Group**	**Accountable in ELA**	**Percent Reportable in ELA**	**Accountable in Mathematics**	**Percent Reportable in Mathematics**
7	**All Students**	≥52,130	97.85%	≥52,270	98.66%
	Gender				
	Female	≥25,210	97.90%	≥25,270	98.69%
	Male	≥26,870	97.92%	≥26,950	98.71%
	Ethnicity				
	Hispanic/Latino	≥2,750	97.54%	≥2,770	98.67%
	American Indian or Alaska Native	≥370	97.59%	≥370	98.94%
	Asian	≥800	98.88%	≥800	99.00%
	Black or African American	≥23,230	97.67%	≥23,310	98.55%
	Native Hawaiian or Other Pacific	≥40	97.83%	≥40	97.83%
	White	≥23,950	98.26%	≥23,990	98.83%
	Two or More Races	≥850	98.48%	≥850	99.30%
	Education Classification				
	Regular	≥47,250	97.90%	≥47,380	98.68%
	Special	≥4,870	97.37%	≥4,880	98.47%
	Economic Status				
	Economically Disadvantaged	≥35,630	98.63%	≥35,740	98.69%
	Not Economically Disadvantaged	≥15,180	98.93%	≥15,190	98.99%
	LEP Status				
	Non-LEP	≥51,040	97.91%	≥51,160	98.70%
	LEP	≥1,080	95.02%	≥1,100	96.57%
	Migrant Status				
	Nonmigrant	≥52,000	97.85%	≥52,140	98.66%
	Migrant	≥120	98.45%	≥120	100.00%
	504 Status				
	Non-504	≥46,720	97.83%	≥46,850	98.63%
	504	≥5,410	98.04%	≥5,420	98.95%

Copyright © 2017 by Louisiana Department of Education.

Participation Rates by Grade and Subgroup (continued)					
Grade	Group	Accountable in ELA	Percent Reportable in ELA	Accountable in Mathematics	Percent Reportable in Mathematics
8	**All Students**	≥52,640	97.61%	≥52,810	98.52%
	Gender				
	Female	≥25,630	97.89%	≥25,700	98.67%
	Male	≥26,960	97.48%	≥27,060	98.44%
	Ethnicity				
	Hispanic/Latino	≥2,730	95.79%	≥2,750	97.39%
	American Indian or Alaska Native	≥370	98.93%	≥370	99.20%
	Asian	≥770	99.35%	≥770	99.74%
	Black or African American	≥23,850	97.61%	≥23,940	98.57%
	Native Hawaiian or Other Pacific	≥40	100.00%	≥40	100.00%
	White	≥24,010	97.94%	≥24,050	98.58%
	Two or More Races	≥770	98.20%	≥770	98.97%
	Education Classification				
	Regular	≥47,870	97.60%	≥48,020	98.52%
	Special	≥4,770	97.65%	≥4,790	98.54%
	Economic Status				
	Economically Disadvantaged	≥35,740	98.53%	≥35,870	98.65%
	Not Economically Disadvantaged	≥15,480	98.62%	≥15,490	98.66%
	LEP Status				
	Non-LEP	≥51,610	97.71%	≥51,750	98.58%
	LEP	≥1,030	92.64%	≥1,050	95.65%
	Migrant Status				
	Nonmigrant	≥52,500	97.61%	≥52,670	98.52%
	Migrant	≥130	95.62%	≥130	97.84%
	504 Status				
	Non-504	≥47,490	97.60%	≥47,650	98.51%
	504	≥5,150	97.65%	≥5,160	98.59%

*Eighth grade students who enrolled in Algebra I had the option of taking the Algebra EOC test instead of the LEAP mathematics Grade 8 test.

7.2 Current Administration Data

The ELA and mathematics LEAP assessments were administered to students in grades 3 through 8. Tables 7.2 and 7.3 provide a summary of the scaled scores based on the state population for the 2016 administration of the ELA and mathematics assessments, respectively.

Copyright © 2017 by Louisiana Department of Education.

Table 7.2. State-Level Scaled Score Statistics: English Language Arts

Grade	N	Mean SS	S.D. SS	Percentile				
				10th	25th	50th	75th	90th
3	≥57,160	736.21	36.56	687	712	738	762	783
4	≥53,980	742.91	31.99	701	722	744	765	785
5	≥52,580	741.05	31.32	701	720	740	763	782
6	≥52,650	742.01	28.77	706	723	741	760	780
7	≥50,960	742.53	34.98	697	721	742	766	787
8	≥51,230	747.73	33.00	706	727	747	770	790

Table 7.3.: State-Level Scaled Score Statistics: Mathematics

Grade	N	Mean SS	S.D. SS	Percentile				
				10th	25th	50th	75th	90th
3	≥57,130	745.98	33.43	703	722	746	770	789
4	≥53,990	743.46	32.45	703	718	743	766	786
5	≥52,570	738.10	31.44	699	714	735	759	779
6	≥52,670	734.18	27.63	699	714	731	753	771
7	≥50,840	732.98	25.74	703	714	730	750	768
8*	≥46,930	732.50	31.63	695	708	731	753	775

*Eighth grade students who enrolled in Algebra I had the option of taking the Algebra EOC test instead of the LEAP mathematics Grade 8 test.

Tables 7.4 and 7.5 show the percentage of students in each achievement level based on the state population for the 2016 administration of the ELA and mathematics assessments.

Table 7.4. Comparison of Percentage of Students in Each Achievement Level, ELA 2016 Census Data

Content	Grade	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
ELA	3	≥57,170	17.4	20.6	26.0	33.6	2.4
ELA	4	≥54,000	9.0	20.5	28.4	34.9	7.2
ELA	5	≥52,590	9.0	21.3	31.1	35.1	3.5
ELA	6	≥52,690	6.9	19.5	34.8	33.7	5.0
ELA	7	≥50,980	10.6	18.9	29.1	30.0	11.4
ELA	8	≥51,300	7.4	15.7	28.2	41.0	7.7

Copyright © 2017 by Louisiana Department of Education.

Table 7.5. Comparison of Percentage of Students in Each Achievement Level, Mathematics 2016 Census Data

Content	Grade	N	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
Mathematics	3	≥57,150	9.3	17.0	27.7	36.9	9.1
Mathematics	4	≥53,990	7.3	22.8	29.4	34.4	6.1
Mathematics	5	≥52,580	10.4	25.3	31.0	27.3	6.0
Mathematics	6	≥52,690	10.5	29.7	32.2	23.6	3.9
Mathematics	7	≥50,860	9.2	31.3	33.7	23.2	2.6
Mathematics	8	≥46,940	16.5	27.5	26.2	27.8	2.0

7.3 Reports

Score reports are the primary means of communicating test scores to relevant district 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 in that it states:

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

Interpretations related to the test scores are disseminated in two ways: (1) the individual score report, and (2) the *LEAP Interpretive Guide* (DRC, 2016).

In addition to providing interpretation, it is important that the information is understandable by the target audience. Standard 7.0 states:

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

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

The individual student report is the primary means for sharing student test results with parents. As such, it should be a stand-alone document from which parents can glean relevant information so they understand their child's test score. In the 2015-2016

Copyright © 2017 by Louisiana Department of Education.

administration year, student reports for each school were posted by grade and downloaded and printed from eDIRECT by districts and schools.

7.3.1 Description of Each Type of Report

In this section, descriptions for the following reports are provided: School Roster Report and Student-Level Report.

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

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

School Roster Report

A School Roster Report, which provides summary information about student performance on the ELA and mathematics tests, is available to districts and schools through eDIRECT. Total test scores and achievement levels indictors are displayed for the content area of interest. Claim and subclaim performance ratings are also reported for students. At the school level, the percentage of students at each achievement level and rating by claim and subclaim are summarized. More detail can be found in the interpretive guide on the LDOE website, https://www.louisianabelieves.com/docs/default-source/assessment/grades-3-8-ela-and-math-interpretive-guide.pdf?sfvrsn=3

Student-Level Report

The Individual Student-Level Report (ISR) is another type of report available through the eDIRECT system. The Individual Student-Level Reports may be downloaded and printed by schools to be sent home to the parents. At the top of the page, overall student performance is reported by scaled scores and achievement level. To give context to the student score, the student's district and state averages are presented to the right of the student information. In the middle of the page, claim and subclaim performance indicators are reported. Achievement level descriptors are found at the bottom of the page, as well as the percentage of students in each achievement level by school, district, and the state, allowing comparisons of the student's overall achievement level to those of his or her peers. When a student does not receive a scaled score, then his or her achievement level will be left blank. Reports for students whose scores were invalidated display a blank scaled score for a given content area.

Copyright © 2017 by Louisiana Department of Education.

7.4 Data Structures

A data file referred to as Louisiana Department of Education Student File (LDESTD) was provided to LDOE by DRC. It contains one record for every student tested; each record contains demographic information for each student as well as item responses for multiple-choice items, item scores for non-multiple-choice- items, raw scores, content and process standard raw scores, and scaled score data for each content area.

7.5 Interpreting Test Results

The *LEAP Interpretative Guide,* https://www.louisianabelieves.com/docs/default-source/assessment/grades-3-8-ela-and-math-interpretive-guide.pdf?sfvrsn=3, was written for Louisiana district and school administrators, teachers, parents, and the general public to better understand the LEAP ELA and mathematics tests. The LEAP Interpretative Guide was developed collaboratively by DRC and LDOE staff. LDOE staff had opportunities to review, provide feedback, and give final approval.

This 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 will be presented on the student-level reports. Sample student-level score reports are included. The third section discusses information that is presented on the School Roster Report and an example of the report.

7.6 Summary

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

> ***Standard 5.1*** Test users should be provided with clear explanations of the characteristics, meaning, and intended interpretation of scaled 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 the 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)

Copyright © 2017 by Louisiana Department of Education.

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

Copyright © 2017 by Louisiana Department of Education.

CHAPTER 8: PERFORMANCE LEVEL SETTING

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

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

In spring 2015, English Language Arts (ELA) and mathematics tests measuring different content and constructs than previous tests were administered to Louisiana students. The new tests were built using the Partnership for Assessment of Readiness for College and Careers (PARCC) item bank and fully aligned to the Louisiana Content Standards. The new tests were reported on the new scales and the students were classified into achievement levels based on their knowledge and ability to perform different tasks in relation to the new test content and standards to which the ELA and mathematics assessments were aligned.

In terms of the validity of the LEAP 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, and cut scores in particular.

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

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

8.2 Cut Scores

In this section, we present the cut scores for each grade and content area of LEAP. Tables 8.1 through 8.3 show the cut scores for grades 3–8 ELA and mathematics.

Copyright © 2017 by Louisiana Department of Education.

Table 8.1: English Language Arts Cut Scores

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

Table 8.2: Mathematics Cut Scores

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

8.2.1 Claim Cut Scores

As stated in Section 6.4.4.1 claim-level/subclaim-level scores, student performance on ELA and mathematics claims and subclaims was classified into one of the three performance ratings: *Weak Performance*, *Moderate Performance*, and *Strong Performance*. Detailed rules for calculating performance ratings for ELA and mathematics claims and subclaims can be found in that section.

8.3 Achievement-Level Descriptors

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

- *Advanced*: Students performing at this level have **exceeded** college- and career-readiness expectations and are well prepared for the next level of 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.

Copyright © 2017 by Louisiana Department of Education.

- *Unsatisfactory*: Students performing at this level have **not yet met** the college- and career-readiness expectations and will need extensive support to be prepared for the next level of studies in this content area.

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

Table 8.3: Achievement Level Scaled-Score Ranges

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

8.4 Summary

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

The standard settings process undertaken by PARCC addresses the following from *Standards:*

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

Copyright © 2017 by Louisiana Department of Education.

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 Louisiana Educational Assessment Program (LEAP) validation process. In this section, DRC presents evidence of construct-related validity through studies of test reliability, convergent validity, and divergent validity. All analyses in this section are based on census data.

Chapter 9 of this report demonstrates the adherence to American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) 2014 Standards 1.13, 1.21, 2.3, 2.13, 2.14, 2.15, 2.16, and 2.19. Each standard will be 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 through 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 and Algina, 1986). Reliability is a necessary but not sufficient condition of validity.

Copyright © 2017 by Louisiana Department of Education.

The *Standards* indicates:

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

In accordance with the *Standards* in developing and maintaining tests of the highest quality, DRC has calculated the reliability of each LEAP 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 9.2.1 of this chapter. The subscore reliabilities and SEMs are presented in Section 9.4.3. The SEM of the total score is discussed in Section 9.2.2.

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

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

Copyright © 2017 by Louisiana Department of Education.

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

9.2.1 Test Reliability

The reliability of raw scores by test form was evaluated using Cronbach's (1951) coefficient alpha, which is a lower-bound estimate of test reliability. The reliability coefficient is a ratio of the variance of true test scores to the variance of the total observed scores, with the values ranging from 0 to 1. The closer the value of the reliability coefficient is to 1, the more consistent the scores, where 1 refers to a perfectly consistent test. As a rule of thumb, 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

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

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

Total test reliability measures, such as Cronbach's coefficient alpha and SEM, consider the consistency (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 are reported in Table 9.1. These reliability coefficients were computed using the census data. The reliability statistics ranged from 0.85 to 0.89 for all ELA forms. For mathematics, the reliabilities ranged from 0.86 to 0.91.These results indicate acceptable reliability coefficients for LEAP tests.

Copyright © 2017 by Louisiana Department of Education.

Table 9.1 Reliability in English Language Arts and Mathematics

Content	Grade	Mode	Number of Items	Number of Score Points	SEM	Cronbach's Alpha	N-Count
ELA	3	PBT	31	66	4.30	0.87	≥57,160
ELA	4	CBT	37	74	4.32	0.87	≥450
ELA	4	PBT	37	74	4.52	0.88	≥53,520
ELA	5	CBT	33	77	4.57	0.87	≥1,690
ELA	5	PBT	33	77	4.41	0.86	≥50,890
ELA	6	CBT	33	77	4.92	0.89	≥4,810
ELA	6	PBT	33	77	4.72	0.88	≥47,840
ELA	7	CBT	32	76	5.35	0.86	≥6,310
ELA	7	PBT	32	76	5.15	0.85	≥44,650
ELA	8	CBT	37	77	4.95	0.89	≥5,450
ELA	8	PBT	37	77	4.69	0.88	≥45,780
Mathematics	3	PBT	45	62	3.75	0.90	≥57,130
Mathematics	4	CBT	44	62	3.75	0.91	≥450
Mathematics	4	PBT	44	62	3.74	0.91	≥53,530
Mathematics	5	CBT	44	62	3.64	0.90	≥1,690
Mathematics	5	PBT	44	62	3.66	0.90	≥50,880
Mathematics	6	CBT	44	66	3.48	0.91	≥4,800
Mathematics	6	PBT	44	66	3.46	0.91	≥47,860
Mathematics	7	CBT	42	65	3.73	0.90	≥6,270
Mathematics	7	PBT	42	65	3.75	0.90	≥44,570
Mathematics	8	CBT	45	66	3.35	0.86	≥4,440
Mathematics	8	PBT	45	66	3.41	0.89	≥42,490

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

9.2.2 Standard Error of Measurement (SEM)

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 68% of the time a student's score obtained from a single test administration would fall within one SEM of the student's true score and that 95% of the time the obtained score would fall within approximately two standard errors of the true score. The SEM is an index of the random variability in test scores and is defined as follows:

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

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

Copyright © 2017 by Louisiana Department of Education.

The SEM at the test form level was computed in raw score metric and is also presented in Tables 9.1 for ELA and mathematics.

9.2.3 Conditional Standard Error of Measurement (CSEM)

In contrast to SEM, CSEM expresses the degree of measurement error in scaled score units and is conditioned on the ability of the student. We report the CSEM in support of *Standard* 2.14, which states:

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

In further compliance with Standard 2.14, the CSEM of each cut score is reported in Table 9.2.

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

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

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

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

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

Note that the CSEMs vary in magnitude across the entire range of student ability estimates (i.e., scaled scores) and are smaller in the middle of the score distribution and higher at the tails. This pattern is expected when IRT methods are used. The CSEMs at the four cut scores that define the performance levels are presented in Table 9.2.

Copyright © 2017 by Louisiana Department of Education.

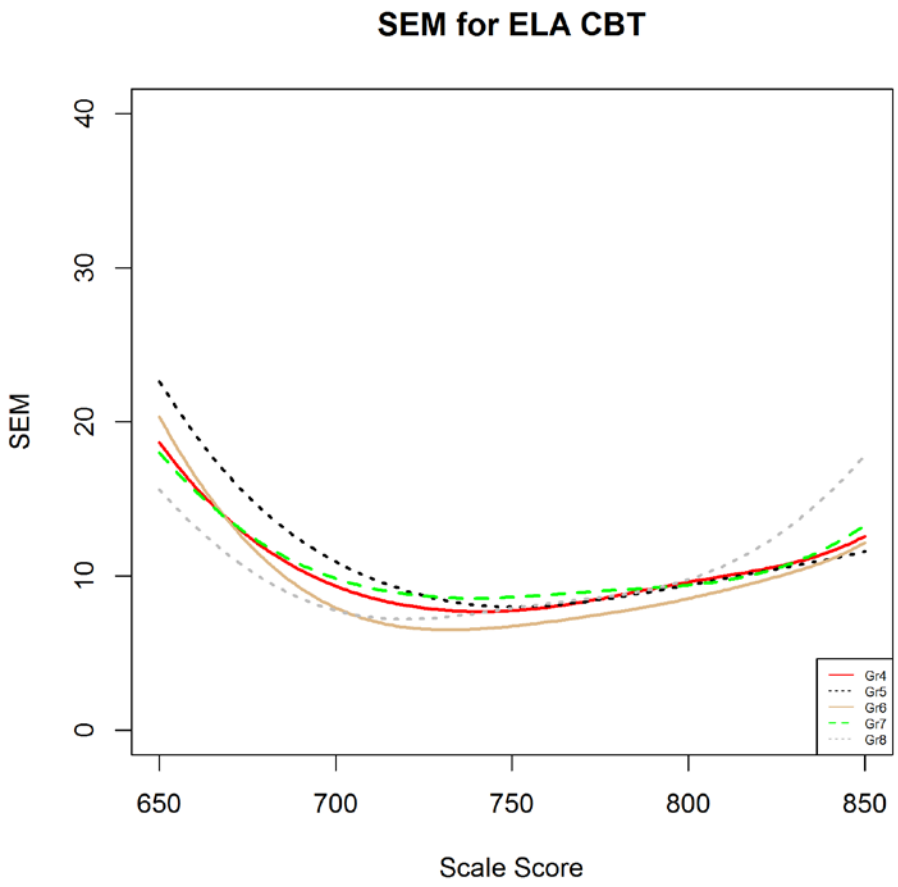
Table 9.2 Conditional Standard Error of Measurement at the Approaching Basic, Basic, Mastery, and Advanced Cut Scores

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

Figures 9.1 through 9.4 display the CSEM curves for each grade/content area by mode. The estimates of measurement error tend to be higher at the low and high ends of the scaled score range. The measurement error increases when there are few observations at a particular ability level. Generally, there are few students with extreme scores, and these score levels cannot be estimated as accurately as levels toward the middle of the ability range. Figures 9.1 through 9.4 demonstrate that the tests are designed so that measurement error is minimized in the middle of the scale range, where the majority of students are located.

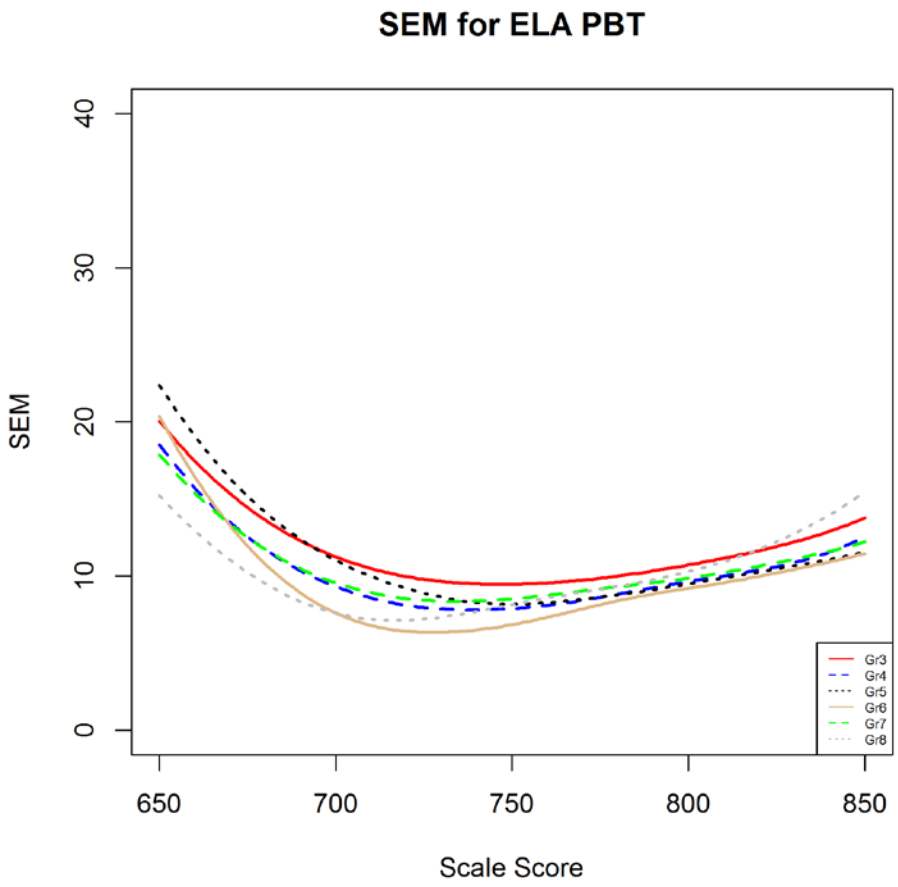
Copyright © 2017 by Louisiana Department of Education.

Figure 9.1 CSEM Curves for ELA Computer-Based Tests (CBT) Grades 4 to 8



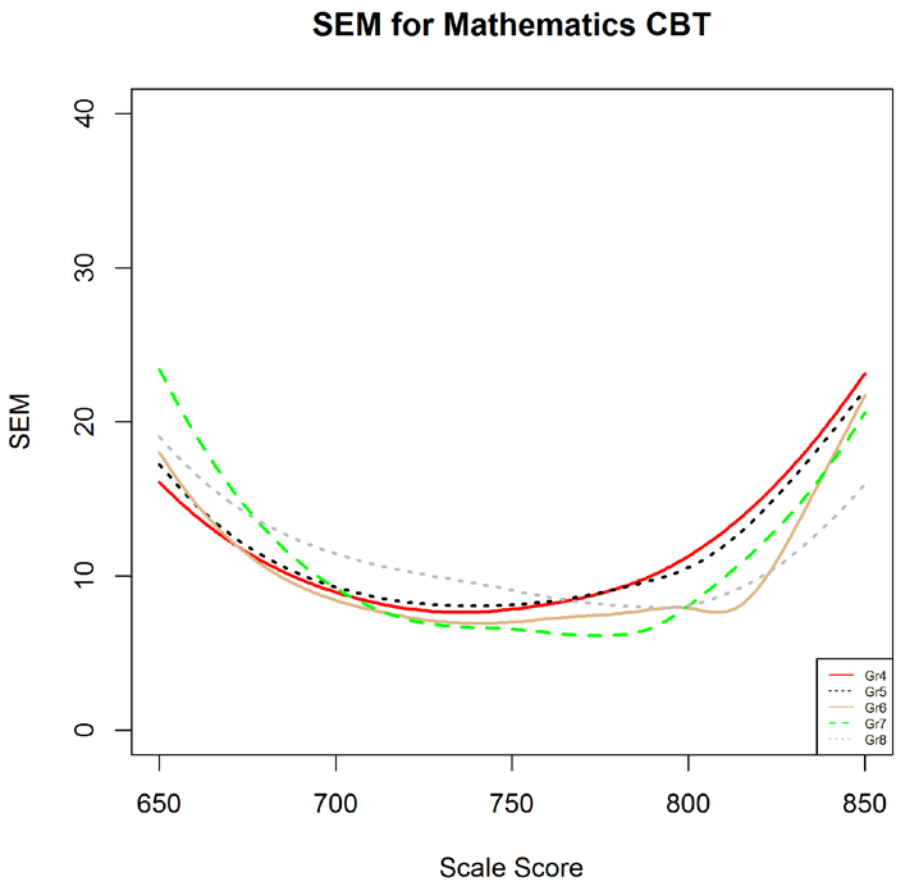
Copyright © 2017 by Louisiana Department of Education.

Figure 9.2 CSEM Curves for ELA Paper-Based Tests (PBT) Grades 3 to 8



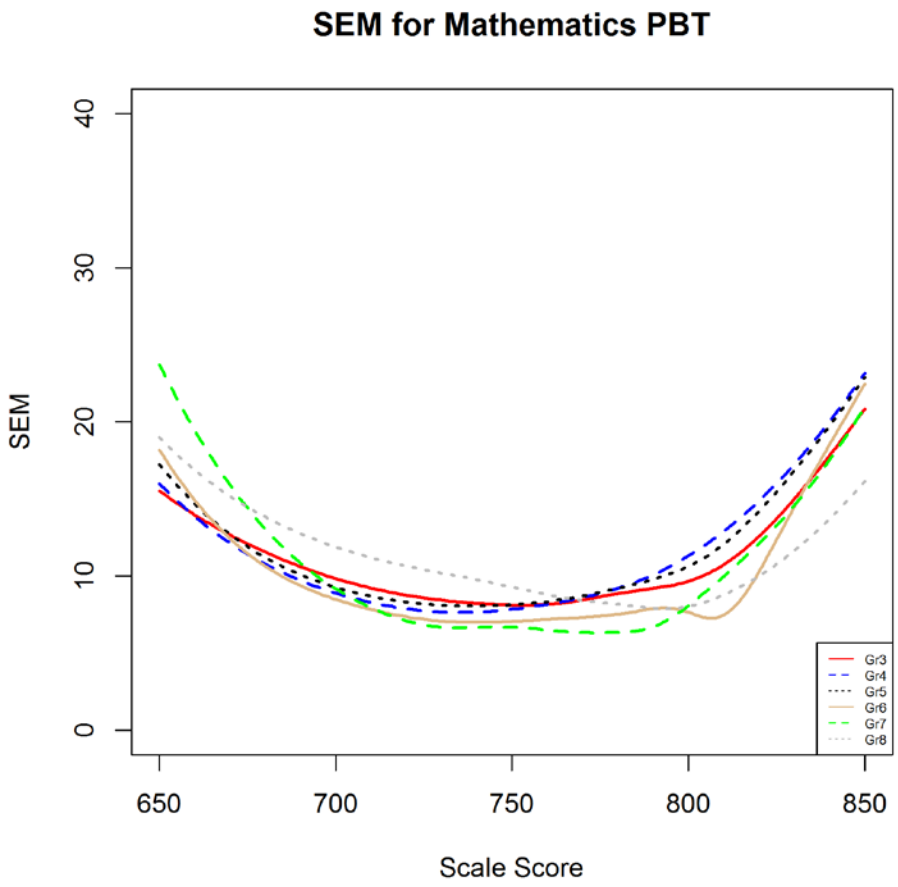
Copyright © 2017 by Louisiana Department of Education.

Figure 9.3 CSEM Curves for Mathematics Computer-Based Tests (CBT) Grades 4 to 8



Copyright © 2017 by Louisiana Department of Education.

Figure 9.4 CSEM Curves for Mathematics Paper-Based Tests (PBT) Grades 3 to 8



9.2.4 Classification Accuracy and Consistency

Decision Accuracy

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

Decision Consistency

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

Copyright © 2017 by Louisiana Department of Education.

parallel form. Therefore, a common practice is to estimate decision consistency from one administration of a test.

The Livingston-Lewis (1995) methodology was used to calculate decision accuracy statistics based on the LEAP 2016 results. The Livingston-Lewis procedure utilizes a beta-binomial model that requires two steps: (1) fitting proportion-correct true scores to a four-parameter beta distribution, and (2) using the binomial distribution to estimate classification accuracy and consistency. All calculations for decision accuracy and consistency are based on census data.

Classification consistency and classification accuracy conditioned on performance level (Table 9.3 and Table 9.4) and on cut score (Table 9.5 and Table 9.6) are presented for the 2016 LEAP in this section of the report. The magnitude of classification consistency and accuracy measures is influenced by several key features of the test design, including the number of items, the location and number of cut scores, the score distribution, and the reliability and associated SEM. As can be seen in Table 9.3, classification accuracy conditioned on achievement level ranged from 0.36 to 0.84 for ELA and 0.53 to 0.84 for mathematics. Classification consistency (Table 9.4) conditioned on achievement level ranged from 0.39 to 0.74 for ELA and 0.46 to 0.77 for mathematics. Table 9.5 shows that classification accuracy at achievement cut points ranged from 0.88 to 0.97 for ELA and 0.87 to 0.99 for mathematics. Classification consistency (Table 9.6) conditioned at achievement cut points ranged from 0.39 to 0.74 for ELA and 0.46 to 0.77 for mathematics. Classification consistency and accuracy at achievement cut points tend to be higher values than those conditioned on performance level. For some ELA tests, classification accuracy and consistency conditioned on *Advanced* level were lower than 0.50. One reason for these relatively low *Advanced* level values is few highly difficult items to distinguish *Advanced* level from other performance levels.

Copyright © 2017 by Louisiana Department of Education.

Table 9.3 Decision Accuracy Conditioned on Level of Achievement

			Decision Accuracy				
Content Area	**Grade**	**Mode**	**Unsatisfactory**	**Approaching Basic**	**Basic**	**Mastery**	**Advanced**
ELA	3	PBT	0.75	0.62	0.62	0.83	0.36
ELA	4	CBT	0.68	0.64	0.71	0.79	0.50
ELA	4	PBT	0.69	0.66	0.70	0.80	0.52
ELA	5	CBT	0.60	0.64	0.74	0.82	0.45
ELA	5	PBT	0.57	0.63	0.73	0.81	0.41
ELA	6	CBT	0.69	0.70	0.76	0.84	0.38
ELA	6	PBT	0.67	0.70	0.74	0.81	0.47
ELA	7	CBT	0.68	0.61	0.65	0.75	0.59
ELA	7	PBT	0.66	0.63	0.64	0.71	0.64
ELA	8	CBT	0.68	0.69	0.71	0.84	0.43
ELA	8	PBT	0.72	0.67	0.70	0.83	0.42
Mathematics	3	PBT	0.64	0.59	0.74	0.81	0.72
Mathematics	4	CBT	0.61	0.72	0.69	0.86	0.52
Mathematics	4	PBT	0.63	0.73	0.74	0.84	0.54
Mathematics	5	CBT	0.63	0.65	0.74	0.80	0.72
Mathematics	5	PBT	0.63	0.65	0.74	0.80	0.71
Mathematics	6	CBT	0.62	0.79	0.74	0.83	0.67
Mathematics	6	PBT	0.62	0.76	0.73	0.82	0.65
Mathematics	7	CBT	0.55	0.76	0.74	0.81	0.66
Mathematics	7	PBT	0.53	0.76	0.74	0.81	0.65
Mathematics	8	CBT	0.56	0.62	0.62	0.81	0.63
Mathematics	8	PBT	0.62	0.63	0.63	0.82	0.66

Copyright © 2017 by Louisiana Department of Education.

Table 9.4 Decision Consistency Conditioned on Level of Achievement

	Decision Consistency						
Content Area	Grade	Mode	Unsatisfactory	Approaching Basic	Basic	Mastery	Advanced
ELA	3	PBT	0.70	0.49	0.49	0.73	0.39
ELA	4	CBT	0.63	0.51	0.58	0.68	0.52
ELA	4	PBT	0.64	0.55	0.56	0.68	0.53
ELA	5	CBT	0.58	0.53	0.58	0.74	0.48
ELA	5	PBT	0.56	0.51	0.57	0.73	0.46
ELA	6	CBT	0.65	0.59	0.64	0.74	0.43
ELA	6	PBT	0.63	0.60	0.62	0.69	0.45
ELA	7	CBT	0.66	0.47	0.52	0.60	0.58
ELA	7	PBT	0.63	0.50	0.51	0.56	0.59
ELA	8	CBT	0.67	0.57	0.56	0.75	0.45
ELA	8	PBT	0.68	0.57	0.56	0.73	0.44
Mathematics	3	PBT	0.61	0.47	0.58	0.74	0.67
Mathematics	4	CBT	0.54	0.60	0.59	0.77	0.51
Mathematics	4	PBT	0.55	0.60	0.63	0.76	0.53
Mathematics	5	CBT	0.54	0.55	0.59	0.74	0.67
Mathematics	5	PBT	0.54	0.55	0.62	0.72	0.66
Mathematics	6	CBT	0.55	0.67	0.65	0.74	0.66
Mathematics	6	PBT	0.59	0.62	0.63	0.73	0.65
Mathematics	7	CBT	0.51	0.62	0.64	0.74	0.63
Mathematics	7	PBT	0.49	0.61	0.63	0.74	0.62
Mathematics	8	CBT	0.52	0.46	0.49	0.74	0.60
Mathematics	8	PBT	0.56	0.49	0.48	0.77	0.63

Perhaps the most important indices for accountability systems are those for the accuracy and consistency of classification decisions made at specific cut points. To evaluate decisions at specific cut points, the joint distribution of all the performance levels is collapsed into a dichotomized distribution around that specific cut point. As an example, the dichotomization at the cut point between the *Basic* and *Mastery* classifications was formed. The proportion of correct classifications below this particular cut point is equal to the sum of all the cells at the levels *Unsatisfactory*, *Approaching Basic*, and *Basic*, and the proportion of correct classifications above that particular cut point is equal to the sum of all the cells at the levels *Mastery* and *Advanced*. Table 9.5 shows the classification accuracy and Table 9.6 shows the consistency estimates when conditioned on LEAP cut scores. The classification accuracy statistics are at or above 0.87, while the classification consistency statistics are at or above 0.82. These results suggest that consistent and accurate performance-level classifications are being made for students in Louisiana based on the LEAP.

Copyright © 2017 by Louisiana Department of Education.

Table 9.5 Decision Accuracy at Achievement Cut Points

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

Copyright © 2017 by Louisiana Department of Education.

Table 9.6 Decision Consistency at Achievement Cut Points

			Decision Consistency			
Content Area	Grade	Mode	Unsatisfactory/ Approaching Basic	Approaching Basic/ Basic	Basic/ Mastery	Mastery/ Advanced
ELA	3	PBT	0.90	0.86	0.85	0.97
ELA	4	CBT	0.93	0.88	0.85	0.93
ELA	4	PBT	0.93	0.87	0.85	0.93
ELA	5	CBT	0.92	0.86	0.86	0.96
ELA	5	PBT	0.92	0.86	0.85	0.96
ELA	6	CBT	0.95	0.89	0.86	0.95
ELA	6	PBT	0.95	0.88	0.85	0.94
ELA	7	CBT	0.92	0.86	0.83	0.91
ELA	7	PBT	0.92	0.86	0.83	0.90
ELA	8	CBT	0.96	0.90	0.85	0.92
ELA	8	PBT	0.95	0.90	0.85	0.92
Mathematics	3	PBT	0.92	0.88	0.87	0.94
Mathematics	4	CBT	0.93	0.89	0.88	0.94
Mathematics	4	PBT	0.93	0.89	0.89	0.94
Mathematics	5	CBT	0.90	0.87	0.89	0.96
Mathematics	5	PBT	0.90	0.86	0.89	0.96
Mathematics	6	CBT	0.92	0.87	0.90	0.97
Mathematics	6	PBT	0.91	0.86	0.90	0.97
Mathematics	7	CBT	0.90	0.86	0.90	0.98
Mathematics	7	PBT	0.90	0.85	0.90	0.98
Mathematics	8	CBT	0.83	0.82	0.88	0.99
Mathematics	8	PBT	0.84	0.84	0.88	0.99

9.2.5 Convergent Validity

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

This technical report summarizes additional statistics that contribute to construct validity (Cronbach's coefficient alpha reported previously in this section and item fit reported in Chapter 6. The internal consistency coefficient (Cronbach's alpha) reported is typically measured via correlations among the test items and indicates of the degree of the same general construct (2015 PARCC Technical Report, page 128). Table 9.1 shows test reliability for ELA and mathematics. The reliability statistics ranged from 0.85 to 0.89 for

Copyright © 2017 by Louisiana Department of Education.

ELA forms and from 0.86 to 0.91 for mathematics forms indicating items on the LEAP 2016 assessments are measuring the same construct or the same content domain. In order for a group of items to be homogeneous, they must measure the same construct (construct validity) or represent the same content domain (content validity). Because IRT models were used to calibrate test items and to report student scores, item fit is also relevant to construct validity. The extent to which test items function as the IRT model prescribes is relevant to the validation of test scores. As shown in Chapter 6, no items were flagged for poor model/data fit.

9.3 Principal Components Analysis

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

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

In this section, we examine the internal structure by evaluating the unidimensionality assumption through Principal Components Analysis (PCA). This analysis seeks evidence that there exists a single primary factor, the first principal component, which accounts for much of the relationship between items. The presence of a single or dominant factor suggests that a test is sufficiently unidimensional (i.e., measures one underlying construct).

A principal components factor analysis was conducted on each grade/content/mode area of LEAP. A large first principal component is evident in each analysis. It is common to have additional eigenvalues greater than 1.0, which may suggest the presence of other factors.

For all grade/content/mode areas of LEAP, the ratio of the variance accounted for by the first factor to the second is sufficiently large, indicating that the unidimensionality assumption holds. All of the LEAP subject-area tests exhibit first principal components accounting for more than 20% of the test variance for ELA (see Table 9.7), and for more than 15% of the test variance for mathematics (see Table 9.8). To further investigate the unidimensionality of the ELA and mathematics, the ratio of the first eigenvalue to the second eigenvalue was explored (see Tables 9.7 and 9.8). These ratios show that the first eigenvalue is at least four times as large as the second eigenvalue for all of the grade/content/mode areas. This substantial difference in magnitude indicates that one factor appears to be dominant and that the ELA and mathematics tests are essentially unidimensional.

Copyright © 2017 by Louisiana Department of Education.

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

Table 9.7 Principal Component Analysis for English Language Arts

Grade	Mode	Components	Eigenvalue	Percent of Variance Explained	Cumulative Percent of Variance Explained
3	PBT	First Component	6.71	23.12	23.12
3	PBT	Second Component	1.31	4.50	27.62
3	PBT	Ratio (First/Second)	5.14		
4	CBT	First Component	7.22	20.07	20.07
4	CBT	Second Component	1.67	4.63	24.69
4	CBT	Ratio (First/Second)	4.34		
4	PBT	First Component	7.48	20.77	20.77
4	PBT	Second Component	1.32	3.67	24.44
4	PBT	Ratio (First/Second)	5.66		
5	CBT	First Component	7.13	23.01	23.01
5	CBT	Second Component	1.36	4.39	27.40
5	CBT	Ratio (First/Second)	5.24		
5	PBT	First Component	6.63	21.37	21.37
5	PBT	Second Component	1.43	4.63	26.00
5	PBT	Ratio (First/Second)	4.62		
6	CBT	First Component	7.79	25.12	25.12
6	CBT	Second Component	1.25	4.02	29.13
6	CBT	Ratio (First/Second)	6.25		
6	PBT	First Component	7.37	23.78	23.78
6	PBT	Second Component	1.40	4.51	28.29
6	PBT	Ratio (First/Second)	5.27		
7	CBT	First Component	6.40	21.32	21.32
7	CBT	Second Component	1.31	4.38	25.70
7	CBT	Ratio (First/Second)	4.87		
7	PBT	First Component	6.13	20.44	20.44
7	PBT	Second Component	1.37	4.57	25.01
7	PBT	Ratio (First/Second)	4.48		
8	CBT	First Component	8.18	22.71	22.71
8	CBT	Second Component	1.45	4.02	26.73
8	CBT	Ratio (First/Second)	5.65		
8	PBT	First Component	8.09	22.47	22.47
8	PBT	Second Component	1.54	4.29	26.75
8	PBT	Ratio (First/Second)	5.24		

Copyright © 2017 by Louisiana Department of Education.

Table 9.8 Principal Component Analysis for Mathematics

Grade	Mode	Components	Eigenvalue	Percent of Variance Explained	Cumulative Percent of Variance Explained
3	PBT	First Component	9.83	21.85	21.85
3	PBT	Second Component	1.95	4.34	26.19
3	PBT	Ratio (First/Second)	5.04		
4	CBT	First Component	9.90	22.51	22.51
4	CBT	Second Component	1.76	3.99	26.50
4	CBT	Ratio (First/Second)	5.64		
4	PBT	First Component	10.33	23.48	23.48
4	PBT	Second Component	1.67	3.79	27.27
4	PBT	Ratio (First/Second)	6.20		
5	CBT	First Component	9.50	21.59	21.59
5	CBT	Second Component	1.68	3.82	25.40
5	CBT	Ratio (First/Second)	5.66		
5	PBT	First Component	9.34	21.23	21.23
5	PBT	Second Component	1.67	3.80	25.03
5	PBT	Ratio (First/Second)	5.59		
6	CBT	First Component	10.29	23.39	23.39
6	CBT	Second Component	1.80	4.09	27.48
6	CBT	Ratio (First/Second)	5.71		
6	PBT	First Component	9.70	22.04	22.04
6	PBT	Second Component	1.82	4.14	26.18
6	PBT	Ratio (First/Second)	5.32		
7	CBT	First Component	9.33	22.22	22.22
7	CBT	Second Component	1.75	4.18	26.40
7	CBT	Ratio (First/Second)	5.32		
7	PBT	First Component	9.00	21.42	21.42
7	PBT	Second Component	1.75	4.16	25.58
7	PBT	Ratio (First/Second)	5.15		
8	CBT	First Component	6.97	15.48	15.48
8	CBT	Second Component	1.63	3.63	19.11
8	CBT	Ratio (First/Second)	4.26		
8	PBT	First Component	8.08	17.96	17.96
8	PBT	Second Component	1.74	3.87	21.84
8	PBT	Ratio (First/Second)	4.64		

9.4 Analyses by Claims and Subclaims

Three sets of analyses were conducted at the claim and subclaim level for ELA and mathematics in another attempt to assess the construct validity of LEAP. First, correlation coefficients that measure the relationship between the claim scores and subclaim scores

Copyright © 2017 by Louisiana Department of Education.

were computed. Second, the reliability of each claim and subclaim was computed. Finally, the SEM was computed for each reportable claim and subclaim.

9.4.1 Correlations among Claims and Subclaims

In this section, we report the strength of the interrelationships among the claims or subclaims by computing correlation between them. Tables 9.9 through 9.11 report the uncorrected Pearson product-moment (PPM) correlation coefficients, as well as the PPM corrected for attenuation (CAPPM), in addition to the reliability coefficients described above. The PPM among the claim or subclaim subscores is presented below the diagonal portion of the matrix, the CAPPM is presented above the diagonal portion of the matrix, and the reliability coefficients used are shown in Tables 9.13 through 9.15.

The uncorrected PPM in Tables 9.10 through 9.12 should be interpreted in the context of the reliability coefficient. In general, we expect to see lower PPM coefficients between variables that are less reliable. In most cases, the PPM coefficients show that performance on one claim or subclaim is moderately to strongly related to performance on another claim or subclaim within the same grade and content area. The value of the correlation coefficients will be affected by the limited number of items measuring each claim or subclaim. So, caution should be used when comparing the PPM coefficients measuring the relationships between claims or subclaims to those measuring the relationships between content areas (Table 9.15). We expect to see a more modest relationship (smaller correlation coefficients) reported between the claims or subclaims as a consequence of the lower number of items measuring each of the reporting categories. The PPM between two claim or subclaim subscores may be artificially low because of measurement error.

Standard 1.21 states:

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

We can correct for the attenuation of the PPM statistically using Spearman's formula:

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

where r_{xy} is the PPM between two claims or GLE Strands, r_{xx} is the reliability of one of those claims or GLE Strands, and r_{yy} is the reliability for the other Claim or GLE Strand.

ELA claim shows moderate relationships between reading and writing claims across all grades, indicating these two claims measure some different traits. Across all tables, the

Copyright © 2017 by Louisiana Department of Education.

CAPPM indicate moderate or strong relationships between subclaims. The CAPPM for reading vocabulary, and written expression and knowledge and use of language shows were moderate. In some cases, the CAPPM is greater than 1.0. “Disattenuated values greater than 1.00 indicate that measurement error is not randomly distributed” (Schumacker, 1996). The moderate or strong relationships suggested by the CAPPM in Tables 9.10 through 9.12 are further evidence of the validity of the test construct. Since the overall content area is comprised of the claims or subclaim subscores and the content area is expected to measure a single dimension, we would expect that these subscores are moderately or highly related.

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

Grade	Mode	No.	Claim	N Items	1	2
3	PBT	1	Reading	27		0.85
	PBT	2	Writing	4	0.70	
4	CBT	1	Reading	33		0.78
	CBT	2	Writing	4	0.65	
	PBT	1	Reading	33		0.82
	PBT	2	Writing	4	0.69	
5	CBT	1	Reading	29		0.87
	CBT	2	Writing	4	0.72	
	PBT	1	Reading	29		0.82
	PBT	2	Writing	4	0.67	
6	CBT	1	Reading	29		0.90
	CBT	2	Writing	4	0.78	
	PBT	1	Reading	29		0.85
	PBT	2	Writing	4	0.73	
7	CBT	1	Reading	28		0.90
	CBT	2	Writing	4	0.76	
	PBT	1	Reading	28		0.89
	PBT	2	Writing	4	0.74	
8	CBT	1	Reading	33		0.86
	CBT	2	Writing	4	0.74	
	PBT	1	Reading	33		0.80
	PBT	2	Writing	4	0.70	

Copyright © 2017 by Louisiana Department of Education.

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

Subclaim Uncorrected and Corrected Correlation Coefficients: English Language Arts									
Grade	**Mode**	**No.**	**Subclaim**	**N Items**	**1**	**2**	**3**	**4**	**5**
3	PBT	1	Reading Literary Text	13		1.04	0.98	1.02	0.89
	PBT	2	Reading Information Text	9	0.67		1.00	1.07	0.90
	PBT	3	Reading Vocabulary	5	0.62	0.59		0.81	0.74
	PBT	4	Written Expression	2	0.63	0.63	0.46		1.26
	PBT	5	Knowledge & Use of Language	2	0.59	0.57	0.45	0.76	
4	CBT	1	Reading Literary Text	13		1.01	0.93	0.89	0.87
	CBT	2	Reading Information Text	15	0.71		1.07	0.97	0.91
	CBT	3	Reading Vocabulary	5	0.45	0.55		0.77	0.77
	CBT	4	Written Expression	2	0.55	0.62	0.35		1.39
	CBT	5	Knowledge & Use of Language	2	0.56	0.61	0.36	0.82	
	PBT	1	Reading Literary Text	13		1.00	0.95	0.88	0.84
	PBT	2	Reading Information Text	15	0.70		1.01	0.99	0.92
	PBT	3	Reading Vocabulary	5	0.47	0.52		0.83	0.78
	PBT	4	Written Expression	2	0.59	0.68	0.40		1.30
	PBT	5	Knowledge & Use of Language	2	0.58	0.65	0.39	0.87	
5	CBT	1	Reading Literary Text	13		1.05	0.94	0.98	0.86
	CBT	2	Reading Information Text	11	0.73		1.02	1.06	0.93
	CBT	3	Reading Vocabulary	5	0.51	0.56		0.83	0.78
	CBT	4	Written Expression	2	0.64	0.70	0.43		1.24
	CBT	5	Knowledge & Use of Language	2	0.59	0.64	0.42	0.80	
	PBT	1	Reading Literary Text	13		1.02	0.99	1.00	0.87
	PBT	2	Reading Information Text	11	0.70		0.99	1.04	0.92
	PBT	3	Reading Vocabulary	5	0.53	0.55		0.82	0.78
	PBT	4	Written Expression	2	0.59	0.63	0.39		1.44
	PBT	5	Knowledge & Use of Language	2	0.53	0.58	0.39	0.79	
6	CBT	1	Reading Literary Text	13		1.07	1.05	1.02	0.98
	CBT	2	Reading Information Text	11	0.72		1.05	1.04	1.02
	CBT	3	Reading Vocabulary	5	0.65	0.65		0.88	0.88
	CBT	4	Written Expression	2	0.72	0.72	0.57		1.27
	CBT	5	Knowledge & Use of Language	2	0.69	0.71	0.57	0.93	
	PBT	1	Reading Literary Text	13		1.03	1.02	0.98	0.94
	PBT	2	Reading Information Text	11	0.68		1.04	1.02	1.01
	PBT	3	Reading Vocabulary	5	0.64	0.64		0.83	0.83
	PBT	4	Written Expression	2	0.67	0.68	0.52		1.38
	PBT	5	Knowledge & Use of Language	2	0.64	0.67	0.52	0.95	

Copyright © 2017 by Louisiana Department of Education.

Subclaim Uncorrected and Corrected Correlation Coefficients: English Language Arts (continued)									
Grade	Mode	No.	Subclaim	N Items	1	2	3	4	5
7	CBT	1	Reading Literary Text	12		1.11	1.01	0.99	0.95
	CBT	2	Reading Information Text	10	0.67		1.06	1.07	1.02
	CBT	3	Reading Vocabulary	6	0.53	0.51		0.82	0.82
	CBT	4	Written Expression	2	0.70	0.69	0.47		1.19
	CBT	5	Knowledge & Use of Language	2	0.67	0.66	0.47	0.90	
	PBT	1	Reading Literary Text	12		1.08	1.01	1.00	0.95
	PBT	2	Reading Information Text	10	0.63		1.07	1.08	1.03
	PBT	3	Reading Vocabulary	6	0.53	0.51		0.84	0.83
	PBT	4	Written Expression	2	0.68	0.66	0.46		1.27
	PBT	5	Knowledge & Use of Language	2	0.65	0.64	0.46	0.92	
8	CBT	1	Reading Literary Text	12		1.01	1.01	0.86	0.87
	CBT	2	Reading Information Text	15	0.70		1.04	1.02	1.00
	CBT	3	Reading Vocabulary	6	0.58	0.61		0.85	0.87
	CBT	4	Written Expression	2	0.61	0.73	0.51		1.25
	CBT	5	Knowledge & Use of Language	2	0.61	0.71	0.52	0.92	
	PBT	1	Reading Literary Text	12		0.99	0.99	0.85	0.86
	PBT	2	Reading Information Text	15	0.69		1.01	0.98	0.96
	PBT	3	Reading Vocabulary	6	0.60	0.61		0.80	0.80
	PBT	4	Written Expression	2	0.59	0.67	0.48		1.38
	PBT	5	Knowledge & Use of Language	2	0.60	0.67	0.49	0.95	

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

Grade	Mode	No.	Subclaim	N Items	1	2	3	4
3	PBT	1	Major Content	29		0.99	1.04	1.02
	PBT	2	Additional & Supporting Con	10	0.75		1.00	1.01
	PBT	3	Expressing Mathematical Rea	3	0.69	0.58		1.14
	PBT	4	Modeling & Application	3	0.75	0.65	0.64	
4	CBT	1	Major Content	28		0.95	1.00	0.99
	CBT	2	Additional & Supporting Con	10	0.72		0.94	0.99
	CBT	3	Expressing Mathematical Rea	3	0.73	0.61		1.08
	CBT	4	Modeling & Application	3	0.74	0.65	0.69	
	PBT	1	Major Content	28		0.95	0.98	1.00
	PBT	2	Additional & Supporting Con	10	0.72		0.94	0.98
	PBT	3	Expressing Mathematical Rea	3	0.73	0.61		1.09
	PBT	4	Modeling & Application	3	0.75	0.64	0.70	

Copyright © 2017 by Louisiana Department of Education.

Subclaim Uncorrected and Corrected Correlation Coefficients: Mathematics (continued)								
Grade	**Mode**	**No.**	**Subclaim**	**N Items**	**1**	**2**	**3**	**4**
5	CBT	1	Major Content	28		0.99	0.98	1.00
	CBT	2	Additional & Supporting Con	10	0.71		1.00	1.05
	CBT	3	Expressing Mathematical Rea	3	0.71	0.62		1.03
	CBT	4	Modeling & Application	3	0.72	0.65	0.64	
	PBT	1	Major Content	28		0.97	0.99	0.97
	PBT	2	Additional & Supporting Con	10	0.69		0.98	1.02
	PBT	3	Expressing Mathematical Rea	3	0.70	0.57		1.02
	PBT	4	Modeling & Application	3	0.70	0.62	0.61	
6	CBT	1	Major Content	28		0.98	1.00	1.00
	CBT	2	Additional & Supporting Con	9	0.74		0.99	1.00
	CBT	3	Expressing Mathematical Rea	4	0.75	0.65		1.07
	CBT	4	Modeling & Application	3	0.72	0.62	0.67	
	PBT	1	Major Content	28		0.99	1.03	0.97
	PBT	2	Additional & Supporting Con	9	0.72		1.02	0.98
	PBT	3	Expressing Mathematical Rea	4	0.76	0.64		1.06
	PBT	4	Modeling & Application	3	0.69	0.59	0.65	
7	CBT	1	Major Content	28		1.01	0.98	1.38
	CBT	2	Additional & Supporting Con	7	0.72		0.98	1.38
	CBT	3	Expressing Mathematical Rea	4	0.75	0.63		1.44
	CBT	4	Modeling & Application	3	0.73	0.62	0.69	
	PBT	1	Major Content	28		1.00	0.99	1.45
	PBT	2	Additional & Supporting Con	7	0.69		1.00	1.45
	PBT	3	Expressing Mathematical Rea	4	0.75	0.62		1.53
	PBT	4	Modeling & Application	3	0.73	0.61	0.70	
8	CBT	1	Major Content	28		0.97	0.97	0.88
	CBT	2	Additional & Supporting Con	10	0.66		0.92	0.85
	CBT	3	Expressing Mathematical Rea	4	0.63	0.52		1.08
	CBT	4	Modeling & Application	3	0.48	0.40	0.49	
	PBT	1	Major Content	28		0.98	0.99	0.86
	PBT	2	Additional & Supporting Con	10	0.70		0.96	0.81
	PBT	3	Expressing Mathematical Rea	4	0.68	0.58		1.02
	PBT	4	Modeling & Application	3	0.57	0.48	0.57	

9.4.2 Reliability of Claims or Subclaims

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

Copyright © 2017 by Louisiana Department of Education.

a small number of items compared to a claim or subclaim assessed by a larger number of items.

9.4.3 Standard Error of Measurement of Claims or Subclaims

In this chapter, we also report the SEM associated with each of the claims or subclaims in Tables 9.12 through 9.14 for ELA and mathematics. These SEMs are reported in the raw score metric.

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

Grade	Mode	Claim	Number of Items	Number of Score Points	Mean Raw Score	Raw Std. Dev.	SEM	Cronbach's Alpha
3	PBT	Reading	27	42	18.97	8.16	3.35	0.83
	PBT	Writing	4	24	6.01	4.58	2.02	0.81
4	CBT	Reading	33	47	20.19	8.18	3.34	0.83
	CBT	Writing	4	27	9.42	4.86	2.07	0.82
	PBT	Reading	33	47	20.06	8.47	3.44	0.84
	PBT	Writing	4	27	10.45	5.39	2.03	0.86
5	CBT	Reading	29	47	18.37	8.45	3.42	0.84
	CBT	Writing	4	30	6.88	5.42	2.22	0.83
	PBT	Reading	29	47	19.28	8.27	3.39	0.83
	PBT	Writing	4	30	6.00	4.77	2.18	0.79
6	CBT	Reading	29	47	22.51	9.05	3.59	0.84
	CBT	Writing	4	30	10.90	6.48	2.09	0.90
	PBT	Reading	29	47	22.59	8.69	3.52	0.84
	PBT	Writing	4	30	10.16	5.80	1.97	0.88
7	CBT	Reading	28	46	20.75	8.29	3.83	0.79
	CBT	Writing	4	30	11.82	6.70	2.14	0.90
	PBT	Reading	28	46	20.79	7.98	3.76	0.78
	PBT	Writing	4	30	11.42	6.11	2.05	0.89
8	CBT	Reading	33	47	25.06	8.56	3.41	0.84
	CBT	Writing	4	30	15.37	7.17	2.36	0.89
	PBT	Reading	33	47	25.32	8.61	3.38	0.85
	PBT	Writing	4	30	15.37	6.34	2.14	0.89

Copyright © 2017 by Louisiana Department of Education.

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

Mean, Standard Deviation, and SEM: English Language Arts								
Grade	Mode	Subclaim	Number of Items	Number of Score Points	Mean Raw Score	Raw Std. Dev.	SEM	Cronbach's Alpha
3	PBT	Reading Literary Text	13	19	7.31	3.75	2.10	0.69
	PBT	Reading Information Text	9	15	6.61	3.36	2.10	0.61
	PBT	Reading Vocabulary	5	8	5.05	2.22	1.45	0.58
	PBT	Written Expression	2	18	4.32	3.52	2.34	0.56
	PBT	Knowledge & Use of Language	2	6	1.70	1.30	0.76	0.65
4	CBT	Reading Literary Text	13	17	8.41	3.67	2.14	0.66
	CBT	Reading Information Text	15	22	8.71	3.95	2.05	0.73
	CBT	Reading Vocabulary	5	8	3.07	1.82	1.46	0.36
	CBT	Written Expression	2	21	7.07	3.74	2.47	0.56
	CBT	Knowledge & Use of Language	2	6	2.36	1.30	0.80	0.62
	PBT	Reading Literary Text	13	17	8.37	3.84	2.15	0.68
	PBT	Reading Information Text	15	22	8.60	4.07	2.14	0.72
	PBT	Reading Vocabulary	5	8	3.09	1.89	1.51	0.36
	PBT	Written Expression	2	21	7.78	4.17	2.47	0.65
	PBT	Knowledge & Use of Language	2	6	2.67	1.36	0.75	0.69
5	CBT	Reading Literary Text	13	20	6.55	3.90	2.18	0.69
	CBT	Reading Information Text	11	19	7.47	3.83	2.10	0.70
	CBT	Reading Vocabulary	5	8	4.34	1.94	1.45	0.44
	CBT	Written Expression	2	24	4.77	4.16	2.56	0.62
	CBT	Knowledge & Use of Language	2	6	2.11	1.48	0.85	0.67
	PBT	Reading Literary Text	13	20	7.04	3.83	2.23	0.66
	PBT	Reading Information Text	11	19	7.74	3.77	2.03	0.71
	PBT	Reading Vocabulary	5	8	4.51	1.91	1.43	0.44
	PBT	Written Expression	2	24	4.09	3.67	2.53	0.53
	PBT	Knowledge & Use of Language	2	6	1.91	1.31	0.86	0.57
6	CBT	Reading Literary Text	13	19	8.39	3.88	2.19	0.68
	CBT	Reading Information Text	11	20	8.90	4.18	2.42	0.66
	CBT	Reading Vocabulary	5	8	5.22	2.08	1.35	0.58
	CBT	Written Expression	2	24	8.17	4.99	2.57	0.73
	CBT	Knowledge & Use of Language	2	6	2.73	1.57	0.81	0.73
	PBT	Reading Literary Text	13	19	8.14	3.74	2.12	0.68
	PBT	Reading Information Text	11	20	9.45	3.97	2.38	0.64
	PBT	Reading Vocabulary	5	8	5.00	2.12	1.37	0.58
	PBT	Written Expression	2	24	7.56	4.44	2.49	0.68
	PBT	Knowledge & Use of Language	2	6	2.60	1.42	0.79	0.69

Copyright © 2017 by Louisiana Department of Education.

Mean, Standard Deviation, and SEM: English Language Arts (continued)								
Grade	**Mode**	**Subclaim**	**Number of Items**	**Number of Score Points**	**Mean Raw Score**	**Raw Std. Dev.**	**SEM**	**Cronbach's Alpha**
7	CBT	Reading Literary Text	12	18	8.07	3.86	2.24	0.66
	CBT	Reading Information Text	10	18	6.84	3.56	2.40	0.54
	CBT	Reading Vocabulary	6	10	5.84	2.28	1.74	0.42
	CBT	Written Expression	2	24	8.75	5.18	2.56	0.76
	CBT	Knowledge & Use of Language	2	6	3.06	1.64	0.80	0.76
	PBT	Reading Literary Text	12	18	8.17	3.71	2.21	0.64
	PBT	Reading Information Text	10	18	6.79	3.42	2.34	0.53
	PBT	Reading Vocabulary	6	10	5.84	2.28	1.73	0.43
	PBT	Written Expression	2	24	8.41	4.68	2.51	0.71
	PBT	Knowledge & Use of Language	2	6	3.00	1.53	0.80	0.72
8	CBT	Reading Literary Text	12	16	8.86	3.14	1.79	0.68
	CBT	Reading Information Text	15	23	11.48	4.70	2.60	0.69
	CBT	Reading Vocabulary	6	8	4.72	1.85	1.32	0.49
	CBT	Written Expression	2	24	11.85	5.63	2.90	0.73
	CBT	Knowledge & Use of Language	2	6	3.52	1.65	0.85	0.73
	PBT	Reading Literary Text	12	16	9.32	3.33	1.82	0.70
	PBT	Reading Information Text	15	23	11.10	4.54	2.54	0.69
	PBT	Reading Vocabulary	6	8	4.90	1.89	1.30	0.53
	PBT	Written Expression	2	24	11.61	4.87	2.73	0.69
	PBT	Knowledge & Use of Language	2	6	3.76	1.53	0.84	0.70

Copyright © 2017 by Louisiana Department of Education.

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

Mean, Standard Deviation, and SEM: Mathematics								
Grade	Mode	Subclaim	Number of Items	Number of Score Points	Mean Raw Score	Raw Std. Dev.	SEM	Cronbach's Alpha
3	PBT	Major Content	29	30	17.76	5.80	2.15	0.86
	PBT	Additional & Supporting Content	10	10	6.48	2.23	1.30	0.66
	PBT	Expressing Mathematical Reasoning	3	10	2.88	2.05	1.43	0.51
	PBT	Modeling & Application	3	12	3.41	3.33	2.06	0.62
4	CBT	Major Content	28	30	18.46	6.04	2.30	0.86
	CBT	Additional & Supporting Content	10	10	6.44	2.26	1.30	0.67
	CBT	Expressing Mathematical Reasoning	3	10	3.02	2.51	1.53	0.63
	CBT	Modeling & Application	3	12	4.64	2.98	1.78	0.64
	PBT	Major Content	28	30	18.52	6.28	2.29	0.87
	PBT	Additional & Supporting Content	10	10	5.80	2.36	1.37	0.67
	PBT	Expressing Mathematical Reasoning	3	10	3.03	2.54	1.53	0.64
	PBT	Modeling & Application	3	12	4.45	2.95	1.74	0.65
5	CBT	Major Content	28	30	17.17	6.13	2.40	0.85
	CBT	Additional & Supporting Content	10	10	4.98	1.97	1.22	0.62
	CBT	Expressing Mathematical Reasoning	3	10	2.18	2.49	1.52	0.63
	CBT	Modeling & Application	3	12	3.30	2.55	1.59	0.61
	PBT	Major Content	28	30	16.93	6.31	2.39	0.86
	PBT	Additional & Supporting Content	10	10	5.01	1.98	1.27	0.59
	PBT	Expressing Mathematical Reasoning	3	10	2.32	2.42	1.57	0.58
	PBT	Modeling & Application	3	12	3.35	2.52	1.57	0.62
6	CBT	Major Content	28	30	16.52	6.12	2.28	0.86
	CBT	Additional & Supporting Content	9	10	3.90	2.24	1.31	0.66
	CBT	Expressing Mathematical Reasoning	4	14	3.63	2.89	1.68	0.66
	CBT	Modeling & Application	3	12	2.14	1.99	1.27	0.59
	PBT	Major Content	28	30	15.77	5.95	2.31	0.85
	PBT	Additional & Supporting Con	9	10	3.63	2.11	1.30	0.62
	PBT	Expressing Mathematical Rea	4	14	3.66	2.79	1.68	0.64
	PBT	Modeling & Application	3	12	1.80	1.90	1.22	0.59

Copyright © 2017 by Louisiana Department of Education.

Mean, Standard Deviation, and SEM: Mathematics								
Grade	Mode	Subclaim	Number of Items	Number of Score Points	Mean Raw Score	Raw Std. Dev.	SEM	Cronbach's Alpha
7	CBT	Major Content	28	30	12.77	6.03	2.38	0.84
	CBT	Additional & Supporting Content	7	9	3.80	2.14	1.33	0.61
	CBT	Expressing Mathematical Reasoning	4	14	3.80	2.63	1.47	0.69
	CBT	Modeling & Application	3	12	3.31	2.44	1.99	0.33
	PBT	Major Content	28	30	11.85	5.91	2.36	0.84
	PBT	Additional & Supporting Content	7	9	3.62	2.04	1.34	0.57
	PBT	Expressing Mathematical Reasoning	4	14	3.61	2.62	1.48	0.68
	PBT	Modeling & Application	3	12	3.10	2.46	2.05	0.31
8	CBT	Major Content	28	30	14.04	5.04	2.37	0.78
	CBT	Additional & Supporting Content	10	10	4.57	2.11	1.36	0.59
	CBT	Expressing Mathematical Reasoning	4	14	2.10	2.31	1.55	0.55
	CBT	Modeling & Application	3	12	0.65	1.29	1.02	0.37
	PBT	Major Content	28	30	13.31	5.33	2.35	0.81
	PBT	Additional & Supporting Content	10	10	4.36	2.22	1.34	0.64
	PBT	Expressing Mathematical Reasoning	4	14	2.28	2.47	1.60	0.58
	PBT	Modeling & Application	3	12	0.83	1.67	1.12	0.55

9.5 Divergent (Discriminant) Validity

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

To assess the divergent validity of the LEAP tests, correlations were computed between the ELA and mathematics scaled scores for students who took more than one LEAP subject area test in 2016. These correlations are based on the census data, and the results are shown in Table 9.15. The correlation coefficients ranged from 0.68 (between ELA and mathematics in grade 8) to 0.73 (between ELA and mathematics in grade 3). The correlation coefficients suggest that individual student scores for ELA and mathematics are moderately related, indicating that these two tests measure a similar knowledge base or general underlying ability, but still measure some different traits as planned.

Copyright © 2017 by Louisiana Department of Education.

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

Grade	ELA/Mathematics
3	0.73
4	0.72
5	0.69
6	0.72
7	0.71
8	0.68

9.6 Summary

In summary, the overall purpose of establishing construct validity is to ensure that the meaning of test scores is supported. Evidence of validity is necessary to justify the use of the LEAP. This evidence has been generated and provided through best practices of the testing industry and in particular as they relate 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. (420
>
> ***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 2.19*** 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)
>
> ***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

Copyright © 2017 by Louisiana Department of Education.

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)

Copyright © 2017 by Louisiana Department of Education.

CHAPTER 10: FAIRNESS

As noted in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], and 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. Also discussed is test performance among varying subgroups assessed by LEAP. It should be noted that differences in test performance among subgroups does not mean that a test is unfair—it simply means that groups perform differently on the test. Even when a test is carefully and properly constructed, differences may exist among subgroups as a result of differences in curriculum or learning by students in the subgroup.

This chapter is particularly relevant to AERA, APA, and NCME Standards 3.1 through 3.6. These standards are from Chapter 3 of the *Standards*, which is titled "Fairness in Testing." Each of these standards, and how each is addressed, is presented in this chapter.

Standard 3.6 states:

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

There is no particular research on LEAP showing that the test scores of examinee subgroups differ in meaning; however, this is an ongoing concern in any large-scale testing program. To lessen the possibility of differences in test score meaning, DRC follows several steps in item development and selections, as is explained in Section 10.1 of this chapter. In addition, LDOE content and assessment experts conduct content and bias reviews on items during the selection process as explained in Chapter 3. These practices adhere to Standard 3.3, which states, "Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test" (64).

PARCC conducted differential item functioning (DIF) studies of their items prior to PARCC operational administrations. Typically items are evaluated for possible DIF in the field test phase of the test development, and items flagged for DIF are further examined to determine possible bias. During the ELA and mathematics test development, DRC content experts tried to avoid including PARCC operational items flagged for DIF. Section 10.2 of this chapter explains the steps taken to evaluate LEAP items through the use of DIF in order to adhere to this standard.

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

Copyright © 2017 by Louisiana Department of Education.

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

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

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

This section explains the steps taken by DRC to minimize words, phrases, and content that may be regarded as offensive by members of particular demographic subgroups. Section 3.2.5 of Chapter 3 discusses the Content and Bias Review conducted for LEAP. This review is also critical in fulfilling Standards 3.1 and 3.2. The PARCC operational items used in LEAP 2016 forms were critical to the forms construction process. Refer to the PARCC website, http://parcconline.org/, for the Bias and Sensitivity guidelines used and the processes and procedures followed by PARCC pertaining to these items.

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. By eliminating irrelevant skills or knowledge from the items, the possibility of bias is reduced.

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

Copyright © 2017 by Louisiana Department of Education.

Third, careful attention is typically given to item statistics throughout the test development process. As part of the test assembly process, attempts are made to avoid using or reusing items with poor statistical fit or distractors with positive point biserial correlations, since this may indicate that an item is tapping an ability that is irrelevant to the construct being measured. DIF statistics are also examined during test construction. Items that have exhibited significant DIF against one or more subgroups are removed from further consideration unless it is essential to include them in order to meet content specifications. The focus of statistical review during LEAP 2016 forms construction was limited to the PARCC items. The PARCC items were the critical link to the LEAP 2015 assessment and the PARCC scale and provided the majority of item points in each form. Statistical review of the DRC-CCR items was hampered due to low student participation in a pilot study. Instead a content-focused process was followed by DRC/LDOE when selecting DRC-CCR items for LEAP 2016 forms.

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 (Tables 10.1 and 10.2) indicate the degree to which members of a particular subgroup perform better or worse than expected on each item as compared to the reference group. The DIF procedures used and the results of these analyses are detailed in this section. It should be noted, however, that all items included on the LEAP were thoroughly reviewed for content and bias by LDOE and DRC content experts to ensure that the items do not tap knowledge or specific 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 and 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, language, 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 common to 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 and Shepard, 1994; Green, 1975).

Copyright © 2017 by Louisiana Department of Education.

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

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

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

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

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

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

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

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

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

For selected-response items, the MH (χ^2_{MH}) statistic was used to evaluate potential DIF items. In the MH procedure, subgroups are matched by their raw total test score, using a

Copyright © 2017 by Louisiana Department of Education.

contingency table with *K* ability levels. When applying the MH procedure, the log-odds ratio α is assumed to be constant across the *K* matched levels. The χ^2_{MH}, then, estimates a pooled common-odds ratio. Taking the natural logarithm of the common-odds ratio and its confidence limits and multiplying these with the constant –2.35, the resulting values may then be placed on the MH delta metric (Δ_{MH}) for interpretive purposes. Items were flagged for DIF using the following criteria:

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

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

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

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

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

A positive DIF value indicates that the item favors the focal group, while a negative value indicates that the item disadvantages the focal group. Tables 10.1 and 10.2 show the DIF results for the following subgroups:

Gender: Focal group is Females; Reference group is Males.

Ethnicity: Focal groups are Hispanic/Latino, American Indian or Alaska Native, Asian, Black or African American, Two or more races; Reference group is White.

Education Classification: Focal group is students who are classified as special education; Reference group is all others.

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

Copyright © 2017 by Louisiana Department of Education.

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

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

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

Tables 10.1 and 10.2 summarize the number of DIF flags by grade and test form for each focal group that included at least 200 students for ELA and mathematics, respectively. If group count was insufficient, then NR (not reported) was placed in the cell. The analyses were conducted by test form. All grades, except for grade 3, administered two forms, one form for each administration mode.

Grade 3 ELA, PBT form (see Table 10.1), can be considered as an example. In this form, one item was flagged for DIF for the female subgroup: it exhibited moderate positive DIF. One item was flagged for DIF for the Asian subgroup. This item exhibited moderate negative DIF. Two items were flagged for the special education subgroup: one displaying moderate negative DIF and one showing moderate positive DIF. Lastly, one item was flagged for the LEP subgroup: it exhibited moderate negative DIF.

Table 10.1 2016 LEAP DIF Statistics: Number of Flagged Items, English Language Arts

DIF Statistics: English Language Arts					Count of Items at DIF Magnitude			
					Moderate		Large	
Grade	Mode	Number of Items	Category	Group	B-	B+	C-	C+
3	PBT	29	Gender	Female	0	1	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	1	0	0	0
			Ethnicity	Black or African American	0	0	0	0
			Ethnicity	Two or more races	0	0	0	0
			Education Classification	Special	1	1	0	0
			LEP Status	LEP	1	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			504 Status	504	0	0	0	0

Copyright © 2017 by Louisiana Department of Education.

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

Copyright © 2017 by Louisiana Department of Education.

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

Copyright © 2017 by Louisiana Department of Education.

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

Copyright © 2017 by Louisiana Department of Education.

Table 10.2 2016 LEAP DIF Statistics: Number of Flagged Items, Mathematics

DIF Statistics: Mathematics					Count of Items at DIF Magnitude			
					Moderate		Large	
Grade	Mode	Number of Items	Category	Group	B-	B+	C-	C+
3	PBT	45	Gender	Female	0	0	0	0
			Ethnicity	Hispanic/Latino	0	0	1	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	0	0	0	1
			Ethnicity	Black or African American	1	0	1	0
			Ethnicity	Two or more races	0	0	0	0
			Education Classification	Special	4	0	1	2
			LEP Status	LEP	1	0	0	0
			Economic Status	Economically Disadvantaged	0	0	1	0
			504 Status	504	0	0	0	0
4	CBT	44	Gender	Female	NR	NR	NR	NR
			Ethnicity	Hispanic/Latino	NR	NR	NR	NR
			Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
			Ethnicity	Asian	NR	NR	NR	NR
			Ethnicity	Black or African American	NR	NR	NR	NR
			Ethnicity	Two or more races	NR	NR	NR	NR
			Education Classification	Special	NR	NR	NR	NR
			LEP Status	LEP	NR	NR	NR	NR
			Economic Status	Economically Disadvantaged	NR	NR	NR	NR
			504 Status	504	NR	NR	NR	NR
	PBT	44	Gender	Female	1	0	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	1	0	0	0
			Ethnicity	Black or African American	1	0	0	0
			Ethnicity	Two or more races	0	0	0	0
			Education Classification	Special	3	2	1	3
			LEP Status	LEP	1	0	0	0
			Economic Status	Economically Disadvantaged	1	0	0	0
			504 Status	504	0	2	0	0

Copyright © 2017 by Louisiana Department of Education.

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

Copyright © 2017 by Louisiana Department of Education.

DIF Statistics: Mathematics (continued)					Count of Items at DIF Magnitude			
					Moderate		Large	
Category	Group	Number of Items	Category	Group	B-	B+	C-	C+
7	CBT	42	Gender	Female	2	2	0	1
			Ethnicity	Hispanic/Latino	2	0	0	0
			Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
			Ethnicity	Asian	NR	NR	NR	NR
			Ethnicity	Black or African American	2	1	0	0
			Ethnicity	Two or more races	NR	NR	NR	NR
			Education Classification	Special	2	1	2	2
			LEP Status	LEP	NR	NR	NR	NR
			Economic Status	Economically Disadvantaged	1	0	0	0
			504 Status	504	0	0	1	0
	PBT	42	Gender	Female	0	2	1	1
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	2	0	0	2
			Ethnicity	Black or African American	1	1	0	0
			Ethnicity	Two or more races	0	0	0	0
			Education Classification	Special	1	2	2	1
			LEP Status	LEP	1	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			504 Status	504	1	0	0	0
8	CBT	45	Gender	Female	2	1	0	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	NR	NR	NR	NR
			Ethnicity	Asian	NR	NR	NR	NR
			Ethnicity	Black or African American	1	0	1	0
			Ethnicity	Two or more races	NR	NR	NR	NR
			Education Classification	Special	5	0	1	2
			LEP Status	LEP	NR	NR	NR	NR
			Economic Status	Economically Disadvantaged	0	0	0	0
			504 Status	504	0	0	0	0
	PBT	45	Gender	Female	0	1	1	0
			Ethnicity	Hispanic/Latino	0	0	0	0
			Ethnicity	American Indian or Alaska Native	0	0	0	0
			Ethnicity	Asian	1	2	0	0
			Ethnicity	Black or African American	2	0	0	0
			Ethnicity	Two or more races	0	0	0	0
			Education Classification	Special	5	2	0	0
			LEP Status	LEP	0	0	0	0
			Economic Status	Economically Disadvantaged	0	0	0	0
			504 Status	504	0	0	0	0

Copyright © 2017 by Louisiana Department of Education.

10.3 Evaluating Bias through Impact Analysis

The impact of achievement testing on subgroups can be determined and reported in the form of average scores and also in terms of test score reliability. Tables 10.3 through 10.13 present the number of students, test form reliability statistics (coefficient alpha, see Chapter 9) scaled score means and standard deviations, effect size (Cohen's *d*) and for the various subgroups of interest by form.

10.3.1 Reliability

Tables 10.3 through 10.13 show the regular test form reliability coefficients and SEM by student gender, ethnicity, education classification, LEP status, and economic status. The reliability coefficients for English Language Arts forms ranged from 0.80 to 0.91, except for grade 7 online reliability coefficient of 0.76 for the LEP group. For mathematics the reliability coefficients ranged from 0.83 to 0.92. This analysis shows that the test reliability is of acceptable magnitude for all of the subgroups. Note that the reliability coefficients are not reported (NR) for subgroups smaller than 10 students.

Copyright © 2017 by Louisiana Department of Education.

Table 10.3 Grade 3 Paper-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥57,160	0.87	4.30	≥57,130	0.90	3.75
Gender						
Female	≥27,920	0.86	4.39	≥27,900	0.90	3.77
Male	≥29,220	0.87	4.19	≥29,200	0.91	3.74
Ethnicity						
Hispanic/Latino	≥3,790	0.87	4.29	≥3,800	0.90	3.72
American Indian or Alaska Native	≥340	0.85	4.30	≥340	0.90	3.74
Asian	≥800	0.88	4.43	≥800	0.90	3.82
Black or African American	≥25,190	0.85	4.25	≥25,190	0.89	3.55
Native Hawaiian or Other Pacific	≥50	0.80	4.72	≥50	0.86	3.78
White	≥25,350	0.86	4.35	≥25,340	0.89	3.80
Two or more races	≥1,550	0.86	4.33	≥1,550	0.90	3.76
Education Classification						
Regular	≥51,110	0.86	4.32	≥51,090	0.90	3.76
Special	≥6,040	0.86	3.99	≥6,040	0.90	3.50
LEP Status						
Non-LEP	≥54,840	0.87	4.30	≥54,820	0.90	3.76
LEP	≥2,310	0.85	4.18	≥2,310	0.90	3.60
Economic Status						
Economically Disadvantaged	≥41,360	0.86	4.26	≥41,340	0.90	3.66
Not Economically Disadvantaged	≥14,870	0.85	4.38	≥14,880	0.89	3.81
504 Status						
Non-504	≥52,330	0.87	4.31	≥52,300	0.90	3.77
504	≥4,820	0.83	4.06	≥4,830	0.89	3.47

Copyright © 2017 by Louisiana Department of Education.

Table 10.4 Grade 4 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥450	0.87	4.32	≥450	0.91	3.75
Gender						
Female	≥250	0.86	4.41	≥250	0.90	3.78
Male	≥200	0.88	4.18	≥200	0.92	3.69
Ethnicity						
Hispanic/Latino	≥10	0.90	4.32	≥10	0.90	3.89
American Indian or Alaska Native	<10	NR	NR	<10	NR	NR
Asian	<10	NR	NR	<10	NR	NR
Black or African American	≥200	0.84	4.39	≥200	0.90	3.58
Native Hawaiian or Other Pacific	<10	NR	NR	<10	NR	NR
White	≥220	0.85	4.24	≥220	0.89	3.73
Two or more races	≥10	0.91	4.08	≥10	0.89	3.50
Education Classification						
Regular	≥400	0.86	4.31	≥400	0.90	3.72
Special	≥50	0.90	3.97	≥50	0.91	3.77
LEP Status						
Non-LEP	≥450	0.87	4.32	≥450	0.91	3.74
LEP	<10	NR	NR	<10	NR	NR
Economic Status						
Economically Disadvantaged	≥250	0.84	4.35	≥250	0.88	3.66
Not Economically Disadvantaged	≥200	0.85	4.27	≥200	0.91	3.72
504 Status						
Non-504	≥400	0.86	4.33	≥400	0.91	3.75
504	≥50	0.82	4.06	≥50	0.81	3.40

Copyright © 2017 by Louisiana Department of Education.

Table 10.5 Grade 4 Paper-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥53,520	0.88	4.52	≥53,530	0.91	3.74
Gender						
Female	≥25,930	0.87	4.56	≥25,930	0.91	3.76
Male	≥27,550	0.87	4.44	≥27,560	0.92	3.71
Ethnicity						
Hispanic/Latino	≥3320	0.88	4.54	≥3,340	0.91	3.73
American Indian or Alaska Native	≥320	0.86	4.58	≥320	0.90	3.77
Asian	≥810	0.89	4.57	≥810	0.92	3.55
Black or African American	≥24,060	0.86	4.52	≥24,070	0.90	3.69
Native Hawaiian or Other Pacific	≥40	0.85	4.57	≥50	0.89	3.68
White	≥23,600	0.86	4.51	≥23,600	0.90	3.69
Two or more races	≥1,280	0.86	4.50	≥1,280	0.90	3.73
Education Classification						
Regular	≥47,860	0.87	4.51	≥47,870	0.91	3.73
Special	≥5,650	0.86	4.30	≥5,660	0.90	3.53
LEP Status						
Non-LEP	≥51,670	0.87	4.52	≥51,670	0.91	3.74
LEP	≥1,840	0.86	4.46	≥1,860	0.91	3.66
Economic Status						
Economically Disadvantaged	≥38,570	0.86	4.51	≥38,580	0.90	3.73
Not Economically Disadvantaged	≥14,220	0.85	4.52	≥14,210	0.90	3.64
504 Status						
Non-504	≥47,800	0.87	4.53	≥47,810	0.91	3.74
504	≥5,710	0.84	4.34	≥5,710	0.89	3.63

Copyright © 2017 by Louisiana Department of Education.

Table 10.6 Grade 5 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥1,690	0.87	4.57	≥1,690	0.90	3.64
Gender						
Female	≥850	0.88	4.64	≥850	0.90	3.67
Male	≥830	0.87	4.48	≥830	0.91	3.61
Ethnicity						
Hispanic/Latino	≥90	0.89	4.57	≥90	0.91	3.63
American Indian or Alaska Native	≥50	0.84	4.39	≥50	0.88	3.46
Asian	≥10	0.84	4.61	≥10	0.88	3.87
Black or African American	≥570	0.84	4.41	≥570	0.88	3.45
Native Hawaiian or Other Pacific	<10	NR	NR	<10	NR	NR
White	≥890	0.87	4.63	≥890	0.90	3.71
Two or more races	≥50	0.89	4.56	≥50	0.90	3.61
Education Classification						
Regular	≥1,500	0.87	4.60	≥1,500	0.90	3.66
Special	≥190	0.84	4.03	≥190	0.89	3.24
LEP Status						
Non-LEP	≥1,650	0.87	4.58	≥1,650	0.90	3.65
LEP	≥40	0.81	4.48	≥40	0.83	3.43
Economic Status						
Economically Disadvantaged	≥1,100	0.86	4.54	≥1,100	0.89	3.53
Not Economically Disadvantaged	≥570	0.87	4.69	≥570	0.90	3.76
504 Status						
Non-504	≥1,490	0.87	4.61	≥1,490	0.90	3.67
504	≥190	0.80	4.06	≥190	0.84	3.24

Copyright © 2017 by Louisiana Department of Education.

Table 10.7 Grade 5 Paper-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥50,890	0.86	4.41	≥50,880	0.90	3.66
Gender						
Female	≥24,620	0.86	4.51	≥24,620	0.90	3.69
Male	≥26,250	0.87	4.27	≥26,250	0.91	3.62
Ethnicity						
Hispanic/Latino	≥3,040	0.87	4.42	≥3,050	0.90	3.67
American Indian or Alaska Native	≥290	0.86	4.40	≥290	0.89	3.62
Asian	≥780	0.89	4.53	≥780	0.92	3.82
Black or African American	≥22,750	0.84	4.39	≥22,760	0.88	3.50
Native Hawaiian or Other Pacific	≥40	0.87	4.87	≥40	0.91	3.77
White	≥22,910	0.86	4.46	≥22,890	0.90	3.72
Two or more races	≥1,010	0.87	4.37	≥1,010	0.90	3.70
Education Classification						
Regular	≥45,380	0.86	4.44	≥45,370	0.90	3.67
Special	≥5,500	0.82	3.93	≥5,510	0.86	3.33
LEP Status						
Non-LEP	≥49,510	0.86	4.42	≥49,500	0.90	3.66
LEP	≥1,370	0.81	4.20	≥1,370	0.87	3.47
Economic Status						
Economically Disadvantaged	≥36,210	0.85	4.37	≥36,190	0.89	3.57
Not Economically Disadvantaged	≥13,950	0.85	4.52	≥13,950	0.90	3.76
504 Status						
Non-504	≥45,030	0.86	4.43	≥45,020	0.90	3.67
504	≥5,850	0.81	4.22	≥5,860	0.86	3.46

Copyright © 2017 by Louisiana Department of Education.

Table 10.8 Grade 6 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥4,810	0.89	4.92	≥4,800	0.91	3.48
Gender						
Female	≥2,370	0.88	4.95	≥2,370	0.91	3.50
Male	≥2,430	0.89	4.80	≥2,430	0.92	3.46
Ethnicity						
Hispanic/Latino	≥250	0.89	4.92	≥250	0.91	3.47
American Indian or Alaska Native	≥90	0.91	4.62	≥90	0.89	3.34
Asian	≥40	0.91	4.95	≥40	0.92	3.61
Black or African American	≥1,690	0.87	4.84	≥1,690	0.89	3.33
Native Hawaiian or Other Pacific	<10	NR	NR	<10	NR	NR
White	≥2,580	0.88	4.94	≥2,580	0.91	3.54
Two or more races	≥130	0.89	4.93	≥130	0.92	3.49
Education Classification						
Regular	≥4,290	0.88	4.90	≥4,280	0.91	3.49
Special	≥510	0.85	4.57	≥510	0.87	3.12
LEP Status						
Non-LEP	≥4,740	0.89	4.92	≥4,740	0.91	3.48
LEP	≥60	0.82	4.63	≥60	0.88	3.22
Economic Status						
Economically Disadvantaged	≥2,880	0.88	4.86	≥2,870	0.90	3.37
Not Economically Disadvantaged	≥1,900	0.87	4.92	≥1,900	0.90	3.56
504 Status						
Non-504	≥4,330	0.89	4.92	≥4,320	0.91	3.49
504	≥470	0.84	4.77	≥470	0.88	3.27

Copyright © 2017 by Louisiana Department of Education.

Table 10.9 Grade 6 Paper-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥47,840	0.88	4.72	≥47,860	0.91	3.46
Gender						
Female	≥23,370	0.87	4.73	≥23,370	0.90	3.48
Male	≥24,460	0.88	4.60	≥24,460	0.91	3.43
Ethnicity						
Hispanic/Latino	≥2,690	0.89	4.74	≥2,720	0.91	3.48
American Indian or Alaska Native	≥310	0.87	4.83	≥310	0.90	3.37
Asian	≥780	0.90	4.70	≥780	0.92	3.69
Black or African American	≥22,080	0.86	4.76	≥22,070	0.88	3.33
Native Hawaiian or Other Pacific	≥40	0.91	4.73	≥40	0.93	3.63
White	≥21,100	0.87	4.68	≥21,090	0.91	3.52
Two or more races	≥790	0.87	4.76	≥790	0.91	3.51
Education Classification						
Regular	≥42,870	0.87	4.71	≥42,880	0.90	3.47
Special	≥4,970	0.85	4.49	≥4970	0.87	3.12
LEP Status						
Non-LEP	≥46,760	0.88	4.72	≥46,740	0.91	3.46
LEP	≥1,080	0.85	4.71	≥1,110	0.86	3.17
Economic Status						
Economically Disadvantaged	≥34,450	0.87	4.74	≥34,430	0.89	3.39
Not Economically Disadvantaged	≥12,680	0.86	4.64	≥12,680	0.91	3.56
504 Status						
Non-504	≥42,290	0.88	4.71	≥42,300	0.91	3.47
504	≥5,550	0.84	4.68	≥5,550	0.87	3.27

Copyright © 2017 by Louisiana Department of Education.

Table 10.10 Grade 7 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	N Count	Cronbach's Alpha	SEM	N Count	Cronbach's Alpha	SEM
All Students	≥6,310	0.86	5.35	≥6,270	0.90	3.73
Gender						
Female	≥3,020	0.85	5.32	≥3,010	0.90	3.74
Male	≥3,280	0.85	5.26	≥3,260	0.90	3.70
Ethnicity						
Hispanic/Latino	≥300	0.86	5.17	≥290	0.88	3.73
American Indian or Alaska Native	≥80	0.87	5.06	≥80	0.91	3.69
Asian	≥80	0.86	5.48	≥80	0.92	3.76
Black or African American	≥2,180	0.83	5.32	≥2,180	0.87	3.57
Native Hawaiian or Other Pacific	<10	NR	NR	<10	NR	NR
White	≥3,500	0.85	5.38	≥3,470	0.90	3.77
Two or more races	≥140	0.82	5.46	≥140	0.90	3.73
Education Classification						
Regular	≥5,740	0.84	5.34	≥5,700	0.90	3.72
Special	≥560	0.80	4.81	≥560	0.86	3.21
LEP Status						
Non-LEP	≥6,220	0.85	5.35	≥6,180	0.90	3.73
LEP	≥90	0.76	5.20	≥80	0.88	3.58
Economic Status						
Economically Disadvantaged	≥3,850	0.84	5.31	≥3,840	0.88	3.64
Not Economically Disadvantaged	≥2,410	0.84	5.37	≥2,390	0.90	3.76
504 Status						
Non-504	≥5,810	0.85	5.36	≥5,770	0.90	3.74
504	≥500	0.83	5.07	≥500	0.88	3.50

Copyright © 2017 by Louisiana Department of Education.

Table 10.11 Grade 7 Paper-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥44,650	0.85	5.15	≥44,570	0.90	3.75
Gender						
Female	≥21,620	0.84	5.14	≥21,580	0.89	3.76
Male	≥23,010	0.85	5.05	≥22,960	0.90	3.71
Ethnicity						
Hispanic/Latino	≥2,430	0.88	5.16	≥2,440	0.90	3.77
American Indian or Alaska Native	≥270	0.83	5.06	≥270	0.89	3.70
Asian	≥710	0.86	5.24	≥710	0.92	3.87
Black or African American	≥20,430	0.82	5.12	≥20,400	0.86	3.57
Native Hawaiian or Other Pacific	≥30	0.83	5.33	≥30	0.87	3.64
White	≥20,010	0.83	5.18	≥19,950	0.90	3.82
Two or more races	≥690	0.84	5.16	≥690	0.89	3.77
Education Classification						
Regular	≥40,470	0.84	5.14	≥40,390	0.89	3.75
Special	≥4170	0.79	4.85	≥4,170	0.83	3.24
LEP Status						
Non-LEP	≥43,640	0.84	5.15	≥43,540	0.90	3.75
LEP	≥1000	0.82	4.95	≥1,020	0.85	3.44
Economic Status						
Economically Disadvantaged	≥31,330	0.83	5.13	≥31,240	0.88	3.66
Not Economically Disadvantaged	≥12,590	0.82	5.15	≥12,530	0.90	3.82
504 Status						
Non-504	≥39,850	0.85	5.16	≥39,780	0.90	3.76
504	≥4,790	0.80	5.04	≥4,790	0.85	3.51

Copyright © 2017 by Louisiana Department of Education.

Table 10.12 Grade 8 Computer-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥5,450	0.89	4.95	≥4,440	0.86	3.35
Gender						
Female	≥2,630	0.88	4.89	≥2,080	0.86	3.40
Male	≥2,810	0.89	4.86	≥2,350	0.87	3.30
Ethnicity						
Hispanic/Latino	≥290	0.91	4.99	≥230	0.87	3.36
American Indian or Alaska Native	≥70	0.87	4.75	≥70	0.91	3.35
Asian	≥70	0.88	4.79	≥30	0.88	3.65
Black or African American	≥1,920	0.88	4.95	≥1,730	0.83	3.15
Native Hawaiian or Other Pacific	<10	NR	NR	<10	NR	NR
White	≥2,970	0.87	4.89	≥2,260	0.86	3.48
Two or more races	≥110	0.86	5.13	≥90	0.86	3.40
Education Classification						
Regular	≥5,010	0.87	4.93	≥4,010	0.86	3.38
Special	≥440	0.88	4.66	≥420	0.83	2.96
LEP Status						
Non-LEP	≥5,380	0.88	4.95	≥4,370	0.86	3.36
LEP	≥60	0.86	4.86	≥60	0.89	3.19
Economic Status						
Economically Disadvantaged	≥3,260	0.88	4.94	≥2,940	0.85	3.25
Not Economically Disadvantaged	≥2,160	0.86	4.87	≥1,470	0.85	3.54
504 Status						
Non-504	≥5,040	0.89	4.94	≥4,050	0.86	3.36
504	≥410	0.86	4.89	≥380	0.87	3.24

Copyright © 2017 by Louisiana Department of Education.

Table 10.13 Grade 8 Paper-Based Test Administration Reliability and SEM by Subgroup

	ELA			Mathematics		
Group	**N Count**	**Cronbach's Alpha**	**SEM**	**N Count**	**Cronbach's Alpha**	**SEM**
All Students	≥45,780	0.88	4.69	≥42,490	0.89	3.41
Gender						
Female	≥22,380	0.88	4.60	≥20,590	0.88	3.46
Male	≥23,380	0.89	4.64	≥21,870	0.89	3.36
Ethnicity						
Hispanic/Latino	≥2,390	0.91	4.84	≥2250	0.89	3.37
American Indian or Alaska Native	≥280	0.86	4.67	≥260	0.88	3.50
Asian	≥700	0.90	4.54	≥530	0.92	4.02
Black or African American	≥21,300	0.87	4.76	≥20,680	0.86	3.20
Native Hawaiian or Other Pacific	≥30	0.91	4.69	≥30	0.93	3.45
White	≥20,380	0.86	4.58	≥18,080	0.88	3.59
Two or more races	≥640	0.86	4.71	≥590	0.88	3.50
Education Classification						
Regular	≥41,560	0.87	4.64	≥38,300	0.88	3.45
Special	≥4,210	0.86	4.70	≥4,190	0.83	2.92
LEP Status						
Non-LEP	≥44,800	0.88	4.67	≥41,470	0.88	3.42
LEP	≥970	0.87	4.85	≥1,020	0.85	2.95
Economic Status						
Economically Disadvantaged	≥31,970	0.88	4.73	≥30,630	0.87	3.30
Not Economically Disadvantaged	≥13,070	0.85	4.51	≥11,060	0.88	3.69
504 Status						
Non-504	≥41,180	0.88	4.68	≥37,970	0.89	3.44
504	≥4,590	0.86	4.73	≥4,520	0.85	3.15

10.3.2 Effect Size

One way to evaluate the magnitude of the differences is to calculate the effect size. Cohen's *d* was used to calculate the effect size. Cohen's *d* is given by the following formula:

$$d = \frac{\overline{x_a} - \overline{x_b}}{\sqrt{\frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{(n_a + n_b) - 2}}},$$

Copyright © 2017 by Louisiana Department of Education.

where $\overline{x_a}$ is the mean score of group A, $\overline{x_b}$ is the mean score of group B, s_a^2 is the variance of group A, s_b^2 is the variance of group B, n_a is the number of students in group A, and n_b is the number of students in group B.

Cohen's *d*, then, expresses the difference in group means in terms of the standard deviation. For example if *d*=.34 for two groups, then it may be interpreted that the mean difference between the two groups is .34 of the pooled standard deviation. Cohen (1988) offered guidelines for interpreting the meaning of the *d* statistic: *d*=.20 is a small effect size, *d*=.50 is a medium effect size, and *d*=.80 is a large effect size.

Using Cohen's (1988) guidelines, certain trends become apparent in Tables 10.14 through 10.24. Results are not reported (NR) for subgroups with less than ten students. On the ELA test in all grades, there are medium differences in mean test scores between girls and boys where girls outperform boys. Although there were no effect sizes larger than a small effect size, |0.20|, for mathematics, girls tend to perform better than boys in all tests except for grade 4 CBT. For most ELA and mathematics tests, mean scaled scores and effect size show that Asian, Native Hawaiian, and White students tend to outperform other ethnicity groups across grades. For most ELA and mathematics tests, there were clear performance differences between regular and special education students in education classification, between not economically disadvantaged and economically disadvantaged in economic status, non-LEP and LEP in Limited English Proficient status, and non-migrant and migrant in migrant status.

Copyright © 2017 by Louisiana Department of Education.

Table 10.14 Impact Analysis, Grade 3 Paper-Based Test Administration

	ELA				Mathematics			
Group	**N**	**Scalded Score Mean**	**Scalded Score Std. Dev.**	**Effect Size**	**N**	**Scalded Score Mean**	**Scaled Score Std. Dev.**	**Effect Size**
All Students	≥57,160	736.21	36.56		≥57,130	745.98	33.43	
Gender								
Male	≥29,220	732.49	36.04		≥29,200	744.43	34.23	
Female	≥27,920	740.11	36.68	-0.21	≥27,900	747.62	32.49	-0.10
Ethnicity								
White	≥25,350	747.24	34.78		≥25,340	756.56	32.10	
Hispanic/Latino	≥3,790	729.60	37.99	0.50	≥3,800	742.62	32.97	0.43
American Indian or Alaska Native	≥340	734.38	34.42	0.37	≥340	744.25	31.87	0.38
Asian	≥800	757.87	40.01	-0.30	≥800	772.94	35.35	-0.51
Black or African American	≥25,190	725.07	34.28	0.64	≥25,190	734.79	30.64	0.69
Native Hawaiian or Other Pacific	≥50	750.78	31.12	-0.10	≥50	756.52	29.02	0.00
Two or more races	≥1,550	742.18	35.65	0.15	≥1,550	749.87	34.25	0.21
Education Classification								
Regular	≥51,110	738.92	35.82		≥51,090	748.27	32.77	
Special	≥6,040	713.28	34.68	0.72	≥6,040	726.60	32.68	0.66
Economic Status								
Not Economically Disadvantaged	≥14,870	754.79	34.41		≥14,880	762.81	32.15	
Economically Disadvantaged	≥41,360	729.88	34.97	0.72	≥41,340	740.31	31.75	0.71
LEP Status								
Non-LEP	≥54,840	737.02	36.38		≥54,820	746.39	33.38	
LEP	≥2,310	716.87	35.49	0.55	≥2,310	736.18	33.05	0.31
Migrant Status								
Nonmigrant	≥56,990	736.25	36.55		≥56,960	746.00	33.43	
Migrant	≥170	721.41	35.95	0.41	≥170	739.18	32.82	0.20
504 Status								
Non-504	≥52,330	737.86	36.57		≥52,300	747.53	33.35	
504	≥4,820	718.25	31.24	0.54	≥4,830	729.20	29.47	0.55

Copyright © 2017 by Louisiana Department of Education.

Table 10.15 Impact Analysis, Grade 4 Computer-Based Test Administration

	ELA				Mathematics			
Group	**N**	**Scaled Score Mean**	**Scaled Score Std.Dev.**	**Effect Size**	**N**	**Scaled Score Mean**	**Scaled Score Std.Dev.**	**Effect Size**
All Students	≥450	740.93	29.46		≥450	745.86	31.98	
Gender								
Male	≥200	738.21	30.04		≥200	746.44	33.76	
Female	≥250	743.14	28.92	-0.17	≥250	745.62	30.38	0.03
Ethnicity								
White	≥220	751.12	26.65		≥220	758.10	31.46	
Hispanic/Latino	≥10	739.08	37.32	0.44	≥10	749.92	31.04	0.26
American Indian or Alaska Native	<10	NR	NR	NR	<10	NR	NR	NR
Asian	<10	NR	NR	NR	<10	NR	NR	NR
Black or African American	≥200	729.90	27.85	0.78	≥200	732.64	27.29	0.86
Native Hawaiian or Other Pacific	<10	NR	NR	NR	<10	NR	NR	NR
Two or more races	≥10	733.80	33.47	0.64	≥10	734.40	25.26	0.76
Education Classification								
Regular	≥400	743.61	28.02		≥400	747.95	31.35	
Special	≥50	719.69	32.18	0.84	≥50	729.24	32.45	0.59
Economic Status								
Not Economically Disadvantaged	≥200	753.05	27.16		≥200	760.04	33.21	
Economically Disadvantaged	≥250	731.11	27.74	0.80	≥250	734.64	25.77	0.87
LEP Status								
Non-LEP	≥450	741.03	29.34		≥450	745.91	31.94	
LEP	<10	NR	NR	NR	<10	NR	NR	NR
Migrant Status								
Nonmigrant	≥450	740.93	29.46		≥450	745.86	31.98	
504 Status								
Non-504	≥400	743.60	28.82		≥400	748.72	32.18	
504	≥50	720.21	26.20	0.82	≥50	723.60	19.32	0.81

Copyright © 2017 by Louisiana Department of Education.

Table 10.16 Impact Analysis, Grade 4 Paper-Based Test Administration

	ELA				Mathematics			
Group	**N**	**Scaled Score Mean**	**Scaled Score Std.Dev.**	**Effect Size**	**N**	**Scaled Score Mean**	**Scaled Score Std.Dev.**	**Effect Size**
All Students	≥53,520	742.92	32.01		≥53,530	743.44	32.46	
Gender								
Male	≥27,550	738.84	31.37		≥27,560	742.12	32.84	
Female	≥25,930	747.26	32.11	-0.27	≥25,930	744.83	31.98	-0.08
Ethnicity								
White	≥23,600	752.99	30.22		≥23,600	754.63	31.10	
Hispanic/Latino	≥3,320	738.68	33.89	0.47	≥3,340	742.05	32.75	0.40
American Indian or Alaska Native	≥320	744.50	29.65	0.28	≥320	746.30	30.82	0.27
Asian	≥810	761.36	36.04	-0.28	≥810	770.06	35.89	-0.49
Black or African American	≥24,060	732.74	29.97	0.67	≥24,070	731.55	29.09	0.77
Native Hawaiian or Other Pacific	≥40	747.10	28.37	0.19	≥50	748.62	27.27	0.19
Two or more races	≥1,280	747.75	29.97	0.17	≥1,280	746.54	30.61	0.26
Education Classification								
Regular	≥47,860	745.83	30.97		≥47,870	745.96	32.01	
Special	≥5,650	718.33	30.04	0.89	≥5,660	722.14	28.13	0.75
Economic Status								
Not Economically Disadvantaged	≥14,220	759.93	29.54		≥14,210	760.61	31.33	
Economically Disadvantaged	≥38,570	736.89	30.54	0.76	≥38,580	737.39	30.50	0.76
LEP Status								
Non-LEP	≥51,670	743.65	31.79		≥51,670	743.84	32.45	
LEP	≥1,840	722.75	31.47	0.66	≥1,860	732.21	30.53	0.36
Migrant Status								
Nonmigrant	≥53,390	742.95	32.00		≥53,400	743.46	32.46	
Migrant	≥130	730.45	31.75	0.39	≥130	736.31	31.88	0.22
504 Status								
Non-504	≥47,800	744.76	32.03		≥47,810	745.23	32.57	
504	≥5,710	727.58	27.37	0.54	≥5,710	728.45	27.24	0.52

Copyright © 2017 by Louisiana Department of Education.

Table 10.17 Impact Analysis, Grade 5 Computer-Based Test Administration

	ELA				Mathematics			
Group	**N**	**Scaled Score Mean**	**Scaled Score Std.Dev.**	**Effect Size**	**N**	**Scaled Score Mean**	**Scaled Score Std.Dev.**	**Effect Size**
All Students	≥1,690	741.19	33.28		≥1,690	738.72	31.32	
Gender								
Male	≥830	738.30	32.96		≥830	737.74	31.64	
Female	≥850	744.02	33.37	-0.17	≥850	739.67	31.00	-0.06
Ethnicity								
White	≥890	750.23	32.15		≥890	746.76	31.11	
Hispanic/Latino	≥90	736.36	36.58	0.43	≥90	742.74	33.28	0.13
American Indian or Alaska Native	≥50	733.92	28.56	0.51	≥50	728.27	26.98	0.60
Asian	≥10	761.82	29.87	-0.36	≥10	765.12	34.33	-0.59
Black or African American	≥570	727.56	29.69	0.73	≥570	725.94	26.84	0.71
Native Hawaiian or Other Pacific	<10	NR	NR	NR	<10	NR	NR	NR
Two or more races	≥50	746.18	35.23	0.13	≥50	736.96	29.71	0.32
Education Classification								
Regular	≥1,500	744.62	32.20		≥1,500	741.41	30.82	
Special	≥190	714.43	29.30	0.95	≥190	717.64	26.98	0.78
Economic Status								
Not Economically Disadvantaged	≥570	755.10	32.15		≥570	753.09	31.48	
Economically Disadvantaged	≥1,100	734.03	31.58	0.66	≥1,100	731.33	28.53	0.74
LEP Status								
Non-LEP	≥1,650	741.89	33.07		≥1,650	739.01	31.48	
LEP	≥40	714.23	30.38	0.84	≥40	727.33	21.59	0.37
Migrant Status								
Nonmigrant	≥1,670	741.49	33.22		≥1,670	738.86	31.40	
Migrant	≥10	705.29	18.64	1.09	≥10	721.36	11.22	0.56
504 Status								
Non-504	≥1,490	744.26	32.93		≥1,490	741.43	31.39	
504	≥190	717.99	26.04	0.82	≥190	718.27	21.89	0.76

Copyright © 2017 by Louisiana Department of Education.

Table 10.18 Impact Analysis, Grade 5 Paper-Based Test Administration

	ELA				Mathematics			
Group	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
All Students	≥50,890	741.05	31.25		≥50,880	738.08	31.44	
Gender								
Male	≥26,250	737.24	30.68		≥26,250	736.48	31.85	
Female	≥24,620	745.11	31.34	-0.25	≥24,620	739.80	30.91	-0.11
Ethnicity								
White	≥22,910	749.88	29.82		≥22,890	747.80	31.19	
Hispanic/Latino	≥3,040	739.41	32.75	0.35	≥3,050	738.07	30.83	0.31
American Indian or Alaska Native	≥290	739.23	31.04	0.36	≥290	734.58	29.68	0.42
Asian	≥780	761.87	36.09	-0.40	≥780	768.25	38.75	-0.65
Black or African American	≥22,750	731.50	29.25	0.62	≥22,760	727.17	27.32	0.70
Native Hawaiian or Other Pacific	≥40	742.38	36.03	0.25	≥40	737.24	34.18	0.34
Two or more races	≥1,010	745.48	30.83	0.15	≥1,010	742.24	31.60	0.18
Education Classification								
Regular	≥45,380	744.24	30.16		≥45,370	740.62	31.24	
Special	≥5,500	714.75	27.42	0.99	≥5,510	717.22	24.59	0.76
Economic Status								
Not Economically Disadvantaged	≥13,950	756.78	29.89		≥13,950	754.98	31.89	
Economically Disadvantaged	≥36,210	735.25	29.61	0.73	≥36,190	731.87	28.78	0.78
LEP Status								
Non-LEP	≥49,510	741.67	31.12		≥49,500	738.48	31.48	
LEP	≥1,370	718.72	27.74	0.74	≥1,370	723.94	26.43	0.46
Migrant Status								
Nonmigrant	≥50,780	741.05	31.25		≥50,770	738.09	31.44	
Migrant	≥110	737.56	33.70	0.11	≥110	736.39	32.84	0.05
504 Status								
Non-504	≥45,030	742.96	31.30		≥45,020	739.86	31.72	
504	≥5,850	726.33	26.63	0.54	≥5,860	724.44	25.33	0.50

Copyright © 2017 by Louisiana Department of Education.

Table 10.19 Impact Analysis, Grade 6 Computer-Based Test Administration

	ELA				Mathematics			
Group	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
All Students	≥4,810	743.84	30.74		≥4,800	737.23	28.49	
Gender								
Male	≥2,430	738.53	30.73		≥2,430	735.67	28.96	
Female	≥2,370	749.29	29.79	-0.36	≥2,370	738.84	27.90	-0.11
Ethnicity								
White	≥2,580	751.07	29.75		≥2,580	744.66	28.05	
Hispanic/Latino	≥250	741.34	30.46	0.33	≥250	737.91	27.73	0.24
American Indian or Alaska Native	≥90	741.80	32.20	0.31	≥90	728.20	24.20	0.59
Asian	≥40	760.85	35.70	-0.33	≥40	760.75	29.02	-0.57
Black or African American	≥1,690	732.63	28.24	0.63	≥1,690	725.63	25.00	0.71
Native Hawaiian or Other Pacific	<10	NR	NR	NR	<10	NR	NR	NR
Two or more races	≥130	746.77	33.96	0.14	≥130	737.59	29.90	0.25
Education Classification								
Regular	≥4,290	747.41	29.27		≥4,280	740.07	27.75	
Special	≥510	714.26	26.39	1.14	≥510	713.72	23.10	0.97
Economic Status								
Not Economically Disadvantaged	≥1,900	757.74	28.80		≥1,900	751.21	26.90	
Economically Disadvantaged	≥2,880	734.77	28.55	0.80	≥2,870	728.05	25.65	0.89
LEP Status								
Non-LEP	≥4,740	744.27	30.61		≥4,740	737.47	28.46	
LEP	≥60	712.75	23.81	1.03	≥60	718.49	23.90	0.67
Migrant Status								
Nonmigrant	≥4,800	743.90	30.71		≥4,790	737.28	28.47	
Migrant	<10	NR	NR	NR	<10	NR	NR	NR
504 Status								
Non-504	≥4,330	745.76	30.68		≥4,320	738.86	28.52	
504	≥470	726.56	25.42	0.64	≥470	722.47	23.50	0.58

Copyright © 2017 by Louisiana Department of Education.

Table 10.20 Impact Analysis, Grade 6 Paper-Based Test Administration

	ELA				Mathematics			
Group	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
All Students	≥47,840	741.83	28.56		≥47,860	733.88	27.53	
Gender								
Male	≥24,460	736.12	27.81		≥24,460	731.86	27.93	
Female	≥23,370	747.82	28.10	-0.42	≥23,370	736.01	26.93	-0.15
Ethnicity								
White	≥21,100	749.08	27.86		≥21,090	742.53	27.19	
Hispanic/Latino	≥2,690	739.54	31.17	0.34	≥2,720	732.50	28.29	0.37
American Indian or Alaska Native	≥310	741.30	28.71	0.28	≥310	731.12	26.17	0.42
Asian	≥780	762.63	34.11	-0.48	≥780	758.54	31.32	-0.59
Black or African American	≥22,080	734.25	26.42	0.55	≥22,070	724.84	24.22	0.69
Native Hawaiian or Other Pacific	≥40	751.20	33.36	-0.08	≥40	741.93	31.62	0.02
Two or more races	≥790	747.65	27.80	0.05	≥790	737.64	27.53	0.18
Education Classification								
Regular	≥42,870	744.86	27.41		≥42,880	736.36	26.96	
Special	≥4,970	715.72	24.69	1.07	≥4,970	712.43	22.68	0.90
Economic Status								
Not Economically Disadvantaged	≥12,680	756.37	27.59		≥12,680	749.01	27.25	
Economically Disadvantaged	≥34,450	736.80	26.97	0.72	≥34,430	728.66	25.46	0.78
LEP Status								
Non-LEP	≥46,760	742.44	28.32		≥46,740	734.33	27.48	
LEP	≥1,080	715.59	26.48	0.95	≥1,110	714.89	22.43	0.71
Migrant Status								
Nonmigrant	≥47,730	741.86	28.55		≥47,740	733.89	27.53	
Migrant	≥110	730.85	32.58	0.39	≥110	726.29	25.48	0.28
504 Status								
Non-504	≥42,290	743.63	28.63		≥42,300	735.51	27.65	
504	≥5,550	728.13	24.01	0.55	≥5,550	721.42	23.08	0.52

Copyright © 2017 by Louisiana Department of Education.

Table 10.21 Impact Analysis, Grade 7 Computer-Based Test Administration

	ELA				Mathematics			
Group	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
All Students	≥6,310	743.03	36.97		≥6,270	735.89	25.77	
Gender								
Male	≥3,280	736.39	36.43		≥3,260	734.73	26.11	
Female	≥3,020	750.25	36.20	-0.38	≥3,010	737.15	25.35	-0.09
Ethnicity								
White	≥3,500	750.86	36.16		≥3,470	741.54	26.01	
Hispanic/Latino	≥300	738.77	36.35	0.33	≥290	732.94	22.83	0.33
American Indian or Alaska Native	≥80	741.32	36.25	0.26	≥80	728.64	27.73	0.50
Asian	≥80	769.95	39.94	-0.53	≥80	760.92	28.82	-0.74
Black or African American	≥2,180	730.12	34.39	0.58	≥2,180	726.82	22.24	0.60
Native Hawaiian or Other Pacific	<10	NR	NR	NR	<10	NR	NR	NR
Two or more races	≥140	742.87	34.22	0.22	≥140	732.49	25.93	0.35
Education Classification								
Regular	≥5,740	746.97	35.07		≥5,700	738.19	25.04	
Special	≥560	702.86	31.57	1.27	≥560	712.60	21.12	1.04
Economic Status								
Not Economically Disadvantaged	≥2,410	758.62	35.01		≥2,390	747.47	25.54	
Economically Disadvantaged	≥3,850	733.40	34.84	0.72	≥3,840	728.79	23.24	0.77
LEP Status								
Non-LEP	≥6,220	743.51	36.84		≥6,180	736.08	25.76	
LEP	≥90	710.87	31.10	0.89	≥80	722.80	23.02	0.52
Migrant Status								
Nonmigrant	≥6,300	743.10	36.95		≥6,260	735.92	25.76	
Migrant	≥10	702.40	30.66	1.10	<10	NR	NR	NR
504 Status								
Non-504	≥5,810	744.64	36.87		≥5,770	737.02	25.65	
504	≥500	724.44	32.90	0.55	≥500	722.81	23.41	0.56

Copyright © 2017 by Louisiana Department of Education.

Table 10.22 Impact Analysis, Grade 7 Paper-Based Test Administration

	ELA				Mathematics			
Group	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
All Students	≥44,650	742.46	34.69		≥44,570	732.57	25.71	
Gender								
Male	≥23,010	735.98	34.06		≥22,960	731.10	26.21	
Female	≥21,620	749.36	34.01	-0.39	≥21,580	734.15	25.06	-0.12
Ethnicity								
White	≥20,010	752.24	33.13		≥19,950	740.57	25.66	
Hispanic/Latino	≥2,430	736.58	40.35	0.46	≥2,440	730.73	26.84	0.38
American Indian or Alaska Native	≥270	744.17	31.55	0.24	≥270	734.32	23.52	0.24
Asian	≥710	766.99	38.18	-0.44	≥710	758.46	30.39	-0.69
Black or African American	≥20,430	732.46	32.11	0.61	≥20,400	723.97	22.14	0.69
Native Hawaiian or Other Pacific	≥30	751.92	33.45	0.01	≥30	741.46	20.47	-0.03
Two or more races	≥690	749.79	34.05	0.07	≥690	735.42	24.55	0.20
Education Classification								
Regular	≥40,470	745.81	33.36		≥40,390	734.77	25.17	
Special	≥4,170	709.99	30.29	1.08	≥4,170	711.27	20.62	0.95
Economic Status								
Not Economically Disadvantaged	≥12,590	759.88	32.13		≥12,530	746.45	25.60	
Economically Disadvantaged	≥31,330	735.93	33.12	0.73	≥31,240	727.39	23.59	0.79
LEP Status								
Non-LEP	≥43,640	743.32	34.21		≥43,540	732.98	25.63	
LEP	≥1,000	704.96	34.30	1.12	≥1,020	715.24	22.97	0.69
Migrant Status								
Nonmigrant	≥44,530	742.48	34.69		≥44,450	732.58	25.70	
Migrant	≥110	732.56	33.96	0.29	≥110	729.42	26.62	0.12
504 Status								
Non-504	≥39,850	744.51	34.63		≥39,780	734.02	25.80	
504	≥4,790	725.42	30.14	0.56	≥4,790	720.48	21.35	0.53

Copyright © 2017 by Louisiana Department of Education.

Table 10.23 Impact Analysis, Grade 8 Computer-Based Test Administration

	ELA				Mathematics			
Group	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
All Students	≥5,450	748.70	34.32		≥4,440	734.60	28.40	
Gender								
Male	≥2,810	741.42	33.61		≥2,350	732.85	28.51	
Female	≥2,630	756.47	33.37	-0.45	≥2,080	736.57	28.17	-0.13
Ethnicity								
White	≥2,970	757.76	32.18		≥2,260	741.29	27.93	
Hispanic/Latino	≥290	744.33	38.92	0.41	≥230	732.51	29.97	0.31
American Indian or Alaska Native	≥70	740.09	29.35	0.55	≥70	736.08	33.56	0.19
Asian	≥70	772.15	35.11	-0.45	≥30	756.26	28.99	-0.54
Black or African American	≥1,920	734.27	31.63	0.73	≥1,730	725.46	25.81	0.59
Native Hawaiian or Other Pacific	<10	NR	NR	NR	<10	NR	NR	NR
Two or more races	≥110	756.89	33.47	0.03	≥90	737.67	27.53	0.13
Education Classification								
Regular	≥5,010	752.11	32.45		≥4,010	737.04	27.54	
Special	≥440	709.90	31.06	1.31	≥420	711.60	26.06	0.93
Economic Status								
Not Economically Disadvantaged	≥2,160	763.51	31.61		≥1,470	745.21	26.90	
Economically Disadvantaged	≥3,260	738.96	32.56	0.76	≥2,940	729.34	27.62	0.58
LEP Status								
Non-LEP	≥5,380	749.24	34.03		≥4,370	734.80	28.30	
LEP	≥60	705.79	30.71	1.28	≥60	720.64	31.97	0.50
Migrant Status								
Nonmigrant	≥5,440	748.74	34.30		≥4,420	734.61	28.39	
Migrant	≥10	727.64	40.46	0.62	≥10	728.09	35.50	0.23
504 Status								
Non-504	≥5,040	750.11	34.35		≥4,050	735.19	28.33	
504	≥410	731.34	28.90	0.55	≥380	728.37	28.46	0.24

Copyright © 2017 by Louisiana Department of Education.

Table 10.24 Impact Analysis, Grade 8 Paper-Based Test Administration

	ELA				Mathematics			
Group	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size	N	Scale Score Mean	Scale Score Std. Dev.	Effect Size
All Students	≥45,780	747.61	32.84		≥42,490	732.28	31.94	
Gender								
Male	≥23,380	741.13	31.89		≥21,870	730.22	32.16	
Female	≥22,380	754.39	32.45	-0.41	≥20,590	734.48	31.55	-0.13
Ethnicity								
White	≥20,380	757.79	30.71		≥18,080	742.28	30.88	
Hispanic/Latino	≥2,390	740.10	39.51	0.56	≥2,250	728.39	33.47	0.45
American Indian or Alaska Native	≥280	752.21	30.69	0.18	≥260	734.36	31.09	0.26
Asian	≥700	769.63	37.64	-0.38	≥530	762.11	38.45	-0.64
Black or African American	≥21,300	737.73	30.43	0.66	≥20,680	723.02	29.29	0.64
Native Hawaiian or Other Pacific	≥30	759.13	37.69	-0.04	≥30	742.41	38.31	0.00
Two or more races	≥640	754.81	30.27	0.10	≥590	738.67	30.88	0.12
Education Classification								
Regular	≥41,560	750.87	31.48		≥38,300	735.01	31.25	
Special	≥4,210	715.44	28.23	1.14	≥4,190	707.29	26.95	0.90
Economic Status								
Not Economically Disadvantaged	≥13,070	763.84	30.26		≥11,060	747.97	31.38	
Economically Disadvantaged	≥31,970	741.51	31.38	0.72	≥30,630	727.13	30.15	0.68
LEP Status								
Non-LEP	≥44,800	748.50	32.31		≥41,470	732.82	31.84	
LEP	≥970	706.83	30.84	1.29	≥1,020	710.32	28.06	0.71
Migrant Status								
Nonmigrant	≥45,650	747.63	32.84		≥42,380	732.28	31.93	
Migrant	≥120	740.20	32.34	0.23	≥110	732.49	35.11	-0.01
504 Status								
Non-504	≥41,180	749.40	32.87		≥37,970	733.77	32.06	
504	≥4,590	731.58	27.84	0.55	≥4,520	719.79	28.01	0.44

Additional mean scaled-score data is provided in Tables 10.25 and 10.26. These tables report the number of students, scaled-score means, and standard deviations for special education classification. Groups that have fewer than 50 students are not reported (NR).

Copyright © 2017 by Louisiana Department of Education.

Table 10.25 Special Education Classification Scaled-Score Means and Standard Deviations: English Language Arts

Special Education Classification Scaled-Score Means and Standard Deviations: English Lanuage Arts							
		Yes			No		
Grade	Group	N	Mean	Std. Dev.	N	Mean	Std. Dev.
3	Gifted	≥1,260	791.44	25.83	≥55,900	734.96	35.79
	Talented	≥500	768.02	32.97	≥56,650	735.92	36.46
	Autism	≥260	706.82	35.72	≥56,890	736.34	36.51
	Deaf-Blindness	<50	NR	NR	≥57,160	736.21	36.56
	Developmental Delay	≥490	702.54	30.58	≥56,660	736.50	36.47
	Emotional Disturbance	≥120	709.40	33.02	≥57,030	736.26	36.54
	HI - Deaf	<50	NR	NR	≥57,140	736.22	36.55
	HI - Hard-of-Hearing	≥60	714.25	31.53	≥57,090	736.23	36.55
	Mild Mental Disability	≥330	683.79	22.69	≥56,820	736.52	36.40
	Moderate Mental Disability	<50	NR	NR	≥57,140	736.23	36.54
	Orthopedic Impairment	≥60	723.69	35.05	≥57,090	736.22	36.56
	Other Health Impairment	≥770	708.31	31.01	≥56,380	736.59	36.48
	Specific Learning Disability	≥2,010	706.73	27.98	≥55,140	737.28	36.38
	Speech or Language Impairment	≥1,810	732.09	36.81	≥55,340	736.34	36.54
	Traumatic Brain Injury	<50	NR	NR	≥57,150	736.21	36.56
	Visual Impairment	<50	NR	NR	≥57,110	736.21	36.56
	Other	<50	NR	NR	≥57,150	736.21	36.56
	HI - Hearing Impairment	<50	NR	NR	≥57,160	736.21	36.56
	Unknown	<50	NR	NR	≥57,150	736.21	36.56
4	Gifted	≥1,380	791.56	23.55	≥52,590	741.63	31.17
	Talented	≥780	771.22	27.02	≥53,190	742.49	31.87
	Autism	≥270	713.38	30.79	≥53,700	743.06	31.92
	Deaf-Blindness	<50	NR	NR	≥53,980	742.91	31.99
	Developmental Delay	<50	NR	NR	≥53,950	742.92	31.98
	Emotional Disturbance	≥130	712.97	27.71	≥53,850	742.98	31.96
	HI - Deaf	<50	NR	NR	≥53,950	742.92	31.98
	HI - Hard-of-Hearing	≥60	724.15	31.08	≥53,910	742.93	31.98
	Mild Mental Disability	≥300	692.71	20.10	≥53,670	743.20	31.81
	Moderate Mental Disability	<50	NR	NR	≥53,970	742.92	31.98
	Orthopedic Impairment	≥70	732.28	30.70	≥53,900	742.92	31.99
	Other Health Impairment	≥970	716.57	27.49	≥53,010	743.39	31.86
	Specific Learning Disability	≥2,430	711.39	24.84	≥51,540	744.39	31.52
	Speech or Language Impairment	≥1,340	738.59	30.92	≥52,630	743.02	32.01
	Traumatic Brain Injury	<50	NR	NR	≥53,970	742.91	31.99
	Visual Impairment	<50	NR	NR	≥53,950	742.91	31.99
	Other	<50	NR	NR	≥53,970	742.91	31.99
	HI - Hearing Impairment	<50	NR	NR	≥53,970	742.91	31.99
	Unknown	<50	NR	NR	≥53,980	742.91	31.99

Copyright © 2017 by Louisiana Department of Education.

Special Education Classification Scaled-Score Means and Standard Deviations: English Lanuage Arts (continued)							
		Yes			No		
Grade	Group	N	Mean	Std. Dev.	N	Mean	Std. Dev.
5	Gifted	≥1,600	785.64	23.89	≥50,980	739.65	30.48
	Talented	≥1,030	764.83	27.92	≥51,550	740.57	31.20
	Autism	≥290	716.24	30.70	≥52,290	741.19	31.27
	Deaf-Blindness	<50	NR	NR	≥52,580	741.05	31.32
	Developmental Delay	<50	NR	NR	≥52,530	741.08	31.31
	Emotional Disturbance	≥180	716.98	28.92	≥52,400	741.13	31.30
	HI - Deaf	<50	NR	NR	≥52,570	741.06	31.31
	HI - Hard-of-Hearing	≥60	724.91	27.24	≥52,520	741.07	31.32
	Mild Mental Disability	≥340	693.19	17.49	≥52,240	741.37	31.15
	Moderate Mental Disability	<50	NR	NR	≥52,570	741.06	31.31
	Orthopedic Impairment	≥50	728.65	34.74	≥52,530	741.06	31.31
	Other Health Impairment	≥1,050	714.06	25.28	≥51,530	741.60	31.19
	Specific Learning Disability	≥2,650	709.83	22.74	≥49,920	742.71	30.84
	Speech or Language Impairment	≥930	735.28	30.77	≥51,650	741.15	31.32
	Traumatic Brain Injury	<50	NR	NR	≥52,570	741.06	31.32
	Visual Impairment	<50	NR	NR	≥52,540	741.06	31.32
	Other	<50	NR	NR	≥52,580	741.05	31.32
	HI - Hearing Impairment	<50	NR	NR	≥52,580	741.05	31.32
	Unknown	<50	NR	NR	≥52,580	741.05	31.32
6	Gifted	≥1,700	783.65	24.49	≥50,950	740.62	27.85
	Talented	≥1,360	764.35	26.96	≥51,290	741.42	28.58
	Autism	≥240	716.58	27.47	≥52,410	742.13	28.73
	Deaf-Blindness	<50	NR	NR	≥52,650	742.01	28.77
	Developmental Delay	<50	NR	NR	≥52,630	742.03	28.77
	Emotional Disturbance	≥150	713.64	24.49	≥52,500	742.10	28.74
	HI - Deaf	<50	NR	NR	≥52,630	742.03	28.77
	HI - Hard-of-Hearing	≥50	724.21	27.54	≥52,600	742.03	28.77
	Mild Mental Disability	≥280	695.97	18.94	≥52,370	742.26	28.62
	Moderate Mental Disability	<50	NR	NR	≥52,650	742.02	28.77
	Orthopedic Impairment	≥50	734.75	32.06	≥52,600	742.02	28.77
	Other Health Impairment	≥1,050	715.28	24.03	≥51,600	742.56	28.60
	Specific Learning Disability	≥2,830	712.29	21.17	≥49,820	743.70	28.22
	Speech or Language Impairment	≥690	734.43	27.49	≥51,960	742.11	28.78
	Traumatic Brain Injury	<50	NR	NR	≥52,640	742.02	28.77
	Visual Impairment	<50	NR	NR	≥52,610	742.02	28.77
	Other	<50	NR	NR	≥52,650	742.01	28.77
	HI - Hearing Impairment	<50	NR	NR	≥52,650	742.01	28.77
	Unknown	<50	NR	NR	≥52,650	742.01	28.77

Copyright © 2017 by Louisiana Department of Education.

Special Education Classification Scaled-Score Means and Standard Deviations: English Lanuage Arts (continued)							
		Yes			No		
Grade	Group	N	Mean	Std. Dev.	N	Mean	Std. Dev.
7	Gifted	≥1,760	791.22	28.29	≥49,200	740.78	33.92
	Talented	≥1,440	768.29	31.32	≥49,510	741.77	34.79
	Autism	≥220	716.07	36.20	≥50,730	742.64	34.93
	Deaf-Blindness	<50	NR	NR	≥50,960	742.53	34.98
	Developmental Delay	<50	NR	NR	≥50,960	742.53	34.98
	Emotional Disturbance	≥180	705.69	29.02	≥50,770	742.66	34.93
	HI - Deaf	<50	NR	NR	≥50,940	742.54	34.97
	HI - Hard-of-Hearing	≥60	718.86	34.39	≥50,890	742.56	34.97
	Mild Mental Disability	≥190	681.84	20.87	≥50,760	742.76	34.82
	Moderate Mental Disability	<50	NR	NR	≥50,950	742.53	34.97
	Orthopedic Impairment	≥80	731.26	37.92	≥50,880	742.54	34.97
	Other Health Impairment	≥920	710.71	29.34	≥50,040	743.11	34.80
	Specific Learning Disability	≥2,460	704.92	26.76	≥48,490	744.44	34.26
	Speech or Language Impairment	≥490	730.42	31.74	≥50,460	742.65	34.99
	Traumatic Brain Injury	<50	NR	NR	≥50,950	742.53	34.98
	Visual Impairment	<50	NR	NR	≥50,920	742.53	34.98
	Other	<50	NR	NR	≥50,960	742.53	34.98
	HI - Hearing Impairment	<50	NR	NR	≥50,960	742.53	34.98
	Unknown	<50	NR	NR	≥50,960	742.53	34.98
8	Gifted	≥1,630	792.99	26.23	≥49,590	746.23	32.13
	Talented	≥1,590	770.65	27.81	≥49,640	746.99	32.89
	Autism	≥180	724.76	36.84	≥51,050	747.81	32.96
	Deaf-Blindness	<50	NR	NR	≥51,230	747.73	33.00
	Developmental Delay	<50	NR	NR	≥51,230	747.73	33.00
	Emotional Disturbance	≥180	713.63	29.09	≥51,050	747.85	32.95
	HI - Deaf	<50	NR	NR	≥51,220	747.73	33.00
	HI - Hard-of-Hearing	≥60	719.35	27.37	≥51,170	747.76	32.99
	Mild Mental Disability	≥210	690.58	20.56	≥51,010	747.97	32.83
	Moderate Mental Disability	<50	NR	NR	≥51,230	747.73	33.00
	Orthopedic Impairment	≥50	740.04	33.93	≥51,180	747.73	33.00
	Other Health Impairment	≥900	716.04	27.46	≥50,330	748.30	32.81
	Specific Learning Disability	≥2,570	712.20	25.50	≥48,650	749.61	32.28
	Speech or Language Impairment	≥380	733.79	31.45	≥50,850	747.83	32.99
	Traumatic Brain Injury	<50	NR	NR	≥51,220	747.73	33.00
	Visual Impairment	<50	NR	NR	≥51,190	747.74	32.99
	Other	<50	NR	NR	≥51,230	747.73	33.00
	HI - Hearing Impairment	<50	NR	NR	≥51,230	747.73	33.00
	Unknown	<50	NR	NR	≥51,230	747.73	33.00

Copyright © 2017 by Louisiana Department of Education.

Table 10.26 Special Education Classification Scaled-Score Means and Standard Deviations: Mathematics

Special Education Classification Scaled-Score Means and Standard Deviations: Mathematics							
		Yes			No		
Grade	Group	N	Mean	Std. Dev.	N	Mean	Std. Dev.
3	Gifted	≥1,250	800.57	24.23	≥55,870	744.75	32.57
	Talented	≥500	770.46	29.31	≥56,620	745.76	33.38
	Autism	≥260	723.85	37.16	≥56,870	746.08	33.38
	Deaf-Blindness	<50	NR	NR	≥57,130	745.98	33.43
	Developmental Delay	≥490	718.85	29.61	≥56,640	746.21	33.36
	Emotional Disturbance	≥120	724.30	34.88	≥57,010	746.02	33.41
	HI - Deaf	<50	NR	NR	≥57,120	745.98	33.43
	HI - Hard-of-Hearing	≥60	732.28	28.44	≥57,070	745.99	33.43
	Mild Mental Disability	≥330	694.91	21.28	≥56,800	746.28	33.25
	Moderate Mental Disability	<50	NR	NR	≥57,120	746.00	33.41
	Orthopedic Impairment	≥60	725.77	35.85	≥57,070	746.00	33.42
	Other Health Impairment	≥770	719.55	28.37	≥56,360	746.34	33.35
	Specific Learning Disability	≥2,000	720.48	24.95	≥55,120	746.91	33.33
	Speech or Language Impairment	≥1,810	745.07	34.17	≥55,320	746.01	33.40
	Traumatic Brain Injury	<50	NR	NR	≥57,120	745.98	33.43
	Visual Impairment	<50	NR	NR	≥57,090	745.99	33.43
	Other	<50	NR	NR	≥57,130	745.98	33.43
	HI - Hearing Impairment	<50	NR	NR	≥57,130	745.98	33.43
	Unknown	<50	NR	NR	≥57,130	745.98	33.43
4	Gifted	≥1,380	794.94	24.90	≥52,600	742.11	31.51
	Talented	≥780	765.79	29.13	≥53,200	743.13	32.38
	Autism	≥270	718.75	30.20	≥53,710	743.59	32.42
	Deaf-Blindness	<50	NR	NR	≥53,990	743.46	32.45
	Developmental Delay	<50	NR	NR	≥53,960	743.47	32.45
	Emotional Disturbance	≥120	716.09	25.84	≥53,860	743.52	32.44
	HI - Deaf	<50	NR	NR	≥53,960	743.47	32.45
	HI - Hard-of-Hearing	≥60	731.27	29.88	≥53,920	743.47	32.45
	Mild Mental Disability	≥310	699.13	15.43	≥53,680	743.72	32.35
	Moderate Mental Disability	<50	NR	NR	≥53,980	743.47	32.45
	Orthopedic Impairment	≥70	727.22	31.07	≥53,910	743.48	32.45
	Other Health Impairment	≥960	720.02	25.04	≥53,020	743.89	32.42
	Specific Learning Disability	≥2,430	715.20	21.03	≥51,550	744.80	32.29
	Speech or Language Impairment	≥1,340	742.31	31.84	≥52,640	743.49	32.47
	Traumatic Brain Injury	<50	NR	NR	≥53,980	743.46	32.45
	Visual Impairment	<50	NR	NR	≥53,960	743.47	32.45
	Other	<50	NR	NR	≥53,980	743.46	32.45
	HI - Hearing Impairment	<50	NR	NR	≥53,980	743.46	32.45
	Unknown	<50	NR	NR	≥53,990	743.46	32.45

Copyright © 2017 by Louisiana Department of Education.

Special Education Classification Scaled-Score Means and Standard Deviations: Mathematics (continued)							
		Yes			No		
Grade	Group	N	Mean	Std. Dev.	N	Mean	Std. Dev.
5	Gifted	≥1,600	789.84	27.48	≥50,970	736.47	30.15
	Talented	≥1,030	757.19	29.47	≥51,540	737.72	31.36
	Autism	≥290	720.87	28.20	≥52,280	738.20	31.43
	Deaf-Blindness	<50	NR	NR	≥52,570	738.10	31.44
	Developmental Delay	<50	NR	NR	≥52,530	738.12	31.44
	Emotional Disturbance	≥180	715.13	25.50	≥52,390	738.18	31.43
	HI - Deaf	<50	NR	NR	≥52,560	738.11	31.44
	HI - Hard-of-Hearing	≥60	726.91	25.03	≥52,510	738.12	31.44
	Mild Mental Disability	≥340	699.97	14.23	≥52,230	738.36	31.37
	Moderate Mental Disability	<50	NR	NR	≥52,570	738.11	31.44
	Orthopedic Impairment	≥50	723.19	30.07	≥52,520	738.12	31.44
	Other Health Impairment	≥1,040	715.45	22.24	≥51,530	738.56	31.43
	Specific Learning Disability	≥2,660	713.37	19.46	≥49,910	739.42	31.41
	Speech or Language Impairment	≥930	735.06	31.12	≥51,640	738.16	31.44
	Traumatic Brain Injury	<50	NR	NR	≥52,560	738.11	31.44
	Visual Impairment	<50	NR	NR	≥52,540	738.11	31.44
	Other	<50	NR	NR	≥52,570	738.10	31.44
	HI - Hearing Impairment	<50	NR	NR	≥52,570	738.10	31.44
	Unknown	<50	NR	NR	≥52,570	738.10	31.44
6	Gifted	≥1,690	779.54	21.63	≥50,970	732.68	26.51
	Talented	≥1,360	751.27	25.91	≥51,300	733.73	27.53
	Autism	≥240	717.27	25.92	≥52,420	734.26	27.62
	Deaf-Blindness	<50	NR	NR	≥52,670	734.18	27.63
	Developmental Delay	<50	NR	NR	≥52,640	734.19	27.63
	Emotional Disturbance	≥150	710.02	24.66	≥52,510	734.25	27.61
	HI - Deaf	<50	NR	NR	≥52,640	734.19	27.63
	HI - Hard-of-Hearing	≥50	722.95	24.34	≥52,610	734.19	27.64
	Mild Mental Disability	≥270	695.20	15.89	≥52,390	734.39	27.54
	Moderate Mental Disability	<50	NR	NR	≥52,660	734.19	27.63
	Orthopedic Impairment	≥50	724.22	28.79	≥52,610	734.19	27.63
	Other Health Impairment	≥1,050	712.01	22.18	≥51,610	734.63	27.55
	Specific Learning Disability	≥2,830	709.51	18.74	≥49,830	735.59	27.40
	Speech or Language Impairment	≥690	728.97	26.87	≥51,970	734.25	27.64
	Traumatic Brain Injury	<50	NR	NR	≥52,650	734.19	27.63
	Visual Impairment	<50	NR	NR	≥52,620	734.19	27.63
	Other	<50	NR	NR	≥52,660	734.18	27.63
	HI - Hearing Impairment	<50	NR	NR	≥52,660	734.18	27.63
	Unknown	<50	NR	NR	≥52,660	734.18	27.63

Copyright © 2017 by Louisiana Department of Education.

Special Education Classification Scaled-Score Means and Standard Deviations: Mathematics (continued)							
		Yes			No		
Grade	Group	N	Mean	Std. Dev.	N	Mean	Std. Dev.
7	Gifted	≥1,710	775.38	22.28	≥49,120	731.50	24.56
	Talented	≥1,430	748.32	23.72	≥49,400	732.53	25.66
	Autism	≥220	720.63	25.59	≥50,610	733.03	25.72
	Deaf-Blindness	<50	NR	NR	≥50,840	732.98	25.74
	Developmental Delay	<50	NR	NR	≥50,840	732.98	25.74
	Emotional Disturbance	≥180	708.97	19.85	≥50,650	733.07	25.71
	HI - Deaf	<50	NR	NR	≥50,820	732.98	25.73
	HI - Hard-of-Hearing	≥60	716.26	22.88	≥50,770	733.00	25.73
	Mild Mental Disability	≥190	696.51	13.00	≥50,640	733.12	25.67
	Moderate Mental Disability	<50	NR	NR	≥50,830	732.99	25.73
	Orthopedic Impairment	≥80	719.12	24.74	≥50,760	733.00	25.73
	Other Health Impairment	≥910	710.86	20.23	≥49,920	733.39	25.65
	Specific Learning Disability	≥2,460	708.38	17.38	≥48,370	734.23	25.46
	Speech or Language Impairment	≥500	727.68	23.83	≥50,340	733.03	25.75
	Traumatic Brain Injury	<50	NR	NR	≥50,830	732.98	25.74
	Visual Impairment	<50	NR	NR	≥50,800	732.99	25.73
	Other	<50	NR	NR	≥50,840	732.98	25.74
	HI - Hearing Impairment	<50	NR	NR	≥50,840	732.98	25.74
	Unknown	<50	NR	NR	≥50,840	732.98	25.74
8	Gifted	≥820	783.43	27.76	≥46,100	731.58	30.94
	Talented	≥1,260	749.84	28.68	≥45,660	732.02	31.57
	Autism	≥180	723.60	35.65	≥46,750	732.53	31.61
	Deaf-Blindness	<50	NR	NR	≥46,930	732.50	31.63
	Developmental Delay	<50	NR	NR	≥46,930	732.50	31.63
	Emotional Disturbance	≥180	702.53	27.69	≥46,740	732.62	31.59
	HI - Deaf	<50	NR	NR	≥46,910	732.50	31.63
	HI - Hard-of-Hearing	≥60	715.22	26.53	≥46,870	732.52	31.63
	Mild Mental Disability	≥210	688.21	18.57	≥46,710	732.71	31.53
	Moderate Mental Disability	<50	NR	NR	≥46,920	732.50	31.63
	Orthopedic Impairment	≥50	725.13	25.32	≥46,880	732.51	31.64
	Other Health Impairment	≥900	708.72	26.22	≥46,030	732.96	31.55
	Specific Learning Disability	≥2,560	705.22	24.53	≥44,360	734.08	31.28
	Speech or Language Impairment	≥360	723.78	30.03	≥46,560	732.57	31.64
	Traumatic Brain Injury	<50	NR	NR	≥46,920	732.50	31.63
	Visual Impairment	<50	NR	NR	≥46,890	732.52	31.63
	Other	<50	NR	NR	≥46,930	732.50	31.63
	HI - Hearing Impairment	<50	NR	NR	≥46,930	732.50	31.63
	Unknown	<50	NR	NR	≥46,930	732.50	31.63

Copyright © 2017 by Louisiana Department of Education.

10.4 Mode Effect Study

In addition to evaluating fairness by examining performance among subgroups, it is also important to evaluate fairness in test administration. LEAP 2016 ELA and mathematics tests were administered as both paper-based tests (PBT) and computer-based tests (CBT), except for grade 3, which was only administered as a PBT. The *Standards* indicate that results across different testing modes should be comparableThe mode comparability for LEAP 2017 CBT and PBT grades 4 through 8 was investigated using the following steps:

- The mode effect study was performed using CBT as the focal group and PBT as the reference group.
- The study was based on equivalent groups design. Equivalent PBT students that match CBT students were selected using propensity score matching (PSM).
- At the item level, DIF analysis was performed using the PSM samples.
- At the test level, effect size based on difference scores of scaled scores between CBT and PBT were used to examine the mode effect.
- Following PARCC's decision to not apply mode adjustment, it was decided to not apply any mode adjustment to the LEAP.

10.4.1 Sampling Using PSM Method

CBT was administered to a smaller number of students than PBT; therefore, CBT was designated as the focal group for propensity score matching, or PSM (Rosenbaum and Rubin, 1983), and PBT was considered the reference group. That is, all CBT students and their matching PBT students were selected using covariates (matching variables), such as LEAP 2015 ELA and mathematics raw scores and 2016 bio-demographic information, such as gender, ethnicity, free-lunch, and ELL. Note that LEAP 2015 was administered as PBT only. A sample of matching PBT students was drawn using the R package, MatchIt for PSM.

Table 10.27 shows the number of equivalent CBT and PBT students matched by PSM method. Grade 4 had a small number of CBT students, making its matching PBT student count small. For mathematics grade 4, there were 405 CBT students and 47,507 PBT students who had all PSM covariate information, such as bio-demographics, 2015 ELA/mathematics performance information, and 2016 mode information. Out of the 47,507 PBT test takers, equivalent 409 PBT students were selected by considering all covariates. There were 1541, 4437, 5765, and 4007 matched students for grades 5 to 8.

Copyright © 2017 by Louisiana Department of Education.

Table 10.27 Number of Students Used for Propensity Score Matching

Content	Grade	CBT Total	PBT Total	PBT Selected
Mathematics	3			
	4	405	47,507	405
	5	1541	45,381	1541
	6	4437	41,998	4437
	7	5765	38,995	5765
	8	4007	37,173	4007
ELA	3			
	4	405	47,528	405
	5	1542	45,363	1542
	6	4429	41,992	4429
	7	5788	39,061	5788
	8	4986	40,350	4986

Total: Number of students who have information for all covariates

At the item level, differential item functioning (DIF) analysis was performed using Mantel-Haenszel (MH) by Holland and Thayer (1988). Table 10.28 shows the number of mode DIF items flagged using the same rules that are used in NAEP. For mathematics, there were three C flag items and four B flag items for grade 4, and two C flag items and two B flag items for grade 5. For ELA, there were three C flag items and two B flag items for grade 4, and three B flag items for grade 5, one B flag item for grade 8.

Table 10.28 2016 LEAP Mode DIF Statistics: Number of Flagged Items

			P-value					DIF			
Content	Grade	N of Items	C-	C	|C|	B-	B	C-	C	B-	B
Mathematics	4	43	0	0	0	3	1	1	2	3	1
	5	43	0	0	0	0	0	1	1	1	1
	6	43	0	0	0	0	0	0	0	1	0
	7	40	0	0	0	0	0	0	0	0	0
	8	44	0	0	0	0	0	0	0	0	0
ELA	4	33	0	0	0	1	0	0	2	1	1
	5	29	0	0	0	0	0	0	0	1	2
	6	29	0	0	0	0	0	0	0	0	0
	7	29	0	0	0	0	0	0	0	0	0
	8	33	0	0	0	0	0	0	0	0	1

Scaled scores of CBT and PBT were estimated using the item parameters for score reporting, and their difference scores were calculated. Effect size of the difference scores were calculated:

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

Copyright © 2017 by Louisiana Department of Education.

Table 10.29 shows scaled score mean and standard deviation of the CBT and PBT forms. When the mean scaled scores were compared, CBT appeared slightly more difficult than PBT across all grades except for mathematics grade 8. When a flag criterion, |0.2|, which can be considered as small difference criterion, was applied, only ELA grade 4 was flagged. However, its statistical significance was not certain because of the small sample size—approximately 400.

Table 10.29 Mode Study Scaled Score Differences and Effect Size

			CBT		PBT				
Content	Grade	N of Items	Mean	Std. Dev.	Mean	Std. Dev.	Mean Diff	ES	Flag>|0.2|
Mathematics	4	43	747.46	31.99	750.72	31.37	-3.26	-0.10	
	5	43	739.09	31.37	739.77	31.72	-0.68	-0.02	
	6	43	737.75	28.38	741.08	28.33	-3.34	-0.12	
	7	40	736.77	25.61	737.35	26.72	-0.58	-0.02	
	8	44	735.46	27.99	735.15	31.27	0.32	0.01	
ELA	4	33	741.91	29.88	749.46	34.20	-7.55	-0.24	Y
	5	29	741.30	33.51	742.26	30.64	-0.96	-0.03	
	6	29	744.53	30.49	748.35	28.5	-3.82	-0.13	
	7	29	745.15	35.55	748.88	33.64	-3.73	-0.11	
	8	33	752.87	34.56	758.33	33.83	-5.46	-0.16	

10.5 Summary

In summary, the overall purpose of this chapter is to address fairness concerns that are relevant to the administration of LEAP. The information in this chapter addresses multiple best practices of the testing industry and in particular is 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)

Copyright © 2017 by Louisiana Department of Education.

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)

Copyright © 2017 by Louisiana Department of Education.

CHAPTER 11: REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

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

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Belmont, CA: Wadsworth.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.

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

DRC (2016). *Interpretive Guide: Grades 3-8 ELA and Math.* Maple Grove, MN: Author.

Educational Testing Service & Measured Progress (2016). *Final Technical Report for 2015Administration*. PARCC.

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.

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

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

Copyright © 2017 by Louisiana Department of Education.

Kolen, M. J., & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking.* New York: Springer-Verlag.

Landis, J.R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.

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

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

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

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

Pearson (2015). *Performance level Setting Technical Report.* PARCC.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70 (1), 41-55.

Schumacker, R.E. (1996). Disattenuating correlation coefficients. *Rasch Measurement Transactions*, 10, 479.

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

Thompson, S. & Thurlow, M. (2002). Universally designed assessments: Better tests for everyone! (Policy Directions. No. 14). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [November 9, 2009], from: http://www.cehd.umn.edu/NCEO/OnlinePUbs/Policy14.htm

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233–251.

Copyright © 2017 by Louisiana Department of Education.