AASA
Arizona ELA & Mathematics Assessments

2024
Technical Report

Submitted to the
Arizona Department of Education
December 2024

TABLE OF CONTENTS

Chapter 1: Introduction 10

1.1. Assessment Overview 10
1.2. Participation 10
1.3. Purpose and Intended Use of Test Scores 11
1.4. Educator Involvement 11

Chapter 2: Test Design 12

2.1. Arizona Academic Standards 12
2.2. Item Specifications 13
2.3. Test Blueprint 13
2.4. Depth of Knowledge 15
2.5. Item Types 15
2.6. Test Designs 17
2.6.1. ELA 18
2.6.1.1. Extended Writing Items 18
2.6.1.2. Grade 3 ORF Items 19
2.6.2. Mathematics 19

Chapter 3: Test Development 21

3.1. Content Development and Management Tool 22
3.2. Item Bank Analysis 22
3.3. Passage and Item Development 23
3.4. Item Review 25
3.5. Form Construction 25
3.5.1. Preparation for Item Selection 26
3.5.2. Item Selection and Positioning 26
3.5.3. Sampling Plan 26
3.6. Data Review 26
3.7. Accommodated Forms 28

Chapter 4: Test Administration 29

4.1. Test Units 29
4.2. Administration Materials 30
4.3. Administration Training 30
4.4. Sample Tests 31
4.5. Accommodations 32
4.6. Universal Test Administration Conditions 33
4.7. Universal Test Tools 34
4.8. Pearson Customer Support 35
4.9. Test Security 35

Chapter 5: Scoring and Reporting 38

5.1. Human Scoring of Open-Ended Items 38
5.1.1. Scorer Recruitment 38
5.1.2. Training 39
5.1.2.1. Writing 39
5.1.2.2. Mathematics and Reading 39
5.1.3. Quality Control 40
5.1.4. Security 41
5.2. Automated Scoring of ELA Writing Prompts 41
5.2.1. Calibration of IEA 41
5.2.2. Smart Routing 42
5.2.3. Quality Control 43

5.3. Reporting 44
Chapter 6: Classical Item Analysis 48
6.1. Data 48
6.2. Descriptive Statistics 49
6.3. Classical Item Analysis 50
6.4. Distractor Analysis 51
Chapter 7: Calibration, Equating, and Scaling 53
7.1. Calibration Sample 53
7.2. Calibration Methods 53
7.3. Calibration Results 54
7.4. Equating 55
7.5. Scaling Methods 56
7.6. IRT Assumptions 56
7.6.1. Unidimensionality 57
7.6.2. Local Item Independence 58
7.6.3. Item Fit 58
Chapter 8: Test Results 60
Chapter 9: Reliability and Validity 67
9.1. Reliability 67
9.1.1. Internal Consistency 67
9.1.2. Inter-rater Reliability 70
9.2. Differential Item Functioning 72
9.3. Correlations Among Reporting Categories 74
9.4. Validity Evidence 77
9.4.1. Evidence Based on Test Content 78
9.4.2. Evidence Based on Response Processes 79
9.4.3. Evidence Based on Internal Structure 79
9.4.4. Evidence Based on Performance Standards 80
9.4.5. Evidence Based on Relation to Other Variables 80
9.4.6. Summary 81
Chapter 10: Classification into Performance Levels 82
10.1. Standard Setting 82
10.2. Classification Consistency and Accuracy 83
10.3. MOWR Policy 85
References 86
Appendix A: Item-Level CTT Statistics 89
Appendix B: Item-Level IRT Statistics 115
Appendix C: Spring 2024 Administration Results 185
Appendix D: Summer 2024 Administration Results 198

LIST OF TABLES

Table 1.1. Schedule of Major Events 11
Table 2.1. AASA ELA Blueprint, Grades 3–8 14
Table 2.2. AASA Mathematics Blueprint, Grades 3–5 14
Table 2.3. AASA Mathematics Blueprint, Grades 6-7 14
Table 2.4. AASA Mathematics Blueprint, Grade 8 14
Table 2.5. DOK Levels 15
Table 2.6. Percentage of Points by DOK Level 15
Table 2.7. Item Types 16
Table 2.8. AASA Test Design—ELA 17
Table 2.9. AASA Test Design—Mathematics 17
Table 3.1. Number of Newly Developed Items 23
Table 3.2. Passage Lexile Measures and Word Count 23
Table 3.3. Item Statistical Flagging Criteria 27
Table 3.4. Data Review Results: Number of Field Tested Items 27
Table 4.1. Estimated Testing Time by Test Unit 30
Table 4.2. Administration Materials 30
Table 4.3. Administration Trainings 31
Table 4.4. Number of Items on the AASA Sample Tests 32
Table 4.5. Available Accommodations 32
Table 4.6. Universal Test Tools 34
Table 5.1. Scoring Qualification Standards 39
Table 6.1. Number of Students in the Calibration Sample by Subgroup—ELA 48
Table 6.2. Number of Students in the Calibration Sample by Subgroup—Mathematics 49
Table 6.3. Classical Test Analysis Statistics 49
Table 6.4. Classical Item Analysis Summary 50
Table 6.5. Distractor Analysis Summary: Point-Biserial Correlations for Correct Options 51
Table 6.6. Distractor Analysis Summary: Point-Biserial Correlations for Incorrect Options 52
Table 7.1. IRT Statistics Summary 54
Table 7.2. Summary of Anchor Items 56
Table 7.3. Eigenvalues from PCA 57
Table 7.4. Q3 Statistics 58
Table 7.5. IRT Item Fit Summary Statistics 59
Table 8.1. Overall Test Results by Year 60
Table 8.2. Performance Distributions by Reporting Category: Percentage of Students at each Level of Mastery—ELA 61
Table 8.3. Performance Distributions by Reporting Category: Percentage of Students at each Level of Mastery—Mathematics 62
Table 8.4. Test Results by Accommodation—ELA 62
Table 8.5. Test Results by Accommodation—Mathematics 64
Table 8.6. Scale Score Distribution by Performance Level—ELA 65
Table 8.7. Scale Score Distribution by Performance Level—Mathematics 66
Table 9.1. Coefficient Alpha and SEM by Total and Reporting Category Score—ELA, Form 1 68
Table 9.2. Coefficient Alpha and SEM by Total and Reporting Category Score—ELA, Form 2 68
Table 9.3. Coefficient Alpha and SEM by Total and Reporting Category Score—Mathematics 69
Table 9.4. Inter-rater Reliability Statistics 71
Table 9.5. DIF Flag Categories 73
Table 9.6. Number of Items Exhibiting Strong DIF 74
Table 9.7. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—ELA Form 1 75

Table 9.8. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—ELA Form 2 ... 76
Table 9.9. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—Mathematics Grades 3–5 ... 76
Table 9.10. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—Mathematics Grades 6 and 7 ... 77
Table 9.11. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—Mathematics Grade 8 ... 77
Table 9.12. Correlation between AASA ELA and Mathematics Scale Scores ... 80
Table 10.1. Performance Level Cut Scores ... 82
Table 10.2. CSEM at Performance Level Cuts ... 83
Table 10.3. Classification Consistency for the *Proficient* Cut ... 84
Table 10.4. Classification Accuracy for the *Proficient* Cut ... 84
Table 10.5. Classification Consistency and Accuracy Results ... 85
Table A.1. Item-Level CTT Statistics, ELA Grade 3 ... 89
Table A.2. Item-Level CTT Statistics, ELA Grade 4 ... 90
Table A.3. Item-Level CTT Statistics, ELA Grade 5 ... 91
Table A.4. Item-Level CTT Statistics, ELA Grade 6 ... 92
Table A.5. Item-Level CTT Statistics, ELA Grade 7 ... 93
Table A.6. Item-Level CTT Statistics, ELA Grade 8 ... 95
Table A.7. Item-Level CTT Statistics, Mathematics Grade 3 ... 96
Table A.8. Item-Level CTT Statistics, Mathematics Grade 4 ... 97
Table A.9. Item-Level CTT Statistics, Mathematics Grade 5 ... 98
Table A.10. Item-Level CTT Statistics, Mathematics Grade 6 ... 99
Table A.11. Item-Level CTT Statistics, Mathematics Grade 7 ... 100
Table A.12. Item-Level CTT Statistics, Mathematics Grade 8 ... 101
Table A.13. Distractor Analysis of Multiple-Choice Items, ELA Grade 3 ... 103
Table A.14. Distractor Analysis of Multiple-Choice Items, ELA Grade 4 ... 104
Table A.15. Distractor Analysis of Multiple-Choice Items, ELA Grade 5 ... 105
Table A.16. Distractor Analysis of Multiple-Choice Items, ELA Grade 6 ... 106
Table A.17. Distractor Analysis of Multiple-Choice Items, ELA Grade 7 ... 107
Table A.18. Distractor Analysis of Multiple-Choice Items, ELA Grade 8 ... 108
Table A.19. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 3 ... 109
Table A.20. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 4 ... 110
Table A.21. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 5 ... 111
Table A.22. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 6 ... 112
Table A.23. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 7 ... 113
Table A.24. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 8 ... 114
Table B.1. Item-Level IRT Statistics, ELA Grade 3 ... 115
Table B.2. Item-Level IRT Statistics, ELA Grade 4 ... 116
Table B.3. Item-Level IRT Statistics, ELA Grade 5 ... 117
Table B.4. Item-Level IRT Statistics, ELA Grade 6 ... 118
Table B.5. Item-Level IRT Statistics, ELA Grade 7 ... 120
Table B.6. Item-Level IRT Statistics, ELA Grade 8 ... 121
Table B.7. Item-Level IRT Statistics, Mathematics Grade 3 ... 122
Table B.8. Item-Level IRT Statistics, Mathematics Grade 4 ... 123
Table B.9. Item-Level IRT Statistics, Mathematics Grade 5 ... 124
Table B.10. Item-Level IRT Statistics, Mathematics Grade 6 ... 125
Table B.11. Item-Level IRT Statistics, Mathematics Grade 7 ... 126
Table B.12. Item-Level IRT Statistics, Mathematics Grade 8 ... 127
Table B.13. Raw-to-Scale Score Conversion, ELA Grade 3 ... 129
Table B.14. Raw-to-Scale Score Conversion, ELA Grade 4 ... 131

Table B.15. Raw-to-Scale Score Conversion, ELA Grade 5 133
Table B.16. Raw-to-Scale Score Conversion, ELA Grade 6 135
Table B.17. Raw-to-Scale Score Conversion, ELA Grade 7 138
Table B.18. Raw-to-Scale Score Conversion, ELA Grade 8 140
Table B.19. Raw-to-Scale Score Conversion, Mathematics Grade 3 142
Table B.20. Raw-to-Scale Score Conversion, Mathematics Grade 4 143
Table B.21. Raw-to-Scale Score Conversion, Mathematics Grade 5 144
Table B.22. Raw-to-Scale Score Conversion, Mathematics Grade 6 145
Table B.23. Raw-to-Scale Score Conversion, Mathematics Grade 7 146
Table B.24. Raw-to-Scale Score Conversion, Mathematics Grade 8 147
Table C.1. Test Results by Subgroup, ELA Grade 3 185
Table C.2. Test Results by Subgroup, ELA Grade 4 186
Table C.3. Test Results by Subgroup, ELA Grade 5 186
Table C.4. Test Results by Subgroup, ELA Grade 6 187
Table C.5. Test Results by Subgroup, ELA Grade 7 187
Table C.6. Test Results by Subgroup, ELA Grade 8 188
Table C.7. Test Results by Subgroup, Mathematics Grade 3 188
Table C.8. Test Results by Subgroup, Mathematics Grade 4 189
Table C.9. Test Results by Subgroup, Mathematics Grade 5 189
Table C.10. Test Results by Subgroup, Mathematics Grade 6 190
Table C.11. Test Results by Subgroup, Mathematics Grade 7 190
Table C.12. Test Results by Subgroup, Mathematics Grade 8 191

LIST OF FIGURES

Figure 3.1. Item Development Process 21
Figure 3.2. Text Complexity Worksheet Example 24
Figure 4.1. Test Security Agreement 36
Figure 5.1. Dynamic Model Development and Deployment 42
Figure 5.2. Smart Routing 43
Figure 5.3. Sample Reports—Confidential Student Score Report 45
Figure 5.4. Sample Reports—Confidential Roster Report with Summary 47
Figure B.1. Item-Person Map, ELA Grade 3, Form 1 149
Figure B.2. Item-Person Map, ELA Grade 3, Form 2 149
Figure B.3. Item-Person Map, ELA Grade 4, Form 1 150
Figure B.4. Item-Person Map, ELA Grade 4, Form 2 150
Figure B.5. Item-Person Map, ELA Grade 5, Form 1 151
Figure B.6. Item-Person Map, ELA Grade 5, Form 2 151
Figure B.7. Item-Person Map, ELA Grade 6, Form 1 152
Figure B.8. Item-Person Map, ELA Grade 6, Form 2 152
Figure B.9. Item-Person Map, ELA Grade 7, Form 1 153
Figure B.10. Item-Person Map, ELA Grade 7, Form 2 153
Figure B.11. Item-Person Map, ELA Grade 8, Form 1 154
Figure B.12. Item-Person Map, ELA Grade 8, Form 2 154
Figure B.13. Item-Person Map, Mathematics Grade 3 155
Figure B.14. Item-Person Map, Mathematics Grade 4 155
Figure B.15. Item-Person Map, Mathematics Grade 5 156
Figure B.16. Item-Person Map, Mathematics Grade 6 156
Figure B.17. Item-Person Map, Mathematics Grade 7 157
Figure B.18. Item-Person Map, Mathematics Grade 8 157
Figure B.19. TCC, ELA Grade 3, Form 1 158
Figure B.20. CSEM, ELA Grade 3, Form 1 158

Figure B.21. TCC, ELA Grade 4, Form 1 159
Figure B.22. CSEM, ELA Grade 4, Form 1 159
Figure B.23. TCC, ELA Grade 5, Form 1 160
Figure B.24. CSEM, ELA Grade 5, Form 1 160
Figure B.25. TCC, ELA Grade 6, Form 1 161
Figure B.26. CSEM, ELA Grade 6, Form 1 161
Figure B.27. TCC, ELA Grade 7, Form 1 162
Figure B.28. CSEM, ELA Grade 7, Form 1 162
Figure B.29. TCC, ELA Grade 8, Form 1 163
Figure B.30. CSEM, ELA Grade 8, Form 1 163
Figure B.31. TCC, ELA Grade 3, Form 2 164
Figure B.32. CSEM, ELA Grade 3, Form 2 164
Figure B.33. TCC, ELA Grade 4, Form 2 165
Figure B.34. CSEM, ELA Grade 4, Form 2 165
Figure B.35. TCC, ELA Grade 5, Form 2 166
Figure B.36. CSEM, ELA Grade 5, Form 2 166
Figure B.37. TCC, ELA Grade 6, Form 2 167
Figure B.38. CSEM, ELA Grade 6, Form 2 167
Figure B.39. TCC, ELA Grade 7, Form 2 168
Figure B.40. CSEM, ELA Grade 7, Form 2 168
Figure B.41. TCC, ELA Grade 8, Form 2 169
Figure B.42. CSEM, ELA Grade 8, Form 2 169
Figure B.43. TCC, Mathematics Grade 3 170
Figure B.44. CSEM, Mathematics Grade 3 170
Figure B.45. TCC, Mathematics Grade 4 171
Figure B.46. CSEM, Mathematics Grade 4 171
Figure B.47. TCC, Mathematics Grade 5 172
Figure B.48. CSEM, Mathematics Grade 5 172
Figure B.49. TCC, Mathematics Grade 6 173
Figure B.50. CSEM, Mathematics Grade 6 173
Figure B.51. TCC, Mathematics Grade 7 174
Figure B.52. CSEM, Mathematics Grade 7 174
Figure B.53. TCC, Mathematics Grade 8 175
Figure B.54. CSEM, Mathematics Grade 8 175
Figure B.55. Scree Plot, ELA Grade 3, Form 1 176
Figure B.56. Scree Plot, ELA Grade 4, Form 1 176
Figure B.57. Scree Plot, ELA Grade 5, Form 1 177
Figure B.58. Scree Plot, ELA Grade 6, Form 1 177
Figure B.59. Scree Plot, ELA Grade 7, Form 1 178
Figure B.60. Scree Plot, ELA Grade 8, Form 1 178
Figure B.61. Scree Plot, ELA Grade 3, Form 2 179
Figure B.62. Scree Plot, ELA Grade 4, Form 2 179
Figure B.63. Scree Plot, ELA Grade 5, Form 2 180
Figure B.64. Scree Plot, ELA Grade 6, Form 2 180
Figure B.65. Scree Plot, ELA Grade 7, Form 2 181
Figure B.66. Scree Plot, ELA Grade 8, Form 2 181
Figure B.67. Scree Plot, Mathematics Grade 3 182
Figure B.68. Scree Plot, Mathematics Grade 4 182
Figure B.69. Scree Plot, Mathematics Grade 5 183
Figure B.70. Scree Plot, Mathematics Grade 6 183
Figure B.71. Scree Plot, Mathematics Grade 7 184
Figure B.72. Scree Plot, Mathematics Grade 8 184

Figure C.1. Total Scale Score Distribution, ELA Grade 3 192
Figure C.2. Total Scale Score Distribution, ELA Grade 4 192
Figure C.3. Total Scale Score Distribution, ELA Grade 5 193
Figure C.4. Total Scale Score Distribution, ELA Grade 6 193
Figure C.5. Total Scale Score Distribution, ELA Grade 7 194
Figure C.6. Total Scale Score Distribution, ELA Grade 8 194
Figure C.7. Total Scale Score Distribution, Mathematics Grade 3 195
Figure C.8. Total Scale Score Distribution, Mathematics Grade 4 195
Figure C.9. Total Scale Score Distribution, Mathematics Grade 5 196
Figure C.10. Total Scale Score Distribution, Mathematics Grade 6 196
Figure C.11. Total Scale Score Distribution, Mathematics Grade 7 197
Figure C.12. Total Scale Score Distribution, Mathematics Grade 8 197

Chapter 1: INTRODUCTION

This technical report documents the design, development, administration, technical processes, and results of the Spring 2024 administration of Arizona's Academic Standards Assessment (AASA) in English language arts (ELA) and mathematics in Grades 3–8 to support test users in evaluating the intended purposes, uses, and interpretations of the test scores. The technical information herein is intended for use by those who evaluate tests, interpret scores, or use test results in making educational decisions. It is assumed that the reader has technical knowledge of test construction and measurement procedures, as stated in the *Standards for Educational and Psychological Testing* (AERA et al., 2014).

1.1. Assessment Overview

AASA is the statewide achievement test for Arizona students in ELA and mathematics in Grades 3–8 aligned with the Arizona Academic Standards as described in state and federal law (State Law ARS 15-741; Federal Law: 34 CFR 200.2 *Participation in Assessments*). It is a summative, criterion-referenced assessment designed to promote increasingly higher academic outcomes for students to prepare them for a broader array of post-secondary outcomes. It is available as a computer-based test (CBT) or paper-based test (PBT), with CBT as the default administration mode.

In November 2014, the State Board of Education adopted Arizona's Measurement of Educational Readiness to Inform Teaching (AzMERIT) to measure student mastery of the Arizona academic standards and progress toward college and career readiness, with the first administration in Spring 2015. The current Arizona Academic Standards were adopted by the Arizona State Board of Education in December 2016. Beginning in 2019–2020, AzMERIT was renamed AzM2. Beginning in 2021–2022, AzM2 was renamed to AASA. AASA is aligned to the 2016 academic content standards and has the same cut scores.

A Writing standalone field test (SAFT) was administered in 2022 to all students in Grades 3–8 to build Arizona's item bank for extended writing items. Oral Reading Fluency (ORF) field test items were also embedded on the Grade 3 operational AASA test in Spring 2022 to enhance coverage of the Grade 3 ELA standards. They were field tested again in Spring 2023 and Spring 2024 to further explore their functioning and performance. An AASA ELA Reading test administration was available for the first time in Summer 2024 for Grade 3 students who failed to meet the Move on When Reading (MOWR) indicator on the Spring 2024 AASA ELA administration and who did not qualify for one of the good cause exemptions. Appendix D presents the results from that administration.

1.2. Participation

Students in Grades 3–8 participate in the spring administration of the AASA test. The state and federal laws mandate that all public school students participate in the assessments that measure student achievement of grade-level content standards. Students with significant cognitive disabilities whose Individualized Education Program (IEP) designates them as eligible for an alternate assessment, the Multi-State Alternate Assessment (MSAA), should not be administered the AASA assessment.

1.3. Purpose and Intended Use of Test Scores

The primary intended score interpretation of AASA is that AASA test scores provide reliable and valid information about important knowledge and skills in grade-level numeracy and literacy that students are attaining. Furthermore, while ultimate use of the test scores is determined by Arizona educators and other stakeholders, the primary intended uses of the AASA test scores include the following:

- Schools and districts use the AASA assessment and its results to (a) monitor trends in student performance and (b) design professional development for teachers.
- Teachers use the AASA assessment and its results to integrate assessment with their instructional planning.
- Parents/guardians use the AASA assessment and its results to get information about (a) what their child knows and can do and (b) their child's progress from year to year.

1.4. Educator Involvement

This section addresses the involvement of Arizona educators in test development as indicated by Standard 4.8 of the *Standards for Educational and Psychological Testing* (AERA et al., 2014). Arizona educators were involved in many steps of the process, as shown in Table 1.1 that presents the major events regarding the development, administration, and reporting of the Spring 2024 AASA assessments.

Arizona educators participated in meetings and provided feedback on assets developed for field testing. These meetings were held virtually and included educators from across the state. The committee meetings included a passage review that enabled educators to review ELA passages for content, grade-level appropriateness, and bias and sensitivity; a content and bias item review that enabled educators to review items for content, standard alignment, grade-level appropriateness, and bias and sensitivity; and a bias and sensitivity community review that enabled community members, including past and present Arizona educators, to evaluate items for bias and sensitivity concerns.

Table 1.1. Schedule of Major Events

Event	Date(s)
ELA Passage Review	February 15, 2023
ELA Community Passage Review	February 16, 2023
Content and Bias Item Review	June 20–23, 2023
Bias and Sensitivity Community Review	July 10–11, 2023
Technical Advisory Committee (TAC) Meeting	June 16, 2023
Administration Training	December 12, 2023
Spring 2024 CBT Administration Window	April 1–26, 2024
Spring 2024 PBT Administration Window	April 1–10, 2024
Spring 2024 ELA Writing Administration Window	April 1–12, 2024
Summer 2024 Administration Window	June 17 – July 19, 2024
Release of Grade 3 Electronic Score Reports	May 15, 2024
Release of Grades 4–8 Electronic Score Reports	May 24, 2024
Release of Grades 3–8 Paper Reports to Districts	June 13, 2024
Data Review	June 17–19, 2024

Chapter 2: TEST DESIGN

This chapter provides information regarding test design as indicated by Standards 1.11, 4.0, 4.1, 4.12, 12.4, and 12.8 (AERA et al., 2014). AASA is designed to be administered online, with paper accommodated forms available as needed. The needs of the student are also addressed through other supports, such as assessment features built into the online platform and accommodations such as using assistive technology, a scribe, and/or sign language (see Chapter 4: for more information). ELA includes 42 operational items consisting of multiple-choice and technology-enhanced item types, along with an open-response writing prompt at each grade level. In Grade 3 ELA only, three short Oral Reading Fluency (ORF) passages are included. Mathematics consists of 53–55 operational multiple-choice and technology-enhanced items. Field test items are also embedded on each assessment that do not count toward students' scores.

Accessibility was the foundation of the AASA test design to make sure all students have access to the content based on the college and career ready Arizona Academic Standards, which begins with rigorous curriculum, instructional resources, and training for teachers. Principles of Universal Design are adhered to throughout the item and test creation process to accommodate the needs and abilities of all learners. AASA is available to be administered in online settings including group, small group, or one-on-one settings. AASA is also available in appropriate accommodations including ASL, Braille, Large Print, or Regular Print format.

2.1. Arizona Academic Standards

In 2016, the State Board of Education adopted new academic content standards in ELA and mathematics that reflect high expectations of all Arizona students and strive to ensure that high school graduates are college and career ready. The Arizona Academic Standards define the knowledge, understanding, and skills that need to be taught and learned so all students are ready to succeed in credit-bearing, college-entry courses and/or in the workplace.

The ELA standards describe the reading, writing, language, speaking, and listening skills that students should acquire from Grades K–12, and the mathematics standards describe expectations for learning in Grades K–8 and the first three high school courses (Algebra I, Geometry, Algebra II; Mathematics 1, 2, 3), plus specific standards that could be included in a fourth high school credit mathematics course. The standards are located on the Arizona Department of Education (ADE) website at https://www.azed.gov/standards-practices.

The standards work together in a clear progression from Grades K–12. Each standard builds on the standard that came before and toward the standard that comes in the next grade level. They are the foundation to guide the construction and evaluation of programs in Arizona K–12 schools and the broader Arizona community. The Arizona Academic Standards are (a) focused in coherent progressions across Grades K–12, (b) aligned with college and workforce expectations, (c) inclusive of rigorous content and applications of knowledge through higher-order thinking, (d) research and evidence based, (e) broad in nature, allowing for the widest possible range of student learning, and (f) designed as an integrated approach to literacy (ELA).

2.2. Item Specifications

AASA item specifications are available for each grade and content area on the ADE website at https://www.azed.gov/assessment/aasa. These item specifications, refined by Pearson and ADE content experts, are used to guide the item development process by defining the content limit, model tasks, and response types for a specific standard. During each level of review, items are compared to the item specifications to ensure their alignment to the standard, grade-level appropriateness, and adherence to the content limits set forth in the item specifications.

The item specifications were developed using a vertical alignment for each standard, wherein the suggested task demands and cognitive complexity of items build upon those of the previous grade level, just as the standards themselves do. The item specifications also provide models for item writers that include item samples that target different Depth of Knowledge (DOK) and difficulty levels. These item models annotate the information to communicate the intent of the standard and DOK and clarify how to manipulate the item difficulty while keeping the cognitive demands the same for the writer. The item specifications document includes the following:

- **Content Limits.** This section delineates the specific content measured by the standard and the extent to which the content is different across grade levels. For example, in Grade 3 Mathematics, fraction denominators are limited to 2, 3, 4, 6, and 8.
- **Acceptable Response Mechanisms.** This section identifies the various ways in which students may respond to a prompt (e.g., multiple choice, graphic response, equation response, matching, multi-select).
- **Task Demands.** In this section, the standards are broken down into specific task demands aligned to the standard. In addition, each task demand is assigned a common item format relevant to that particular task demand.

2.3. Test Blueprint

The test blueprint, in concert with the item specifications, defines the content and structure of the test. Table 2.1 – Table 2.4 present a summary of the blueprints based on the 2016 standards for Grades 3–8 in ELA and mathematics. External, public-facing blueprints are available on the ADE website at https://www.azed.gov/assessment/aasa. More detailed blueprints are used internally by ADE and the vendor. The blueprint defines the standards to be assessed for each test form, the number of items per standard, the number of item types, the number of points per item type, and the total number of items and points per test form. Inherent in the number of points per test is the relative weighting associated with the standards and the reporting categories being assessed.

Table 2.1. AASA ELA Blueprint, Grades 3–8

Reporting Category	Grades 3–5		Grades 6–8	
	Min.	Max.	Min.	Max.
Reading Standards for Literature	26%	35%	24%	31%
Reading Standards for Informational Text	26%	35%	30%	38%
Reading for Informational Text	26%	22%	30%	25%
Listening Comprehension	0%	13%	0%	13%
Writing and Language	26%	38%	30%	38%
Writing	13%	19%	17%	19%
Language	13%	19%	13%	19%

Note. Listening standards are only assessed on the online assessment.

Table 2.2. AASA Mathematics Blueprint, Grades 3–5

Reporting Category	Grade 3		Grade 4		Grade 5	
	Min.	Max.	Min.	Max.	Min.	Max.
Operations and Algebraic Thinking and Numbers and Operations in Base Ten	49%	53%	46%	54%	38%	42%
Operations and Algebraic Thinking	38%	42%	22%	26%	4%	8%
Numbers in Base Ten	9%	13%	24%	28%	31%	35%
Numbers and Operations – Fractions	18%	22%	29%	33%	31%	35%
Measurement and Data and Geometry	26%	30%	15%	19%	24%	28%
Measurement and Data	26%	28%	9%	13%	18%	20%
Geometry	1%	4%	4%	7%	7%	11%

Table 2.3. AASA Mathematics Blueprint, Grades 6-7

Reporting Category	Grade 6		Grade 7	
	Min.	Max.	Min.	Max.
Ratios and Proportions	19%	23%	19%	23%
The Number System	28%	32%	19%	23%
Expressions and Equations	29%	33%	23%	27%
Geometry and Statistics and Probability	15%	19%	27%	35%
Geometry	6%	15%	15%	19%
Statistics and Probability	6%	11%	12%	16%

Table 2.4. AASA Mathematics Blueprint, Grade 8

Reporting Category	Grade 8	
	Min.	Max.
Functions	21%	25%
Expressions & Equations	29%	33%
Geometry	17%	21%
Statistics and Probability and The Number System	19%	27%
Statistics and Probability	4%	8%
The Number System	15%	19%

2.4. Depth of Knowledge

All items are aligned according to DOK, the cognitive complexity of the item, and the cognitive demands on the student. DOK refers to the level of rigor or sophistication of the task in an item designed to reflect the complexity of the Arizona Academic Standards. Table 2.5 presents a description of the DOK levels as provided in the item specifications, and Table 2.6 presents the percentage of points by DOK level as provided in the blueprints.

Table 2.5. DOK Levels

DOK Level	ELA	Mathematics
Level 1: Recall	Focuses on basic tasks such as correcting grammatical and spelling errors, defining terms, and locating details or facts in texts.	Focuses on the recall of information, such as definitions, terms, and simple procedures.
Level 2: Skill/Concept	Requires a greater degree of engagement and cognitive processing than DOK 1 items. DOK 2 items may require students to show relationships or identify examples, use context to identify meaning, identify structures or features of texts, or distinguish between facts and opinions.	Requires students to make decisions, solve problems, or recognize patterns. In general, DOK 2 items require a greater degree of engagement and cognitive processing than DOK 1 items.
Level 3: Strategic Thinking	Features higher-order cognitive tasks that assess students' capacities to read complex texts and think abstractly and focuses on critical thinking, developing, and assessing logical arguments, making inferences, and citing evidence to support claims or conclusions.	Features higher-order cognitive tasks that assess students' capacities to approach abstract or complex problems.
Level 4: Extended Thinking (Writing only)	Requires creativity, extensive planning, and/or sophisticated reasoning in the composition and organization of written essays.	N/A

Table 2.6. Percentage of Points by DOK Level

DOK Level	ELA	Mathematics
Level 1	10–20%	10–20%
Level 2	50–60%	60–70%
Level 3	15–25%	12–30%
Level 4	16–19% (Writing)	N/A

2.5. Item Types

The AASA assessments include traditional multiple-choice items and technology-enhanced items (TEIs), as shown in Table 2.7. Examples of each item type may be found in the AASA sample tests accessed through TestNav (see Section 4.4 for more information).

TEIs require students to interact with test content to select, construct, and/or support their responses and are better able to assess a deeper level of understanding. For paper-based assessments, TEIs are modified or replaced with another item type that assesses the same standards so they can be scanned and scored electronically or hand scored. For example, gap match/gap match table, and short-constructed response items may be replaced with another item type that assesses the same standard and can be scanned and scored electronically. Inline choice items are modified so the student fills in a circle to indicate the correct word or phrase, and hot text items are modified so the student fills in a circle to indicate a selection.

Table 2.7. Item Types

Item Type	Description
Multiple-Choice (MC)	The student selects only one correct answer from among a number of options.
Multiple-Select (MS)	The student selects all of the correct answers from among a number of options.
Evidence-Based Selected Response (EBSR) (ELA only)	• MC/MS Format: The student answers a Part A multiple-choice item based on a passage and then provides evidence in support of that answer by completing another Part B multiple-choice item or a Part B multi-select item. • MC/TEI Format: The student answers a Part A multiple-choice item based on a passage and then provides evidence in support of that answer by completing a Part B technology-enhanced item.
Bar Graph (mathematics only)	The student drags bars vertically or horizontally along numerical values. Individual bars, histograms, and clusters are supported.
Equation Editor (mathematics only)	The student uses a palette of buttons to enter a numerical response or to create mathematical expressions.
Fraction Model (mathematics only)	The student divides a shape (circle or rectangle) into varying numbers of segments by clicking a 'Fewer' or 'More' button and selects those segments to shade those segments with a solid color.
Point Graph (mathematics only)	The student plots points, line segments, continuous lines, and/or polygons. Point graph items can use one or multiple graph interactions (composite graphs).
Shape Transformation (mathematics only)	The student chooses one of four variants of a single shape, drags it onto a four-quadrant grid, and positions it on the grid.
Inline Choice (IC)	The student selects a single text option from a drop-down menu within a table or inline text, similar to a fill-in-the-blank item. The item may contain multiple blanks.
Gap Match (GM)	Certain numbers, words, phrases, or sentences may be designated "draggable" in this item type. The student can click on the option, hold down the mouse button, and drag it to a graphic or other format.
Gap Match Table (GMT)	Same as the gap match item except the drop zone is in a table format.
Match – Table Grid (MTG)	The student selects radio buttons or checks boxes in cells to indicate if information from a column header matches information from a row.
Hot Text (HT) (ELA only)	The student selects one or more areas called hot spots on an image. For ELA, excerpted sentences from the text are presented in this item type. Certain words, phrases, or sentences are highlighted to indicate that the text is selectable ("hot"). The student can then click on an option to select it.
Hot Spot (mathematics only)	The student selects one or more areas called hot spots on an image. An example for mathematics is selecting a point on a number line. The student can click on an option to select it.
Short Constructed Response (SCR) (ELA only)	The student uses the keyboard to enter a response into a text field. These items can usually be answered in a sentence or two.
Writing Prompt (ELA only)	These items may require the student to use features of an online word processor. The student can perform various tasks within the online word processor such as bold text, use bullet points, underline, etc.

2.6. Test Designs

Table 2.8 and Table 2.9 present the test designs for the ELA and mathematics assessments. As shown in the tables, the AASA test consists of the following test units:

- ELA Oral Reading Fluency (ORF) test unit (Grade 3 only)
- ELA Writing test unit
- ELA Reading/Language Test Unit 1 and Test Unit 2
- Math Test Unit 1 and Test Unit 2

Each grade-level ELA and mathematics test form includes the same operational items but a different set of embedded field test items. The ELA assessments consist of three test units (Writing, Reading/Language Test Unit 1, Reading/Language Test Unit 2), with a fourth ORF unit for Grade 3 only. The mathematics assessments consist of two test units (Math Test Unit 1 and Math Test Unit 2). The tables indicate the number of operational and field test items included on the test form for each unit. Given the nature of passage-based item sets in Reading/Language, field test items are confined to their associated set in only one unit of the test.

Table 2.8. AASA Test Design—ELA

		Overall				#Items by Test Unit											
			#Items			Writing			Reading/Language Test Unit 1			Reading/Language Test Unit 2			Oral Reading Fluency (ORF)		
Grade	#Forms	#Passages	OP	FT	Total	OP	FT	Total	OP	FT	Total	OP	FT	Total	OP	FT	Total
3	18	8	42	10	52	1	–	1	15	7	22	26	–	26	–	3	3
4	18	8	42	7	49	1	–	1	16	7	23	25	–	25	–	–	–
5	18	7	42	7	49	1	–	1	18	7	25	23	–	23	–	–	–
6	18	7	42	7	49	1	–	1	17	7	24	24	–	24	–	–	–
7	18	7	42	7	49	1	–	1	16	7	23	25	–	25	–	–	–
8	18	7	42	7	50	1	–	1	16	7	23	25	–	25	–	–	–

Note. Each writing prompt is worth 10 points. The test design for ELA is based on the number of items, and the total points per operational form vary from 52–56 points. For Grade 3, the ORF passages are worth 2 points each. The #Passages are specific to the two Reading test units.

Table 2.9. AASA Test Design—Mathematics

		#Items								
		Overall			Test Unit 1			Test Unit 2		
Grade	#Forms	Total	OP	FT	Total	OP	FT	Total	OP	FT
3	11	53	45	8	27	23	4	26	22	4
4	11	53	45	8	27	23	4	26	22	4
5	11	53	45	8	27	23	4	26	22	4
6	11	55	47	8	27	23	4	28	24	4
7	11	55	47	8	27	23	4	28	24	4
8	11	55	47	8	27	23	4	28	24	4

Note. Each operational item is worth 1 point. Grades 3–5 have 45 points possible, and Grades 6–8 have 47 points possible.

2.6.1. ELA

The ELA test has a Writing part and a Reading Part 1 and Part 2 for all grade levels. Writing consists of one writing prompt, which is an extended text/essay response. The Reading/Language is a long test, so it is split into two units. Each unit includes both reading and language items. Grade 3 also includes an ORF test unit.

The ELA passages represent a variety of genres and topics in the reading/language test units. Pearson's content experts develop informational texts from multiple content areas, such as history, science, and technical subjects. Literary texts represent authentic pieces from multiple genres, including stories, poetry, and drama. The ratio of informational to literary texts increases at each grade band, with a greater percentage of informational texts in the upper grades. The AASA uses both single passages and passage sets in which students are asked to synthesize information across texts. The number of items associated with each varies depending on the actual set and what standards are assessed.

2.6.1.1. Extended Writing Items

The AASA ELA assessment is designed to reflect the importance of using evidence and reading complex texts outlined in the Arizona Academic Standards. It includes extended writing items that provide students with meaningful contexts in which to construct their responses. Each writing prompt presents students with various stimuli (at least 2–3 per task) that serve as a springboard for an informed piece of writing. Students are given research articles, charts and graphs, and narratives to serve as the basis for their written responses. Students can then use this information, along with their own reasoning, to formulate an essay that is a clear and coherent expression of their own thinking while being grounded in research and evidence.

Each student is administered a single informative/explanatory or opinion/argumentative writing essay. While each student will only see one type of writing essay, both types are administered operationally at every grade level each year. Informative/explanatory writing is focused on conveying information accurately and seeks to enlighten the reader about processes or procedures, phenomena, states of affairs, and terminology. To produce this kind of writing, students draw from what they already know and from primary and secondary sources and develop a main idea and a primary focus as they relate facts, details, and examples.

Opinion (Grades 3–5) and argumentative (Grades 6–11) prompts ask students to analyze primary and secondary sources, make sound judgments, and present their opinions or arguments in a coherent manner that weaves personal opinions with evidence from the texts. The stimuli present opposing points of view about a topic so that students have enough information to take a stand. The stimuli are followed by a prompt that asks students to write an opinion or argumentative essay. The students must synthesize information across the passages to write the essay and cite specific details to support the ideas they present. For example, the prompt might require students to describe the steps in a process or describe problems that need to be solved.

The reading level of the stimulus does not exceed the easy Lexile range for the grade level to enable the students to attend to the content of the passages and not struggle with unfamiliar language and non-content-related vocabulary. Moreover, this helps ensure that students are assessed on their writing skills and not their reading abilities.

A Writing standalone field test (SAFT) was administered in 2022 to all students in Grades 3–8 to build Arizona's item bank for extended writing items. In line with this work, AASA writing rubrics and guides were made available in August 2022. The rubrics can be used in classrooms to score students' work to prepare them for the AASA Writing test unit. The writing guides are included in the test environment with each prompt as a reminder to students of key pieces from the rubric to include in their essays. They can be used in classrooms with assignments or to help students complete classroom or district essays throughout the school year.

2.6.1.2. Grade 3 ORF Items

The ELA ORF items were field tested again in Spring 2024 and did not count toward students' scores. The ELA ORF test unit consists of three items (i.e., short passages) that students read aloud to measure oral reading fluency (i.e., a student's ability to read words aloud). Students receive points based on their words correct per minute (WCPM), or the number of words a student reads correctly per minute. Each Grade 3 student receives three ORF items worth 2 points each. Students have one minute to read each passage, and they receive a score of 0, 1, or 2. Word counts for ORF passages range between 250 and 400 words, and each of the three passages has a different difficulty level of low, medium, or high based on their pre-determined Lexile® ranges established by ADE (600L – 650L, 650L – 700L, and 700L – 750L). Each student receives a combination of fiction and nonfiction genres.

2.6.2. Mathematics

The mathematics items are created to address key components of the Arizona mathematics standards that check a student's conceptual understanding of mathematics and their procedural skills. The standards for Math Practices are embedded within all AASA items. The items are written in accordance with the item specifications to address key components of the standards and assess a range of important skills and performance levels based on the performance level descriptors (PLDs) that provide a standard level description of the level of knowledge and skills required at each performance level of the assessment. This provides an opportunity for students at all performance levels to show their understanding of the mathematics standards in the assessment. Each item is also aligned to a DOK level and the overall percentage of points by DOK level, as outlined in Table 2.6.

Equation editor items are an item type unique to mathematics. For an equation editor response, students type with a keyboard or use a palette of buttons to enter a response that could be a number, an expression, or an equation. The response may contain scaffolding where students are given part of a solution and fill in the missing parts. Two types of palettes are used in equation editor items that provides quick access to mathematical operators and symbols. For numerical responses, an abbreviated palette is given that contains the digits 0–9, a decimal point, a negative sign, a button to add a fraction, and a button to add a mixed number button. For expression or equation responses, the palette contains everything from the abbreviated palette plus additional mathematical operators and symbols depending on the grade level.

Calculators are not allowed for the mathematics assessments in Grades 3–6. For the Grades 7 and 8 assessments, where calculator use is allowed for some item types, the items are grouped into two units administered separately to students: calculator and no calculator. The construct of the items dictates in which section they are to be assessed.

Arizona has determined content emphasis in the standards at the cluster level for each grade and course. Major clusters are considered as groups of related standards that require greater emphasis than some of the others due to the depth of the ideas and the time it takes to master these groups of related standards. Supporting clusters are considered as groups of related standards that support standards within the major cluster in and across grade levels. Supporting clusters also encompass pre-requisite and extension of grade-level content. Arizona suggests instructional time encompass a range of at least 65%–75% for major clusters and a range of 25%–35% for supporting cluster instruction. Content emphasis can be found at the beginning of all grade-level standards documents at https://www.azed.gov/standards-practices/k-12standards/mathematics-standards. The major and supporting clusters align with the test blueprints for AASA mathematics.

Chapter 3: TEST DEVELOPMENT

This chapter addresses Standards 1.11, 3.2, 3.6, 4.0, 4.1, 4.4, 4.6, 4.7, 4.8, 4.10, 4.12, 7.0, 7.2, 12.4, and 12.8 (AERA et al., 2014) regarding item development and test construction. ADE and Pearson worked together to construct the AASA tests based on the steps depicted in Figure 3.1.

Figure 3.1. Item Development Process

Items used to develop the Spring 2024 operational test forms were drawn from the item pool of Arizona-owned items and writing prompts custom-developed to align to the Arizona Academic Standards. The item development process is iterative, allowing for multiple opportunities for review of the items by various stakeholders including ADE and external passage and item content and bias review participants. Newly developed items are then field tested during the spring administration, followed by a data analysis and data review process with Arizona stakeholders. Items that pass data review are added to the operational item bank.

This multistage development and review process provides ample opportunity to evaluate items for their accessibility, appropriateness, and adherence to the principles of Universal Design. In this way, accessibility serves as a primary area of consideration throughout the item development process. This focus on accessibility is critical in developing an assessment that allows for the widest range of student participation as educators seek to provide access to the general education curriculum and foster higher expectations for students.

3.1. Content Development and Management Tool

The item pool, as well as content development and test construction processes, are managed within Pearson's Assessment Banking and Building solutions for Interoperable assessments tool (ABBI) that acts as a content development and management tool, item bank, and publication system supporting both paper-pencil and online publication. The item development workflow is designed to move items and assets from inception through a series of content, fairness, graphic, and other reviews to final publication. The system captures the outcomes at each review and maintains previous versions of each item. As items travel through the review process, every version of each asset is archived, along with each comment received in any review. Reviewers have immediate access to all older versions, providing version control throughout development.

ABBI allows remote internet access by item writers and reviewers while ensuring security with individualized passwords for all users, limited access for external users, and strong encryption of all information. Forms are also built in ABBI. After items are used, ABBI stores the resulting statistics, including exposure statistics and classical and item response theory (IRT) statistics.

The item development process is predicated on a high level of interaction between test developers at Pearson and ADE, as well as with Arizona educators and stakeholders. Pearson's ABBI manages item content throughout the entire lifecycle of an item. It also manages item content beyond the operational life of the item, including items identified for use in sample tests or other training materials. ABBI provides on-demand reports of the content and item bank status. Each item is directed through a sequence of reviews and approvals by Pearson and ADE before it is identified for field test or operational administration.

3.2. Item Bank Analysis

Pearson conducted an item bank analysis at the start of the test development cycle to identify gaps that were then used to determine the priorities for new item development. For ELA, the gap analysis examined the Arizona-owned items in the bank eligible for operational use. A comparison to the blueprint requirements revealed the standards underrepresented in the bank as the focus for new development. For mathematics, the gap analysis identified areas of need for standards coverage as the focus for new item development.

An item development plan was created based on the item bank analysis that outlines the number of items needed to be developed by item type, standard, and DOK. Table 3.1 presents the number of newly developed items that varied by grade and content area depending on the needs of the bank. Standards that were underrepresented in the item bank, or were represented by items with poorly performing statistics, were identified as candidates for item development. Blueprint requirements were also used to determine which standards most needed new item development.

Table 3.1. Number of Newly Developed Items

Assessment	#Items for FT
ELA 3	136
ELA 4	136
ELA 5	138
ELA 6	137
ELA 7	142
ELA 8	141
Mathematics 3	60
Mathematics 4	57
Mathematics 5	56
Mathematics 6	58
Mathematics 7	60
Mathematics 8	61

3.3. Passage and Item Development

Item development for ELA began with the development of reading passages. All new reading passages are commissioned by professional writers who are current or retired educators, while some legacy passages are permissioned. To ensure that all passages align to the correct grade level and provide sufficient complexity for close analytical reading, test developers adhere to detailed passage specifications. The passage specifications call for a close examination of both quantitative measures, such as word counts and Lexile readabilities as shown in Table 3.2, and qualitative measures such as passage structure and levels of meaning, all of which are defined as important measures of text complexity. For example, content experts use passage text complexity worksheets based on the passage specifications to analyze each passage in-depth, as illustrated in Figure 3.2. Table 3.2 also presents the Lexile measures and word count for passages used in the Grade 3 ORF test.

Table 3.2. Passage Lexile Measures and Word Count

Grade	Lexile Range	Word Count Range	ORF Lexile Range	ORF Word Count
3	420–820	100–700	600–750	250–400
4	740–1010	100–900	–	–
5	740–1010	200–1,000	–	–
6	925–1185	200–1,100	–	–
7	925–1185	300–1,100	–	–
8	925–1185	350–1,200	–	–

Note. ORF = Oral Reading Fluency

Figure 3.2. Text Complexity Worksheet Example

UIN:	**Word Count:**	
Title:		
Genre:	**Sub-Genre:**	
Quantitative Measures	**Flesch-Kincaid:**	**Lexile:**
Qualitative Considerations		
Identify the theme and/or central message and describe how it is adequately developed. *(Theme and central message should be similar or the same across paired texts.)*		
Briefly describe how the characters are adequately developed, including how they respond to an event or how they change throughout the story.		
Describe the overall structure of a text and how it contributes to the development of the theme, setting, or plot.		
Briefly describe additional plot elements *(setting, climax, rising and falling action)* **that demonstrate clear plot development and how they are similar and/or different across multiple texts. (***Paired text only***.)**		
Explain how you, the author, develops the points of view from which each text is narrated.		
Compare/contrast the differences between the texts when considering genre, theme, and topic.		
Identify one higher level words used in the passage(s) and identify its text support for understanding meaning.		
List grade-level appropriate examples of literary devices used throughout the passage *(e.g., metaphor, onomatopoeia, flashback, foreshadowing, voice, irony, symbolism).*		
Identify a phrase from the text that has a figurative or connotative meaning and describe the text support.		
Holistically, this text should be considered: **ACCESSIBLE** **MODERATELY COMPLEX** **HIGHLY COMPLEX** for grade____.		

The next step of item development for ELA and the first step for mathematics was training item writers and introducing them to project requirements. Writers relied on existing item specifications and the Arizona Academic Standards to guide item development. Items were submitted in batches and revised as needed based on feedback from Pearson, with open communication throughout the writing process. Queries were addressed in a timely manner to facilitate a deeper understanding of the Arizona standards and ADE expectations.

Throughout all steps, Pearson responded to ADE feedback, revised, and resubmitted for approval as needed. An integral part of this process was a review by Pearson research librarians who verified accuracy of information and by Pearson copyeditors who reviewed for clarity and correct use of grammar, punctuation, and spelling. All asset creators and reviewers also apply the principles of Universal Design to meet the goal of maximizing accessibility and minimizing construct-irrelevant demands for all items. To meet these goals, text complexity was controlled, graphics were designed to be clear, and subject matter that might affect the student's performance was monitored. Pearson also paid close attention to respecting the diverse cultures of the American Indian tribes in Arizona, particularly to the presentation of topics related to animals.

3.4. Item Review

ADE pre-review was the first of several external reviews of the newly developed passages and items. Educators and community members also had opportunities to participate in review committees. Content and bias review allowed educators to apply their familiarity with Arizona students and the Arizona Academic Standards to provide feedback on the accuracy and appropriateness of the item and stimulus content. A bias and sensitivity community review also allowed parents and other community stakeholders to review assets.

Prior to beginning review, committee members received training from Pearson assessment specialists and were provided resources, including a checklist, to guide the review process. All feedback was recorded in ABBI. The overall goals for both committees were to confirm alignment to the standards, ensure that assets had no bias or sensitivity issues, and revise the assets as needed to be appropriate for Arizona students. An additional benefit of these interactions was that Pearson gained insight to help guide future item development.

ADE and Pearson engaged in a reconciliation process to review committee feedback. Pearson revised assets based on ADE guidance and made the newly edited versions available for ADE review. With ADE approval, the assets went through a final editorial review at Pearson to confirm that they met style expectations and that no errors had unintentionally been introduced.

3.5. Form Construction

Once the newly developed items were ready for field testing, the next step was to construct the test forms, beginning with selecting and positioning the items.

3.5.1. Preparation for Item Selection

Parameters based on the test construction blueprint for each grade were loaded into ABBI by Pearson psychometricians and verified by Pearson assessment specialists. Different test map views were configured based on the specific needs of various users, including the Pearson assessment specialists, ADE and Pearson psychometricians, and Pearson publishing teams. Test maps for each stage were maintained throughout all steps of production. Pearson updated the test maps when any replacements or changes to items or item metadata were made.

Pearson psychometricians had previously loaded statistics from the previous administrations, and Pearson assessment specialists had updated the ABBI item status used to indicate eligibility for operational or field test selection based on the results from data review. Item statistics included, but were not limited to, classical difficulty (*p*-value) and item response theory difficulty (Rasch), item discrimination (point-biserial correlation by total score and by reporting category score), the Rasch model fit indices (infit/outfit), differential item functioning (DIF) flags as a measure of possible bias, coefficient alpha, kappa, and distractor analysis.

3.5.2. Item Selection and Positioning

The overriding goal in selecting items for the forms was adhering to the blueprint requirements. Additional criteria for item selection included item positioning and both content and statistical considerations. For each grade, a Pearson assessment specialist did an initial pull of operational items using the tools embedded in ABBI to verify blueprint alignment and acceptable statistics according to the test construction specifications. A different assessment specialist reviewed the form and provided feedback, identifying issues such as clueing. After issues were resolved, Pearson and ADE psychometricians reviewed the form and provided feedback based on statistical considerations. This process repeated until the form met psychometric approval.

The form is also reviewed by the ADE content and psychometrics teams who work with Pearson throughout the process, including final item selection for each form (including the paper and braille versions) and ensuring the psychometric thresholds. Revisions were made based on ADE feedback, and ADE provided the final approval. Once the operational forms were approved, Pearson selected the field test items, with ADE reviewing the field test selections and Pearson revising as needed.

3.5.3. Sampling Plan

All grades for ELA had 18 forms, and all grades for mathematics had 11 forms. The operational items were the same on all forms within a grade. The test forms were randomly assigned at a student level within a testing group, by TestNav, Pearson's online test delivery platform. Only one paper-pencil version was available per grade.

3.6. Data Review

Field tested items were flagged based on the criteria in Table 3.3. During data review, committee members reviewed the flagged items and their item statistics to determine whether they were eligible for the operational item pool. Two different committees meet for data review. One committee group focused solely on the items flagged for DIF, while another group reviewed the items flagged by the remaining statistics (e.g., item difficulty, point biserial, distractor analysis and Rasch values). The DIF committee looks at the possibility of bias in each item flagged for DIF.

The meeting began with a training session that introduced the item review process, including an overview of the item statistics and how they should be used to evaluate items. Decisions about an item's quality cannot be made on statistics alone; the item itself and the content it measures should also be considered. Thus, the groups also reviewed the content of the items and how the items functioned according to the statistics before making a consensus decision about whether the item should be accepted or rejected for operational use. Revisions were recommended for the rejected items if applicable. The decisions were documented in ABBI for all items that were reviewed.

Table 3.3. Item Statistical Flagging Criteria

Statistic	Criterion	Possible Indication
P-value	< 0.2 or > 0.9	Very difficult or easy item
Point-biserial correlation	< 0.25	Poorly discriminating item
Distractor point-biserial correlation (MC only)	> 0.05	Possible miskey*
Omit rate	> 2%	Skipped item
Rasch difficulty	< -3 or > 3	Easy or difficult item
Item fit statistics	< 0.6 or > 1.4	Poor fit
Score point percentage (multi-point items only)	< 1%**	Very few students got a certain score
Differential item functioning (DIF)	B, C	Item could be biased toward a certain student demographic group

*Possible miskey because the key should have a positive point-biserial correlation
**I.e., there should be at least 1% of students at each score point (multi-point items only)

Table 3.4 presents the data review results based on the Spring 2024 data. Committee members made these decisions based on the item content, using the item statistics to guide their discussion. Accepted items were added to the operational item pool for future use. The ELA reading items cannot be re-field tested because they are tied to passages, so the only options are to accept or reject the items. Because the data review committee only reviewed the flagged field tested items, this table does not reflect the total number of field tested items because many did not have any statistical issues or they had fatal statistical issues (e.g., negative point-biserial) that removed them from the item pool.

Table 3.4. Data Review Results: Number of Field Tested Items

Assessment	#Accepted	#Accepted w/Edits	#Rejected
ELA 3	105	0	21
ELA 4	99	0	27
ELA 5	96	0	30
ELA 6	97	0	29
ELA 7	102	0	24
ELA 8	107	0	19
Mathematics 3	25	2	0
Mathematics 4	19	1	3
Mathematics 5	31	2	3
Mathematics 6	25	6	3
Mathematics 7	32	2	2
Mathematics 8	32	3	5

3.7. Accommodated Forms

Each grade and content area had one form of the paper-pencil Special Paper Version (SPV). The Pearson content team worked with ADE to produce paper-equivalent versions of the items used on the online test form. Upon approval of the item set, the Pearson publishing team worked with ADE to determine an approved paper-based test template for each grade. There were three rounds of review between ADE and Pearson before the document was approved to print. A final PDF printer proof was provided to ADE.

Upon approval of the paper-pencil form, Pearson began work on the Large Print and Braille forms. The Large Print forms are enlarged versions of the paper-pencil test forms. The publishing team enlarged the entire test book file to reach an 18-point font equivalent. The final Large Print printer proof file was posted for ADE's review and approval.

The Inkprint Braille version of the test was modified based on the Braille modification document to reflect any item omissions or modifications on the Student Braille Test Book. Pearson Braille Services reviewed all forms presented for Braille to determine if forms were well-suited for Braille testers. Any recommended modifications were reviewed in conjunction with ADE to arrive at final decisions. ADE then reviewed the Inkprint Test Book, the Student Braille Test Book proof, the Braille Test Administration Directions, and the Braille memo before production of the Braille material commenced.

Each grade and content area also had one form created for ASL testers. After approval by ADE of the online test form, Pearson ASL team began work for ASL translation. The Pearson ASL team created scripts to be used for filming of the ASL translation by professional ASL signers. Video sessions for ASL Filming were attended by the Pearson ASL team as well as Pearson content for any questions that arose during translation. ADE had final approval of any modifications necessary for successful ASL filming. All ASL videos and test forms were reviewed and approved by ADE before final production.

Chapter 4: TEST ADMINISTRATION

This chapter describes how the AASA assessments were administered, including the procedures used to ensure that the test administration was conducted in a secure and standardized manner, as indicated by Standards 1.10, 3.1, 3.9, 3.10, 4.2, 4.5, 4.15, 4.16, 4.21, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 7.0, and 7.8 (AERA et al., 2014). The AASA assessment is administered online via TestNav, Pearson's online testing platform that students use to access the assessment, with accommodated forms available as needed. PearsonAccess^{next} (PAN) is the student test management portal that test administrators use to manage student tests and registrations and order materials if needed.

District Test Coordinators (DTCs), School Test Coordinators (STCs), and Test Administrators (TAs) received online training and the supporting documents to ensure fidelity of implementation and the validity of the assessment results and to help prevent, detect, and respond to irregularities in academic testing and maintain testing integrity practices for technology-based assessments. For example, TAs were instructed to use the *Test Administration Directions* (TAD) for the online and paper administrations, as well as for the Special Paper Version (SPV) tests and entering student responses into TestNav.

When all TAs use the same well-defined administration procedures and are provided the same training, manuals, and supporting documents, administration is optimally standardized and poised to be fair to all students. DTCs were responsible for supporting the TAs in understanding and following the administration procedures. Comprehensive test coordinator training and materials targeted to their role and responsibility ensure that they are appropriately prepared to support the test administrators.

4.1. Test Units

Table 4.1 presents the estimated time to complete each test unit. A test unit must be completed prior to starting the next one. All ELA Writing and Reading test units must be administered to receive an ELA score, and both mathematics test units must be administered to receive a mathematics score. The ELA Writing test must be administered on a separate day than the ELA Reading and mathematics units. ELA Reading and mathematics test units could be administered in any order, with no more than two test units plus the Grade 3 ORF unit in a single day. If two test units were administered on the same day, there must be a significant break between them. ADE requires that a test unit be submitted within the day that it is started. Any test that is not complete at the end of the testing day is marked complete and submitted for scoring by Pearson.

As part of the operational test administration, Grade 3 students also participated in the ORF test unit that was field tested in Spring 2022 and again in Spring 2023 and Spring 2024; thus, the items were not included in scoring. Each student read three separate passages, with a time limit of one minute per passage. The ORF online test unit was to be administered in small groups, with no more than six students testing simultaneously in a classroom or a computer lab environment. For paper-based testing, ORF was administered one-on-one on a speaker telephone.

Table 4.1. Estimated Testing Time by Test Unit

Unit	Testing Time
ELA Writing	60–90 minutes
ELA Reading Test Unit 1	45–75 minutes
ELA Reading Test Unit 2	45–75 minutes
Grade 3 Oral Reading Fluency (ORF)	15 minutes
Math Test Unit 1	60–85 minutes
Math Test Unit 2	60–85 minutes

Note. The testing time is the same for the CBT and PBT administrations.

4.2. Administration Materials

Table 4.2 describes the materials provided to support the standardized administration of the AASA assessments and ensure fair testing for all students. The TAD and *Test Coordinator's Manual* (TCM) were produced in collaboration with ADE. The Pearson program team drafted each manual using the previous year's version as a template. The manuals were then composed in desktop publishing software and sent for an editorial review. After a review of all comments and edits by the program team, the file was delivered for ADE review. There were multiple rounds of review between ADE and Pearson before the document was approved to print. ADE was provided with a final web-ready 508 compliant version in addition to the final printer's proof. Hard copies were sent automatically to all participating schools, and a limited number were available for additional order during the additional order window. The materials are available on the ADE website at https://www.azed.gov/assessment/aasa.

Table 4.2. Administration Materials

Material	Description
Test Administration Directions (TAD)	Provides an overview of the AASA test administration, including the user roles in PAN and the test administration schedule, and directions about what to do before, during, and after testing. Provided for both the CBT and PBT assessments.
Test Coordinator's Manual (TCM)	Indicates the responsibilities of the DTCs before, during, and after testing and explains the procedures for test administration. DTCs must review the TCM and the TAD well in advance of training STCs and TAs and before administering the tests. DTCs are responsible for ensuring the appropriate and correct administration of the AASA in all schools within the district or under the same charter.
PAN User's Guide	Explains how to navigate PAN and the tasks related to the AASA test administration.
Arizona Accommodation Manual	Lists the current accommodations, accessibility features, and tools available on Arizona's achievement assessments.

4.3. Administration Training

Mandatory test administration training was provided by ADE and Pearson and delivered through Pearson's online Training Management System (TMS) that contained the training modules summarized in Table 4.3 that were required for DTCs, STCs, TAs, and other school staff involved in testing or test results.

The online training modules were available prior to the beginning of the testing window and throughout the testing window. The training modules addressed the specific responsibilities of the DTC and provided important information from the three documents TAs are required to use (i.e., the TAD, TCM, and *PAN User's Guide*). These training modules are updated for each test administration in correspondence with the updates to the required documents. Each module requires approximately 30–45 minutes to complete. DTCs are required to view the training modules in sequence and to successfully complete a final quiz after viewing all modules. DTCs must obtain a score of 80% or higher on the final quiz to be certified to access the secure test administration materials. DTCs are allowed multiple attempts to obtain a score of 80% or higher on the final quiz.

Table 4.3. Administration Trainings

Training	Description
AASA Training for Test Coordinators	This training covered the AASA test administration for Grades 3–8, including an overview of the test administration, websites and resources, and responsibilities before, during, and after testing.
Accommodations	This training covered the test accommodations. This was required for all DTCs but could be shared with staff members.
Achievement Test Administration Responsibilities	This training covered the test administration of AASA and AzSCI for all employees who administered, proctored, or were in contact with test materials. The purpose of this training was to provide guidance on consistent test administration across the state, increase the number of valid student tests, reduce test improprieties, and limit staff exposure to accusations of testing violations and discipline.
Test Security and Ethics	This training covered policies and practices to ensure the security and confidentiality of testing materials and the reliability and validity of test score interpretation. This training module was required for all employees who administered, proctored, or came in contact with testing materials.
PearsonAccess^{next} (PAN)	This training covered PAN and was required for DTCs, STCs, and other testing staff who assisted with registering students or managing test sessions in PAN.
Technology Training	This training outlined the critical steps necessary to prepare the network, testing devices, and other technology related items required for a successful test administration.

4.4. Sample Tests

In addition to the module training, TAs are instructed to become familiar with the online system by accessing sample items. Sample tests are available in TestNav year-round to help TAs and students become familiar with the AASA item types. The sample tests were created following Pearson's standard item and test development process, including item content and bias review by Arizona educators and community members. The sample tests reflect the AASA test specifications and blueprints and had 1–25 items on each test, as shown in Table 4.4. Because the sample tests do not include an item for each of the aligned Arizona Academic Standards and do not provide scores for students, they should NOT be used to evaluate a student's performance level. Students access the test as a guest, so no personal information needs to be provided.

There is a sample test for each grade and content area, and every eligible item type is represented, and an accompanying scoring guide identifies standard and DOK alignment. The portal and scoring guides are both available on ADE website at https://www.azed.gov/assessment/aasa. Scoring guides for the sample tests are also available.

Table 4.4. Number of Items on the AASA Sample Tests

Grade	ELA	Writing	ORF	Mathematics
3	24	1	3	25
4	24	1	–	25
5	24	1	–	25
6	24	1	–	25
7	24	1	–	25
8	24	1	–	25

4.5. Accommodations

Accommodations are specific practices and procedures that provide students with equitable access during the assessment. They are made to provide a student equal access to learning and equal opportunity to demonstrate what is known and are intended to reduce or even eliminate the effects of a student's disability. Accommodations can be changes in the presentation, response, setting, and timing/scheduling of educational activities. There should be a direct connection between a student's disability, special education need, or language need and the accommodation(s) provided to the student during educational activities, including assessment.

Students should receive the same accommodations for classroom instruction, classroom assessments, district assessments, and state assessments. No accommodations should be provided during assessments that are not also provided during instruction. However, not all accommodations appropriate for instruction are appropriate for use during a standardized state assessment. Table 4.5 presents the accommodations available to students while testing on Arizona assessments.

Table 4.5. Available Accommodations

Accommodation	Description
Abacus	Students may use an abacus without restrictions for any mathematics test (for students taking the Braille test only). Students may use an abacus without restrictions for any mathematics test or a talking calculator for students taking Unit 1 of the Grades 7 or 8 mathematics test.
Adult Scribe	A student who requires one-on-one adult assistance during daily instruction may orally dictate or use gestures to indicate a selected response for multiple-choice items only while an adult enters this in the test. The adult may not ask or answer any questions during the session or influence student responses in any way.
American Sign Language (ASL)	ASL requires the use of a different test form that must be indicated in PearsonAccess^{next} (PAN). The ASL test form must be requested using the Additional Accommodations online request form.
Braille test booklet	Braille tests must be requested using the special paper version (SPV) test online request form. Requires adult transcription: An adult must transfer the student's response exactly as written into the TestNav system.
Large print test booklet	Large Print tests must be requested using the special paper version (SPV) test online request form. The 504 plan or IEP must clearly state the font size used for instruction and the type of materials teachers enlarge for the student. Requires adult Transcription: An adult must transfer the student's response exactly as written into the TestNav system.

Accommodation	Description
Paper test booklet	A student who cannot access the computer for classroom work due to injury, illness, or vision impairments may need a paper test in lieu of taking the test with peers on the computer. Requires adult transcription: An adult must transfer the student's response exactly as written into the TestNav system.
Sign test content	Any student who requires signing of content during daily instruction may have any of the content of writing, mathematics, and science signed.
Simplified test administration directions	The test administrator may provide verbal directions in simplified English for the scripted directions from the *Test Administration Directions* manual. This must take place in a setting that does not disturb other students.
Translated test administration directions	Exact oral translation, in the student's native language, of the scripted directions from the *Test Administration Directions* manual are permitted. No test content or directions embedded within the test may be translated.
Translation dictionary	During testing, students may use the word-for-word published paper translation dictionary that is used regularly for classroom instruction. Students with a visual impairment may use an electronic dictionary with other features turned off.

4.6. Universal Test Administration Conditions

The following Universal Test Administration Conditions are testing situations and conditions that may be offered to any student to provide a comfortable and distraction-free testing environment. They do not require an accommodations request. While some of the items listed as Universal Test Administration Conditions might be included in an IEP or 504 plan as an accommodation, for achievement testing purposes these are not considered testing accommodations and are available to any student who needs them.

- Testing in a small group, 1:1, or in a separate location on campus or in a study carrel
- Being seated in a specific location within the testing room or at special furniture
- Having the test administered by a familiar test administrator
- Using a special pencil or pencil grip
- Using a placeholder
- Read-aloud (text-to-speech or human reader) content of the ELA writing, mathematics, and science assessments
- Using devices that allow the student to see the test: glasses, contacts, magnification, and special lighting
- Using different contrast settings or color overlays
- Using devices that allow the student to hear the test directions: hearing aids and amplification
- Wearing noise buffers after the scripted directions from the *Test Administration Directions* manual have been read
- Signing the scripted directions from the *Test Administration Directions* manual
- Repeating the scripted directions from the *Test Administration Directions* manual
- Having assistance with logging into an online test
- Reading the test quietly to themselves as long as other students are not disrupted
- A phone or electronic device needed for medical care is permitted. The phone needs to stay close to the Test Administrator or proctor as well as the student and should be monitored to assure the device is only being used for medical purposes during testing

- Individual students may take a stretch break (1 or 2 minutes) during the test session (students may not talk, use electronic devices, go to lunch, or leave the testing room)
    - Paper test booklet and scratch paper must be collected
    - Students must sign out of TestNav without submitting the test. The test administrator will need to resume the student's test session using PAN.
- Students may use the restroom (only one student at a time)
    - The TA must collect the student's paper test booklet and scratch paper.
    - Students must sign out of TestNav without submitting the test. The test administrator will need to resume the student's test session using PAN.
- The use of scratch paper (plain, lined, or graph; school provided). Scratch paper must be securely shredded at the conclusion of testing
- Each testing session must be completed in the same school day in which it was started. The AASA and AzSCI are untimed. Do not start a test unit unless there is sufficient time to complete the test in the same school day.
- Students cannot leave for lunch during a test session. Test units should be scheduled in a way that provides the student more than adequate time to complete the test.

4.7. Universal Test Tools

The Universal Test Tools provided in Table 4.6 are available to all students taking the AASA assessment and cannot be disabled.

Table 4.6. Universal Test Tools

Universal Test Tool	Description
Alternate Mouse Pointer	There are six alternate mouse pointers available for students in TestNav. Alternate options include a medium, large, or extra-large sized white pointer, and extra-large sized black, green, or yellow pointer.
Answer Masking	Allows student to electronically cover and reveal individual answer choices.
Answer Eliminator	Cross out answer options for multiple-choice and multi-select items.
Area Boundaries	Allows student to click anywhere on the selected response text or button for multiple choice items.
Bookmark for Review	Mark an item for review so that it can be easily found later.
Contrast	Allows the student to change the background and text color based on need or preference. The Contrast setting will not change images or artwork. The options are white background with black text; cream background with black text; light blue background with black text; black background with white text; light magenta background with black text; and blue background with yellow text.
Expand/Collapse Passage	Expand a passage for easier readability. Expanded passages can also be collapsed.
Highlighter	Highlight text in a passage or item.
Line Reader	An adjustable box allows the student to focus on one line or a few lines at a time. The box can be adjusted to increase or decrease the number of lines shown. The Line Reader and Magnifier tools may be used simultaneously.
Magnifier	Allows the student to make part of the screen larger. When in use, the magnifier can be moved around the screen as needed.
Pause and Restart	Students may sign out of TestNav. Before the student can resume testing, the Test Administrator will need to resume the student's session in TestNav.

Universal Test Tool	Description
Notes/Comments	Allows student to open an on-screen notepad and take notes or make comments. Notes carry over within a passage set. In non-passage items, notes are attached to the specific test item on which they are entered.
Review Test	Allows student to review the test before submitting it.
System Settings	Adjust audio (volume) during the test.
Text-to-Speech	Text-to-Speech for content of writing, mathematics, and science.
Tutorial	Learn and practice using TestNav tools and responding to each item type.
Writing Tools	Editing tools (cut, copy, and paste) and basic text formatting tools (bold, underline, and italic) for extended response items.
Zoom In/Zoom Out	Enlarge the font and images in the test up to 200%. Undo zoom in and return the font and images in the test to original size.

4.8. Pearson Customer Support

To provide support to schools before, during, and after testing, Pearson provides tiered technical support Monday – Friday from 7:00 a.m. to 7:00 p.m. CST. DTCs, STCs, and TAs can contact the customer support line with questions pertaining to the TestNav and PAN system and test administration procedures. The toll-free support number, e-mail address, and chat link are disseminated to the field through the AASA system and related communications.

4.9. Test Security

All test coordinators, test administrators, and proctors must be trained in proper test security procedures, must sign an Achievement Tests Staff Security Agreement form (as shown in Figure 4.1), and must adhere to test security procedures. Test materials should be secured prior to, and at the conclusion of, all testing sessions. Test Administrators and proctors may not assist students in answering test items and may not translate, reword, or explain any test content. No test content may ever be discussed before, during, or after test administration. It is unethical and shall be viewed as a violation of test security for any person to:

- Log into TestNav as a student unless assisting student with log in procedures
- Share their username/password for PAN
- Capture images of any part of the test via any electronic device
- Duplicate in any way any part of the test
- Examine, read, or review the content of any portion of the test
- Disclose, or allow to be disclosed, the content of any portion of the test before, during, or after test administration
- Discuss any test item before, during, or after test administration
- Allow students access to test content prior to testing
- Provide any reference sheets to students during the mathematics test administration or graphic organizers during the Writing test administration
- Allow students to share information during test administration
- Read any parts of the test to students, except as indicated in TAD or as part of an approved accommodation
- Influence students' responses by making any kind of gestures (e.g., pointing to items, holding up fingers to signify item numbers or answer options) while students are taking the test

- Instruct students to go back and reread/redo responses after they have finished their test since this instruction may only be given before the students take the test
- Review students' responses
- Change students' answer choices
- Read or review students' scratch paper
- Participate in, direct, aid, counsel, assist in, encourage, or fail to report any violations of these test administration security procedures

Figure 4.1. Test Security Agreement

ASSESSMENTS Achievement

Achievement Tests (AASA, AzSCI, ACT Aspire, and ACT)
School Year 2023–2024 Staff Test Security Agreement

I acknowledge that all Achievement Tests are secure tests and agree to the following conditions of use to ensure the security of the test. For this document, Achievement Tests refers to AASA, AzSCI, ACT Aspire, and ACT.

1. I shall take necessary precautions to safeguard test materials.
   a. I shall sign an *Achievement Tests Staff Security Agreement* for School Year 2023-2024.
   b. Access to test materials, including online tests, is restricted. I shall not attempt to gain access to test materials beyond that which is granted to me by my school/district test coordinator, superintendent, or charter representative.
   c. If test materials are distributed to me, I shall keep them under lock and key except during actual test times. This includes any student data sheets or student information sheets provided to me.
   d. I shall not permit students to remove test material from the testing room except under the supervision of staff.
   e. I shall not examine, read, or review the Achievement Tests.
      i. I shall not disclose, nor allow to be disclosed, the content of the test.
      ii. I shall not discuss any test item at any time.
      iii. I shall not examine, read, or review any student responses.
      iv. I shall not log into any student online test.
   f. I shall not erase or change any student responses or any marks (including stray marks) on a scorable test booklet or answer document.
   g. If test materials are distributed to me, I shall return all test materials to the school/district test coordinator immediately upon the completion of testing.
   h. I shall not use any test materials for instruction before or after test administration. I shall follow *Test Preparation and Administration Practices*, the guidelines approved by the State Board of Education in January 2003 and updated in December 2007.
   i. I shall not provide prohibited or inappropriate resources to students during testing, including but not limited to graphic organizers, reference sheets, and calculators, except for tests and test sections where calculators are allowed.
2. I understand that the district superintendent or charter representative will develop, distribute, and enforce disciplinary procedures for the violation of test security by staff.

Individuals who will administer or proctor Achievement Tests for school year 2023-2024 must also agree to the following conditions to ensure the correct administration of the tests.

3. I shall participate in training activities prior to administering the tests.
4. I shall review the appropriate Test Administration Directions prior to administering the test.
5. I shall follow all instructions in the appropriate Test Administration Directions including reading the directions to students exactly as scripted.

By signing my name to this document, I am assuring my district/charter and the Arizona Department of Education that I will abide by the above conditions and that anyone I supervise, who will have access to the Achievement Tests, will also sign a Test Security Agreement.

Signed By:_____ Date: _____

Printed Name: _____

Title:_____ School: _____

Please return signed copy as per instructions from your school/district test coordinator.
Signed copies will be maintained by school/district administrators for 6 years.

In addition to test security procedures required of all educators involved in the testing process, TestNav has built-in security features for the test content and personal data that relies on multiple levels of protection, including restricted user access, encryption of data in transit and at rest, systems monitoring for abnormal behavior, application, server, and network security testing, and qualified, verified and trusted support personnel.

Pearson uses Advanced Encryption Standard (AES) encryption for data at rest and Hypertext Transfer Protocol Secure (HTTPS) to provide encryption and data-in-motion security for online testing by creating a secure channel on the network with the Secure Socket Layer (SSL) /Transport Layer Security (TLS) protocols. Test content can only be viewed through a valid test registration and login, all of which are logged within the platform's audit trail system and cannot be deleted.

TestNav also locks down the student's desktop during testing to prevent students from accessing outside resources that could be used for cheating, such as email, instant messaging, or internet browsing. TestNav will stop students' tests if another background application attempts to interfere with or take "focus" away from the secure testing environment. These types of interruption cannot be blocked during testing and therefore could present additional opportunities for students to access unauthorized resources. However, TestNav also has a blocklist feature that prevents students from starting their test if certain applications that pose a threat to disrupt testing are running at the time TestNav is launched. In these situations, the student and/or proctor are prompted to shut down the offending application before attempting to start TestNav again.

Chapter 5: SCORING AND REPORTING

This chapter describes the human-scoring procedures used by the Pearson Performance Scoring Center (PSC) to score the AASA writing, reading, and mathematics open-ended items, as well as the automated scoring procedures for the writing prompts. This section addresses Standards 2.7, 4.18, 4.19, 4.20, 6.8, and 6.9 (AERA et al., 2014) regarding the scoring of the assessments.

The AASA machine-scored items were scored with maximum likelihood estimation (MLE) scoring, with an attemptedness rule that a student needed to answer one item in each operational unit. Both ELA and mathematics have their own scale score ranges. Students received a scale score in each content area, and student performance was reported as one of four performance levels: Level 1: *Minimally Proficient*, Level 2: *Partially Proficient*, Level 3: *Proficient*, and Level 4: *Highly Proficient*.

Student performance on reporting categories is reported as one of three levels of mastery: *Below Mastery*, *At/Near Mastery*, or *Above Mastery*. Students who score *Below Mastery* demonstrate performance in the reporting category that was clearly below *Proficient*. Students who score *At/Near Mastery* demonstrate performance in the reporting category that was exactly at or immediately above/below *Proficient*. Students who score *Above Mastery* demonstrate performance in the reporting category that was clearly *Proficient* or higher.

5.1. Human Scoring of Open-Ended Items

The AASA assessments contain open-ended items that prompt students to write a short answer or extended response (i.e., a paragraph or multi-paragraph essay) that require scoring by professionally trained scorers. These items were the writing prompts on the ELA Writing test, short constructed-response items on the ELA Reading test, and the paper-equivalent of the technology-enhanced (TE) items on the ELA Reading and mathematics assessments. Writing was scored via a distributed scoring model (i.e., scorers were trained in a self-paced model), whereas Reading and mathematics were scored using a synchronous model (i.e., scorers were trained by instructors). Human scoring was conducted in Pearson's scoring platform known as OSCAR (Online Scoring and Reporting).

5.1.1. Scorer Recruitment

Scorers are recruited by Pearson, with scorers who have extensive experience scoring this type of rubric on previous projects being given priority. Scorers receive performance ratings based on internal quality metrics of inter-rater reliability and validity. Those who have achieved a high-performance rating on previous writing, reading, and mathematics responses are recruited for the AASA assessment. Upon being hired, scorers sign a confidentiality agreement in which they pledge to keep all information and student responses confidential.

Scoring supervisors are chosen based on demonstrated expertise in the scoring process, including strong organizational abilities and training, practical skills, leadership abilities, and sensitivity to interpersonal communication requirements. Supervisors also possess the essential capability of helping scorers understand the AASA scoring requirements. Supervisors provide continuous feedback to the scorers through the validity and calibration process and monitor the quality of their assigned scorers. All scoring, including the scorers and supervisors, is supervised by a content specialist who is responsible for training and leading the entirety of the project.

5.1.2. Training

Scorers and scoring supervisors were trained to learn the rubric and score responses according to the AASA scoring guidelines. At the beginning of the scoring project, all scoring supervisors and scorers completed project-specific training consisting of a review of the rubric and prompts for the items being scored and a review of the anchor responses selected and approved by ADE for each prompt. Training for the ELA Writing prompts differed than the training for the Reading and mathematics open-ended items. Writing established training materials that could be inserted into modules for self-paced training, whereas training materials for Reading and mathematics were created as the students completed testing. This could be accomplished because the Reading and mathematics open-ended items were only 0,1 score point items.

5.1.2.1. Writing

The training for ELA Writing was conducted in a distributed environment using online modules designed to take scorers through the background of the assessment and the rubric and anchor sets for each item. A module is an online set of training materials that can be delivered to scorers individually at their own pace. These modules are embedded into the OSCAR system and are set up so as not to allow scorers to advance in their training until all proceeding modules are complete and correct.

Scoring supervisors and scorers were both required to take one set of practice papers and two sets of qualification papers once they completed the item-specific modules. They must have passed one of the two qualification sets for the items they were assigned before they could score on the project based on the criteria in Table 5.1. Their scores were compared to the "true score" approved by ADE for each training response. Once the scorer completed the item-specific training and had qualified, they were allowed to score live responses for that item or set of items. Different scoring rubrics are used for the different item types and are posted on the ADE website at https://www.azed.gov/assessment/aasa.

Table 5.1. Scoring Qualification Standards

Reporting Category	Score Points	Qualification %Perfect/Adjacent Agreement	#Sets
Writing Multi-trait	1–4	60/90 for each trait at least once across the two sets	2

5.1.2.2. Mathematics and Reading

Prior to scorer training for reading and mathematics, scoring directors reviewed items/passages and rubrics and selected actual student responses to review and discuss at range finding sessions with ADE staff. The range finding sessions allowed Pearson and ADE to discuss any questions regarding possible correct answers and assign final scores for the student responses. These scored student responses from the range finding sessions used to create an anchor and practice set for reading and "prototype" items for mathematics used as initial training items for an item type that included an anchor set and practice set. The sets were shared with ADE and adjusted as needed for final approval.

Training was conducted in the train-score-train-score model live via online conferencing where scoring directors trained scorers on the content for a single item and worked with the team to score that item before moving to train the second item. There were two separate ELA teams and two separate mathematics teams, each led by a scoring director. Mathematics scoring directors began content training on a prototype item, reviewing the prompt, rubric, and the anchor set for the item. The team then took and discussed a practice set to test their knowledge of rubric application before moving into live scoring. Subsequent similar items were trained with bridge sets. For such items, the scoring director would prepare the team by covering the prompt, rubric, and bridge set. Reading scoring directors began content training on every item, reviewing the prompt and passages, the rubric, and anchor set. The team then took and discussed a practice set to test their knowledge of rubric application before moving into live scoring.

5.1.3. Quality Control

A variety of reports are produced throughout the scoring process to monitor the progress of the project, the reliability of scores assigned, and individual scorers' work:

- Daily and Cumulative Interrater Reliability Reports by item and scorer that indicate how many times scorers were in exact agreement or assigned adjacent scores. The reliability is computed and is monitored daily and cumulatively for the project.
- Daily and Cumulative Validity Reports by item and scorer that indicate how many times scorers were in exact agreement or assigned adjacent scores to responses deemed True Scores. The validity is computed and monitored daily and cumulatively for the project.
- Daily and Cumulative Frequency Distributions that show how many times each score point has been assigned to the item being scored. The frequency distributions are produced daily and cumulatively for the entire scoring project. This report allows scoring supervisors and directors to see whether scorers tend to score consistently high or low.

The most immediate method of monitoring a scorer's performance is through backreading by scoring supervisors and directors. If a scoring supervisor discovers that a scorer is consistently assigning scores other than those the scoring supervisor would assign, they can send a message to that scorer using the backreading function and through the OSCAR instant messaging system.

With the help of the individual scorer reliability metrics and through backreading, the scoring staff can closely monitor each scorer's performance. Scorers are also monitored using the scorer exception process for validity and scoring rate. A scorer must meet and maintain the quality metrics established for AASA in the designated area to continue scoring the project. If a scorer fails to maintain the established validity perfect agreement and perfect plus adjacent agreement percentage, they will receive a targeted calibration set consisting of 10 anchor-type responses similar to a qualification set. If the scorer fails to pass the calibration set, they will be locked out of scoring and dismissed from the project.

Scorers with low inter-rater reliability or a lower- or higher-than-desired scoring rate are closely monitored in backreading and through reports. If, in the opinion of the scoring director and content specialist, these scorers are still performing below acceptable standards after receiving sufficient feedback and being given every reasonable opportunity to improve, they are manually locked out of the system and dismissed from the project.

5.1.4. Security

To ensure that test security is never compromised, the following safeguards are employed:

- Scorers and scoring staff personnel must sign a non-disclosure and confidentiality form in which they agree not to use or divulge any information concerning the tests.
- All contact with the press is handled through ADE.
- OSCAR is accessed via a secure website with login credentials required for each user. Only Pearson project support staff can issue user IDs to scorers to access OSCAR.

5.2. Automated Scoring of ELA Writing Prompts

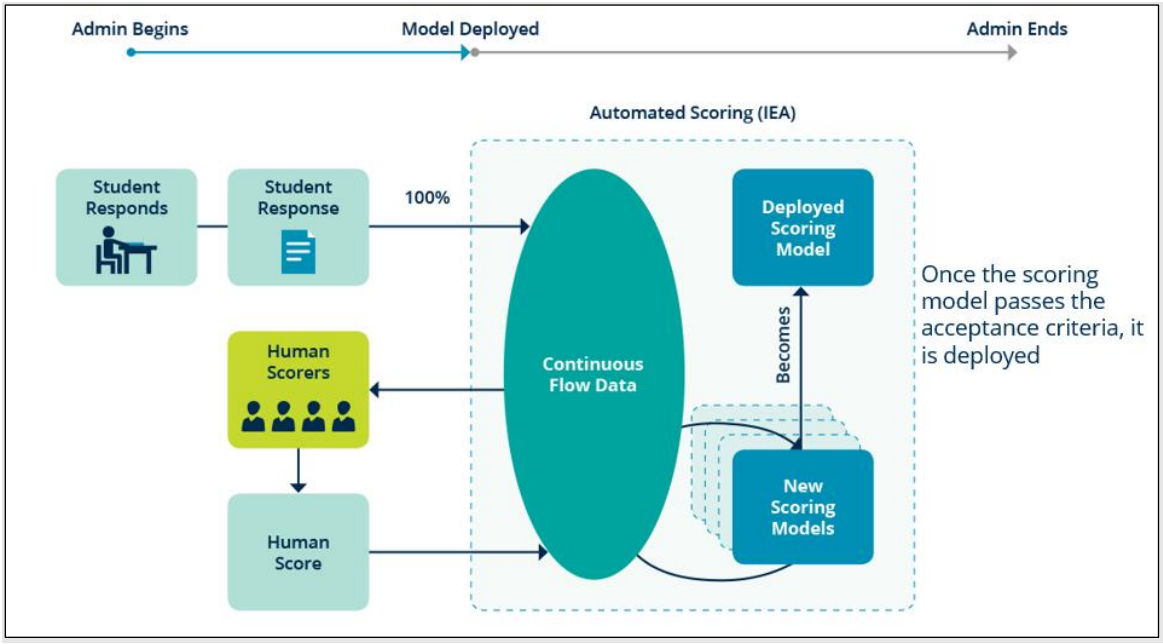
Pearson's automated scoring engine, the Intelligent Essay Assessor (IEA), is the default option for scoring the AASA ELA writing prompts. For the operational writing prompts in Spring 2024, the automated scoring engine was calibrated based on previously tested and human-scored field test responses. During the scoring window, human-scored student responses were used to continue improving and validating the scoring models via Continuous Flow. All the ELA prompts were scored at least in part by IEA in the spring. For 10% of responses, a second reliability score was assigned by human scorers to provide data for evaluating the consistency of scoring, which is done by evaluating scoring agreement.

IEA is trained by humans anytime a new writing prompt is introduced and follows the Continuous Flow process that incorporates human scoring to ensure the highest-quality scores. Responses flow between the engine and human scorers so the engine can learn from humans in real time and challenging responses can be instantly routed to human scorers (known as Smart Routing). When the engine is less confident in scoring a response, the response is marked with a low confidence flag that automatically routes it to human scorers. Human scoring is applied to responses that are scored while IEA is being trained, as well as to the Smart Routing responses. When multiple scores are assigned for a given response, the IEA score is reported operationally if it is a high confidence score. If the IEA score is low confidence, the human score is assigned.

5.2.1. Calibration of IEA

With Continuous Flow, human scorers begin the scoring process and IEA learns from them. This process can begin with previously tested and human scored field test items or during the operational scoring window. For the writing prompts, IEA used a combination of human-scored field test and operational data to calibrate the automated scoring engine. The field test data were used to build initial models. Some prompts had enough data to build a full scoring model that passed the criteria described in Section 5.2.3, while other prompts required additional data to meet the criteria using the Continuous Flow process to supplement the models with human-scored operation data until it met all the quality criteria. Figure 5.1 presents scoring model development and deployment in the Continuous Flow scoring approach.

Figure 5.1. Dynamic Model Development and Deployment



The early performance of human scoring was monitored based on the following characteristics to verify that an appropriate set of data was available for training IEA:

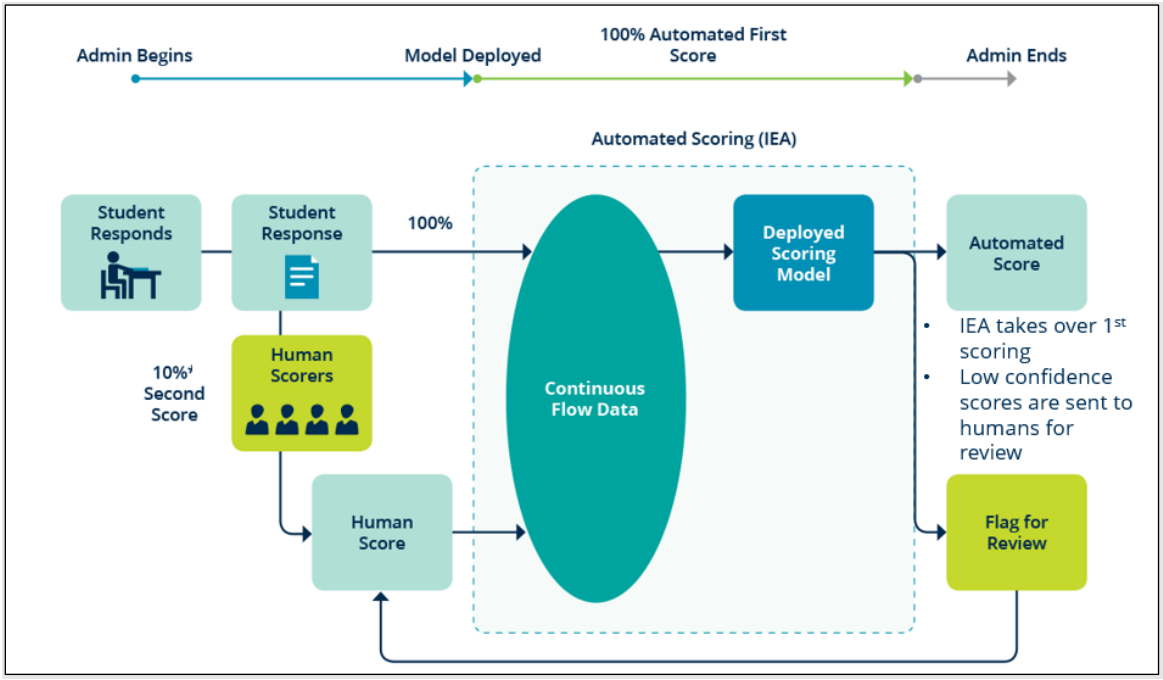
- Exact agreement between human scorers (with the goal of at least 65%)
- Exact agreement between human scores conditioned on score point (with the goal of at least 50%)
- The number of responses at each score point
- The number of responses with two human scores assigned (IEA via Continuous Flow "ordered" additional scoring of responses during the sampling period as needed)

Although the desired characteristics of the training data were easily achieved for some prompts, they were more challenging to achieve for others. For some prompts, a subset of scores were reset and clarifying directions were provided to scorers to improve human-human agreement. A healthy percentage of responses were also backread during the sampling period. These scores in addition to the double human scores were all part of the data used to train IEA.

5.2.2. Smart Routing

As illustrated in Figure 5.2, once IEA is trained, it takes over first scoring with human scorers providing the 10% second score for reliability. Smart Routing refers to the practice of using automated scoring results to detect responses that are likely to be challenging to score and applying automated routing rules to obtain one or more additional human scores on those responses. Smart Routing can be applied prompt-by-prompt to the extent needed to meet scoring quality criteria for automated scoring. When the engine is less confident in scoring a response, the response is marked with a low confidence flag that automatically routes it for human scorers.

Figure 5.2. Smart Routing



5.2.3. Quality Control

IEA performance on the writing prompts was evaluated based on IEA-human exact agreement and compared to agreement based on responses that were double-scored by humans. The following industry-standard measures were computed between pairs of human scores and between IEA and humans to evaluate scoring performance:

- Pearson correlation between IEA-human should be at least 0.70 and within 0.1 of human-human.
- Quadratic-weighted kappa between IEA-human should be at least 0.70 and within 0.1 of human-human.
- Standardized mean difference between IEA-human should be less than |0.15|.
- With Smart Routing applied as needed, exact agreement between IEA-human should meet the inter-rater reliability requirement of at least 65% and be within 5.25% of human-human exact agreement. If the IEA-human agreement is within 5.25% of the human-human agreement, IEA can be deployed operationally. This is the primary criterion for evaluating IEA.

In addition to the overall comparison, the following performance thresholds were targeted in the test data set: (1) at least 65% overall IEA-human agreement and (2) 50% IEA-human agreement by score point (conditioned on the human score).

5.3. Reporting

The following AASA reports were available in PAN at https://az.pearsonaccessnext.com. PDF versions of the reports and district-wide electronic student data files were also available for downloading. District-level user roles provided access to all school-level reports and district-level reports, including all Confidential Student Score Reports for students who tested in the district. School-level user roles provided access to all school-level reports and all Confidential Student Score Reports for students who tested in the school. A Family Guide for interpreting reports was also available for download. Figure 5.3 and Figure 5.4 present sample reports.

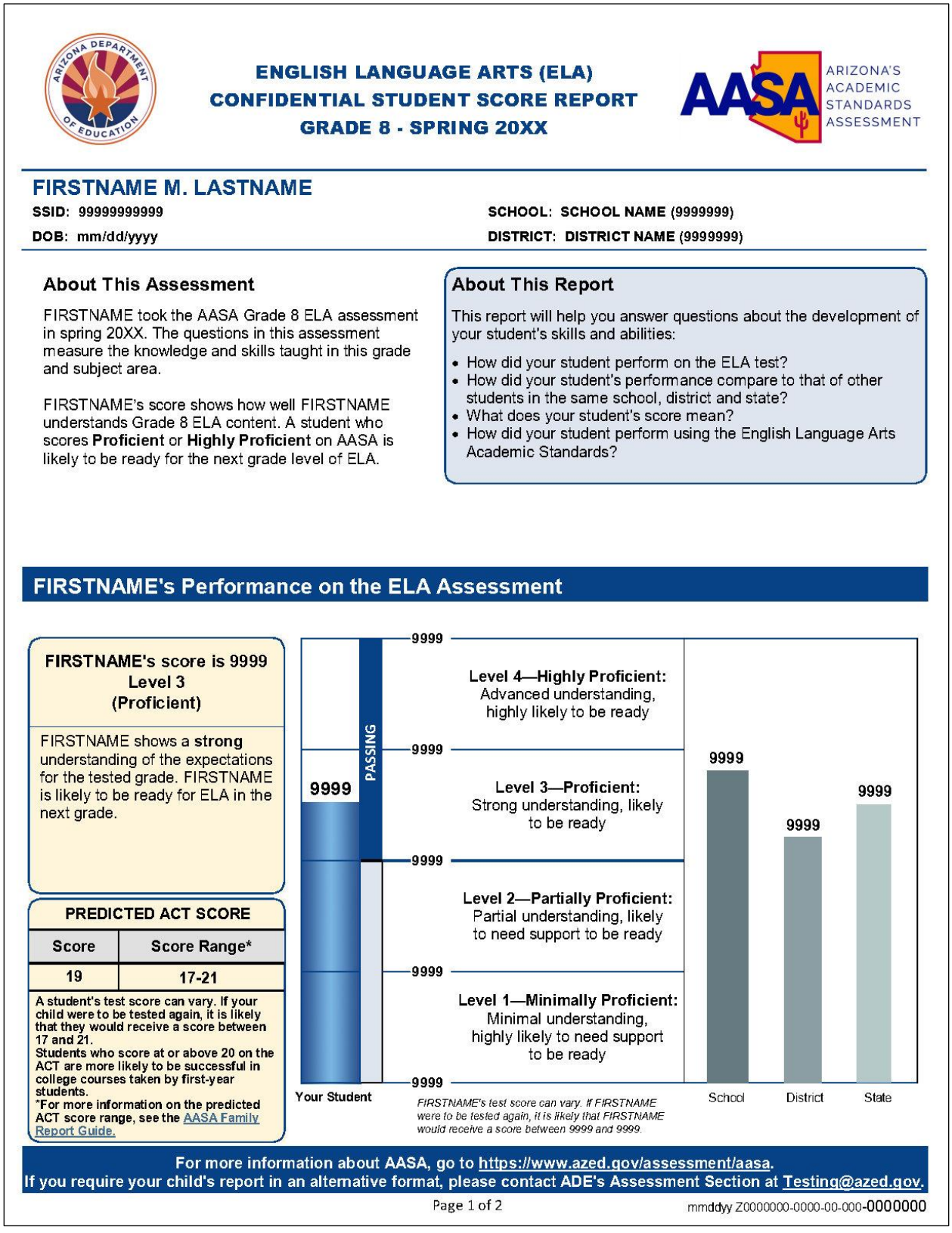
- District-level
    - District Confidential Roster Report with Summary (district-level, student roster by grade and content area
    - Student Data File
- School-level
    - Confidential Student Score Report (individual student report)
    - Informe del Estudiante (individual student report in Spanish)
    - Confidential Roster Report with Summary (school-level, student roster by grade and content area

AASA reports have been designed with the user's comprehension in mind. The goal of these reports is to deliver accurate assessment data and ensure that it is correctly interpreted and understood. Similar colors are used for groups of similar elements, such as performance levels, throughout the design to guide the user to compare like elements and avoid comparison of dissimilar elements. All score report data are based on the total number of students whose tests have been scored. All score report data in PAN, except for individual students' score reports, can be disaggregated into testing groups if they were set up by the school during the specified timeframe.

The Confidential Student Score Report (individual student report) includes the average scale scores for the school, district, and state to allow for visual comparison. Two copies of the printed Confidential Student Score Report and Family Report Guide were also provided. Printed reports are packed by the school and shipped to participating districts. An ACT predicted score is also included for Grade 8 students on the Confidential Student Score Reports. Students who score at or above the ACT score are more likely to be successful in college courses taken by first-year students.

The AASA score reports are also available in the Parent Portal, which is an optional resource for schools and districts to use that allows families to securely access and view their student's online individual student report. After creating a user account, families enter the student's information, including the student's claim code, to retrieve the AASA Student Report. The claim codes file (in CSV format) is available for request in PAN for authorized district and school users. The *Parent Portal Access Guide* is also available to families and includes the steps that should be followed to access their student's information on the Parent Portal.

Figure 5.3. Sample Reports—Confidential Student Score Report



ENGLISH LANGUAGE ARTS (ELA)
CONFIDENTIAL STUDENT SCORE REPORT
GRADE 8 - SPRING 20XX

ARIZONA'S ACADEMIC STANDARDS ASSESSMENT

FIRSTNAME M. LASTNAME

SSID: 99999999999
DOB: mm/dd/yyyy
SCHOOL: SCHOOL NAME (9999999)
DISTRICT: DISTRICT NAME (9999999)

About This Assessment

FIRSTNAME took the AASA Grade 8 ELA assessment in spring 20XX. The questions in this assessment measure the knowledge and skills taught in this grade and subject area.

FIRSTNAME's score shows how well FIRSTNAME understands Grade 8 ELA content. A student who scores **Proficient** or **Highly Proficient** on AASA is likely to be ready for the next grade level of ELA.

About This Report

This report will help you answer questions about the development of your student's skills and abilities:

- How did your student perform on the ELA test?
- How did your student's performance compare to that of other students in the same school, district and state?
- What does your student's score mean?
- How did your student perform using the English Language Arts Academic Standards?

FIRSTNAME's Performance on the ELA Assessment

FIRSTNAME's score is 9999
Level 3
(Proficient)

FIRSTNAME shows a **strong** understanding of the expectations for the tested grade. FIRSTNAME is likely to be ready for ELA in the next grade.

PREDICTED ACT SCORE	
Score	Score Range*
19	17-21

A student's test score can vary. If your child were to be tested again, it is likely that they would receive a score between 17 and 21.
Students who score at or above 20 on the ACT are more likely to be successful in college courses taken by first-year students.
*For more information on the predicted ACT score range, see the AASA Family Report Guide.

9999
PASSING
Level 4—Highly Proficient: Advanced understanding, highly likely to be ready
Level 3—Proficient: Strong understanding, likely to be ready
Level 2—Partially Proficient: Partial understanding, likely to need support to be ready
Level 1—Minimally Proficient: Minimal understanding, highly likely to need support to be ready
Your Student
School
District
State
FIRSTNAME's test score can vary. If FIRSTNAME were to be tested again, it is likely that FIRSTNAME would receive a score between 9999 and 9999.

For more information about AASA, go to https://www.azed.gov/assessment/aasa.
If you require your child's report in an alternative format, please contact ADE's Assessment Section at Testing@azed.gov.

Page 1 of 2

mmddyy Z0000000-0000-00-000-0000000

GRADE 8 **FIRSTNAME's ELA CONFIDENTIAL SCORE REPORT - page 2**

Legend: Reporting Categories
= Below Mastery = At/Near Mastery = Above Mastery

ELA Reporting Categories

Reading for Information

FIRSTNAME performed **above mastery** in Reading for Information.

What was assessed?
Students explain how reasoning and evidence shape and support the main idea of a text. They examine how a text makes connections between different individuals, ideas, or events. They show how an author of a text responds to evidence that does not support his or her point of view.

What do these results mean?
Your student gives an objective summary of a text; uses evidence from a text to make and support conclusions; explains how an author addresses a conflicting viewpoint; determines if information in a text is needed; evaluates how presentation (like text or audio) affects information.

Reading for Literature

FIRSTNAME performed **at or near mastery** in Reading for Literature.

What was assessed?
Students find the main idea of a text and examine how it is developed. They determine how specific words and phrases can change the meaning and tone of a text. They analyze how a character's point of view affects a text. They recognize the influence of other literature on a text.

What do these results mean?
Your student often uses supporting details to explain the theme or main idea; shows how a story moves forward; describes the effect of point of view on a text; recognizes the influences of other literature on a text; compares the structure of two or more texts.

Writing and Language

FIRSTNAME performed **below mastery** in Writing and Language.

What was assessed?
Students write to inform or make an argument. They use evidence and clear reasoning to support their writing. Their evidence comes from many different sources. They determine the meaning of new words and figurative language. They spell correctly and use correct grammar.

What do these results mean?
Your student may have trouble stating a claim clearly and providing supporting details to make an argument when writing; using citations correctly when doing research; using verb tenses and punctuation correctly; using other words or word parts to figure out the meaning of new words.

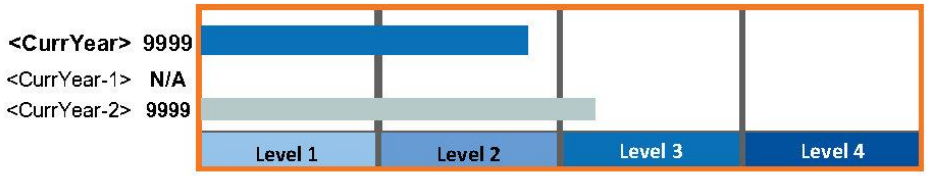
The Writing and Language portion of the ELA assessment requires that each student complete an essay. The essay is evaluated on three criteria.

Writing Essay Performance

Statement of Purpose, Focus & Organization	Evidence & Elaboration	Conventions & Editing
Your student earned 3 out of 4 possible points. In general, your student's essay stays on topic and is focused. The main idea of the topic is given context and addresses the audience and purpose for writing. The response is organized and develops connections between ideas. It uses transitions and has an introduction and conclusion.	Your student earned 2 out of 4 possible points. Your student's essay includes some support or evidence for the main idea. It uses some facts and details from other sources but does not use citations regularly. The response does not expand on ideas or make clear connections between ideas. It uses simple and sometimes inappropriate language for the audience and purpose.	Your student earned 2 out of 2 possible points. Your student's essay shows an understanding of sentence formation and other conventions. The response may have some mistakes, but they are not repeated often in the text. It uses correct punctuation, capitalization, and spelling.

FIRSTNAME's ELA Assessment Progress

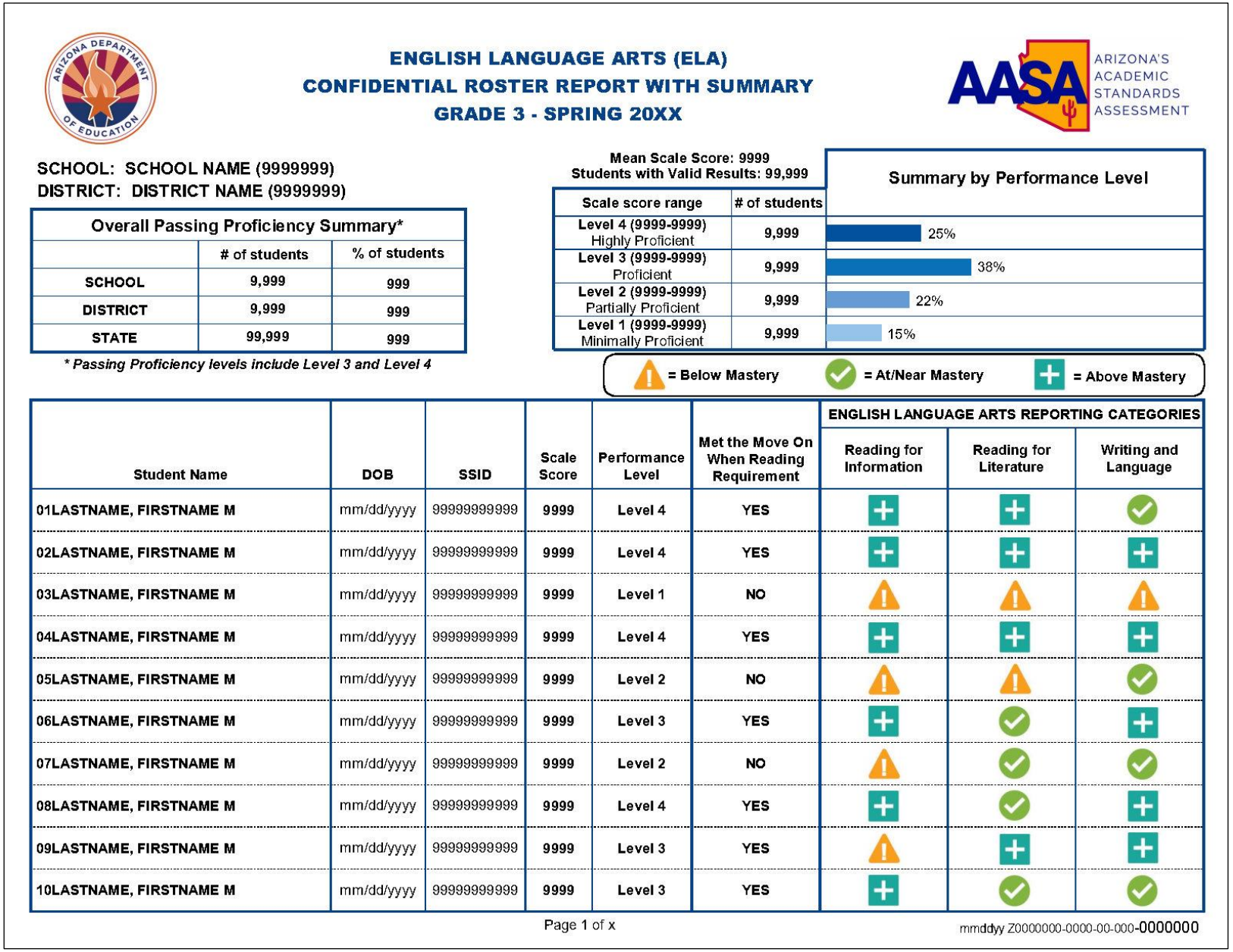
This chart displays your student's performance in ELA assessments over time. It reports the proficiency level for the most recently completed tests in ELA (if available). You can use this information to determine your student's progress in ELA.



For more information about AASA, go to https://www.azed.gov/assessment/aasa.
If you require your child's report in an alternative format, please contact ADE's Assessment Section at Testing@azed.gov.

Page 2 of 2

Figure 5.4. Sample Reports—Confidential Roster Report with Summary



ENGLISH LANGUAGE ARTS (ELA)
CONFIDENTIAL ROSTER REPORT WITH SUMMARY
GRADE 3 - SPRING 20XX

ARIZONA'S ACADEMIC STANDARDS ASSESSMENT

SCHOOL: SCHOOL NAME (9999999)
DISTRICT: DISTRICT NAME (9999999)

Overall Passing Proficiency Summary*		
	# of students	% of students
SCHOOL	9,999	999
DISTRICT	9,999	999
STATE	99,999	999

** Passing Proficiency levels include Level 3 and Level 4*

Mean Scale Score: 9999
Students with Valid Results: 99,999

Scale score range	# of students
Level 4 (9999-9999) Highly Proficient	9,999
Level 3 (9999-9999) Proficient	9,999
Level 2 (9999-9999) Partially Proficient	9,999
Level 1 (9999-9999) Minimally Proficient	9,999

Summary by Performance Level

= Below Mastery = At/Near Mastery = Above Mastery

Student Name	DOB	SSID	Scale Score	Performance Level	Met the Move On When Reading Requirement	ENGLISH LANGUAGE ARTS REPORTING CATEGORIES		
						Reading for Information	Reading for Literature	Writing and Language
01LASTNAME, FIRSTNAME M	mm/dd/yyyy	99999999999	9999	Level 4	YES	+	+	✓
02LASTNAME, FIRSTNAME M	mm/dd/yyyy	99999999999	9999	Level 4	YES	+	+	+
03LASTNAME, FIRSTNAME M	mm/dd/yyyy	99999999999	9999	Level 1	NO	!	!	!
04LASTNAME, FIRSTNAME M	mm/dd/yyyy	99999999999	9999	Level 4	YES	+	+	+
05LASTNAME, FIRSTNAME M	mm/dd/yyyy	99999999999	9999	Level 2	NO	!	!	✓
06LASTNAME, FIRSTNAME M	mm/dd/yyyy	99999999999	9999	Level 3	YES	+	✓	+
07LASTNAME, FIRSTNAME M	mm/dd/yyyy	99999999999	9999	Level 2	NO	!	✓	✓
08LASTNAME, FIRSTNAME M	mm/dd/yyyy	99999999999	9999	Level 4	YES	+	✓	+
09LASTNAME, FIRSTNAME M	mm/dd/yyyy	99999999999	9999	Level 3	YES	!	+	+
10LASTNAME, FIRSTNAME M	mm/dd/yyyy	99999999999	9999	Level 3	YES	+	✓	✓

Page 1 of x

mmddyy Z0000000-0000-00-000-0000000

Chapter 6: CLASSICAL ITEM ANALYSIS

This chapter presents classical statistics for the data used for calibration, equating, and scaling of the Spring 2024 AASA assessments as indicated by Standards 1.8, 1.10, 2.5, 2.19, 3.6, 4.14, and 7.4 (AERA et al., 2014). Each grade in ELA had two core online forms with different embedded field test sets for a total of 18 field test forms. The core online forms differed by only a writing prompt. Nine forms had one writing prompt (referred to as Form 1), and the other nine forms had another writing prompt (referred to as Form 2). Where appropriate, statistics are reported for both ELA core online forms. Mathematics only had one core online form with different embedded field test sets for each grade, with 11 online forms total.

6.1. Data

The classical item analysis was conducted based on the calibration samples described in Section 7.1. Table 6.1 and Table 6.2 present demographic information of the students included in the calibration sample by gender, ethnicity (Hispanic or Non-Hispanic), race, and special education, English learner (EL), and low socioeconomic status (SES). Because only a few students took the accommodated forms, these students were not included in the item analysis. Students who did not complete the test were also excluded.

Table 6.1. Number of Students in the Calibration Sample by Subgroup—ELA

Subgroup	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
All	74,077	77,076	77,580	76,717	77,917	79,101
Male	37,421	39,066	39,359	38,985	39,553	40,330
Female	36,656	38,010	38,221	37,732	38,364	38,771
Hispanic	36,763	38,259	38,000	37,323	37,960	38,743
Non-Hispanic	37,314	38,817	39,580	39,394	39,957	40,358
American Indian	3,928	4,143	4,243	4,184	4,529	4,465
Asian	2,124	2,206	2,268	2,307	2,307	2,251
Black or African American	5,494	5,646	5,846	5,592	5,672	5,812
Multi-racial	4,855	4,873	4,856	4,615	4,615	4,529
Native Hawaiian or Other Pacific Islander	434	450	436	430	476	446
White	56,424	58,937	59,036	58,329	58,912	59,925
Missing	818	821	895	1,260	1,406	1,673
Special Education	10,776	11,480	11,160	10,240	9,742	9,485
English Learner (EL)	10,571	8,312	8,583	7,496	6,938	6,717
Low Socioeconomic Status (SES)	38,549	39,580	39,755	38,477	38,081	37,686

Table 6.2. Number of Students in the Calibration Sample by Subgroup—Mathematics

Subgroup	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
All	75,267	77,793	78,156	77,439	78,660	79,736
Male	38,199	39,516	39,687	39,422	39,985	40,684
Female	37,068	38,277	38,469	38,017	38,675	39,052
Hispanic	37,441	38,673	38,345	37,745	38,370	39,085
Non-Hispanic	37,826	39,120	39,811	39,694	40,290	40,651
American Indian	4,027	4,188	4,280	4,228	4,592	4,501
Asian	2,152	2,225	2,284	2,331	2,322	2,263
Black or African American	5,620	5,724	5,911	5,661	5,745	5,857
Multi-racial	4,928	4,912	4,883	4,635	4,640	4,544
Native Hawaiian or Other Pacific Islander	445	454	439	429	476	450
White	57,192	59,402	59,401	58,831	59,409	60,394
Missing	903	888	958	1,324	1,476	1,727
Special Education	11,154	11,707	11,295	10,384	9,870	9,581
English Learner (EL)	10,823	8,447	8,699	7,604	7,064	6,806
Low Socioeconomic Status (SES)	39,187	39,930	39,993	38,798	38,390	37,941

6.2. Descriptive Statistics

Table 6.3 presents the descriptive statistics on total raw scores for the spring AASA assessment, including the number of students included in the classical analysis, the number of operational items on the assessment, the maximum possible raw score, the mean raw score, the standard deviation (SD) of the raw score, and the minimum/maximum obtained raw score.

Table 6.3. Classical Test Analysis Statistics

Assessment	#Students	#Items	Max. Possible Raw Score	Mean Raw Score	SD Raw Score	Min. Raw Score	Max. Raw Score
ELA 3, Form 1	37,077	44	55	26.64	10.93	1	55
ELA 4, Form 1	38,497	44	56	29.33	11.70	3	56
ELA 5, Form 1	38,817	44	55	29.41	11.19	2	55
ELA 6, Form 1	38,354	44	55	29.71	11.19	3	55
ELA 7, Form 1	39,153	44	55	30.11	11.47	4	55
ELA 8, Form 1	39,314	44	55	29.62	10.57	4	55
ELA 3, Form 2	37,000	44	55	27.18	11.18	2	55
ELA 4, Form 2	38,579	44	56	29.28	11.47	2	56
ELA 5, Form 2	38,763	44	55	29.47	11.05	4	55
ELA 6, Form 2	38,363	44	55	29.79	11.25	3	55
ELA 7, Form 2	38,764	44	55	30.06	11.24	2	55
ELA 8, Form 2	39,787	44	55	29.31	10.66	4	55
Mathematics 3	75,267	45	45	25.35	11.17	0	45
Mathematics 4	77,793	45	45	23.21	11.08	0	45
Mathematics 5	78,156	45	45	20.81	10.27	0	45
Mathematics 6	77,439	47	47	19.73	11.24	0	47
Mathematics 7	78,660	47	47	20.33	11.13	1	47
Mathematics 8	79,736	47	47	18.93	10.10	0	47

6.3. Classical Item Analysis

Classical item analysis was conducted to show how the items performed for each grade-level assessment. Item difficulty is measured by the *p*-value bounded by 0.0 and 1.0 that indicates how easy or hard an item is for students. The *p*-value for 1-point items is based on the proportion of students who answered an item correctly and is derived by dividing the number of students who got the item correct by the total number of students who answered it. For multiple-point items, the *p*-value is the average item score (i.e., the sum of student scores on an item divided by the total number of students who responded to the item) divided by the number of possible score points on the item. A high *p*-value indicates that an item is easy (high proportion of students answered it correctly), whereas a low *p*-value indicates that an item is difficult. For example, a *p*-value of 0.79 indicates that 79% of students answered the item correctly. Easy and hard items are both necessary to include on an assessment to balance the test difficulty. The AASA assessment targets *p*-values in the range of 0.20 to 0.90.

Item discrimination is represented by the point-biserial correlation bounded by -1.0 and 1.0 that indicates how well an item discriminates, or distinguishes, between low-performing and high-performing students. The point-biserial correlation is based on the relationship between student performance on a specific item and performance on the entire test based on their test score. Students who do well on a test are expected to select the right answer to any given item, and students who do poorly are expected to select the wrong answer. This means that for a highly discriminating item, students who get the item correct will have a higher average test score than students who get the item incorrect. An item with a high positive point-biserial correlation discriminates between low-performing and high-performing students better than an item with a point-biserial correlation near zero. A negative point-biserial correlation indicates that lower-performing students did better on that item than higher-performing students. The AASA assessment targets point-biserial correlations of 0.25 or higher.

Table 6.4 presents a summary of the classical item analysis, and Appendix A presents the statistics for each item. If the classical item statistics for the operational items were outside of the item selection criteria presented in Table 3.3, the items will be reviewed during test construction of the next testing cycle for possible replacement in future administrations.

Table 6.4. Classical Item Analysis Summary

Assessment	#Items	Mean *P*-Value	Mean Point-Biserial
ELA 3, Form 1	44	0.48	0.48
ELA 4, Form 1	44	0.52	0.50
ELA 5, Form 1	44	0.53	0.48
ELA 6, Form 1	44	0.53	0.48
ELA 7, Form 1	44	0.54	0.48
ELA 8, Form 1	44	0.52	0.45
ELA 3, Form 2	44	0.48	0.48
ELA 4, Form 2	44	0.52	0.50
ELA 5, Form 2	44	0.53	0.48
ELA 6, Form 2	44	0.53	0.48
ELA 7, Form 2	44	0.54	0.48
ELA 8, Form 2	44	0.52	0.45

Assessment	#Items	Mean *P*-Value	Mean Point-Biserial
Mathematics 3	45	0.56	0.54
Mathematics 4	45	0.52	0.53
Mathematics 5	45	0.46	0.48
Mathematics 6	47	0.42	0.51
Mathematics 7	47	0.43	0.51
Mathematics 8	47	0.40	0.47

6.4. Distractor Analysis

Table 6.5 and Table 6.6 present the point-biserial correlations associated with a correct option and the incorrect options at various percentiles. As expected, the point-biserial correlation for a correct option was around 0.20 or higher for most items, whereas the point-biserial correlation for incorrect options was negative or very close to zero. The results show that students with higher proficiency tended to choose a correct option, and students with lower proficiency tended to choose an incorrect option. This indicates that the distractors appear to perform appropriately.

Table 6.5. Distractor Analysis Summary: Point-Biserial Correlations for Correct Options

Assessment	#MC Items	Min.	P25	P50	P75	Max.
ELA 3, Form 1	27	0.24	0.39	0.45	0.53	0.57
ELA 4, Form 1	25	0.30	0.39	0.48	0.51	0.59
ELA 5, Form 1	28	0.26	0.39	0.44	0.50	0.59
ELA 6, Form 1	27	0.29	0.38	0.44	0.52	0.64
ELA 7, Form 1	30	0.24	0.36	0.44	0.48	0.60
ELA 8, Form 1	30	0.22	0.35	0.43	0.50	0.57
ELA 3, Form 2	27	0.24	0.39	0.45	0.53	0.57
ELA 4, Form 2	25	0.30	0.39	0.48	0.51	0.59
ELA 5, Form 2	28	0.26	0.39	0.44	0.50	0.59
ELA 6, Form 2	27	0.29	0.38	0.44	0.52	0.64
ELA 7, Form 2	30	0.24	0.36	0.44	0.48	0.60
ELA 8, Form 2	30	0.22	0.35	0.43	0.50	0.57
Mathematics 3	19	0.27	0.42	0.52	0.57	0.63
Mathematics 4	21	0.32	0.45	0.49	0.51	0.59
Mathematics 5	23	0.29	0.37	0.43	0.47	0.54
Mathematics 6	21	0.28	0.34	0.42	0.48	0.56
Mathematics 7	25	0.26	0.38	0.45	0.49	0.59
Mathematics 8	28	0.29	0.36	0.40	0.49	0.63

Note. Min.= minimum, P25 = 25th percentile, P50 = 50th percentile (median), P75 = 75th percentile, Max. = maximum. This analysis is conducted for MC items only.

Table 6.6. Distractor Analysis Summary: Point-Biserial Correlations for Incorrect Options

Assessment	#MC Items	Min.	P25	P50	P75	Max.
ELA 3, Form 1	27	-0.36	-0.26	-0.21	-0.15	0.00
ELA 4, Form 1	25	-0.37	-0.29	-0.23	-0.17	0.01
ELA 5, Form 1	28	-0.38	-0.27	-0.23	-0.18	0.08
ELA 6, Form 1	27	-0.40	-0.30	-0.23	-0.17	0.05
ELA 7, Form 1	30	-0.37	-0.29	-0.21	-0.14	0.05
ELA 8, Form 1	30	-0.35	-0.26	-0.21	-0.16	0.08
ELA 3, Form 2	27	-0.36	-0.26	-0.21	-0.15	0.00
ELA 4, Form 2	25	-0.37	-0.29	-0.23	-0.17	0.01
ELA 5, Form 2	28	-0.38	-0.27	-0.23	-0.18	0.08
ELA 6, Form 2	27	-0.40	-0.30	-0.23	-0.17	0.05
ELA 7, Form 2	30	-0.37	-0.29	-0.21	-0.14	0.05
ELA 8, Form 2	30	-0.35	-0.26	-0.21	-0.16	0.08
Mathematics 3	19	-0.45	-0.30	-0.24	-0.16	-0.07
Mathematics 4	21	-0.44	-0.28	-0.23	-0.17	-0.06
Mathematics 5	23	-0.38	-0.26	-0.20	-0.14	0.06
Mathematics 6	21	-0.34	-0.25	-0.18	-0.11	0.05
Mathematics 7	25	-0.39	-0.26	-0.19	-0.15	0.04
Mathematics 8	28	-0.37	-0.24	-0.19	-0.13	0.02

Note. Min.= minimum, P25 = 25th percentile, P50 = 50th percentile (median), P75 = 75th percentile, Max. = maximum. This analysis is conducted for MC items only.

A distractor analysis was also conducted for each multiple-choice item, as presented in Appendix A. The response distribution for an item across all possible choices (e.g., a correct option and distractors) was calculated. The point-biserial correlation associated with each response option was calculated as well. Typically, a negative point-biserial correlation is sought for distractors because less-proficient students should be more likely to choose an incorrect option.

Chapter 7: CALIBRATION, EQUATING, AND SCALING

This chapter describes the calibration, equating, and scaling procedures that took place for the Spring 2024 AASA assessments, addressing Standards 1.10, 5.1, 5.2, 5.3, 7.2, 7.4, and 12.9 (AERA et al., 2014).

7.1. Calibration Sample

To ensure valid calibration results, several data cleaning steps occurred upon receipt of raw data from the scanning and scoring processes. These steps allowed for calibration to be conducted on valid student responses. The cleaning process removed the following records from the calibration datasets for each grade level:

- Records with invalidated tests that are marked Do Not Report (DNR) in PearsonAccess^{next} (PAN)
- Records that indicate the student took an accommodated form
- Records with non-valid attempts noted by less than one response
- Duplicate records (e.g., students indicated as taking the test more than once)
- Records in which a student was enrolled in an exclusionary school list from ADE

7.2. Calibration Methods

Item response theory (IRT) models were used in the item calibration. All tests were calibrated separately by grade. If there was more than one operational form, all operational forms were calibrated concurrently. All calibration activities were replicated with two psychometricians independently as a quality control measure. The calibration results were also reviewed independently by a senior-level psychometrician at Pearson.

The Rasch model (Rasch, 1960) was used for 1-point items and the partial-credit model (Masters, 1982) was used for multiple-point items for calibration. Parameter estimation for items was implemented using Winsteps 4.8.1.0 (Linacre, 2022b) that uses joint maximum likelihood estimation (JMLE) as described by Wright and Masters (1982).

The Rasch model estimates item difficulty and student ability on the same scale. Under the Rasch model, the probability that student j with ability θ answers item i with difficulty of b correctly is as follows:

$$P_i(\theta_j) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)}$$

The partial-credit model is an extension of the Rasch model for items in which students may receive partial credit. Thus, the partial-credit model reduces to the Rasch model when items have only two response categories (i.e., 0 or 1). According to the partial-credit model, the probability that student j scores x on item i, which has a maximum possible score of m ($k = m+1$ possible response categories), is expressed as follows:

$$P_{ix}(\theta_j) = \frac{\exp \sum_{l=0}^{x} (\theta_j - D_{il})}{\sum_{k=0}^{m_i} [\exp \sum_{l=0}^{k} (\theta_j - D_{il})]}$$

where $x = 0, 1, \ldots, m_i$, D_{il} is a step difficulty for score l and by definition,

$$\sum_{l=0}^{0}(\theta_j - D_{il}) = 0$$

The step difficulty D_{il} can be decomposed such that

$$D_{il} = b_i + h_{il}$$

where b_i is an overall difficulty for item i, and h_{il} is a threshold for score l (Embretson & Reise, 2000; Linacre, 2022a). This parameterization allows b_i in the partial-credit model to be comparable to b_i in the Rasch model.

7.3. Calibration Results

All items converged during calibration using typical procedures for Winsteps software. Standard error of estimates for the Rasch difficulty measures indicated that the parameters were well-estimated. Table 7.1 presents a summary of the IRT statistics, and Appendix B presents the item-level IRT statistics resulting from the calibration of the spring AASA assessments.

Table 7.1. IRT Statistics Summary

Assessment	#Items	Mean Rasch
ELA 3, Form 1	44	0.00
ELA 4, Form 1	44	0.27
ELA 5, Form 1	44	0.00
ELA 6, Form 1	44	0.03
ELA 7, Form 1	44	0.05
ELA 8, Form 1	44	-0.01
ELA 3, Form 2	44	-0.01
ELA 4, Form 2	44	0.28
ELA 5, Form 2	44	-0.01
ELA 6, Form 2	44	0.03
ELA 7, Form 2	44	0.07
ELA 8, Form 2	44	0.01
Mathematics 3	45	0.12
Mathematics 4	45	0.05
Mathematics 5	45	0.07
Mathematics 6	47	0.00
Mathematics 7	47	0.07
Mathematics 8	47	-0.10

An item-person map shows the distribution of item difficulty and the distribution of student ability in one graph, as they are on the same scale. This graph is useful for Rasch models to evaluate the extent to which the item difficulty and student ability distributions are aligned because they assume the probability of a correct answer is affected only by a student's ability and the item difficulty. Figure B.1 – Figure B.18 in Appendix B present the item difficulty distribution on the lefthand side and the student ability distribution on the right. Each marker in the item difficulty distribution is an item, and the item difficulty values are rounded with an increment of 0.20 before they are plotted. Horizontal dotted lines represent the three performance level cuts (i.e., *Partially Proficient*, *Proficient*, and *Highly Proficient*) for the total test.

In addition to the item-person map, two more graphs are presented to summarize the characteristics of each operational assessment. The test characteristic curve (TCC) shows an expected total raw score across different student abilities, whereas the CSEM curve presents an amount of standard error across different student abilities. The CSEM has an inverse relationship with the test information function (TIF) as follows:

$$SE(\theta) = \frac{1}{TI(\theta)}$$

where $SE(\theta)$ is the CSEM, and $TI(\theta)$ is the TIF (Embretson & Reise, 2000). Because the CSEM can be interpreted on the ability scale, the CSEM curve is presented over the TIF curve in this technical report.

7.4. Equating

The Spring 2024 AASA tests were equated and placed on the operational AASA scale using a non-equivalent groups anchor item (NEAT) design. A set of anchor items was selected from the existing item bank. The anchor items were selected such that they contributed approximately 30% of the total score points and their content representation was as similar as possible to the blueprint. The location of all anchor items stayed within three positions from where they were in the previous year.

A fixed anchor parameter equating was implemented within Winsteps to place the tests on the operational reporting scale. This was implemented by constraining the parameter estimates in the existing item bank for the anchor items to equal the final parameter estimates obtained in the original AASA calibration analyses. The displacement statistic, which estimates the difference between the fixed parameter and the estimate had the item parameter not been constrained, was evaluated for each anchor item.

Items with a displacement statistic greater than 0.30 or less than -0.30 were reiteratively removed from the anchor set. The criterion of 0.30 has been used to flag displaced anchor items under a common item, non-equivalent group equating design for many state programs (Miller et al., 2004). If more than one anchor item was flagged, the item with the largest magnitude of displacement value was dropped from the anchor set. The displacement values of the remaining anchor items were then re-estimated by implementing the fixed anchor parameter equating with the remaining anchor items. This process was repeated until all the anchor items had displacement values of a magnitude smaller than 0.30 and greater than -0.30.

Table 7.2 presents the number of items for the initial anchor set of each grade and the number of items dropped from each initial anchor set.

Table 7.2. Summary of Anchor Items

Assessment	#Items in Initial Anchor Set	#Items Dropped from Anchor
ELA 3	14	2
ELA 4	17	1
ELA 5	15	1
ELA 6	15	2
ELA 7	14	0
ELA 8	14	1
Mathematics 3	15	1
Mathematics 4	16	2
Mathematics 5	15	1
Mathematics 6	16	2
Mathematics 7	18	1
Mathematics 8	16	0

7.5. Scaling Methods

The AASA reporting scale was established in 2015 when the first administration took place (known as the AzMERIT statewide achievement assessment at that time). These tests were placed on a vertical scale for the total score as a result of a previous study (American Institutes for Research, 2015, Appendix J). Scaling constants for the total score were determined such that the vertically scaled theta score, based on the total test, was transformed by solving the following equation:

$$Scale\ Score = VS_A \times \theta + VS_B$$

where VS_A and VS_B are scaling constants on the vertical scale that are used to transform θ, which are the performance level cuts on the theta (ability) scale, into scale scores. For reporting, θ is truncated at -3.5 and 3.5 for the lower and upper ends, respectively.

The AASA reporting scale ranged from 2395 to 2658 across grades for ELA and from 3395 to 3776 across grades for mathematics. In addition to a total score, a subscore was also calculated for each reporting category by grade using the same formula. The scaling constants were applied to a theta score based on items associated with a reporting category to transform it to a scale score. Table B.13 – Table B.24 in Appendix B presents the raw-to-scale score conversion tables for each content area and grade.

7.6. IRT Assumptions

It is important to evaluate how the Rasch models fit the data because reported scale scores are derived from theta estimated under the IRT models. Three major assumptions are investigated: unidimensionality, local item independence, and item fit.

7.6.1. Unidimensionality

An assumption under the Rasch models is unidimensionality, that there is exactly one latent variable (e.g., mathematics proficiency) that an instrument intends to measure. This is a more traditional and strict definition of the unidimensionality assumption. On the other hand, essential unidimensionality, in which there is one dominant latent variable with some minor latent variable(s), is a more practically applicable assumption (Stout, 1990).

Principal component analysis (PCA) is a statistical technique widely applied to investigate the dimensionality of data (Jackson, 1993; Velicer & Jackson, 1990). Many decision rules have been proposed to determine the number of dimensions using PCA results. Horn's (1965) parallel analysis is a Monte Carlo simulation technique used to determine the number of factors to retain from a PCA. Parallel analysis compares the observed eigenvalues from a correlation matrix to be analyzed with those obtained from uncorrelated normal variables (Ledesma & Valero-Mora, 2007). In other words, expected eigenvalues are obtained by simulating normal, random samples that "parallel" the observed data in terms of sample size and number of variables. Numerous studies have shown parallel analysis to be an effective and appropriate method to determine the number of factors underlying a construct (Glorfeld, 1995; Humphreys & Montanelli, 1975; Zwick & Velicer, 1986), including the least variability and sensitivity to different factors.

PCA was conducted for the operational form in each content area and grade. Table 7.3 presents the first 10 eigenvalues from PCA for each operational form, as well as the percentage of total variance explained by the first component (%Var). Reckase (1979) claimed that at least 20% of the total variance should be accounted for by the first principal component to obtain acceptable parameter estimates in a unidimensional model. Because the same blueprint was used to construct the operational forms, only one set of eigenvalues from the parallel analysis is presented. The graphical presentations of eigenvalues (i.e., scree plot) are presented in Figure B.55 – Figure B.72 in Appendix B. The PCA results with the parallel analysis criterion and Reckase's index show only one significant dimension for each grade, which supports unidimensionality.

Table 7.3. Eigenvalues from PCA

Assessment	1	2	3	4	5	6	7	8	9	10	%Var
ELA 3, Form 1	15.72	1.58	1.21	1.01	0.95	0.94	0.90	0.87	0.83	0.82	36%
ELA 4, Form 1	17.21	1.51	1.22	1.02	1.01	0.88	0.86	0.83	0.82	0.78	39%
ELA 5, Form 1	15.66	1.46	1.13	1.02	0.94	0.92	0.91	0.89	0.85	0.84	36%
ELA 6, Form 1	15.86	1.47	1.18	0.98	0.96	0.92	0.90	0.88	0.87	0.84	36%
ELA 7, Form 1	15.63	1.34	1.18	1.00	0.94	0.88	0.88	0.87	0.85	0.84	36%
ELA 8, Form 1	14.11	1.71	1.15	1.08	0.99	0.94	0.93	0.90	0.88	0.86	32%
ELA 3, Form 2	15.97	1.61	1.11	1.01	0.95	0.93	0.89	0.86	0.84	0.83	36%
ELA 4, Form 2	17.22	1.53	1.23	1.03	0.96	0.88	0.85	0.84	0.81	0.79	39%
ELA 5, Form 2	15.57	1.45	1.17	0.97	0.95	0.94	0.92	0.89	0.84	0.84	35%
ELA 6, Form 2	16.06	1.45	1.14	0.97	0.96	0.90	0.89	0.87	0.85	0.84	37%
ELA 7, Form 2	15.48	1.38	1.20	0.99	0.93	0.90	0.88	0.87	0.86	0.84	35%
ELA 8, Form 2	14.21	1.70	1.20	1.06	0.99	0.93	0.92	0.89	0.87	0.86	32%

Assessment	1	2	3	4	5	6	7	8	9	10	%Var
Mathematics 3	21.57	1.66	1.22	0.96	0.86	0.83	0.79	0.77	0.74	0.71	48%
Mathematics 4	20.60	1.70	1.11	0.99	0.91	0.85	0.81	0.77	0.74	0.71	46%
Mathematics 5	16.78	1.61	1.31	1.14	1.05	0.92	0.88	0.86	0.84	0.82	37%
Mathematics 6	20.02	1.67	1.37	0.98	0.95	0.93	0.91	0.84	0.83	0.83	43%
Mathematics 7	20.17	1.52	1.06	0.98	0.93	0.90	0.86	0.83	0.82	0.81	43%
Mathematics 8	17.66	1.87	1.17	1.02	0.99	0.94	0.92	0.88	0.84	0.83	38%

7.6.2. Local Item Independence

Local item independence is another assumption under the Rasch models that assumes any item pair is uncorrelated, conditioned on the latent trait an instrument is intended to measure (e.g., mathematics proficiency). A violation of local item dependence would impact parameter estimation under the Rasch models because JMLE performed by Winsteps (Linacre, 2022b) relies on uncorrelated item pairs. Winsteps produces raw score residual correlations for pairs of items on a test, which are analogous to Yen's Q3 statistics (Yen, 1984). For an item pair with the residual correlation greater than 0.70, only one item is needed on the test (Linacre, 2022a).

Table 7.4 summarizes the distribution of the residual correlations. Most residual correlations are slightly negative or slightly positive, and only eight in ELA (out of more than 1,000 per grade) are greater than 0.70, which indicates that the local item independence assumption holds for the AASA tests.

Table 7.4. Q3 Statistics

Assessment	#Item Pairs	Mean	SD	Min.	P10	P25	P50	P75	P90	Max.	#Items Exceeding 0.70
ELA 3	1,081	-0.02	0.05	-0.10	-0.06	-0.04	-0.02	-0.01	0.01	0.71	1
ELA 4	1,081	-0.02	0.05	-0.10	-0.06	-0.04	-0.02	-0.01	0.01	0.81	2
ELA 5	1,081	-0.02	0.05	-0.09	-0.05	-0.04	-0.03	-0.01	0.01	0.84	2
ELA 6	1,081	-0.02	0.05	-0.09	-0.06	-0.04	-0.02	-0.01	0.01	0.72	1
ELA 7	1,081	-0.02	0.05	-0.10	-0.06	-0.04	-0.02	-0.01	0.01	0.77	2
ELA 8	1,081	-0.02	0.04	-0.09	-0.05	-0.04	-0.02	-0.01	0.01	0.63	0
Mathematics 3	990	-0.02	0.03	-0.11	-0.06	-0.04	-0.02	-0.01	0.02	0.17	0
Mathematics 4	990	-0.02	0.03	-0.13	-0.06	-0.04	-0.03	-0.01	0.01	0.35	0
Mathematics 5	990	-0.02	0.03	-0.10	-0.06	-0.04	-0.03	-0.01	0.01	0.28	0
Mathematics 6	1,081	-0.02	0.04	-0.13	-0.06	-0.04	-0.02	0.00	0.02	0.48	0
Mathematics 7	1,081	-0.02	0.04	-0.14	-0.06	-0.04	-0.02	0.00	0.02	0.33	0
Mathematics 8	1,081	-0.02	0.03	-0.13	-0.06	-0.04	-0.02	0.00	0.02	0.21	0

Note. SD = standard deviation, min. = minimum, P10 = 10th percentile, P25 = 25th percentile, P50 = 50th percentile, P75 = 75th percentile, P90 = 90th percentile, max. = maximum

7.6.3. Item Fit

Item fit was monitored using weighted mean-square (MNSQ) that indicates the degree of accuracy and predictability with which the data fit the model (Linacre, 2022b). In Winsteps and Rasch literature, weighted mean-square is also referred to as infit MNSQ. The infit MNSQ is sensitive to unexpected responses at or near the item's calibrated level. Items were flagged for misfit using a set of conservative criteria. For infit MNSQ, values less than 0.60 or greater than 1.40 were flagged, in accordance with Wright and Linacre's (1994) recommendation.

Table 7.5 presents a summary of the item fit statistics, and Table B.1 – Table B.12 in Appendix B present the statistics for each item. Items flagged by Winsteps' infit statistics are reviewed during test construction for possible replacement in future administrations.

Table 7.5. IRT Item Fit Summary Statistics

Assessment	#Items	#Flagged Items by Infit	%Flagged
ELA 3	47	0	0
ELA 4	47	0	0
ELA 5	47	0	0
ELA 6	47	0	0
ELA 7	47	0	0
ELA 8	47	0	0
Mathematics 3	45	0	0
Mathematics 4	45	0	0
Mathematics 5	45	0	0
Mathematics 6	47	0	0
Mathematics 7	47	0	0
Mathematics 8	47	0	0

Chapter 8: TEST RESULTS

This chapter presents the test results of the Spring 2024 AASA administration, addressing Standards 1.8, 2.11, 2.15, 3.1, 3.3, 3.6, 3.15, 5.3, 7.4, 12.17, and 12.18 (AERA et al., 2014). The results, summarized below, are based on the population data contained within the final electronic data files (note that the data in this chapter are different from the calibration sample). The results in this section of the technical report may differ slightly from the final testing results presented on the ADE website due to small differences in the application of exclusion rules. Official results typically use more detailed school-level information than is used to conduct research analyses. Please note that the results in the following tables are presented as evidence of reliability and validity of the test scores and should not be used for state accountability purposes.

- Table 8.1 presents the test results for all students by content area and grade, including the mean and standard deviation (SD) of the scale scores and the percentage of students in the overall performance levels. Overall performance levels are determined based on the performance levels for the total score. The table also presents the percentage of Grade 3 students who met the MOWR policy requirements (see Section 10.3 for more information). Results from the last three years are included to show longitudinal performance. The percentage of students who are on-grade (Levels 3 and 4) are consistent across years in a given content area and grade.
- Table 8.2 and Table 8.3 present the percentage of students in each level of mastery by reporting category.
- Appendix C presents the test results by demographics. Histograms of the scale score distribution for the total score are also presented.
- Table 8.4 and Table 8.5 present the mean and standard deviation of the scale scores and the performance level distributions by accommodation for students who used the available accommodations. These tables only include the accommodations captured in the student data file (i.e., accommodations used by students during the Spring 2024 administration).
- Table 8.6 and Table 8.7 present the frequency distribution statistics for total scale score by performance level. Results indicate that average scale scores increase when moving from lower to higher performance levels across all grades and content areas.

Table 8.1. Overall Test Results by Year

Assessment	Year	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4	%MOWR Met
ELA 3	2024	77,828	2497.30	35.66	50.4	10.5	25.8	13.4	95.3
	2023	80,814	2498.68	35.51	47.4	11.7	27.1	13.8	95.9
	2022	79,804	2500.64	35.45	47.6	11.6	26.1	14.7	96.6
ELA 4	2024	80,867	2518.04	37.42	41.6	12.6	29.5	16.2	–
	2023	80,659	2519.03	33.92	41.1	14.2	30.5	14.3	–
	2022	79,949	2519.10	34.04	42.0	13.8	30.8	13.5	–
ELA 5	2024	81,113	2528.98	35.09	39.5	22.5	28.7	9.3	–
	2023	80,917	2528.74	35.21	40.2	22.6	28.3	8.9	–
	2022	80,649	2529.81	35.65	39.7	21.3	28.7	10.4	–
ELA 6	2024	80,561	2542.44	35.78	38.3	18.8	35.4	7.5	–
	2023	81,369	2543.19	34.76	35.7	22.4	35.5	6.5	–
	2022	81,041	2542.68	30.12	35.9	25.4	35.2	3.6	–

Assessment	Year	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4	%MOWR Met
ELA 7	2024	81,605	2551.34	34.29	40.7	18.1	32.3	9.0	–
	2023	82,061	2552.90	31.66	38.9	19.6	32.8	8.7	–
	2022	83,804	2554.47	34.18	38.1	19.2	32.0	10.8	–
ELA 8	2024	82,472	2557.26	32.54	42.8	22.0	26.2	9.0	–
	2023	85,232	2557.60	34.17	41.4	22.2	26.5	9.8	–
	2022	87,227	2558.97	33.04	42.1	22.4	25.8	9.7	–
Mathematics 3	2024	79,059	3516.30	46.80	31.4	26.1	29.5	13.1	–
	2023	81,986	3517.02	48.72	31.2	25.5	28.3	14.9	–
	2022	80,808	3515.80	44.57	33.2	27.3	27.6	11.9	–
Mathematics 4	2024	81,598	3545.79	45.86	37.5	26.9	22.0	13.5	–
	2023	81,480	3543.68	49.02	39.3	21.9	26.2	12.6	–
	2022	80,600	3545.07	51.24	37.7	23.1	25.4	13.8	–
Mathematics 5	2024	81,727	3580.40	38.12	36.8	30.1	23.3	9.9	–
	2023	81,451	3578.26	44.04	36.5	27.3	25.0	11.2	–
	2022	81,283	3577.68	44.21	39.4	23.7	25.5	11.4	–
Mathematics 6	2024	81,288	3605.44	41.58	47.8	22.7	18.7	10.8	–
	2023	82,066	3606.26	41.62	48.7	20.9	20.8	9.6	–
	2022	81,769	3607.68	41.96	48.3	21.2	19.8	10.8	–
Mathematics 7	2024	82,360	3626.98	41.96	54.7	16.3	15.5	13.6	–
	2023	82,799	3625.99	46.01	54.1	15.8	15.3	14.8	–
	2022	84,940	3626.67	41.14	55.6	17.3	14.4	12.6	–
Mathematics 8	2024	83,137	3653.89	37.67	54.6	17.8	16.5	11.2	–
	2023	86,031	3653.61	37.08	53.9	19.1	16.6	10.4	–
	2022	88,301	3653.19	36.48	55.2	18.0	16.6	10.2	–

Note. SS = scale score, SD = standard deviation, Level 1 = *Minimally Proficient*, Level 2 = *Partially Proficient*, Level 3 = *Proficient*, Level 4 = *Highly Proficient*

Table 8.2. Performance Distributions by Reporting Category: Percentage of Students at each Level of Mastery—ELA

Grade	Reporting Category	N	%Level 1	%Level 2	%Level 3
3	Reading for Information	77,828	47.1	28.9	23.9
	Reading for Literature	77,828	49.4	26.3	24.3
	Writing and Language	77,828	43.0	30.0	27.0
4	Reading for Information	80,867	41.3	26.9	31.8
	Reading for Literature	80,867	44.7	27.7	27.6
	Writing and Language	80,867	41.4	27.3	31.4
5	Reading for Information	81,113	48.7	28.3	23.0
	Reading for Literature	81,113	44.4	33.7	21.9
	Writing and Language	81,113	42.6	35.3	22.1
6	Reading for Information	80,561	45.9	26.9	27.3
	Reading for Literature	80,561	44.3	29.8	25.8
	Writing and Language	80,561	40.9	34.8	24.3
7	Reading for Information	81,605	44.2	29.5	26.3
	Reading for Literature	81,605	48.6	28.7	22.7
	Writing and Language	81,605	45.0	26.8	28.2

Grade	Reporting Category	N	%Level 1	%Level 2	%Level 3
8	Reading for Information	82,472	50.2	30.3	19.5
	Reading for Literature	82,472	51.1	32.2	16.7
	Writing and Language	82,472	44.1	35.5	20.3

Note. Level 1 = *Below Mastery*, Level 2 = *At or Around Mastery*, Level 3 = *Above Mastery*

Table 8.3. Performance Distributions by Reporting Category: Percentage of Students at each Level of Mastery—Mathematics

Grade	Reporting Category	N	%Level 1	%Level 2	%Level 3
3	Operations, Algebraic Thinking, and Numbers in Base Ten	79,059	47.5	20.9	31.6
	Numbers and Operations – Fractions	79,059	43.2	27.1	29.7
	Measurement, Data, and Geometry	79,059	46.1	33.8	20.1
4	Operations, Algebraic Thinking, and Numbers in Base Ten	81,598	54.6	19.9	25.5
	Numbers and Operations – Fractions	81,598	52.2	22.0	25.8
	Measurement, Data, and Geometry	81,598	37.5	43.3	19.2
5	Operations, Algebraic Thinking, and Numbers in Base Ten	81,727	53.3	23.6	23.1
	Numbers and Operations – Fractions	81,727	49.1	30.8	20.1
	Measurement, Data, and Geometry	81,727	43.8	37.1	19.1
6	Ratio and Proportional Relationships	81,288	59.2	19.8	21.0
	The Number System	81,288	56.4	25.7	17.9
	Expressions and Equations	81,288	61.7	19.3	19.0
	Geometry, Statistics and Probability	81,288	60.8	23.3	15.9
7	Ratio and Proportional Relationships	82,360	57.1	28.8	14.2
	The Number System	82,360	50.2	27.5	22.3
	Expressions & Equations	82,360	57.5	21.4	21.0
	Geometry, Statistics and Probability	82,360	57.0	28.1	14.9
8	Expressions and Equations	83,137	61.6	16.3	22.1
	Functions	83,137	56.5	29.9	13.6
	Geometry	83,137	56.5	23.2	20.3
	Statistics and Probability and The Number System	83,137	62.5	23.2	14.2

Note. Level 1 = *Below Mastery*, Level 2 = *At or Around Mastery*, Level 3 = *Above Mastery*

Table 8.4. Test Results by Accommodation—ELA

Grade	Accommodation	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
3	Adult Transcription	5	–	–	–	–	–	–
	American Sign Language	0	–	–	–	–	–	–
	Assistive Technology	4	–	–	–	–	–	–
	Braille Test Booklet	0	–	–	–	–	–	–
	Large Print Test Booklet	14	2478.50	32.58	78.6	0.0	14.3	7.1
	Read Aloud Content	112	2470.17	25.70	84.8	5.4	9.8	0.0
	Sign Test Content	1	–	–	–	–	–	–
	Simplified Directions	359	2463.09	23.45	91.1	5.0	3.1	0.8
	Translate Directions	93	2460.29	17.47	96.8	3.2	0.0	0.0
	Translation Dictionary	117	2464.09	18.74	94.9	3.4	1.7	0.0

Grade	Accommodation	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
4	Adult Transcription	5	–	–	–	–	–	–
	American Sign Language	17	2476.29	32.78	88.2	0.0	5.9	5.9
	Assistive Technology	6	–	–	–	–	–	–
	Braille Test Booklet	4	–	–	–	–	–	–
	Large Print Test Booklet	10	–	–	–	–	–	–
	Read Aloud Content	79	2482.33	29.20	84.8	6.3	7.6	1.3
	Sign Test Content	2	–	–	–	–	–	–
	Simplified Directions	340	2479.71	24.68	87.9	6.8	4.7	0.6
	Translate Directions	71	2477.32	23.06	90.1	4.2	5.6	0.0
	Translation Dictionary	99	2477.74	21.94	88.9	7.1	4.0	0.0
5	Adult Transcription	4	–	–	–	–	–	–
	American Sign Language	23	2493.57	34.69	87.0	0.0	4.3	8.7
	Assistive Technology	5	–	–	–	–	–	–
	Braille Test Booklet	8	–	–	–	–	–	–
	Large Print Test Booklet	13	2508.69	38.18	53.8	23.1	23.1	0.0
	Read Aloud Content	74	2495.11	28.43	79.7	13.5	6.8	0.0
	Sign Test Content	2	–	–	–	–	–	–
	Simplified Directions	353	2492.93	23.38	85.8	10.8	3.4	0.0
	Translate Directions	68	2489.93	23.25	88.2	8.8	2.9	0.0
	Translation Dictionary	93	2494.30	24.54	83.9	12.9	3.2	0.0
6	Adult Transcription	3	–	–	–	–	–	–
	American Sign Language	24	2503.08	27.94	79.2	12.5	8.3	0.0
	Assistive Technology	2	–	–	–	–	–	–
	Braille Test Booklet	8	–	–	–	–	–	–
	Large Print Test Booklet	8	–	–	–	–	–	–
	Read Aloud Content	36	2511.47	22.12	80.6	16.7	2.8	0.0
	Sign Test Content	1	–	–	–	–	–	–
	Simplified Directions	258	2510.22	25.94	79.8	10.5	9.3	0.4
	Translate Directions	79	2497.42	24.06	89.9	7.6	2.5	0.0
	Translation Dictionary	122	2505.12	25.14	82.8	10.7	6.6	0.0
7	Adult Transcription	3	–	–	–	–	–	–
	American Sign Language	19	2511.74	17.86	94.7	0.0	5.3	0.0
	Assistive Technology	2	–	–	–	–	–	–
	Braille Test Booklet	6	–	–	–	–	–	–
	Large Print Test Booklet	8	–	–	–	–	–	–
	Read Aloud Content	24	2530.71	29.87	66.7	12.5	20.8	0.0
	Sign Test Content	0	–	–	–	–	–	–
	Simplified Directions	258	2515.76	23.58	87.6	7.4	4.7	0.4
	Translate Directions	79	2506.97	17.77	96.2	2.5	1.3	0.0
	Translation Dictionary	143	2509.62	20.77	93.0	4.2	2.8	0.0

Grade	Accommodation	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
8	Adult Transcription	1	–	–	–	–	–	–
	American Sign Language	37	2521.35	18.67	94.6	2.7	2.7	0.0
	Assistive Technology	2	–	–	–	–	–	–
	Braille Test Booklet	4	–	–	–	–	–	–
	Large Print Test Booklet	5	–	–	–	–	–	–
	Read Aloud Content	21	2533.05	29.89	66.7	23.8	9.5	0.0
	Sign Test Content	3	–	–	–	–	–	–
	Simplified Directions	235	2525.22	21.14	86.4	11.5	1.7	0.4
	Translate Directions	74	2519.08	17.43	91.9	8.1	0.0	0.0
	Translation Dictionary	133	2518.20	19.62	91.7	6.0	2.3	0.0

Note. SS = scale score, SD = standard deviation, Level 1 = *Minimally Proficient*, Level 2 = *Partially Proficient*, Level 3 = *Proficient*, Level 4 = *Highly Proficient*. Statistics for subgroups with less than 11 students are omitted in compliance with FERPA regulations. Read aloud is for Writing only.

Table 8.5. Test Results by Accommodation—Mathematics

Grade	Accommodation	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
3	Adult Transcription	4	–	–	–	–	–	–
	American Sign Language	28	3448.79	21.98	96.4	3.6	0.0	0.0
	Assistive Technology	6	–	–	–	–	–	–
	Braille Test Booklet	6	–	–	–	–	–	–
	Large Print Test Booklet	12	3502.67	51.27	41.7	25.0	33.3	0.0
	Read Aloud Content	135	3494.82	43.34	48.9	25.9	20.7	4.4
	Sign Test Content	0	–	–	–	–	–	–
	Simplified Directions	341	3475.50	41.12	68.3	21.7	7.9	2.1
	Translate Directions	96	3482.56	38.75	64.6	22.9	10.4	2.1
	Translation Dictionary	121	3489.36	41.59	57.9	24.0	14.0	4.1
4	Adult Transcription	5	–	–	–	–	–	–
	American Sign Language	17	3497.41	43.62	82.4	5.9	5.9	5.9
	Assistive Technology	5	–	–	–	–	–	–
	Braille Test Booklet	4	–	–	–	–	–	–
	Large Print Test Booklet	12	3487.00	23.46	91.7	8.3	0.0	0.0
	Read Aloud Content	99	3508.73	35.60	67.7	26.3	4.0	2.0
	Sign Test Content	2	–	–	–	–	–	–
	Simplified Directions	345	3502.66	34.16	77.4	17.7	3.8	1.2
	Translate Directions	70	3504.80	28.25	75.7	20.0	4.3	0.0
	Translation Dictionary	99	3504.52	29.98	77.8	18.2	3.0	1.0
5	Adult Transcription	4	–	–	–	–	–	–
	American Sign Language	23	3535.96	43.96	87.0	4.3	0.0	8.7
	Assistive Technology	6	–	–	–	–	–	–
	Braille Test Booklet	8	–	–	–	–	–	–
	Large Print Test Booklet	14	3557.64	29.47	71.4	14.3	14.3	0.0
	Read Aloud Content	88	3556.66	36.39	61.4	26.1	9.1	3.4
	Sign Test Content	1	–	–	–	–	–	–
	Simplified Directions	342	3549.45	23.62	76.9	19.6	2.6	0.9
	Translate Directions	68	3551.13	24.00	73.5	23.5	1.5	1.5
	Translation Dictionary	86	3552.66	25.98	69.8	24.4	4.7	1.2

Grade	Accommodation	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
6	Adult Transcription	2	–	–	–	–	–	–
	American Sign Language	24	3579.79	35.97	83.3	12.5	0.0	4.2
	Assistive Technology	1	–	–	–	–	–	–
	Braille Test Booklet	8	–	–	–	–	–	–
	Large Print Test Booklet	8	–	–	–	–	–	–
	Read Aloud Content	32	3587.88	28.98	68.8	18.8	9.4	3.1
	Sign Test Content	4	–	–	–	–	–	–
	Simplified Directions	236	3576.55	32.02	79.7	13.1	5.1	2.1
	Translate Directions	69	3566.32	34.30	88.4	7.2	1.4	2.9
	Translation Dictionary	102	3579.33	40.06	72.5	15.7	6.9	4.9
7	Adult Transcription	0	–	–	–	–	–	–
	American Sign Language	20	3589.15	26.70	90.0	5.0	5.0	0.0
	Assistive Technology	1	–	–	–	–	–	–
	Braille Test Booklet	6	–	–	–	–	–	–
	Large Print Test Booklet	9	–	–	–	–	–	–
	Read Aloud Content	22	3598.59	25.97	86.4	9.1	4.5	0.0
	Sign Test Content	0	–	–	–	–	–	–
	Simplified Directions	241	3587.72	24.57	92.9	5.4	1.2	0.4
	Translate Directions	79	3582.23	19.70	97.5	1.3	1.3	0.0
	Translation Dictionary	107	3591.35	30.44	87.9	4.7	4.7	2.8
8	Adult Transcription	0	–	–	–	–	–	–
	American Sign Language	40	3619.83	12.55	97.5	2.5	0.0	0.0
	Assistive Technology	3	–	–	–	–	–	–
	Braille Test Booklet	4	–	–	–	–	–	–
	Large Print Test Booklet	5	–	–	–	–	–	–
	Read Aloud Content	17	3641.12	28.62	70.6	17.6	5.9	5.9
	Sign Test Content	2	–	–	–	–	–	–
	Simplified Directions	205	3625.45	18.59	91.7	5.4	2.4	0.5
	Translate Directions	74	3624.92	16.65	91.9	6.8	1.4	0.0
	Translation Dictionary	92	3625.80	18.93	90.2	7.6	2.2	0.0

Note. SS = scale score, SD = standard deviation, Level 1 = *Minimally Proficient*, Level 2 = *Partially Proficient*, Level 3 = *Proficient*, Level 4 = *Highly Proficient*. Statistics for subgroups with less than 11 students are omitted in compliance with FERPA regulations.

Table 8.6. Scale Score Distribution by Performance Level—ELA

Grade	Performance Level	N	Average Scale Score	%	Cumulative %
3	Level 1	39,190	2467.93	50.4	50.4
	Level 2	8,156	2501.72	10.5	60.8
	Level 3	20,082	2522.18	25.8	86.6
	Level 4	10,400	2556.50	13.4	100.0
4	Level 1	33,678	2481.87	41.6	41.6
	Level 2	10,182	2515.09	12.6	54.2
	Level 3	23,886	2538.41	29.5	83.8
	Level 4	13,121	2576.06	16.2	100.0
5	Level 1	32,044	2493.73	39.5	39.5
	Level 2	18,245	2530.07	22.5	62.0
	Level 3	23,284	2556.30	28.7	90.7
	Level 4	7,540	2591.80	9.3	100.0

Grade	Performance Level	N	Average Scale Score	%	Cumulative %
6	Level 1	30,825	2505.37	38.3	38.3
	Level 2	15,150	2541.15	18.8	57.1
	Level 3	28,553	2569.31	35.4	92.5
	Level 4	6,033	2607.88	7.5	100.0
7	Level 1	33,187	2517.38	40.7	40.7
	Level 2	14,730	2551.39	18.1	58.7
	Level 3	26,343	2577.05	32.3	91.0
	Level 4	7,345	2612.49	9.0	100.0
8	Level 1	35,327	2526.52	42.8	42.8
	Level 2	18,146	2560.56	22.0	64.8
	Level 3	21,577	2584.80	26.2	91.0
	Level 4	7,422	2615.48	9.0	100.0

Note. 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table 8.7. Scale Score Distribution by Performance Level—Mathematics

Grade	Performance Level	N	Average Scale Score	%	Cumulative %
3	Level 1	24,805	3460.82	31.4	31.4
	Level 2	20,609	3511.63	26.1	57.4
	Level 3	23,325	3547.65	29.5	86.9
	Level 4	10,320	3588.14	13.1	100.0
4	Level 1	30,631	3499.38	37.5	37.5
	Level 2	21,983	3543.89	26.9	64.5
	Level 3	17,970	3579.36	22.0	86.5
	Level 4	11,014	3623.87	13.5	100.0
5	Level 1	30,050	3543.08	36.8	36.8
	Level 2	24,562	3577.44	30.1	66.8
	Level 3	19,030	3611.11	23.3	90.1
	Level 4	8,085	3655.86	9.9	100.0
6	Level 1	38,874	3570.37	47.8	47.8
	Level 2	18,476	3612.64	22.7	70.6
	Level 3	15,190	3641.22	18.7	89.2
	Level 4	8,748	3683.92	10.8	100.0
7	Level 1	45,023	3595.48	54.7	54.7
	Level 2	13,418	3637.70	16.3	71.0
	Level 3	12,751	3661.98	15.5	86.4
	Level 4	11,168	3701.17	13.6	100.0
8	Level 1	45,362	3626.56	54.6	54.6
	Level 2	14,758	3659.74	17.8	72.3
	Level 3	13,685	3685.90	16.5	88.8
	Level 4	9,332	3730.53	11.2	100.0

Note. 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Chapter 9: RELIABILITY AND VALIDITY

This chapter provides evidence supporting the reliability and validity of scores on the Spring 2024 AASA assessments, addressing Standards 1.8, 1.9, 1.21, 2.3, 2.7, 2.8, 2.11, 2.15, 2.19, 3.1, 3.3, 3.6, 3.15, and 7.4 (AERA et al., 2014).

9.1. Reliability

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) refer to reliability as the "consistency of scores across replications of a testing procedure" (p. 33). A reliable test produces stable scores, meaning that very similar score distributions would result if the test were administered repeatedly under similar conditions to the same students without memory or fatigue affecting the scores. The level of reliability/precision of scores has implications for validity in that the scores must be consistent and precise enough to be useful for intended purposes. If scores are to be meaningful, tests should produce stable scores if the same group of students were to take the same test repeatedly without any fatigue or memory of the test. The range of certainty around the score should also be small enough to support educational decisions.

9.1.1. Internal Consistency

Reliability was evaluated based on the internal consistency for all tests. For test reliability, coefficient alpha, which is based on classical test theory (CTT), is a frequently used measure of internal consistency. Coefficient alpha is computed as follows:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_X^2}\right)$$

where k is the number of items, σ_X^2 is the variance of the total score, and σ_i^2 is the variance of item i (Crocker & Algina, 1986; Cronbach, 1951).

Typically, a test score is obtained from a single observation of performance and represents an estimate of the trait being measured. As an estimate, an observed test score contains some measurement error and does not perfectly reflect an individual's true score. The degree of measurement error in a test score can be estimated using a statistic called the standard error of measurement (SEM), which is calculated as follows:

$$SEM = \sigma_X \sqrt{1 - r}$$

where σ_X is a standard deviation of total score X, and r is a reliability coefficient, such as the coefficient alpha (Crocker & Algina, 1986).

Table 9.1, Table 9.2, and Table 9.3 present coefficient alphas and SEMs (computed based on the calibration sample) for the total and reporting category scores. These results suggest that the AASA assessments produce reliable scores.

Table 9.1. Coefficient Alpha and SEM by Total and Reporting Category Score—ELA, Form 1

Grade	Reporting Category	N	#Items	Coefficient Alpha	SEM
3	Total	36,987	44	0.92	3.08
	Reading for Information	36,989	18	0.78	1.87
	Reading for Literature	36,996	17	0.84	1.75
	Writing and Language	36,987	9	0.81	1.60
4	Total	38,391	44	0.93	3.13
	Reading for Information	38,388	18	0.85	1.73
	Reading for Literature	38,400	17	0.81	1.89
	Writing and Language	38,391	9	0.82	1.68
5	Total	38,723	44	0.92	3.13
	Reading for Information	38,726	19	0.82	1.88
	Reading for Literature	38,732	16	0.80	1.72
	Writing and Language	38,723	9	0.82	1.70
6	Total	38,279	44	0.92	3.12
	Reading for Information	38,286	20	0.83	1.95
	Reading for Literature	38,285	15	0.77	1.63
	Writing and Language	38,279	9	0.83	1.70
7	Total	39,013	44	0.92	3.20
	Reading for Information	39,018	18	0.83	1.82
	Reading for Literature	39,015	17	0.79	1.84
	Writing and Language	39,013	9	0.82	1.76
8	Total	39,271	44	0.91	3.19
	Reading for Information	39,265	18	0.78	1.86
	Reading for Literature	39,277	17	0.77	1.81
	Writing and Language	39,271	9	0.80	1.76

Table 9.2. Coefficient Alpha and SEM by Total and Reporting Category Score—ELA, Form 2

Grade	Reporting Category	N	#Items	Coefficient Alpha	SEM
3	Total	36,908	44	0.92	3.09
	Reading for Information	36,887	18	0.78	1.86
	Reading for Literature	36,912	17	0.84	1.75
	Writing and Language	36,908	9	0.83	1.60
4	Total	38,512	44	0.93	3.09
	Reading for Information	38,499	18	0.85	1.73
	Reading for Literature	38,508	17	0.81	1.89
	Writing and Language	38,512	9	0.81	1.63
5	Total	38,711	44	0.92	3.11
	Reading for Information	38,709	19	0.82	1.89
	Reading for Literature	38,713	16	0.80	1.72
	Writing and Language	38,711	9	0.81	1.67
6	Total	38,324	44	0.92	3.11
	Reading for Information	38,323	20	0.83	1.94
	Reading for Literature	38,331	15	0.78	1.62
	Writing and Language	38,324	9	0.83	1.68

Grade	Reporting Category	N	#Items	Coefficient Alpha	SEM
7	Total	38,728	44	0.92	3.17
	Reading for Information	38,727	18	0.83	1.82
	Reading for Literature	38,736	17	0.79	1.84
	Writing and Language	38,728	9	0.81	1.72
8	Total	39,685	44	0.91	3.19
	Reading for Information	39,692	18	0.78	1.86
	Reading for Literature	39,694	17	0.77	1.81
	Writing and Language	39,685	9	0.81	1.74

Table 9.3. Coefficient Alpha and SEM by Total and Reporting Category Score—Mathematics

Grade	Reporting Category	N	#Items	Coefficient Alpha	SEM
3	Total	75,148	45	0.94	2.66
	Operations, Algebraic Thinking, and Numbers in Base Ten	75,148	23	0.91	1.87
	Numbers and Operations – Fractions	75,117	9	0.77	1.17
	Measurement, Data, and Geometry	75,166	13	0.79	1.46
4	Total	77,701	45	0.94	2.71
	Operations, Algebraic Thinking, and Numbers in Base Ten	77,731	23	0.89	1.91
	Numbers and Operations – Fractions	77,701	14	0.85	1.50
	Measurement, Data, and Geometry	77,671	8	0.70	1.18
5	Total	78,100	45	0.92	2.87
	Operations, Algebraic Thinking, and Numbers in Base Ten	78,100	18	0.83	1.82
	Numbers and Operations – Fractions	78,082	15	0.78	1.68
	Measurement, Data, and Geometry	78,052	12	0.79	1.42
6	Total	77,374	47	0.94	2.83
	Ratio and Proportional Relationships	77,333	10	0.76	1.30
	The Number System	77,202	14	0.85	1.50
	Expressions and Equations	77,378	15	0.82	1.59
	Geometry, Statistics and Probability	77,374	8	0.65	1.19
7	Total	78,557	47	0.94	2.81
	Ratio and Proportional Relationships	78,608	10	0.75	1.30
	The Number System	78,609	10	0.82	1.23
	Expressions and Equations	78,601	12	0.80	1.44
	Geometry, Statistics and Probability	78,557	15	0.79	1.60
8	Total	79,698	47	0.92	2.83
	Functions	79,549	15	0.84	1.56
	Expressions and Equations	79,689	11	0.66	1.41
	Geometry	79,676	9	0.74	1.18
	Statistics and Probability and The Number System	79,698	12	0.71	1.46

In contrast to the CTT-based SEM, an IRT-based SEM (i.e., CSEM) varies across an ability continuum. The CSEM should be lower around important performance level cuts (e.g., *Proficient*), which indicates higher measurement precision. The CSEM tends to be higher for upper and lower ends of the ability continuum because there are usually fewer items that measure those difficulty levels. Figure B.19 – Figure B.54 in Appendix B present the TCC and CSEM curves of the assessments. As expected, the CSEMs around the performance level cuts were the lowest.

9.1.2. Inter-rater Reliability

For the handscored ELA writing prompts, the consistency with which two raters assign scores to student responses is determined by inter-rater agreement, also referred to as rater agreement, which indicates the level of agreement between two scores assigned to student responses. It is the measure of how often scorers agree with each other. Rater agreement is calculated between the human-scored and IEA-scored prompts, and rater agreement statistics include the percentage of exact and adjacent scores for each item that received two scores. For 10% of responses, a second "reliability" score was assigned by a second scorer.

The expectation is an inter-rater agreement of 65% or higher between the first and second scores. When IEA provided a high confidence score, the second reliability score was from a human rater. For the subset of responses where IEA provided a low confidence score, the first and second score were both from human raters. Pearson scoring staff used inter-rater agreement indices as one factor in determining the needs for continuing training and intervention on both individual and group levels.

Two other statistical indices are also used to measure reliability in the handscoring process: Cohen's kappa and intraclass correlation. The quadratic weighted kappa (Cohen, 1968) allows rater disagreements to be weighted differentially (e.g., magnitude of a 1-point difference in ratings versus a 2-point difference) and is calculated with the weighted differences included, which are defined by the following formulas:

$$w_{ij} = \frac{\left(|i-j|\right)^2}{\left(k-1\right)^2}$$

$$\kappa_w = 1 - \frac{\sum w_{ij} O_{ij}}{\sum w_{ij} E_{ij}}$$

where $|i-j|$ is the number of categories by which raters disagree, k is the total number of score categories, and w_{ij} is the weighted level of disagreement. E_{ij} is the expected matrix, and O_{ij} is the observed matrix. The quadratic weighed kappa ranges from -1.0 to 1.0, with higher, more positive values indicative of greater rater agreement.

The intraclass correlation is defined by Shrout and Fleiss (1979) as "the correlation between one measurement (either a single rating or a mean of ratings) on a target and another measurement obtained on that target" (p. 422). In the context of the AASA assessments, the "target" was the student response and each measurement was obtained by a rater randomly assigned to that response. Therefore, *ICC*(1,1) was used to estimate the intraclass correlation. *ICC*(1,1) is estimated as follows (Shrout & Fleiss, 1979):

$$ICC(1,1) = \frac{BMS - WMS}{BMS + \left(k-1\right)WMS}$$

where *BMS* is the between-targets mean square, *WMS* is the within-targets mean square, and *k* is the number of raters rating each target. Table 9.4 presents the quadratic weighted kappa and intraclass correlation by reporting category. Items with a kappa statistic lower than 0.20, considered as slight rater agreement (Landis & Koch, 1977) and of which there were none, were flagged for potential replacement in future administrations.

Table 9.4. Inter-rater Reliability Statistics

Grade	OE Item	Trait	Score Range	N	Quadratic Kappa	ICC	%Exact Agreement	%Adjacent Agreement
3	WR 1	Statement of Purpose, Focus & Organization	1–4	3,746	0.66	0.66	0.64	0.35
		Evidence & Elaboration	1–4	3,746	0.65	0.65	0.68	0.31
		Conventions & Editing	0–2	3,746	0.67	0.67	0.78	0.20
	WR 2	Statement of Purpose, Focus & Organization	1–4	4,024	0.57	0.57	0.63	0.35
		Evidence & Elaboration	1–4	4,024	0.58	0.58	0.67	0.31
		Conventions & Editing	0–2	4,024	0.72	0.72	0.84	0.16
4	WR 1	Statement of Purpose, Focus & Organization	1–4	3,925	0.57	0.56	0.57	0.40
		Evidence & Elaboration	1–4	3,925	0.57	0.56	0.58	0.39
		Conventions & Editing	0–2	3,925	0.56	0.56	0.87	0.11
	WR 2	Statement of Purpose, Focus & Organization	1–4	4,199	0.74	0.74	0.72	0.26
		Evidence & Elaboration	1–4	4,199	0.77	0.77	0.77	0.22
		Conventions & Editing	0–2	4,199	0.79	0.79	0.81	0.18
5	WR 1	Statement of Purpose, Focus & Organization	1–4	3,892	0.59	0.58	0.58	0.41
		Evidence & Elaboration	1–4	3,892	0.55	0.53	0.55	0.44
		Conventions & Editing	0–2	3,892	0.71	0.71	0.90	0.09
	WR 2	Statement of Purpose, Focus & Organization	1–4	4,021	0.81	0.81	0.79	0.20
		Evidence & Elaboration	1–4	4,021	0.80	0.80	0.82	0.17
		Conventions & Editing	0–2	4,021	0.73	0.73	0.72	0.27
6	WR 1	Statement of Purpose, Focus & Organization	1–4	3,785	0.71	0.70	0.64	0.34
		Evidence & Elaboration	1–4	3,785	0.69	0.69	0.65	0.33
		Conventions & Editing	0–2	3,785	0.67	0.67	0.77	0.20
	WR 2	Statement of Purpose, Focus & Organization	1–4	4,000	0.84	0.84	0.66	0.33
		Evidence & Elaboration	1–4	4,000	0.83	0.83	0.68	0.31
		Conventions & Editing	0–2	4,000	0.79	0.79	0.80	0.19
7	WR 1	Statement of Purpose, Focus & Organization	1–4	3,837	0.81	0.81	0.71	0.28
		Evidence & Elaboration	1–4	3,837	0.79	0.79	0.69	0.30
		Conventions & Editing	0–2	3,837	0.76	0.76	0.81	0.18
	WR 2	Statement of Purpose, Focus & Organization	1–4	3,914	0.55	0.54	0.59	0.38
		Evidence & Elaboration	1–4	3,914	0.57	0.57	0.59	0.39
		Conventions & Editing	0–2	3,914	0.65	0.65	0.88	0.11
8	WR 1	Statement of Purpose, Focus & Organization	1–4	3,951	0.80	0.80	0.68	0.31
		Evidence & Elaboration	1–4	3,951	0.82	0.82	0.77	0.22
		Conventions & Editing	0–2	3,951	0.67	0.67	0.75	0.25
	WR 2	Statement of Purpose, Focus & Organization	1–4	3,914	0.72	0.72	0.61	0.36
		Evidence & Elaboration	1–4	3,914	0.76	0.76	0.68	0.29
		Conventions & Editing	0–2	3,914	0.60	0.60	0.76	0.24

Note. OE = open-ended, ICC = intraclass correlation

9.2. Differential Item Functioning

Because test scores can have many sources of variation, the test developers' task is to create assessments that measure the intended abilities and skills without introducing extraneous elements or construct-irrelevant variance. When tests measure something other than what they are intended to measure, test scores will reflect these unintended skills and knowledge, as well as what is purportedly assessed by the test. If this occurs, these tests can be called biased (Angoff, 1993; Camilli & Shepard, 1994; Green, 1975; Zumbo, 1999). One of the factors that may render test scores biased is differing cultural and socioeconomic experiences.

Analysis of DIF is a statistical method to detect potential bias of an item. DIF is defined as a difference between groups (e.g., male and female) in the probability of answering an item correctly. DIF analyses are conditioned on the ability that the assessment is intended to measure (e.g., mathematics proficiency). DIF is an indicator that the item might exhibit bias for one group over the other, not that it actually does. If DIF exists on an item, a committee composed of subject experts reviews the item to determine whether it actually shows bias.

The Mantel-Haenszel (MH) method (Holland & Thayer, 1988; Mantel & Haenszel, 1959) was used to investigate DIF on one-point items. The MH method is frequently used and efficient in terms of statistical power (Clauser & Mazor, 1998). The Mantel-Haenszel chi-square statistic is computed as follows:

$$MH - \chi^2 = \frac{(\sum_k F_k - \sum_k E(F_k))^2}{\sum_k Var(F_k)}$$

where F_k is the sum of scores for the focal group at the *k*th level of the matching variable (Zwick et al., 1993). 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 MH delta statistic (ΔMH) was computed. Educational Testing Service (ETS) first developed the ΔMH DIF statistic. To compute the ΔMH DIF, the MH alpha (the odds ratio) is first computed:

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

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*. The *ΔMH DIF* is computed as follows:

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

Positive values of *ΔMH DIF* indicate items that favor the focal group, whereas negative values indicate items that favor the reference group. The MH chi-square statistic and the *ΔMH DIF* were used in combination to identify both the operational and field test items that exhibit strong, weak, or no DIF for single-point items.

The standardized mean difference (SMD) is another DIF method applied to multiple-point items (Dorans & Schmitt, 1991; Zwick et al., 1993). The SMD is an effect size index of DIF that compares the mean scores of the reference and focal groups for an item, adjusting for the distribution of the reference and focal groups on the conditioned variable, which for the analyses is the raw score. The SMD is computed as follows:

$$SMD = \sum_k P_{F_k}(m_{F_k} - m_{R_k})$$

where P_{F_k} is the proportion of the focal group at the *k*th level of the matching variable, m_{F_k} is the mean score on the item for the focal group at the *k*th level of the matching variable, and m_{R_k} is the mean score on the item for the reference group at the *k*th level of the matching variable (Zwick et al., 1993). A negative SMD value indicates an item in which the focal group has a lower mean than the reference group, conditioned on the matching variable (e.g., science proficiency), whereas a positive SMD value indicates an item for which the reference group has a lower mean than the focal group, conditioned on the matching variable.

Table 9.5 presents the summary of DIF classification criteria for both the MH method and SMD. An alpha level of 0.05 was used for all MH and SMD statistics.

Table 9.5. DIF Flag Categories

Category	Description	MH Criterion	SMD Criterion
A	No DIF	MH chi-square not significantly different from 0 ($p < 0.05$) or $\lvert\Delta MH\ DIF\rvert < 1.0$	MH chi-square not significantly different from 0 ($p < 0.05$) or $\lvert SMD\rvert \leq 0.17$
B	Weak DIF	MH chi-square significantly different from 0 ($p < 0.05$) and $1.0 \leq \lvert\Delta MH\ DIF\rvert < 1.5$	MH chi-square significantly different from 0 ($p < 0.05$) and $0.17 < \lvert SMD\rvert \leq 0.25$
C	Strong DIF	MH chi-square significantly higher than 1 ($p < 0.05$) and $\lvert\Delta MH\ DIF\rvert \geq 1.5$	MH chi-square significantly different from 0 ($p < 0.05$) and $\lvert SMD\rvert > 0.25$

The DIF analysis was conducted for 10 different group pairs:

1. Female vs. Male
2. Hispanic vs. Non-Hispanic
3. American Indian vs. White
4. Asian vs. White
5. Black or African American vs. White
6. Native Hawaiian or Other Pacific Islander vs. White
7. Multi-racial vs. White
8. Students with Disability vs. Students without Disability

9. Economically Disadvantaged vs. Not Economically Disadvantaged
10. English Learner vs. English as a First Language

Table 9.6 presents the number of operational items exhibiting strong DIF between any two groups. Any items that display strong DIF are flagged for possible replacement in the future administration, as strong DIF is one of the holistic item replacement evaluation criteria used for item selection. DIF results with a sample size of less than 200 per group should not be considered statistically reliable (Clauser & Mazor, 1998; Mazor et al., 1992).

Table 9.6. Number of Items Exhibiting Strong DIF

Assessment	Total #Items	#Items with Strong DIF
ELA 3, Form 1	44	0
ELA 4, Form 1	44	0
ELA 5, Form 1	44	2
ELA 6, Form 1	44	2
ELA 7, Form 1	44	1
ELA 8, Form 1	44	3
ELA 3, Form 2	44	0
ELA 4, Form 2	44	0
ELA 5, Form 2	44	3
ELA 6, Form 2	44	2
ELA 7, Form 2	44	1
ELA 8, Form 2	44	2
Mathematics 3	45	1
Mathematics 4	45	1
Mathematics 5	45	1
Mathematics 6	47	1
Mathematics 7	47	4
Mathematics 8	47	1

9.3. Correlations Among Reporting Categories

Correlations were examined between the total raw score and the reporting category raw scores. The data used to calculate the correlations were based on the calibration sample described in Chapter 7:. Disattenuated correlations between were also computed, calculated based on the following formula:

$$r_{T_{xy}} = \frac{r_{xy}}{\sqrt{r_x r_y}}$$

where $r_{T_{xy}}$ is a corrected correlation for attenuation between scores *x* and *y*, r_{xy} is an observed correlation between the scores *x* and *y*, and r_x and r_y are reliabilities for *x* and *y*, respectively. Coefficient alphas (presented in Table 9.1, Table 9.2, and Table 9.3) were used to calculate the corrected correlation coefficients for attenuation. The disattenuated correlations could be greater than 1.00.

Table 9.7 – Table 9.11 present the test correlations and disattenuated correlations between the total raw score and the reporting category raw scores. The numbers in the lower diagonal of the table are the disattenuated correlations.

Table 9.7. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—ELA Form 1

Grade	Score	Total	Reading for Information	Reading for Literature	Writing and Language
3	Total	1.00	0.91	0.93	0.88
	Reading for Information	1.07	1.00	0.79	0.69
	Reading for Literature	1.06	0.98	1.00	0.73
	Writing and Language	1.02	0.87	0.88	1.00
4	Total	1.00	0.94	0.92	0.89
	Reading for Information	1.06	1.00	0.82	0.75
	Reading for Literature	1.06	0.99	1.00	0.72
	Writing and Language	1.02	0.90	0.88	1.00
5	Total	1.00	0.93	0.91	0.88
	Reading for Information	1.07	1.00	0.79	0.71
	Reading for Literature	1.06	0.98	1.00	0.71
	Writing and Language	1.01	0.87	0.88	1.00
6	Total	1.00	0.94	0.90	0.91
	Reading for Information	1.08	1.00	0.78	0.76
	Reading for Literature	1.07	0.98	1.00	0.73
	Writing and Language	1.04	0.92	0.91	1.00
7	Total	1.00	0.93	0.91	0.90
	Reading for Information	1.06	1.00	0.80	0.74
	Reading for Literature	1.07	0.99	1.00	0.71
	Writing and Language	1.04	0.90	0.88	1.00
8	Total	1.00	0.92	0.91	0.90
	Reading for Information	1.09	1.00	0.77	0.72
	Reading for Literature	1.09	0.99	1.00	0.71
	Writing and Language	1.05	0.91	0.90	1.00

Table 9.8. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—ELA Form 2

Grade	Score	Total	Reading for Information	Reading for Literature	Writing and Language
3	Total	1.00	0.91	0.94	0.89
	Reading for Information	1.07	1.00	0.79	0.69
	Reading for Literature	1.07	0.98	1.00	0.75
	Writing and Language	1.02	0.86	0.90	1.00
4	Total	1.00	0.94	0.93	0.88
	Reading for Information	1.06	1.00	0.82	0.74
	Reading for Literature	1.07	0.99	1.00	0.71
	Writing and Language	1.01	0.89	0.88	1.00
5	Total	1.00	0.93	0.91	0.88
	Reading for Information	1.07	1.00	0.79	0.71
	Reading for Literature	1.06	0.98	1.00	0.71
	Writing and Language	1.02	0.87	0.88	1.00
6	Total	1.00	0.94	0.90	0.91
	Reading for Information	1.08	1.00	0.79	0.77
	Reading for Literature	1.06	0.98	1.00	0.74
	Writing and Language	1.04	0.93	0.92	1.00
7	Total	1.00	0.93	0.91	0.89
	Reading for Information	1.06	1.00	0.79	0.73
	Reading for Literature	1.07	0.98	1.00	0.70
	Writing and Language	1.03	0.89	0.88	1.00
8	Total	1.00	0.92	0.91	0.90
	Reading for Information	1.09	1.00	0.77	0.72
	Reading for Literature	1.09	0.99	1.00	0.71
	Writing and Language	1.05	0.91	0.90	1.00

Table 9.9. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—Mathematics Grades 3–5

Grade	Score	Total	Operations, Algebraic Thinking, and Numbers in Base Ten	Numbers and Operations – Fractions	Measurement, Data, and Geometry
3	Total	1.00	0.97	0.87	0.93
	Operations, Algebraic Thinking, and Numbers in Base Ten	1.05	1.00	0.78	0.84
	Numbers and Operations – Fractions	1.02	0.93	1.00	0.76
	Measurement, Data, and Geometry	1.08	0.99	0.97	1.00
4	Total	1.00	0.97	0.93	0.85
	Operations, Algebraic Thinking, and Numbers in Base Ten	1.06	1.00	0.84	0.75
	Numbers and Operations – Fractions	1.04	0.97	1.00	0.74
	Measurement, Data, and Geometry	1.05	0.95	0.96	1.00
5	Total	1.00	0.94	0.91	0.90
	Operations, Algebraic Thinking, and Numbers in Base Ten	1.08	1.00	0.78	0.77
	Numbers & Operations – Fractions	1.07	0.97	1.00	0.74
	Measurement, Data, and Geometry	1.06	0.95	0.94	1.00

Table 9.10. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—Mathematics Grades 6 and 7

Grade	Score	Total	Ratio and Proportional Relationships	The Number System	Expressions and Equations	Geometry, Statistics and Probability
6	Total	1.00	0.90	0.93	0.93	0.82
	Ratio and Proportional Relationships	1.06	1.00	0.78	0.79	0.68
	The Number System	1.04	0.97	1.00	0.81	0.69
	Expressions and Equations	1.06	1.00	0.97	1.00	0.70
	Geometry, Statistics and Probability	1.05	0.97	0.93	0.96	1.00
7	Total	1.00	0.90	0.92	0.92	0.91
	Ratio and Proportional Relationships	1.07	1.00	0.78	0.77	0.76
	The Number System	1.05	0.99	1.00	0.80	0.78
	Expressions and Equations	1.06	0.99	0.99	1.00	0.77
	Geometry, Statistics and Probability	1.06	0.99	0.97	0.97	1.00

Table 9.11. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—Mathematics Grade 8

Grade	Score	Total	Expressions and Equations	Functions	Geometry	Statistics and Probability and The Number System
8	Total	1.00	0.94	0.85	0.86	0.88
	Expressions and Equations	1.07	1.00	0.73	0.76	0.76
	Functions	1.09	0.98	1.00	0.66	0.67
	Geometry	1.04	0.96	0.94	1.00	0.69
	Statistics and Probability and The Number System	1.09	0.98	0.98	0.95	1.00

9.4. Validity Evidence

According to the *Standards for Educational and Psychological Testing* (AERA et al., 2014), "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests" (p. 11). The purpose of test score validation is not to validate the test itself but to validate interpretations of the test scores for a particular purpose or use.

A validity argument should begin with clear statements regarding the purpose of a test and intended interpretations and uses of the test results. The purpose of the AASA tests is to assess the ELA and mathematics proficiency of students based on the Arizona Academic Standards. The objective of the proceeding sections is to highlight validity evidence for each aspect and to guide interested readers where to look for the evidence. Different aspects of validity evidence, which are in line with the *Standards* (AERA et al., 2014), are considered throughout this technical report. Providing validity evidence is an ongoing activity for any assessment as it matures.

9.4.1. Evidence Based on Test Content

Validity evidence based on test content refers to the extent to which a test is aligned with the construct the assessment is intended to measure (AERA et al., 2014, p. 14). AASA measures a student's level of ELA and mathematics proficiency based on the skills specified in the Arizona Academic Standards. Although the validity of AASA test score interpretations is evaluated along several dimensions as a criterion-referenced system of tests, the meaning of test scores is critically evaluated by the degree to which test content is aligned with the standards. The AASA ELA and Mathematics assessments are rigorously examined in accordance with the guidelines in the *Standards for Educational and Psychological Testing* (AERA et al., 2014). The Elementary and Secondary Education Act (ESEA) legislation also describes the evidence that is necessary to validate assessment scores for their intended purposes based on these standards.

Alignment of content standards is achieved through a rigorous, iterative test development process that proceeds from the content standards and begins with the item specifications and test blueprints, the core documents that ensure that the assessments are aligned to the Arizona Academic Standards. The item specifications define the content limit, model tasks, and response types for a specific standard, and the test blueprint defines the standards to be assessed for each test form, the number of items per standard, the number of item types, the number of points per item type, and the total number of items and points per test form. In addition to ensuring that test items are aligned with their intended content standards, each assessment is intended to measure a representative sample of the knowledge and skills identified in the standards. Thus, the blueprints also represent a policy document specifying the relative importance of content strands and standards in addition to meeting important measurement goals.

Once the item specifications and blueprints are established, item and test development can begin. It was a rigorous and iterative process involving the Pearson content team and ADE to ensure that the AASA assessments meet the test blueprints and other content criteria and psychometric targets, as described in Chapter 3:. Beyond the test blueprint, ADE and Pearson attempted to include items measuring different levels of rigor to cover the Arizona Academic Standards as much as possible.

Alignment of test forms to the test blueprints is a thoughtful, careful task that involves collaboration among assessment specialists, psychometricians, and ADE. Developing test forms is challenging because test blueprints can be highly complex, specifying not only the range of items and points for each reporting category and standard, but also cross-cutting criteria such as distribution across item types, DOK, writing genre, etc. In addition to meeting complex blueprint requirements, test developers worked to meet psychometric goals so that accommodated test forms measure equivalently across the range of student ability.

9.4.2. Evidence Based on Response Processes

Evidence based on response processes refers to the cognitive process engaged in by students when answering test items, or the "evidence concerning the fit between the construct and the detailed nature of performance or response actually engaged in by examinees" (AERA et al., 2014, p. 15). A standalone field test was administered in Spring 2022 for the ELA Writing test to increase the number of eligible writing prompts in the item bank for operational use in future administrations. New items were field tested in Grade 3 for to assess students Oral Reading Fluency (ORF), which is one of the reading foundation standards, in Spring 2022 and again in Spring 2023 and Spring 2024 to further explore their functioning and performance. The ORF items were designed to align with low, medium, and high levels of difficulty (based on Lexiles) and gauge students' ability to read aloud words.

As presented in Chapter 3:, all newly developed items also go through a rigorous item review process, including content, bias, and sensitivity committees with Arizona educators, parents, and community members. Reviewers evaluated the item for its alignment to the Arizona Science Standards, grade appropriateness, editorial completeness and accuracy, and the presence of any content that could be biased or sensitive in nature. Only the items accepted by the committees were considered eligible to be field tested.

9.4.3. Evidence Based on Internal Structure

Validity evidence based on internal structure refers to the extent to which an item or a component of a test ties to the assessment it is intended to measure (AERA et al., 2014, p. 16). AASA is designed to measure students' overall ELA and mathematics proficiency based on the Arizona Academic Standards, which are composed of various reporting categories for each content area. AASA items across all reporting categories were calibrated concurrently under the unidimensional Rasch models (Masters, 1982; Rasch, 1960) as indicated in Chapter 7:. To evaluate the unidimensionality assumption of the Rasch models, PCA was conducted for each operational form. The results of PCA analysis with the parallel analysis criterion (Horn, 1965) and Reckase's index (1979) presented in Table 7.3 indicated that there is one dominant dimension for both ELA and mathematics and the remaining components are non-significant.

Another assumption under the Rasch models is local item independence. The local item independence assumption is typically evaluated using Q3 statistics (Yen, 1984); Winsteps (Linacre, 2022b) produces raw score residual correlations for pairs of items on a test, which are analogous to the Q3 statistics. A distribution of the residual correlations by form, presented in Table 7.4, showed that most statistics are either slightly negative or slightly positive, which indicates that the item independence assumption generally holds for the AASA tests.

In addition to the total scale score, the scale score for each reporting category is reported individually. The scale scores for the reporting categories are generated by including the items associated with each reporting category and using the item parameter estimates from the concurrent calibration across all reporting categories. Details about scaling methods are described in Section 7.5. Correlations between the total score and reporting category score presented in Section 9.3 show that they are at least moderately correlated to each other, if not highly correlated, as expected.

A point-biserial correlation, as an indicator of interrelationship between an item and a construct that it is intended to measure, is calculated as a correlation between an item raw score and a total raw score. The point-biserial correlations should be higher than or equal to 0.25, as any item with a lower correlation is flagged during item selection. It is one of the psychometric criteria considered for item selection. The point-biserial correlation was calculated for distractors of multiple-choice items as well. Table 6.5 and Table 6.6 show that all the multiple-choice items have negative point-biserial correlations, except for a few distractors with a slightly positive correlation close to zero. The results indicate that the distractors work as expected.

Differential item functioning (DIF) analysis is a statistical method to detect potential bias of an item for (or against) a manifest group (e.g., female). DIF is defined as a difference between groups (e.g., male and female) in the probability of getting an item correct, given the same level of ability within the construct that an assessment is intended to measure. Details on DIF analysis are presented in Section 9.2. Items showing strong DIF are flagged for possible replacement in future administrations.

9.4.4. Evidence Based on Performance Standards

Validity evidence concerning performance standards refers to the extent to which passing scores are aligned to performance standards (Kane, 1994). Performance level descriptors (PLDs) highlight the knowledge, skills, and processes students possess at different performance levels (Egan et al., 2012). The PLDs are the foundation of standard setting meetings. The PLDs for AASA, provided on the ADE website at https://www.azed.gov/assessment/aasa, were drafted prior to the 2015 standard setting workshop and included educator input. ADE considered any need for clarification or revision that arose throughout the standard setting process prior to publishing the final versions (American Institutes for Research, 2015). See Section 10.1 for more details on standard setting.

9.4.5. Evidence Based on Relation to Other Variables

Validity evidence concerning a relation to other variables refers to the extent to which test scores are related to other external measures (AERA et al., 2014, p. 16). Because both the ELA and mathematics AASA assessments are administered to all eligible Arizona students, scores on the tests are expected to be positively correlated. Table 9.12 presents the correlation between AASA ELA and mathematics scale scores from the Spring 2024 administration. The correlations range from 0.73 to 0.80.

Table 9.12. Correlation between AASA ELA and Mathematics Scale Scores

Grade	N	Correlation
3	77,638	0.78
4	80,672	0.79
5	80,959	0.77
6	80,314	0.78
7	81,276	0.80
8	82,137	0.73

9.4.6. Summary

Overall, the validity evidence supports the use of AASA scores. The PCA revealed unidimensionality of AASA, which supports the use of unidimensional Rasch models. The AASA ELA and mathematics scores were also positively correlated. Test score validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization and continuing throughout the entire assessment process. Additional evidence should and will be added to the AASA technical report in the future, as appropriate.

Chapter 10: CLASSIFICATION INTO PERFORMANCE LEVELS

This chapter provides information regarding classification of students into performance levels for the Spring 2024 AASA assessments, addressing Standards 1.8, 1.9, 2.13, 2.14, 2.16, 5.5, 5.21, 5.22, 5.23, and 7.4 (AERA et al., 2014).

Scores from the AASA tests are used to classify students into one of four performance levels: *Minimally Proficient, Partially Proficient, Proficient,* and *Highly Proficient*. This section provides information regarding classification of students into these four categories, including the consistency and accuracy with which students who took the Spring 2024 AASA assessment were assigned to the performance levels.

10.1. Standard Setting

Arizona educators made recommendations for cut scores for each performance level on the AASA assessments during the standard setting workshop conducted from July 13–16, 2015, following the first operational administration of the AASA in Spring 2015 (known as the AzMERIT assessments at that time) using the bookmark standard setting procedure. The State Board of Education adopted the panelist-recommended performance standards on August 14, 2015. See the standard setting report for a detailed account of the workshop process and outcomes (American Institutes for Research, 2015).

Table 10.1 presents the final scale score ranges for the AASA performance levels, and Table 10.2 presents the scale score and associated CSEM at the performance level cuts. The CSEM is very similar across all grades and content areas within each cut.

Table 10.1. Performance Level Cut Scores

Assessment	*Minimally Proficient*	*Partially Proficient*	*Proficient*	*Highly Proficient*
ELA 3	2395–2496	2497–2508	2509–2540	2541–2605
ELA 4	2400–2509	2510–2522	2523–2558	2559–2610
ELA 5	2419–2519	2520–2542	2543–2577	2578–2629
ELA 6	2431–2531	2532–2552	2553–2596	2597–2641
ELA 7	2438–2542	2543–2560	2561–2599	2600–2648
ELA 8	2448–2550	2551–2571	2572–2603	2604–2658
Mathematics 3	3395–3494	3495–3530	3531–3572	3573–3605
Mathematics 4	3435–3529	3530–3561	3562–3605	3606–3645
Mathematics 5	3478–3562	3563–3594	3595–3634	3635–3688
Mathematics 6	3512–3601	3602–3628	3629–3662	3663–3722
Mathematics 7	3529–3628	3629–3651	3652–3679	3680–3739
Mathematics 8	3566–3649	3650–3672	3673–3704	3705–3776

Note. The scale score cut for Move on When Reading (MOWR) in Grade 3 is 2446.

Table 10.2. CSEM at Performance Level Cuts

	Partially Proficient Cut		*Proficient* Cut		*Highly Proficient* Cut	
Assessment	Scale Score	CSEM	Scale Score	CSEM	Scale Score	CSEM
ELA 3, Form 1	2497	9	2509	9	2541	11
ELA 4, Form 1	2510	9	2523	9	2559	11
ELA 5, Form 1	2520	9	2543	9	2578	12
ELA 6, Form 1	2532	9	2553	9	2597	12
ELA 7, Form 1	2543	9	2561	9	2600	12
ELA 8, Form 1	2551	9	2572	9	2604	11
ELA 3, Form 2	2497	9	2509	9	2541	11
ELA 4, Form 2	2510	9	2523	9	2559	11
ELA 5, Form 2	2520	9	2543	9	2578	12
ELA 6, Form 2	2532	9	2553	9	2597	13
ELA 7, Form 2	2543	9	2561	9	2600	12
ELA 8, Form 2	2551	9	2572	9	2604	11
Mathematics 3	3495	10	3531	11	3573	14
Mathematics 4	3530	10	3562	10	3606	14
Mathematics 5	3563	10	3595	10	3635	12
Mathematics 6	3602	10	3629	10	3663	11
Mathematics 7	3629	10	3652	10	3680	11
Mathematics 8	3650	10	3673	10	3705	11

Performance classifications for reporting categories are determined by student performance on the reporting categories compared to the respective *Proficient* performance standard. For each reporting category, a mid-range band is established by extending one CSEM below and above the *Proficient* performance standard scale score cut. If a student's scale score for a reporting category is fallen into the mid-range band, the student performance is classified as *At/Near Mastery* for the reporting category. On the other hand, if a student's scale score is above or below the mid-range band, the student performance is classified as *Above Mastery* or *Below Mastery*, respectively.

10.2. Classification Consistency and Accuracy

Classification consistency is the agreement between students' performance level classification from two independent administrations of the same test (or two parallel forms of the test). Classification accuracy refers to the agreement between the actual classifications using observed cut scores and true classifications based on known true cut scores (Livingston & Lewis, 1995).

In conjunction with internal consistency, classification consistency is an important type of reliability and is particularly relevant to high-stakes decisions, such as passing or not passing the AASA tests. As a form of reliability, classification consistency represents how reliably students can be classified into performance levels. For tests such as AASA, classification consistency is most important for students whose ability is near the *Proficient* cut score. Students whose ability is far above or far below the value established for *Proficient* are unlikely to be misclassified because repeated administration of the test will nearly always result in the same classification. Students whose true scores are close to the cut score are a more serious concern. These students' true scores will likely lie within the SEM of the cut score. For this reason, the measurement error at the cut scores should be considered when evaluating the classification consistency of a test.

Classification consistency and accuracy were estimated using the total scale score for the *Proficient* cut based on procedures described by Livingston and Lewis (1995). Classification consistency is calculated as the proportion of students in the diagonal in Table 10.3 (i.e., students classified consistently between two parallel forms, listed in bold). Similarly, classification accuracy is calculated as the proportion of students in the diagonal in Table 10.4 (i.e., students classified the same between observed scores and true scores, listed in bold).

Table 10.3. Classification Consistency for the *Proficient* Cut

		Expected Performance on Parallel Form	
		Not Proficient	*Proficient*
Observed Performance on Actual Form	*Not Proficient*	**Consistent Classification**	Inconsistent Classification
	Proficient	Inconsistent Classification	**Consistent Classification**

Table 10.4. Classification Accuracy for the *Proficient* Cut

		Expected Performance on Test	
		Not Proficient	*Proficient*
Observed Performance on Test	*Not Proficient*	**Accurate Classification**	False Negative
	Proficient	False Positive	**Accurate Classification**

Cohen's kappa (κ) coefficient (Cohen, 1960) is another way of expressing overall consistency. This statistic assesses the proportion of consistent classification expected beyond chance and is therefore most often lower than the unadjusted value of overall consistency. Cohen's kappa is calculated as follows:

$$\kappa = \frac{P - P_c}{1 - P_c}$$

where P_c is the probability of consistent classification by chance, and P is the probability of consistent classification (unadjusted by chance). Students can be misclassified in one of two ways. Students who are truly not *Proficient* but were classified as being *Proficient*, based on the assessment, are false positives. Similarly, students who are truly *Proficient* but were classified as being not *Proficient* are false negatives.

Table 10.5 presents the classification consistency and accuracy results, generated by BB-class (Brennan, 2004). These results are for classifying students into four performance levels using the total score on the assessment for students in the calibration sample. Included in the table are the sample size (N), classification consistency (Consistency), classification inconsistency (Inconsistency), probability of consistent classification by chance (Chance), Cohen's Kappa (κ), classification accuracy (Accuracy), false positive (False Positive), and false negative (False Negative). Inconsistency is defined as one minus Consistency.

Table 10.5. Classification Consistency and Accuracy Results

Assessment	N	Consistency	Inconsistency	Chance	κ	Accuracy	False Positive	False Negative
ELA 3, Form 1	37,054	0.75	0.25	0.35	0.61	0.81	0.10	0.09
ELA 4, Form 1	38,425	0.74	0.26	0.30	0.63	0.81	0.10	0.09
ELA 5, Form 1	38,738	0.71	0.29	0.30	0.59	0.79	0.11	0.10
ELA 6, Form 1	38,310	0.72	0.28	0.32	0.60	0.80	0.11	0.09
ELA 7, Form 1	39,039	0.72	0.28	0.31	0.59	0.79	0.11	0.09
ELA 8, Form 1	39,289	0.71	0.29	0.31	0.58	0.79	0.12	0.09
ELA 3, Form 2	36,963	0.75	0.25	0.36	0.61	0.81	0.10	0.09
ELA 4, Form 2	38,544	0.74	0.26	0.31	0.62	0.81	0.10	0.09
ELA 5, Form 2	38,725	0.71	0.29	0.30	0.59	0.79	0.11	0.09
ELA 6, Form 2	38,347	0.72	0.28	0.32	0.59	0.79	0.12	0.09
ELA 7, Form 2	38,753	0.72	0.28	0.32	0.59	0.79	0.11	0.09
ELA 8, Form 2	39,705	0.72	0.28	0.32	0.58	0.79	0.11	0.09
Mathematics 3	75,267	0.73	0.27	0.27	0.63	0.81	0.10	0.09
Mathematics 4	77,793	0.76	0.24	0.28	0.67	0.83	0.09	0.08
Mathematics 5	78,156	0.74	0.26	0.29	0.63	0.81	0.10	0.08
Mathematics 6	77,439	0.78	0.22	0.34	0.67	0.84	0.09	0.07
Mathematics 7	78,660	0.79	0.21	0.38	0.67	0.85	0.08	0.07
Mathematics 8	79,736	0.78	0.22	0.38	0.65	0.84	0.09	0.07

10.3. MOWR Policy

Arizona's Move On When Reading (MOWR) policy is designed to provide students with evidence-based, effective reading instruction in Grades K–3 to position them for success as they progress through school, college, and career. The heart of the legislation emphasizes early identification and immediate intervention for struggling readers. Grade 3 students must meet the MOWR cut score of 2446 on the AASA ELA Reading portion, as established by the State Board of Education, to be promoted to Grade 4, with some exemptions. Students who are retained receive an extra year of specialized support so they are ready to enter Grade 4 as strong readers. For more information, refer to the ADE website at https://www.azed.gov/mowr/.

REFERENCES

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing*. AERA.

American Institutes for Research (AIR). (2015). *Recommending AzMERIT performance standards: English language arts grades 3–11, math grades 3–8, Algebra I, Geometry, and Algebra II*. https://www.azed.gov/sites/default/files/2016/12/spring-2015-azmerit-standard-setting_091415.pdf?id=5846d5b4aadebe0cf0337f5e

Angoff, W. (1993). Perspective on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3–24). Lawrence Erlbaum Associates.

Brennan, R. L. (2004). BB-CLASS: *A computer program that uses the beta-binomial model for classification consistency and accuracy [computer software] (Version 1.0)*. University of Iowa.

Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement, 41,* 687–699.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Sage.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46. http://dx.doi.org/10.1177/001316446002000104

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin, 70,* 213.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice, 17,* 31–44.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart, and Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 12*, 671–684.

Dorans, N. J., & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach.* ETS Research Report 91-47. Educational Testing Service.

Egan, K. A., Schneider, C., & Ferrara, S. (2012). Performance level descriptors: History, practice, and a proposed work. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 79–106). Routledge.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Erlbaum.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55,* 377–393.

Green, D. R. (1975, December). *Procedures for assessing bias in achievement tests.* Presented at the National Institute of Education Conference on Test Bias, Annapolis, MD.

Holland, P. W., & Thayer, D. T. (1988). Differential item functioning and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Lawrence Erlbaum Associates.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.

Humphreys, L. G., & Montanelli, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research, 10,* 193–206.

Jackson, D. A. (1993). Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology, 74*(8), 2204–2214.

Kane, M. T. (1994). Validating interpretive arguments for licensure and certification examinations. *Evaluation & the Health Professions*, *17*, 133–159.

Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2015). *Handbook of test development*. Routledge.

Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research, and Evaluation, 12,* 2.

Linacre, J. M. (2022a). *Winsteps® Rasch measurement computer program user's guide, Version 4.8.1.0.* Winsteps.com.

Linacre, J. M. (2022b). *Winsteps® (Version 4.8.1.0)* [Computer Software]. http://www.winsteps.com/

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.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149–174.

Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1992). The effect of sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement, 52*(2), 443–451. https://doi.org/10.1177/0013164492052002020

Miller, E, G., Ourania, R., & Twing, J, S. (2004). Evaluation of the 0.3 logits screening criterion in common item equating. *Journal of Applied Measurement, 5*, 172–177.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Danmarks Paedogogiske Institut.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*(3), 207–230.

Stout, W. F. (1990). A new item response theory modelling approach and applications to unidimensionality assessment and ability estimation. *Psychometrika, 55*, 293–325.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25*(1), 1–28.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*, 370.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Mesa Press.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement, 8*(2), 125–145.

Zumbo, B. D. (1999*). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Directorate of Human Resources Research and Evaluation, Department of National Defense.

Zwick, R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 26,* 44–66.

Appendix A: ITEM-LEVEL CTT STATISTICS

This appendix includes the following item-level CTT results:

- Table A.1 – Table A.12 present the item-level CTT statistics for each content area and grade, including item type, maximum number of points possible, number of students (N), *p*-value, and the point-biserial correlation between an item and total raw score.
- Table A.13 – Table A.24 present the item-level distractor analysis for multiple-choice items, including the percentage of students who selected the correct and incorrect response options, the point-biserial correlation associated with each option, and the overall omission rate for the item.

Table A.1. Item-Level CTT Statistics, ELA Grade 3

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	OE	4	37,054	0.14	0.68
2	OE	4	37,054	0.13	0.66
3	OE	2	37,054	0.61	0.69
4	OE	4	36,963	0.23	0.73
5	OE	4	36,963	0.19	0.70
6	OE	2	36,963	0.61	0.72
7	MC	1	74,077	0.51	0.52
8	MX	1	74,077	0.51	0.59
9	MC	1	74,077	0.77	0.53
10	MC	1	74,077	0.50	0.42
11	MC	1	74,077	0.46	0.41
12	MC	1	74,077	0.40	0.32
13	MC	1	74,077	0.60	0.57
14	MC	1	74,077	0.27	0.42
15	MC	1	74,077	0.59	0.53
16	MX	1	74,077	0.23	0.35
17	MC	1	74,077	0.33	0.27
18	XI	1	74,077	0.57	0.47
19	MC	1	74,077	0.67	0.57
20	MX	2	74,077	0.54	0.44
21	MX	2	74,077	0.66	0.55
22	MC	1	74,077	0.70	0.47
23	MC	1	74,077	0.44	0.45
24	MC	1	74,077	0.47	0.39
25	MC	1	74,077	0.44	0.45
26	MC	1	74,077	0.43	0.46
27	MC	1	74,077	0.70	0.52
28	MC	1	74,077	0.53	0.52
29	XI	1	74,077	0.58	0.58
30	MC	1	74,077	0.52	0.39
31	MC	1	74,077	0.44	0.42
32	MX	1	74,077	0.38	0.43
33	MC	1	74,077	0.64	0.54

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
34	MC	1	74,077	0.50	0.41
35	MC	1	74,077	0.32	0.28
36	MC	1	74,077	0.44	0.39
37	MC	1	74,077	0.42	0.24
38	MX	1	74,077	0.37	0.53
39	MC	1	74,077	0.63	0.53
40	MX	1	74,077	0.22	0.51
41	MC	1	74,077	0.54	0.35
42	XI	1	74,077	0.36	0.60
43	MC	1	74,077	0.47	0.53
44	MC	1	74,077	0.50	0.57
45	MC	1	74,077	0.50	0.45
46	MX	2	74,077	0.58	0.57
47	MX	2	74,077	0.57	0.45

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.2. Item-Level CTT Statistics, ELA Grade 4

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	OE	4	38,425	0.23	0.71
2	OE	4	38,425	0.22	0.70
3	OE	2	38,425	0.64	0.67
4	OE	4	38,544	0.19	0.68
5	OE	4	38,544	0.17	0.65
6	OE	2	38,544	0.73	0.66
7	MC	1	77,076	0.76	0.56
8	MC	1	77,076	0.54	0.58
9	MC	1	77,076	0.57	0.42
10	MC	1	77,076	0.69	0.51
11	MC	1	77,076	0.74	0.49
12	MX	1	77,076	0.57	0.59
13	MC	1	77,076	0.50	0.37
14	MC	1	77,076	0.62	0.56
15	XI	1	77,076	0.54	0.60
16	MC	1	77,076	0.80	0.52
17	MX	1	77,076	0.33	0.49
18	XI	1	77,076	0.35	0.48
19	MC	1	77,076	0.43	0.49
20	MC	1	77,076	0.47	0.48
21	MX	2	77,076	0.56	0.54
22	MX	2	77,076	0.50	0.50
23	MC	1	77,076	0.62	0.39
24	MC	1	77,076	0.78	0.50
25	MX	1	77,076	0.28	0.40
26	MC	1	77,076	0.66	0.45
27	MC	1	77,076	0.32	0.46

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
28	MC	1	77,076	0.77	0.51
29	MC	1	77,076	0.60	0.41
30	MX	2	77,076	0.49	0.59
31	MC	1	77,076	0.45	0.33
32	MC	1	77,076	0.33	0.37
33	MC	1	77,076	0.47	0.48
34	MC	1	77,076	0.69	0.59
35	MC	1	77,076	0.46	0.63
36	MC	1	77,076	0.59	0.48
37	MC	1	77,076	0.43	0.33
38	MC	1	77,076	0.27	0.47
39	MC	1	77,076	0.71	0.58
40	MC	1	77,076	0.26	0.30
41	MC	1	77,076	0.28	0.40
42	MC	1	77,076	0.46	0.39
43	XI	1	77,076	0.38	0.53
44	MC	1	77,076	0.67	0.51
45	MC	1	77,076	0.75	0.53
46	MX	2	77,076	0.53	0.53
47	MX	2	77,076	0.72	0.52

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.3. Item-Level CTT Statistics, ELA Grade 5

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	OE	4	38,738	0.31	0.70
2	OE	4	38,738	0.28	0.70
3	OE	2	38,738	0.70	0.71
4	OE	4	38,725	0.32	0.72
5	OE	4	38,725	0.30	0.68
6	OE	2	38,725	0.76	0.69
7	MC	1	77,580	0.84	0.48
8	MC	1	77,580	0.57	0.51
9	MC	1	77,580	0.68	0.57
10	MC	1	77,580	0.55	0.48
11	MC	1	77,580	0.50	0.48
12	MC	1	77,580	0.60	0.52
13	XI	1	77,580	0.49	0.55
14	MC	1	77,580	0.53	0.41
15	MC	1	77,580	0.66	0.35
16	MX	1	77,580	0.56	0.62
17	MC	1	77,580	0.60	0.45
18	MX	1	77,580	0.43	0.44
19	MC	1	77,580	0.42	0.35
20	MC	1	77,580	0.57	0.59
21	MC	1	77,580	0.31	0.36

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
22	MC	1	77,580	0.60	0.44
23	MX	2	77,580	0.54	0.56
24	MX	2	77,580	0.68	0.55
25	MC	1	77,580	0.62	0.43
26	MC	1	77,580	0.23	0.47
27	MC	1	77,580	0.48	0.43
28	MC	1	77,580	0.66	0.37
29	MC	1	77,580	0.63	0.47
30	MC	1	77,580	0.59	0.43
31	MC	1	77,580	0.48	0.35
32	MC	1	77,580	0.45	0.28
33	MC	1	77,580	0.50	0.44
34	MC	1	77,580	0.60	0.57
35	MC	1	77,580	0.49	0.36
36	MX	1	77,580	0.52	0.46
37	MC	1	77,580	0.76	0.56
38	XI	1	77,580	0.32	0.39
39	MC	1	77,580	0.46	0.60
40	MX	1	77,580	0.22	0.46
41	MC	1	77,580	0.55	0.50
42	MC	1	77,580	0.62	0.53
43	MC	1	77,580	0.39	0.26
44	MC	1	77,580	0.55	0.44
45	MC	1	77,580	0.70	0.40
46	MX	2	77,580	0.60	0.59
47	MX	2	77,580	0.41	0.37

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.4. Item-Level CTT Statistics, ELA Grade 6

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	OE	4	38,310	0.40	0.73
2	OE	4	38,310	0.35	0.71
3	OE	2	38,310	0.74	0.70
4	OE	4	38,347	0.43	0.76
5	OE	4	38,347	0.35	0.72
6	OE	2	38,347	0.73	0.68
7	MC	1	76,717	0.63	0.47
8	MC	1	76,717	0.47	0.39
9	MC	1	76,717	0.62	0.49
10	MC	1	76,717	0.81	0.44
11	MC	1	76,717	0.64	0.38
12	MC	1	76,717	0.65	0.49
13	MC	1	76,717	0.45	0.35
14	MX	1	76,717	0.15	0.33
15	MC	1	76,717	0.63	0.53

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
16	MC	1	76,717	0.40	0.36
17	MC	1	76,717	0.74	0.54
18	MC	1	76,717	0.39	0.56
19	XI	1	76,717	0.20	0.25
20	MC	1	76,717	0.58	0.37
21	MC	1	76,717	0.72	0.43
22	MX	2	76,717	0.65	0.63
23	MX	2	76,717	0.33	0.39
24	MC	1	76,717	0.63	0.30
25	MC	1	76,717	0.66	0.60
26	MC	1	76,717	0.42	0.39
27	MC	1	76,717	0.57	0.44
28	MC	1	76,717	0.45	0.32
29	MX	1	76,717	0.45	0.49
30	MC	1	76,717	0.64	0.53
31	MC	1	76,717	0.63	0.47
32	MC	1	76,717	0.38	0.44
33	MC	1	76,717	0.53	0.44
34	MC	1	76,717	0.53	0.49
35	MC	1	76,717	0.83	0.52
36	MC	1	76,717	0.68	0.64
37	MX	1	76,717	0.61	0.39
38	MC	1	76,717	0.48	0.44
39	MX	1	76,717	0.36	0.54
40	MC	1	76,717	0.44	0.39
41	MX	1	76,717	0.52	0.68
42	MC	1	76,717	0.64	0.48
43	MX	1	76,717	0.41	0.54
44	MC	1	76,717	0.38	0.29
45	MC	1	76,717	0.67	0.63
46	MX	2	76,717	0.50	0.50
47	MX	2	76,717	0.53	0.61

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.5. Item-Level CTT Statistics, ELA Grade 7

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	OE	4	39,039	0.39	0.77
2	OE	4	39,039	0.35	0.75
3	OE	2	39,039	0.73	0.71
4	OE	4	38,753	0.37	0.73
5	OE	4	38,753	0.31	0.74
6	OE	2	38,753	0.73	0.70
7	MC	1	77,917	0.64	0.48
8	MC	1	77,917	0.72	0.60
9	MC	1	77,917	0.46	0.35

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
10	MC	1	77,917	0.49	0.55
11	MC	1	77,917	0.52	0.30
12	MC	1	77,917	0.58	0.52
13	XI	1	77,917	0.42	0.49
14	MC	1	77,917	0.62	0.39
15	MC	1	77,917	0.67	0.55
16	MC	1	77,917	0.55	0.47
17	MC	1	77,917	0.62	0.48
18	MC	1	77,917	0.42	0.24
19	MC	1	77,917	0.47	0.30
20	MC	1	77,917	0.74	0.51
21	MX	2	77,917	0.80	0.61
22	MX	2	77,917	0.63	0.55
23	MC	1	77,917	0.69	0.58
24	MX	1	77,917	0.49	0.61
25	MC	1	77,917	0.47	0.47
26	MC	1	77,917	0.58	0.38
27	MC	1	77,917	0.53	0.40
28	MC	1	77,917	0.30	0.35
29	MC	1	77,917	0.53	0.39
30	MC	1	77,917	0.58	0.41
31	MC	1	77,917	0.36	0.54
32	MC	1	77,917	0.44	0.36
33	MC	1	77,917	0.50	0.62
34	MC	1	77,917	0.65	0.47
35	MC	1	77,917	0.62	0.58
36	MC	1	77,917	0.60	0.46
37	MC	1	77,917	0.64	0.60
38	XI	1	77,917	0.24	0.32
39	MC	1	77,917	0.58	0.43
40	MC	1	77,917	0.55	0.39
41	MC	1	77,917	0.52	0.46
42	MC	1	77,917	0.61	0.54
43	MC	1	77,917	0.41	0.47
44	MC	1	77,917	0.47	0.36
45	MC	1	77,917	0.49	0.35
46	MX	2	77,917	0.44	0.44
47	MX	2	77,917	0.50	0.37

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.6. Item-Level CTT Statistics, ELA Grade 8

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	OE	4	39,289	0.47	0.72
2	OE	4	39,289	0.31	0.64
3	OE	2	39,289	0.83	0.62
4	OE	4	39,705	0.39	0.71
5	OE	4	39,705	0.33	0.71
6	OE	2	39,705	0.79	0.66
7	MC	1	79,101	0.66	0.54
8	XI	1	79,101	0.66	0.47
9	MC	1	79,101	0.58	0.37
10	MC	1	79,101	0.66	0.40
11	MX	1	79,101	0.66	0.52
12	MC	1	79,101	0.64	0.48
13	MC	1	79,101	0.34	0.22
14	MC	1	79,101	0.45	0.39
15	MC	1	79,101	0.43	0.37
16	MC	1	79,101	0.40	0.35
17	MC	1	79,101	0.36	0.34
18	MX	1	79,101	0.35	0.35
19	MC	1	79,101	0.73	0.45
20	MX	2	79,101	0.64	0.42
21	MX	2	79,101	0.71	0.59
22	MC	1	79,101	0.79	0.50
23	MX	1	79,101	0.28	0.41
24	MC	1	79,101	0.47	0.29
25	MC	1	79,101	0.44	0.49
26	MC	1	79,101	0.58	0.45
27	MC	1	79,101	0.35	0.29
28	MC	1	79,101	0.50	0.27
29	MC	1	79,101	0.70	0.41
30	MX	1	79,101	0.25	0.37
31	MC	1	79,101	0.59	0.52
32	XI	1	79,101	0.49	0.55
33	MC	1	79,101	0.52	0.52
34	MC	1	79,101	0.38	0.35
35	MC	1	79,101	0.48	0.41
36	MC	1	79,101	0.52	0.46
37	MC	1	79,101	0.45	0.39
38	MC	1	79,101	0.28	0.24
39	MC	1	79,101	0.57	0.52
40	MC	1	79,101	0.59	0.57
41	MX	1	79,101	0.37	0.52
42	MC	1	79,101	0.54	0.48
43	MC	1	79,101	0.55	0.48
44	MC	1	79,101	0.82	0.51
45	MC	1	79,101	0.68	0.53

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
46	MX	2	79,101	0.42	0.43
47	MX	2	79,101	0.49	0.58

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.7. Item-Level CTT Statistics, Mathematics Grade 3

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	MC	1	75,267	0.87	0.33
2	XI	1	75,267	0.76	0.62
3	XI	1	75,267	0.50	0.54
4	XI	1	75,267	0.33	0.61
5	XI	1	75,267	0.65	0.62
6	MC	1	75,267	0.49	0.61
7	XI	1	75,267	0.67	0.57
8	MX	1	75,267	0.34	0.47
9	XI	1	75,267	0.70	0.54
10	MC	1	75,267	0.66	0.63
11	XI	1	75,267	0.37	0.51
12	MC	1	75,267	0.42	0.55
13	MC	1	75,267	0.36	0.27
14	MC	1	75,267	0.59	0.60
15	MC	1	75,267	0.60	0.37
16	MC	1	75,267	0.39	0.50
17	MC	1	75,267	0.57	0.54
18	MC	1	75,267	0.92	0.43
19	XI	1	75,267	0.60	0.64
20	MC	1	75,267	0.57	0.40
21	MC	1	75,267	0.71	0.62
22	MC	1	75,267	0.80	0.53
23	MC	1	75,267	0.63	0.43
24	MC	1	75,267	0.77	0.42
25	MC	1	75,267	0.34	0.62
26	XI	1	75,267	0.69	0.65
27	XI	1	75,267	0.79	0.57
28	XI	1	75,267	0.29	0.47
29	MC	1	75,267	0.57	0.56
30	XI	1	75,267	0.50	0.36
31	MX	1	75,267	0.22	0.49
32	MC	1	75,267	0.41	0.54
33	MC	1	75,267	0.79	0.52
34	XI	1	75,267	0.49	0.65
35	XI	1	75,267	0.72	0.61
36	XI	1	75,267	0.78	0.63
37	MC	1	75,267	0.73	0.58
38	MC	1	75,267	0.46	0.46
39	XI	1	75,267	0.53	0.67

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
40	MC	1	75,267	0.18	0.43
41	MC	1	75,267	0.70	0.54
42	XI	1	75,267	0.41	0.57
43	MC	1	75,267	0.34	0.63
44	MC	1	75,267	0.59	0.67
45	MC	1	75,267	0.56	0.57

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.8. Item-Level CTT Statistics, Mathematics Grade 4

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	MC	1	77,793	0.79	0.47
2	MX	1	77,793	0.59	0.47
3	XI	1	77,793	0.38	0.66
4	MC	1	77,793	0.59	0.50
5	MC	1	77,793	0.40	0.37
6	MX	1	77,793	0.32	0.57
7	MC	1	77,793	0.76	0.46
8	MC	1	77,793	0.69	0.48
9	XI	1	77,793	0.59	0.58
10	MC	1	77,793	0.56	0.58
11	MC	1	77,793	0.52	0.49
12	MC	1	77,793	0.54	0.32
13	XI	1	77,793	0.66	0.62
14	MC	1	77,793	0.22	0.53
15	MC	1	77,793	0.45	0.50
16	XI	1	77,793	0.48	0.64
17	XI	1	77,793	0.37	0.61
18	MC	1	77,793	0.42	0.46
19	XI	1	77,793	0.66	0.58
20	XI	1	77,793	0.70	0.58
21	MC	1	77,793	0.60	0.57
22	MC	1	77,793	0.50	0.34
23	XI	1	77,793	0.70	0.58
24	MC	1	77,793	0.81	0.46
25	XI	1	77,793	0.27	0.59
26	MC	1	77,793	0.58	0.50
27	MC	1	77,793	0.74	0.50
28	XI	1	77,793	0.25	0.56
29	MC	1	77,793	0.70	0.51
30	XI	1	77,793	0.42	0.63
31	XI	1	77,793	0.72	0.43
32	MC	1	77,793	0.48	0.59
33	MC	1	77,793	0.36	0.37
34	MC	1	77,793	0.15	0.42
35	XI	1	77,793	0.23	0.50

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
36	MC	1	77,793	0.59	0.41
37	MC	1	77,793	0.76	0.52
38	XI	1	77,793	0.55	0.69
39	MC	1	77,793	0.32	0.63
40	XI	1	77,793	0.33	0.61
41	XI	1	77,793	0.74	0.62
42	MC	1	77,793	0.34	0.51
43	XI	1	77,793	0.51	0.53
44	MC	1	77,793	0.51	0.45
45	XI	1	77,793	0.36	0.69

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.9. Item-Level CTT Statistics, Mathematics Grade 5

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	XI	1	78,156	0.74	0.45
2	XI	1	78,156	0.55	0.56
3	MC	1	78,156	0.59	0.43
4	MC	1	78,156	0.53	0.44
5	MX	1	78,156	0.32	0.51
6	XI	1	78,156	0.40	0.63
7	MC	1	78,156	0.39	0.38
8	MC	1	78,156	0.29	0.50
9	MC	1	78,156	0.52	0.37
10	MC	1	78,156	0.54	0.52
11	MC	1	78,156	0.60	0.52
12	MC	1	78,156	0.51	0.36
13	XI	1	78,156	0.52	0.53
14	MX	1	78,156	0.40	0.50
15	XI	1	78,156	0.49	0.57
16	XI	1	78,156	0.28	0.58
17	MC	1	78,156	0.65	0.54
18	MC	1	78,156	0.46	0.34
19	XI	1	78,156	0.34	0.64
20	MC	1	78,156	0.50	0.41
21	XI	1	78,156	0.68	0.53
22	MC	1	78,156	0.33	0.45
23	MC	1	78,156	0.52	0.47
24	MC	1	78,156	0.43	0.64
25	MC	1	78,156	0.63	0.38
26	XI	1	78,156	0.25	0.59
27	MC	1	78,156	0.61	0.36
28	XI	1	78,156	0.58	0.40
29	MC	1	78,156	0.55	0.57
30	MX	1	78,156	0.40	0.38
31	MC	1	78,156	0.26	0.43

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
32	MC	1	78,156	0.46	0.53
33	XI	1	78,156	0.47	0.40
34	XI	1	78,156	0.32	0.52
35	XI	1	78,156	0.18	0.56
36	MC	1	78,156	0.66	0.47
37	XI	1	78,156	0.22	0.63
38	MC	1	78,156	0.48	0.44
39	MC	1	78,156	0.45	0.37
40	XI	1	78,156	0.40	0.45
41	MC	1	78,156	0.39	0.29
42	MC	1	78,156	0.74	0.52
43	MC	1	78,156	0.31	0.34
44	XI	1	78,156	0.45	0.59
45	MC	1	78,156	0.42	0.42

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.10. Item-Level CTT Statistics, Mathematics Grade 6

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	MC	1	77,439	0.73	0.48
2	MC	1	77,439	0.53	0.45
3	XI	1	77,439	0.44	0.61
4	MC	1	77,439	0.65	0.28
5	MC	1	77,439	0.66	0.42
6	MC	1	77,439	0.50	0.67
7	XI	1	77,439	0.25	0.62
8	XI	1	77,439	0.40	0.58
9	MC	1	77,439	0.36	0.47
10	MC	1	77,439	0.33	0.33
11	MC	1	77,439	0.50	0.42
12	MC	1	77,439	0.36	0.36
13	MC	1	77,439	0.41	0.53
14	MC	1	77,439	0.28	0.33
15	MC	1	77,439	0.40	0.29
16	XI	1	77,439	0.26	0.54
17	MC	1	77,439	0.36	0.34
18	MC	1	77,439	0.25	0.48
19	XI	1	77,439	0.28	0.66
20	MC	1	77,439	0.42	0.42
21	XI	1	77,439	0.19	0.50
22	XI	1	77,439	0.38	0.62
23	MC	1	77,439	0.48	0.46
24	MC	1	77,439	0.70	0.47
25	XI	1	77,439	0.24	0.61
26	XI	1	77,439	0.52	0.61
27	XI	1	77,439	0.54	0.64

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
28	MC	1	77,439	0.31	0.53
29	MC	1	77,439	0.43	0.55
30	MX	1	77,439	0.60	0.58
31	MC	1	77,439	0.45	0.41
32	MC	1	77,439	0.45	0.28
33	XI	1	77,439	0.36	0.47
34	XI	1	77,439	0.33	0.66
35	XI	1	77,439	0.48	0.60
36	XI	1	77,439	0.32	0.62
37	XI	1	77,439	0.54	0.63
38	MC	1	77,439	0.35	0.53
39	XI	1	77,439	0.54	0.46
40	MC	1	77,439	0.33	0.51
41	XI	1	77,439	0.18	0.47
42	MC	1	77,439	0.45	0.55
43	XI	1	77,439	0.35	0.59
44	XI	1	77,439	0.16	0.57
45	MC	1	77,439	0.72	0.56
46	XI	1	77,439	0.37	0.65
47	MC	1	77,439	0.59	0.45

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.11. Item-Level CTT Statistics, Mathematics Grade 7

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	MC	1	78,660	0.70	0.54
2	MC	1	78,660	0.53	0.38
3	XI	1	78,660	0.38	0.62
4	MC	1	78,660	0.39	0.36
5	XI	1	78,660	0.25	0.58
6	MC	1	78,660	0.54	0.36
7	XI	1	78,660	0.37	0.66
8	MC	1	78,660	0.20	0.41
9	XI	1	78,660	0.16	0.51
10	XI	1	78,660	0.48	0.62
11	MC	1	78,660	0.38	0.38
12	XI	1	78,660	0.36	0.56
13	MC	1	78,660	0.63	0.51
14	MC	1	78,660	0.52	0.38
15	XI	1	78,660	0.34	0.71
16	MC	1	78,660	0.36	0.46
17	XI	1	78,660	0.26	0.65
18	XI	1	78,660	0.38	0.58
19	XI	1	78,660	0.26	0.63
20	MC	1	78,660	0.69	0.59
21	MC	1	78,660	0.66	0.49

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
22	XI	1	78,660	0.38	0.53
23	XI	1	78,660	0.58	0.62
24	MC	1	78,660	0.66	0.45
25	XI	1	78,660	0.32	0.62
26	MC	1	78,660	0.53	0.36
27	MC	1	78,660	0.58	0.40
28	MC	1	78,660	0.42	0.47
29	XI	1	78,660	0.20	0.53
30	MC	1	78,660	0.44	0.53
31	MC	1	78,660	0.39	0.41
32	XI	1	78,660	0.33	0.67
33	MC	1	78,660	0.52	0.47
34	XI	1	78,660	0.25	0.64
35	XI	1	78,660	0.29	0.56
36	MC	1	78,660	0.35	0.48
37	XI	1	78,660	0.28	0.51
38	MC	1	78,660	0.68	0.55
39	MC	1	78,660	0.57	0.48
40	XI	1	78,660	0.20	0.59
41	MC	1	78,660	0.40	0.64
42	MC	1	78,660	0.72	0.26
43	MC	1	78,660	0.57	0.40
44	MC	1	78,660	0.54	0.41
45	XI	1	78,660	0.30	0.37
46	MC	1	78,660	0.73	0.56
47	MX	1	78,660	0.27	0.32

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.12. Item-Level CTT Statistics, Mathematics Grade 8

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
1	MC	1	79,736	0.68	0.41
2	XI	1	79,736	0.34	0.63
3	XI	1	79,736	0.38	0.62
4	MC	1	79,736	0.24	0.43
5	MC	1	79,736	0.47	0.53
6	MC	1	79,736	0.52	0.52
7	MC	1	79,736	0.24	0.37
8	MC	1	79,736	0.29	0.33
9	MX	1	79,736	0.32	0.52
10	MC	1	79,736	0.79	0.39
11	MC	1	79,736	0.27	0.29
12	MX	1	79,736	0.37	0.29
13	MC	1	79,736	0.36	0.46
14	XI	1	79,736	0.28	0.65
15	XI	1	79,736	0.21	0.63

Item Number	Item Type	Max. Points	N	*P*-Value	Point-Biserial
16	MC	1	79,736	0.61	0.35
17	MX	1	79,736	0.41	0.45
18	MC	1	79,736	0.24	0.53
19	XI	1	79,736	0.23	0.39
20	XI	1	79,736	0.12	0.52
21	MC	1	79,736	0.47	0.50
22	MC	1	79,736	0.59	0.57
23	MC	1	79,736	0.50	0.38
24	MC	1	79,736	0.47	0.33
25	XI	1	79,736	0.49	0.35
26	MC	1	79,736	0.44	0.34
27	XI	1	79,736	0.19	0.59
28	MC	1	79,736	0.81	0.37
29	MC	1	79,736	0.76	0.35
30	MC	1	79,736	0.50	0.46
31	MC	1	79,736	0.20	0.64
32	MC	1	79,736	0.68	0.44
33	XI	1	79,736	0.25	0.61
34	MC	1	79,736	0.41	0.55
35	XI	1	79,736	0.42	0.50
36	MC	1	79,736	0.49	0.37
37	MC	1	79,736	0.54	0.34
38	MC	1	79,736	0.42	0.49
39	MC	1	79,736	0.15	0.50
40	MC	1	79,736	0.43	0.51
41	MC	1	79,736	0.37	0.47
42	MC	1	79,736	0.22	0.38
43	XI	1	79,736	0.34	0.64
44	MC	1	79,736	0.34	0.63
45	MC	1	79,736	0.46	0.37
46	XI	1	79,736	0.38	0.66
47	MC	1	79,736	0.25	0.39

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.13. Distractor Analysis of Multiple-Choice Items, ELA Grade 3

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
7	51.02	0.52	9.46	-0.19	20.63	-0.25	18.80	-0.27	0.08
9	77.13	0.53	3.54	-0.23	9.84	-0.26	9.41	-0.36	0.08
10	49.52	0.42	20.14	-0.24	15.12	-0.18	15.09	-0.14	0.12
11	46.27	0.41	24.07	-0.11	15.82	-0.24	13.74	-0.21	0.10
12	40.26	0.32	19.61	-0.18	18.06	-0.15	21.89	-0.07	0.19
13	59.51	0.57	15.08	-0.30	11.88	-0.29	13.35	-0.22	0.18
15	59.26	0.53	12.76	-0.21	18.10	-0.28	9.68	-0.27	0.20
17	33.06	0.27	28.17	-0.06	16.79	-0.16	21.79	-0.10	0.19
19	67.12	0.57	10.30	-0.24	9.98	-0.26	12.42	-0.34	0.19
22	69.90	0.47	12.22	-0.31	10.73	-0.22	7.07	-0.18	0.09
23	43.74	0.45	22.55	-0.14	19.72	-0.21	13.88	-0.23	0.11
24	46.80	0.39	25.58	-0.19	18.30	-0.23	9.19	-0.08	0.13
25	44.34	0.45	20.15	-0.16	23.19	-0.21	12.18	-0.21	0.15
26	43.19	0.46	17.74	-0.23	18.77	-0.21	20.21	-0.15	0.10
27	70.27	0.52	9.80	-0.25	12.97	-0.26	6.85	-0.29	0.10
28	53.02	0.52	10.16	-0.20	10.56	-0.17	26.16	-0.34	0.09
30	52.14	0.39	23.80	-0.08	14.02	-0.31	9.91	-0.18	0.13
31	44.23	0.42	30.47	-0.16	18.12	-0.20	7.05	-0.24	0.12
33	63.54	0.54	10.26	-0.27	12.28	-0.29	13.80	-0.23	0.12
34	50.10	0.41	16.73	-0.19	20.72	-0.15	12.29	-0.23	0.16
35	32.29	0.28	25.36	-0.16	24.95	-0.08	17.23	-0.06	0.17
36	44.40	0.39	16.36	-0.14	15.43	-0.28	23.65	-0.09	0.16
37	42.03	0.24	19.18	-0.15	19.14	-0.08	19.50	-0.06	0.16
39	62.96	0.53	12.97	-0.19	15.89	-0.32	8.02	-0.26	0.16
41	54.29	0.35	25.81	0.00	12.23	-0.33	7.50	-0.25	0.17
43	47.24	0.53	30.03	-0.14	13.31	-0.30	9.27	-0.32	0.15
45	49.65	0.45	13.68	-0.25	23.07	-0.21	13.47	-0.14	0.14

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.14. Distractor Analysis of Multiple-Choice Items, ELA Grade 4

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
7	76.28	0.56	8.90	-0.34	10.68	-0.34	4.10	-0.18	0.03
9	56.59	0.42	15.49	-0.24	17.49	-0.22	10.39	-0.12	0.03
10	69.39	0.51	11.60	-0.29	9.99	-0.23	8.96	-0.26	0.05
11	74.16	0.49	4.64	-0.25	13.62	-0.27	7.54	-0.26	0.04
13	50.16	0.37	26.75	-0.23	11.47	-0.07	11.57	-0.19	0.04
14	62.07	0.56	14.74	-0.28	11.81	-0.36	11.30	-0.17	0.08
16	80.47	0.52	6.74	-0.29	8.97	-0.33	3.75	-0.21	0.08
19	42.78	0.49	26.73	-0.20	21.29	-0.13	9.12	-0.34	0.08
20	47.36	0.48	23.33	-0.19	8.92	-0.32	20.28	-0.16	0.11
23	62.21	0.39	14.94	-0.23	15.54	-0.18	7.29	-0.15	0.02
24	78.13	0.50	5.90	-0.21	11.09	-0.30	4.86	-0.29	0.02
26	65.61	0.46	6.45	-0.24	18.39	-0.25	9.50	-0.19	0.06
28	76.80	0.51	7.67	-0.28	10.46	-0.31	5.02	-0.21	0.04
29	60.28	0.41	10.32	-0.31	18.90	-0.14	10.45	-0.17	0.04
31	45.37	0.33	14.36	-0.26	25.11	-0.09	15.12	-0.10	0.04
32	32.60	0.37	14.72	-0.15	36.36	-0.14	16.26	-0.15	0.05
33	47.36	0.48	19.47	-0.18	17.39	-0.23	15.72	-0.22	0.06
34	68.88	0.59	9.69	-0.31	14.42	-0.33	6.93	-0.25	0.07
36	58.71	0.48	19.33	-0.15	14.90	-0.32	7.00	-0.25	0.06
37	43.39	0.33	13.67	-0.15	27.29	-0.15	15.56	-0.12	0.09
39	70.52	0.58	10.03	-0.30	11.82	-0.37	7.52	-0.20	0.11
40	25.61	0.30	10.04	-0.26	50.40	-0.03	13.87	-0.12	0.09
42	45.83	0.39	12.48	-0.25	16.56	-0.31	25.05	0.01	0.09
44	67.22	0.51	12.90	-0.19	13.40	-0.35	6.40	-0.23	0.08
45	74.84	0.53	10.25	-0.33	7.76	-0.30	7.07	-0.19	0.08

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.15. Distractor Analysis of Multiple-Choice Items, ELA Grade 5

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
7	84.10	0.48	5.66	-0.26	6.80	-0.31	3.44	-0.21	0.01
8	57.37	0.51	13.46	-0.15	12.94	-0.27	16.21	-0.30	0.02
9	67.97	0.57	5.71	-0.27	11.61	-0.35	14.68	-0.26	0.03
10	54.82	0.48	10.36	-0.24	20.31	-0.22	14.48	-0.22	0.03
11	49.58	0.48	18.91	-0.28	14.76	-0.24	16.72	-0.12	0.03
12	60.09	0.52	9.48	-0.28	12.62	-0.25	17.78	-0.24	0.03
14	53.31	0.41	15.35	-0.21	15.77	-0.17	15.52	-0.19	0.04
15	65.76	0.35	13.62	-0.10	13.00	-0.24	7.57	-0.18	0.05
17	59.59	0.45	8.86	-0.27	10.34	-0.33	21.17	-0.10	0.04
19	42.01	0.35	11.38	-0.36	28.07	-0.09	18.49	-0.05	0.06
20	56.50	0.59	15.76	-0.27	13.62	-0.31	14.08	-0.25	0.04
22	59.61	0.44	7.24	-0.27	12.01	-0.38	21.10	-0.05	0.05
25	61.57	0.43	17.26	-0.27	14.79	-0.17	6.37	-0.19	0.01
27	48.19	0.43	12.37	-0.25	18.10	-0.15	21.31	-0.19	0.03
28	65.82	0.37	6.75	-0.20	11.15	-0.27	16.25	-0.12	0.02
29	62.90	0.47	9.85	-0.22	13.93	-0.26	13.30	-0.22	0.02
30	59.13	0.43	24.74	-0.13	9.70	-0.30	6.41	-0.26	0.03
31	47.55	0.35	10.25	-0.30	33.05	-0.06	9.13	-0.19	0.02
32	45.37	0.28	15.57	-0.19	15.09	-0.29	23.93	0.08	0.03
33	49.78	0.44	22.39	-0.19	11.19	-0.29	16.60	-0.14	0.03
34	60.10	0.57	15.61	-0.27	16.84	-0.33	7.42	-0.22	0.03
35	49.35	0.36	18.34	-0.05	17.33	-0.22	14.96	-0.22	0.03
37	76.42	0.56	5.67	-0.25	12.32	-0.36	5.57	-0.26	0.03
41	54.99	0.50	11.76	-0.35	26.00	-0.19	7.21	-0.20	0.04
42	62.18	0.53	10.62	-0.19	16.67	-0.24	10.49	-0.36	0.04
43	38.64	0.26	24.17	-0.01	25.74	-0.11	11.40	-0.23	0.04
44	54.77	0.44	12.65	-0.33	16.55	-0.20	16.00	-0.09	0.04
45	69.74	0.40	11.31	-0.22	4.50	-0.29	14.41	-0.15	0.04

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.16. Distractor Analysis of Multiple-Choice Items, ELA Grade 6

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
7	63.04	0.47	7.56	-0.17	18.64	-0.21	10.74	-0.33	0.02
8	46.84	0.39	21.75	-0.25	26.56	-0.13	4.83	-0.15	0.02
9	62.49	0.49	8.38	-0.25	13.86	-0.30	15.25	-0.18	0.02
10	80.51	0.44	10.28	-0.24	4.48	-0.27	4.71	-0.20	0.02
11	64.22	0.38	7.19	-0.17	16.55	-0.20	12.01	-0.20	0.04
12	64.79	0.49	7.17	-0.20	8.09	-0.32	19.94	-0.23	0.02
13	45.30	0.35	18.67	-0.20	25.07	-0.10	10.91	-0.17	0.04
15	63.35	0.53	7.23	-0.31	16.86	-0.30	12.52	-0.18	0.04
16	40.19	0.36	10.96	-0.12	17.87	-0.22	30.93	-0.11	0.05
17	73.87	0.54	8.15	-0.23	8.73	-0.31	9.20	-0.30	0.05
20	57.94	0.37	12.30	-0.31	9.00	-0.35	20.72	0.05	0.05
21	72.45	0.43	14.23	-0.22	5.91	-0.29	7.35	-0.18	0.07
24	63.37	0.30	19.47	-0.14	6.80	-0.29	10.34	-0.04	0.02
25	66.29	0.60	15.01	-0.31	11.51	-0.31	7.16	-0.29	0.02
26	42.25	0.39	18.06	-0.18	24.55	-0.19	15.11	-0.11	0.03
28	45.11	0.32	15.84	-0.06	23.64	-0.19	15.38	-0.16	0.04
30	64.00	0.53	10.49	-0.29	13.75	-0.36	11.74	-0.12	0.02
31	63.28	0.47	14.47	-0.34	12.00	-0.26	10.21	-0.08	0.04
33	53.29	0.44	11.62	-0.26	23.97	-0.22	11.08	-0.13	0.04
34	53.31	0.49	8.85	-0.23	22.22	-0.34	15.58	-0.11	0.04
35	82.58	0.52	6.27	-0.29	6.00	-0.30	5.11	-0.24	0.04
36	67.63	0.64	10.82	-0.33	9.44	-0.38	12.06	-0.26	0.04
38	47.50	0.44	17.44	-0.15	23.40	-0.30	11.60	-0.11	0.06
40	43.57	0.39	13.05	-0.33	21.80	-0.23	21.51	0.02	0.07
42	63.55	0.48	9.86	-0.27	20.20	-0.21	6.33	-0.27	0.07
44	37.83	0.29	9.84	-0.04	38.20	-0.22	14.07	-0.06	0.06
45	66.84	0.63	13.44	-0.40	6.98	-0.33	12.68	-0.23	0.06

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.17. Distractor Analysis of Multiple-Choice Items, ELA Grade 7

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
7	63.77	0.48	8.90	-0.29	7.85	-0.29	19.45	-0.17	0.03
9	46.42	0.35	26.43	-0.20	12.18	-0.11	14.95	-0.14	0.02
11	51.50	0.30	16.98	0.05	16.36	-0.15	15.11	-0.31	0.04
12	58.48	0.52	17.10	-0.30	15.96	-0.30	8.43	-0.11	0.02
14	61.96	0.39	7.81	-0.14	22.78	-0.20	7.41	-0.25	0.05
15	66.55	0.55	11.16	-0.32	10.79	-0.17	11.44	-0.34	0.05
16	55.48	0.47	27.53	-0.18	12.69	-0.31	4.24	-0.24	0.06
17	61.60	0.48	14.70	-0.20	14.73	-0.30	8.92	-0.19	0.05
18	41.55	0.24	16.25	-0.10	21.62	-0.14	20.51	-0.06	0.06
19	47.13	0.30	32.89	-0.11	9.93	-0.24	9.99	-0.10	0.06
20	74.22	0.51	8.69	-0.28	8.75	-0.30	8.28	-0.22	0.07
23	68.99	0.58	11.40	-0.22	12.42	-0.37	7.17	-0.30	0.03
25	47.31	0.47	21.48	-0.10	11.18	-0.28	19.99	-0.27	0.04
26	58.41	0.38	12.60	-0.13	17.02	-0.21	11.92	-0.21	0.04
27	53.18	0.40	31.13	-0.11	6.95	-0.29	8.70	-0.27	0.04
28	30.17	0.35	16.51	-0.14	22.36	-0.19	30.93	-0.06	0.03
29	53.10	0.39	14.44	-0.20	18.06	-0.21	14.35	-0.12	0.05
30	57.90	0.41	17.60	-0.12	11.18	-0.24	13.28	-0.24	0.04
32	43.59	0.36	21.43	-0.12	12.06	-0.26	22.88	-0.11	0.03
34	64.57	0.47	12.68	-0.15	9.96	-0.37	12.75	-0.19	0.04
35	61.82	0.58	20.91	-0.23	9.33	-0.33	7.88	-0.34	0.05
36	60.16	0.46	13.46	-0.30	21.81	-0.16	4.54	-0.25	0.04
37	63.93	0.60	14.53	-0.26	11.41	-0.34	10.08	-0.28	0.04
39	58.14	0.43	13.90	-0.15	16.54	-0.30	11.37	-0.14	0.06
40	54.55	0.39	15.00	-0.15	16.47	-0.33	13.91	-0.05	0.06
41	51.69	0.46	10.80	-0.32	14.57	-0.29	22.88	-0.07	0.06
42	61.40	0.54	20.75	-0.26	10.13	-0.33	7.67	-0.21	0.05
43	41.34	0.47	15.72	-0.20	21.59	-0.17	21.30	-0.21	0.05
44	47.21	0.36	20.69	-0.12	23.12	-0.17	8.92	-0.21	0.05
45	48.81	0.35	17.98	-0.21	8.60	-0.30	24.56	-0.02	0.05

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.18. Distractor Analysis of Multiple-Choice Items, ELA Grade 8

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
7	66.38	0.54	17.32	-0.30	4.00	-0.23	12.28	-0.29	0.01
9	57.71	0.37	8.11	-0.18	7.52	-0.21	26.64	-0.17	0.01
10	66.35	0.40	7.78	-0.10	16.78	-0.26	9.05	-0.23	0.03
12	64.37	0.48	10.84	-0.24	9.12	-0.32	15.64	-0.18	0.03
13	33.64	0.22	18.03	-0.07	27.71	-0.11	20.60	-0.08	0.03
14	44.50	0.39	28.07	-0.16	13.78	-0.15	13.63	-0.21	0.03
15	43.13	0.37	21.34	-0.16	21.10	-0.09	14.40	-0.22	0.03
16	40.42	0.35	21.58	-0.06	22.74	-0.17	15.24	-0.20	0.02
17	36.00	0.34	17.08	-0.24	24.24	-0.11	22.64	-0.07	0.04
19	72.90	0.45	18.76	-0.25	5.39	-0.30	2.92	-0.20	0.04
22	78.60	0.50	8.63	-0.30	5.92	-0.27	6.84	-0.22	0.02
24	46.69	0.29	8.66	-0.17	16.72	-0.20	27.89	-0.05	0.04
25	43.51	0.49	17.23	-0.26	22.43	-0.17	16.79	-0.20	0.03
26	58.23	0.45	9.25	-0.21	15.08	-0.19	17.41	-0.25	0.03
27	34.62	0.29	11.00	-0.19	28.35	-0.11	26.00	-0.07	0.03
28	49.71	0.27	6.41	-0.16	31.34	-0.15	12.51	-0.07	0.03
29	70.23	0.41	19.29	-0.22	6.78	-0.27	3.68	-0.18	0.02
31	59.36	0.52	17.43	-0.16	12.73	-0.26	10.46	-0.35	0.02
33	52.12	0.52	8.57	-0.28	25.70	-0.14	13.60	-0.34	0.02
34	38.27	0.35	33.84	-0.07	17.42	-0.25	10.45	-0.12	0.03
35	47.73	0.41	18.79	-0.13	23.48	-0.20	9.97	-0.23	0.03
36	52.21	0.46	15.27	-0.28	17.61	-0.26	14.88	-0.07	0.04
37	44.66	0.39	24.43	0.02	13.00	-0.29	17.87	-0.27	0.04
38	27.60	0.24	17.87	-0.14	38.09	0.08	16.39	-0.25	0.04
39	57.10	0.52	13.91	-0.27	19.78	-0.26	9.17	-0.21	0.05
40	58.73	0.57	12.87	-0.26	15.84	-0.33	12.52	-0.22	0.04
42	54.21	0.48	15.56	-0.17	19.55	-0.24	10.64	-0.28	0.04
43	54.79	0.48	14.29	-0.19	18.53	-0.21	12.37	-0.28	0.03
44	82.27	0.51	7.96	-0.29	4.69	-0.27	5.05	-0.26	0.03
45	68.11	0.53	8.67	-0.29	13.02	-0.33	10.16	-0.17	0.03

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.19. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 3

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
1	86.76	0.33	5.16	-0.24	2.69	-0.16	5.31	-0.15	0.07
10	65.73	0.63	8.62	-0.38	8.08	-0.25	17.43	-0.33	0.14
12	42.47	0.55	28.32	-0.28	13.91	-0.24	15.08	-0.16	0.22
13	36.09	0.27	26.06	-0.12	27.58	-0.07	9.99	-0.13	0.28
14	58.54	0.60	15.12	-0.45	14.97	-0.20	11.24	-0.20	0.13
15	60.15	0.37	23.48	-0.23	6.30	-0.15	9.90	-0.15	0.16
16	39.45	0.50	21.31	-0.27	23.35	-0.16	15.65	-0.17	0.24
17	56.90	0.54	25.74	-0.30	10.60	-0.30	6.55	-0.16	0.21
18	91.90	0.43	2.49	-0.24	2.52	-0.24	2.86	-0.24	0.22
20	57.20	0.40	20.76	-0.21	14.46	-0.22	7.29	-0.12	0.29
21	71.15	0.62	12.22	-0.29	10.39	-0.37	6.01	-0.30	0.22
22	80.03	0.53	7.67	-0.29	6.68	-0.30	5.39	-0.25	0.23
23	62.59	0.43	7.54	-0.29	7.95	-0.28	21.68	-0.13	0.24
24	76.57	0.42	12.81	-0.30	6.04	-0.16	4.53	-0.18	0.05
29	56.64	0.56	22.17	-0.36	15.91	-0.28	5.18	-0.11	0.11
33	78.99	0.52	6.15	-0.30	7.63	-0.32	7.09	-0.21	0.13
37	72.91	0.58	10.18	-0.37	9.26	-0.34	7.48	-0.17	0.18
40	18.39	0.43	18.27	-0.21	55.42	-0.11	7.76	-0.11	0.16
45	55.64	0.57	24.45	-0.32	14.71	-0.28	5.03	-0.20	0.16

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.20. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 4

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
1	79.43	0.47	12.30	-0.33	6.67	-0.27	1.56	-0.13	0.03
4	58.81	0.50	7.37	-0.25	17.08	-0.32	16.70	-0.17	0.05
5	40.42	0.37	29.62	-0.19	14.20	-0.17	15.70	-0.10	0.06
7	75.88	0.46	11.50	-0.28	8.99	-0.22	3.56	-0.23	0.06
10	55.91	0.58	14.46	-0.15	21.41	-0.40	8.14	-0.25	0.07
11	51.52	0.49	17.56	-0.21	25.86	-0.28	4.97	-0.17	0.08
12	53.77	0.32	22.72	-0.23	15.01	-0.11	8.38	-0.06	0.12
15	44.94	0.50	18.19	-0.26	26.62	-0.20	10.14	-0.19	0.11
18	42.04	0.46	18.11	-0.23	29.79	-0.25	9.92	-0.08	0.15
21	60.11	0.57	17.70	-0.20	11.72	-0.33	10.33	-0.31	0.14
22	50.26	0.34	26.05	-0.07	13.77	-0.24	9.75	-0.18	0.18
24	80.61	0.46	15.43	-0.39	2.78	-0.18	1.14	-0.11	0.04
26	57.92	0.50	13.32	-0.24	17.52	-0.32	11.19	-0.15	0.05
27	74.22	0.50	12.17	-0.31	8.36	-0.25	5.18	-0.21	0.08
29	69.70	0.51	7.19	-0.28	15.67	-0.28	7.36	-0.23	0.07
32	48.47	0.59	7.40	-0.25	21.84	-0.44	22.22	-0.12	0.06
33	36.11	0.37	11.86	-0.24	10.60	-0.15	41.34	-0.11	0.07
36	59.35	0.41	11.17	-0.18	22.52	-0.23	6.82	-0.19	0.14
37	75.52	0.52	15.27	-0.34	5.41	-0.28	3.68	-0.20	0.12
42	34.03	0.51	18.72	-0.27	22.94	-0.20	24.18	-0.12	0.12
44	50.75	0.45	10.03	-0.10	33.69	-0.28	5.45	-0.28	0.08

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.21. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 5

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
3	58.81	0.43	20.32	-0.28	14.14	-0.16	6.70	-0.17	0.03
4	52.91	0.44	16.67	-0.18	22.72	-0.26	7.64	-0.17	0.07
7	39.05	0.38	28.73	-0.08	14.18	-0.23	17.97	-0.18	0.07
9	51.98	0.37	7.47	-0.18	27.40	-0.20	13.09	-0.14	0.06
10	53.93	0.52	12.87	-0.25	23.61	-0.28	9.52	-0.19	0.07
11	59.77	0.52	13.71	-0.20	12.09	-0.31	14.37	-0.23	0.07
12	50.72	0.36	32.26	-0.14	11.72	-0.25	5.21	-0.16	0.09
17	64.66	0.54	3.75	-0.19	16.95	-0.30	14.57	-0.31	0.08
18	46.45	0.34	28.56	-0.14	13.76	-0.28	11.12	-0.04	0.11
20	49.90	0.41	28.07	-0.38	8.15	-0.21	13.79	0.06	0.10
22	33.02	0.45	17.83	-0.27	17.88	-0.25	31.18	-0.02	0.10
23	52.30	0.47	15.91	-0.26	21.45	-0.28	10.23	-0.07	0.11
25	63.41	0.38	31.80	-0.30	3.55	-0.17	1.23	-0.12	0.01
27	60.63	0.36	5.24	-0.21	12.02	-0.18	22.09	-0.17	0.02
31	26.26	0.43	19.22	0.03	25.33	-0.20	29.17	-0.24	0.03
32	45.60	0.53	17.54	-0.31	28.25	-0.24	8.57	-0.12	0.03
36	65.98	0.47	9.56	-0.26	17.03	-0.30	7.37	-0.14	0.06
38	48.29	0.44	27.65	-0.23	16.29	-0.17	7.70	-0.20	0.07
39	44.83	0.37	35.92	-0.10	10.86	-0.29	8.32	-0.18	0.06
41	39.26	0.29	24.32	-0.01	23.07	-0.23	13.28	-0.13	0.07
42	74.18	0.52	6.41	-0.26	14.53	-0.34	4.82	-0.21	0.06
43	30.61	0.34	29.81	-0.10	27.42	-0.18	12.07	-0.09	0.09
45	42.43	0.42	12.74	-0.14	29.45	-0.23	15.31	-0.15	0.07

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.22. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 6

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
1	73.46	0.48	11.46	-0.27	6.65	-0.21	8.42	-0.26	0.01
2	53.02	0.45	28.32	-0.23	12.34	-0.27	6.28	-0.12	0.03
4	65.23	0.28	18.75	-0.17	5.70	-0.16	10.29	-0.09	0.02
5	66.18	0.42	12.42	-0.23	14.02	-0.27	7.33	-0.11	0.05
10	32.79	0.33	27.04	-0.23	21.22	-0.05	18.87	-0.09	0.07
11	49.88	0.42	22.52	-0.23	17.22	-0.16	10.28	-0.17	0.11
12	35.82	0.36	34.35	-0.18	16.49	-0.06	13.26	-0.18	0.09
14	28.37	0.33	8.09	-0.07	30.70	-0.17	32.76	-0.11	0.08
15	40.45	0.29	19.00	-0.09	22.85	-0.20	17.63	-0.06	0.07
17	35.86	0.34	25.11	-0.06	22.88	-0.32	16.07	-0.01	0.08
18	24.56	0.48	24.14	-0.16	23.57	-0.23	27.64	-0.09	0.09
20	41.57	0.42	22.79	-0.28	26.62	-0.15	8.90	-0.07	0.12
23	47.61	0.46	15.88	-0.25	20.71	-0.15	15.69	-0.21	0.12
24	70.39	0.47	12.80	-0.34	6.66	-0.25	10.13	-0.13	0.02
28	31.28	0.53	14.11	-0.24	20.98	-0.08	33.60	-0.28	0.03
29	42.52	0.55	22.42	-0.29	21.56	-0.22	13.45	-0.18	0.05
31	45.42	0.41	15.12	-0.22	30.48	-0.18	8.94	-0.15	0.04
32	45.18	0.28	20.84	-0.24	7.53	-0.24	26.42	0.05	0.03
38	34.56	0.53	17.06	-0.29	14.24	-0.30	34.09	-0.07	0.06
45	72.36	0.56	8.37	-0.23	8.82	-0.30	10.37	-0.34	0.08
47	59.34	0.45	20.45	-0.28	13.45	-0.22	6.68	-0.13	0.08

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.23. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 7

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
1	70.03	0.54	4.75	-0.21	18.53	-0.39	6.66	-0.21	0.03
2	53.28	0.38	13.56	-0.16	21.24	-0.19	11.88	-0.17	0.04
4	39.09	0.36	35.54	-0.15	13.17	-0.11	12.16	-0.20	0.04
6	54.33	0.36	13.82	-0.12	17.15	-0.19	14.62	-0.18	0.08
8	19.98	0.41	27.37	0.04	35.68	-0.23	16.88	-0.18	0.09
11	37.56	0.38	11.56	-0.12	34.04	-0.30	16.76	-0.02	0.09
13	62.88	0.51	8.08	-0.23	18.11	-0.35	10.82	-0.14	0.11
14	52.42	0.38	12.62	-0.15	28.35	-0.28	6.49	-0.04	0.11
16	35.84	0.46	23.56	-0.08	25.74	-0.33	14.75	-0.11	0.10
20	69.49	0.59	14.07	-0.34	11.32	-0.33	5.00	-0.22	0.11
21	66.31	0.49	3.87	-0.15	23.55	-0.35	6.17	-0.22	0.11
24	66.05	0.45	13.89	-0.28	13.59	-0.24	6.46	-0.15	0.03
26	53.01	0.36	23.48	-0.04	18.78	-0.33	4.68	-0.15	0.05
27	58.27	0.40	8.31	-0.25	12.70	-0.28	20.68	-0.08	0.04
28	41.60	0.47	22.25	-0.22	15.75	-0.18	20.37	-0.19	0.03
30	43.90	0.53	11.99	-0.26	22.61	-0.17	21.45	-0.26	0.04
31	39.42	0.41	15.95	-0.16	20.10	-0.14	24.49	-0.19	0.04
33	51.66	0.47	16.39	-0.19	17.97	-0.27	13.93	-0.17	0.06
36	34.78	0.48	10.49	-0.21	35.23	-0.17	19.44	-0.21	0.07
38	68.34	0.55	16.52	-0.33	9.53	-0.30	5.56	-0.20	0.05
39	56.71	0.48	14.17	-0.26	21.02	-0.19	8.04	-0.26	0.06
42	71.89	0.26	1.90	-0.09	6.05	-0.15	20.11	-0.17	0.06
43	56.83	0.40	9.37	-0.21	23.66	-0.19	10.06	-0.19	0.08
44	53.95	0.41	11.07	-0.23	23.37	-0.20	11.54	-0.16	0.08
46	72.69	0.56	13.34	-0.33	7.07	-0.27	6.83	-0.26	0.06

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table A.24. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 8

Item Number	Correct Option		Distractor 1		Distractor 2		Distractor 3		
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%Omit
1	67.74	0.41	11.00	-0.25	8.94	-0.21	12.30	-0.16	0.02
4	23.57	0.43	23.65	-0.04	23.78	-0.07	28.97	-0.30	0.04
5	46.75	0.53	23.13	-0.24	16.81	-0.24	13.26	-0.23	0.05
7	23.88	0.37	40.16	-0.09	18.18	-0.17	17.73	-0.12	0.05
8	28.93	0.33	38.43	-0.16	27.62	-0.14	4.97	-0.04	0.05
10	78.98	0.39	8.03	-0.22	6.49	-0.19	6.44	-0.20	0.06
11	27.26	0.29	11.34	-0.12	44.11	-0.18	17.25	0.00	0.05
13	35.78	0.46	27.83	-0.19	26.05	-0.22	10.26	-0.13	0.08
16	61.38	0.35	22.77	-0.13	9.66	-0.27	6.13	-0.15	0.05
18	23.83	0.53	28.40	-0.23	30.64	-0.19	17.06	-0.09	0.08
21	47.30	0.50	22.80	-0.19	17.17	-0.27	12.67	-0.21	0.07
22	58.79	0.57	18.97	-0.29	13.13	-0.28	9.04	-0.25	0.07
23	50.30	0.38	18.28	-0.11	21.39	-0.21	9.95	-0.21	0.09
24	46.59	0.33	32.43	-0.13	10.96	-0.26	9.99	-0.08	0.03
26	44.03	0.34	13.11	-0.19	29.05	-0.19	13.77	-0.05	0.04
28	80.91	0.37	5.35	-0.20	10.91	-0.24	2.81	-0.15	0.02
29	75.89	0.35	6.36	-0.21	13.12	-0.20	4.60	-0.13	0.02
30	50.15	0.46	19.15	-0.25	10.11	-0.21	20.55	-0.17	0.04
32	67.77	0.44	11.39	-0.22	14.84	-0.27	5.94	-0.17	0.06
34	41.11	0.55	8.91	-0.19	34.16	-0.26	15.77	-0.26	0.05
36	48.78	0.37	11.97	-0.04	23.97	-0.29	15.24	-0.12	0.05
37	53.57	0.34	18.65	-0.11	18.73	-0.17	9.00	-0.19	0.05
38	41.88	0.49	14.36	-0.14	22.42	-0.19	21.30	-0.28	0.05
40	43.32	0.51	25.36	-0.26	21.63	-0.25	9.62	-0.11	0.08
41	36.72	0.47	28.52	0.02	12.21	-0.24	22.48	-0.37	0.06
44	34.41	0.63	20.18	-0.19	24.01	-0.28	21.33	-0.26	0.06
45	46.05	0.37	16.42	-0.16	22.47	-0.22	15.00	-0.10	0.06
47	25.21	0.39	44.00	-0.25	24.53	-0.09	6.21	-0.04	0.05

Note. The item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Appendix B: ITEM-LEVEL IRT STATISTICS

This appendix includes the following item-level IRT statistics:

- Table B.1 – Table B.12 present the IRT statistics, including item type, Rasch difficulty, standard error (SE) of Rasch, and infit values.
- Table B.13 – Table B.24 present the raw-to-scale score conversion tables.
- Figure B.1 – Figure B.18 present the item-person map for each post-equated operational form.
- Figure B.19 – Figure B.54 present the test characteristic curve (TCC) and conditional standard error of measurement (CSEM) curve for each post-equated operational form.
- Figure B.55 – Figure B.72 present the scree plot from the principal component analysis (PCA) for each operational form. The scree plot shows only the first 10 components.

Table B.1. Item-Level IRT Statistics, ELA Grade 3

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	OE	2.6889	0.0102	0.76
2	OE	2.8093	0.0107	0.77
3	OE	-0.6989	0.0084	0.80
4	OE	2.2236	0.0093	0.73
5	OE	2.5405	0.0097	0.77
6	OE	-0.6691	0.0083	0.76
7	MC	-0.1954	0.0084	0.94
8	MX	-0.1786	0.0084	0.86
9	MC	-1.6720	0.0096	0.82
10	MC	-0.1179	0.0084	1.08
11	MC	0.0521	0.0084	1.09
12	MC	0.2279	0.0085	1.18
13	MC	-0.6417	0.0085	0.87
14	MC	1.0918	0.0092	1.00
15	MC	-0.5722	0.0084	0.91
16	MX	1.5143	0.0100	1.13
17	MC	0.7924	0.0089	1.22
18	XI	-0.3018	0.0084	1.00
19	MC	-1.0014	0.0087	0.83
20	MX	-0.3143	0.0067	1.17
21	MX	-1.1192	0.0063	1.05
22	MC	-1.2171	0.0089	0.95
23	MC	0.1852	0.0085	1.03
24	MC	0.0237	0.0084	1.11
25	MC	0.1529	0.0084	1.03
26	MC	0.2144	0.0085	1.02
27	MC	-1.2391	0.0089	0.88
28	MC	-0.2991	0.0084	0.94
29	XI	-0.5602	0.0084	0.85
30	MC	-0.2543	0.0084	1.11
31	MC	0.1589	0.0084	1.06

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
32	MX	0.5094	0.0086	1.05
33	MC	-0.8577	0.0086	0.89
34	MC	-0.1491	0.0084	1.08
35	MC	0.8174	0.0089	1.20
36	MC	0.1488	0.0084	1.11
37	MC	0.2753	0.0085	1.30
38	MX	0.5364	0.0087	0.92
39	MC	-0.8358	0.0086	0.92
40	MX	1.4787	0.0099	0.91
41	MC	-0.3670	0.0084	1.15
42	XI	0.5727	0.0087	0.82
43	MC	0.0005	0.0084	0.94
44	MC	-0.1811	0.0084	0.88
45	MC	-0.1251	0.0084	1.04
46	MX	-0.5748	0.0061	1.03
47	MX	-0.5762	0.0065	1.16

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.2. Item-Level IRT Statistics, ELA Grade 4

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	OE	2.5356	0.0088	0.78
2	OE	2.5419	0.0090	0.79
3	OE	-0.3717	0.0084	0.87
4	OE	2.8988	0.0100	0.78
5	OE	2.9153	0.0101	0.82
6	OE	-0.8319	0.0088	0.82
7	MC	-1.1279	0.0094	0.82
8	MC	0.0219	0.0083	0.90
9	MC	0.0754	0.0083	1.09
10	MC	-0.6906	0.0088	0.93
11	MC	-0.8981	0.0091	0.89
12	MX	0.0108	0.0083	0.86
13	MC	0.3698	0.0083	1.16
14	MC	-0.2718	0.0085	0.90
15	XI	0.1763	0.0083	0.86
16	MC	-1.4313	0.0100	0.84
17	MX	1.2880	0.0087	0.97
18	XI	1.1939	0.0087	0.99
19	MC	0.7643	0.0084	1.00
20	MC	0.7958	0.0084	1.05
21	MX	0.0309	0.0059	1.13
22	MX	0.3502	0.0061	1.15
23	MC	-0.2778	0.0085	1.12
24	MC	-1.2569	0.0096	0.88
25	MX	1.6200	0.0091	1.06

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
26	MC	-0.4692	0.0086	1.01
27	MC	1.3689	0.0088	1.01
28	MC	-1.1638	0.0095	0.88
29	MC	-0.1009	0.0084	1.08
30	MX	0.5217	0.0054	1.22
31	MC	0.6029	0.0083	1.22
32	MC	1.0400	0.0085	1.06
33	MC	0.4876	0.0083	1.01
34	MC	-0.6259	0.0088	0.82
35	MC	0.4611	0.0083	0.81
36	MC	-0.0609	0.0084	1.01
37	MC	0.7066	0.0084	1.21
38	MC	1.6629	0.0092	0.98
39	MC	-0.7606	0.0089	0.84
40	MC	1.7784	0.0093	1.13
41	MC	1.6174	0.0091	1.06
42	MC	0.5993	0.0083	1.14
43	XI	1.0350	0.0085	0.93
44	MC	-0.5641	0.0087	0.94
45	MC	-1.0338	0.0093	0.86
46	MX	0.2945	0.0059	1.17
47	MX	-0.8331	0.0063	1.07

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.3. Item-Level IRT Statistics, ELA Grade 5

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	OE	1.5373	0.0083	0.80
2	OE	1.9339	0.0084	0.80
3	OE	-0.9279	0.0086	0.73
4	OE	1.6387	0.0085	0.77
5	OE	1.9563	0.0089	0.80
6	OE	-1.3187	0.0093	0.72
7	MC	-1.9440	0.0106	0.86
8	MC	-0.2603	0.0082	0.95
9	MC	-0.8399	0.0086	0.85
10	MC	-0.1270	0.0082	0.99
11	MC	0.1438	0.0082	0.99
12	MC	-0.4039	0.0083	0.93
13	XI	0.1741	0.0082	0.91
14	MC	-0.0495	0.0082	1.08
15	MC	-0.7137	0.0085	1.13
16	MX	-0.1997	0.0082	0.81
17	MC	-0.3777	0.0083	1.02
18	MX	0.6657	0.0083	1.05
19	MC	0.5174	0.0082	1.14

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
20	MC	-0.2150	0.0082	0.85
21	MC	1.2005	0.0088	1.09
22	MC	-0.2379	0.0082	1.02
23	MX	-0.0142	0.0058	1.09
24	MX	-0.9021	0.0062	1.02
25	MC	-0.4830	0.0083	1.04
26	MC	1.6251	0.0094	0.92
27	MC	0.2156	0.0082	1.04
28	MC	-0.7168	0.0085	1.10
29	MC	-0.5550	0.0084	0.98
30	MC	-0.3533	0.0083	1.04
31	MC	0.1600	0.0082	1.15
32	MC	0.3206	0.0082	1.23
33	MC	0.1335	0.0082	1.04
34	MC	-0.4039	0.0083	0.88
35	MC	0.2012	0.0082	1.13
36	MX	0.0152	0.0082	1.02
37	MC	-1.4313	0.0094	0.86
38	XI	1.0922	0.0087	1.04
39	MC	0.3476	0.0082	0.84
40	MX	1.5012	0.0092	0.86
41	MC	-0.1364	0.0082	0.96
42	MC	-0.5167	0.0084	0.92
43	MC	0.7174	0.0083	1.24
44	MC	-0.1059	0.0082	1.04
45	MC	-0.9430	0.0087	1.04
46	MX	-0.4121	0.0058	1.02
47	MX	0.6550	0.0062	1.34

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.4. Item-Level IRT Statistics, ELA Grade 6

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	OE	0.8327	0.0076	0.80
2	OE	1.1883	0.0082	0.78
3	OE	-1.1020	0.0090	0.74
4	OE	0.8485	0.0079	0.72
5	OE	1.3695	0.0081	0.78
6	OE	-1.1326	0.0093	0.77
7	MC	-0.4980	0.0085	0.99
8	MC	0.3599	0.0083	1.11
9	MC	-0.4673	0.0085	0.99
10	MC	-1.6036	0.0100	0.94
11	MC	-0.5633	0.0085	1.10
12	MC	-0.5951	0.0086	0.97
13	MC	0.3566	0.0083	1.16

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
14	MX	2.3672	0.0110	1.04
15	MC	-0.5154	0.0085	0.93
16	MC	0.8816	0.0085	1.17
17	MC	-1.0592	0.0091	0.85
18	MC	0.7713	0.0084	0.89
19	XI	1.9833	0.0100	1.17
20	MC	-0.2555	0.0083	1.13
21	MC	-1.0511	0.0091	1.00
22	MX	-0.5827	0.0060	0.94
23	MX	1.2331	0.0065	1.24
24	MC	-0.5162	0.0085	1.21
25	MC	-0.6810	0.0086	0.83
26	MC	0.6012	0.0083	1.10
27	MC	-0.1582	0.0083	1.05
28	MC	0.4504	0.0083	1.19
29	MX	0.4460	0.0083	0.99
30	MC	-0.5515	0.0085	0.93
31	MC	-0.5114	0.0085	0.99
32	MC	0.8249	0.0084	1.03
33	MC	0.0233	0.0083	1.05
34	MC	0.0223	0.0083	0.99
35	MC	-1.7698	0.0104	0.84
36	MC	-0.7591	0.0087	0.78
37	MX	-0.3626	0.0084	1.10
38	MC	0.3218	0.0082	1.03
39	MX	0.9434	0.0085	0.91
40	MC	0.5212	0.0083	1.10
41	MX	0.1497	0.0082	0.75
42	MC	-0.4650	0.0085	0.98
43	MX	0.5972	0.0083	0.90
44	MC	0.9370	0.0085	1.21
45	MC	-0.7133	0.0087	0.79
46	MX	0.1710	0.0060	1.16
47	MX	-0.0494	0.0059	1.00

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.5. Item-Level IRT Statistics, ELA Grade 7

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	OE	0.9901	0.0072	0.72
2	OE	1.2043	0.0074	0.74
3	OE	-0.9847	0.0087	0.70
4	OE	1.2654	0.0077	0.78
5	OE	1.6833	0.0080	0.74
6	OE	-0.9898	0.0088	0.72
7	MC	-0.4587	0.0083	0.96
8	MC	-0.9404	0.0088	0.79
9	MC	0.3856	0.0081	1.13
10	MC	0.3505	0.0081	0.90
11	MC	0.2907	0.0081	1.21
12	MC	-0.0625	0.0082	0.92
13	XI	0.6267	0.0082	0.98
14	MC	-0.3911	0.0083	1.07
15	MC	-0.6436	0.0085	0.87
16	MC	-0.0514	0.0081	0.99
17	MC	-0.3720	0.0083	0.98
18	MC	0.6684	0.0082	1.25
19	MC	0.3786	0.0081	1.19
20	MC	-1.1006	0.0090	0.89
21	MX	-1.2626	0.0064	0.82
22	MX	-0.4898	0.0060	1.03
23	MC	-0.7815	0.0086	0.83
24	MX	0.2591	0.0081	0.83
25	MC	0.3695	0.0081	0.99
26	MC	-0.2031	0.0082	1.09
27	MC	0.0682	0.0081	1.08
28	MC	1.2985	0.0087	1.06
29	MC	0.1787	0.0081	1.09
30	MC	-0.0175	0.0081	1.06
31	MC	0.7146	0.0082	0.86
32	MC	0.3790	0.0081	1.12
33	MC	0.1548	0.0081	0.82
34	MC	-0.5328	0.0084	0.98
35	MC	-0.3840	0.0083	0.86
36	MC	-0.2956	0.0082	1.00
37	MC	-0.4977	0.0084	0.83
38	XI	1.6608	0.0092	1.12
39	MC	-0.1890	0.0082	1.05
40	MC	-0.0030	0.0081	1.09
41	MC	0.1446	0.0081	1.01
42	MC	-0.3615	0.0083	0.91
43	MC	0.6794	0.0082	0.98
44	MC	0.3742	0.0081	1.13
45	MC	0.3485	0.0081	1.14

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
46	MX	0.4899	0.0059	1.21
47	MX	0.2932	0.0060	1.33

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.6. Item-Level IRT Statistics, ELA Grade 8

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	OE	0.3647	0.0073	0.78
2	OE	1.1550	0.0070	0.97
3	OE	-1.8472	0.0103	0.75
4	OE	0.8889	0.0072	0.81
5	OE	1.2108	0.0072	0.83
6	OE	-1.5485	0.0095	0.72
7	MC	-0.7382	0.0083	0.87
8	XI	-0.7054	0.0083	0.95
9	MC	-0.2817	0.0080	1.09
10	MC	-0.7375	0.0083	1.02
11	MX	-0.7097	0.0083	0.89
12	MC	-0.6298	0.0082	0.94
13	MC	0.8516	0.0083	1.18
14	MC	0.3584	0.0080	1.06
15	MC	0.5562	0.0081	1.10
16	MC	0.5928	0.0081	1.09
17	MC	0.5266	0.0081	1.06
18	MX	0.8866	0.0083	1.12
19	MC	-0.9556	0.0086	0.90
20	MX	-0.6283	0.0064	1.13
21	MX	-0.9164	0.0059	0.91
22	MC	-1.4505	0.0093	0.84
23	MX	1.2974	0.0088	1.00
24	MC	0.2446	0.0080	1.18
25	MC	0.4095	0.0080	0.94
26	MC	-0.3255	0.0081	0.99
27	MC	0.9586	0.0084	1.18
28	MC	0.1211	0.0080	1.21
29	MC	-0.9560	0.0086	0.99
30	MX	1.4467	0.0090	1.06
31	MC	-0.3667	0.0081	0.92
32	XI	0.1486	0.0080	0.89
33	MC	0.0008	0.0080	0.92
34	MC	0.7054	0.0082	1.08
35	MC	0.2209	0.0080	1.03
36	MC	-0.0041	0.0080	0.99
37	MC	0.3753	0.0080	1.06
38	MC	1.3048	0.0088	1.17
39	MC	-0.2514	0.0080	0.92

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
40	MC	-0.3349	0.0081	0.86
41	MX	0.7901	0.0082	0.92
42	MC	-0.1052	0.0080	0.96
43	MC	-0.1339	0.0080	0.96
44	MC	-1.7565	0.0100	0.83
45	MC	-0.8344	0.0084	0.88
46	MX	0.5336	0.0058	1.21
47	MX	0.1700	0.0057	0.98

Note. OE = open-ended, MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.7. Item-Level IRT Statistics, Mathematics Grade 3

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	MC	-2.0209	0.0120	1.09
2	XI	-1.1096	0.0101	0.81
3	XI	0.5728	0.0089	1.05
4	XI	1.5809	0.0093	0.84
5	XI	-0.3147	0.0092	0.89
6	MC	0.6355	0.0089	0.93
7	XI	-0.6020	0.0095	0.99
8	MX	1.5475	0.0092	1.08
9	XI	-0.6824	0.0096	1.00
10	MC	-0.4119	0.0093	0.88
11	XI	1.3005	0.0091	1.06
12	MC	1.0007	0.0089	0.99
13	MC	1.3892	0.0091	1.39
14	MC	0.0417	0.0090	0.94
15	MC	-0.0585	0.0091	1.35
16	MC	1.0418	0.0089	1.05
17	MC	0.1403	0.0090	1.06
18	MC	-2.8417	0.0149	0.86
19	XI	-0.0479	0.0091	0.88
20	MC	0.1213	0.0090	1.30
21	MC	-0.8415	0.0098	0.89
22	MC	-1.4722	0.0108	0.93
23	MC	-0.2121	0.0092	1.24
24	MC	-1.1807	0.0102	1.14
25	MC	1.6195	0.0093	0.83
26	XI	-0.6284	0.0095	0.81
27	XI	-1.3803	0.0106	0.86
28	XI	1.8402	0.0095	1.06
29	MC	-0.0694	0.0091	1.05
30	XI	0.6796	0.0089	1.37
31	MX	2.3280	0.0102	0.94
32	MC	1.0975	0.0090	1.04
33	MC	-1.3793	0.0106	0.97

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
34	XI	0.6133	0.0089	0.86
35	XI	-0.8684	0.0098	0.85
36	XI	-1.2121	0.0103	0.72
37	MC	-0.7664	0.0097	0.91
38	MC	0.8026	0.0089	1.20
39	XI	0.3761	0.0089	0.83
40	MC	2.6667	0.0109	0.91
41	MC	-0.7203	0.0096	1.00
42	XI	1.0725	0.0090	0.97
43	MC	1.5406	0.0092	0.82
44	MC	0.0749	0.0090	0.83
45	MC	0.2169	0.0089	1.00

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.8. Item-Level IRT Statistics, Mathematics Grade 4

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	MC	-1.7361	0.0101	0.93
2	MX	-0.5185	0.0087	1.14
3	XI	0.8284	0.0090	0.83
4	MC	-0.3864	0.0087	1.06
5	MC	0.7896	0.0089	1.31
6	MX	1.2425	0.0093	0.96
7	MC	-1.4709	0.0096	1.00
8	MC	-1.0142	0.0091	1.01
9	XI	-0.4122	0.0087	0.92
10	MC	-0.2174	0.0087	0.96
11	MC	0.0377	0.0086	1.12
12	MC	-0.0941	0.0086	1.40
13	XI	-0.8413	0.0089	0.82
14	MC	1.9534	0.0103	0.95
15	MC	0.4229	0.0087	1.08
16	XI	0.1355	0.0087	0.87
17	XI	0.9143	0.0090	0.90
18	MC	0.5958	0.0088	1.16
19	XI	-1.0077	0.0091	0.92
20	XI	-1.0726	0.0091	0.85
21	MC	-0.4650	0.0087	0.94
22	MC	0.1094	0.0087	1.37
23	XI	-1.1063	0.0092	0.83
24	MC	-1.8328	0.0102	0.93
25	XI	1.5507	0.0097	0.91
26	MC	-0.3345	0.0087	1.07
27	MC	-1.3555	0.0095	0.97
28	XI	1.7485	0.0100	0.91
29	MC	-1.0561	0.0091	0.99

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
30	XI	0.5770	0.0088	0.90
31	XI	-1.1904	0.0093	1.08
32	MC	0.2162	0.0087	0.94
33	MC	1.1661	0.0092	1.35
34	MC	2.6812	0.0119	1.06
35	XI	1.9322	0.0103	1.01
36	MC	-0.3783	0.0087	1.21
37	MC	-1.4487	0.0096	0.90
38	XI	-0.1478	0.0086	0.78
39	MC	1.2512	0.0093	0.86
40	XI	1.1837	0.0093	0.89
41	XI	-1.3734	0.0095	0.74
42	MC	1.0968	0.0092	1.06
43	XI	0.1922	0.0087	1.06
44	MC	-0.0153	0.0086	1.17
45	XI	1.0134	0.0091	0.77

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.9. Item-Level IRT Statistics, Mathematics Grade 5

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	XI	-1.4512	0.0089	0.93
2	XI	-0.4958	0.0082	0.87
3	MC	-0.6147	0.0082	1.04
4	MC	-0.1505	0.0082	1.06
5	MX	0.8991	0.0089	0.99
6	XI	0.4045	0.0084	0.83
7	MC	0.4356	0.0085	1.15
8	MC	1.0408	0.0091	0.98
9	MC	-0.2566	0.0082	1.15
10	MC	-0.3592	0.0082	0.95
11	MC	-0.6665	0.0082	0.93
12	MC	-0.1920	0.0082	1.16
13	XI	-0.2746	0.0082	0.94
14	MX	0.3740	0.0084	1.00
15	XI	-0.0237	0.0082	0.90
16	XI	1.0037	0.0091	0.84
17	MC	-0.9310	0.0084	0.87
18	MC	0.0327	0.0082	1.20
19	XI	0.7229	0.0087	0.82
20	MC	-0.1986	0.0082	1.09
21	XI	-1.1367	0.0086	0.86
22	MC	0.7844	0.0088	1.06
23	MC	-0.2750	0.0082	1.02
24	MC	0.2690	0.0083	0.82
25	MC	-0.8610	0.0083	1.07

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
26	XI	1.2923	0.0095	0.87
27	MC	-0.8369	0.0083	1.15
28	XI	-0.6364	0.0082	1.08
29	MC	-0.3931	0.0082	0.88
30	MX	0.3888	0.0084	1.16
31	MC	1.2197	0.0094	1.06
32	MC	0.0794	0.0083	0.96
33	XI	-0.0006	0.0082	1.12
34	XI	0.8567	0.0089	0.97
35	XI	1.9380	0.0108	0.92
36	MC	-1.0038	0.0084	0.95
37	XI	1.5369	0.0099	0.78
38	MC	-0.1844	0.0082	1.07
39	MC	0.1392	0.0083	1.16
40	XI	0.4315	0.0085	1.06
41	MC	0.4238	0.0085	1.27
42	MC	-1.4873	0.0090	0.83
43	MC	0.9331	0.0090	1.17
44	XI	0.0893	0.0083	0.88
45	MC	0.2494	0.0083	1.10

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.10. Item-Level IRT Statistics, Mathematics Grade 6

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	MC	-1.8247	0.0091	0.91
2	MC	-0.5536	0.0084	1.10
3	XI	-0.1486	0.0085	0.87
4	MC	-1.3975	0.0087	1.28
5	MC	-1.3801	0.0087	1.06
6	MC	-0.5048	0.0084	0.76
7	XI	1.0509	0.0097	0.84
8	XI	0.1530	0.0087	0.94
9	MC	0.3107	0.0088	1.09
10	MC	0.5102	0.0090	1.26
11	MC	-0.4752	0.0084	1.16
12	MC	0.3241	0.0088	1.24
13	MC	0.0038	0.0086	0.99
14	MC	0.7981	0.0094	1.25
15	MC	0.0521	0.0086	1.35
16	XI	0.9508	0.0096	0.96
17	MC	0.3665	0.0089	1.28
18	MC	1.0668	0.0098	1.01
19	XI	0.8057	0.0094	0.79
20	MC	0.0743	0.0086	1.17
21	XI	1.4973	0.0106	0.95

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
22	XI	0.2149	0.0087	0.86
23	MC	-0.3505	0.0084	1.10
24	MC	-1.6313	0.0089	0.96
25	XI	1.0921	0.0098	0.83
26	XI	-0.6209	0.0084	0.85
27	XI	-0.5250	0.0084	0.79
28	MC	0.6070	0.0091	0.98
29	MC	-0.1403	0.0085	0.95
30	MX	-1.0507	0.0085	0.86
31	MC	-0.2277	0.0085	1.17
32	MC	-0.2140	0.0085	1.36
33	XI	0.4728	0.0090	1.11
34	XI	0.5234	0.0090	0.80
35	XI	-0.3649	0.0084	0.87
36	XI	0.5220	0.0090	0.85
37	XI	-0.6890	0.0084	0.81
38	MC	0.3367	0.0088	0.98
39	XI	-0.7318	0.0084	1.05
40	MC	0.4962	0.0090	1.03
41	XI	1.6132	0.0109	0.99
42	MC	-0.1791	0.0085	0.96
43	XI	0.3913	0.0089	0.91
44	XI	1.7902	0.0113	0.80
45	MC	-1.7560	0.0091	0.79
46	XI	0.0398	0.0086	0.80
47	MC	-1.0889	0.0085	1.07

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.11. Item-Level IRT Statistics, Mathematics Grade 7

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	MC	-1.4840	0.0088	0.84
2	MC	-0.5590	0.0083	1.19
3	XI	0.3035	0.0087	0.87
4	MC	0.2586	0.0086	1.26
5	XI	1.2041	0.0097	0.91
6	MC	-0.5669	0.0083	1.24
7	XI	0.2534	0.0087	0.79
8	MC	1.5711	0.0104	1.08
9	XI	1.8888	0.0112	0.95
10	XI	-0.1984	0.0084	0.85
11	MC	0.4638	0.0088	1.24
12	XI	0.6692	0.0090	1.04
13	MC	-1.0690	0.0085	0.97
14	MC	-0.4913	0.0083	1.20
15	XI	0.5737	0.0089	0.73

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
16	MC	0.4523	0.0088	1.11
17	XI	1.1117	0.0096	0.80
18	XI	0.2634	0.0086	0.92
19	XI	1.1041	0.0096	0.83
20	MC	-1.5791	0.0089	0.81
21	MC	-1.2652	0.0086	0.96
22	XI	0.3351	0.0087	1.01
23	XI	-0.7828	0.0084	0.80
24	MC	-1.2477	0.0086	1.02
25	XI	0.9471	0.0094	0.98
26	MC	-0.5221	0.0083	1.23
27	MC	-0.8100	0.0084	1.12
28	MC	0.1140	0.0085	1.09
29	XI	1.5742	0.0104	0.95
30	MC	-0.0491	0.0085	0.99
31	MC	0.0916	0.0085	1.17
32	XI	0.6263	0.0090	0.80
33	MC	-0.4477	0.0084	1.08
34	XI	1.1255	0.0096	0.79
35	XI	0.8707	0.0093	0.95
36	MC	0.3188	0.0087	1.04
37	XI	1.0104	0.0094	1.01
38	MC	-1.3826	0.0087	0.84
39	MC	-0.7253	0.0084	1.04
40	XI	1.5495	0.0104	0.84
41	MC	0.2225	0.0086	0.84
42	MC	-1.5448	0.0088	1.20
43	MC	-0.7319	0.0084	1.14
44	MC	-0.5406	0.0083	1.15
45	XI	0.8466	0.0092	1.24
46	MC	-1.6496	0.0090	0.79
47	MX	1.0085	0.0094	1.29

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.12. Item-Level IRT Statistics, Mathematics Grade 8

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
1	MC	-1.6084	0.0083	0.97
2	XI	0.1888	0.0086	0.82
3	XI	-0.0608	0.0084	0.83
4	MC	0.8711	0.0097	1.07
5	MC	-0.5165	0.0081	0.92
6	MC	-0.8414	0.0080	0.91
7	MC	0.8483	0.0096	1.15
8	MC	0.5570	0.0091	1.25
9	MX	0.2859	0.0087	0.97

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
10	MC	-2.2891	0.0093	0.92
11	MC	0.6845	0.0093	1.28
12	MX	-0.1094	0.0083	1.24
13	MC	0.0801	0.0085	1.05
14	XI	0.4998	0.0090	0.79
15	XI	1.0941	0.0101	0.81
16	MC	-1.2668	0.0081	1.08
17	MX	-0.1611	0.0083	1.06
18	MC	0.8512	0.0096	0.93
19	XI	0.8433	0.0096	1.09
20	XI	1.9673	0.0125	0.87
21	MC	-0.5707	0.0081	0.95
22	MC	-1.1337	0.0080	0.82
23	MC	-0.6997	0.0080	1.11
24	MC	-0.5078	0.0081	1.19
25	XI	-0.5994	0.0080	1.16
26	MC	-0.3731	0.0081	1.19
27	XI	1.2197	0.0104	0.87
28	MC	-2.4253	0.0096	0.93
29	MC	-2.0838	0.0089	0.99
30	MC	-0.6912	0.0080	1.01
31	MC	1.1179	0.0102	0.80
32	MC	-1.6092	0.0083	0.94
33	XI	0.7936	0.0095	0.84
34	MC	-0.2171	0.0082	0.92
35	XI	-0.2976	0.0082	0.98
36	MC	-0.6213	0.0080	1.13
37	MC	-0.8777	0.0080	1.15
38	MC	-0.2507	0.0082	1.00
39	MC	1.5791	0.0113	0.98
40	MC	-0.3360	0.0082	0.97
41	MC	0.0265	0.0085	1.03
42	MC	1.0026	0.0099	1.17
43	XI	0.3671	0.0089	0.87
44	MC	0.1201	0.0086	0.80
45	MC	-0.4797	0.0081	1.14
46	XI	0.0565	0.0085	0.80
47	MC	0.7521	0.0094	1.12

Note. MC = multiple-choice, MX = multi-part, XI = technology-enhanced. Item number does not indicate item location on an operational test form, as field test items were embedded on the form but not included in the analysis.

Table B.13. Raw-to-Scale Score Conversion, ELA Grade 3

Form	Raw Score	Scale Score	CSEM	Performance Level
1	2	2395	21	1
1	3	2395	21	1
1	4	2395	21	1
1	5	2406	18	1
1	6	2415	16	1
1	7	2423	15	1
1	8	2430	14	1
1	9	2435	13	1
1	10	2441	12	1
1	11	2445	12	1
1	12	2450	11	1
1	13	2454	11	1
1	14	2457	11	1
1	15	2461	10	1
1	16	2464	10	1
1	17	2468	10	1
1	18	2471	10	1
1	19	2474	10	1
1	20	2477	9	1
1	21	2480	9	1
1	22	2483	9	1
1	23	2486	9	1
1	24	2489	9	1
1	25	2491	9	1
1	26	2494	9	1
1	27	2497	9	2
1	28	2500	9	2
1	29	2502	9	2
1	30	2505	9	2
1	31	2509	9	3
1	32	2511	9	3
1	33	2514	9	3
1	34	2517	9	3
1	35	2520	10	3
1	36	2523	10	3
1	37	2526	10	3
1	38	2529	10	3
1	39	2533	10	3
1	40	2536	10	3
1	41	2541	11	4
1	42	2544	11	4
1	43	2548	11	4
1	44	2553	12	4
1	45	2558	12	4
1	46	2563	13	4
1	47	2569	14	4
1	48	2576	15	4
1	49	2584	16	4

Form	Raw Score	Scale Score	CSEM	Performance Level
1	50	2593	17	4
1	51	2603	19	4
1	52	2605	19	4
1	53	2605	19	4
1	54	2605	19	4
1	55	2605	19	4
2	2	2395	21	1
2	3	2395	21	1
2	4	2395	21	1
2	5	2405	18	1
2	6	2415	16	1
2	7	2423	15	1
2	8	2429	14	1
2	9	2435	13	1
2	10	2440	12	1
2	11	2445	12	1
2	12	2449	11	1
2	13	2453	11	1
2	14	2457	10	1
2	15	2460	10	1
2	16	2464	10	1
2	17	2467	10	1
2	18	2470	10	1
2	19	2473	10	1
2	20	2476	9	1
2	21	2479	9	1
2	22	2482	9	1
2	23	2485	9	1
2	24	2487	9	1
2	25	2490	9	1
2	26	2493	9	1
2	27	2496	9	1
2	28	2498	9	2
2	29	2501	9	2
2	30	2504	9	2
2	31	2507	9	2
2	32	2509	9	3
2	33	2512	9	3
2	34	2515	9	3
2	35	2518	9	3
2	36	2521	10	3
2	37	2524	10	3
2	38	2528	10	3
2	39	2531	10	3
2	40	2534	10	3
2	41	2538	11	3
2	42	2542	11	4
2	43	2546	11	4
2	44	2550	12	4

Form	Raw Score	Scale Score	CSEM	Performance Level
2	45	2555	12	4
2	46	2560	13	4
2	47	2566	14	4
2	48	2573	14	4
2	49	2580	16	4
2	50	2589	17	4
2	51	2600	19	4
2	52	2605	20	4
2	53	2605	20	4
2	54	2605	20	4
2	55	2605	20	4

Table B.14. Raw-to-Scale Score Conversion, ELA Grade 4

Form	Raw Score	Scale Score	CSEM	Performance Level
1	2	2400	24	1
1	3	2400	24	1
1	4	2406	22	1
1	5	2420	18	1
1	6	2429	16	1
1	7	2437	15	1
1	8	2443	13	1
1	9	2449	13	1
1	10	2454	12	1
1	11	2459	12	1
1	12	2463	11	1
1	13	2467	11	1
1	14	2471	10	1
1	15	2474	10	1
1	16	2478	10	1
1	17	2481	10	1
1	18	2484	10	1
1	19	2487	10	1
1	20	2490	9	1
1	21	2493	9	1
1	22	2496	9	1
1	23	2499	9	1
1	24	2502	9	1
1	25	2504	9	1
1	26	2507	9	1
1	27	2510	9	2
1	28	2512	9	2
1	29	2515	9	2
1	30	2518	9	2
1	31	2520	9	2
1	32	2523	9	3
1	33	2526	9	3
1	34	2529	9	3
1	35	2532	9	3
1	36	2534	9	3

Form	Raw Score	Scale Score	CSEM	Performance Level
1	37	2537	10	3
1	38	2540	10	3
1	39	2544	10	3
1	40	2547	10	3
1	41	2550	10	3
1	42	2554	11	3
1	43	2559	11	4
1	44	2562	11	4
1	45	2566	12	4
1	46	2571	12	4
1	47	2576	13	4
1	48	2581	13	4
1	49	2588	14	4
1	50	2595	15	4
1	51	2603	17	4
1	52	2610	18	4
1	53	2610	18	4
1	54	2610	18	4
1	55	2610	18	4
1	56	2610	18	4
2	2	2400	24	1
2	3	2400	24	1
2	4	2406	22	1
2	5	2419	18	1
2	6	2429	16	1
2	7	2436	15	1
2	8	2443	13	1
2	9	2449	13	1
2	10	2454	12	1
2	11	2458	12	1
2	12	2462	11	1
2	13	2466	11	1
2	14	2470	10	1
2	15	2474	10	1
2	16	2477	10	1
2	17	2480	10	1
2	18	2484	10	1
2	19	2487	10	1
2	20	2490	9	1
2	21	2493	9	1
2	22	2496	9	1
2	23	2498	9	1
2	24	2501	9	1
2	25	2504	9	1
2	26	2507	9	1
2	27	2510	9	2
2	28	2512	9	2
2	29	2515	9	2
2	30	2518	9	2

Form	Raw Score	Scale Score	CSEM	Performance Level
2	31	2520	9	2
2	32	2523	9	3
2	33	2526	9	3
2	34	2529	9	3
2	35	2532	9	3
2	36	2535	10	3
2	37	2538	10	3
2	38	2541	10	3
2	39	2545	10	3
2	40	2548	10	3
2	41	2552	11	3
2	42	2556	11	3
2	43	2560	11	4
2	44	2564	12	4
2	45	2569	12	4
2	46	2574	13	4
2	47	2579	13	4
2	48	2586	14	4
2	49	2592	15	4
2	50	2600	16	4
2	51	2609	17	4
2	52	2610	17	4
2	53	2610	17	4
2	54	2610	17	4
2	55	2610	17	4
2	56	2610	17	4

Table B.15. Raw-to-Scale Score Conversion, ELA Grade 5

Form	Raw Score	Scale Score	CSEM	Performance Level
1	2	2419	22	1
1	3	2419	22	1
1	4	2419	22	1
1	5	2432	18	1
1	6	2442	16	1
1	7	2450	14	1
1	8	2456	13	1
1	9	2462	13	1
1	10	2467	12	1
1	11	2471	11	1
1	12	2475	11	1
1	13	2479	11	1
1	14	2483	10	1
1	15	2486	10	1
1	16	2490	10	1
1	17	2493	10	1
1	18	2496	9	1
1	19	2499	9	1
1	20	2502	9	1
1	21	2504	9	1

Form	Raw Score	Scale Score	CSEM	Performance Level
1	22	2507	9	1
1	23	2510	9	1
1	24	2513	9	1
1	25	2515	9	1
1	26	2518	9	1
1	27	2521	9	2
1	28	2523	9	2
1	29	2526	9	2
1	30	2529	9	2
1	31	2531	9	2
1	32	2534	9	2
1	33	2537	9	2
1	34	2540	9	2
1	35	2543	9	3
1	36	2545	9	3
1	37	2548	10	3
1	38	2552	10	3
1	39	2555	10	3
1	40	2558	10	3
1	41	2562	10	3
1	42	2565	11	3
1	43	2569	11	3
1	44	2573	11	3
1	45	2578	12	4
1	46	2583	12	4
1	47	2588	13	4
1	48	2594	14	4
1	49	2601	15	4
1	50	2609	16	4
1	51	2619	18	4
1	52	2629	20	4
1	53	2629	20	4
1	54	2629	20	4
1	55	2629	20	4
2	2	2419	21	1
2	3	2419	21	1
2	4	2419	21	1
2	5	2431	18	1
2	6	2440	16	1
2	7	2448	15	1
2	8	2454	13	1
2	9	2460	13	1
2	10	2465	12	1
2	11	2470	11	1
2	12	2474	11	1
2	13	2478	11	1
2	14	2482	10	1
2	15	2485	10	1
2	16	2488	10	1

Form	Raw Score	Scale Score	CSEM	Performance Level
2	17	2492	10	1
2	18	2495	10	1
2	19	2498	9	1
2	20	2501	9	1
2	21	2503	9	1
2	22	2506	9	1
2	23	2509	9	1
2	24	2512	9	1
2	25	2515	9	1
2	26	2517	9	1
2	27	2520	9	2
2	28	2523	9	2
2	29	2525	9	2
2	30	2528	9	2
2	31	2531	9	2
2	32	2533	9	2
2	33	2536	9	2
2	34	2539	9	2
2	35	2543	9	3
2	36	2545	10	3
2	37	2548	10	3
2	38	2551	10	3
2	39	2555	10	3
2	40	2558	10	3
2	41	2562	10	3
2	42	2565	11	3
2	43	2569	11	3
2	44	2574	12	3
2	45	2578	12	4
2	46	2583	13	4
2	47	2589	13	4
2	48	2595	14	4
2	49	2602	15	4
2	50	2611	17	4
2	51	2621	19	4
2	52	2629	20	4
2	53	2629	20	4
2	54	2629	20	4
2	55	2629	20	4

Table B.16. Raw-to-Scale Score Conversion, ELA Grade 6

Form	Raw Score	Scale Score	CSEM	Performance Level
1	2	2431	22	1
1	3	2431	22	1
1	4	2431	22	1
1	5	2443	18	1
1	6	2453	16	1
1	7	2460	14	1
1	8	2467	13	1

Form	Raw Score	Scale Score	CSEM	Performance Level
1	9	2473	13	1
1	10	2478	12	1
1	11	2482	11	1
1	12	2486	11	1
1	13	2490	11	1
1	14	2494	10	1
1	15	2498	10	1
1	16	2501	10	1
1	17	2504	10	1
1	18	2507	10	1
1	19	2510	9	1
1	20	2513	9	1
1	21	2516	9	1
1	22	2519	9	1
1	23	2522	9	1
1	24	2525	9	1
1	25	2527	9	1
1	26	2530	9	1
1	27	2533	9	2
1	28	2536	9	2
1	29	2538	9	2
1	30	2541	9	2
1	31	2544	9	2
1	32	2547	9	2
1	33	2549	9	2
1	34	2553	9	3
1	35	2555	9	3
1	36	2558	10	3
1	37	2561	10	3
1	38	2565	10	3
1	39	2568	10	3
1	40	2571	10	3
1	41	2575	10	3
1	42	2578	11	3
1	43	2582	11	3
1	44	2586	11	3
1	45	2591	12	3
1	46	2597	12	4
1	47	2601	13	4
1	48	2606	13	4
1	49	2612	14	4
1	50	2619	15	4
1	51	2628	17	4
1	52	2638	19	4
1	53	2641	19	4
1	54	2641	19	4
1	55	2641	19	4
2	2	2431	22	1
2	3	2431	22	1

Form	Raw Score	Scale Score	CSEM	Performance Level
2	4	2431	22	1
2	5	2443	18	1
2	6	2452	16	1
2	7	2460	15	1
2	8	2466	13	1
2	9	2472	13	1
2	10	2477	12	1
2	11	2482	12	1
2	12	2486	11	1
2	13	2490	11	1
2	14	2494	10	1
2	15	2497	10	1
2	16	2501	10	1
2	17	2504	10	1
2	18	2507	10	1
2	19	2510	9	1
2	20	2513	9	1
2	21	2516	9	1
2	22	2519	9	1
2	23	2522	9	1
2	24	2524	9	1
2	25	2527	9	1
2	26	2530	9	1
2	27	2532	9	2
2	28	2535	9	2
2	29	2538	9	2
2	30	2541	9	2
2	31	2543	9	2
2	32	2546	9	2
2	33	2549	9	2
2	34	2553	9	3
2	35	2555	9	3
2	36	2558	10	3
2	37	2561	10	3
2	38	2564	10	3
2	39	2567	10	3
2	40	2571	10	3
2	41	2574	10	3
2	42	2578	11	3
2	43	2582	11	3
2	44	2586	11	3
2	45	2591	12	3
2	46	2597	12	4
2	47	2601	13	4
2	48	2607	14	4
2	49	2614	15	4
2	50	2622	16	4
2	51	2631	17	4
2	52	2641	19	4

Form	Raw Score	Scale Score	CSEM	Performance Level
2	53	2641	19	4
2	54	2641	19	4
2	55	2641	19	4

Table B.17. Raw-to-Scale Score Conversion, ELA Grade 7

Form	Raw Score	Scale Score	CSEM	Performance Level
1	2	2438	23	1
1	3	2438	23	1
1	4	2443	22	1
1	5	2456	18	1
1	6	2465	16	1
1	7	2472	14	1
1	8	2479	13	1
1	9	2484	12	1
1	10	2489	12	1
1	11	2493	11	1
1	12	2497	11	1
1	13	2501	10	1
1	14	2504	10	1
1	15	2508	10	1
1	16	2511	10	1
1	17	2514	9	1
1	18	2517	9	1
1	19	2520	9	1
1	20	2522	9	1
1	21	2525	9	1
1	22	2528	9	1
1	23	2530	9	1
1	24	2533	9	1
1	25	2536	9	1
1	26	2538	9	1
1	27	2541	9	1
1	28	2543	9	2
1	29	2546	9	2
1	30	2548	9	2
1	31	2551	9	2
1	32	2554	9	2
1	33	2556	9	2
1	34	2559	9	2
1	35	2562	9	3
1	36	2564	9	3
1	37	2567	9	3
1	38	2570	9	3
1	39	2573	10	3
1	40	2576	10	3
1	41	2579	10	3
1	42	2583	10	3
1	43	2586	11	3
1	44	2590	11	3

Form	Raw Score	Scale Score	CSEM	Performance Level
1	45	2594	11	3
1	46	2600	12	4
1	47	2604	12	4
1	48	2609	13	4
1	49	2615	14	4
1	50	2622	15	4
1	51	2630	16	4
1	52	2640	19	4
1	53	2648	21	4
1	54	2648	21	4
1	55	2648	21	4
2	2	2438	23	1
2	3	2438	23	1
2	4	2443	22	1
2	5	2456	18	1
2	6	2465	16	1
2	7	2472	14	1
2	8	2478	13	1
2	9	2484	12	1
2	10	2489	12	1
2	11	2493	11	1
2	12	2497	11	1
2	13	2501	10	1
2	14	2504	10	1
2	15	2508	10	1
2	16	2511	10	1
2	17	2514	10	1
2	18	2517	9	1
2	19	2520	9	1
2	20	2522	9	1
2	21	2525	9	1
2	22	2528	9	1
2	23	2531	9	1
2	24	2533	9	1
2	25	2536	9	1
2	26	2538	9	1
2	27	2541	9	1
2	28	2543	9	2
2	29	2546	9	2
2	30	2549	9	2
2	31	2551	9	2
2	32	2554	9	2
2	33	2557	9	2
2	34	2559	9	2
2	35	2562	9	3
2	36	2565	9	3
2	37	2568	9	3
2	38	2571	10	3
2	39	2574	10	3

Form	Raw Score	Scale Score	CSEM	Performance Level
2	40	2577	10	3
2	41	2580	10	3
2	42	2584	10	3
2	43	2588	11	3
2	44	2592	11	3
2	45	2596	12	3
2	46	2601	12	4
2	47	2606	13	4
2	48	2612	14	4
2	49	2618	15	4
2	50	2626	16	4
2	51	2635	18	4
2	52	2647	20	4
2	53	2648	20	4
2	54	2648	20	4
2	55	2648	20	4

Table B.18. Raw-to-Scale Score Conversion, ELA Grade 8

Form	Raw Score	Scale Score	CSEM	Performance Level
1	2	2448	21	1
1	3	2448	21	1
1	4	2448	21	1
1	5	2457	18	1
1	6	2467	16	1
1	7	2475	15	1
1	8	2482	14	1
1	9	2487	13	1
1	10	2493	12	1
1	11	2497	12	1
1	12	2502	11	1
1	13	2506	11	1
1	14	2510	11	1
1	15	2513	10	1
1	16	2517	10	1
1	17	2520	10	1
1	18	2523	10	1
1	19	2526	10	1
1	20	2529	9	1
1	21	2532	9	1
1	22	2535	9	1
1	23	2538	9	1
1	24	2541	9	1
1	25	2543	9	1
1	26	2546	9	1
1	27	2549	9	1
1	28	2551	9	2
1	29	2554	9	2
1	30	2557	9	2
1	31	2559	9	2

Form	Raw Score	Scale Score	CSEM	Performance Level
1	32	2562	9	2
1	33	2565	9	2
1	34	2568	9	2
1	35	2570	9	2
1	36	2573	9	3
1	37	2576	9	3
1	38	2579	10	3
1	39	2582	10	3
1	40	2585	10	3
1	41	2589	10	3
1	42	2592	10	3
1	43	2596	11	3
1	44	2599	11	3
1	45	2604	11	4
1	46	2608	12	4
1	47	2612	12	4
1	48	2618	13	4
1	49	2623	14	4
1	50	2630	15	4
1	51	2638	16	4
1	52	2648	18	4
1	53	2658	21	4
1	54	2658	21	4
1	55	2658	21	4
2	2	2448	21	1
2	3	2448	21	1
2	4	2448	21	1
2	5	2459	18	1
2	6	2469	16	1
2	7	2477	15	1
2	8	2483	14	1
2	9	2489	13	1
2	10	2494	12	1
2	11	2499	12	1
2	12	2503	11	1
2	13	2507	11	1
2	14	2511	10	1
2	15	2514	10	1
2	16	2518	10	1
2	17	2521	10	1
2	18	2524	10	1
2	19	2527	9	1
2	20	2530	9	1
2	21	2533	9	1
2	22	2536	9	1
2	23	2538	9	1
2	24	2541	9	1
2	25	2544	9	1
2	26	2547	9	1

Form	Raw Score	Scale Score	CSEM	Performance Level
2	27	2549	9	1
2	28	2552	9	2
2	29	2555	9	2
2	30	2557	9	2
2	31	2560	9	2
2	32	2563	9	2
2	33	2565	9	2
2	34	2568	9	2
2	35	2572	9	3
2	36	2574	9	3
2	37	2577	9	3
2	38	2580	10	3
2	39	2583	10	3
2	40	2586	10	3
2	41	2589	10	3
2	42	2593	10	3
2	43	2597	11	3
2	44	2600	11	3
2	45	2605	11	4
2	46	2609	12	4
2	47	2614	12	4
2	48	2620	13	4
2	49	2626	14	4
2	50	2633	15	4
2	51	2641	17	4
2	52	2651	19	4
2	53	2658	20	4
2	54	2658	20	4
2	55	2658	20	4

Table B.19. Raw-to-Scale Score Conversion, Mathematics Grade 3

Raw Score	Scale Score	CSEM	Performance Level
0	3395	22	1
1	3395	22	1
2	3395	22	1
3	3408	19	1
4	3418	17	1
5	3427	15	1
6	3434	14	1
7	3440	13	1
8	3446	13	1
9	3451	12	1
10	3456	12	1
11	3461	12	1
12	3465	11	1
13	3469	11	1
14	3473	11	1
15	3477	11	1
16	3481	11	1

Raw Score	Scale Score	CSEM	Performance Level
17	3485	10	1
18	3488	10	1
19	3492	10	1
20	3495	10	2
21	3499	10	2
22	3502	10	2
23	3505	10	2
24	3509	10	2
25	3512	10	2
26	3516	10	2
27	3519	10	2
28	3523	10	2
29	3527	11	2
30	3531	11	3
31	3534	11	3
32	3538	11	3
33	3542	11	3
34	3547	12	3
35	3551	12	3
36	3556	12	3
37	3561	13	3
38	3567	13	3
39	3573	14	4
40	3580	15	4
41	3589	17	4
42	3599	19	4
43	3605	20	4
44	3605	20	4
45	3605	20	4

Table B.20. Raw-to-Scale Score Conversion, Mathematics Grade 4

Raw Score	Scale Score	CSEM	Performance Level
0	3435	22	1
1	3435	22	1
2	3435	22	1
3	3448	18	1
4	3458	16	1
5	3466	15	1
6	3473	14	1
7	3479	13	1
8	3485	13	1
9	3490	12	1
10	3494	12	1
11	3499	11	1
12	3503	11	1
13	3507	11	1
14	3511	11	1
15	3515	11	1
16	3518	10	1

Raw Score	Scale Score	CSEM	Performance Level
17	3522	10	1
18	3525	10	1
19	3530	10	2
20	3532	10	2
21	3535	10	2
22	3539	10	2
23	3542	10	2
24	3546	10	2
25	3549	10	2
26	3553	10	2
27	3556	10	2
28	3560	10	2
29	3563	10	3
30	3567	11	3
31	3571	11	3
32	3575	11	3
33	3579	11	3
34	3583	12	3
35	3588	12	3
36	3593	12	3
37	3598	13	3
38	3606	13	4
39	3610	14	4
40	3617	15	4
41	3625	17	4
42	3636	19	4
43	3645	21	4
44	3645	21	4
45	3645	21	4

Table B.21. Raw-to-Scale Score Conversion, Mathematics Grade 5

Raw Score	Scale Score	CSEM	Performance Level
0	3478	24	1
1	3478	24	1
2	3485	22	1
3	3498	18	1
4	3508	16	1
5	3516	15	1
6	3522	14	1
7	3528	13	1
8	3533	12	1
9	3538	12	1
10	3542	11	1
11	3547	11	1
12	3550	11	1
13	3554	10	1
14	3558	10	1
15	3561	10	1
16	3564	10	2

Raw Score	Scale Score	CSEM	Performance Level
17	3568	10	2
18	3571	10	2
19	3574	10	2
20	3577	10	2
21	3580	10	2
22	3583	10	2
23	3586	10	2
24	3589	10	2
25	3592	10	2
26	3595	10	3
27	3598	10	3
28	3602	10	3
29	3605	10	3
30	3608	10	3
31	3612	10	3
32	3615	10	3
33	3619	11	3
34	3623	11	3
35	3627	11	3
36	3631	12	3
37	3636	12	4
38	3641	13	4
39	3647	14	4
40	3654	15	4
41	3662	16	4
42	3672	18	4
43	3685	22	4
44	3688	23	4
45	3688	23	4

Table B.22. Raw-to-Scale Score Conversion, Mathematics Grade 6

Raw Score	Scale Score	CSEM	Performance Level
0	3512	23	1
1	3512	23	1
2	3514	22	1
3	3527	18	1
4	3537	16	1
5	3545	15	1
6	3552	14	1
7	3558	13	1
8	3563	12	1
9	3568	12	1
10	3572	11	1
11	3576	11	1
12	3580	11	1
13	3584	10	1
14	3587	10	1
15	3591	10	1
16	3594	10	1

Raw Score	Scale Score	CSEM	Performance Level
17	3598	10	1
18	3602	10	2
19	3604	10	2
20	3607	10	2
21	3610	9	2
22	3613	9	2
23	3616	9	2
24	3619	9	2
25	3622	9	2
26	3625	9	2
27	3629	10	3
28	3631	10	3
29	3634	10	3
30	3637	10	3
31	3640	10	3
32	3643	10	3
33	3647	10	3
34	3650	10	3
35	3654	11	3
36	3658	11	3
37	3663	11	4
38	3666	12	4
39	3671	12	4
40	3676	13	4
41	3682	14	4
42	3689	15	4
43	3697	16	4
44	3706	18	4
45	3720	22	4
46	3722	23	4
47	3722	23	4

Table B.23. Raw-to-Scale Score Conversion, Mathematics Grade 7

Raw Score	Scale Score	CSEM	Performance Level
0	3529	23	1
1	3529	23	1
2	3531	22	1
3	3544	18	1
4	3554	16	1
5	3562	15	1
6	3569	14	1
7	3575	13	1
8	3580	12	1
9	3585	12	1
10	3590	11	1
11	3594	11	1
12	3598	11	1
13	3602	11	1
14	3606	10	1

Raw Score	Scale Score	CSEM	Performance Level
15	3609	10	1
16	3613	10	1
17	3616	10	1
18	3619	10	1
19	3623	10	1
20	3626	10	1
21	3629	10	2
22	3632	10	2
23	3635	10	2
24	3638	10	2
25	3641	10	2
26	3644	10	2
27	3647	10	2
28	3652	10	3
29	3654	10	3
30	3657	10	3
31	3660	10	3
32	3664	10	3
33	3667	10	3
34	3671	11	3
35	3675	11	3
36	3680	11	4
37	3683	11	4
38	3687	12	4
39	3692	12	4
40	3697	13	4
41	3703	14	4
42	3710	15	4
43	3718	16	4
44	3728	18	4
45	3739	22	4
46	3739	22	4
47	3739	22	4

Table B.24. Raw-to-Scale Score Conversion, Mathematics Grade 8

Raw Score	Scale Score	CSEM	Performance Level
0	3566	21	1
1	3566	21	1
2	3566	21	1
3	3575	19	1
4	3585	16	1
5	3593	15	1
6	3600	14	1
7	3606	13	1
8	3612	13	1
9	3617	12	1
10	3621	12	1
11	3626	11	1
12	3630	11	1

Raw Score	Scale Score	CSEM	Performance Level
13	3634	11	1
14	3637	10	1
15	3641	10	1
16	3644	10	1
17	3648	10	1
18	3651	10	2
19	3654	10	2
20	3657	10	2
21	3661	10	2
22	3664	10	2
23	3667	10	2
24	3670	10	2
25	3673	10	3
26	3676	10	3
27	3679	10	3
28	3682	10	3
29	3685	10	3
30	3689	10	3
31	3692	10	3
32	3695	10	3
33	3699	10	3
34	3702	11	3
35	3706	11	4
36	3710	11	4
37	3714	11	4
38	3719	12	4
39	3724	12	4
40	3729	13	4
41	3735	14	4
42	3741	15	4
43	3749	16	4
44	3759	18	4
45	3773	22	4
46	3776	23	4
47	3776	23	4

Figure B.1. Item-Person Map, ELA Grade 3, Form 1

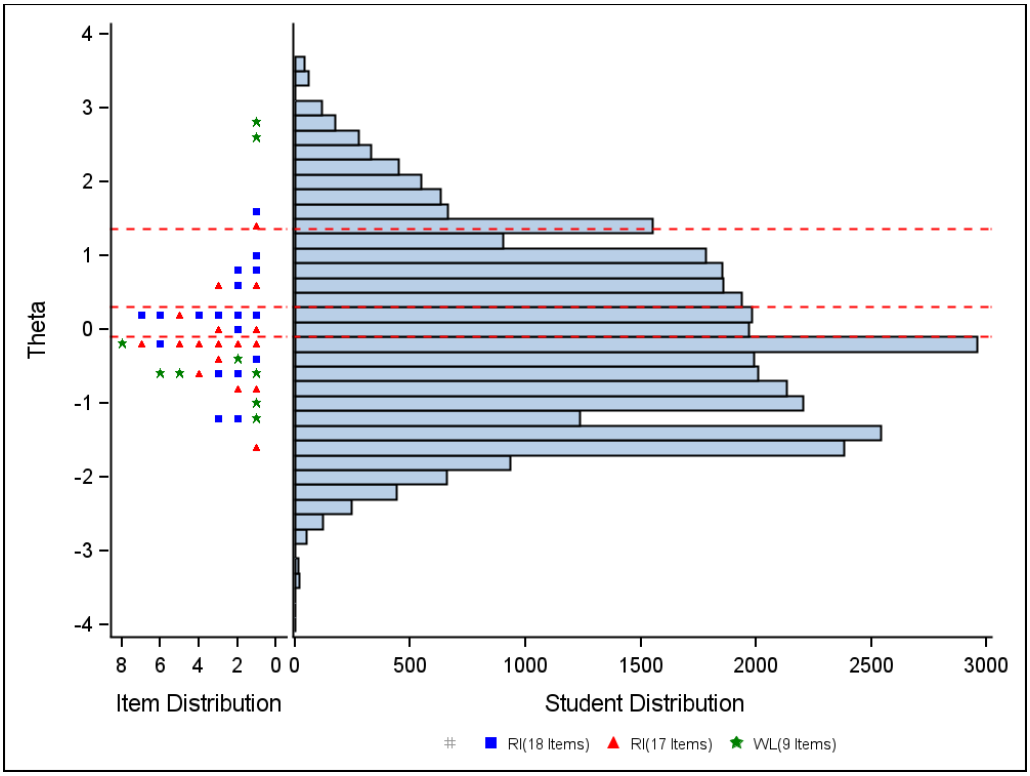


Figure B.2. Item-Person Map, ELA Grade 3, Form 2

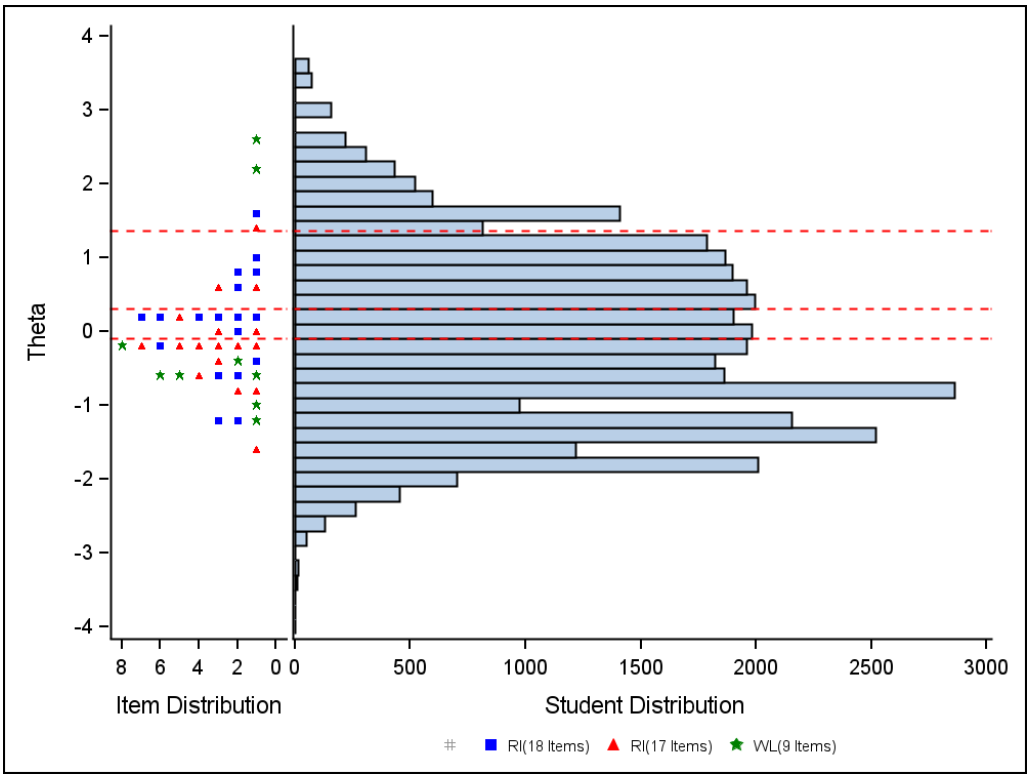


Figure B.3. Item-Person Map, ELA Grade 4, Form 1

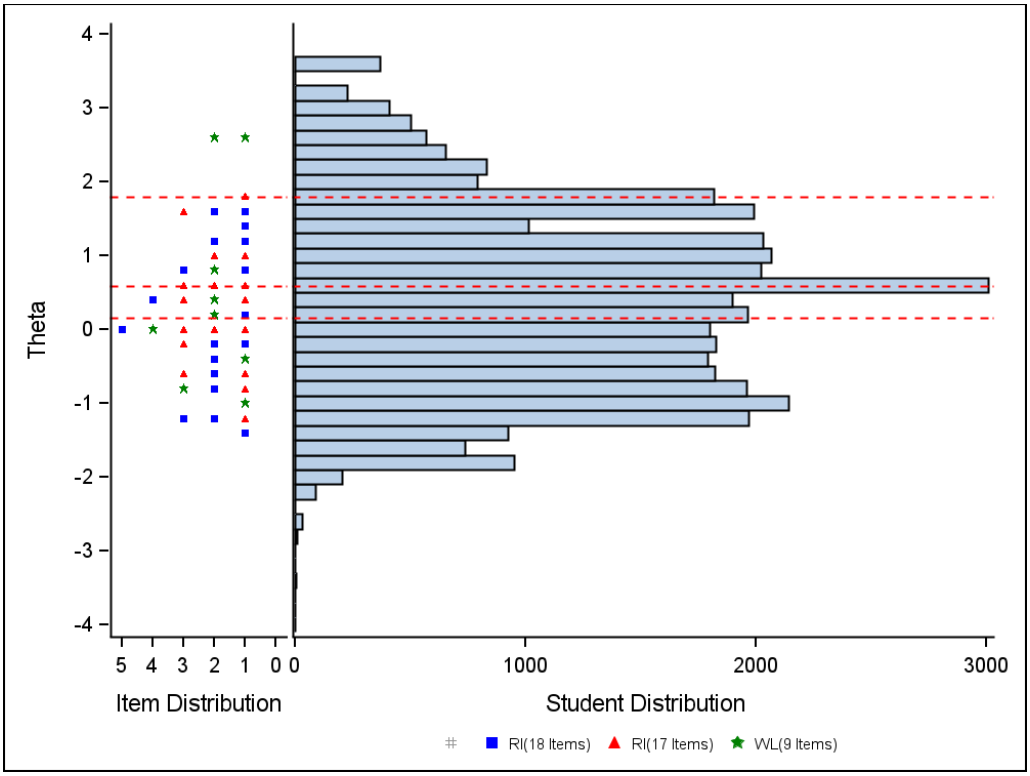


Figure B.4. Item-Person Map, ELA Grade 4, Form 2

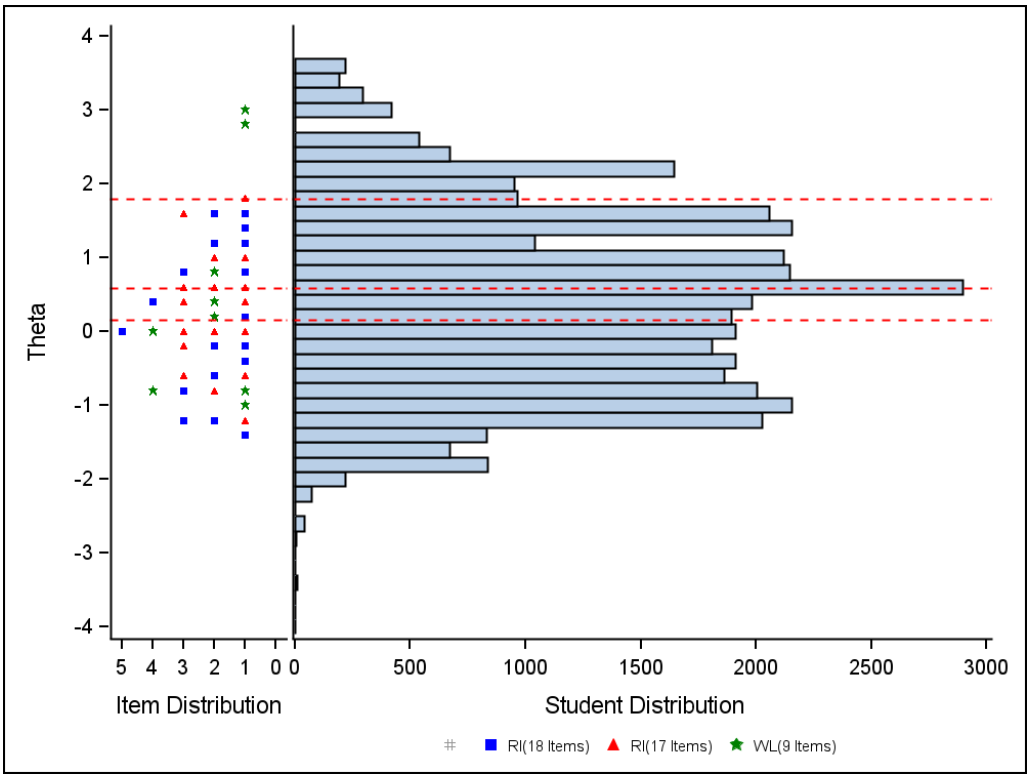


Figure B.5. Item-Person Map, ELA Grade 5, Form 1

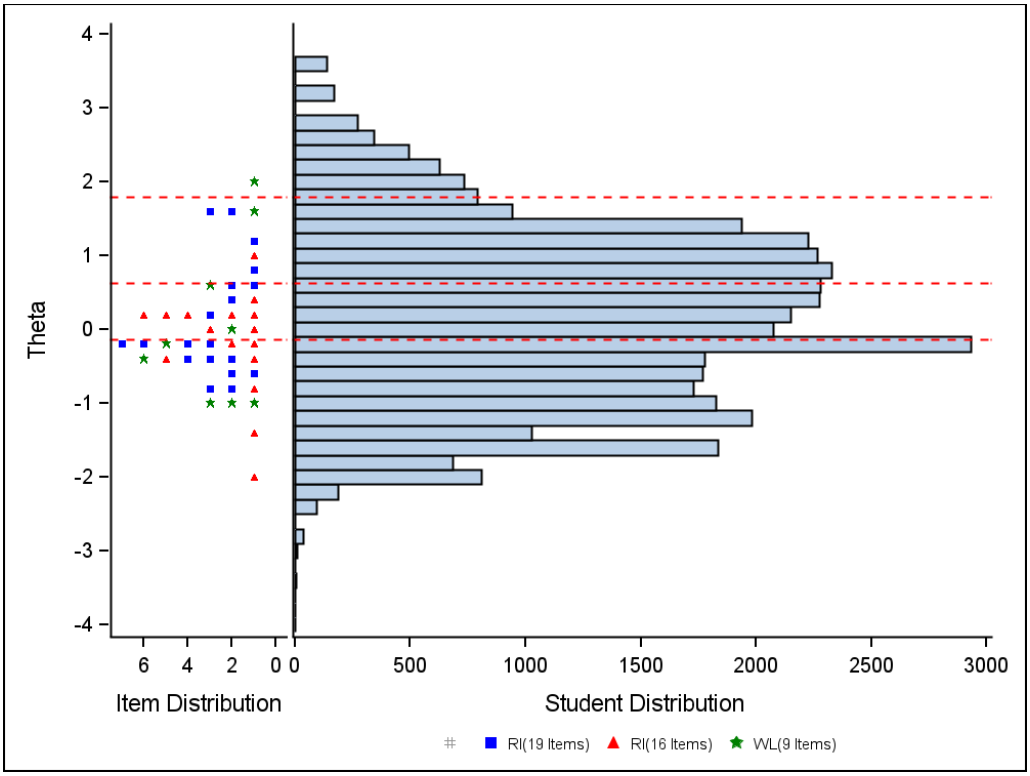


Figure B.6. Item-Person Map, ELA Grade 5, Form 2

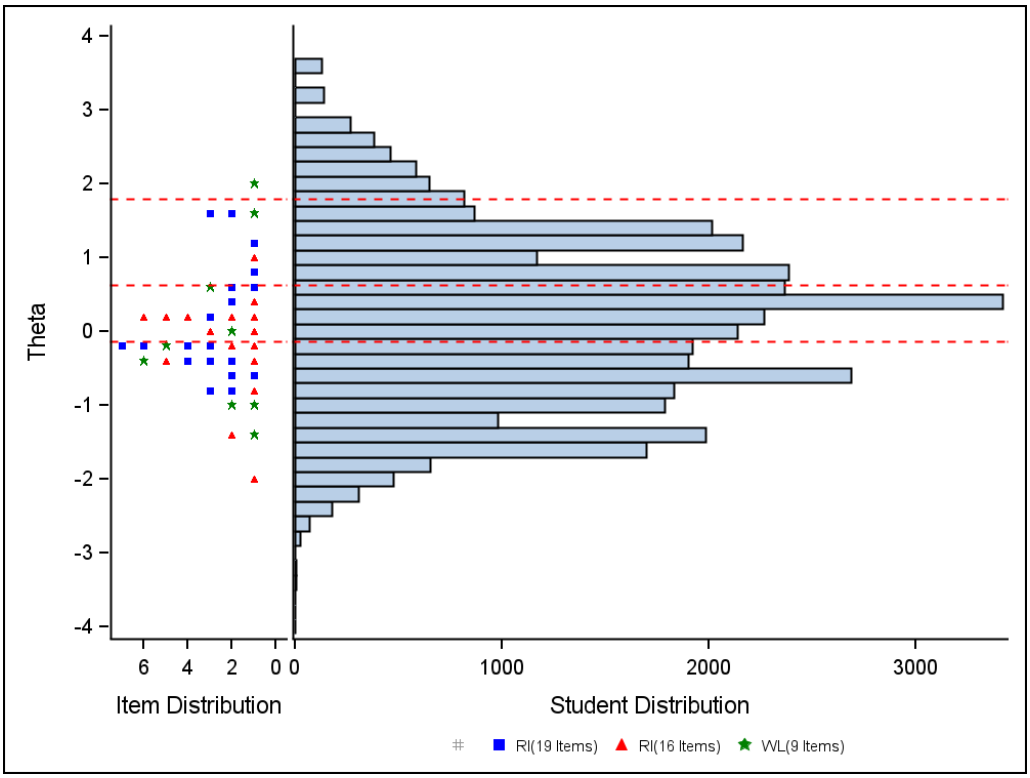


Figure B.7. Item-Person Map, ELA Grade 6, Form 1

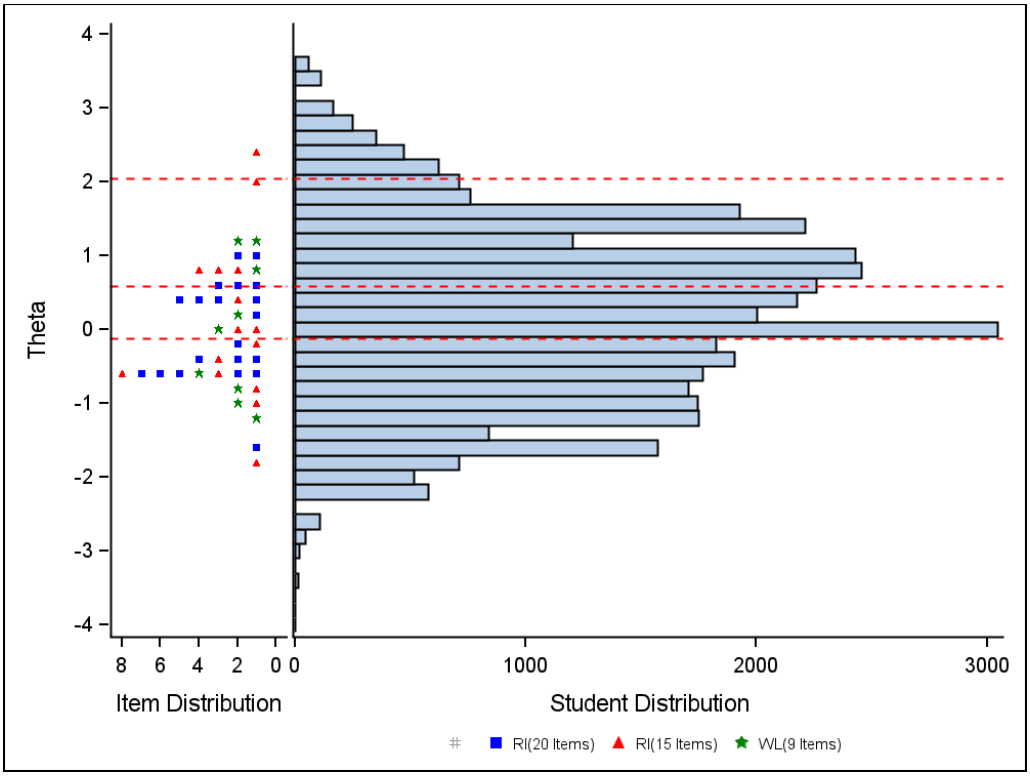


Figure B.8. Item-Person Map, ELA Grade 6, Form 2

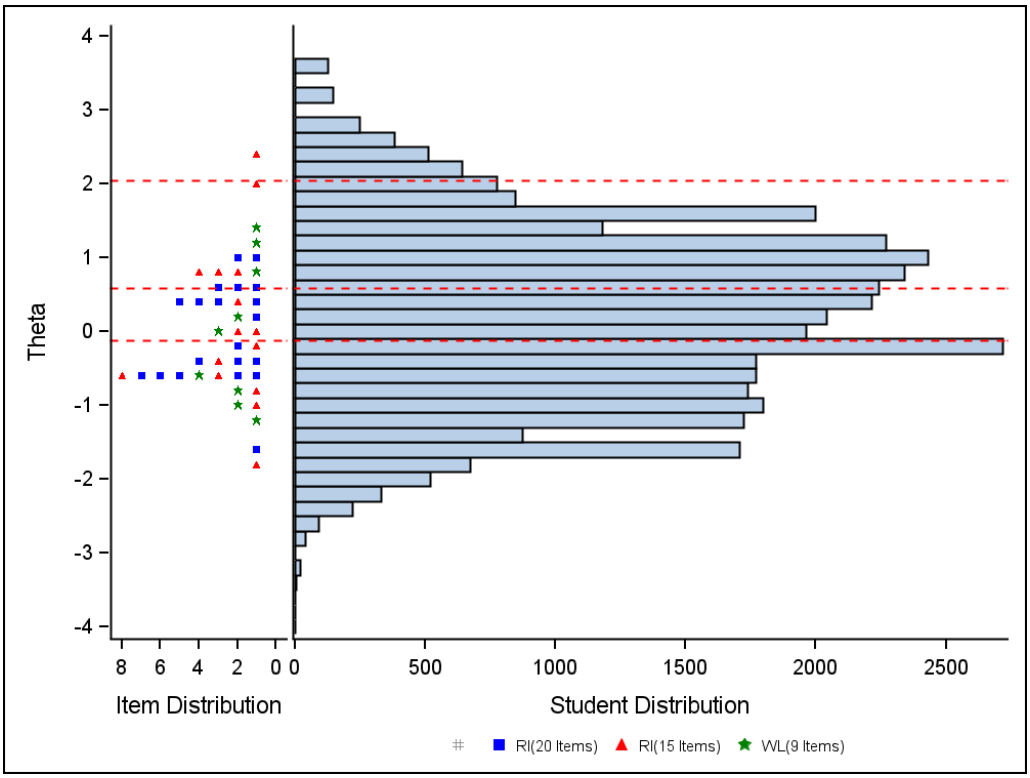


Figure B.9. Item-Person Map, ELA Grade 7, Form 1

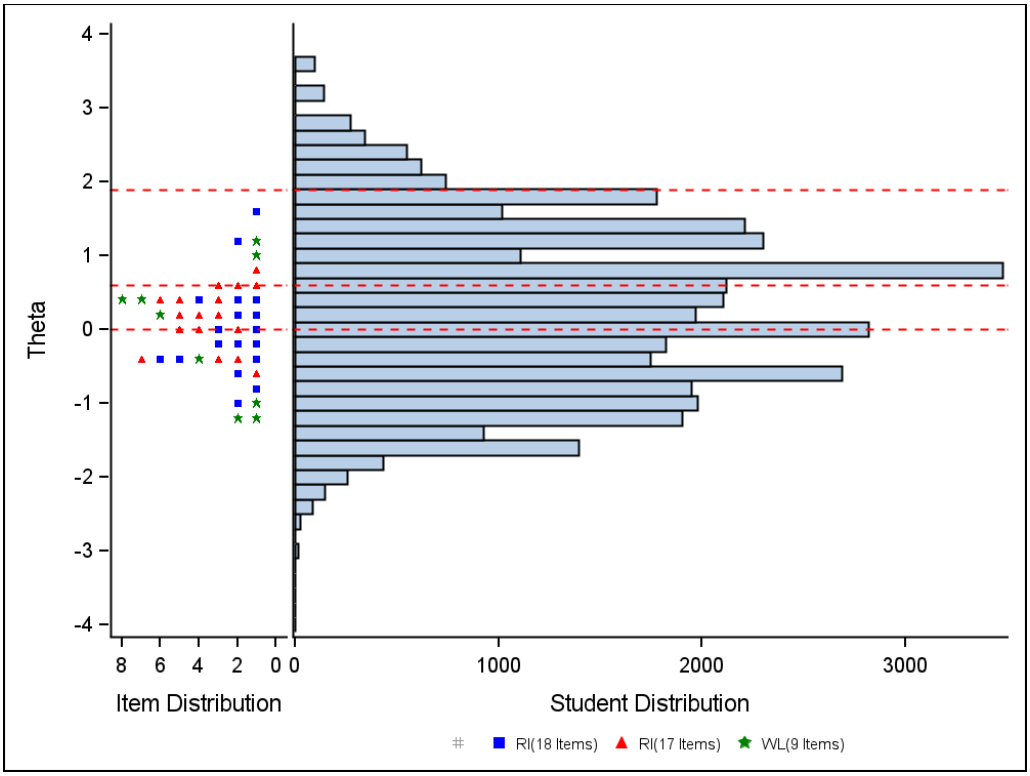


Figure B.10. Item-Person Map, ELA Grade 7, Form 2

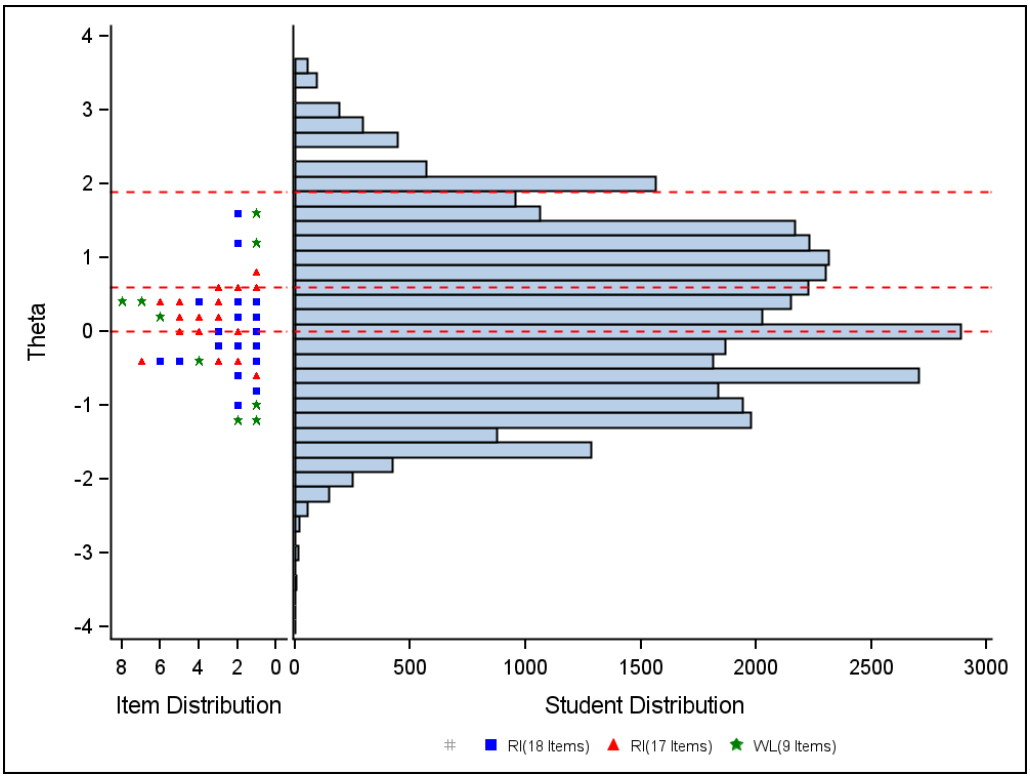


Figure B.11. Item-Person Map, ELA Grade 8, Form 1

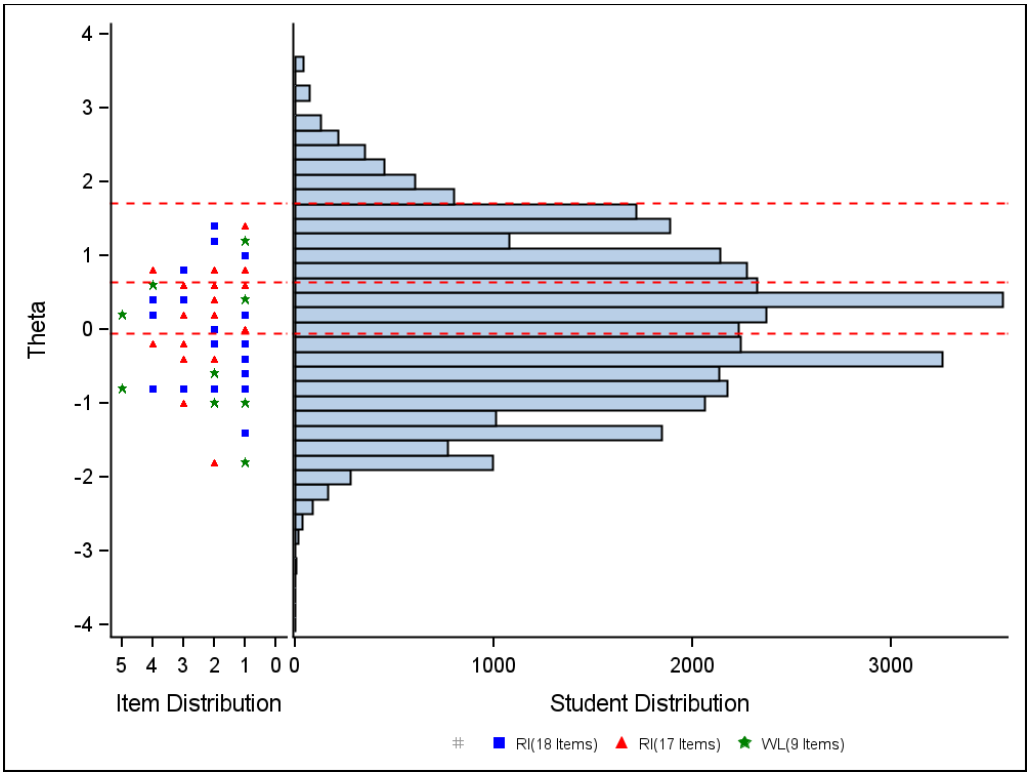


Figure B.12. Item-Person Map, ELA Grade 8, Form 2

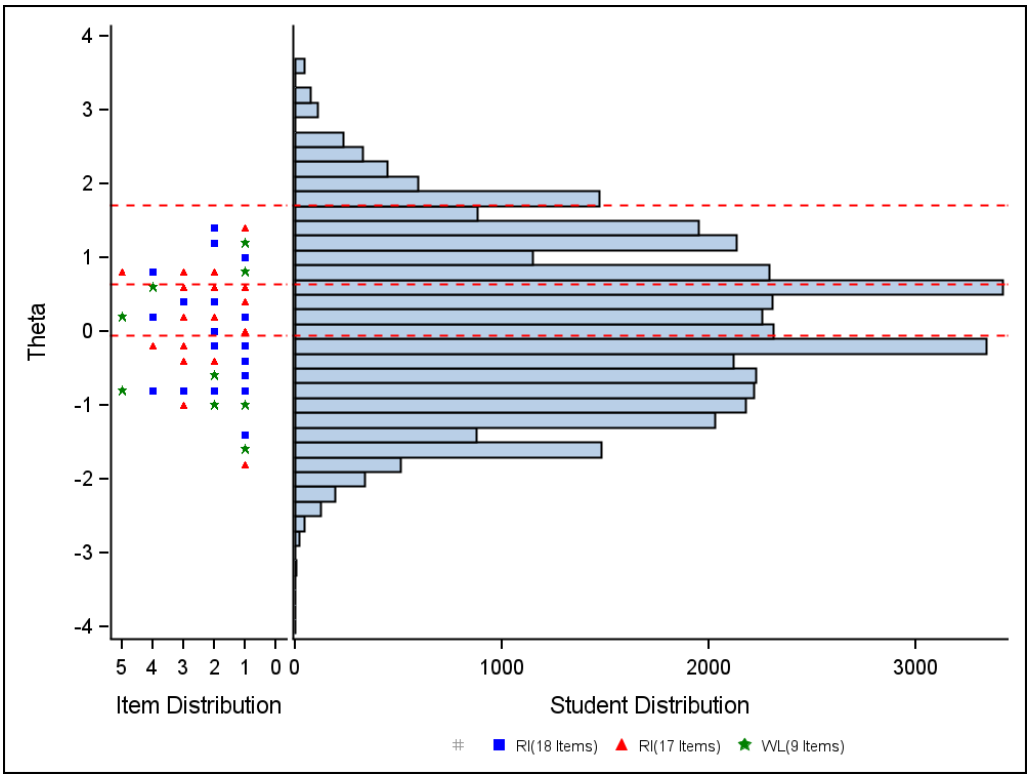


Figure B.13. Item-Person Map, Mathematics Grade 3

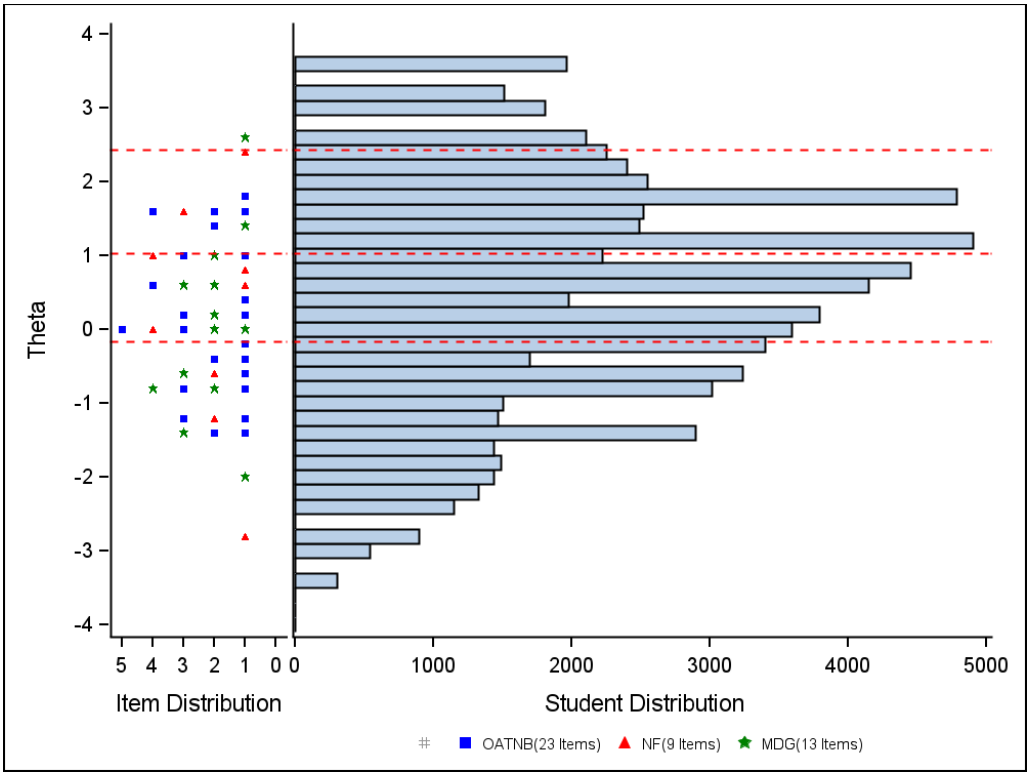


Figure B.14. Item-Person Map, Mathematics Grade 4

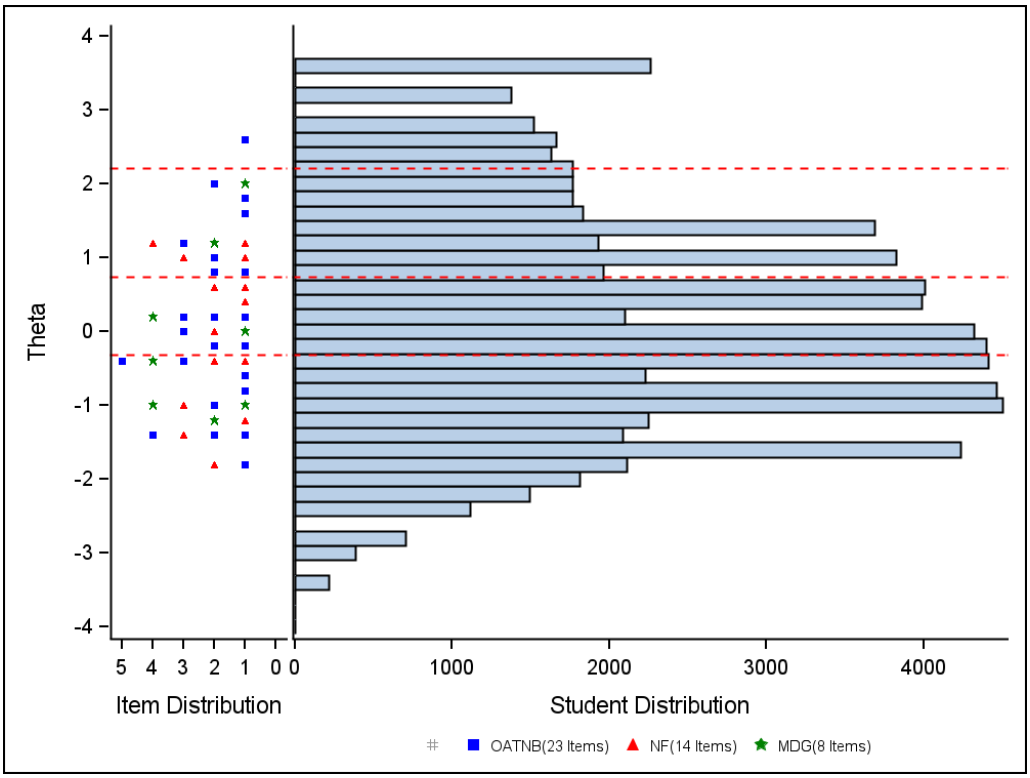


Figure B.15. Item-Person Map, Mathematics Grade 5

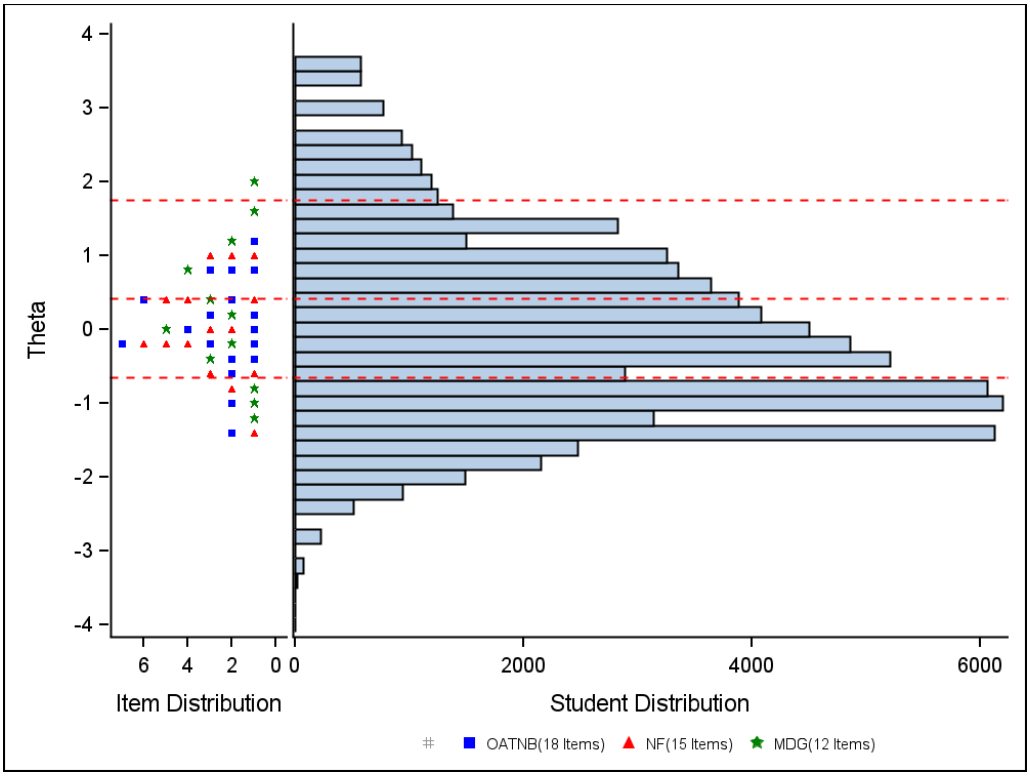


Figure B.16. Item-Person Map, Mathematics Grade 6

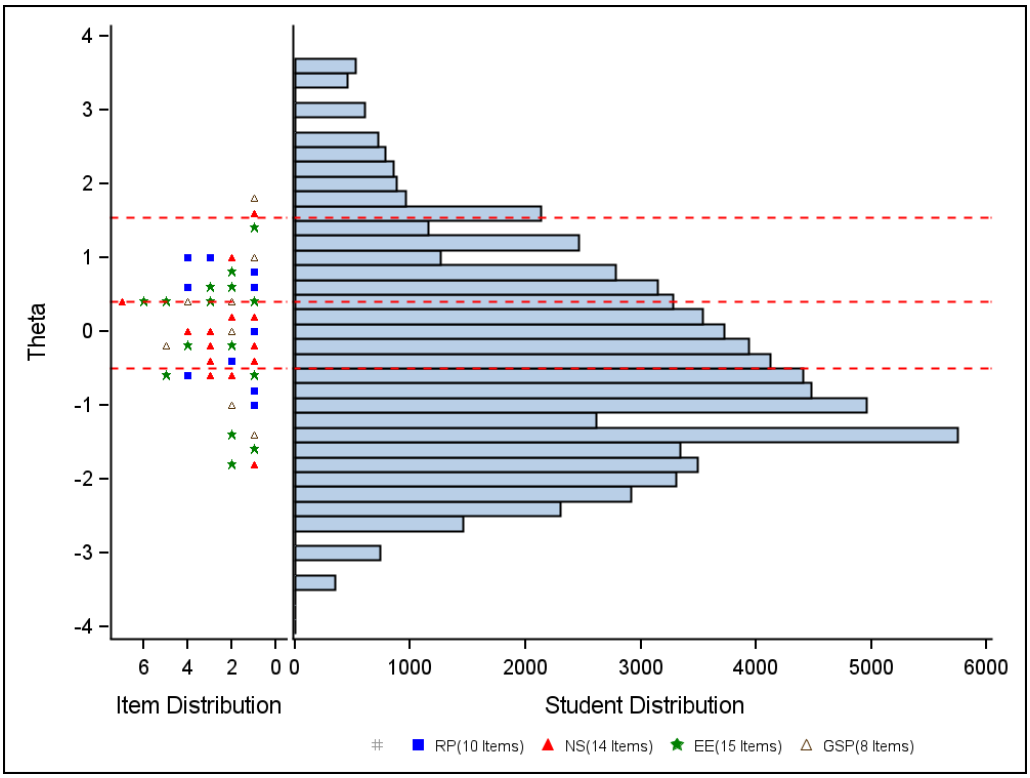


Figure B.17. Item-Person Map, Mathematics Grade 7

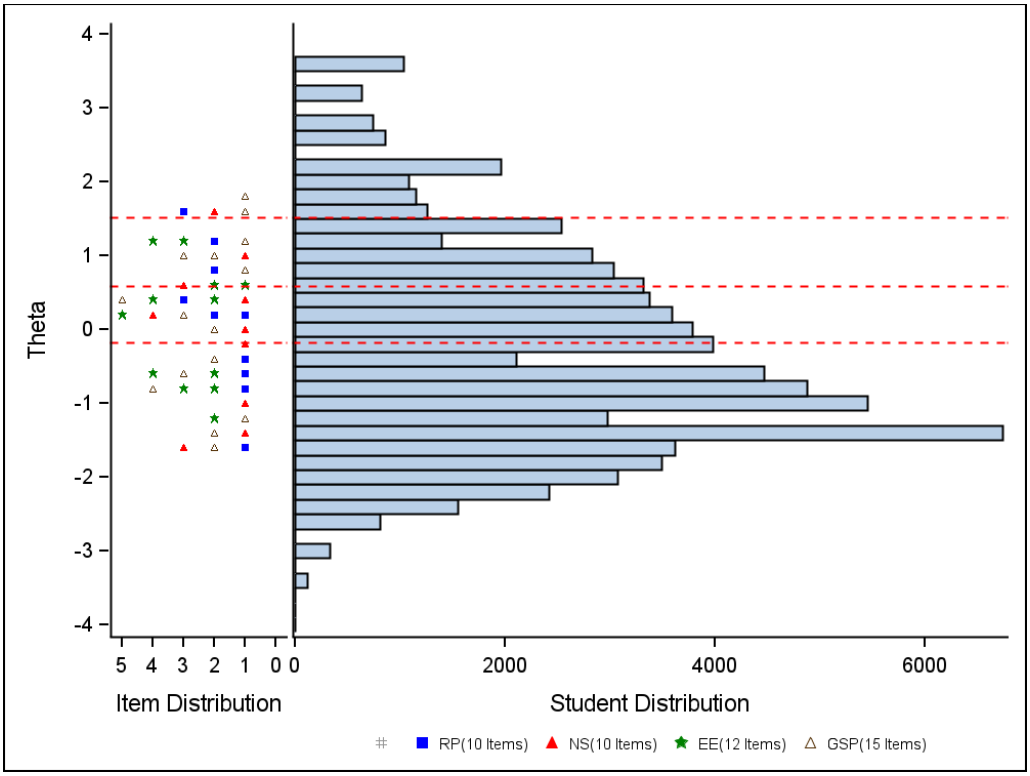


Figure B.18. Item-Person Map, Mathematics Grade 8

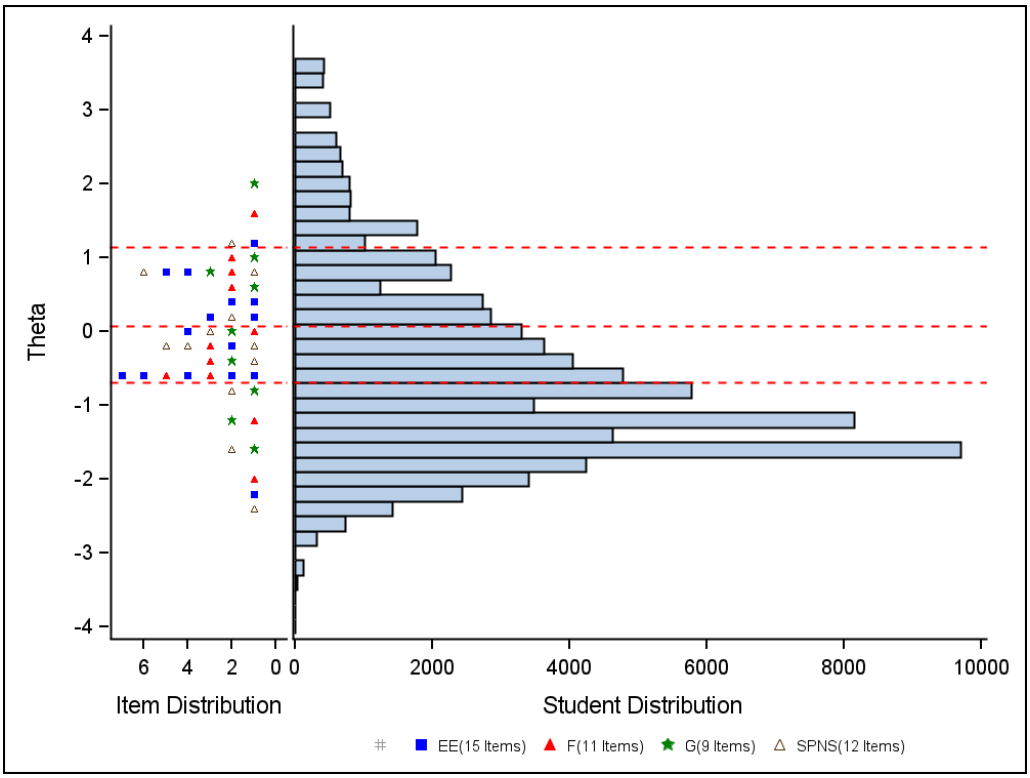


Figure B.19. TCC, ELA Grade 3, Form 1

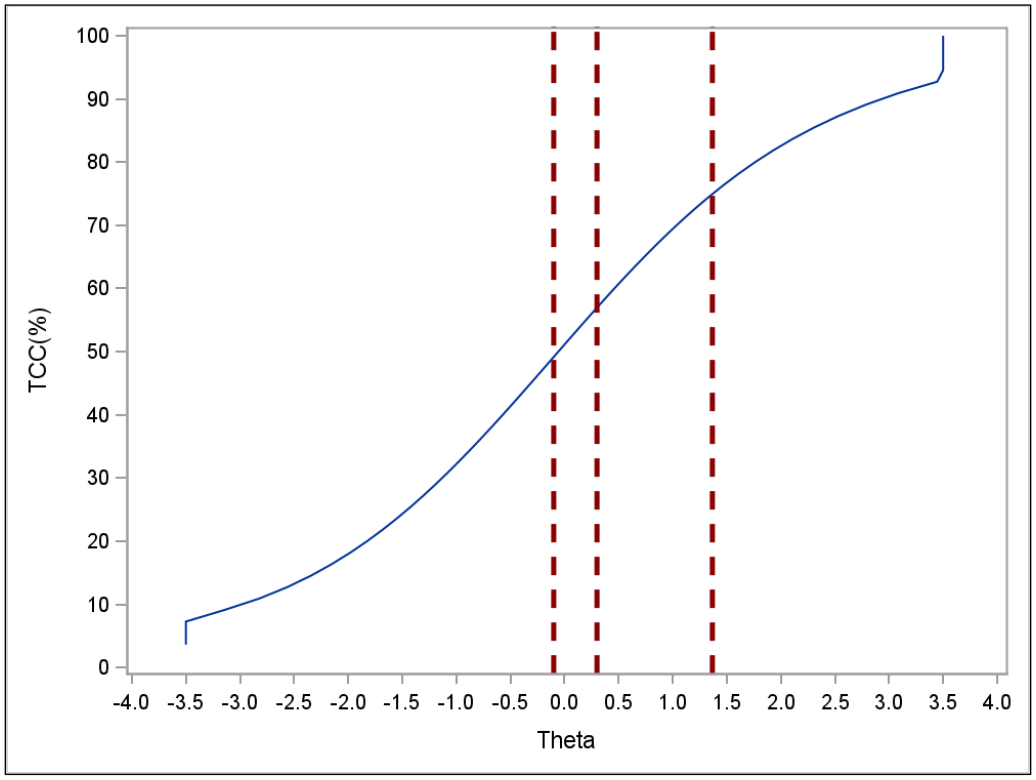


Figure B.20. CSEM, ELA Grade 3, Form 1

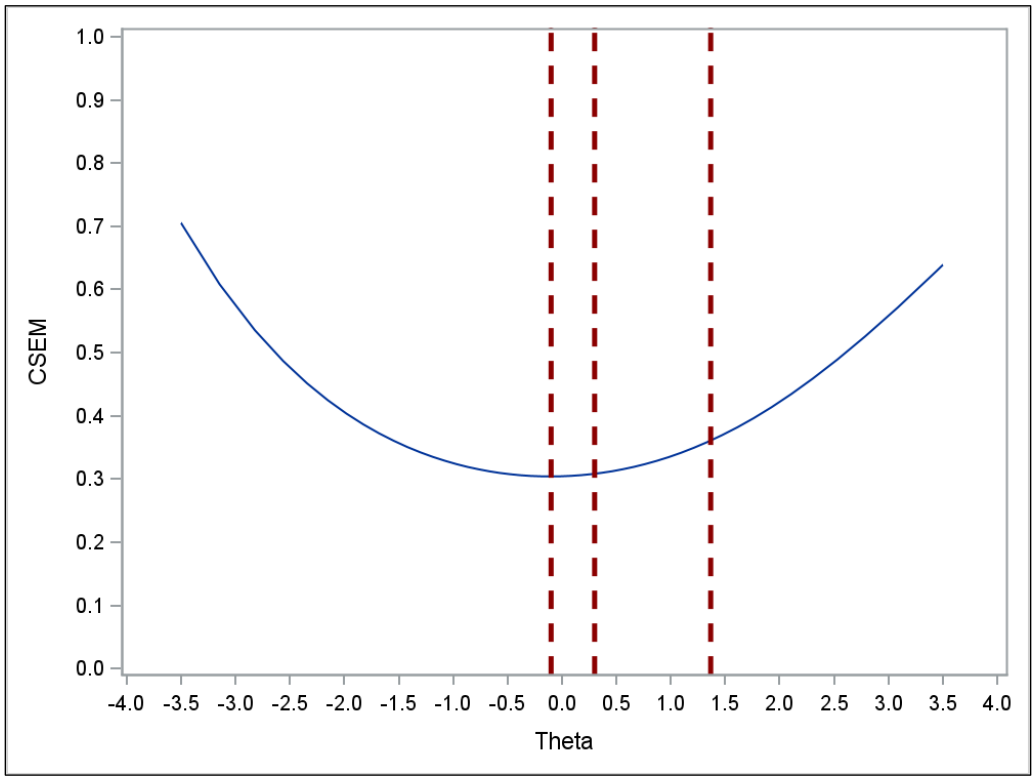


Figure B.21. TCC, ELA Grade 4, Form 1

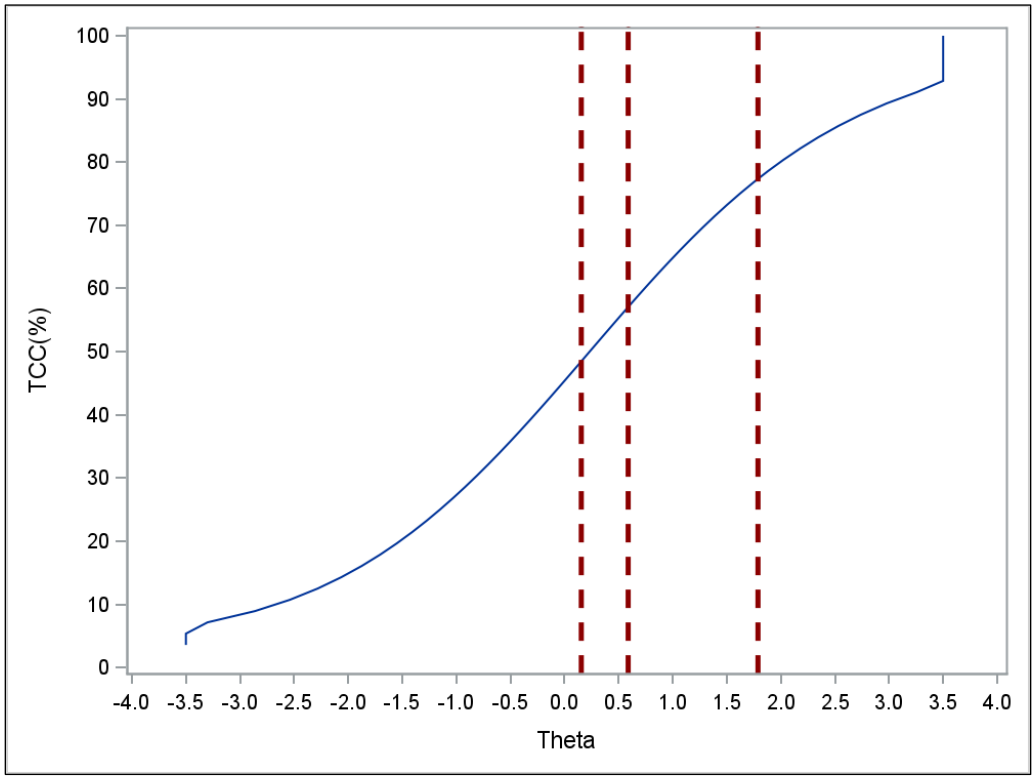


Figure B.22. CSEM, ELA Grade 4, Form 1

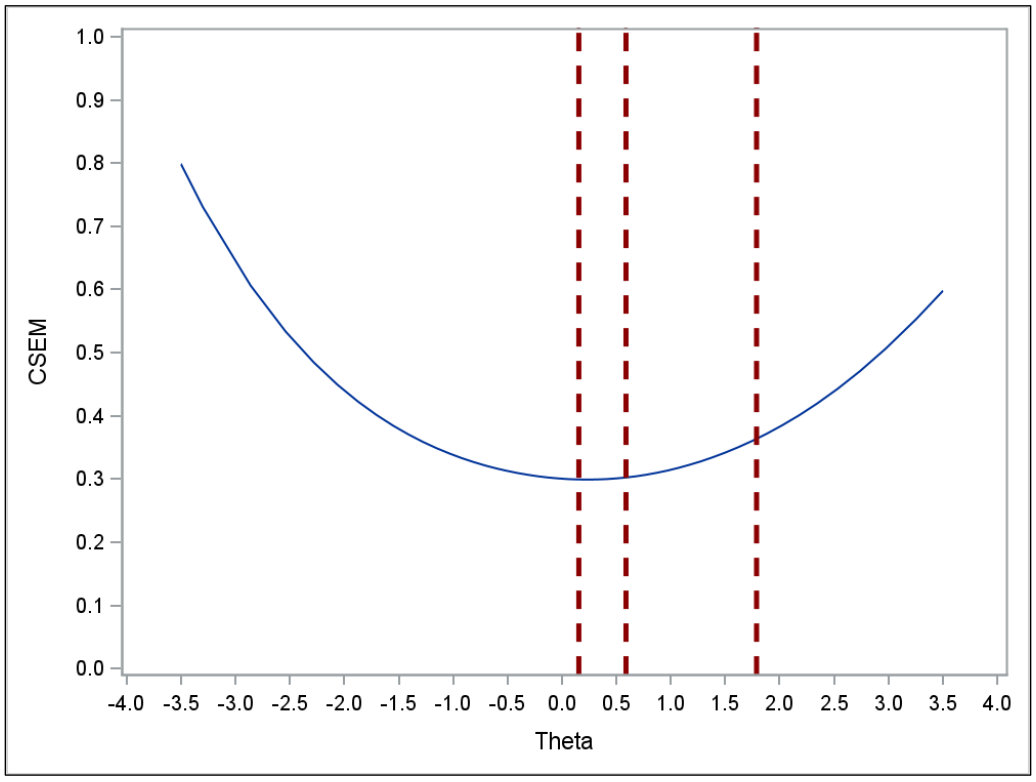


Figure B.23. TCC, ELA Grade 5, Form 1

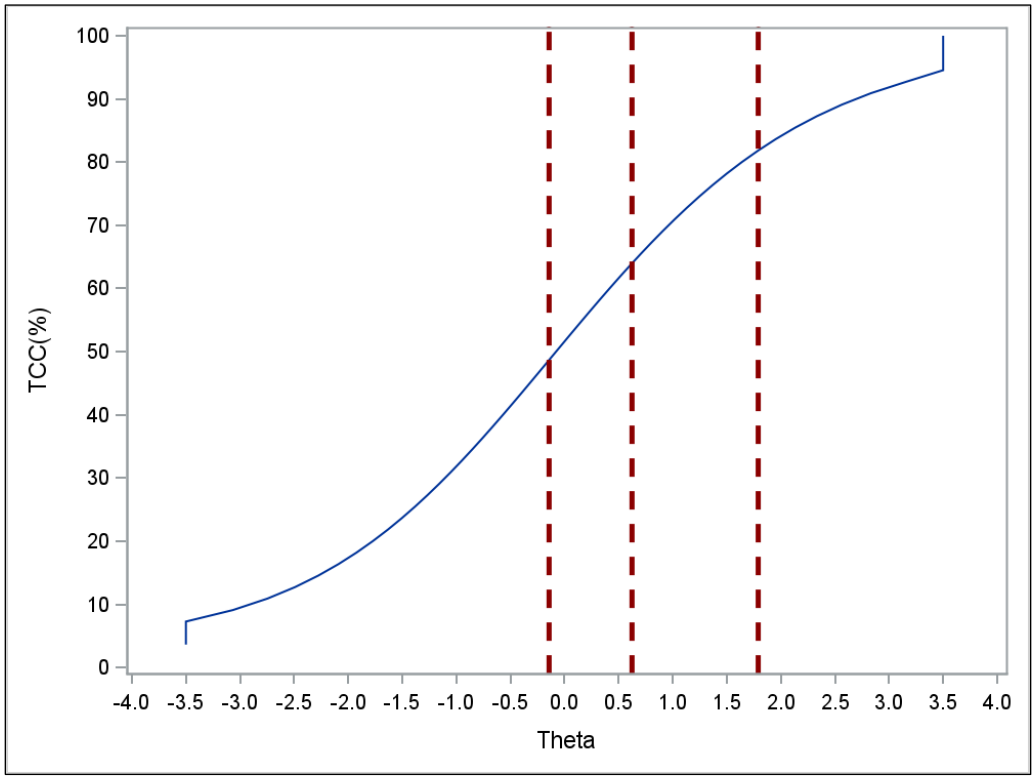


Figure B.24. CSEM, ELA Grade 5, Form 1

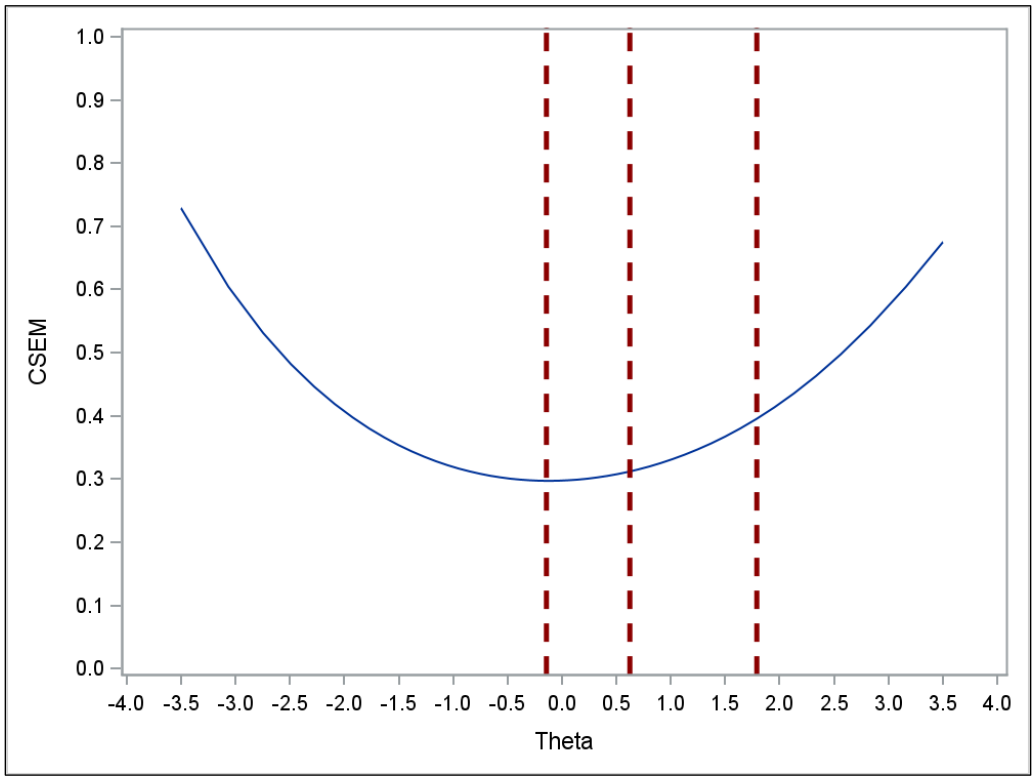


Figure B.25. TCC, ELA Grade 6, Form 1

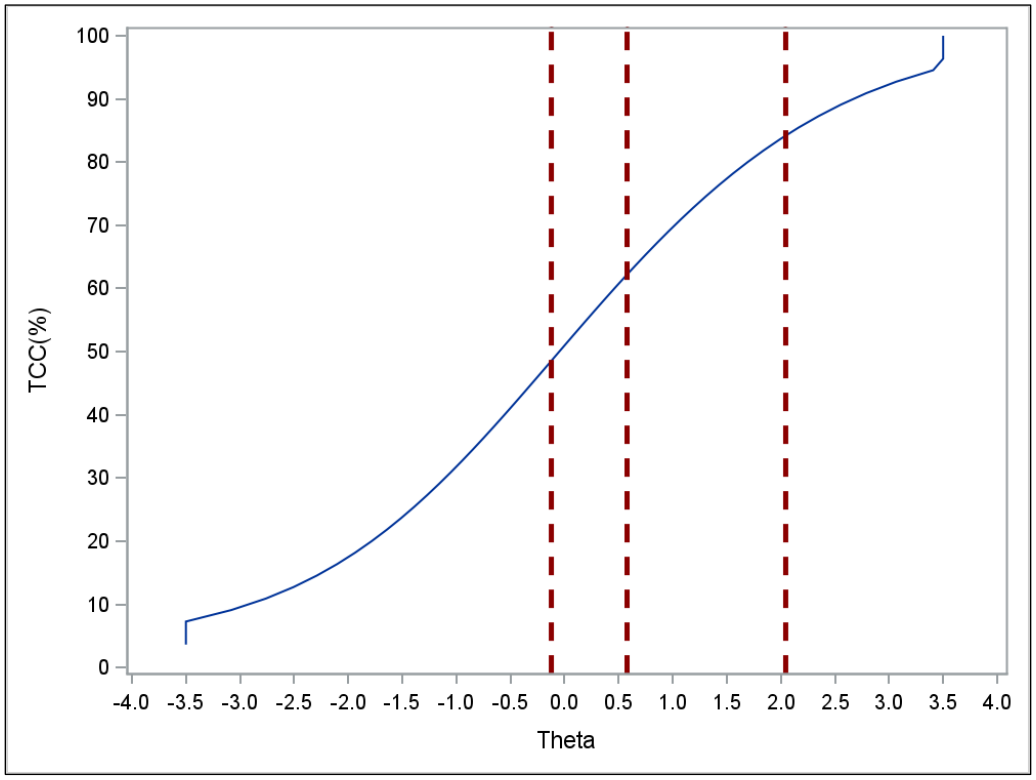


Figure B.26. CSEM, ELA Grade 6, Form 1

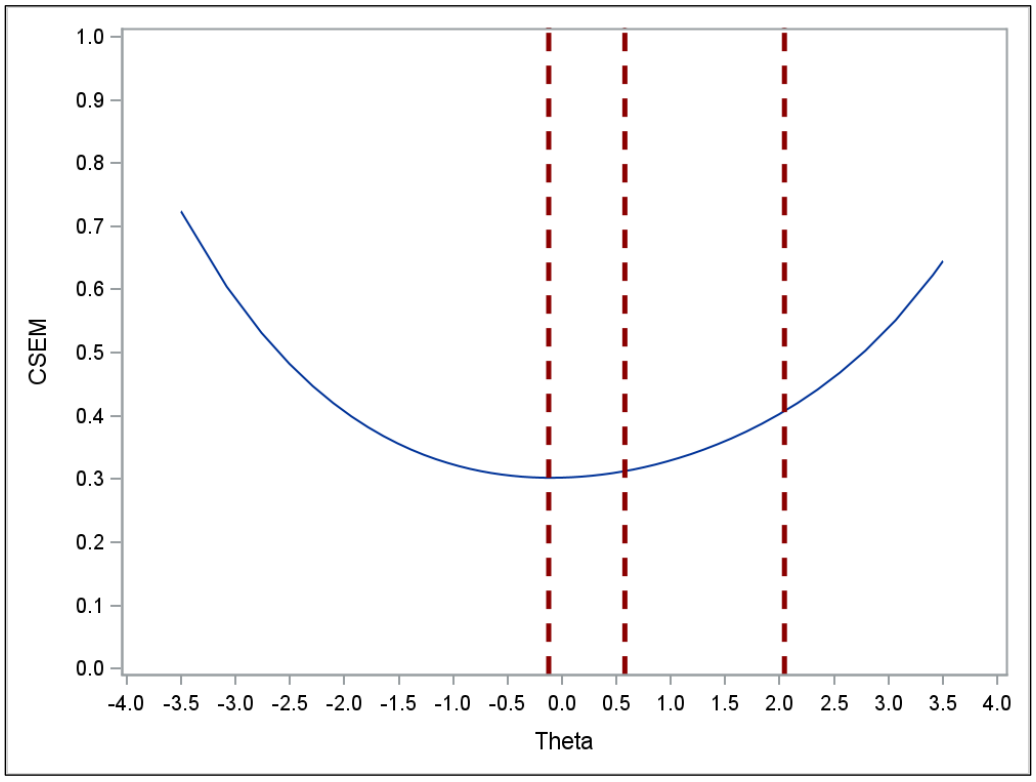


Figure B.27. TCC, ELA Grade 7, Form 1

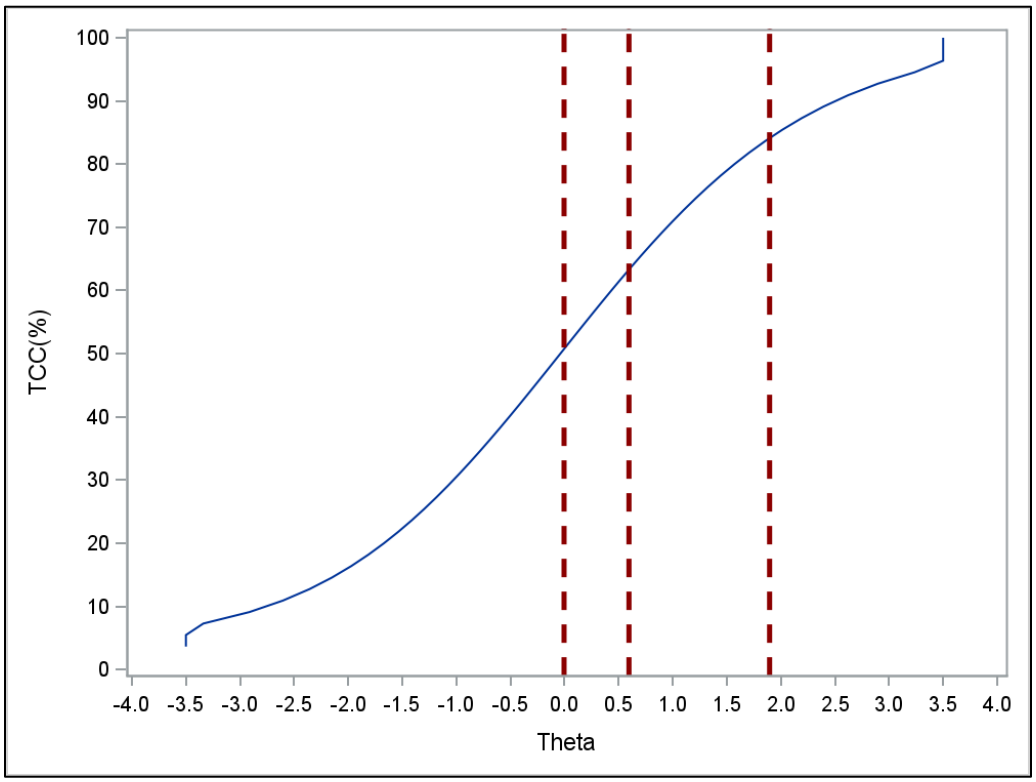


Figure B.28. CSEM, ELA Grade 7, Form 1

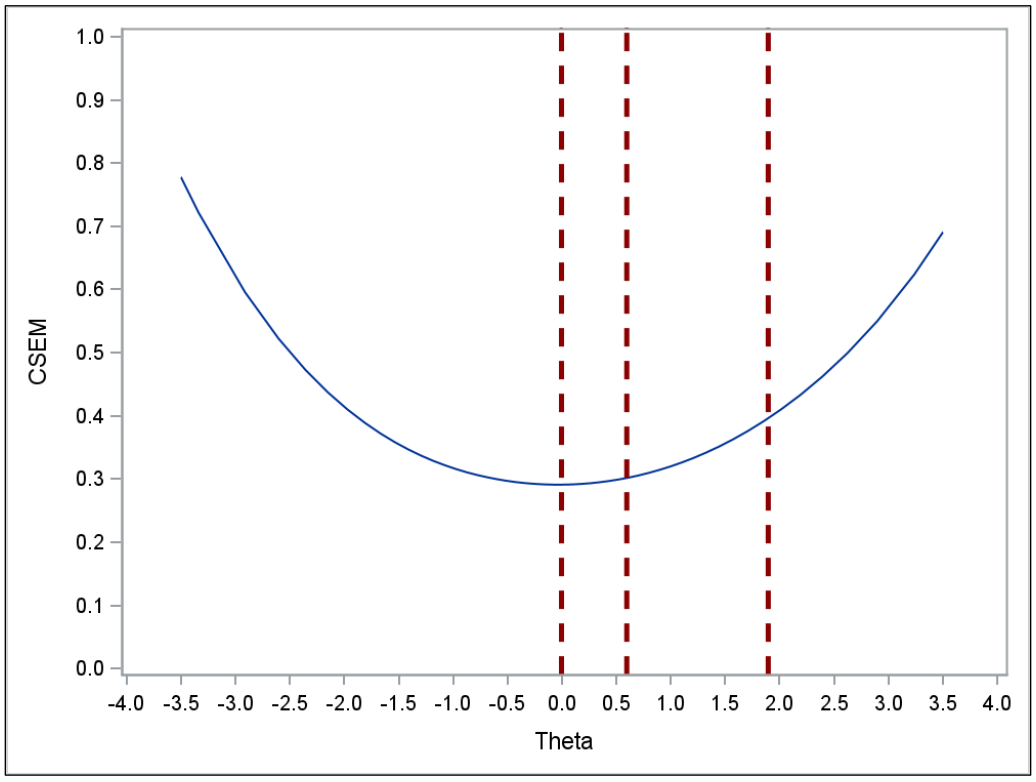


Figure B.29. TCC, ELA Grade 8, Form 1

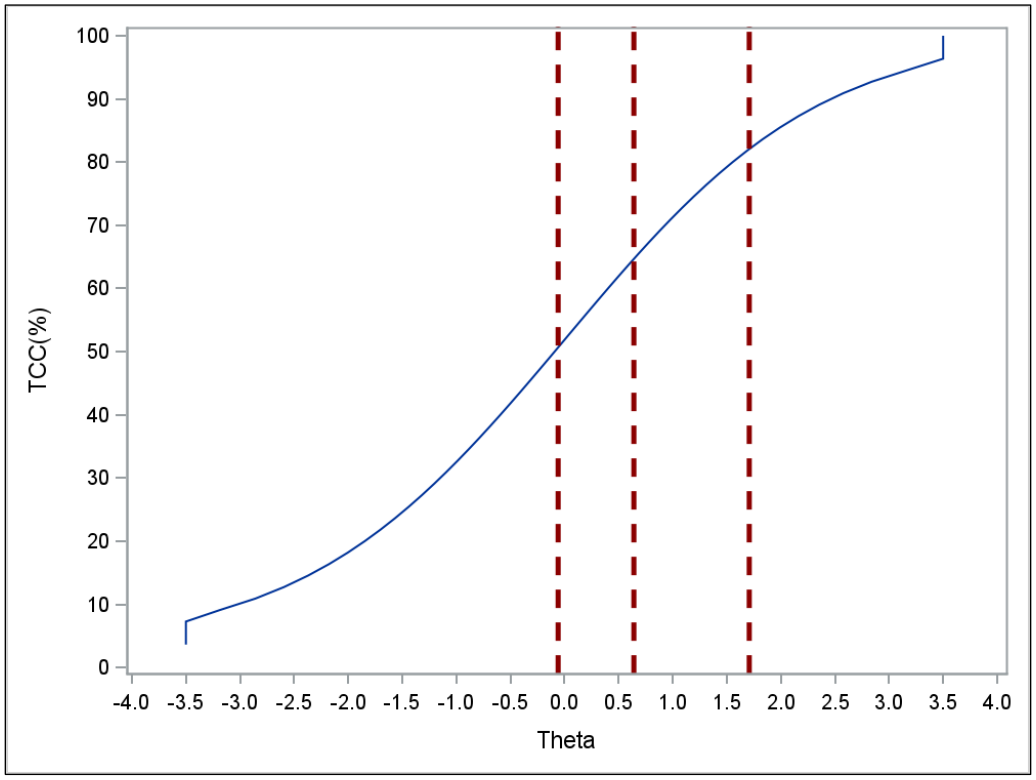


Figure B.30. CSEM, ELA Grade 8, Form 1

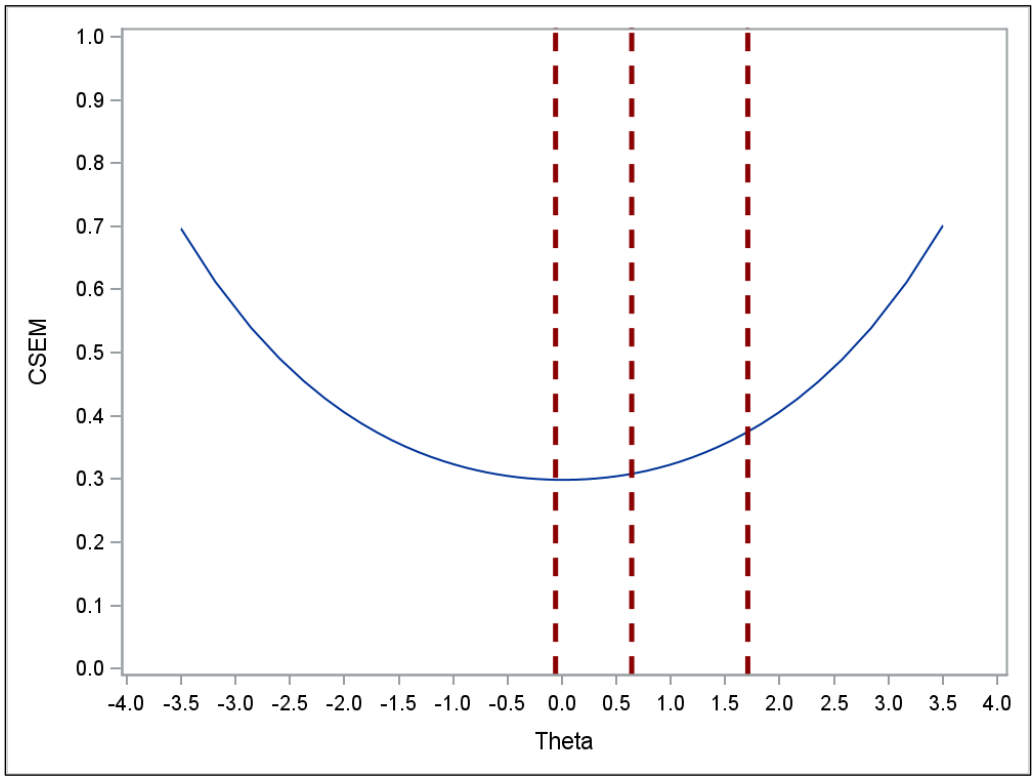


Figure B.31. TCC, ELA Grade 3, Form 2

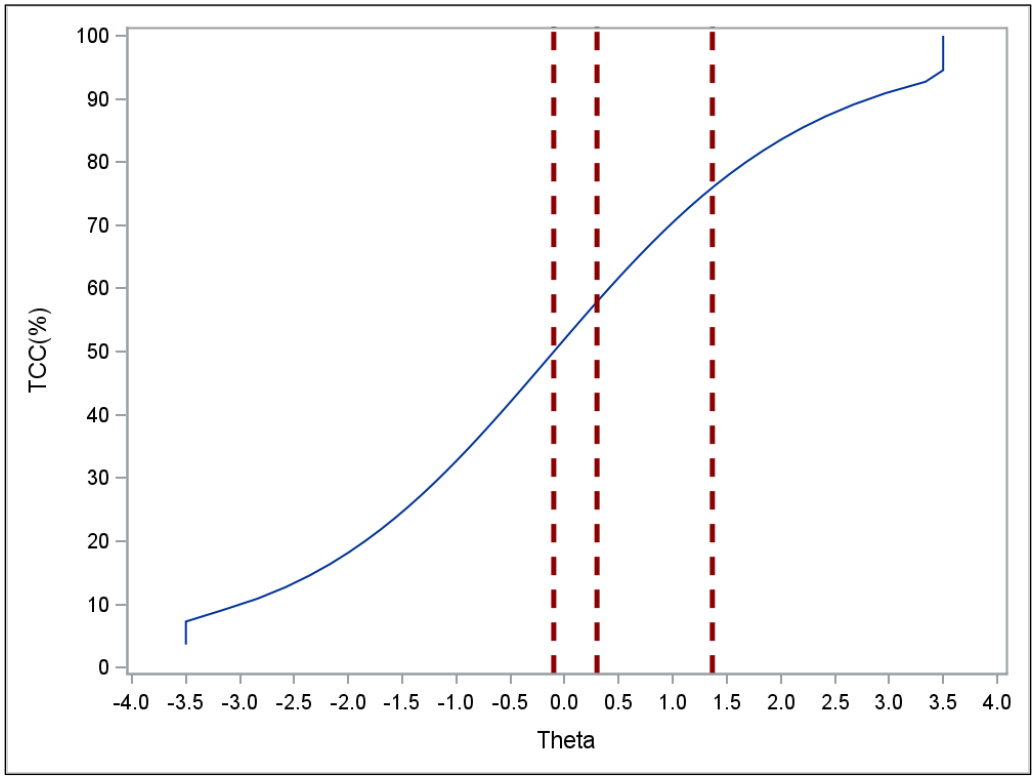


Figure B.32. CSEM, ELA Grade 3, Form 2

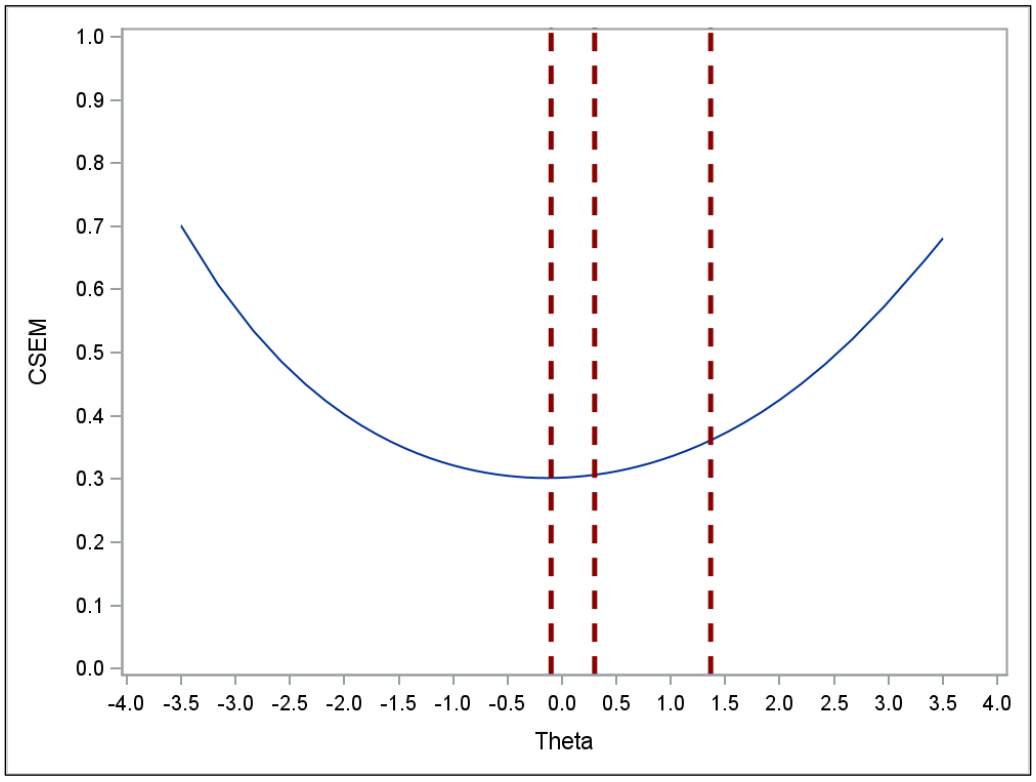


Figure B.33. TCC, ELA Grade 4, Form 2

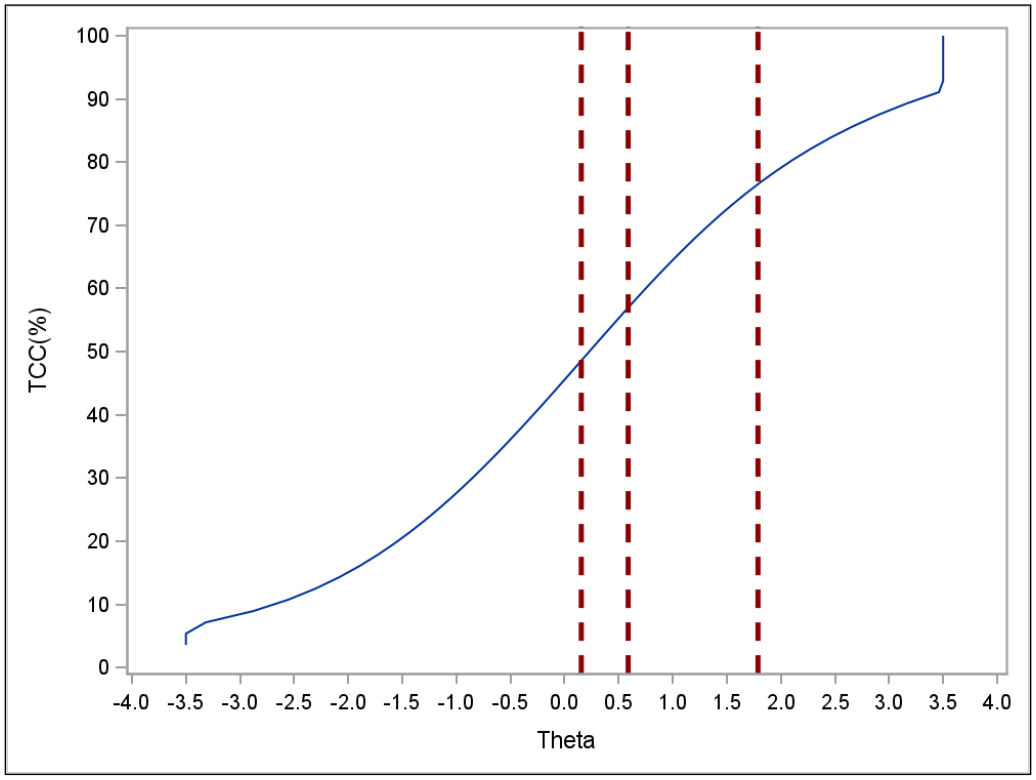


Figure B.34. CSEM, ELA Grade 4, Form 2

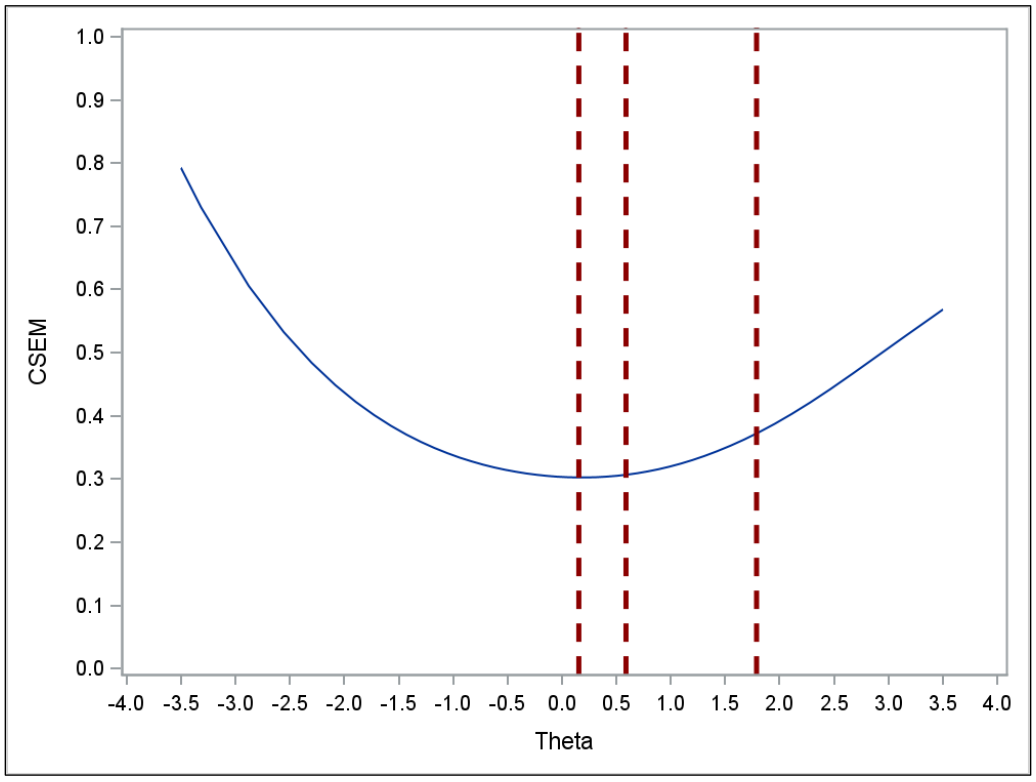


Figure B.35. TCC, ELA Grade 5, Form 2

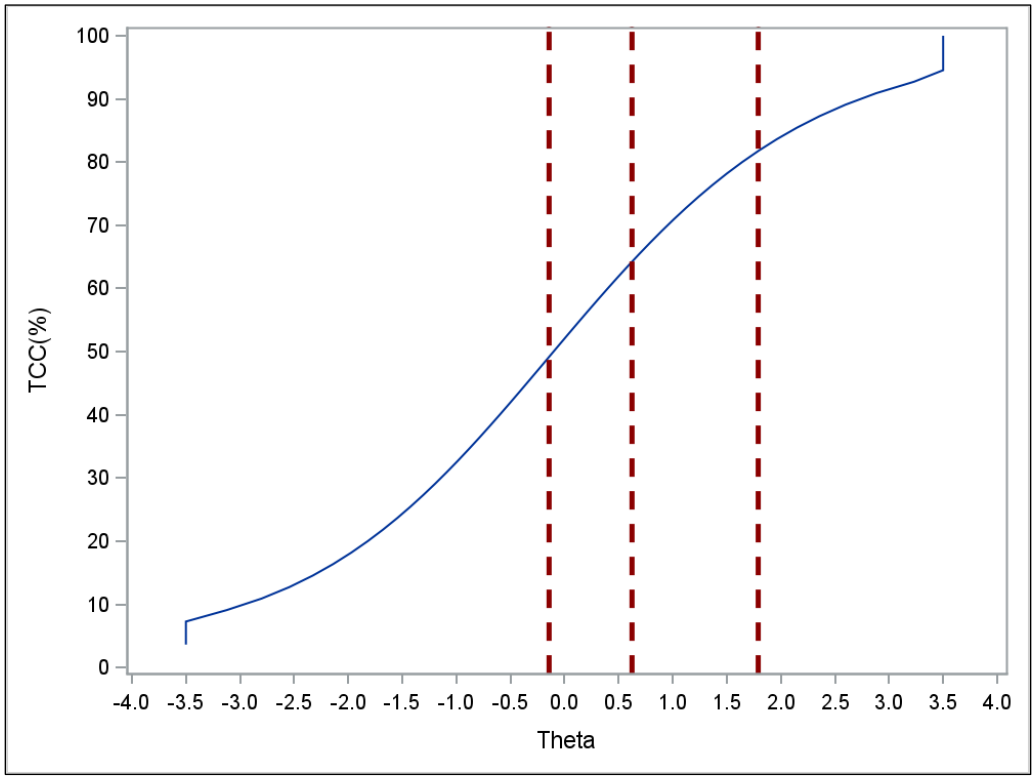


Figure B.36. CSEM, ELA Grade 5, Form 2

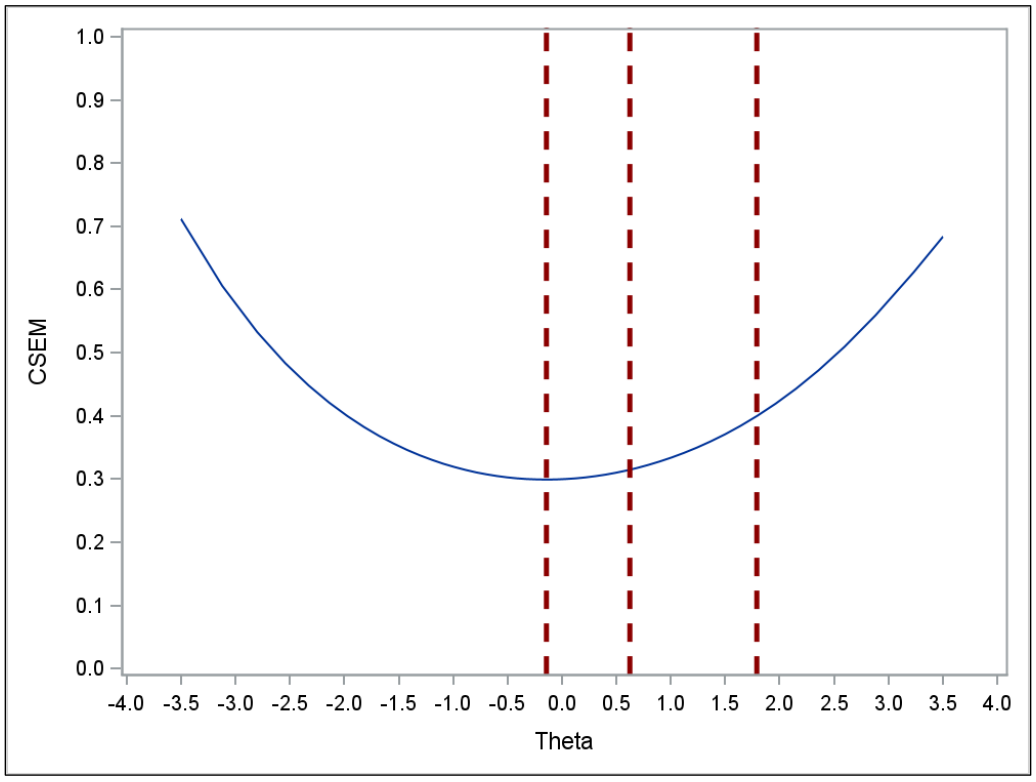


Figure B.37. TCC, ELA Grade 6, Form 2

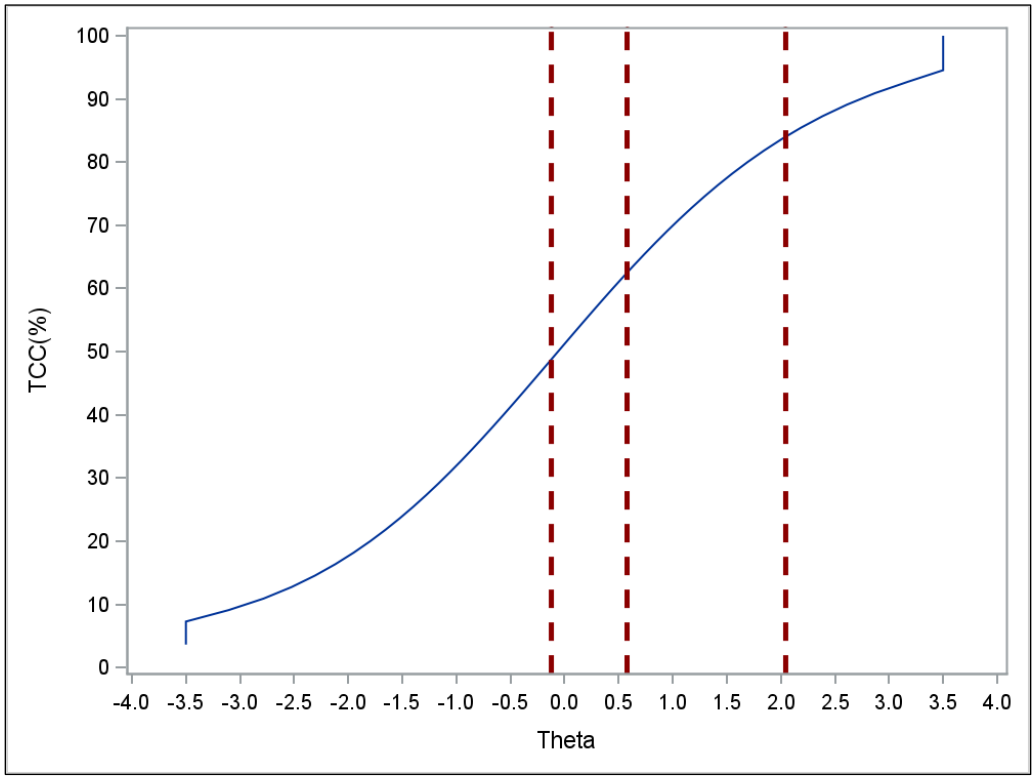


Figure B.38. CSEM, ELA Grade 6, Form 2

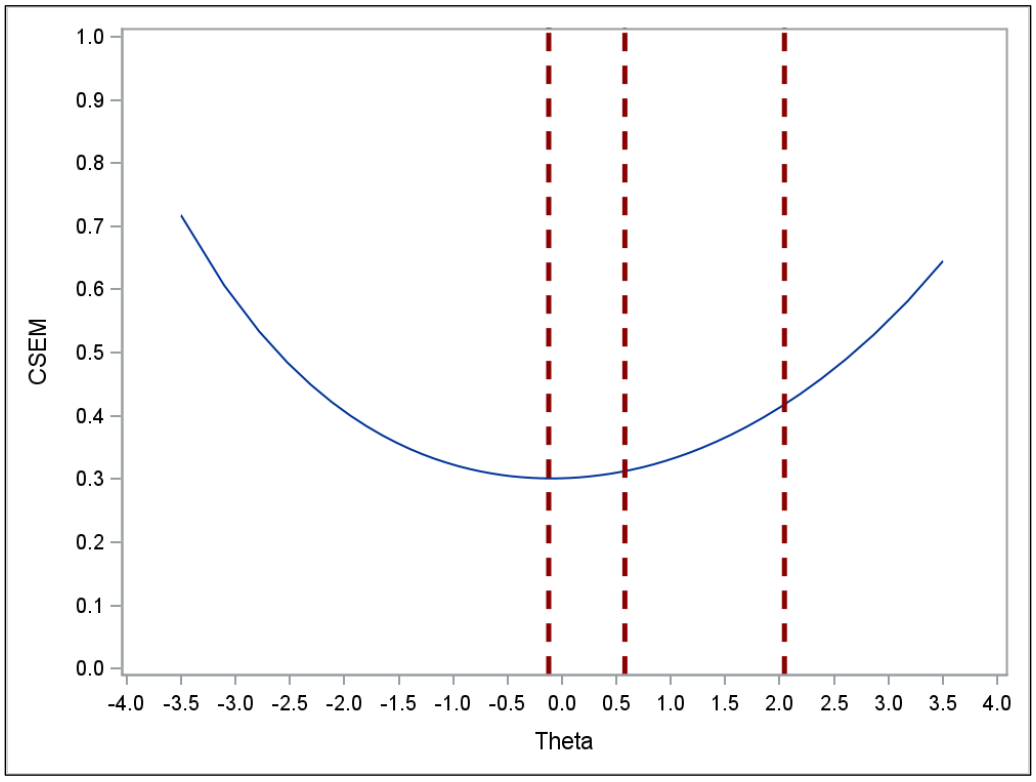


Figure B.39. TCC, ELA Grade 7, Form 2

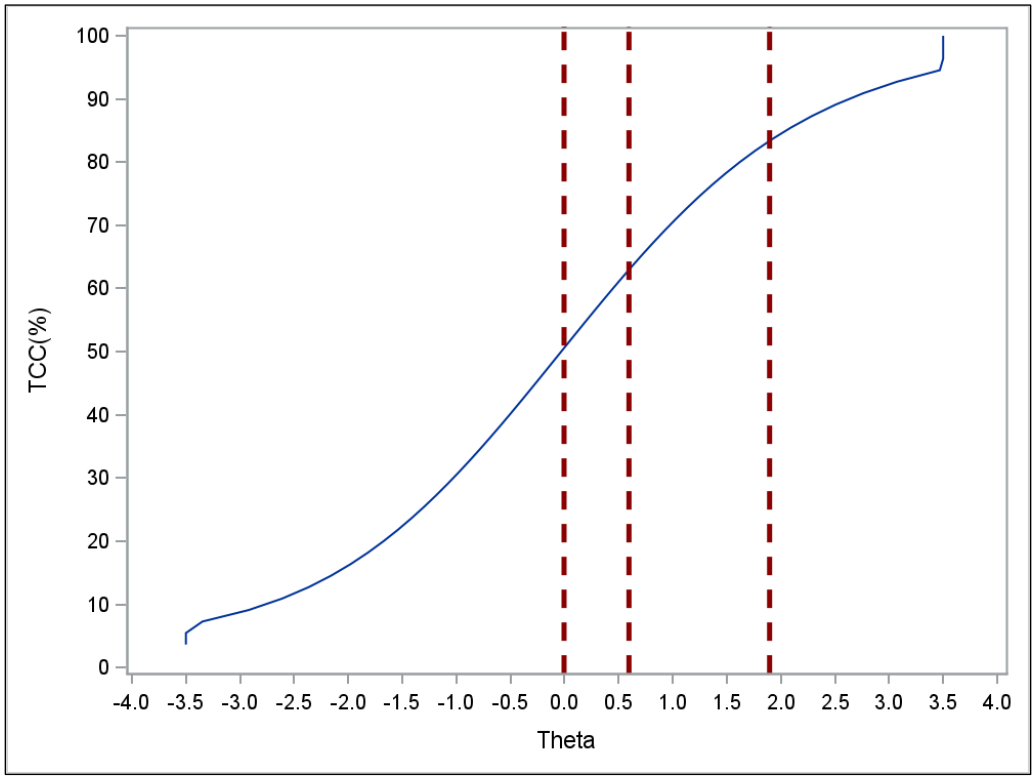


Figure B.40. CSEM, ELA Grade 7, Form 2

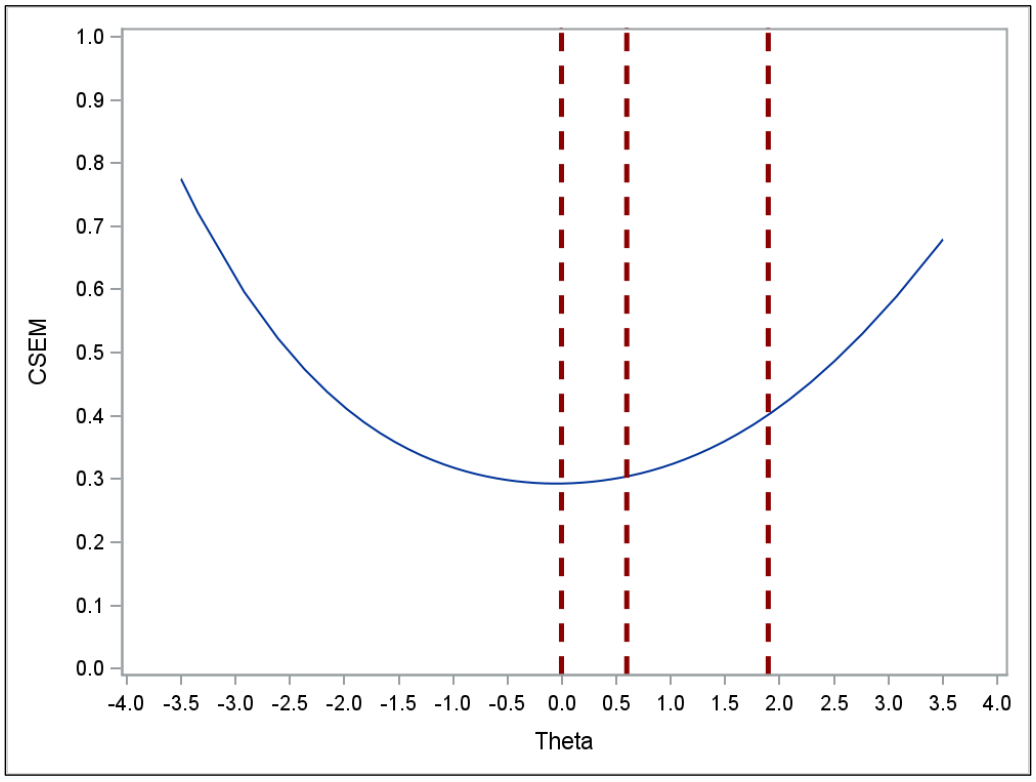


Figure B.41. TCC, ELA Grade 8, Form 2

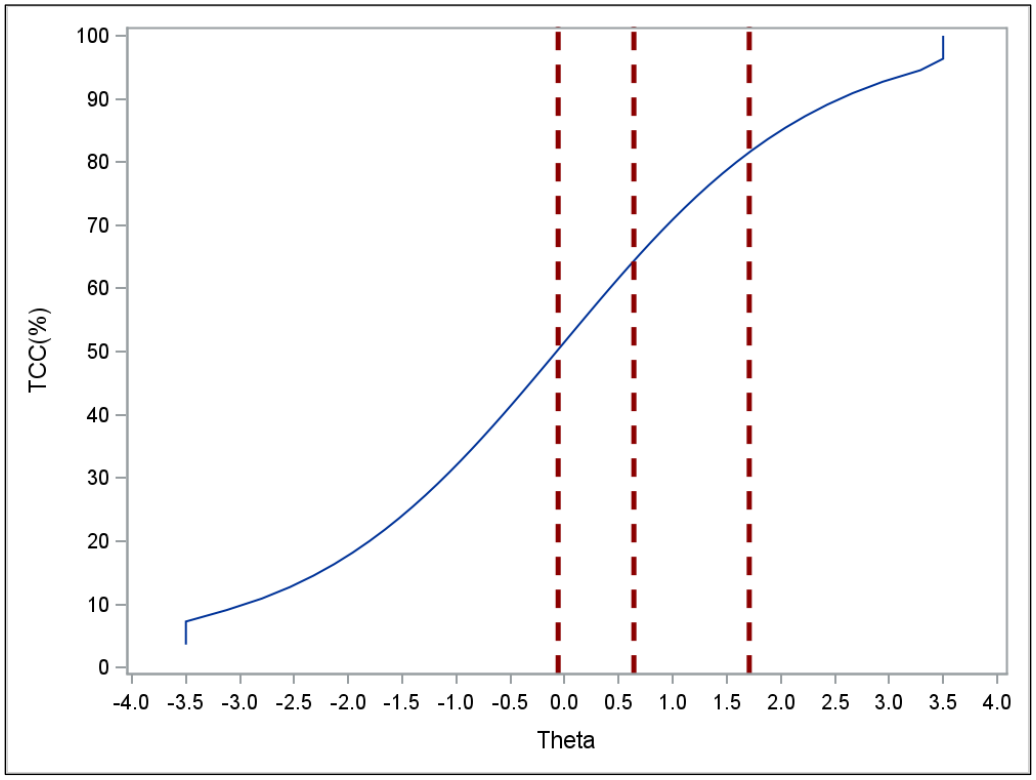


Figure B.42. CSEM, ELA Grade 8, Form 2

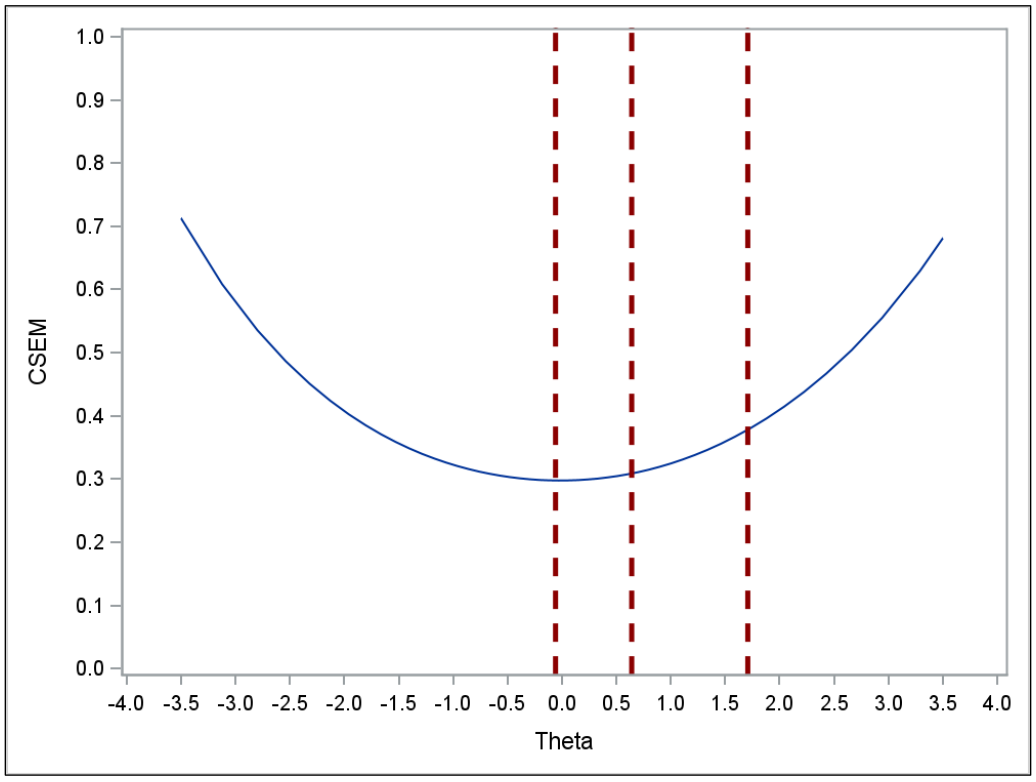


Figure B.43. TCC, Mathematics Grade 3

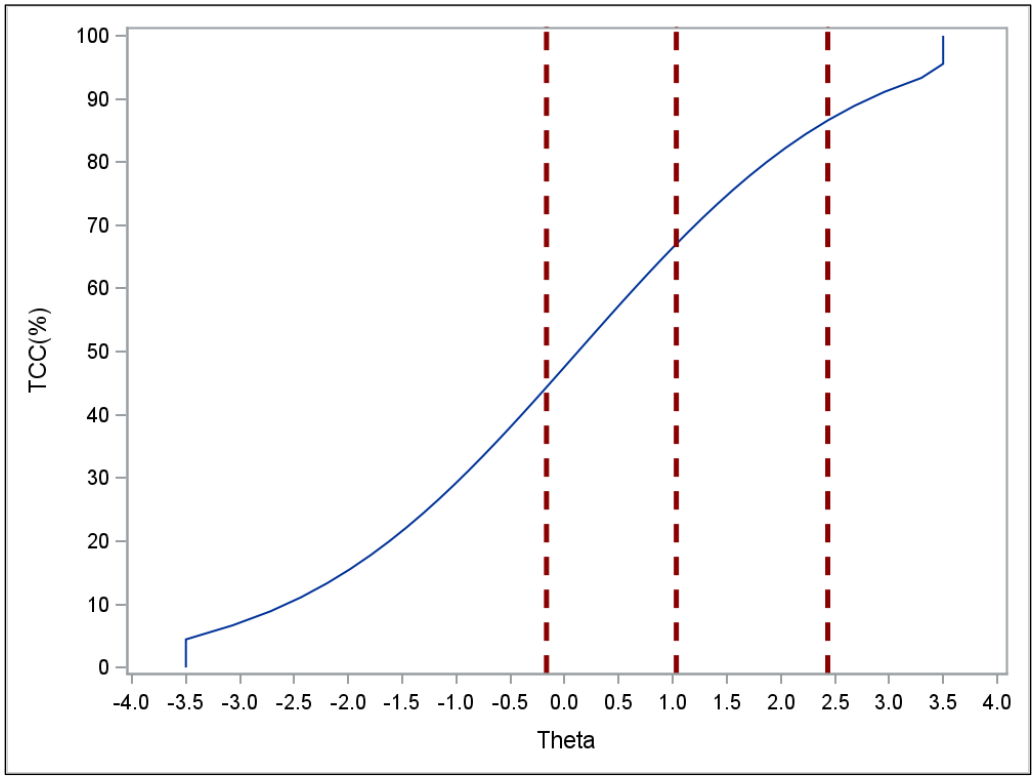


Figure B.44. CSEM, Mathematics Grade 3

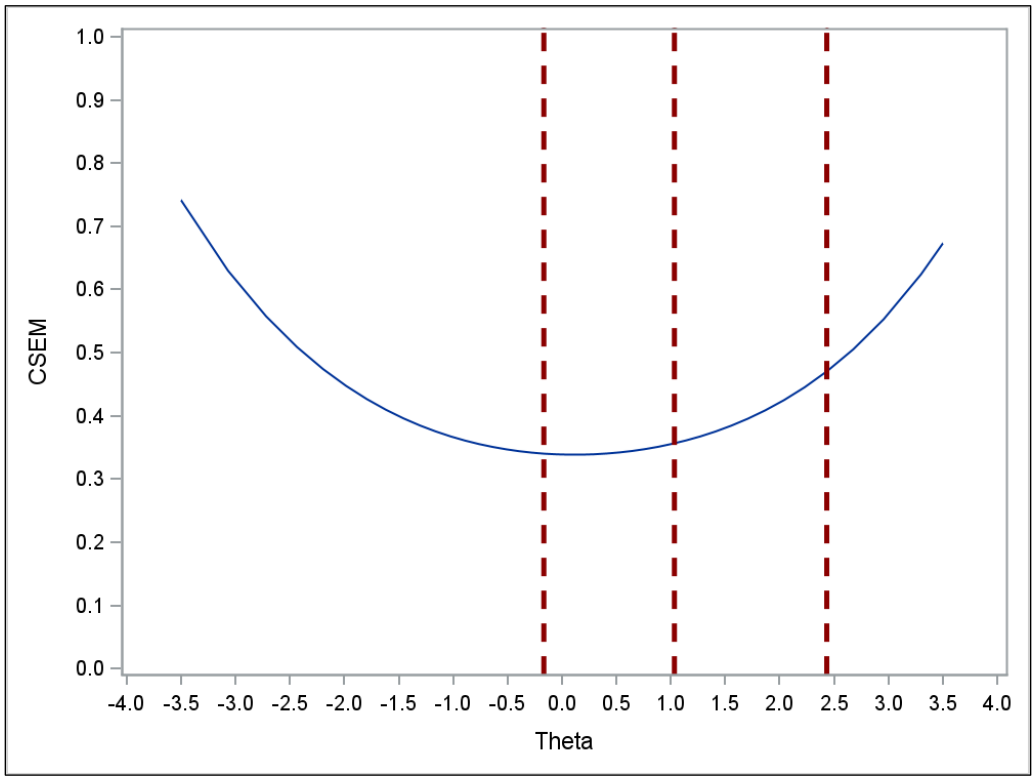


Figure B.45. TCC, Mathematics Grade 4

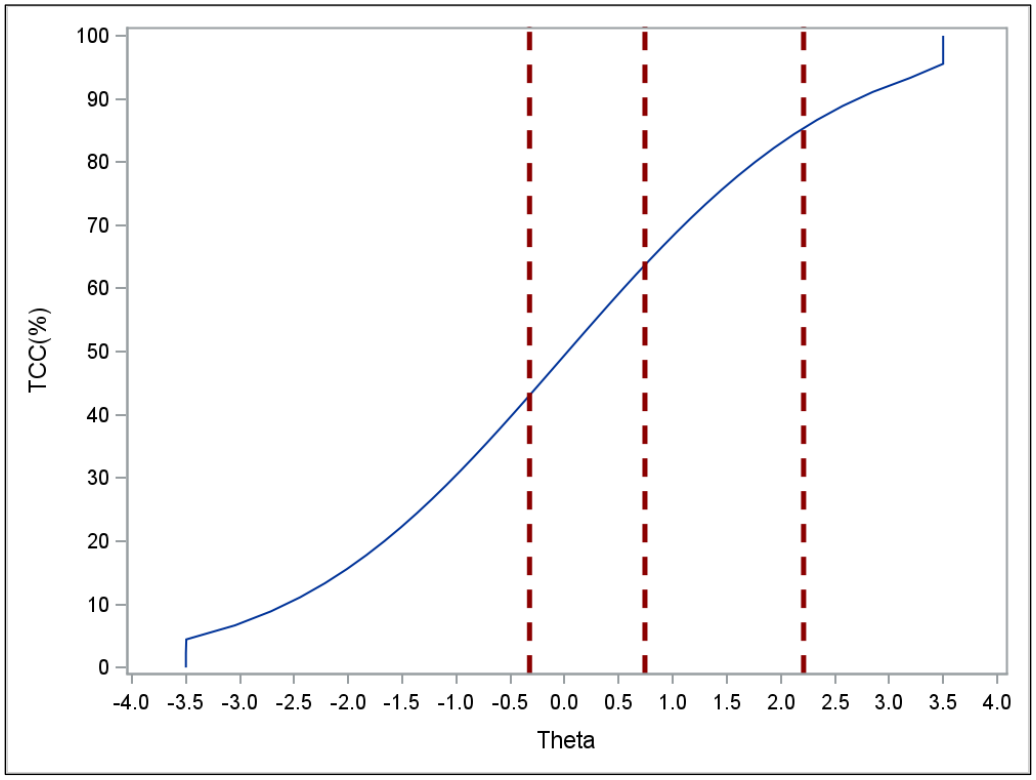


Figure B.46. CSEM, Mathematics Grade 4

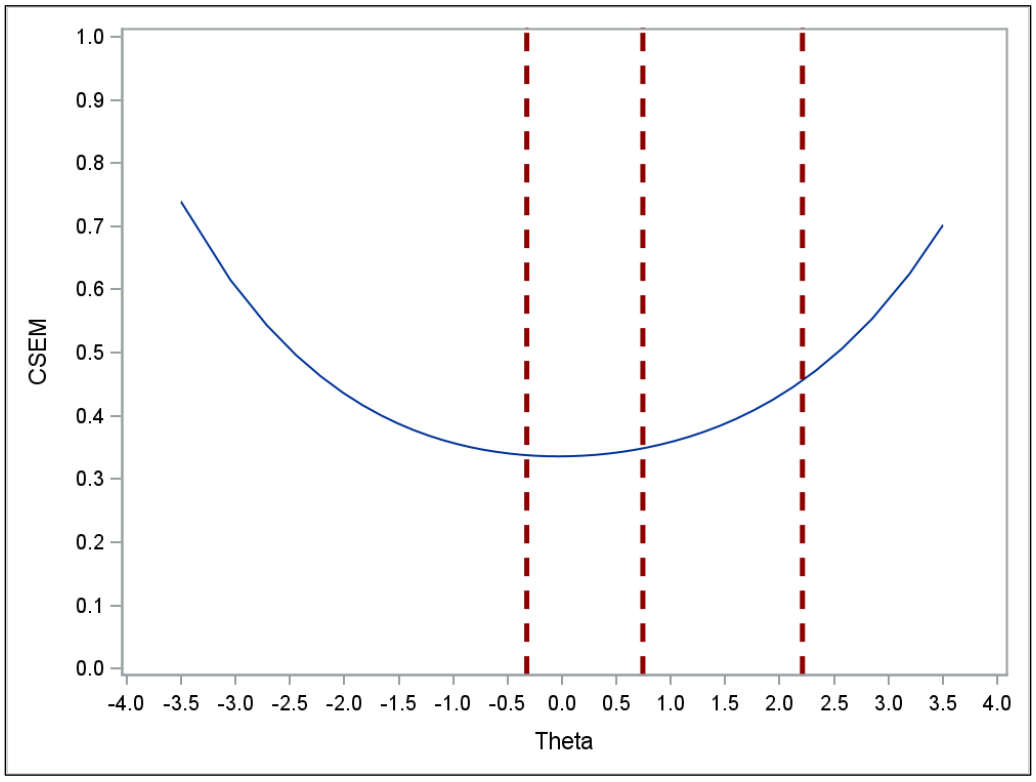


Figure B.47. TCC, Mathematics Grade 5

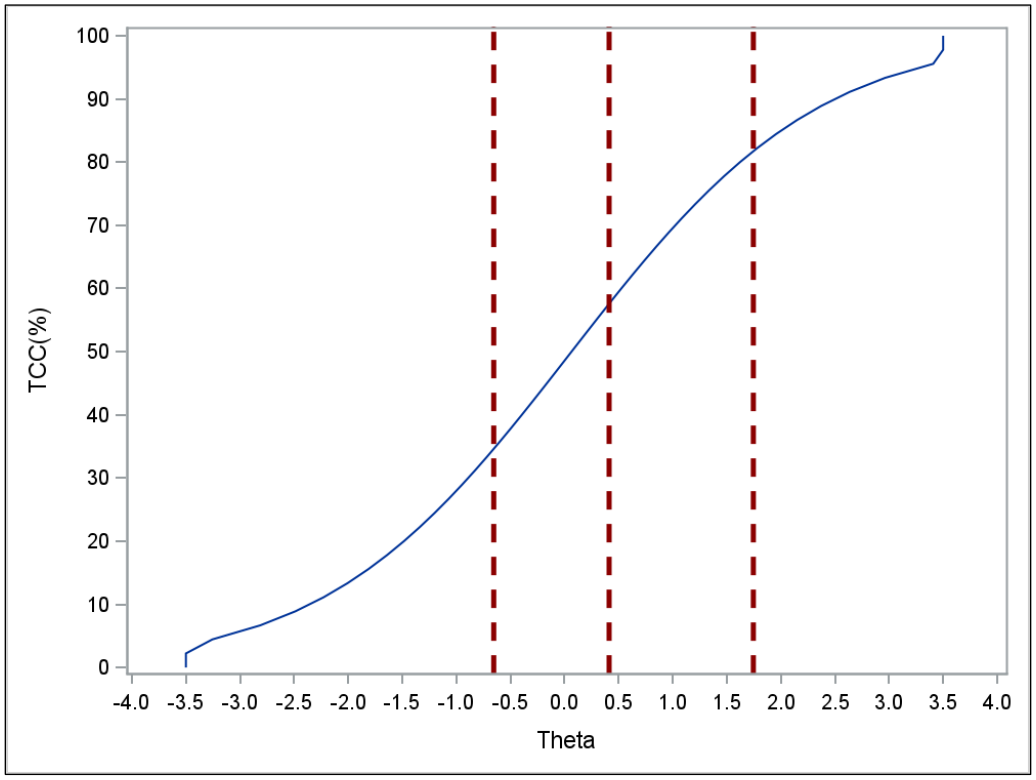


Figure B.48. CSEM, Mathematics Grade 5

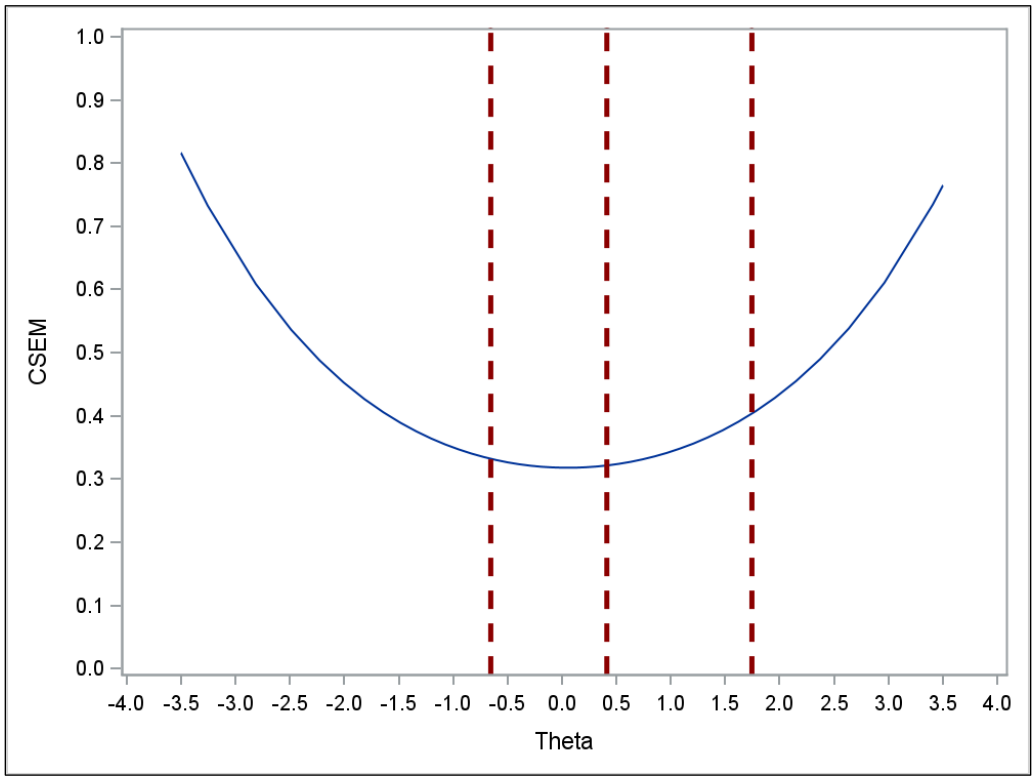


Figure B.49. TCC, Mathematics Grade 6

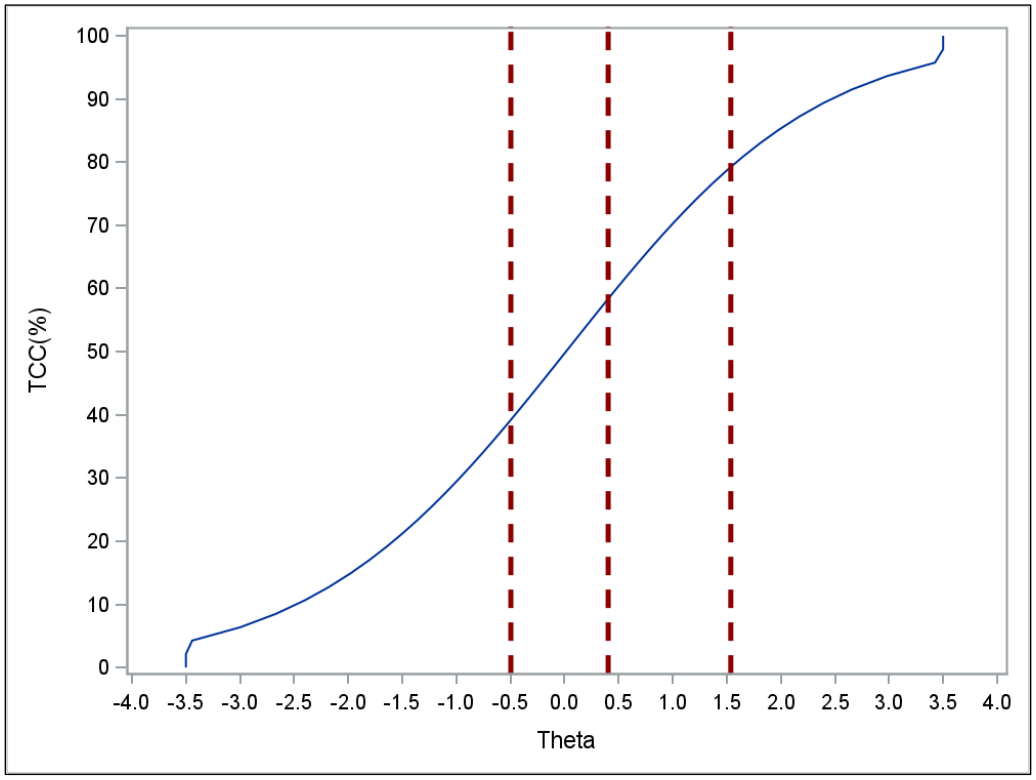


Figure B.50. CSEM, Mathematics Grade 6

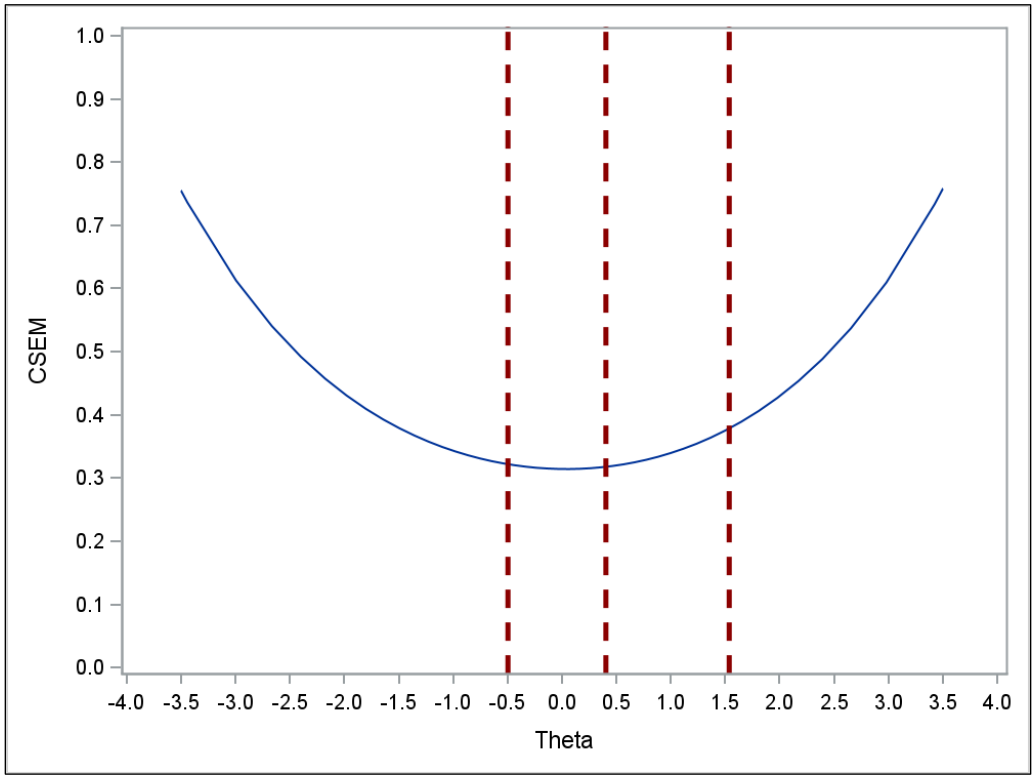


Figure B.51. TCC, Mathematics Grade 7

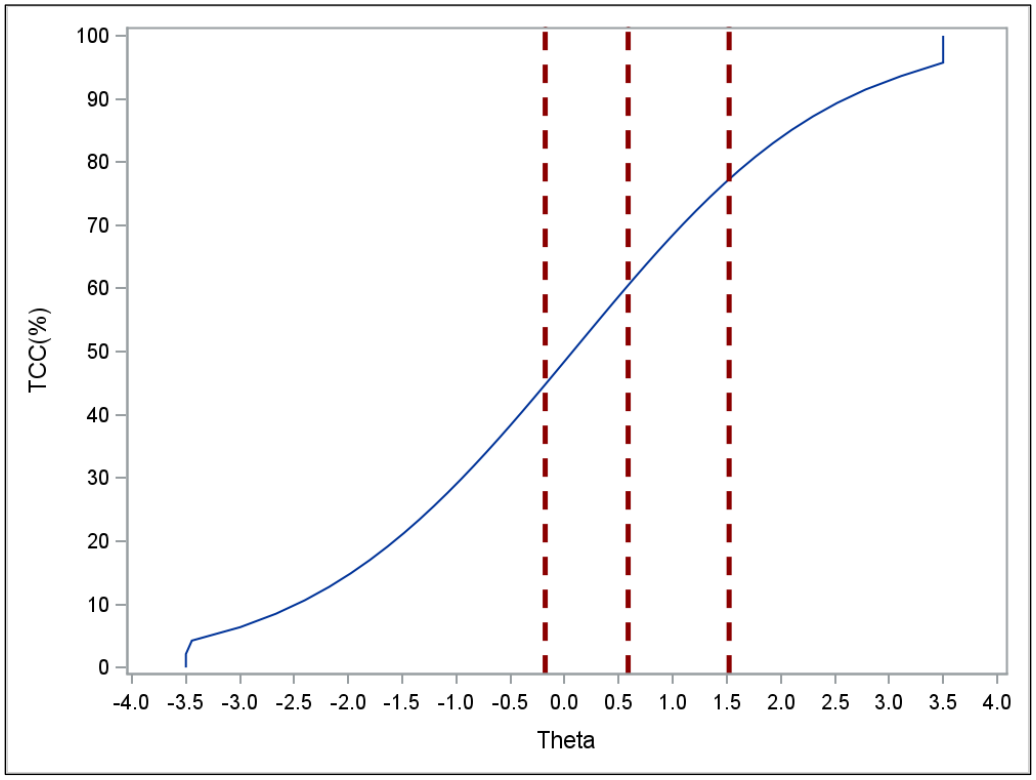


Figure B.52. CSEM, Mathematics Grade 7

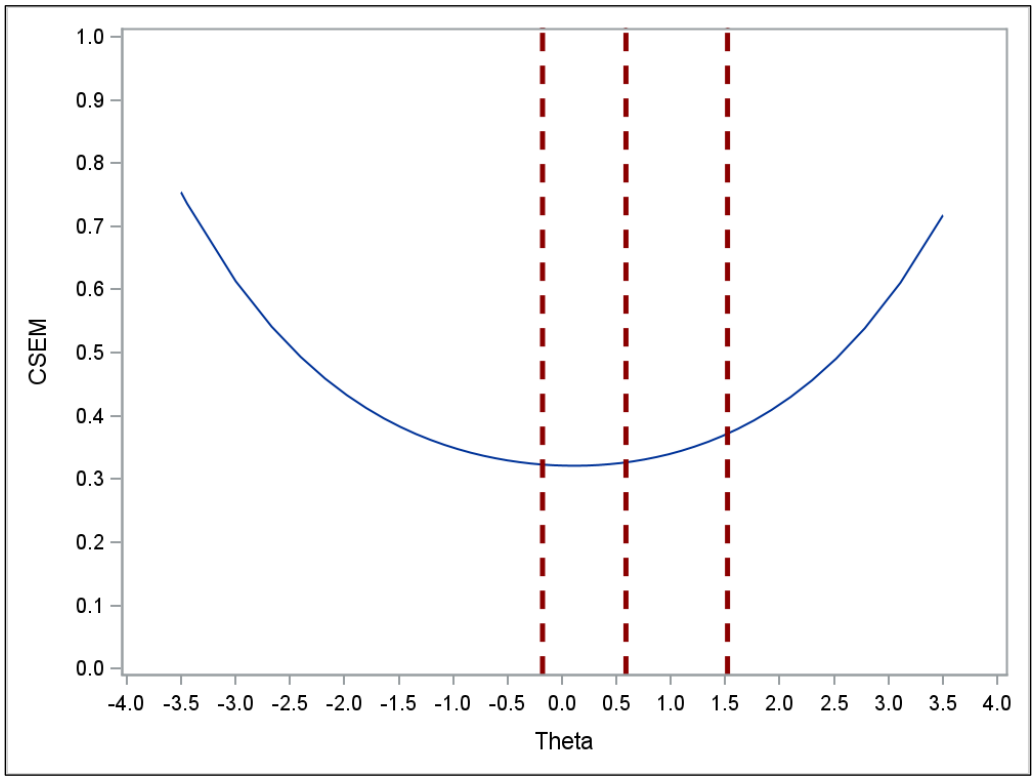


Figure B.53. TCC, Mathematics Grade 8

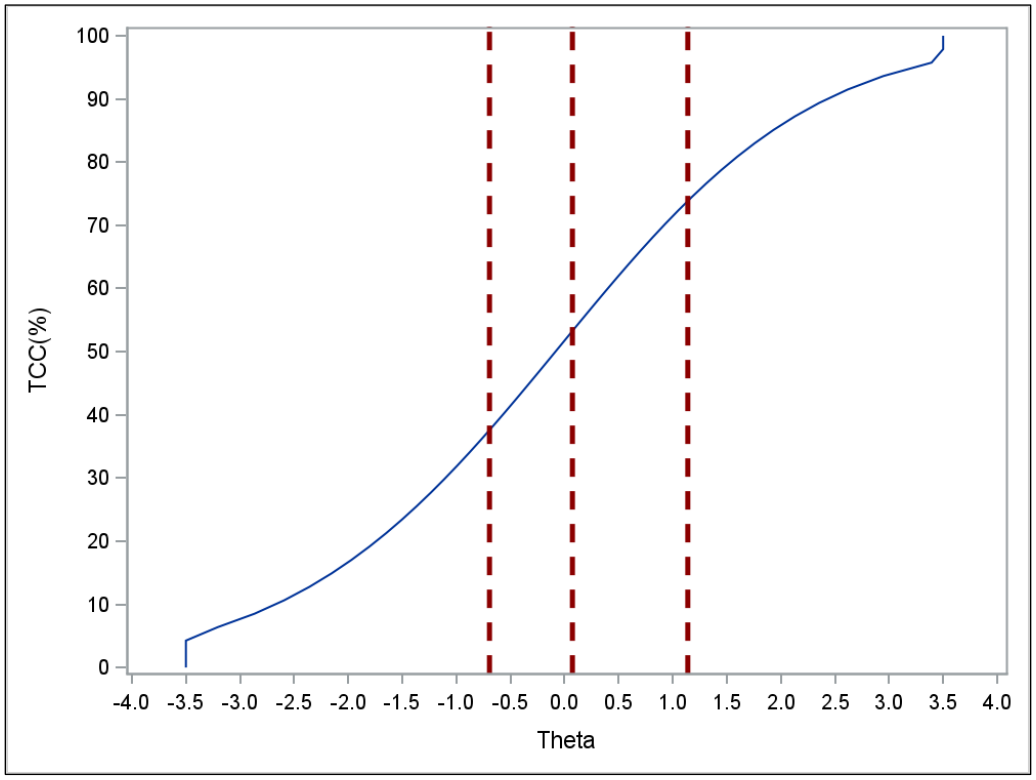


Figure B.54. CSEM, Mathematics Grade 8

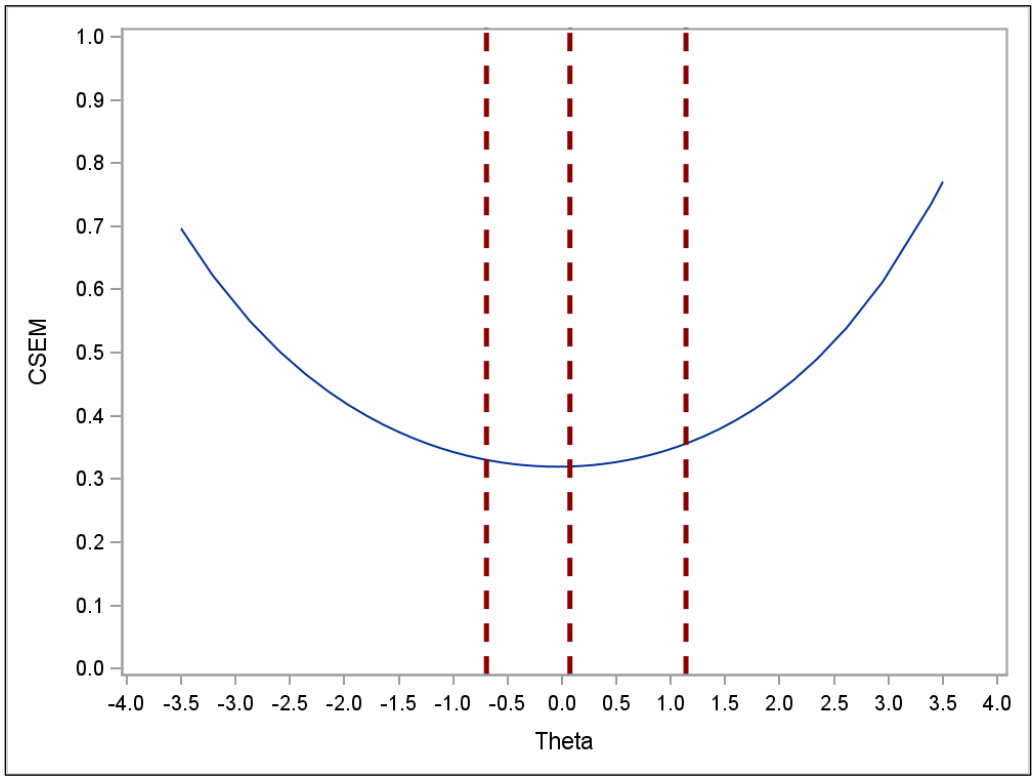


Figure B.55. Scree Plot, ELA Grade 3, Form 1

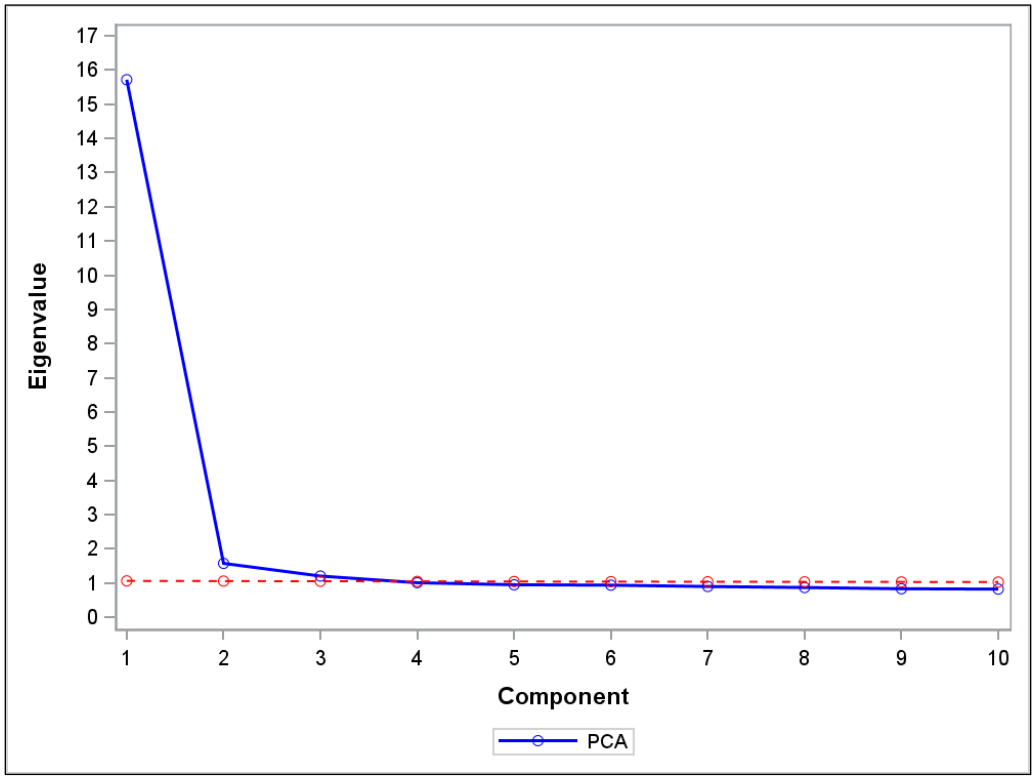


Figure B.56. Scree Plot, ELA Grade 4, Form 1

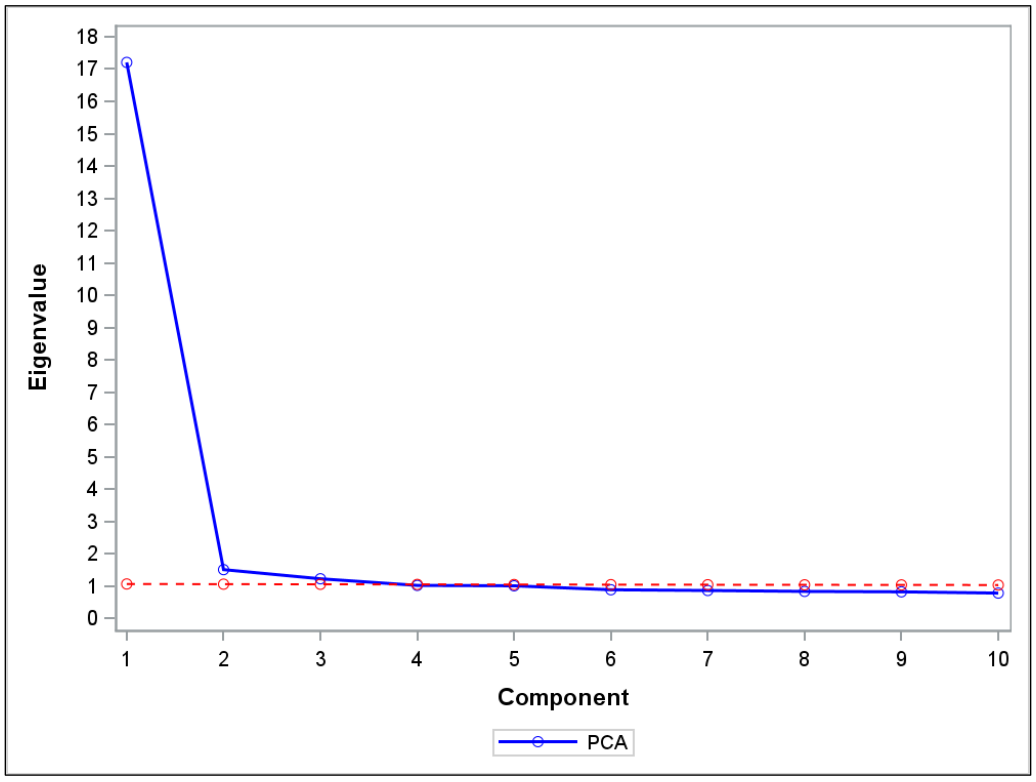


Figure B.57. Scree Plot, ELA Grade 5, Form 1

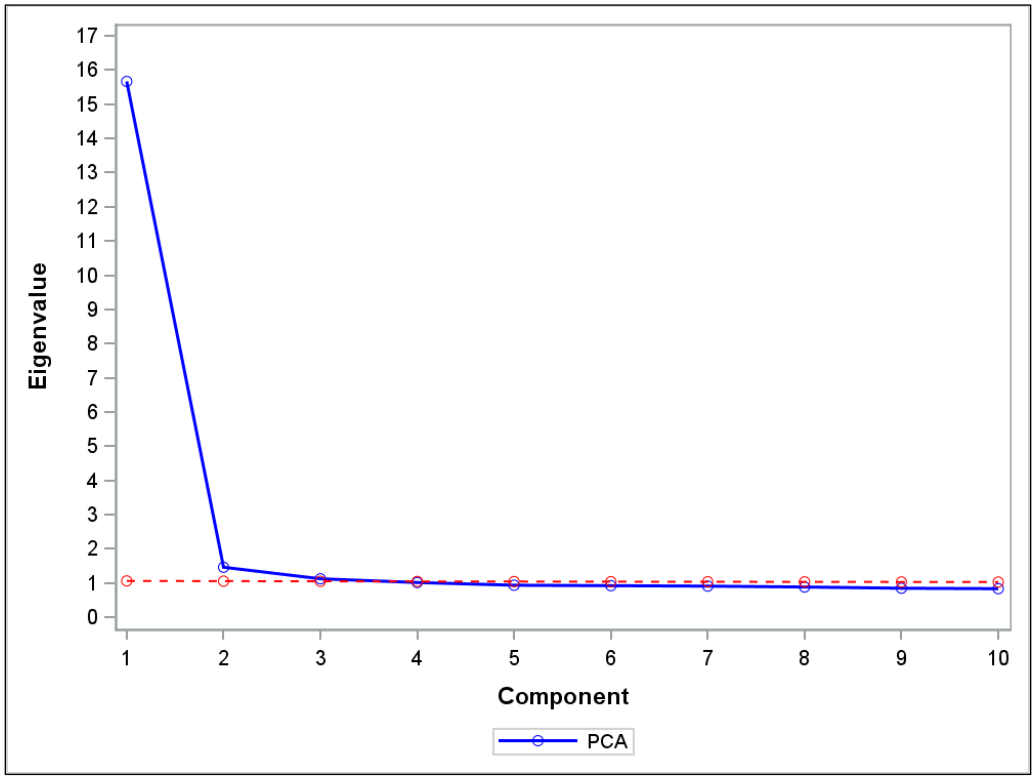


Figure B.58. Scree Plot, ELA Grade 6, Form 1

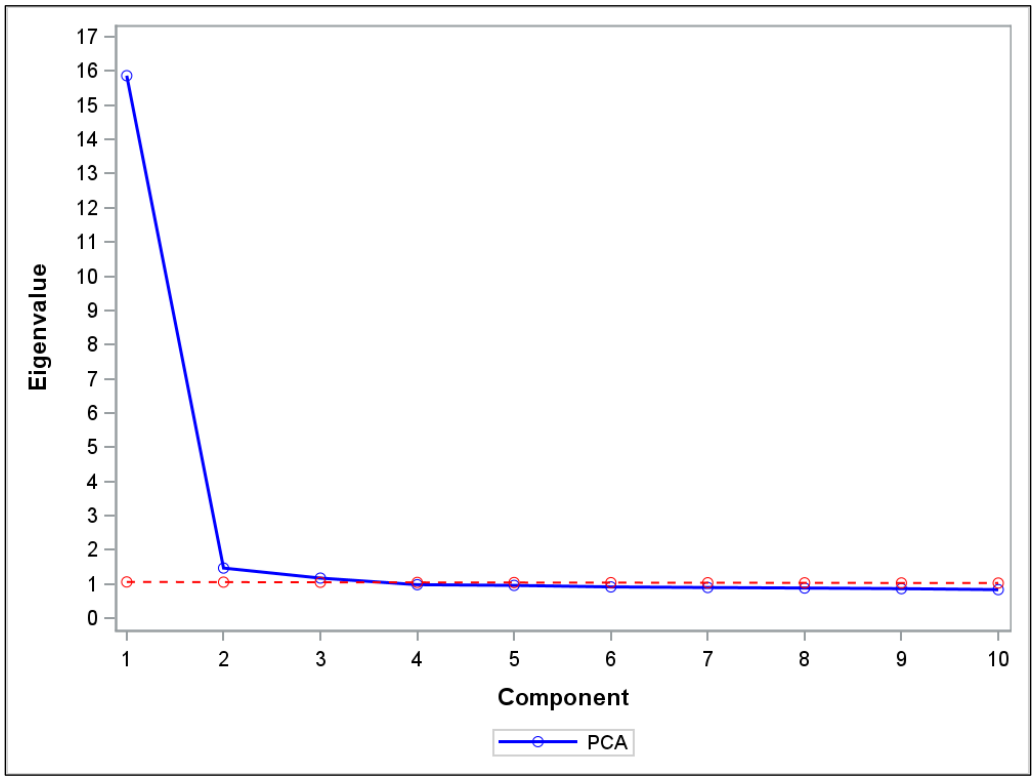


Figure B.59. Scree Plot, ELA Grade 7, Form 1

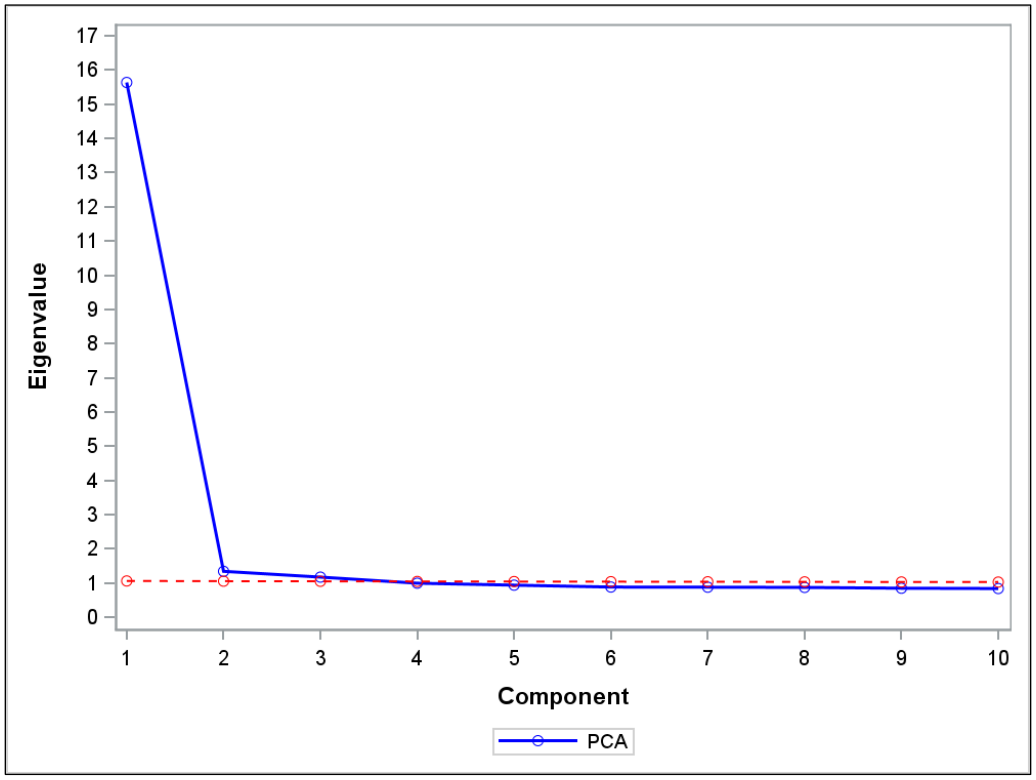


Figure B.60. Scree Plot, ELA Grade 8, Form 1

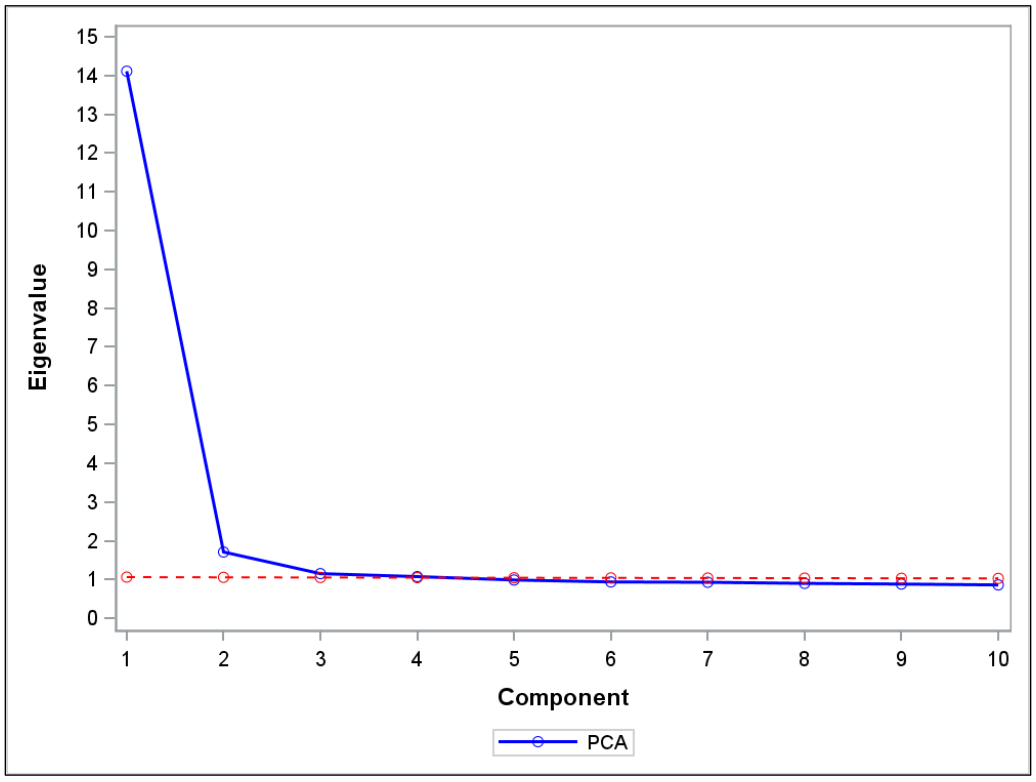


Figure B.61. Scree Plot, ELA Grade 3, Form 2

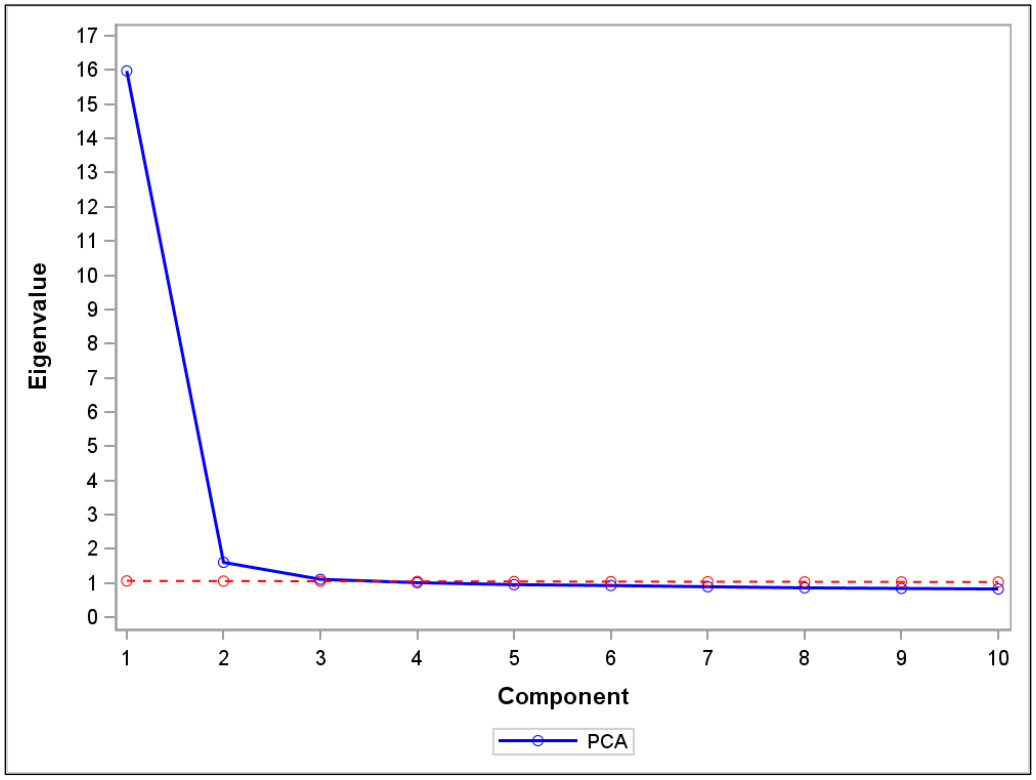


Figure B.62. Scree Plot, ELA Grade 4, Form 2

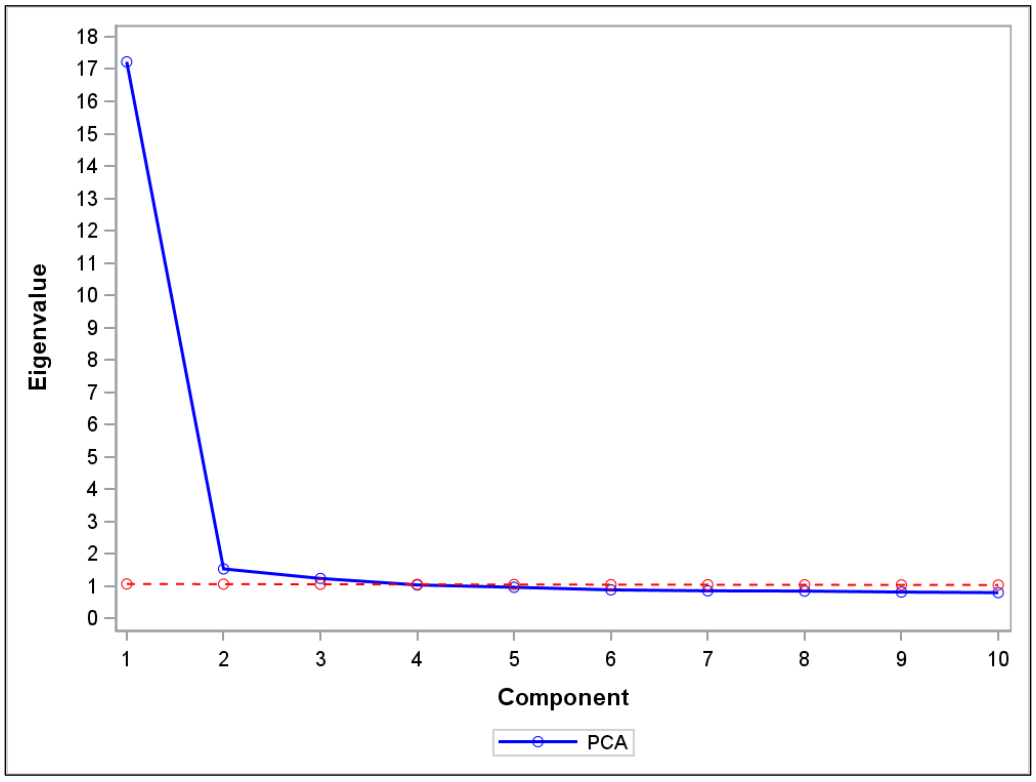


Figure B.63. Scree Plot, ELA Grade 5, Form 2

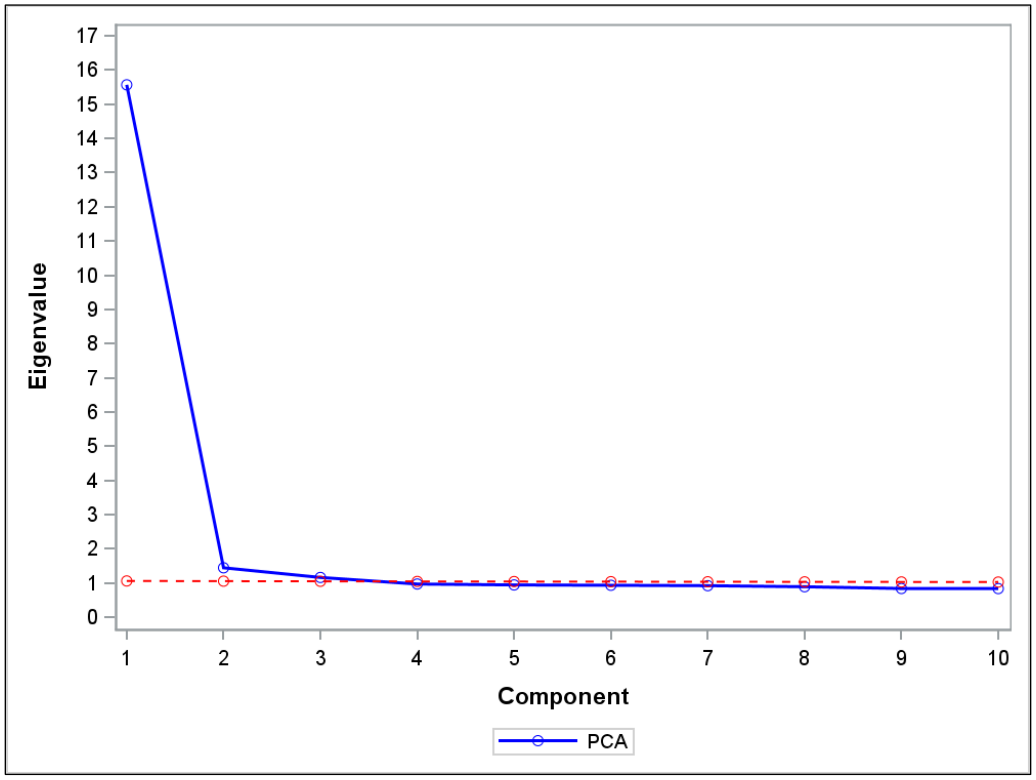


Figure B.64. Scree Plot, ELA Grade 6, Form 2

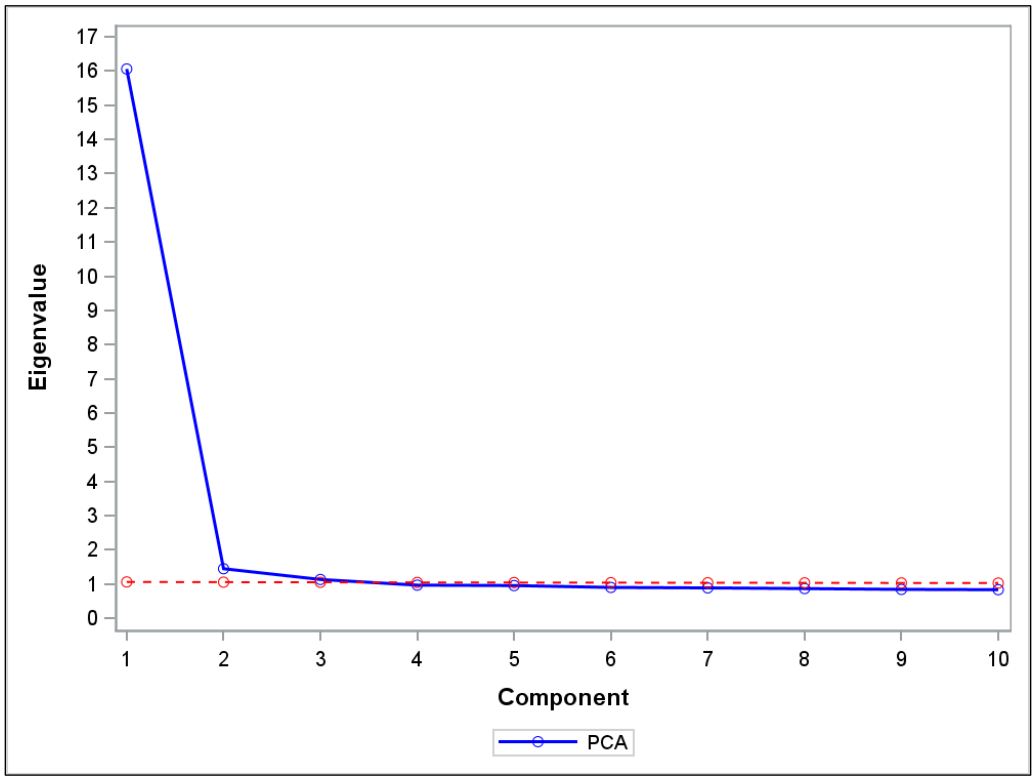


Figure B.65. Scree Plot, ELA Grade 7, Form 2

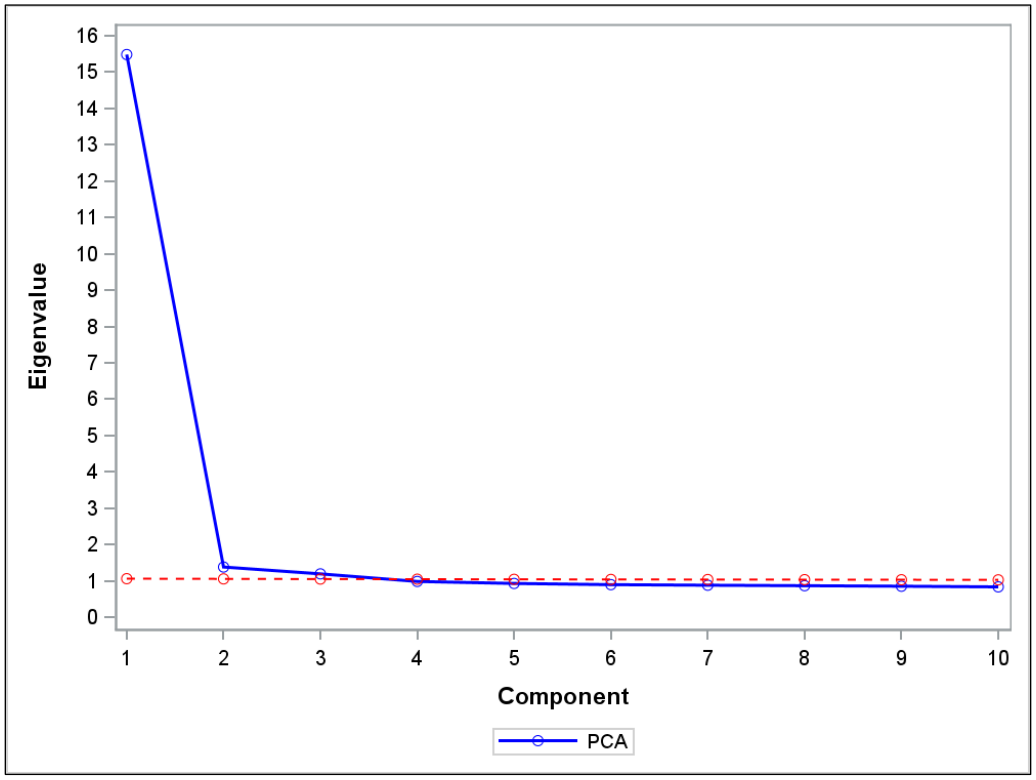


Figure B.66. Scree Plot, ELA Grade 8, Form 2

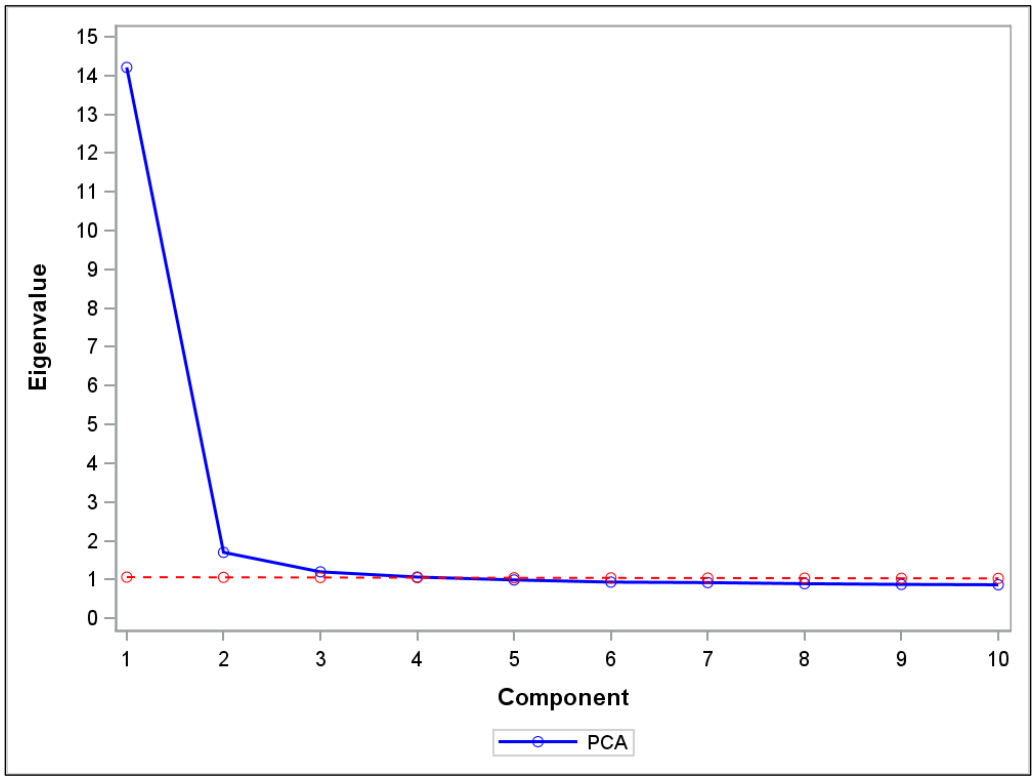


Figure B.67. Scree Plot, Mathematics Grade 3

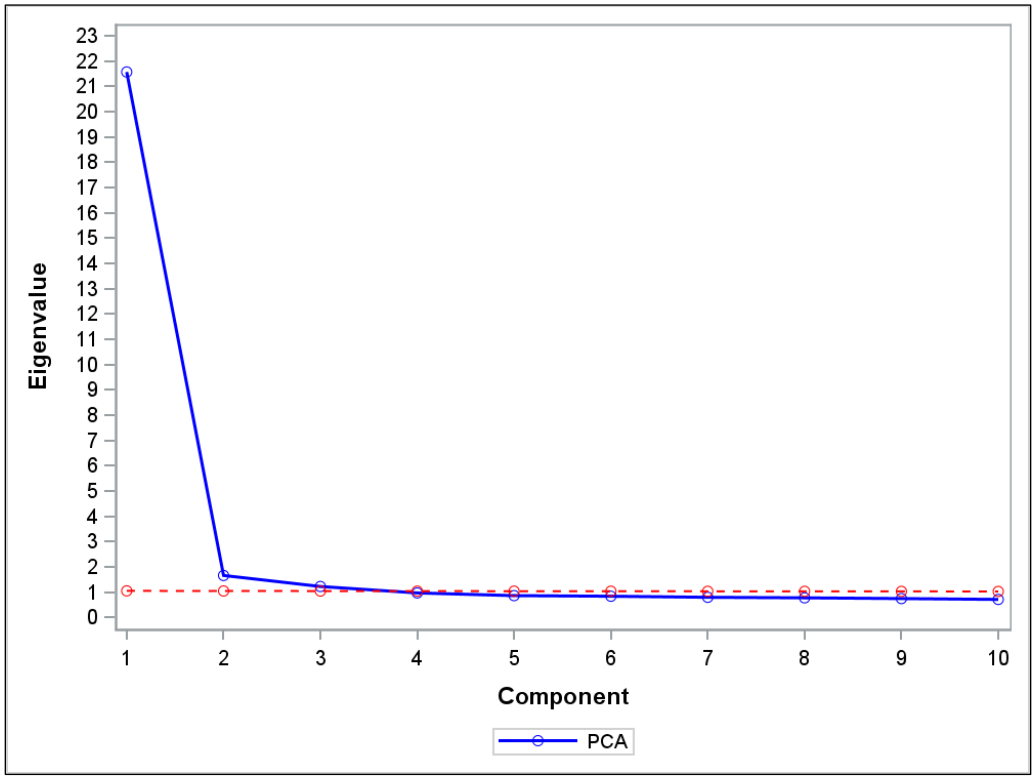


Figure B.68. Scree Plot, Mathematics Grade 4

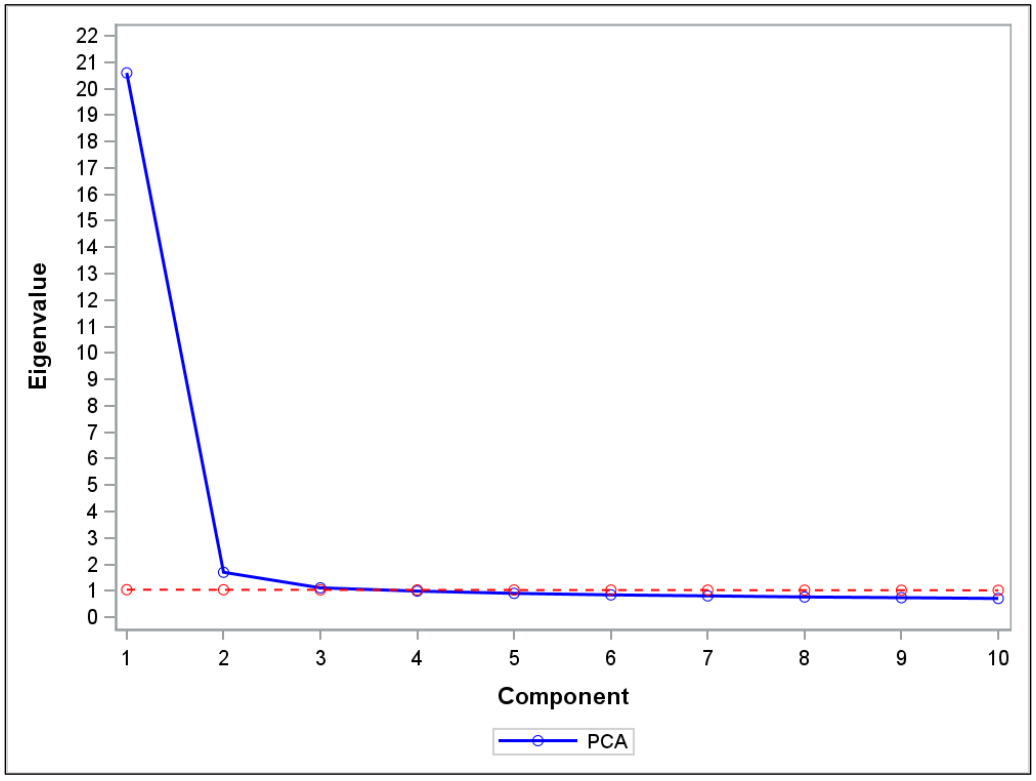


Figure B.69. Scree Plot, Mathematics Grade 5

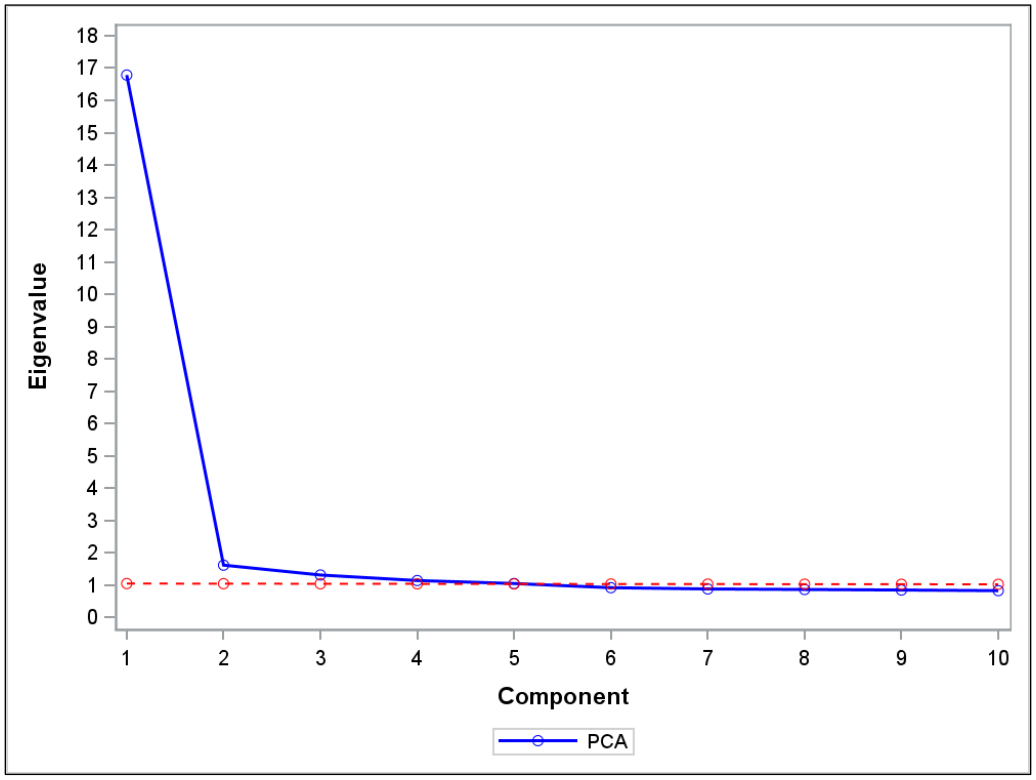


Figure B.70. Scree Plot, Mathematics Grade 6

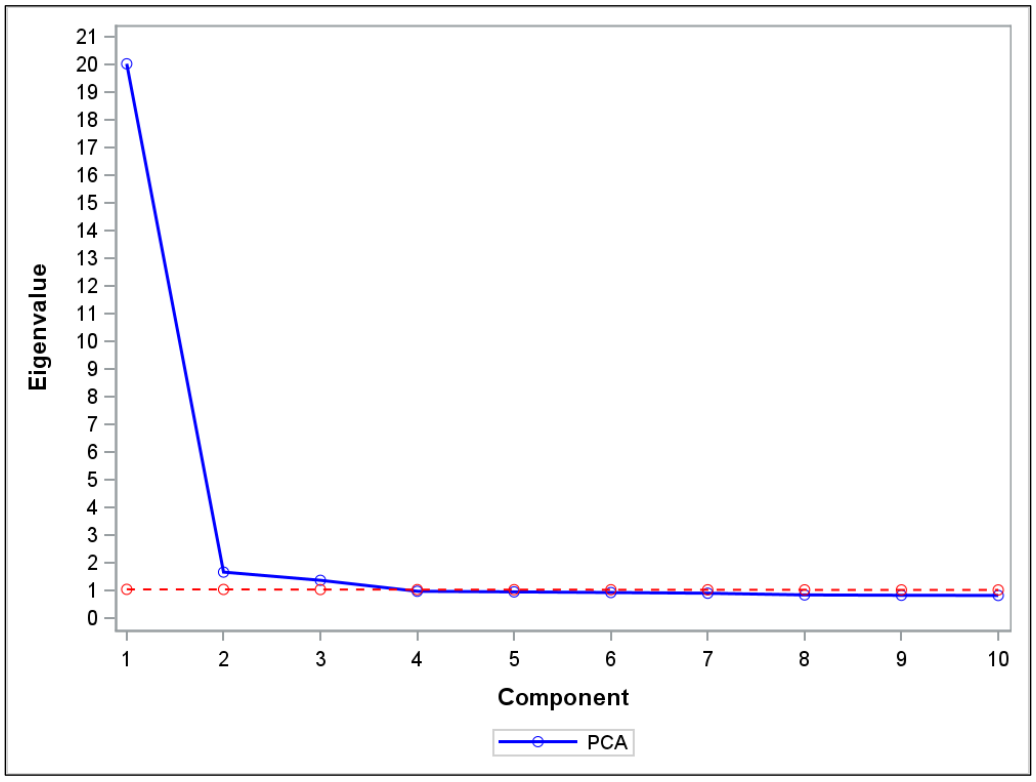


Figure B.71. Scree Plot, Mathematics Grade 7

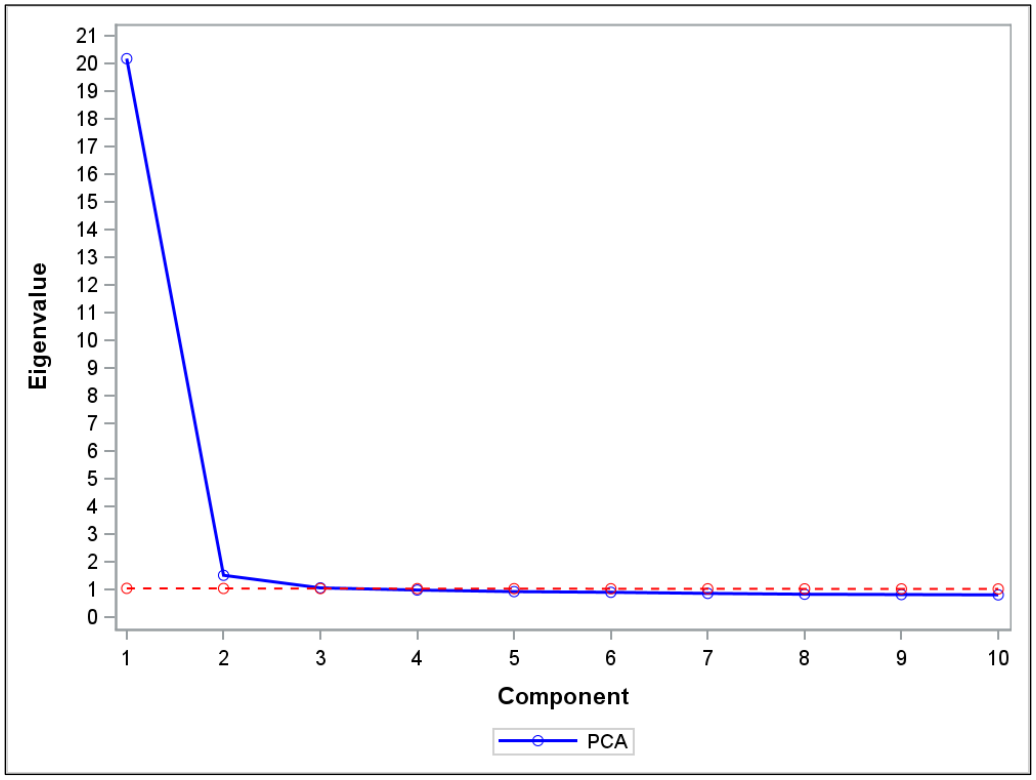
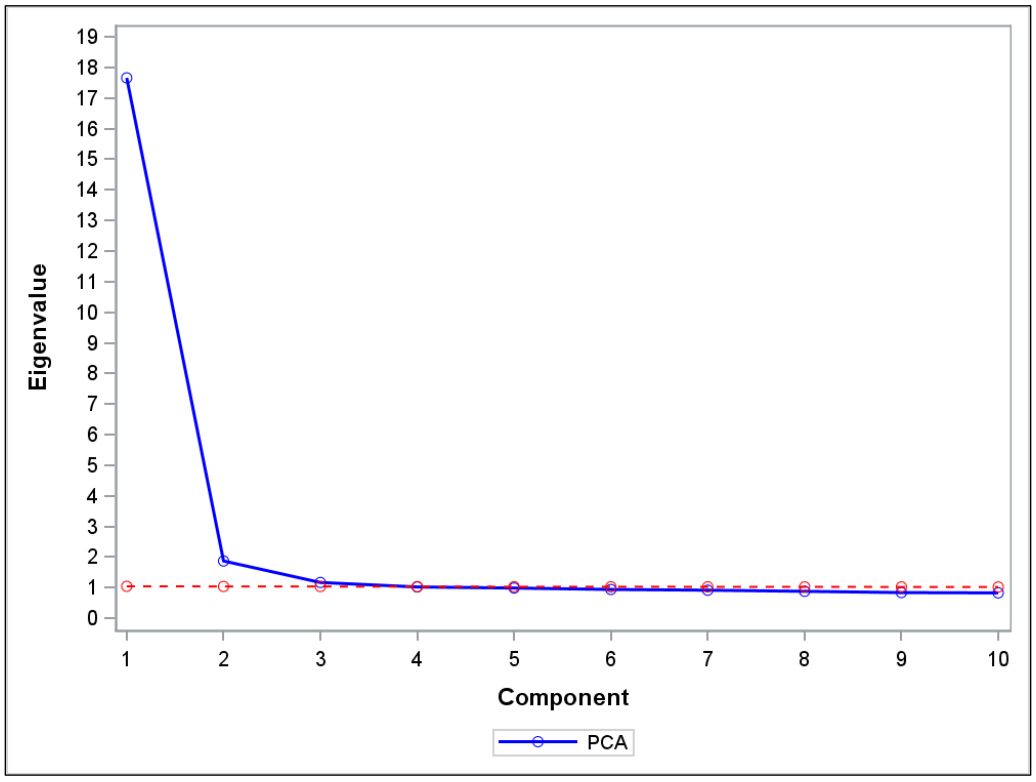


Figure B.72. Scree Plot, Mathematics Grade 8



Appendix C: SPRING 2024 ADMINISTRATION RESULTS

This appendix presents the Spring 2024 AASA results for all students and subgroups by gender, ethnicity (Hispanic or Not-Hispanic), race, and special education, English learner (EL), and low socioeconomic status. Specifically:

- Table C.1 – Table C.12 present the overall results by subgroup, including the sample size, mean and standard deviation (SD) of the total scale score (SS), and percentage of students at each performance level overall.
- Figure C.1 – Figure C.12 present histograms of the total scale score distribution.

Table C.1. Test Results by Subgroup, ELA Grade 3

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4	%MOWR Met
All	77,828	2497.30	35.66	50.4	10.5	25.8	13.4	95.3
Male	39,260	2495.12	35.54	52.8	10.0	25.0	12.2	94.8
Female	38,568	2499.52	35.65	47.8	11.0	26.6	14.6	95.8
Hispanic	37,872	2488.83	33.06	60.5	10.3	21.5	7.7	93.9
Non-Hispanic	39,956	2505.34	36.17	40.8	10.7	29.9	18.7	96.7
American Indian	4,096	2480.94	29.80	71.1	9.6	15.6	3.7	91.9
Asian	2,878	2518.99	35.53	26.2	9.1	33.4	31.3	98.0
Black or African American	5,794	2487.76	32.32	61.7	10.6	21.0	6.7	93.6
Multi-racial	5,244	2504.26	35.67	42.0	10.7	30.1	17.2	96.6
Native Hawaiian or Other Pacific Islander	442	2493.87	31.70	55.7	11.5	24.9	7.9	96.6
White	59,330	2497.74	35.61	49.7	10.6	26.2	13.5	95.5
Missing	44	2477.20	31.88	72.7	9.1	13.6	4.5	93.2
Special Education	12,832	2473.31	31.25	79.0	5.8	10.9	4.3	88.6
English Learner (EL)	11,223	2468.38	22.80	87.1	6.9	5.6	0.4	88.1
Low Socioeconomic Status (SES)	42,227	2487.11	32.24	62.6	10.4	20.5	6.5	93.7
Migrant	439	2473.94	28.62	79.3	5.5	12.1	3.2	88.6

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.2. Test Results by Subgroup, ELA Grade 4

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	80,867	2518.04	37.42	41.6	12.6	29.5	16.2
Male	40,970	2515.50	37.75	44.5	12.3	28.3	15.0
Female	39,897	2520.65	36.90	38.8	12.9	30.8	17.5
Hispanic	39,371	2508.90	34.65	51.2	13.4	26.1	9.3
Non-Hispanic	41,496	2526.71	37.89	32.5	11.9	32.8	22.8
American Indian	4,313	2499.99	30.97	63.6	12.8	19.1	4.5
Asian	3,000	2543.18	38.32	18.8	9.3	33.4	38.5
Black or African American	5,941	2507.88	34.69	53.0	12.7	24.9	9.4
Multi-racial	5,232	2524.20	37.24	34.4	12.7	32.9	20.0
Native Hawaiian or Other Pacific Islander	463	2517.23	35.40	40.8	14.5	31.5	13.2
White	61,879	2518.56	37.17	40.7	12.7	30.2	16.3
Missing	39	2482.69	30.79	82.1	7.7	10.3	0.0
Special Education	12,904	2490.00	32.43	75.4	7.8	12.1	4.7
English Learner (EL)	8,839	2481.42	22.57	87.2	7.4	5.1	0.3
Low Socioeconomic Status (SES)	43,224	2507.21	33.90	53.4	13.3	25.0	8.4
Migrant	417	2494.73	31.96	71.0	7.4	17.7	3.8

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.3. Test Results by Subgroup, ELA Grade 5

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	81,113	2528.98	35.09	39.5	22.5	28.7	9.3
Male	41,103	2525.56	35.18	43.5	21.8	26.9	7.8
Female	40,010	2532.50	34.65	35.4	23.2	30.5	10.8
Hispanic	39,050	2520.71	32.72	48.7	23.5	22.8	5.0
Non-Hispanic	42,063	2536.67	35.47	31.0	21.6	34.2	13.3
American Indian	4,403	2511.90	30.49	60.5	21.1	16.1	2.2
Asian	3,035	2549.49	35.53	18.7	18.8	39.3	23.2
Black or African American	6,117	2518.95	32.48	51.1	22.7	21.7	4.4
Multi-racial	5,178	2534.12	34.55	33.6	22.8	32.5	11.0
Native Hawaiian or Other Pacific Islander	453	2527.04	32.81	39.7	28.7	23.4	8.2
White	61,883	2529.78	34.91	38.4	22.7	29.5	9.5
Missing	44	2507.70	32.53	65.9	15.9	15.9	2.3
Special Education	12,366	2500.42	29.36	76.1	13.4	8.6	1.9
English Learner (EL)	9,126	2496.34	23.37	82.0	13.9	4.0	0.1
Low Socioeconomic Status (SES)	43,460	2519.03	32.12	50.6	23.6	21.5	4.3
Migrant	440	2510.84	33.96	62.5	17.3	15.9	4.3

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.4. Test Results by Subgroup, ELA Grade 6

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	80,561	2542.44	35.78	38.3	18.8	35.4	7.5
Male	40,871	2539.12	35.86	41.6	18.6	33.8	6.1
Female	39,690	2545.86	35.37	34.9	19.1	37.2	8.9
Hispanic	38,364	2533.46	33.56	48.0	20.0	28.4	3.6
Non-Hispanic	42,197	2550.60	35.77	29.5	17.8	41.8	11.0
American Indian	4,370	2524.70	31.13	59.6	18.5	20.4	1.5
Asian	3,024	2565.50	34.16	15.2	14.9	50.1	19.8
Black or African American	5,869	2532.30	33.44	49.3	20.5	26.7	3.4
Multi-racial	5,061	2548.39	35.03	32.0	18.5	40.0	9.6
Native Hawaiian or Other Pacific Islander	445	2541.48	32.62	38.0	22.0	34.8	5.2
White	61,747	2543.07	35.60	37.3	18.9	36.3	7.5
Missing	45	2514.73	26.19	73.3	17.8	8.9	0.0
Special Education	11,260	2511.89	30.44	76.2	11.6	10.9	1.3
English Learner (EL)	7,972	2504.32	23.33	86.6	9.8	3.5	0.0
Low Socioeconomic Status (SES)	41,955	2531.91	33.04	50.0	20.0	26.8	3.2
Migrant	456	2523.38	34.50	59.9	16.4	21.3	2.4

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.5. Test Results by Subgroup, ELA Grade 7

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	81,605	2551.34	34.29	40.7	18.1	32.3	9.0
Male	41,388	2547.68	34.38	45.0	17.5	29.8	7.6
Female	40,217	2555.11	33.78	36.2	18.6	34.8	10.5
Hispanic	38,938	2543.03	32.29	50.5	18.5	25.9	5.0
Non-Hispanic	42,667	2558.93	34.31	31.7	17.6	38.1	12.7
American Indian	4,621	2534.02	29.14	63.0	17.9	17.1	2.1
Asian	3,006	2576.35	33.22	15.6	12.0	46.5	25.8
Black or African American	5,931	2542.41	32.06	51.4	18.3	25.5	4.8
Multi-racial	5,030	2557.22	33.23	33.2	19.2	36.7	10.9
Native Hawaiian or Other Pacific Islander	501	2547.41	31.72	45.3	18.0	31.9	4.8
White	62,466	2551.84	34.08	39.7	18.2	33.0	9.0
Missing	50	2529.26	33.30	66.0	14.0	18.0	2.0
Special Education	10,550	2521.58	27.38	79.5	10.5	8.7	1.4
English Learner (EL)	7,378	2515.49	21.22	88.8	7.6	3.6	0.1
Low Socioeconomic Status (SES)	41,687	2541.35	31.50	52.6	18.7	24.6	4.1
Migrant	469	2529.28	30.77	71.0	11.9	13.9	3.2

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.6. Test Results by Subgroup, ELA Grade 8

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	82,472	2557.26	32.54	42.8	22.0	26.2	9.0
Male	41,971	2553.03	32.61	48.4	21.0	23.3	7.4
Female	40,501	2561.65	31.88	37.1	23.0	29.1	10.7
Hispanic	39,668	2549.73	30.45	52.0	22.4	20.9	4.7
Non-Hispanic	42,804	2564.24	32.87	34.3	21.7	31.0	13.0
American Indian	4,533	2542.15	27.63	63.4	20.8	13.6	2.2
Asian	2,918	2580.44	32.57	17.7	16.6	39.2	26.4
Black or African American	6,039	2550.15	30.30	51.4	22.5	21.5	4.6
Multi-racial	4,881	2562.44	31.55	36.0	23.5	29.5	11.1
Native Hawaiian or Other Pacific Islander	470	2554.09	31.51	46.4	22.8	23.8	7.0
White	63,560	2557.59	32.42	42.2	22.2	26.7	9.0
Missing	71	2543.96	32.57	54.9	25.4	15.5	4.2
Special Education	10,182	2529.22	25.96	80.6	11.6	6.6	1.1
English Learner (EL)	7,111	2523.10	20.79	89.5	8.5	1.9	0.1
Low Socioeconomic Status (SES)	41,582	2548.33	29.97	54.0	22.2	19.6	4.2
Migrant	449	2536.59	28.01	67.5	19.2	12.2	1.1

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.7. Test Results by Subgroup, Mathematics Grade 3

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	79,059	3516.30	46.80	31.4	26.1	29.5	13.1
Male	40,066	3518.05	48.14	30.4	24.8	29.9	14.9
Female	38,993	3514.50	45.31	32.4	27.4	29.1	11.2
Hispanic	38,559	3505.97	44.68	39.0	28.5	24.9	7.6
Non-Hispanic	40,500	3526.14	46.65	24.1	23.8	33.8	18.3
American Indian	4,194	3494.05	42.58	50.1	27.4	18.6	3.9
Asian	2,920	3549.00	42.57	10.9	15.7	38.4	35.0
Black or African American	5,931	3501.64	45.65	42.5	27.5	23.7	6.4
Multi-racial	5,325	3522.03	45.87	26.6	26.0	31.9	15.5
Native Hawaiian or Other Pacific Islander	454	3509.21	44.60	35.5	30.0	26.2	8.4
White	60,185	3517.29	46.28	30.3	26.3	30.2	13.1
Missing	50	3477.88	43.17	60.0	28.0	8.0	4.0
Special Education	13,264	3485.31	47.57	58.7	21.1	15.3	4.9
English Learner (EL)	11,575	3484.87	40.06	58.3	27.2	12.7	1.8
Low Socioeconomic Status (SES)	42,974	3503.78	44.39	40.8	28.7	23.9	6.7
Migrant	451	3493.24	44.65	48.1	29.3	18.4	4.2

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.8. Test Results by Subgroup, Mathematics Grade 4

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	81,598	3545.79	45.86	37.5	26.9	22.0	13.5
Male	41,427	3549.76	47.56	34.7	25.8	23.0	16.5
Female	40,171	3541.69	43.67	40.4	28.1	21.0	10.4
Hispanic	39,784	3534.30	41.85	47.0	28.1	17.5	7.4
Non-Hispanic	41,814	3556.72	46.83	28.6	25.8	26.3	19.3
American Indian	4,366	3522.70	37.57	59.2	25.4	12.0	3.4
Asian	3,023	3580.00	45.79	14.2	19.2	30.4	36.2
Black or African American	6,025	3528.41	40.49	53.0	26.2	15.5	5.4
Multi-racial	5,276	3552.00	45.73	31.9	27.2	24.7	16.1
Native Hawaiian or Other Pacific Islander	468	3542.05	41.70	36.8	34.4	19.7	9.2
White	62,400	3546.96	45.53	36.1	27.4	22.8	13.7
Missing	40	3504.13	31.80	75.0	22.5	0.0	2.5
Special Education	13,160	3515.77	42.01	67.2	18.3	9.8	4.7
English Learner (EL)	9,037	3508.52	32.51	75.0	18.4	5.4	1.1
Low Socioeconomic Status (SES)	43,646	3532.39	41.21	48.5	28.3	16.7	6.5
Migrant	427	3526.57	39.32	53.6	27.9	15.7	2.8

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.9. Test Results by Subgroup, Mathematics Grade 5

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	81,727	3580.40	38.12	36.8	30.1	23.3	9.9
Male	41,453	3581.73	39.58	36.5	28.6	23.7	11.3
Female	40,274	3579.04	36.51	37.1	31.6	22.9	8.4
Hispanic	39,403	3571.18	33.64	45.5	31.6	17.8	5.1
Non-Hispanic	42,324	3588.99	40.00	28.7	28.6	28.4	14.4
American Indian	4,439	3562.62	30.30	56.4	29.4	11.6	2.5
Asian	3,073	3609.32	41.46	13.4	22.5	36.0	28.1
Black or African American	6,187	3566.33	32.80	52.2	28.9	15.4	3.6
Multi-racial	5,208	3584.33	38.48	32.4	30.6	25.4	11.6
Native Hawaiian or Other Pacific Islander	457	3575.30	34.82	41.8	32.8	18.6	6.8
White	62,316	3581.37	37.80	35.3	30.5	24.1	10.0
Missing	47	3553.40	31.69	66.0	19.1	14.9	0.0
Special Education	12,531	3554.80	31.21	68.7	20.5	8.2	2.6
English Learner (EL)	9,297	3552.91	25.33	70.3	23.3	5.7	0.7
Low Socioeconomic Status (SES)	43,779	3569.49	33.08	47.4	31.4	16.8	4.4
Migrant	448	3567.38	32.02	50.2	28.6	16.7	4.5

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.10. Test Results by Subgroup, Mathematics Grade 6

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	81,288	3605.44	41.58	47.8	22.7	18.7	10.8
Male	41,305	3607.06	43.16	46.7	21.8	19.2	12.4
Female	39,983	3603.77	39.82	49.0	23.6	18.2	9.1
Hispanic	38,780	3594.24	36.63	59.2	22.0	13.7	5.1
Non-Hispanic	42,508	3615.66	43.17	37.4	23.4	23.3	15.9
American Indian	4,413	3584.79	33.51	70.3	17.6	9.4	2.7
Asian	3,052	3641.89	43.93	17.5	17.6	29.3	35.6
Black or African American	5,943	3588.83	35.37	65.4	19.6	11.2	3.9
Multi-racial	5,085	3610.05	41.28	43.3	23.5	21.2	12.1
Native Hawaiian or Other Pacific Islander	444	3601.53	37.93	50.9	25.7	16.0	7.4
White	62,302	3606.38	41.07	46.4	23.6	19.4	10.7
Missing	49	3570.53	33.88	85.7	6.1	6.1	2.0
Special Education	11,433	3576.51	33.41	79.9	11.5	5.9	2.6
English Learner (EL)	8,149	3570.63	26.89	86.5	9.9	3.0	0.7
Low Socioeconomic Status (SES)	42,323	3592.52	36.02	61.2	21.5	12.9	4.5
Migrant	462	3591.20	40.25	62.3	19.7	11.5	6.5

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.11. Test Results by Subgroup, Mathematics Grade 7

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	82,360	3626.98	41.96	54.7	16.3	15.5	13.6
Male	41,830	3628.98	43.22	52.8	16.1	15.8	15.3
Female	40,530	3624.92	40.52	56.5	16.5	15.2	11.8
Hispanic	39,347	3615.56	37.01	66.7	15.0	11.3	7.0
Non-Hispanic	43,013	3637.44	43.47	43.6	17.5	19.3	19.6
American Indian	4,683	3604.41	32.20	78.3	12.0	6.6	3.1
Asian	3,028	3666.21	44.34	19.9	13.7	23.1	43.3
Black or African American	6,013	3611.24	35.62	70.8	14.4	9.4	5.4
Multi-racial	5,058	3633.17	41.42	47.9	17.5	18.4	16.1
Native Hawaiian or Other Pacific Islander	503	3620.96	38.05	58.8	20.1	12.9	8.2
White	63,027	3627.85	41.46	53.5	16.8	16.1	13.5
Missing	48	3606.96	36.31	77.1	12.5	6.3	4.2
Special Education	10,703	3596.12	30.88	85.9	7.4	4.2	2.5
English Learner (EL)	7,573	3591.65	25.40	91.2	5.7	2.2	0.9
Low Socioeconomic Status (SES)	42,083	3613.67	35.95	68.6	14.8	10.5	6.0
Migrant	477	3602.12	31.50	81.8	9.6	5.5	3.1

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Table C.12. Test Results by Subgroup, Mathematics Grade 8

Subgroup	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
All	83,137	3653.89	37.67	54.6	17.8	16.5	11.2
Male	42,337	3655.16	39.46	53.8	16.8	16.6	12.8
Female	40,800	3652.57	35.67	55.4	18.8	16.3	9.5
Hispanic	40,012	3644.48	32.33	65.6	16.3	12.2	5.9
Non-Hispanic	43,125	3662.61	40.10	44.3	19.1	20.4	16.2
American Indian	4,564	3637.34	27.79	74.7	13.8	8.6	3.0
Asian	2,941	3692.24	45.78	21.3	14.1	23.5	41.1
Black or African American	6,086	3641.87	30.29	68.6	16.0	10.9	4.5
Multi-racial	4,903	3657.74	37.48	49.4	19.2	19.0	12.3
Native Hawaiian or Other Pacific Islander	475	3650.78	34.50	56.6	20.6	14.5	8.2
White	64,095	3654.20	37.19	53.7	18.2	17.1	11.0
Missing	73	3630.89	28.52	82.2	11.0	2.7	4.1
Special Education	10,305	3629.13	25.20	86.3	7.6	4.2	1.9
English Learner (EL)	7,250	3626.78	21.79	89.2	7.1	2.7	1.0
Low Socioeconomic Status (SES)	41,904	3643.23	31.36	67.2	16.3	11.4	5.2
Migrant	461	3634.91	29.21	77.4	12.6	6.7	3.3

Note. SS = scale score, SD = standard deviation, 1 = *Minimally Proficient*, 2 = *Partially Proficient*, 3 = *Proficient*, 4 = *Highly Proficient*

Figure C.1. Total Scale Score Distribution, ELA Grade 3

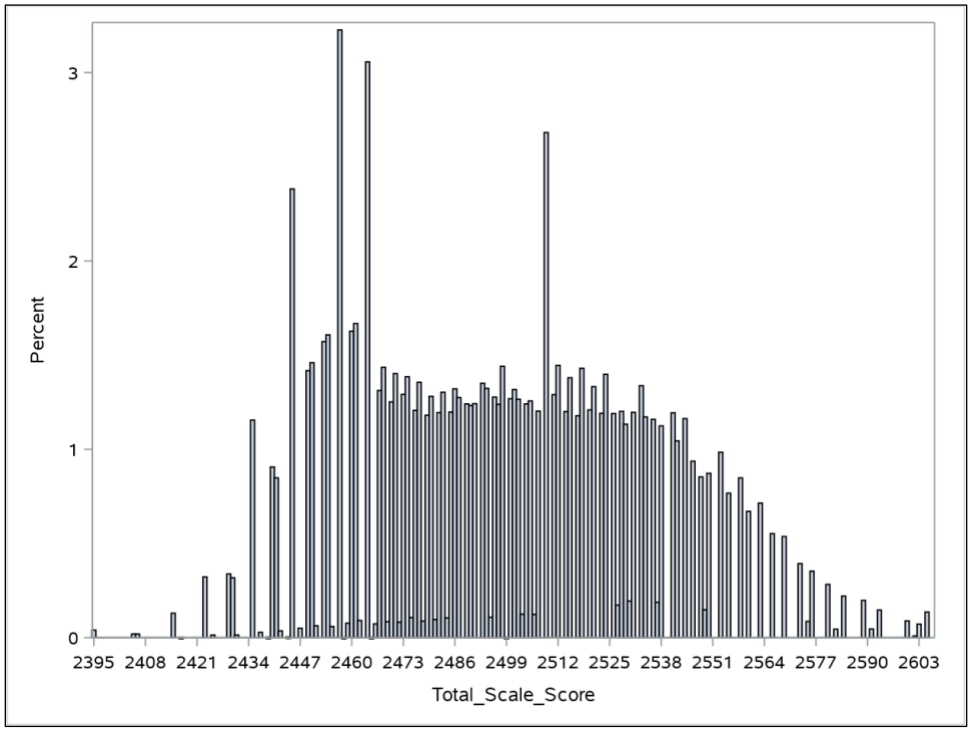


Figure C.2. Total Scale Score Distribution, ELA Grade 4

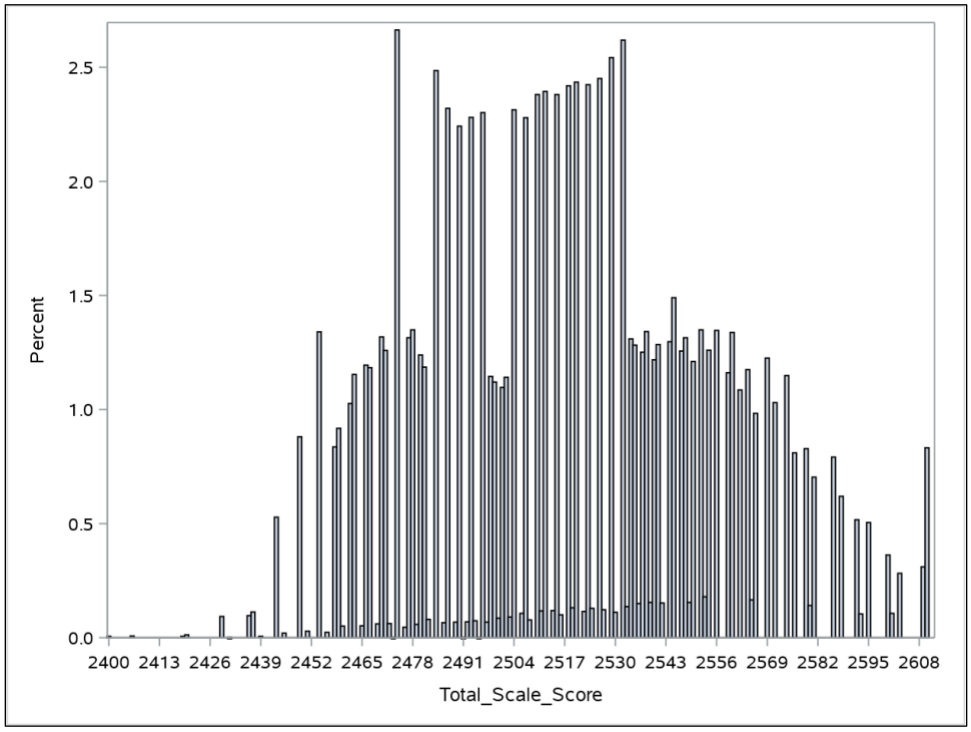


Figure C.3. Total Scale Score Distribution, ELA Grade 5

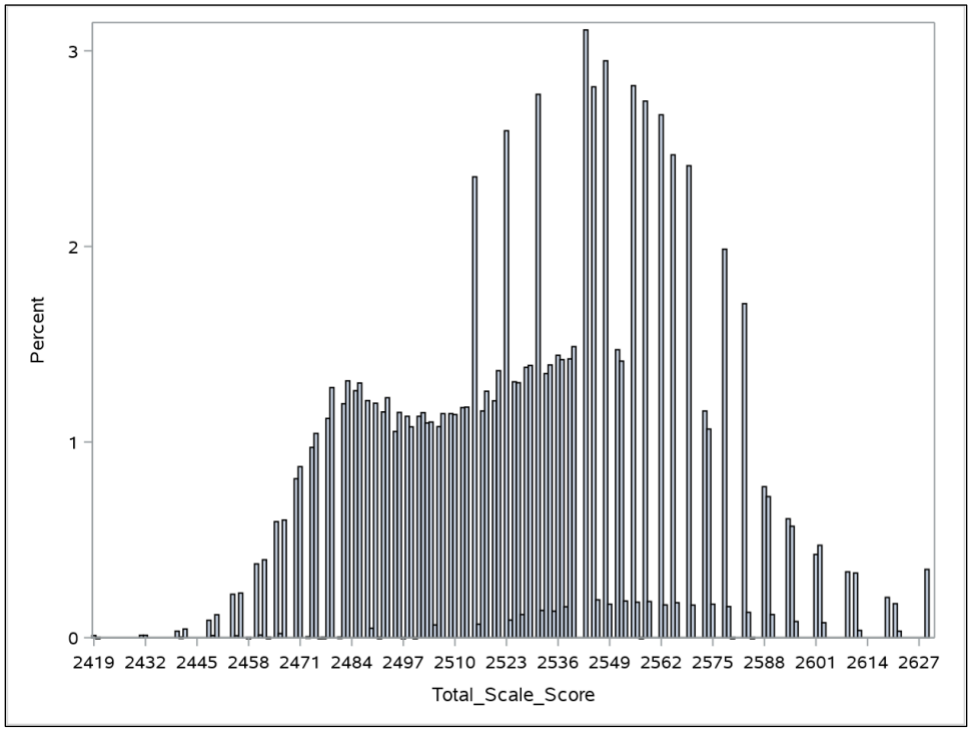


Figure C.4. Total Scale Score Distribution, ELA Grade 6

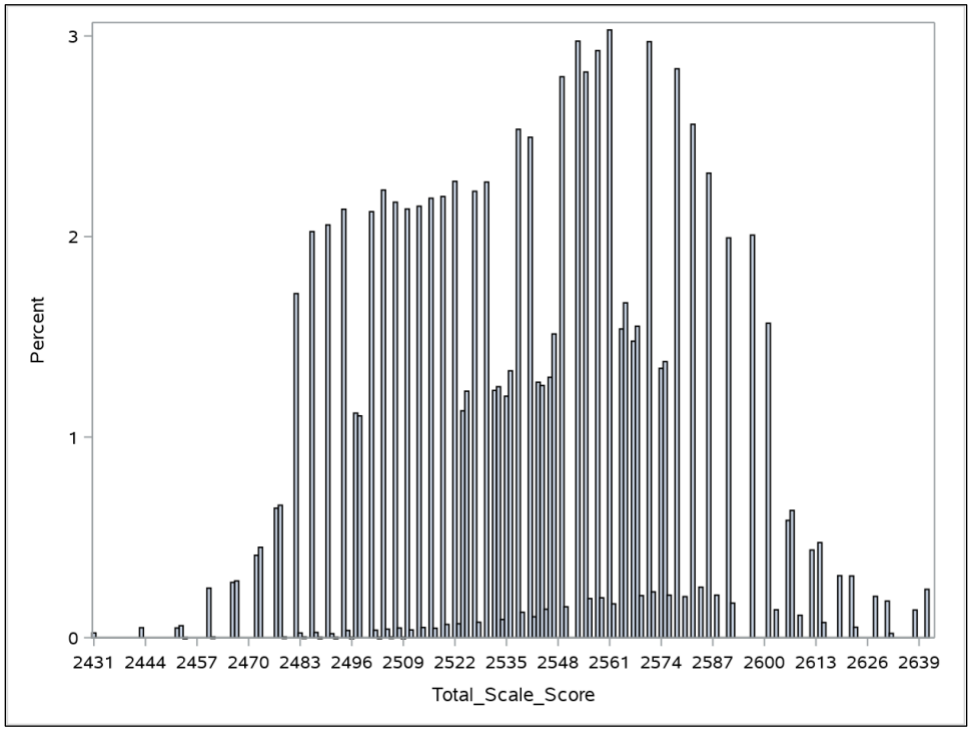


Figure C.5. Total Scale Score Distribution, ELA Grade 7

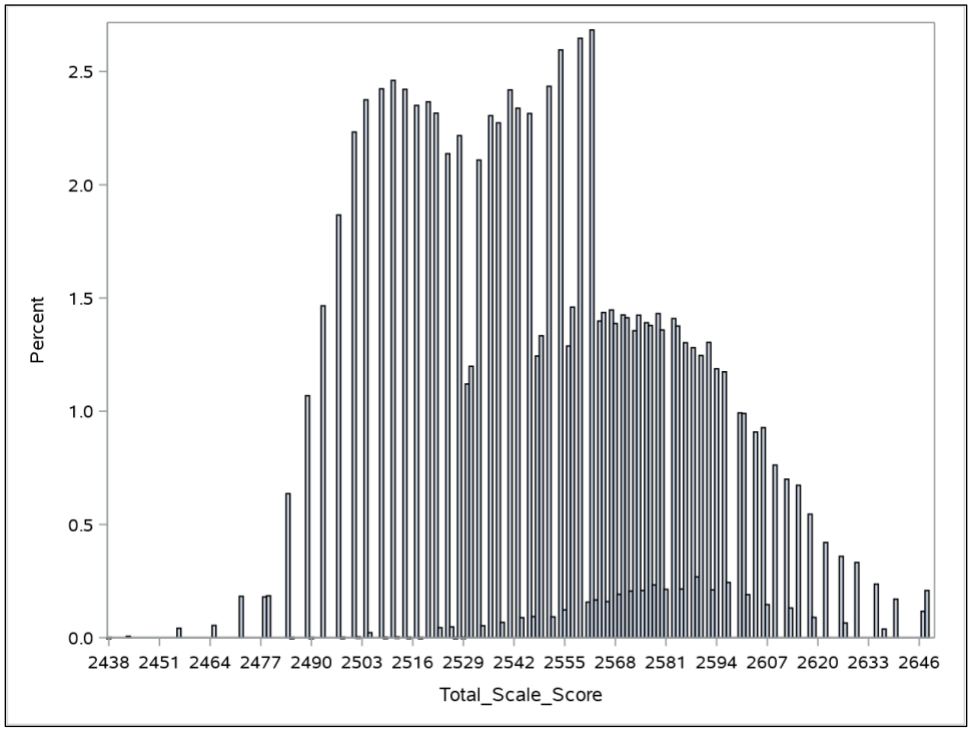


Figure C.6. Total Scale Score Distribution, ELA Grade 8

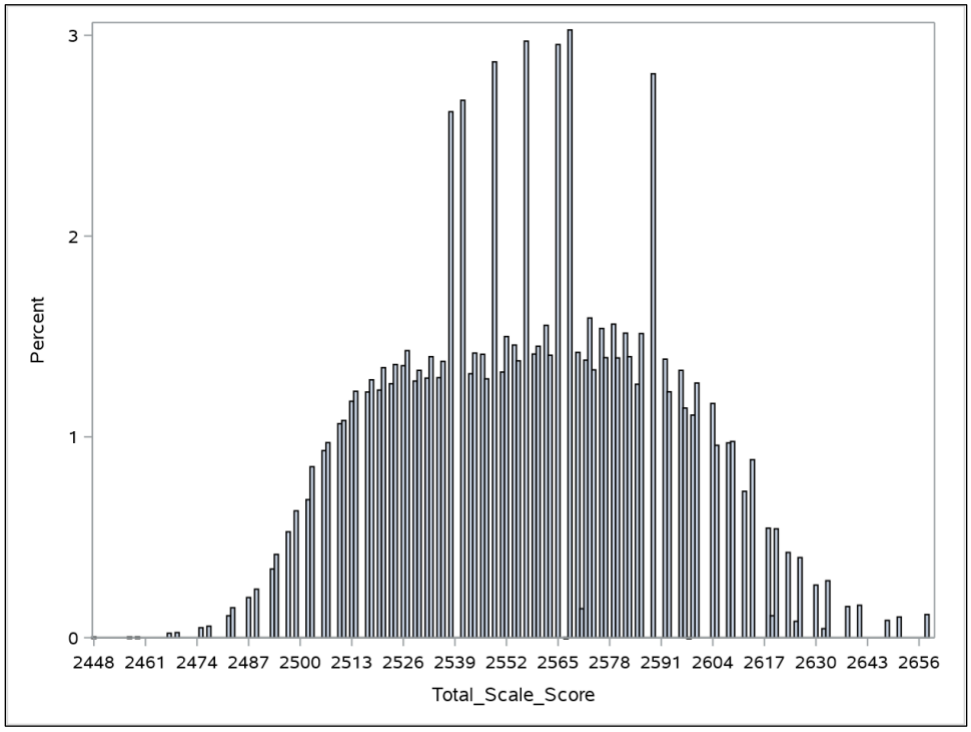


Figure C.7. Total Scale Score Distribution, Mathematics Grade 3

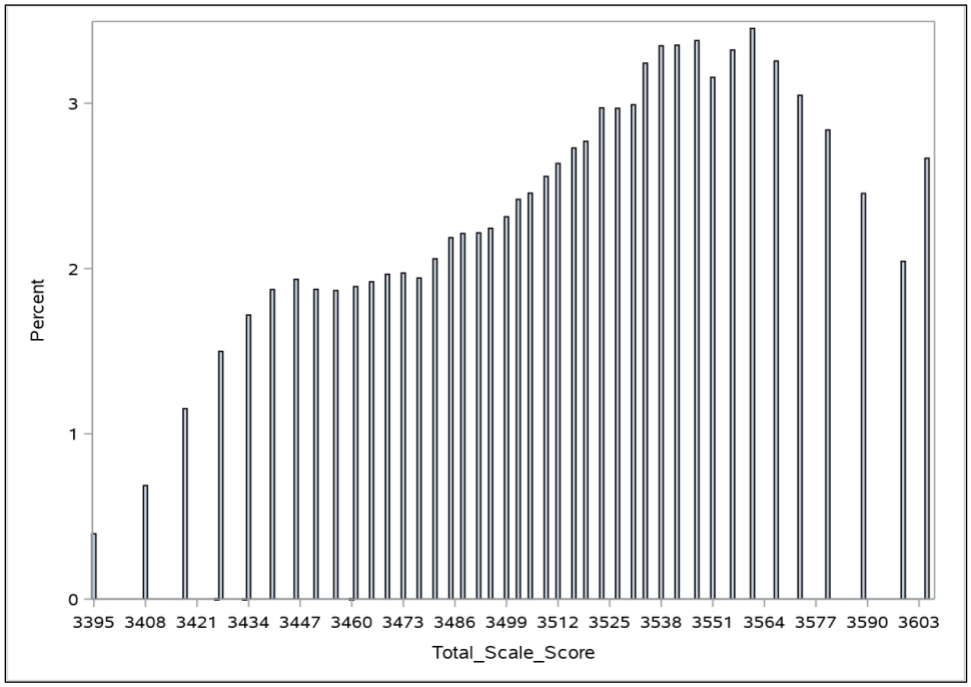


Figure C.8. Total Scale Score Distribution, Mathematics Grade 4

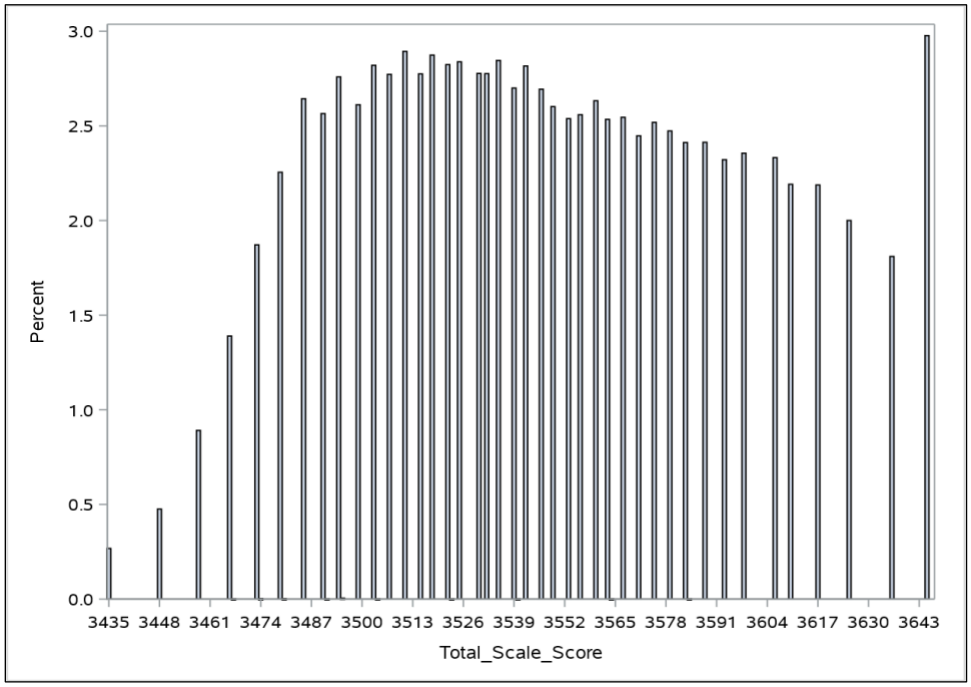


Figure C.9. Total Scale Score Distribution, Mathematics Grade 5

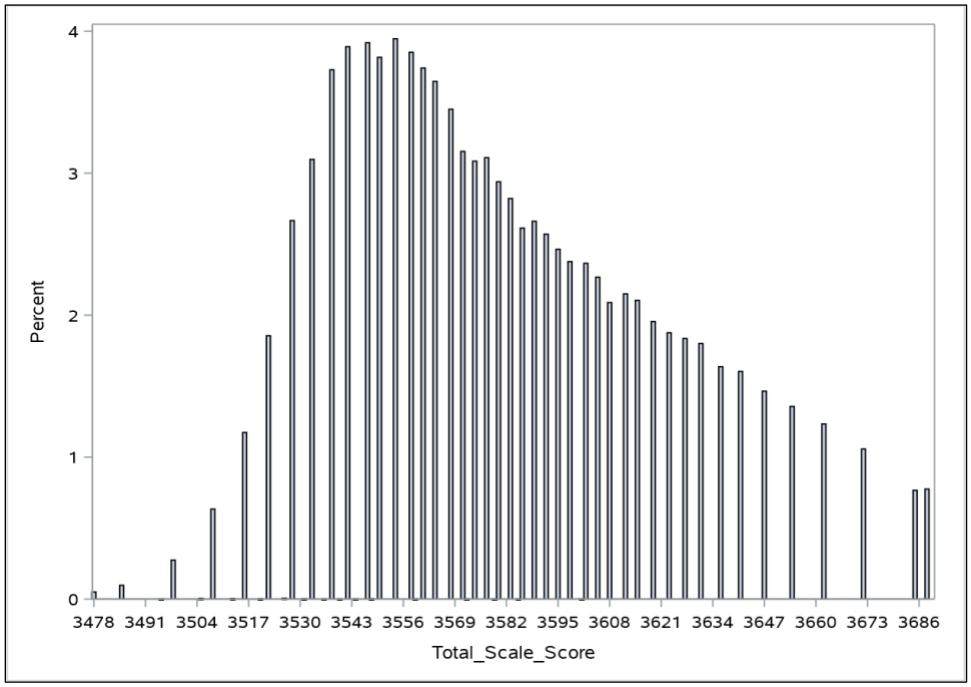


Figure C.10. Total Scale Score Distribution, Mathematics Grade 6

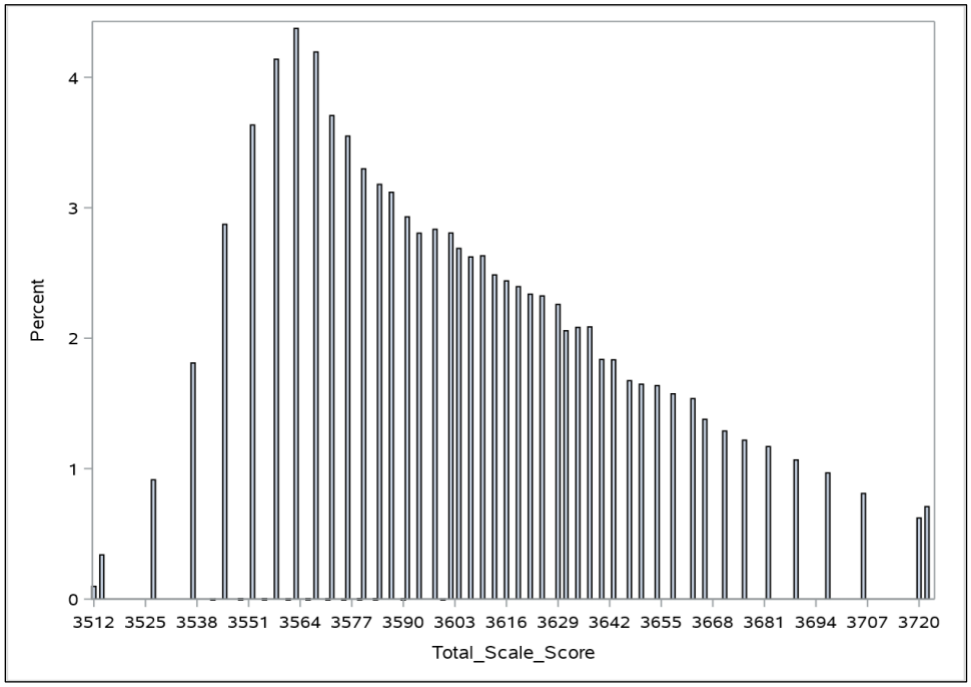


Figure C.11. Total Scale Score Distribution, Mathematics Grade 7

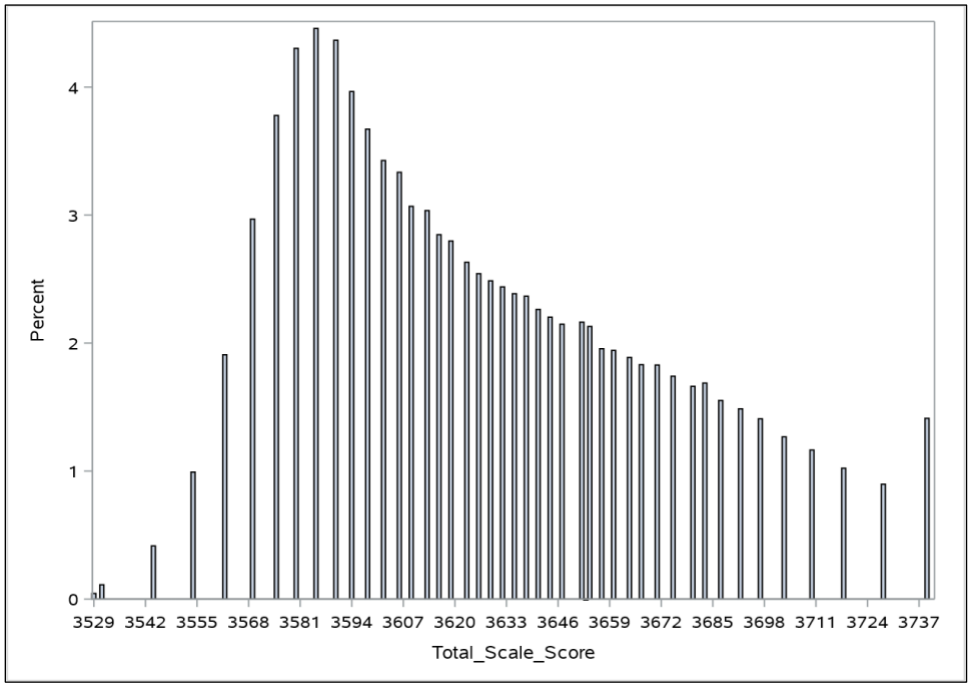
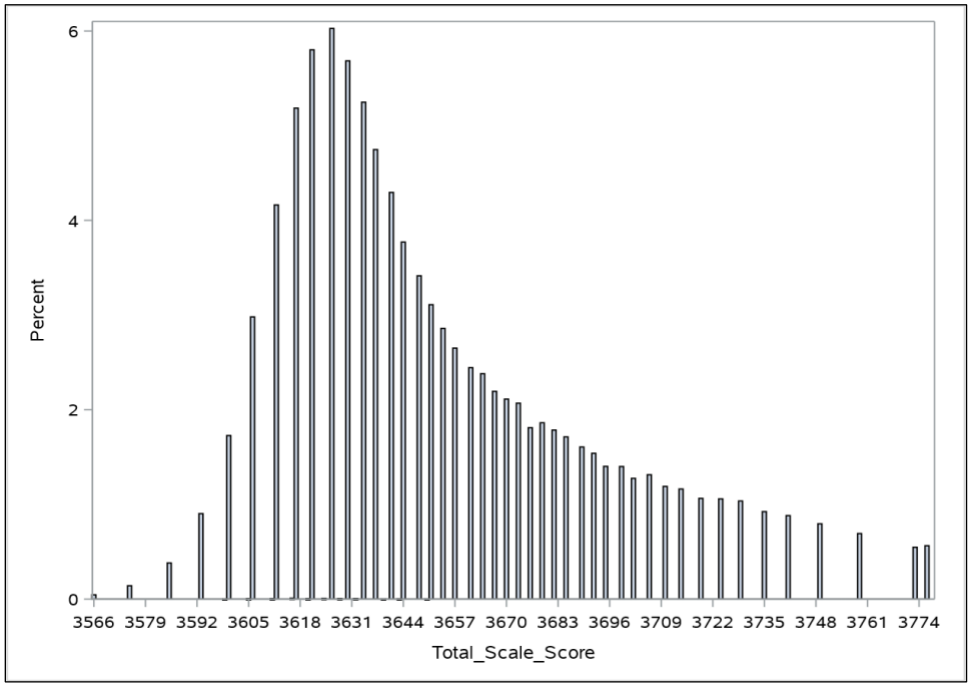


Figure C.12. Total Scale Score Distribution, Mathematics Grade 8



Appendix D: SUMMER 2024 ADMINISTRATION RESULTS

On September 25, 2023, the Arizona State Board of Education approved an additional reading demonstration tool to meet the requirements in R7-2-318, the AASA Grade 3 ELA-Reading summer administration. Grade 3 students who fail to meet the Move on When Reading (MOWR) indicator on the spring AASA ELA-Reading administration and who do not qualify for one of the good cause exemptions related to students with IEPs, a diagnosis of dyslexia, or being an English learner (EL) are eligible.

The current MOWR cut scores were used to determine if the student met the MOWR indicator for the Spring 2024 administration and for the summer administration. Grade 3 students participating in the summer administration were only administered the reading sections of the ELA test (i.e., no writing prompt and no ORF items). The summer administration was available as a computer-based test (CBT). Table D.1 and Table D.2 present the Summer 2024 administration results.

Table D.1. Summer 2024 Overall Test Results

Content Area	Grade	N	%MOWR Met
ELA – Reading	3	159	84.9

Table D.2. Summer 2024 Performance Level Distributions by Reporting Category: Percentage of Students at each Level of Mastery

Content Area	Grade	Reporting Category	N	%Level 1	%Level 2	%Level 3
ELA – Reading	3	Reading for Information	159	93.1	6.3	0.6
		Reading For Literature	159	92.5	6.9	0.6

Note. Level 1 = *Below Mastery*, Level 2 = *At or Around Mastery*, Level 3 = *Above Mastery*