

AASA
Arizona ELA & Mathematics Assessments

2025
Technical Report

Submitted to the
Arizona Department of Education
October 2025



FOREWORD

This technical report provides information on the Arizona Academic Standards Assessment (AASA), the statewide accountability assessment administered to Arizona students in grades 3–8 each spring to assess their performance on the Arizona Academic Standards in English language arts (ELA) and mathematics. 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).

The Arizona Department of Education (ADE) is a service organization committed to raising academic outcomes and empowering parents. The ADE Assessments agency is responsible for statewide assessment of students enrolled in Arizona public schools, working closely with educators in the development and administration of statewide assessments. All Arizona public schools, including district schools and charter schools, are required to properly administer state and federally mandated assessments.

Arizona Department of Education

1535 West Jefferson Street

Phoenix, AZ 85007

(602) 542-5057 (Assessments)

<https://www.azed.gov/>



ARIZONA DEPARTMENT OF
EDUCATION

Copyright © 2025 by the Arizona Department of Education (ADE). All rights reserved. This report was created by Pearson under contract with ADE. The content and format of this report is determined by ADE. Only State of Arizona educators and citizens may copy, download, and/or print the document. Any other use or reproduction of this document, in whole or in part, requires written permission of ADE.

TABLE OF CONTENTS

Chapter 1: Introduction	10
1.1. Assessment Overview	10
1.2. Background.....	11
1.3. Testing Requirements	11
1.4. Educator Involvement.....	11
Chapter 2: Test Design	13
2.1. Academic Content Standards	13
2.2. Item Specifications.....	13
2.3. Test Blueprint	14
2.4. Item Types	15
2.5. Cognitive Complexity	16
2.6. Test Structure	17
2.6.1. ELA.....	19
2.6.1.1. Extended Writing Items	19
2.6.1.2. Grade 3 ORF.....	20
2.6.2. Mathematics	20
Chapter 3: Test Development	22
3.1. Content Development and Management Tool	23
3.2. Item Bank Analysis.....	23
3.3. Passage and Item Development	24
3.4. Item Review	26
3.5. Form Construction	26
3.5.1. Preparation for Item Selection.....	26
3.5.2. Item Selection and Positioning.....	26
3.5.3. Sampling Plan	27
3.6. Data Review.....	27
3.7. Accommodated Forms	28
Chapter 4: Test Administration.....	30
4.1. Test Units.....	30
4.2. Administration Materials	31
4.3. Administration Training.....	31
4.4. Sample Tests	32
4.5. Accommodations	33
4.6. Universal Test Administration Conditions.....	34
4.7. Universal Test Tools	35
4.8. Pearson Customer Support.....	37
4.9. Test Security	37
Chapter 5: Scoring.....	40
5.1. Human Scoring of Open-Ended Items	40
5.1.1. Scorer Recruitment.....	40
5.1.2. Training	41
5.1.2.1. Writing	41
5.1.2.2. Mathematics and Reading.....	41
5.1.2.3. Grade 3 ORF.....	42
5.1.3. Quality Control.....	42
5.1.4. Security	43
5.2. IEA Scoring of ELA Writing Prompts	43
5.2.1. Calibration of IEA.....	44
5.2.2. Smart Routing	44

5.2.3. Quality Control.....	45
5.3. Versant Scoring of Grade 3 ORF Items	45
5.3.1. Training.....	46
5.3.2. ORF Score Points	46
Chapter 6: Reporting.....	47
Chapter 7: Classical Item Analysis.....	52
7.1. Data.....	52
7.2. Descriptive Statistics.....	53
7.3. Classical Item Analysis.....	54
7.4. Distractor Analysis.....	55
Chapter 8: Calibration, Equating, and Scaling.....	57
8.1. Calibration Sample.....	57
8.2. Calibration Methods.....	57
8.3. Calibration Results.....	58
8.4. Equating.....	59
8.5. Scaling Methods.....	60
8.6. IRT Assumptions	60
8.6.1. Unidimensionality	61
8.6.2. Local Item Independence	62
8.6.3. Item Fit.....	63
Chapter 9: Test Results	64
Chapter 10: Reliability and Validity	72
10.1. Reliability.....	72
10.1.1. Internal Consistency	72
10.1.2. Inter-rater Reliability	75
10.2. Differential Item Functioning	77
10.3. Correlations Among Reporting Categories	79
10.4. Validity Evidence.....	82
10.4.1. Evidence Based on Test Content	83
10.4.2. Evidence Based on Response Processes.....	84
10.4.3. Evidence Based on Internal Structure	84
10.4.4. Evidence Based on Performance Standards	85
10.4.5. Evidence Based on Relation to Other Variables	85
10.4.6. Summary	86
Chapter 11: Classification into Performance Levels	87
11.1. Standard Setting.....	87
11.2. Classification Consistency and Accuracy	88
References.....	91
Appendix A: Item-Level CTT Statistics.....	94
Appendix B: Item-Level IRT Statistics.....	120
Appendix C: Spring 2025 Administration Results.....	190

LIST OF TABLES

Table 1.1. Schedule of Major Events.....	12
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	15
Table 2.4. AASA Mathematics Blueprint, Grade 8	15
Table 2.5. Item Types	15
Table 2.6. DOK Levels.....	17
Table 2.7. Percentage of Points by DOK Level.....	17
Table 2.8. AASA Test Structure—ELA	18
Table 2.9. AASA Test Structure—Mathematics	18
Table 3.1. Number of Newly Developed Items	23
Table 3.2. Passage Lexile Measures and Word Count.....	24
Table 3.3. Item Statistical Flagging Criteria	27
Table 3.4. Data Review Results: Number of Field-Tested Items	28
Table 4.1. Estimated Testing Time by Test Unit	31
Table 4.2. Administration Materials	31
Table 4.3. Administration Trainings.....	32
Table 4.4. Number of Items on the AASA Sample Tests	33
Table 4.5. Available Accommodations.....	33
Table 4.6. Universal Test Tools.....	35
Table 5.1. Scoring Qualification Standards	41
Table 7.1. Number of Students in the Calibration Sample by Subgroup—ELA	52
Table 7.2. Number of Students in the Calibration Sample by Subgroup—Mathematics	53
Table 7.3. Classical Test Analysis Statistics.....	53
Table 7.4. Classical Item Analysis Summary	54
Table 7.5. Distractor Analysis Summary: Point-Biserial Correlations for Correct Options	55
Table 7.6. Distractor Analysis Summary: Point-Biserial Correlations for Incorrect Options.....	56
Table 8.1. IRT Statistics Summary	58
Table 8.2. Summary of Anchor Items.....	60
Table 8.3. Eigenvalues from PCA	61
Table 8.4. Q3 Statistics.....	62
Table 8.5. IRT Item Fit Summary Statistics	63
Table 9.1. Overall Test Results by Year.....	65
Table 9.2. Performance Distributions by Reporting Category: Percentage of Students at each Level of Mastery—ELA.....	66
Table 9.3. Performance Distributions by Reporting Category: Percentage of Students at each Level of Mastery—Mathematics	66
Table 9.4. Test Results by Accommodation—ELA	67
Table 9.5. Test Results by Accommodation—Mathematics.....	68
Table 9.6. Scale Score Distribution by Performance Level—ELA	70
Table 9.7. Scale Score Distribution by Performance Level—Mathematics	71
Table 10.1. Coefficient Alpha and SEM by Total and Reporting Category Score—ELA, Form 1	73
Table 10.2. Coefficient Alpha and SEM by Total and Reporting Category Score—ELA, Form 2	73
Table 10.3. Coefficient Alpha and SEM by Total and Reporting Category Score—Mathematics.....	74
Table 10.4. Inter-rater Reliability Statistics.....	76
Table 10.5. DIF Flag Categories.....	78
Table 10.6. Number of Items Exhibiting Strong DIF	79
Table 10.7. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—ELA Form 1	80

Table 10.8. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—ELA Form 2	81
Table 10.9. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—Mathematics Grades 3–5	81
Table 10.10. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—Mathematics Grades 6 and 7	82
Table 10.11. Correlations and Disattenuated Correlations between Total and Reporting Category Raw Score—Mathematics Grade 8	82
Table 10.12. Correlation between AASA ELA and Mathematics Scale Scores	85
Table 11.1. Performance Level Cut Scores	87
Table 11.2. CSEM at Performance Level Cuts	88
Table 11.3. Classification Consistency and Accuracy for the <i>Proficient</i> Cut	89
Table 11.4. Classification Consistency and Accuracy Results	90
Table A.1. Item-Level CTT Statistics, ELA Grade 3	94
Table A.2. Item-Level CTT Statistics, ELA Grade 4	95
Table A.3. Item-Level CTT Statistics, ELA Grade 5	96
Table A.4. Item-Level CTT Statistics, ELA Grade 6	97
Table A.5. Item-Level CTT Statistics, ELA Grade 7	98
Table A.6. Item-Level CTT Statistics, ELA Grade 8	100
Table A.7. Item-Level CTT Statistics, Mathematics Grade 3	101
Table A.8. Item-Level CTT Statistics, Mathematics Grade 4	102
Table A.9. Item-Level CTT Statistics, Mathematics Grade 5	103
Table A.10. Item-Level CTT Statistics, Mathematics Grade 6	104
Table A.11. Item-Level CTT Statistics, Mathematics Grade 7	105
Table A.12. Item-Level CTT Statistics, Mathematics Grade 8	106
Table A.13. Distractor Analysis of Multiple-Choice Items, ELA Grade 3	108
Table A.14. Distractor Analysis of Multiple-Choice Items, ELA Grade 4	109
Table A.15. Distractor Analysis of Multiple-Choice Items, ELA Grade 5	110
Table A.16. Distractor Analysis of Multiple-Choice Items, ELA Grade 6	111
Table A.17. Distractor Analysis of Multiple-Choice Items, ELA Grade 7	112
Table A.18. Distractor Analysis of Multiple-Choice Items, ELA Grade 8	113
Table A.19. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 3	114
Table A.20. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 4	115
Table A.21. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 5	116
Table A.22. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 6	117
Table A.23. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 7	118
Table A.24. Distractor Analysis of Multiple-Choice Items, Mathematics Grade 8	119
Table B.1. Item-Level IRT Statistics, ELA Grade 3	120
Table B.2. Item-Level IRT Statistics, ELA Grade 4	121
Table B.3. Item-Level IRT Statistics, ELA Grade 5	122
Table B.4. Item-Level IRT Statistics, ELA Grade 6	123
Table B.5. Item-Level IRT Statistics, ELA Grade 7	125
Table B.6. Item-Level IRT Statistics, ELA Grade 8	126
Table B.7. Item-Level IRT Statistics, Mathematics Grade 3	127
Table B.8. Item-Level IRT Statistics, Mathematics Grade 4	128
Table B.9. Item-Level IRT Statistics, Mathematics Grade 5	129
Table B.10. Item-Level IRT Statistics, Mathematics Grade 6	130
Table B.11. Item-Level IRT Statistics, Mathematics Grade 7	131
Table B.12. Item-Level IRT Statistics, Mathematics Grade 8	132
Table B.13. Raw-to-Scale Score Conversion, ELA Grade 3	134
Table B.14. Raw-to-Scale Score Conversion, ELA Grade 4	136
Table B.15. Raw-to-Scale Score Conversion, ELA Grade 5	138

Table B.16. Raw-to-Scale Score Conversion, ELA Grade 6	141
Table B.17. Raw-to-Scale Score Conversion, ELA Grade 7	143
Table B.18. Raw-to-Scale Score Conversion, ELA Grade 8	145
Table B.19. Raw-to-Scale Score Conversion, Mathematics Grade 3	147
Table B.20. Raw-to-Scale Score Conversion, Mathematics Grade 4	148
Table B.21. Raw-to-Scale Score Conversion, Mathematics Grade 5	149
Table B.22. Raw-to-Scale Score Conversion, Mathematics Grade 6	150
Table B.23. Raw-to-Scale Score Conversion, Mathematics Grade 7	151
Table B.24. Raw-to-Scale Score Conversion, Mathematics Grade 8	152
Table C.1. Test Results by Subgroup, ELA Grade 3	190
Table C.2. Test Results by Subgroup, ELA Grade 4	191
Table C.3. Test Results by Subgroup, ELA Grade 5	191
Table C.4. Test Results by Subgroup, ELA Grade 6	192
Table C.5. Test Results by Subgroup, ELA Grade 7	192
Table C.6. Test Results by Subgroup, ELA Grade 8	193
Table C.7. Test Results by Subgroup, Mathematics Grade 3	193
Table C.8. Test Results by Subgroup, Mathematics Grade 4	194
Table C.9. Test Results by Subgroup, Mathematics Grade 5	194
Table C.10. Test Results by Subgroup, Mathematics Grade 6	195
Table C.11. Test Results by Subgroup, Mathematics Grade 7	195
Table C.12. Test Results by Subgroup, Mathematics Grade 8	196
Table C.13. State Mean Raw Score Points by Reporting Category and Cluster—ELA Online	197
Table C.14. State Mean Raw Score Points by Reporting Category and Cluster—Mathematics Online	199

LIST OF FIGURES

Figure 3.1. Item Development Process	22
Figure 3.2. Text Complexity Worksheet Example	24
Figure 4.1. Test Security Agreement	38
Figure 6.1. Sample Reports—Confidential Student Score Report	49
Figure 6.2. Sample Reports—Confidential Roster Report with Summary	51
Figure B.1. Item-Person Map, ELA Grade 3, Form 1	154
Figure B.2. Item-Person Map, ELA Grade 3, Form 2	154
Figure B.3. Item-Person Map, ELA Grade 4, Form 1	155
Figure B.4. Item-Person Map, ELA Grade 4, Form 2	155
Figure B.5. Item-Person Map, ELA Grade 5, Form 1	156
Figure B.6. Item-Person Map, ELA Grade 5, Form 2	156
Figure B.7. Item-Person Map, ELA Grade 6, Form 1	157
Figure B.8. Item-Person Map, ELA Grade 6, Form 2	157
Figure B.9. Item-Person Map, ELA Grade 7, Form 1	158
Figure B.10. Item-Person Map, ELA Grade 7, Form 2	158
Figure B.11. Item-Person Map, ELA Grade 8, Form 1	159
Figure B.12. Item-Person Map, ELA Grade 8, Form 2	159
Figure B.13. Item-Person Map, Mathematics Grade 3	160
Figure B.14. Item-Person Map, Mathematics Grade 4	160
Figure B.15. Item-Person Map, Mathematics Grade 5	161
Figure B.16. Item-Person Map, Mathematics Grade 6	161
Figure B.17. Item-Person Map, Mathematics Grade 7	162
Figure B.18. Item-Person Map, Mathematics Grade 8	162
Figure B.19. TCC, ELA Grade 3, Form 1	163
Figure B.20. TIF and CSEM, ELA Grade 3, Form 1	163

Figure B.21. TCC, ELA Grade 4, Form 1	164
Figure B.22. TIF and CSEM, ELA Grade 4, Form 1.....	164
Figure B.23. TCC, ELA Grade 5, Form 1	165
Figure B.24. TIF and CSEM, ELA Grade 5, Form 1.....	165
Figure B.25. TCC, ELA Grade 6, Form 1	166
Figure B.26. TIF and CSEM, ELA Grade 6, Form 1.....	166
Figure B.27. TCC, ELA Grade 7, Form 1	167
Figure B.28. TIF and CSEM, ELA Grade 7, Form 1.....	167
Figure B.29. TCC, ELA Grade 8, Form 1	168
Figure B.30. TIF and CSEM, ELA Grade 8, Form 1.....	168
Figure B.31. TCC, ELA Grade 3, Form 2	169
Figure B.32. TIF and CSEM, ELA Grade 3, Form 2.....	169
Figure B.33. TCC, ELA Grade 4, Form 2	170
Figure B.34. TIF and CSEM, ELA Grade 4, Form 2.....	170
Figure B.35. TCC, ELA Grade 5, Form 2	171
Figure B.36. TIF and CSEM, ELA Grade 5, Form 2.....	171
Figure B.37. TCC, ELA Grade 6, Form 2	172
Figure B.38. TIF and CSEM, ELA Grade 6, Form 2.....	172
Figure B.39. TCC, ELA Grade 7, Form 2	173
Figure B.40. TIF and CSEM, ELA Grade 7, Form 2.....	173
Figure B.41. TCC, ELA Grade 8, Form 2	174
Figure B.42. TIF and CSEM, ELA Grade 8, Form 2.....	174
Figure B.43. TCC, Mathematics Grade 3	175
Figure B.44. TIF and CSEM, Mathematics Grade 3	175
Figure B.45. TCC, Mathematics Grade 4	176
Figure B.46. TIF and CSEM, Mathematics Grade 4	176
Figure B.47. TCC, Mathematics Grade 5	177
Figure B.48. TIF and CSEM, Mathematics Grade 5	177
Figure B.49. TCC, Mathematics Grade 6	178
Figure B.50. TIF and CSEM, Mathematics Grade 6	178
Figure B.51. TCC, Mathematics Grade 7	179
Figure B.52. TIF and CSEM, Mathematics Grade 7	179
Figure B.53. TCC, Mathematics Grade 8	180
Figure B.54. TIF and CSEM, Mathematics Grade 8	180
Figure B.55. Scree Plot, ELA Grade 3, Form 1	181
Figure B.56. Scree Plot, ELA Grade 4, Form 1	181
Figure B.57. Scree Plot, ELA Grade 5, Form 1	182
Figure B.58. Scree Plot, ELA Grade 6, Form 1	182
Figure B.59. Scree Plot, ELA Grade 7, Form 1	183
Figure B.60. Scree Plot, ELA Grade 8, Form 1	183
Figure B.61. Scree Plot, ELA Grade 3, Form 2.....	184
Figure B.62. Scree Plot, ELA Grade 4, Form 2.....	184
Figure B.63. Scree Plot, ELA Grade 5, Form 2.....	185
Figure B.64. Scree Plot, ELA Grade 6, Form 2.....	185
Figure B.65. Scree Plot, ELA Grade 7, Form 2.....	186
Figure B.66. Scree Plot, ELA Grade 8, Form 2.....	186
Figure B.67. Scree Plot, Mathematics Grade 3.....	187
Figure B.68. Scree Plot, Mathematics Grade 4.....	187
Figure B.69. Scree Plot, Mathematics Grade 5.....	188
Figure B.70. Scree Plot, Mathematics Grade 6.....	188
Figure B.71. Scree Plot, Mathematics Grade 7.....	189
Figure B.72. Scree Plot, Mathematics Grade 8.....	189

Figure C.1. Total Scale Score Distribution, ELA Grade 3	202
Figure C.2. Total Scale Score Distribution, ELA Grade 4	202
Figure C.3. Total Scale Score Distribution, ELA Grade 5	203
Figure C.4. Total Scale Score Distribution, ELA Grade 6	203
Figure C.5. Total Scale Score Distribution, ELA Grade 7	204
Figure C.6. Total Scale Score Distribution, ELA Grade 8	204
Figure C.7. Total Scale Score Distribution, Mathematics Grade 3	205
Figure C.8. Total Scale Score Distribution, Mathematics Grade 4	205
Figure C.9. Total Scale Score Distribution, Mathematics Grade 5	206
Figure C.10. Total Scale Score Distribution, Mathematics Grade 6	206
Figure C.11. Total Scale Score Distribution, Mathematics Grade 7	207
Figure C.12. Total Scale Score Distribution, Mathematics Grade 8	207
Figure C.13. Trend in % of Students On-Grade, ELA Grade 3	208
Figure C.14. Trend in % of Students On-Grade, ELA Grade 4	208
Figure C.15. Trend in % of Students On-Grade, ELA Grade 5	209
Figure C.16. Trend in % of Students On-Grade, ELA Grade 6	209
Figure C.17. Trend in % of Students On-Grade, ELA Grade 7	210
Figure C.18. Trend in % of Students On-Grade, ELA Grade 8	210
Figure C.19. Trend in % of Students On-Grade, Mathematics Grade 3	211
Figure C.20. Trend in % of Students On-Grade, Mathematics Grade 4	211
Figure C.21. Trend in % of Students On-Grade, Mathematics Grade 5	212
Figure C.22. Trend in % of Students On-Grade, Mathematics Grade 6	212
Figure C.23. Trend in % of Students On-Grade, Mathematics Grade 7	213
Figure C.24. Trend in % of Students On-Grade, Mathematics Grade 8	213

Chapter 1: INTRODUCTION

This technical report documents evidence of reliability and validity for the spring 2025 administration of the Arizona 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 test scores. The evidence includes descriptions of the test design, development, and administration procedures; student test results; and psychometric analyses, including calibration, equating, and scaling to ensure comparable test results across different test forms and administrations. The report concludes with a synthesis of the evidence to support the validity argument for the AASA assessments.

1.1. Assessment Overview

AASA is the statewide summative achievement test administered each spring to Arizona students in ELA and mathematics in grades 3–8. It is a criterion-referenced assessment aligned to the Arizona Academic Standards, designed to promote increasingly higher academic outcomes for students to prepare them for a broader array of post-secondary outcomes.

AASA is administered online as a computer-based test (CBT) or as a paper-based test (PBT). The needs of the student are also addressed through other supports, including accessibility features built into the online platform and accommodations such as assistive technology, a scribe, and/or sign language. The ELA assessment includes 42 operational items consisting of multiple-choice and technology-enhanced items, along with an open-response writing prompt. Grade 3 ELA also includes two operational short Oral Reading Fluency (ORF) passages. Mathematics consists of 53–55 operational multiple-choice and technology-enhanced items. Field test items are embedded on each assessment that do not count toward students' scores.

Student performance is reported as an overall scale score and an accompanying performance level classification. The four performance levels are *Minimally Proficient*, *Partially Proficient*, *Proficient*, and *Highly Proficient*. Student performance on reporting categories is reported as one of three levels of mastery: *Below Mastery*, *At/Near Mastery*, or *Above Mastery*. The primary intended score interpretation of the assessment is that AASA test scores provide reliable and valid information about important knowledge and skills that students are attaining in grade-level literacy and numeracy. 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.

1.2. Background

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. In December 2016, the Board adopted the current Arizona Academic 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. Beginning in 2019–2020, AzMERIT was renamed AzM2. Beginning in 2021–2022, AzM2 was renamed to AASA, which 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. ORF field test items were also embedded on the grade 3 operational AASA test beginning 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. On September 2024, the Board approved the operational use of the ORF test items on the AASA ELA grade 3 assessment. As such, the grade 3 ORF test unit became an operational test unit in spring 2025 and is included in each grade 3 student’s ELA score and in the Move on When Reading (MOWR) indicator.

1.3. Testing Requirements

AASA is an accountability assessment developed to fulfill requirements of both state and federal laws (State Law ARS 15-741; Federal Law: 34 CFR 200.2 *Participation in Assessments*) that mandate 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.

Arizona’s 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/>.

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 AASA assessments in 2024–2025.

Arizona educators have several opportunities to participate in meetings and provide feedback on assets developed for field testing. The content and bias and sensitivity community reviews enable community members, including past and present Arizona educators, to evaluate the ELA passages and items for content, standards alignment, grade-level appropriateness, and bias and sensitivity. These meetings represent a continuation of stakeholder involvement in the development process. Arizona educators are also involved in the development of the performance level descriptors (PLDs) and the establishment of cut scores during the standard setting meetings as needed.

Table 1.1. Schedule of Major Events

Event	Date(s)
Educator ELA Passage Review Committee	February 8, 2024
Community and Parent ELA Passage Review Committee	February 9, 2024
Educator Item Review Committee	June 10–13, 2024
Community and Parent Item Review	June 20–21, 2024
Technical Advisory Committee (TAC) Meeting	September 3, 2024
Technical Advisory Committee (TAC) Meeting	April 7, 2025
Administration Training	November 18, 2024 (live date)
Spring 2025 CBT Administration Window (ELA Reading and Math)	March 31 – April 25, 2025
Spring 2025 PBT Administration Window	March 31 – April 9, 2025
Spring 2025 CBT ELA Writing and Grade 3 ORF Administration Window	March 31 – April 11, 2025
Release of Grade 3 Electronic Score Reports	May 15, 2025
Release of Grades 4–8 Electronic Score Reports	May 23, 2025
Release of Grades 3–8 Paper Reports to Districts	June 12, 2025
Data Review Committee	June 16–18, 2025

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). This transparency into the intended construct and rationale behind the assessments supports the valid interpretations of the test scores and alignment with Arizona’s standards-based educational system.

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 were adhered to throughout the item and test creation process to accommodate the needs and abilities of all learners. AASA was available to be administered in online settings including group, small group, or one-on-one settings. It was also available in appropriate accommodations including American Sign Language (ASL), Braille, large print, or Special Paper Version (SPV). An Accommodated-CBT format (ELA grade 3 only) was also available.

2.1. Academic Content Standards

The Arizona Academic Standards, available online at <https://www.azed.gov/standards-practices>, define the knowledge, understanding, and skills that need to be taught and learned so all students are ready to succeed in 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 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

The AASA item specifications, available online at <https://www.azed.gov/assessment/aasa>, guide the item development process by defining the content limits, model tasks, and response types for a specific standard. During the item development and review processes, items are compared to the item specifications to ensure their alignment to the standard, grade-level appropriateness, and adherence to the content limits. The item specifications were developed by Pearson and ADE content experts using a vertical alignment for each standard, where 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 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 content and structure of the test. 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 – 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 online at <https://www.azed.gov/assessment/aasa>, while more detailed blueprints are used internally by ADE and Pearson.

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. Item Types

The AASA assessments include traditional multiple-choice items and technology-enhanced items (TEIs), as shown in Table 2.5. 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 2.5. 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.

Item Type	Description
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.

2.5. Cognitive Complexity

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 2.6 presents a description of the DOK levels as provided in the item specifications, and Table 2.7 presents the percentage of points by DOK level as provided in the blueprints.

Table 2.6. 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 2.7. 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

2.6. Test Structure

Table 2.8 and Table 2.9 present the test structure of the AASA assessments that consist of the following test units:

- ELA 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 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 Structure—ELA

Grade	#Forms	#Passages	Overall			#Items by Test Unit											
			#Items			Writing			Reading/Language Test Unit 1			Reading/Language Test Unit 2			Oral Reading Fluency (ORF)		
			OP	FT	Total	OP	FT	Total	OP	FT	Total	OP	FT	Total	OP	FT	Total
3	18	8	44	8	52	1	—	1	16	7	23	25	—	25	2	1	3
4	18	8	42	7	49	1	—	1	16	7	23	25	—	25	—	—	—
5	18	8	42	7	49	1	—	1	16	7	23	25	—	25	—	—	—
6	18	8	42	7	49	1	—	1	16	7	23	25	—	25	—	—	—
7	18	8	42	7	49	1	—	1	16	7	23	25	—	25	—	—	—
8	18	8	42	7	49	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 Structure—Mathematics

Grade	#Forms	#Items								
		Overall			Test Unit 1			Test Unit 2		
		Total	OP	FT	Total	OP	FT	Total	OP	FT
3	8	53	45	8	27	23	4	26	22	4
4	8	53	45	8	27	23	4	26	22	4
5	8	53	45	8	27	23	4	26	22	4
6	8	55	47	8	27	23	4	28	24	4
7	8	55	47	8	27	23	4	28	24	4
8	8	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 three parts for all grade levels: Writing and Reading/Language Parts 1 and 2. Writing consists of one writing prompt, which is an extended text/essay response. The Reading/Language is split into two units given its length and includes both reading and language items in both units. Grade 3 also includes an ORF test unit. The ELA passages represent a variety of genres and topics in the Reading/Language units. Informational texts are developed from multiple subjects (e.g., history, science, technical subjects). Literary texts are authentic pieces from multiple genres (e.g., stories, poetry, drama). The ratio of informational to literary texts increases with grade, with a greater percentage of informational texts in the upper grades. Both single passages and passage sets are used, in which students synthesize information across texts. The number of items associated with each passage or passage set varies.

2.6.1.1. Extended Writing Items

The ELA assessment is designed to reflect the importance of using evidence and reading complex texts as outlined in the Arizona Academic Standards. It includes extended writing items that provide meaningful contexts in which to construct responses. Each writing prompt presents students with various stimuli (at least 2–3 per task) in the form of research articles, charts and graphs that serve as a springboard for an informed piece of writing. Students use the stimuli 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 one informative/explanatory or opinion/argumentative writing prompt. While each student will only see one prompt type, both types are administered operationally at every grade level each year. Informative/explanatory writing focuses 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. Opinion (grades 3–5) and argumentative (grades 6–8) 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 and are followed by a prompt that asks students to write an opinion or argumentative essay. The students must synthesize information across the passages and cite details to support the ideas they present.

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

The ELA ORF items were field tested in spring 2022, 2023, and 2024 and did not count toward students' scores. The first operational administration of the ORF items occurred in spring 2025, with the items contributing to students' test scores. The ELA ORF test unit consists of two operational items and one field test item (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 two operational 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. Grade 3 students who take the American Sign Language (ASL) or Braille form do not take the ORF test unit.

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 2.7.

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.

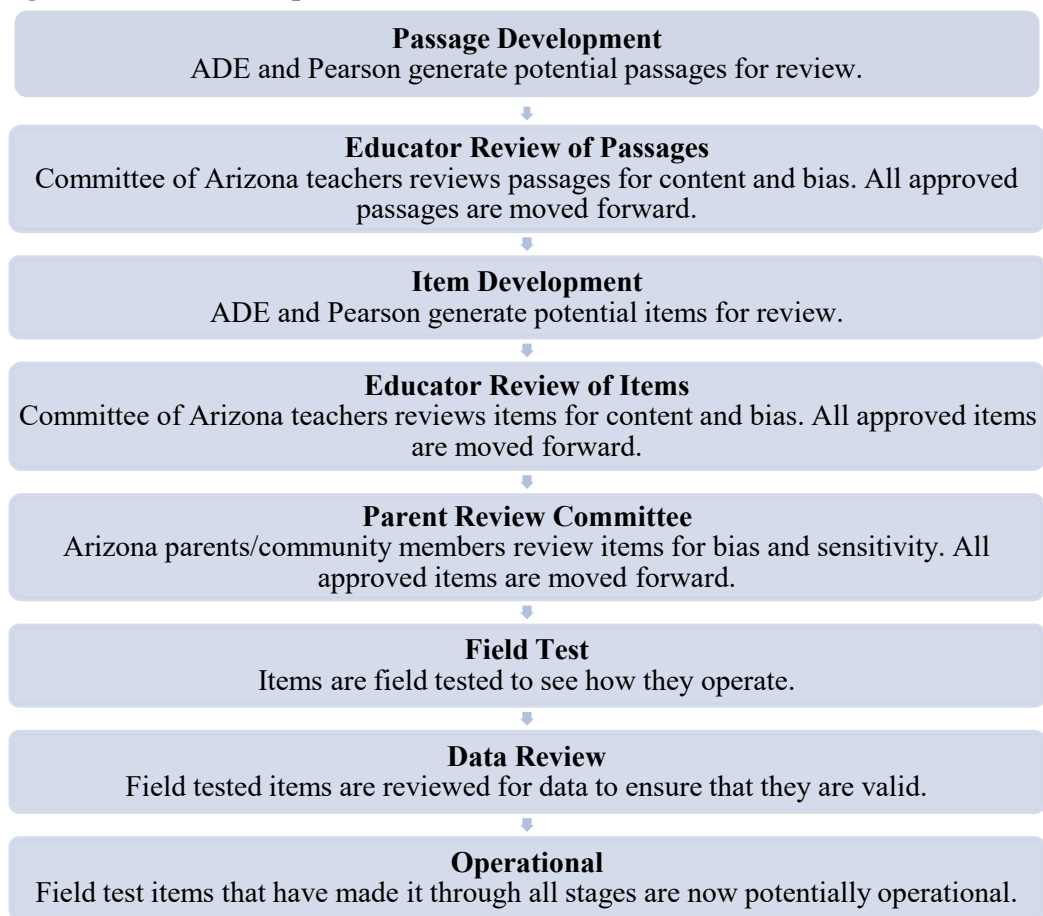
Calculators are not allowed for the mathematics assessments in grades 3–6. For the grades 7 and 8 assessments, where calculator use is allowed for some item types, the items are grouped into two units administered separately to students: calculator and no calculator. The construct of the items dictates 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 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 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 and test development process 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	140
ELA 4	138
ELA 5	126
ELA 6	140
ELA 7	138
ELA 8	138

Assessment	#Items for FT
Mathematics 3	45
Mathematics 4	45
Mathematics 5	45
Mathematics 6	45
Mathematics 7	45
Mathematics 8	45

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

Assessment	Lexile Range	Word Count Range	ORF Lexile Range	ORF Word Count
ELA 3	420–820	100–700	600–750	250–400
ELA 4	740–1010	100–900	–	–
ELA 5	740–1010	200–1,000	–	–
ELA 6	925–1185	200–1,100	–	–
ELA 7	925–1185	300–1,100	–	–
ELA 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 (<i>setting, climax, rising and falling action</i>) that demonstrate clear plot development and how they are similar and/or different across multiple texts. (<i>Paired text only.</i>)
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 (<i>e.g., metaphor, onomatopoeia, flashback, foreshadowing, voice, irony, symbolism</i>).
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 copy editors 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 specifications 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 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 prior administrations, and Pearson assessment specialists updated the ABBI item status to indicate eligibility for operational or field test selection based on the data review results. Item statistics included classical difficulty (p -value) and IRT difficulty (Rasch), item discrimination (point-biserial correlation), 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 main goal in item selection was adhering to the blueprints. Additional criteria included item positioning and both content and statistical considerations. For each test form, a Pearson assessment specialist initially pulled operational items using the tools embedded in ABBI to verify blueprint alignment and acceptable statistics according to the test construction specifications. Another assessment specialist reviewed the form, identifying issues such as clueing. After the issues were resolved, Pearson and ADE psychometricians reviewed the forms based on statistical considerations. This process repeated until each form met psychometric approval.

The forms were also reviewed by the ADE content and psychometrics teams who worked with Pearson throughout the process, including final item selection for each form (including the paper, ASL, 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 8 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

Following the spring 2025 administration, item analysis was conducted to generate statistics for the field-tested items that were flagged based on the criteria in Table 3.3. During data review, committee members reviewed the flagged items and their statistics to determine their eligibility for the operational item pool. Two different committees met for data review. One 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 looked 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. Decisions for all items reviewed were documented in ABBI.

Table 3.3. Item Statistical Flagging Criteria

Statistic	Criterion	Possible Indication
<i>P</i> -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 2025 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	29	0	13
ELA 4	27	0	13
ELA 5	24	0	13
ELA 6	24	0	13
ELA 7	32	0	15
ELA 8	26	0	16
Mathematics 3	12	8	2
Mathematics 4	14	5	2
Mathematics 5	14	5	3
Mathematics 6	13	11	5
Mathematics 7	17	5	5
Mathematics 8	15	8	7

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. Grade 3 Braille does not include ORF. In 2025, contracted and uncontracted Braille was available for the ELA Reading units in all grades. 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.

The Accommodated-CBT version was available for grade 3 ELA, with ADE approval, and was based on the online test form. This accommodated version was designed for students who are selective mute or who use an augmentative and alternate communication (AAC) device. The Pearson content team worked with ADE to identify items that would not be appropriate for this population and replaced them with equivalent items.

Each grade and content area also had one form created for ASL testers. Grade 3 ASL does not include ORF. 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 and 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, or as a PBT, 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 operational ORF test unit. Each student read three separate passages (i.e., two operational and one field test), with a time limit of one minute per passage. The ORF online test unit was administered in small groups, with no more than six students testing simultaneously in a classroom or a computer lab environment for both CBT and PBT schools.

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)	30 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 were available online 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 Arizona Learning Management System (AzLMS) 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, STCs, and TAs and provided important information from the three documents testing staff 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 all training modules and to successfully complete a final quiz at the end of each module. DTCs, STCs, and TAs must obtain a score of 80% or higher on each module's 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 each module's quiz. DTCs are required to ensure that all testing staff have viewed the applicable training modules and to track staff training completion.

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. This training was required to be completed by DTCs and STCs.
Accommodations	This training covered the test accommodations. This was required for all DTCs and STCs 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 or proctored the test. 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 or the test environment.
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. It was required for all DTCs, STCs, and Technology Coordinators (TCs).

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 guests, so no personal information needs to be provided.

There is a sample test for each grade and content area, every eligible item type is represented, and an accompanying scoring guide identifies standard and DOK alignment. ASL and Braille sample tests are also available. The portal and scoring guides are both available on the 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	26
4	24	1	–	26
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 with 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 who are blind or have low vision may use a talking calculator for Unit 1 of the grades 7 or 8 mathematics test in a one-on-one testing environment.
Accommodated – Computer-Based Test (grade 3 ELA only)	Requires the use of a different test form that must be indicated in PAN. This form must be requested using the Additional Accommodations online request form and must be approved by ADE.
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 and multi-select 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. An adult scribe may not be used for the ELA writing test.

Accommodation	Description
American Sign Language (ASL)	ASL requires the use of a different test form that must be indicated in 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 SPV test online request form. Requires adult transcription: An adult must transfer the student's responses exactly as written into the TestNav system.
Large print test booklet	Large Print tests must be requested using the 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 responses exactly as written into the TestNav system.
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 responses 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 <i>Test Administration Directions</i> 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 <i>Test Administration Directions</i> manual are permitted. No test content or directions embedded within the test may be translated.
Translation dictionary	During testing, English learner 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, CCTV, dome magnifiers, enlarged monitors, enlarged keyboards, and special lighting
- Using different contrast settings or color overlays
- Using devices that allow the student to hear the test directions: hearing aids, cochlear implants, 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. Lunch may be brought to the student if necessary.

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	Six alternate mouse pointers are 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	Student can electronically cover and reveal individual answer choices.
Answer Eliminator	Student can cross out answer options for multiple-choice and multi-select items.
Area Boundaries	Student can click anywhere on the selected response text or button for multiple choice items.
Bookmark for Review	Student can mark an item for review so that it can be easily found later.

Universal Test Tool	Description
Contrast	Student can 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 and pale green background with gray text.
Expand/Collapse Passage	Student can expand a passage for easier readability. Expanded passages can also be collapsed.
Highlighter	Student can highlight text in a passage or item.
Key Board Navigation	Can be used in supported item types using the following keyboard commands: • Move forward: Tab • Move backwards: Shift and Tab • Select buttons: Enter or Spacebar • Navigate and select radio buttons: Arrow Up or Arrow Down • Select and unselect boxes: Spacebar • Increase / Decrease screen size: Ctrl + and Ctrl - • Copy: Ctrl + C • Paste: Ctrl + V
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	Student 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.
Notes/Comments	Student can 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	Student can review the test before submitting it.
Student Readiness Tool	Video lessons and interactive practice questions are embedded by grade and content area to provide students a hands-on experience with the tools, questions, and test support in the test environment.
System Settings	Adjust audio (volume) during the test.
Text-to-Speech (TTS)	Student access TTS for content of writing, mathematics, and science.
Writing Tools	Student can use editing tools (cut, copy, and paste) and basic text formatting tools (bold, underline, and italic) for extended response items.
Zoom In/Zoom Out	Student can enlarge the font and images in the test up to 500%. Zoom may be set in PAN up to 200% prior to testing, then the student may use the keys Ctrl + to Zoom up to 500% in TestNav. Undo zoom in and return the font and images in the test to original size using Ctrl -.

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 PAN 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 2024–2025 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 2024-2025.
 - 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.
 - 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](#) approved by the State Board of Education.
 - i. I shall not provide prohibited or inappropriate resources or content support 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 2024-2025 must also agree to the following conditions to ensure the correct administration of the tests.

3. I shall participate in and complete training activities prior to administering the tests.
4. I shall review the Test Administration Directions prior to administering the test.
5. I shall follow all instructions in the 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 or the testing environment, 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

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. In addition, all grade 3 ORF items were scored by Versant, Pearson's automated scoring engine for spoken responses, with low-confidence responses sent to human scorers, along with responses flagged with zero words spoken correctly that required auditing by human scorers. 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 MC, MS, and TE items were machine-scored, with an attemptedness rule that a student needed to answer one item in each operational unit.

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. Beginning in 2025, the low-confidence and audited grade 3 ORF items (part of ELA Reading) were also added to the hand-scored items. Writing was scored via a distributed scoring model (i.e., scorers were trained in a self-paced model), while 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 ELA Reading and Mathematics open-ended items. Writing established training materials that could be inserted into modules for self-paced training, while 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 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.

5.1.2.3. Grade 3 ORF

Unlike reading and mathematics, scoring of the grade 3 ORF items did not require rangefinding or ADE approval of the training sets. The scoring director of the ORF team selected one of the four ORF items and built full training materials (anchor, practice, and qualification set) to be used as a “prototype” item. The scoring director began content training on that item by reviewing the rubric, the passage, and the anchor set. The team then took and discussed the practice set to test their knowledge of applying the rubric to student responses. The team then took a qualification set and were required to meet specific criteria to score the project. Qualified scorers began scoring “live” student responses. For the remaining items, the scoring director reviewed the passage and item-specific anchor before scoring of that item occurred. ORF items differ from math and reading items because only a first score was applied. No Inter-Rater Reliability (IRR) statistics were created. Validity responses were presented.

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 requalification set consisting of 10 responses. If the scorer fails to pass the requalification 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. IEA 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 2025, 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.

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)

Two distinct exact agreement metrics were calculated to evaluate inter-rater reliability. Overall exact agreement represents the percentage of double-scored responses where both raters assigned identical scores across all score points. In contrast, conditional exact agreement by score point is computed separately for each score point on the rubric and subsequently averaged across all score points. The conditional metric provides a more comprehensive assessment of scoring consistency by ensuring equal weight is given to agreement at each score point, regardless of the distribution of student responses. This approach prevents potential masking of inconsistent scoring at less frequently assigned score points. For example, if most responses receive a score of 1 and raters demonstrate high agreement on this score point, overall exact agreement may appear satisfactory even when agreement is poor at score points 2 and 3. The conditional exact agreement metric addresses this limitation by providing equal consideration to scoring consistency across the entire rubric range.

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

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.

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. **Versant Scoring of Grade 3 ORF Items**

The ORF items are short passages that students read aloud to measure their 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. They are included on the grade 3 ELA assessment only, with each student receiving 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 for low, 650L – 700L for medium, and 700L – 750L for high difficulty). Each student receives a combination of fiction and nonfiction genres.

The grade 3 ORF items are scored by Versant, Pearson's automated scoring engine for spoken responses that scores the content (i.e., what was read aloud by students during the assessment) using a speech processing method built to handle different rhythms and varied pronunciations. The system recognizes the words that were uttered and aligns the speech signal to the hypothesized response by locating the part of the signal containing the relevant segments, syllables, and words, allowing the system to assign independent scores based on the content of what is said and the way it is said. Thus, the system can generate scores based on the words used in the spoken responses (the "content").

5.3.1. Training

Prior to the initial spring 2022 field testing, the Pearson content team developed 12 ORF items for the grade 3 ELA assessment with an equal number of low, medium, or high expected difficulty items based on their Lexile ranges (i.e., four items of each). Nine of the ORF items were intended for future operational administrations, while three were developed for the sample test. The Versant engine was then trained to score the ORF items for the first time in spring 2022, during which 218,520 field test responses were collected across the nine ORF items. For the Versant engine training, approximately 600 responses were randomly drawn for each of the nine ORF items, totaling more than 5,800 responses. The selected ORF responses were each transcribed by a group of trained individuals. The transcribed responses were then used to develop speech recognition and scoring models in the Versant system.

As part of developing the capabilities that enable the ORF items to be machine-scored during operational testing, automated speech recognition and scoring models were created. Each item received its own scoring model that counts the number of words in the passage accurately read aloud by the student. These three independently calculated WCPM scores from the low, medium, and high expected difficulty passages for each student form the basis of the ORF measures described in this report. After the engine training was completed, all responses for the nine ORF items were scored by the trained engine, resulting in WCPM raw scores for students.

5.3.2. ORF Score Points

To convert the WCPM raw score performance to a three-point scale (i.e., 0, 1, and 2 points), the threshold for 2 points is 110 WCPM. This was determined based on the national norms developed by Hasbrouck and Tindal (2017) from data compiled across three assessments. The median WCPM in the spring for grade 3 was 112, and the average WCPM measured by aimswebPlus© based on the norm sample in the spring during the 2013–2014 school year for grade 3 was 122.83 (aimswebPlus, 2017). Taking these studies into consideration, ADE decided to use 110 WCPM as the threshold for 2 points.

The threshold for 1 point is 20 WCPM, which was determined by identifying the 15th percentile of WCPM for grade 3 Arizona students in spring 2022. Students below the 15th percentile in WCPM would receive a score of 0.

This results in a maximum of 4 raw score points for the two ORF items on the grade 3 ELA assessment that is worth a total of around 59 points (i.e., the ORF items account for approximately 7% of the total grade 3 ELA raw score).

Chapter 6: 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 were 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 were 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 6.1 and Figure 6.2 present sample reports.

- District-level
 - Confidential Roster Report with Summary (school-level¹, student roster by grade)
 - District Confidential Roster Report with Summary (district-level, student roster by grade and content area)
 - Student Data File
 - Summary Cluster Performance Reports
- School-level
 - Confidential Student Score Report (individual student report)
 - Family Report Guide
 - Informe del Estudiante (individual student report in Spanish)
 - Confidential Roster Report with Summary (school-level, student roster by grade and content area).
 - Summary Cluster Performance Reports

In addition, the following AASA reports were made accessible through the PAN system at <https://az.pearsonaccessnext.com>. The On Demand Reports were available for download, while the Performance Level Dashboards and Historical Student Data reports were available to print from the PAN system.

- On Demand Reports
- Performance Level Dashboard
- Historical Student Data

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 and Confidential Roster Reports, can be disaggregated into reporting groups and testing groups if they were set up by the school during the specified timeframe.

¹ Districts received their own copy of the school-level Confidential Roster Report with Summary. For example, if a district had five schools, they had a copy of all five rosters in one PDF file.

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 school and shipped to participating districts. The MOWR indicator was included in the grade 3 Confidential Student Score Reports. 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 Confidential Student 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* was also available for schools and districts to provide to families and includes the steps that should be followed to access their student's information on the Parent Portal.

Figure 6.1. Sample Reports—Confidential Student Score Report



ENGLISH LANGUAGE ARTS (ELA)
CONFIDENTIAL STUDENT SCORE REPORT
GRADE 3 - SPRING 20XX



ARIZONA'S
ACADEMIC
STANDARDS
ASSESSMENT

FIRSTNAME M. LASTNAME

SSID: 9999999999

DOB: mm/dd/yyyy

SCHOOL: SCHOOL NAME (9999999)

DISTRICT: DISTRICT NAME (9999999)

About This Assessment

FIRSTNAME took the AASA Grade 3 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 3 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.

9999

PASSING

Level 4—Highly Proficient:
Advanced understanding, highly likely to be ready

Level 3—Proficient:
Strong understanding, likely to be ready

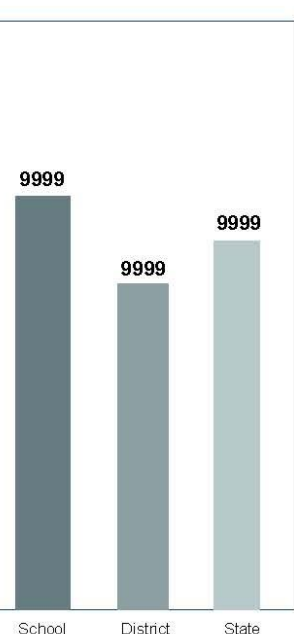
Level 2—Partially Proficient:
Partial understanding, likely to need support to be ready

Level 1—Minimally Proficient:
Minimal understanding, highly likely to need support to be ready

Has FIRSTNAME met the Move On When Reading requirement?

YES

This determination is made based on the Reading component of the ELA assessment.



Entity	Score
Your Student	9999
School	9999
District	9999
State	9999

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 Z0000000-0000-00-000-00000000

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 find the main idea and the supporting details of a text. They connect events, ideas, steps, sentences, paragraphs, and illustrations to one another. They find similarities and differences between two texts on the same topic.

What do these results mean?

Your student almost always finds connections between concepts, ideas, or events; uses the text and pictures to make conclusions to ask and answer questions; and finds the similarities and differences between important ideas and key details in two texts on the same topic.

Reading for Literature



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

What was assessed?

Students ask and answer questions about a text. They tell how characters and their actions affect a story. They explain how pictures help tell a story. They read two texts by one author and tell the similarities and differences. They find the central message of a story.

What do these results mean?

Your student can often find similarities and differences between the settings or plots of stories written by the same author; tell how one part of a story affects another part; use key details to retell a story and find the main idea; and tell the point of view in a story.

Writing and Language



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

What was assessed?

Students write to give information or state opinions. They write on a topic giving supporting details or facts. They use correct capitalization, punctuation, and spelling. They use sentences, a glossary, or a dictionary to figure out the meaning of new words.

What do these results mean?

Your student may have trouble organizing writing for a purpose (like to give information or give opinions); using clues in a text to understand the meaning of new words; spelling commonly used words correctly; and writing simple sentences with correct capitalization and punctuation.

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

Your student earned 3 out of 4 possible points. Your student's essay mainly stays on topic. The opinion is clearly stated and mostly focused. Context supporting the opinion fits the purpose. The response is organized and has few mistakes. There is some variety of transitions used. There is a clear progression of ideas within the essay. There is a clear beginning and end.

Evidence & Elaboration

Your student earned 2 out of 4 possible points. Your student's essay includes details, facts, and sources that somewhat support its opinion. This evidence is unevenly integrated into the response. The words used are sometimes inappropriate for audience and purpose.

Conventions & Editing

Your student earned 2 out of 2 possible points. Your student's essay shows a strong understanding of sentence structure and language conventions. There are few mistakes in punctuation, capitalization, and spelling present in the response.

FIRSTNAME's ELA Oral Reading Fluency

The information below represents your student's Oral Reading Fluency on two passages as measured by the number of Words Correct Per Minute (WCPM).

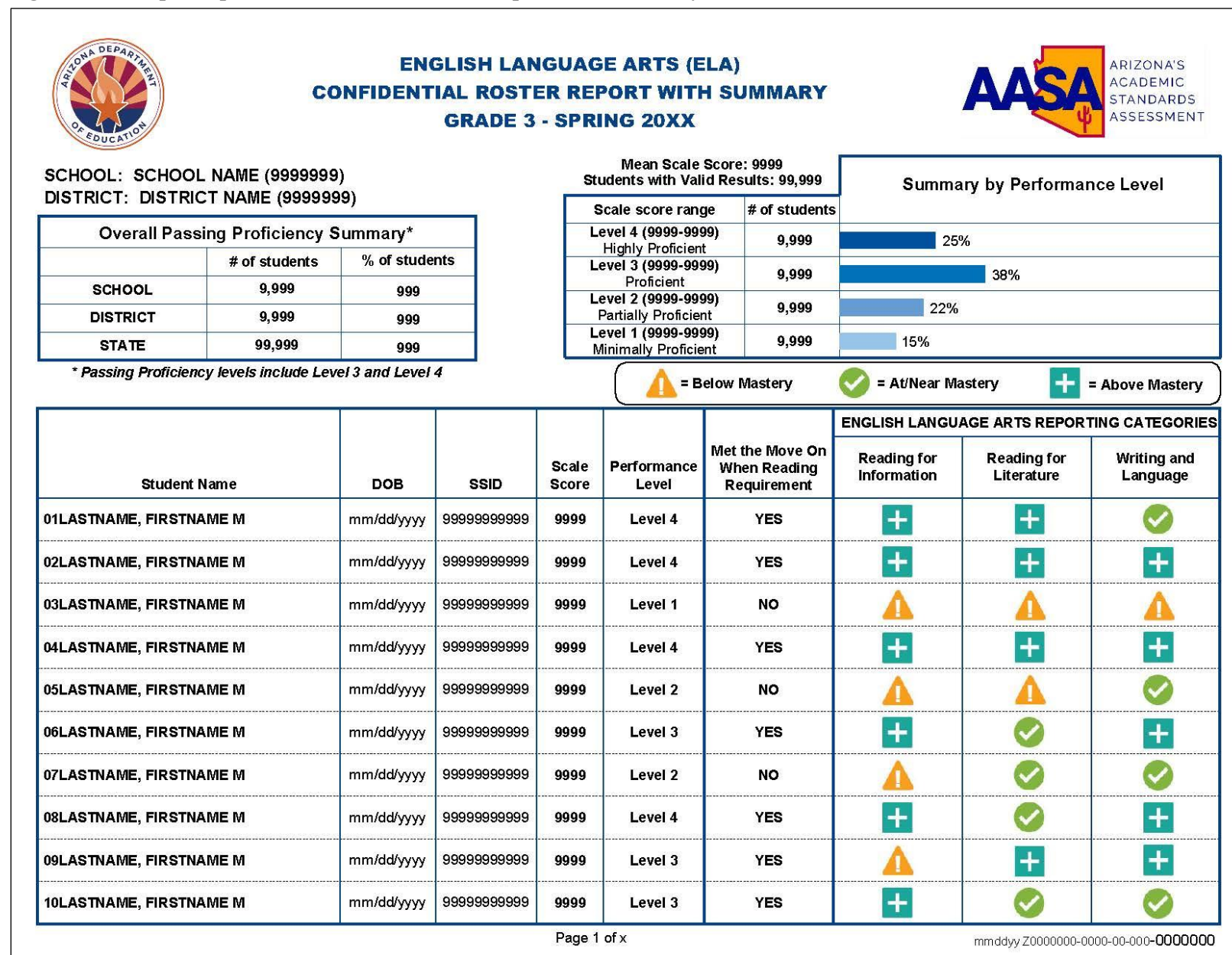
Passage 1: FIRSTNAME received 999 WCPM compared to the state average of 999 WCPM* on this passage.

Passage 2: FIRSTNAME received 999 WCPM compared to the state average of 999 WCPM* on this passage.

*The state average WCPM provided is based on the prior year's data.

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.

Figure 6.2. Sample Reports—Confidential Roster Report with Summary



Chapter 7: CLASSICAL ITEM ANALYSIS

This chapter presents classical statistics for the data used for calibration, equating, and scaling of the spring 2025 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 8 online forms total.

7.1. Data

The classical item analysis was conducted based on the calibration samples described in Section 8.1. Table 7.1 and Table 7.2 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 7.1. Number of Students in the Calibration Sample by Subgroup—ELA

Subgroup	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
All	76,004	74,857	78,081	77,352	77,204	78,671
Male	38,448	37,837	39,593	39,215	39,342	39,948
Female	37,556	37,020	38,488	38,137	37,862	38,723
Hispanic	38,059	37,460	38,989	38,345	37,920	38,592
Non-Hispanic	37,945	37,397	39,092	39,007	39,284	40,079
American Indian	4,246	3,902	4,125	4,200	4,212	4,369
Asian	2,233	2,249	2,338	2,276	2,265	2,197
Black or African American	5,669	5,635	5,832	5,856	5,652	5,799
Multi-racial	5,291	4,838	4,935	4,837	4,673	4,709
Native Hawaiian or Other Pacific Islander	490	438	477	451	452	485
White	55,984	55,581	58,100	57,320	57,502	58,469
Missing	2,091	2,214	2,274	2,412	2,448	2,643
Special Education	10,749	11,720	11,722	10,888	9,809	9,572
English Learner (EL)	10,708	9,767	9,018	8,483	8,115	7,172
Low Socioeconomic Status (SES)	40,819	39,668	40,796	39,967	38,682	38,814

Table 7.2. Number of Students in the Calibration Sample by Subgroup—Mathematics

Subgroup	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
All	77,564	75,510	78,380	77,676	77,670	79,153
Male	39,413	38,256	39,786	39,402	39,627	40,227
Female	38,151	37,254	38,594	38,274	38,043	38,926
Hispanic	38,925	37,782	39,113	38,486	38,151	38,806
Non-Hispanic	38,639	37,728	39,267	39,190	39,519	40,347
American Indian	4,340	3,931	4,163	4,223	4,237	4,402
Asian	2,284	2,262	2,342	2,280	2,281	2,215
Black or African American	5,832	5,719	5,872	5,916	5,685	5,857
Multi-racial	5,378	4,877	4,960	4,858	4,691	4,738
Native Hawaiian or Other Pacific Islander	505	443	481	456	454	496
White	57,061	56,006	58,272	57,500	57,833	58,768
Missing	2,164	2,272	2,290	2,443	2,489	2,677
Special Education	11,220	11,953	11,831	10,986	9,911	9,662
English Learner (EL)	11,067	9,900	9,061	8,549	8,197	7,232
Low Socioeconomic Status (SES)	41,779	39,994	40,982	40,097	38,883	38,998

7.2. Descriptive Statistics

Table 7.3 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 7.3. 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	37,998	46	59	28.63	11.53	2	59
ELA 4, Form 1	37,750	44	55	29.02	11.14	3	55
ELA 5, Form 1	38,951	44	55	28.08	11.29	3	55
ELA 6, Form 1	38,689	44	55	29.67	11.09	2	55
ELA 7, Form 1	38,602	44	55	29.88	10.61	2	55
ELA 8, Form 1	39,561	44	55	30.06	11.13	3	55
ELA 3, Form 2	38,006	46	59	28.67	11.64	1	58
ELA 4, Form 2	37,107	44	55	29.20	11.19	3	55
ELA 5, Form 2	39,130	44	55	28.38	11.29	2	55
ELA 6, Form 2	38,663	44	55	28.85	10.86	4	55
ELA 7, Form 2	38,602	44	55	29.86	11.05	2	55
ELA 8, Form 2	39,110	44	55	29.82	11.12	4	55
Mathematics 3	77,564	45	45	26.29	10.62	0	45
Mathematics 4	75,510	45	45	23.41	11.20	0	45
Mathematics 5	78,380	45	45	21.60	10.40	0	45
Mathematics 6	77,676	47	47	20.67	10.82	0	47
Mathematics 7	77,670	47	47	20.31	10.66	1	47
Mathematics 8	79,153	47	47	19.44	10.15	0	47

7.3. 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 7.4 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 7.4. Classical Item Analysis Summary

Assessment	#Items	Mean P -Value	Mean Point-Biserial
ELA 3, Form 1	46	0.47	0.47
ELA 4, Form 1	44	0.52	0.47
ELA 5, Form 1	44	0.51	0.49
ELA 6, Form 1	44	0.53	0.46
ELA 7, Form 1	44	0.54	0.45
ELA 8, Form 1	44	0.54	0.47
ELA 3, Form 2	46	0.47	0.47
ELA 4, Form 2	44	0.52	0.48
ELA 5, Form 2	44	0.51	0.49
ELA 6, Form 2	44	0.52	0.46
ELA 7, Form 2	44	0.54	0.46
ELA 8, Form 2	44	0.53	0.47

Assessment	#Items	Mean <i>P</i> -Value	Mean Point-Biserial
Mathematics 3	45	0.58	0.51
Mathematics 4	45	0.52	0.53
Mathematics 5	45	0.48	0.49
Mathematics 6	47	0.44	0.49
Mathematics 7	47	0.43	0.48
Mathematics 8	47	0.41	0.46

7.4. Distractor Analysis

Table 7.5 and Table 7.6 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 7.5. Distractor Analysis Summary: Point-Biserial Correlations for Correct Options

Assessment	#MC Items	Min.	P25	P50	P75	Max.
ELA 3, Form 1	30	0.30	0.39	0.43	0.46	0.57
ELA 4, Form 1	26	0.20	0.37	0.46	0.49	0.58
ELA 5, Form 1	25	0.32	0.38	0.41	0.48	0.60
ELA 6, Form 1	29	0.28	0.39	0.44	0.49	0.62
ELA 7, Form 1	30	0.26	0.35	0.45	0.49	0.58
ELA 8, Form 1	28	0.26	0.39	0.47	0.52	0.57
ELA 3, Form 2	30	0.30	0.39	0.43	0.46	0.57
ELA 4, Form 2	26	0.20	0.37	0.46	0.49	0.58
ELA 5, Form 2	25	0.32	0.38	0.41	0.48	0.60
ELA 6, Form 2	29	0.28	0.39	0.44	0.49	0.62
ELA 7, Form 2	30	0.26	0.35	0.45	0.49	0.58
ELA 8, Form 2	28	0.26	0.39	0.47	0.52	0.57
Mathematics 3	22	0.28	0.42	0.50	0.54	0.62
Mathematics 4	22	0.37	0.44	0.50	0.56	0.62
Mathematics 5	20	0.34	0.40	0.47	0.50	0.55
Mathematics 6	23	0.28	0.37	0.41	0.48	0.59
Mathematics 7	28	0.28	0.40	0.45	0.49	0.58
Mathematics 8	28	0.33	0.38	0.43	0.49	0.57

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

Table 7.6. Distractor Analysis Summary: Point-Biserial Correlations for Incorrect Options

Assessment	#MC Items	Min.	P25	P50	P75	Max.
ELA 3, Form 1	30	-0.35	-0.25	-0.20	-0.14	-0.01
ELA 4, Form 1	26	-0.36	-0.29	-0.22	-0.16	0.03
ELA 5, Form 1	25	-0.40	-0.27	-0.22	-0.15	-0.06
ELA 6, Form 1	29	-0.39	-0.27	-0.22	-0.17	0.00
ELA 7, Form 1	30	-0.36	-0.28	-0.24	-0.17	-0.03
ELA 8, Form 1	28	-0.36	-0.28	-0.24	-0.17	0.07
ELA 3, Form 2	30	-0.35	-0.25	-0.20	-0.14	-0.01
ELA 4, Form 2	26	-0.36	-0.29	-0.22	-0.16	0.03
ELA 5, Form 2	25	-0.40	-0.27	-0.22	-0.15	-0.06
ELA 6, Form 2	29	-0.39	-0.27	-0.22	-0.17	0.00
ELA 7, Form 2	30	-0.36	-0.28	-0.24	-0.17	-0.03
ELA 8, Form 2	28	-0.36	-0.28	-0.24	-0.17	0.07
Mathematics 3	22	-0.42	-0.30	-0.24	-0.17	-0.07
Mathematics 4	22	-0.40	-0.30	-0.24	-0.18	0.03
Mathematics 5	20	-0.36	-0.28	-0.22	-0.17	-0.04
Mathematics 6	23	-0.39	-0.25	-0.20	-0.11	-0.02
Mathematics 7	28	-0.37	-0.25	-0.20	-0.15	0.06
Mathematics 8	28	-0.30	-0.24	-0.19	-0.14	0.00

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 8: CALIBRATION, EQUATING, AND SCALING

This chapter describes the calibration, equating, and scaling procedures that took place for the spring 2025 AASA assessments, addressing Standards 1.10, 5.1, 5.2, 5.3, 7.2, 7.4, and 12.9 (AERA et al., 2014).

8.1. 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

8.2. 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.

For calibration, the Rasch model (Rasch, 1960) was used for 1-point items and the partial-credit model (Masters, 1982) was used for multiple-point items such as ORF items, in which the WCPM raw score performance was converted to a 3-point scale (i.e., 0, 1, and 2 points). 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 $x = 0, 1, \dots, m_i$, 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.

8.3. 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 8.1 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 8.1. IRT Statistics Summary

Assessment	#Items	Mean Rasch
ELA 3, Form 1	46	0.02
ELA 4, Form 1	44	0.26
ELA 5, Form 1	44	0.12
ELA 6, Form 1	44	0.06
ELA 7, Form 1	44	0.08
ELA 8, Form 1	44	0.00
ELA 3, Form 2	46	0.06
ELA 4, Form 2	44	0.26
ELA 5, Form 2	44	0.11
ELA 6, Form 2	44	0.09
ELA 7, Form 2	44	0.06
ELA 8, Form 2	44	0.01
Mathematics 3	45	0.06
Mathematics 4	45	0.09
Mathematics 5	45	0.10
Mathematics 6	47	-0.07
Mathematics 7	47	0.11
Mathematics 8	47	-0.14

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, while 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). Figure B.19 – Figure B.54 in Appendix B present the TCC, TIF, and CSEM curves.

8.4. Equating

The spring 2025 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 8.2 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 8.2. Summary of Anchor Items

Assessment	#Items in Initial Anchor Set	#Items Dropped from Anchor
ELA 3	16	0
ELA 4	15	2
ELA 5	14	0
ELA 6	15	0
ELA 7	14	0
ELA 8	14	1
Mathematics 3	19	1
Mathematics 4	18	0
Mathematics 5	15	0
Mathematics 6	16	1
Mathematics 7	15	0
Mathematics 8	16	0

8.5. 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:

$$\text{Scale Score} = VS_A \times \theta + VS_B$$

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.

8.6. 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.

8.6.1. 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 8.3 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 8.3. Eigenvalues from PCA

Assessment	1	2	3	4	5	6	7	8	9	10	%Var
ELA 3, Form 1	15.71	2.00	1.42	1.11	1.00	0.93	0.92	0.87	0.84	0.80	34%
ELA 4, Form 1	15.93	1.50	1.17	1.02	0.96	0.92	0.89	0.85	0.84	0.83	36%
ELA 5, Form 1	16.61	1.55	1.28	0.92	0.90	0.88	0.86	0.84	0.81	0.80	38%
ELA 6, Form 1	15.04	1.52	1.12	0.97	0.92	0.89	0.87	0.85	0.85	0.83	34%
ELA 7, Form 1	14.68	1.60	1.26	1.01	0.96	0.93	0.92	0.88	0.87	0.86	33%
ELA 8, Form 1	15.84	1.57	1.20	1.01	0.97	0.92	0.90	0.88	0.85	0.84	36%
ELA 3, Form 2	15.97	1.94	1.36	1.10	1.00	0.94	0.91	0.85	0.83	0.80	35%
ELA 4, Form 2	15.67	1.52	1.18	1.04	0.97	0.92	0.90	0.88	0.87	0.82	36%
ELA 5, Form 2	16.47	1.58	1.30	0.91	0.91	0.88	0.87	0.86	0.81	0.81	37%
ELA 6, Form 2	14.92	1.49	1.15	0.97	0.92	0.89	0.88	0.85	0.84	0.83	34%
ELA 7, Form 2	14.93	1.55	1.15	1.00	0.94	0.91	0.89	0.87	0.87	0.85	34%
ELA 8, Form 2	15.81	1.57	1.11	1.04	0.99	0.91	0.89	0.87	0.85	0.82	36%

Assessment	1	2	3	4	5	6	7	8	9	10	%Var
Mathematics 3	19.92	1.61	1.17	1.10	0.94	0.84	0.82	0.80	0.74	0.74	44%
Mathematics 4	21.15	1.81	1.19	0.94	0.91	0.83	0.77	0.73	0.70	0.69	47%
Mathematics 5	17.81	1.64	1.43	1.01	1.01	0.96	0.90	0.86	0.84	0.80	40%
Mathematics 6	19.02	1.69	1.18	1.01	0.92	0.88	0.87	0.83	0.83	0.80	40%
Mathematics 7	18.23	1.32	1.15	0.96	0.93	0.90	0.84	0.84	0.81	0.78	39%
Mathematics 8	16.55	1.69	1.14	1.03	0.99	0.94	0.90	0.87	0.85	0.81	35%

8.6.2. 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 8.4 summarizes the distribution of the residual correlations. Most residual correlations are slightly negative or slightly positive, and only six 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 8.4. Q3 Statistics

Assessment	#Item Pairs	Mean	SD	Min.	P10	P25	P50	P75	P90	Max.	#Items Exceeding 0.70
ELA 3	1,275	-0.02	0.05	-0.11	-0.06	-0.04	-0.02	0.00	0.02	0.68	0
ELA 4	1,081	-0.02	0.05	-0.10	-0.06	-0.04	-0.02	-0.01	0.01	0.72	1
ELA 5	1,081	-0.02	0.04	-0.11	-0.06	-0.04	-0.02	-0.01	0.01	0.71	1
ELA 6	1,081	-0.02	0.05	-0.09	-0.05	-0.04	-0.02	-0.01	0.00	0.79	2
ELA 7	1,081	-0.02	0.05	-0.11	-0.06	-0.04	-0.02	-0.01	0.01	0.85	2
ELA 8	1,081	-0.02	0.04	-0.11	-0.05	-0.04	-0.03	-0.01	0.01	0.62	0
Mathematics 3	990	-0.02	0.03	-0.10	-0.06	-0.04	-0.03	-0.01	0.02	0.18	0
Mathematics 4	990	-0.02	0.03	-0.10	-0.05	-0.04	-0.03	-0.01	0.01	0.18	0
Mathematics 5	990	-0.02	0.03	-0.12	-0.05	-0.04	-0.03	-0.01	0.00	0.25	0
Mathematics 6	1,081	-0.02	0.03	-0.11	-0.06	-0.04	-0.02	-0.01	0.02	0.17	0
Mathematics 7	1,081	-0.02	0.02	-0.09	-0.05	-0.04	-0.02	-0.01	0.00	0.10	0
Mathematics 8	1,081	-0.02	0.03	-0.10	-0.05	-0.04	-0.02	-0.01	0.01	0.27	0

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

8.6.3. 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 8.5 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 8.5. IRT Item Fit Summary Statistics

Assessment	#Items	#Flagged Items by Infit	%Flagged
ELA 3	51	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 9: TEST RESULTS

This chapter presents the test results of the spring 2025 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).

Both ELA and mathematics have their own scale score ranges. Students receive a scale score in each content area, and student performance is reported as one of four performance levels: *Minimally Proficient*, *Partially Proficient*, *Proficient*, and *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 is clearly below *Proficient*. Students who score *At/Near Mastery* demonstrate performance in the reporting category that is exactly at or immediately above/below *Proficient*. Students who score *Above Mastery* demonstrate performance in the reporting category that is clearly *Proficient* or higher.

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 9.1 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 1.3 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 9.2 and Table 9.3 present the percentage of students in each level of mastery by reporting category.
- Table 9.4 and Table 9.5 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 administration).
- Table 9.6 and Table 9.7 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.
- Tables C.1-C.12 present the test results by demographics.
- Tables C.13 and C.14 present the state mean raw score points by reporting category and cluster.
- Figures C.1-C.12 present histograms of the scale score distribution for the total score.
- Figures C.13-C.24 present line graphs of trend in % of students on-grade (i.e., *Proficient* and *Highly Proficient* performance levels combined).

Table 9.1. Overall Test Results by Year

Assessment	Year	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4	%MOWR Met
ELA 3	2025	79,299	2497.40	34.73	51.7	12.5	22.8	13.1	96.4
	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	2025	78,164	2517.61	35.84	41.7	12.3	31.5	14.5	—
	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	2025	81,330	2530.45	36.80	39.2	23.1	26.3	11.3	—
	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	2025	81,086	2542.75	34.88	38.2	20.8	33.8	7.2	—
	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	—
ELA 7	2025	80,645	2552.37	33.64	39.1	18.5	32.9	9.4	—
	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	2025	82,002	2559.85	35.32	40.4	19.4	29.0	11.3	—
	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	2025	80,802	3517.85	43.94	29.6	29.2	27.6	13.6	—
	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	2025	78,836	3547.84	46.54	36.8	23.6	25.9	13.7	—
	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	2025	81,638	3583.69	40.17	33.6	29.5	23.9	13.0	—
	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	2025	81,413	3606.43	40.56	47.9	22.8	18.1	11.2	—
	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	2025	81,135	3628.84	39.59	57.0	15.8	13.9	13.2	—
	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	—

Assessment	Year	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4	%MOWR Met
Mathematics 8	2025	82,495	3655.33	36.72	53.4	19.8	15.9	10.9	–
	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 9.2. 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	79,299	48.9	25.8	25.3
	Reading for Literature	79,299	46.2	28.0	25.8
	Writing and Language	79,299	44.3	31.4	24.2
4	Reading for Information	78,164	38.2	27.6	34.3
	Reading for Literature	78,164	35.9	35.9	28.2
	Writing and Language	78,164	38.6	30.7	30.8
5	Reading for Information	81,330	44.6	27.7	27.8
	Reading for Literature	81,330	48.8	29.3	21.9
	Writing and Language	81,330	46.2	28.9	25.0
6	Reading for Information	81,086	48.7	24.8	26.5
	Reading for Literature	81,086	46.1	24.8	29.1
	Writing and Language	81,086	44.6	29.1	26.3
7	Reading for Information	80,645	43.2	26.7	30.1
	Reading for Literature	80,645	44.4	31.2	24.4
	Writing and Language	80,645	43.3	29.7	27.0
8	Reading for Information	82,002	45.2	30.3	24.5
	Reading for Literature	82,002	49.3	24.2	26.5
	Writing and Language	82,002	44.2	34.6	21.2

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

Table 9.3. 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	80,802	44.7	22.7	32.6
	Numbers and Operations – Fractions	80,802	35.3	45.8	18.9
	Measurement, Data, and Geometry	80,802	45.9	34.8	19.3
4	Operations, Algebraic Thinking, and Numbers in Base Ten	78,836	51.9	21.5	26.6
	Numbers and Operations – Fractions	78,836	51.5	21.1	27.4
	Measurement, Data, and Geometry	78,836	39.2	42.1	18.8
5	Operations, Algebraic Thinking, and Numbers in Base Ten	81,638	49.6	24.4	26.0
	Numbers and Operations – Fractions	81,638	49.6	27.2	23.3
	Measurement, Data, and Geometry	81,638	46.3	34.2	19.4
6	Ratio and Proportional Relationships	81,413	47.5	34.9	17.6
	The Number System	81,413	57.4	20.1	22.4
	Expressions and Equations	81,413	55.4	25.9	18.7
	Geometry, Statistics and Probability	81,413	53.8	28.4	17.8

Grade	Reporting Category	N	%Level 1	%Level 2	%Level 3
7	Ratio and Proportional Relationships	81,135	57.7	28.5	13.8
	The Number System	81,135	55.0	25.0	20.1
	Expressions & Equations	81,135	59.5	22.4	18.0
	Geometry, Statistics and Probability	81,135	64.7	18.3	16.9
8	Expressions and Equations	82,495	61.0	16.9	22.1
	Functions	82,495	56.8	28.8	14.4
	Geometry	82,495	51.6	32.2	16.2
	Statistics and Probability and The Number System	82,495	49.1	34.5	16.4

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

Table 9.4. Test Results by Accommodation—ELA

Grade	Accommodation	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
3	Adult Transcription	9	*	*	*	*	*	*
	American Sign Language	23	2450.74	12.30	100.0	0.0	0.0	0.0
	Assistive Technology	8	*	*	*	*	*	*
	Braille Test Booklet	3	*	*	*	*	*	*
	Accommodated-CBT	4	*	*	*	*	*	*
	Large Print Test Booklet	8	*	*	*	*	*	*
	Read Aloud Content	86	2467.37	25.10	89.5	3.5	4.7	2.3
	Sign Test Content	3	*	*	*	*	*	*
	Simplified Directions	335	2461.97	20.64	93.7	3.3	3.0	0.0
	Translate Directions	73	2463.11	19.57	95.9	1.4	2.7	0.0
	Translation Dictionary	95	2471.13	22.44	87.4	6.3	6.3	0.0
4	Adult Transcription	13	2480.46	27.54	92.3	0.0	0.0	7.7
	American Sign Language	23	2466.48	13.82	100.0	0.0	0.0	0.0
	Assistive Technology	23	2487.96	32.48	82.6	4.3	8.7	4.3
	Braille Test Booklet	6	*	*	*	*	*	*
	Large Print Test Booklet	12	2481.83	22.55	83.3	16.7	0.0	0.0
	Read Aloud Content	78	2488.00	29.71	73.1	12.8	11.5	2.6
	Sign Test Content	2	*	*	*	*	*	*
	Simplified Directions	354	2480.49	22.56	89.0	6.8	2.8	1.4
	Translate Directions	116	2477.92	20.77	89.7	6.9	3.4	0.0
	Translation Dictionary	161	2481.55	25.40	84.5	8.7	5.6	1.2
5	Adult Transcription	12	2477.42	14.36	100.0	0.0	0.0	0.0
	American Sign Language	19	2487.79	18.45	94.7	5.3	0.0	0.0
	Assistive Technology	11	2508.18	21.59	72.7	18.2	9.1	0.0
	Braille Test Booklet	6	*	*	*	*	*	*
	Large Print Test Booklet	10	*	*	*	*	*	*
	Read Aloud Content	68	2498.84	30.52	76.5	14.7	7.4	1.5
	Sign Test Content	1	*	*	*	*	*	*
	Simplified Directions	280	2489.38	23.13	89.3	8.2	1.1	1.4
	Translate Directions	67	2485.48	17.84	92.5	7.5	0.0	0.0
	Translation Dictionary	112	2493.79	22.29	84.8	13.4	1.8	0.0

Grade	Accommodation	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
6	Adult Transcription	2	*	*	*	*	*	*
	American Sign Language	19	2498.26	25.48	94.7	0.0	5.3	0.0
	Assistive Technology	23	2518.91	29.82	69.6	17.4	13.0	0.0
	Braille Test Booklet	8	*	*	*	*	*	*
	Large Print Test Booklet	8	*	*	*	*	*	*
	Read Aloud Content	46	2518.41	30.06	67.4	17.4	15.2	0.0
	Sign Test Content	3	*	*	*	*	*	*
	Simplified Directions	265	2504.75	22.39	86.0	11.3	2.6	0.0
	Translate Directions	53	2494.81	19.93	94.3	3.8	1.9	0.0
	Translation Dictionary	127	2502.88	25.52	87.4	8.7	3.1	0.8
7	Adult Transcription	0	*	*	*	*	*	*
	American Sign Language	20	2510.10	23.55	95.0	0.0	5.0	0.0
	Assistive Technology	15	2531.40	16.11	80.0	13.3	6.7	0.0
	Braille Test Booklet	9	*	*	*	*	*	*
	Large Print Test Booklet	6	*	*	*	*	*	*
	Read Aloud Content	15	2519.47	31.82	73.3	13.3	13.3	0.0
	Sign Test Content	0	*	*	*	*	*	*
	Simplified Directions	188	2517.12	22.05	86.2	9.6	4.3	0.0
	Translate Directions	77	2511.61	19.32	93.5	5.2	1.3	0.0
	Translation Dictionary	154	2521.32	24.59	82.5	9.1	7.8	0.6
8	Adult Transcription	3	*	*	*	*	*	*
	American Sign Language	18	2520.33	20.38	94.4	0.0	5.6	0.0
	Assistive Technology	17	2535.06	35.78	70.6	5.9	23.5	0.0
	Braille Test Booklet	6	*	*	*	*	*	*
	Large Print Test Booklet	6	*	*	*	*	*	*
	Read Aloud Content	16	2531.19	28.18	87.5	0.0	12.5	0.0
	Sign Test Content	2	*	*	*	*	*	*
	Simplified Directions	158	2524.06	25.67	86.1	8.2	5.7	0.0
	Translate Directions	57	2514.82	23.66	94.7	3.5	1.8	0.0
	Translation Dictionary	102	2518.84	24.75	90.2	6.9	2.9	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 and replaced with an asterisk (*). Read aloud is for Writing only.

Table 9.5. Test Results by Accommodation—Mathematics

Grade	Accommodation	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
3	Adult Transcription	10	*	*	*	*	*	*
	American Sign Language	24	3453.25	24.37	95.8	0.0	4.2	0.0
	Assistive Technology	10	*	*	*	*	*	*
	Braille Test Booklet	3	*	*	*	*	*	*
	Large Print Test Booklet	9	*	*	*	*	*	*
	Read Aloud Content	82	3503.45	36.85	43.9	26.8	23.2	6.1
	Sign Test Content	1	*	*	*	*	*	*
	Simplified Directions	316	3473.29	34.83	72.5	20.9	6.3	0.3
	Translate Directions	67	3478.16	33.03	73.1	17.9	7.5	1.5
	Translation Dictionary	91	3489.15	42.67	59.3	23.1	11.0	6.6

Grade	Accommodation	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
4	Adult Transcription	11	3486.73	29.52	100.0	0.0	0.0	0.0
	American Sign Language	25	3483.72	22.67	92.0	8.0	0.0	0.0
	Assistive Technology	21	3523.43	47.69	66.7	4.8	23.8	4.8
	Braille Test Booklet	6	*	*	*	*	*	*
	Large Print Test Booklet	12	3500.00	43.19	66.7	16.7	16.7	0.0
	Read Aloud Content	83	3515.18	35.00	62.7	24.1	13.3	0.0
	Sign Test Content	2	*	*	*	*	*	*
	Simplified Directions	340	3502.40	34.16	80.9	12.6	5.3	1.2
	Translate Directions	121	3500.55	30.69	83.5	12.4	4.1	0.0
	Translation Dictionary	138	3510.12	41.72	72.5	14.5	10.1	2.9
5	Adult Transcription	11	3544.64	18.24	81.8	18.2	0.0	0.0
	American Sign Language	20	3530.70	34.39	90.0	0.0	10.0	0.0
	Assistive Technology	12	3557.50	30.03	58.3	25.0	16.7	0.0
	Braille Test Booklet	6	*	*	*	*	*	*
	Large Print Test Booklet	8	*	*	*	*	*	*
	Read Aloud Content	40	3558.20	32.87	62.5	25.0	12.5	0.0
	Sign Test Content	2	*	*	*	*	*	*
	Simplified Directions	262	3544.57	28.42	77.5	17.2	4.6	0.8
	Translate Directions	67	3548.70	27.96	71.6	22.4	4.5	1.5
	Translation Dictionary	112	3556.08	33.05	61.6	25.0	9.8	3.6
6	Adult Transcription	1	*	*	*	*	*	*
	American Sign Language	19	3565.26	19.61	89.5	10.5	0.0	0.0
	Assistive Technology	18	3597.44	30.52	61.1	22.2	11.1	5.6
	Braille Test Booklet	8	*	*	*	*	*	*
	Large Print Test Booklet	8	*	*	*	*	*	*
	Read Aloud Content	48	3577.81	34.20	72.9	14.6	12.5	0.0
	Sign Test Content	2	*	*	*	*	*	*
	Simplified Directions	255	3571.78	27.07	85.5	9.8	4.3	0.4
	Translate Directions	53	3568.92	26.31	88.7	7.5	1.9	1.9
	Translation Dictionary	91	3576.40	33.46	81.3	9.9	3.3	5.5
7	Adult Transcription	0	*	*	*	*	*	*
	American Sign Language	21	3590.48	32.16	90.5	4.8	0.0	4.8
	Assistive Technology	14	3601.43	24.49	85.7	14.3	0.0	0.0
	Braille Test Booklet	9	*	*	*	*	*	*
	Large Print Test Booklet	5	*	*	*	*	*	*
	Read Aloud Content	13	3598.38	23.72	92.3	0.0	7.7	0.0
	Sign Test Content	0	*	*	*	*	*	*
	Simplified Directions	187	3595.96	18.80	92.5	6.4	1.1	0.0
	Translate Directions	82	3594.06	23.98	95.1	2.4	0.0	2.4
	Translation Dictionary	139	3603.16	32.34	82.7	7.2	6.5	3.6

Grade	Accommodation	N	SS Mean	SS SD	%Level 1	%Level 2	%Level 3	%Level 4
8	Adult Transcription	1	*	*	*	*	*	*
	American Sign Language	18	3629.50	26.71	83.3	11.1	0.0	5.6
	Assistive Technology	14	3638.43	31.38	85.7	7.1	0.0	7.1
	Braille Test Booklet	6	*	*	*	*	*	*
	Large Print Test Booklet	6	*	*	*	*	*	*
	Read Aloud Content	11	3631.64	29.06	81.8	0.0	18.2	0.0
	Sign Test Content	2	*	*	*	*	*	*
	Simplified Directions	154	3629.69	20.80	87.7	8.4	2.6	1.3
	Translate Directions	61	3630.10	16.92	86.9	9.8	3.3	0.0
	Translation Dictionary	104	3631.73	19.39	83.7	10.6	4.8	1.0

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 and replaced with an asterisk (*).

Table 9.6. Scale Score Distribution by Performance Level—ELA

Grade	Performance Level	N	Average Scale Score	%	Cumulative %
3	Level 1	40,966	2469.83	51.7	51.7
	Level 2	9,885	2502.28	12.5	64.1
	Level 3	18,082	2522.79	22.8	86.9
	Level 4	10,366	2557.38	13.1	100.0
4	Level 1	32,584	2482.94	41.7	41.7
	Level 2	9,623	2515.44	12.3	54.0
	Level 3	24,626	2538.20	31.5	85.5
	Level 4	11,331	2574.39	14.5	100.0
5	Level 1	31,922	2493.76	39.2	39.2
	Level 2	18,778	2530.10	23.1	62.3
	Level 3	21,430	2557.69	26.3	88.7
	Level 4	9,200	2594.98	11.3	100.0
6	Level 1	31,008	2506.96	38.2	38.2
	Level 2	16,831	2541.48	20.8	59.0
	Level 3	27,374	2569.81	33.8	92.8
	Level 4	5,873	2609.20	7.2	100.0
7	Level 1	31,554	2518.50	39.1	39.1
	Level 2	14,944	2550.62	18.5	57.7
	Level 3	26,566	2576.38	32.9	90.6
	Level 4	7,581	2612.64	9.4	100.0
8	Level 1	33,089	2524.63	40.4	40.4
	Level 2	15,899	2560.61	19.4	59.7
	Level 3	23,782	2585.72	29.0	88.7
	Level 4	9,232	2618.18	11.3	100.0

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

Table 9.7. Scale Score Distribution by Performance Level—Mathematics

Grade	Performance Level	N	Average Scale Score	%	Cumulative %
3	Level 1	23,931	3464.94	29.6	29.6
	Level 2	23,630	3512.33	29.2	58.9
	Level 3	22,282	3546.36	27.6	86.4
	Level 4	10,959	3587.36	13.6	100.0
4	Level 1	29,010	3499.35	36.8	36.8
	Level 2	18,576	3544.10	23.6	60.4
	Level 3	20,433	3579.50	25.9	86.3
	Level 4	10,817	3624.49	13.7	100.0
5	Level 1	27,456	3541.34	33.6	33.6
	Level 2	24,107	3577.97	29.5	63.2
	Level 3	19,496	3612.04	23.9	87.0
	Level 4	10,579	3654.42	13.0	100.0
6	Level 1	38,979	3572.14	47.9	47.9
	Level 2	18,590	3613.32	22.8	70.7
	Level 3	14,713	3641.87	18.1	88.8
	Level 4	9,131	3681.69	11.2	100.0
7	Level 1	46,265	3600.67	57.0	57.0
	Level 2	12,848	3638.59	15.8	72.9
	Level 3	11,316	3662.83	13.9	86.8
	Level 4	10,706	3702.95	13.2	100.0
8	Level 1	44,018	3628.57	53.4	53.4
	Level 2	16,297	3659.32	19.8	73.1
	Level 3	13,157	3687.36	15.9	89.1
	Level 4	9,023	3732.01	10.9	100.0

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

Chapter 10: RELIABILITY AND VALIDITY

This chapter provides evidence supporting the reliability and validity of scores on the spring 2025 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).

10.1. 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.

10.1.1. 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 10.1, Table 10.2, and Table 10.3 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 10.1. Coefficient Alpha and SEM by Total and Reporting Category Score—ELA, Form 1

Grade	Reporting Category	N	#Items	Coefficient Alpha	SEM
3	Total	37,162	46	0.92	3.20
	Reading for Information	37,162	20	0.83	1.95
	Reading for Literature	37,162	17	0.79	1.81
	Writing and Language	37,162	9	0.83	1.64
4	Total	36,638	44	0.92	3.12
	Reading for Information	36,638	19	0.82	1.85
	Reading for Literature	36,638	16	0.78	1.72
	Writing and Language	36,638	9	0.83	1.72
5	Total	38,945	44	0.92	3.10
	Reading for Information	38,945	18	0.83	1.81
	Reading for Literature	38,945	17	0.82	1.73
	Writing and Language	38,945	9	0.80	1.74
6	Total	38,432	44	0.92	3.17
	Reading for Information	38,432	18	0.78	1.83
	Reading for Literature	38,432	17	0.79	1.75
	Writing and Language	38,432	9	0.84	1.75
7	Total	38,502	44	0.91	3.14
	Reading for Information	38,502	20	0.83	1.89
	Reading for Literature	38,502	15	0.73	1.71
	Writing and Language	38,502	9	0.80	1.73
8	Total	39,551	44	0.92	3.13
	Reading for Information	39,551	18	0.81	1.82
	Reading for Literature	39,551	17	0.80	1.73
	Writing and Language	39,551	9	0.82	1.75

Table 10.2. Coefficient Alpha and SEM by Total and Reporting Category Score—ELA, Form 2

Grade	Reporting Category	N	#Items	Coefficient Alpha	SEM
3	Total	37,060	46	0.92	3.21
	Reading for Information	37,060	20	0.83	1.95
	Reading for Literature	37,060	17	0.80	1.81
	Writing and Language	37,060	9	0.82	1.66
4	Total	35,373	44	0.92	3.13
	Reading for Information	35,373	19	0.82	1.85
	Reading for Literature	35,373	16	0.77	1.72
	Writing and Language	35,373	9	0.83	1.71
5	Total	38,606	44	0.92	3.11
	Reading for Information	38,606	18	0.83	1.81
	Reading for Literature	38,606	17	0.82	1.73
	Writing and Language	38,606	9	0.80	1.75
6	Total	36,714	44	0.92	3.15
	Reading for Information	36,714	18	0.78	1.84
	Reading for Literature	36,714	17	0.79	1.75
	Writing and Language	36,714	9	0.83	1.73

Grade	Reporting Category	N	#Items	Coefficient Alpha	SEM
7	Total	35,483	44	0.92	3.22
	Reading for Information	35,483	20	0.83	1.89
	Reading for Literature	35,483	15	0.73	1.71
	Writing and Language	35,483	9	0.82	1.83
8	Total	37,675	44	0.92	3.15
	Reading for Information	37,675	18	0.81	1.82
	Reading for Literature	37,675	17	0.80	1.73
	Writing and Language	37,675	9	0.81	1.78

Table 10.3. Coefficient Alpha and SEM by Total and Reporting Category Score—Mathematics

Grade	Reporting Category	N	#Items	Coefficient Alpha	SEM
3	Total	77,564	45	0.94	2.69
	Operations, Algebraic Thinking, and Numbers in Base Ten	77,564	23	0.90	1.89
	Numbers and Operations – Fractions	77,564	9	0.75	1.18
	Measurement, Data, and Geometry	77,564	13	0.77	1.48
4	Total	75,510	45	0.94	2.67
	Operations, Algebraic Thinking, and Numbers in Base Ten	75,510	23	0.89	1.92
	Numbers and Operations – Fractions	75,510	14	0.87	1.44
	Measurement, Data, and Geometry	75,510	8	0.74	1.14
5	Total	78,380	45	0.93	2.79
	Operations, Algebraic Thinking, and Numbers in Base Ten	78,380	18	0.84	1.79
	Numbers and Operations – Fractions	78,380	15	0.83	1.60
	Measurement, Data, and Geometry	78,380	12	0.77	1.40
6	Total	77,676	47	0.93	2.82
	Ratio and Proportional Relationships	77,676	10	0.75	1.29
	The Number System	77,676	14	0.84	1.52
	Expressions and Equations	77,676	15	0.82	1.57
	Geometry, Statistics and Probability	77,676	8	0.61	1.22
7	Total	77,670	47	0.93	2.85
	Ratio and Proportional Relationships	77,670	10	0.73	1.29
	The Number System	77,670	10	0.78	1.31
	Expressions and Equations	77,670	12	0.79	1.41
	Geometry, Statistics and Probability	77,670	15	0.78	1.65
8	Total	79,153	47	0.92	2.89
	Functions	79,153	11	0.68	1.43
	Expressions and Equations	79,153	15	0.82	1.57
	Geometry	79,153	9	0.70	1.29
	Statistics and Probability and The Number System	79,153	12	0.75	1.45

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, TIF, and CSEM curves of the assessments. As expected, the CSEMs around the performance level cuts were the lowest.

10.1.2. Inter-rater Reliability

For the hand-scored 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 hand-scoring 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:

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

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

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):

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

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 10.4 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 10.4. 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	3,203	0.76	0.76	0.74	0.26
		Evidence & Elaboration	1-4	3,203	0.73	0.73	0.74	0.25
		Conventions & Editing	0-2	3,203	0.91	0.91	0.90	0.09
	WR 2	Statement of Purpose, Focus & Organization	1-4	2,955	0.75	0.75	0.67	0.32
		Evidence & Elaboration	1-4	2,955	0.76	0.76	0.73	0.27
		Conventions & Editing	0-2	2,955	0.93	0.93	0.96	0.03
4	WR 1	Statement of Purpose, Focus & Organization	1-4	2,704	0.78	0.78	0.69	0.31
		Evidence & Elaboration	1-4	2,704	0.78	0.78	0.68	0.31
		Conventions & Editing	0-2	2,704	0.82	0.82	0.90	0.07
	WR 2	Statement of Purpose, Focus & Organization	1-4	3,207	0.83	0.83	0.67	0.32
		Evidence & Elaboration	1-4	3,207	0.82	0.82	0.68	0.31
		Conventions & Editing	0-2	3,207	0.88	0.88	0.90	0.08
5	RD		0-1	7,124	0.90	0.90	0.96	0.04
	WR 1	Statement of Purpose, Focus & Organization	1-4	3,908	0.78	0.78	0.71	0.29
		Evidence & Elaboration	1-4	3,908	0.77	0.77	0.73	0.27
		Conventions & Editing	0-2	3,908	0.93	0.93	0.96	0.03
	WR 2	Statement of Purpose, Focus & Organization	1-4	3,801	0.74	0.74	0.65	0.34
		Evidence & Elaboration	1-4	3,801	0.71	0.71	0.64	0.35
		Conventions & Editing	0-2	3,801	0.90	0.90	0.95	0.05
6	WR 1	Statement of Purpose, Focus & Organization	1-4	3,629	0.81	0.81	0.64	0.36
		Evidence & Elaboration	1-4	3,629	0.81	0.81	0.64	0.36
		Conventions & Editing	0-2	3,629	0.99	0.99	1.00	0.00
	WR 2	Statement of Purpose, Focus & Organization	1-4	2,357	0.76	0.76	0.64	0.36
		Evidence & Elaboration	1-4	2,357	0.70	0.70	0.57	0.43
		Conventions & Editing	0-2	2,357	1.00	1.00	1.00	0.00
7	WR 1	Statement of Purpose, Focus & Organization	1-4	3,892	0.77	0.77	0.69	0.29
		Evidence & Elaboration	1-4	3,892	0.76	0.76	0.67	0.31
		Conventions & Editing	0-2	3,892	0.90	0.90	0.94	0.06
	WR 2	Statement of Purpose, Focus & Organization	1-4	760	0.77	0.76	0.61	0.38
		Evidence & Elaboration	1-4	760	0.74	0.74	0.63	0.35
		Conventions & Editing	0-2	760	0.82	0.82	0.87	0.11
8	WR 1	Statement of Purpose, Focus & Organization	1-4	3,970	0.74	0.74	0.58	0.41
		Evidence & Elaboration	1-4	3,970	0.76	0.76	0.59	0.39
		Conventions & Editing	0-2	3,970	0.84	0.84	0.90	0.09
	WR 2	Statement of Purpose, Focus & Organization	1-4	2,605	0.73	0.73	0.66	0.33
		Evidence & Elaboration	1-4	2,605	0.80	0.80	0.73	0.25
		Conventions & Editing	0-2	2,605	0.74	0.73	0.80	0.19

Note. OE = open-ended, ICC = intraclass correlation, WR = Writing Prompt, RD = Reading

10.2. 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:

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

Positive values of $\Delta MH\ DIF$ indicate items that favor the focal group, whereas negative values indicate items that favor the reference group. The MH chi-square statistic and the $\Delta MH\ DIF$ 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 10.5 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 10.5. DIF Flag Categories

Category	Description	MH Criterion	SMD Criterion
A	No DIF	$ \Delta MH\ DIF $ is not significantly different from 0 ($p < 0.05$) or $ \Delta MH\ DIF < 1.0$	MH chi-square not significantly different from 0 ($p < 0.05$) or $ SMD < 0.17$
B	Weak DIF	$ \Delta MH\ DIF $ is significantly different from 0 ($p < 0.05$) and $1.0 \leq \Delta MH\ DIF < 1.5$ or $ \Delta MH\ DIF $ is significantly different from 0 but not from 1 and $ \Delta MH\ DIF \geq 1.0$	MH chi-square significantly different from 0 ($p < 0.05$) and $0.17 < SMD \leq 0.25$
C	Strong DIF	$ \Delta MH\ DIF $ is significantly different from 1 ($p < 0.05$) and $ \Delta MH\ DIF \geq 1.5$	MH chi-square significantly higher than 0 ($p < 0.05$) and $ SMD > 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 10.6 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 10.6. Number of Items Exhibiting Strong DIF

Assessment	Total #Items	#Items with Strong DIF
ELA 3, Form 1	46	0
ELA 4, Form 1	44	1
ELA 5, Form 1	44	1
ELA 6, Form 1	44	1
ELA 7, Form 1	44	1
ELA 8, Form 1	44	2
ELA 3, Form 2	46	0
ELA 4, Form 2	44	1
ELA 5, Form 2	44	1
ELA 6, Form 2	44	1
ELA 7, Form 2	44	1
ELA 8, Form 2	44	2
Mathematics 3	45	0
Mathematics 4	45	0
Mathematics 5	45	0
Mathematics 6	47	0
Mathematics 7	47	0
Mathematics 8	47	0

10.3. 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 8. 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 10.1, Table 10.2, and Table 10.3) were used to calculate the corrected correlation coefficients for attenuation. The disattenuated correlations could be greater than 1.00.

Table 10.7 – Table 10.11 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 10.7. 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	0.93	0.91	0.86
	Reading for Information	1.06	1.00	0.81	0.69
	Reading for Literature	1.07	1.00	1.00	0.68
	Writing and Language	0.98	0.83	0.84	1.00
4	Total	1.00	0.93	0.91	0.89
	Reading for Information	1.07	1.00	0.79	0.73
	Reading for Literature	1.07	0.99	1.00	0.71
	Writing and Language	1.02	0.88	0.88	1.00
5	Total	1.00	0.93	0.93	0.89
	Reading for Information	1.06	1.00	0.82	0.74
	Reading for Literature	1.07	0.99	1.00	0.73
	Writing and Language	1.04	0.91	0.90	1.00
6	Total	1.00	0.91	0.91	0.91
	Reading for Information	1.07	1.00	0.78	0.73
	Reading for Literature	1.07	0.99	1.00	0.73
	Writing and Language	1.04	0.90	0.90	1.00
7	Total	1.00	0.93	0.89	0.89
	Reading for Information	1.07	1.00	0.76	0.73
	Reading for Literature	1.09	0.98	1.00	0.68
	Writing and Language	1.04	0.90	0.89	1.00
8	Total	1.00	0.92	0.91	0.90
	Reading for Information	1.07	1.00	0.79	0.74
	Reading for Literature	1.06	0.98	1.00	0.73
	Writing and Language	1.04	0.91	0.90	1.00

Table 10.8. 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	0.93	0.92	0.87
	Reading for Information	1.06	1.00	0.81	0.70
	Reading for Literature	1.07	0.99	1.00	0.68
	Writing and Language	1.00	0.85	0.84	1.00
4	Total	1.00	0.93	0.90	0.89
	Reading for Information	1.07	1.00	0.79	0.72
	Reading for Literature	1.07	0.99	1.00	0.70
	Writing and Language	1.02	0.87	0.88	1.00
5	Total	1.00	0.93	0.92	0.89
	Reading for Information	1.06	1.00	0.81	0.72
	Reading for Literature	1.06	0.98	1.00	0.72
	Writing and Language	1.04	0.88	0.89	1.00
6	Total	1.00	0.91	0.91	0.89
	Reading for Information	1.07	1.00	0.78	0.69
	Reading for Literature	1.07	0.99	1.00	0.69
	Writing and Language	1.02	0.86	0.85	1.00
7	Total	1.00	0.92	0.88	0.88
	Reading for Information	1.05	1.00	0.76	0.69
	Reading for Literature	1.07	0.98	1.00	0.65
	Writing and Language	1.01	0.84	0.84	1.00
8	Total	1.00	0.92	0.91	0.89
	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	1.03	0.88	0.88	1.00

Table 10.9. 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	0.86	0.92
	Operations, Algebraic Thinking, and Numbers in Base Ten	1.05	1.00	0.76	0.82
	Numbers and Operations – Fractions	1.02	0.93	1.00	0.73
	Measurement, Data, and Geometry	1.08	0.99	0.96	1.00
4	Total	1.00	0.97	0.94	0.86
	Operations, Algebraic Thinking, and Numbers in Base Ten	1.06	1.00	0.85	0.77
	Numbers and Operations – Fractions	1.04	0.97	1.00	0.76
	Measurement, Data, and Geometry	1.03	0.95	0.95	1.00
5	Total	1.00	0.94	0.93	0.88
	Operations, Algebraic Thinking, and Numbers in Base Ten	1.06	1.00	0.81	0.74
	Numbers & Operations – Fractions	1.06	0.97	1.00	0.74
	Measurement, Data, and Geometry	1.04	0.92	0.93	1.00

Table 10.10. 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	0.89	0.93	0.93	0.81
	Ratio and Proportional Relationships	1.07	1.00	0.77	0.77	0.66
	The Number System	1.05	0.97	1.00	0.81	0.70
	Expressions and Equations	1.06	0.98	0.98	1.00	0.69
	Geometry, Statistics and Probability	1.08	0.98	0.98	0.98	1.00
7	Total	1.00	0.88	0.91	0.91	0.92
	Ratio and Proportional Relationships	1.07	1.00	0.74	0.74	0.74
	The Number System	1.07	0.98	1.00	0.79	0.76
	Expressions and Equations	1.06	0.97	1.01	1.00	0.77
	Geometry, Statistics and Probability	1.08	0.98	0.97	0.98	1.00

Table 10.11. 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	0.93	0.85	0.85	0.89
	Expressions and Equations	1.18	1.00	0.71	0.73	0.77
	Functions	0.98	0.95	1.00	0.63	0.68
	Geometry	1.06	1.06	0.83	1.00	0.68
	Statistics and Probability and The Number System	1.07	1.08	0.87	0.94	1.00

10.4. 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.

10.4.1. 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.

10.4.2. 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 used operationally for the first time in spring 2025. 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 described 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.

10.4.3. 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 8. 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 8.3 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 8.4, 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 8.5. Correlations between the total score and reporting category score presented in Section 10.3 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 7.5 and Table 7.6 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 10.2. Items showing strong DIF are flagged for possible replacement in future administrations.

10.4.4. 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 11.1 for more details on standard setting.

10.4.5. 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 10.12 presents the correlation between AASA ELA and mathematics scale scores from the spring 2025 administration. The correlations range from 0.71 to 0.79.

Table 10.12. Correlation between AASA ELA and Mathematics Scale Scores

Grade	N	Correlation
3	79,052	0.78
4	78,023	0.79
5	81,120	0.77
6	80,802	0.77
7	80,369	0.77
8	81,649	0.71

10.4.6. 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 11: CLASSIFICATION INTO PERFORMANCE LEVELS

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 chapter provides information regarding classification of students into these four categories, including the consistency and accuracy with which students who took the spring 2025 AASA assessment were assigned to the performance levels, 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).

11.1. 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 11.1 presents the final scale score ranges for the AASA performance levels, and Table 11.2 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 11.1. Performance Level Cut Scores

Assessment	<i>Minimally Proficient</i>	<i>Partially Proficient</i>	<i>Proficient</i>	<i>Highly Proficient</i>
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 11.2. CSEM at Performance Level Cuts

Assessment	<i>Partially Proficient</i> Cut		<i>Proficient</i> Cut		<i>Highly Proficient</i> Cut	
	Scale Score	CSEM	Scale Score	CSEM	Scale Score	CSEM
ELA 3, Form 1	2497	9	2509	9	2541	10
ELA 4, Form 1	2510	9	2523	9	2559	11
ELA 5, Form 1	2520	9	2543	9	2578	11
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	10
ELA 4, Form 2	2510	9	2523	9	2559	11
ELA 5, Form 2	2520	9	2543	9	2578	11
ELA 6, Form 2	2532	9	2553	9	2597	12
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	15
Mathematics 4	3530	10	3562	10	3606	13
Mathematics 5	3563	10	3595	10	3635	12
Mathematics 6	3602	10	3629	10	3663	12
Mathematics 7	3629	10	3652	10	3680	11
Mathematics 8	3650	10	3673	9	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.

11.2. 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 the procedures described by Livingston and Lewis (1995). Classification consistency is calculated as the proportion of students classified consistently between two parallel forms, and classification accuracy is calculated as the proportion of students classified the same between observed scores and true scores, as shown in Table 11.3.

Table 11.3. Classification Consistency and Accuracy for the *Proficient* Cut

Observed Performance	Expected Performance	Classification Outcome
Not Proficient	Not Proficient	Consistent Classification
Not Proficient	Proficient	Inconsistent Classification
Proficient	Not Proficient	Inconsistent Classification
Proficient	Proficient	Consistent Classification
Not Proficient	Not Proficient	Accurate Classification
Not Proficient	Proficient	False Negative
Proficient	Not Proficient	False Positive
Proficient	Proficient	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 11.4 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 11.4. Classification Consistency and Accuracy Results

Assessment	N	Consistency	Inconsistency	Chance	κ	Accuracy	False Positive	False Negative
ELA 3, Form 1	37,162	0.76	0.24	0.36	0.62	0.82	0.10	0.08
ELA 4, Form 1	36,638	0.72	0.28	0.31	0.60	0.79	0.11	0.10
ELA 5, Form 1	38,945	0.73	0.27	0.30	0.61	0.80	0.11	0.09
ELA 6, Form 1	38,432	0.73	0.27	0.31	0.60	0.80	0.11	0.09
ELA 7, Form 1	38,502	0.71	0.29	0.30	0.58	0.79	0.12	0.10
ELA 8, Form 1	39,551	0.71	0.29	0.30	0.59	0.79	0.11	0.10
ELA 3, Form 2	37,060	0.75	0.25	0.35	0.62	0.81	0.10	0.08
ELA 4, Form 2	35,373	0.73	0.27	0.32	0.60	0.80	0.11	0.10
ELA 5, Form 2	38,606	0.72	0.28	0.29	0.61	0.80	0.11	0.09
ELA 6, Form 2	36,714	0.73	0.27	0.31	0.61	0.81	0.11	0.09
ELA 7, Form 2	35,483	0.73	0.27	0.32	0.60	0.80	0.11	0.09
ELA 8, Form 2	37,675	0.71	0.29	0.31	0.59	0.79	0.11	0.10
Mathematics 3	77,564	0.73	0.27	0.27	0.63	0.81	0.10	0.09
Mathematics 4	75,510	0.75	0.25	0.28	0.66	0.82	0.09	0.08
Mathematics 5	78,380	0.74	0.26	0.28	0.64	0.82	0.10	0.08
Mathematics 6	77,676	0.77	0.23	0.34	0.65	0.84	0.09	0.07
Mathematics 7	77,670	0.80	0.20	0.40	0.67	0.85	0.08	0.06
Mathematics 8	79,153	0.78	0.22	0.37	0.65	0.84	0.09	0.07

REFERENCES

- aimswbPlus. (2017). *AimswbPlus: Technical Manual*. Pearson
https://app.aimswbplus.com/help/fo_help/Content/Resources/PDF's/Guides%20and%20Manuals/aimswbPlus%20Technical%20Manual.pdf
- 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.
- Hasbrouck, J. & Tindal, G. (2017). *An update to compiled ORF norms* (Technical Report No. 1702). Eugene, OR, Behavioral Research and Teaching, University of Oregon.
- 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 Paedagogiske 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	<i>P</i> -Value	Point-Biserial
1	OE	4	37,162	0.28	0.73
2	OE	4	37,162	0.22	0.71
3	OE	2	37,162	0.64	0.69
4	OE	4	37,060	0.24	0.73
5	OE	4	37,060	0.22	0.66
6	OE	2	37,060	0.62	0.69
7	MC	1	76,004	0.52	0.50
8	MC	1	76,004	0.28	0.40
9	MX	1	76,004	0.51	0.58
10	MC	1	76,004	0.78	0.52
11	MC	1	76,004	0.50	0.42
12	MC	1	76,004	0.46	0.40
13	MC	1	76,004	0.55	0.46
14	MC	1	76,004	0.69	0.57
15	XI	1	76,004	0.62	0.59
16	MC	1	76,004	0.45	0.39
17	MC	1	76,004	0.36	0.44
18	XI	1	76,004	0.20	0.40
19	MC	1	76,004	0.49	0.46
20	MC	1	76,004	0.50	0.52
21	MX	2	76,004	0.46	0.45
22	MX	2	76,004	0.68	0.57
23	MC	1	76,004	0.71	0.45
24	MC	1	76,004	0.44	0.46
25	MC	1	76,004	0.47	0.40
26	MC	1	76,004	0.45	0.45
27	MC	1	76,004	0.44	0.46
28	MC	1	76,004	0.71	0.50
29	MC	1	76,004	0.54	0.33
30	MC	1	76,004	0.49	0.39
31	XI	1	76,004	0.36	0.57
32	MC	1	76,004	0.47	0.45
33	MC	1	76,004	0.49	0.42

Item Number	Item Type	Max. Points	N	<i>P</i> -Value	Point-Biserial
34	MC	1	76,004	0.32	0.30
35	MC	1	76,004	0.48	0.41
36	MC	1	76,004	0.42	0.44
37	MC	1	76,004	0.37	0.33
38	MX	1	76,004	0.40	0.55
39	MC	1	76,004	0.52	0.47
40	MX	1	76,004	0.25	0.59
41	MC	1	76,004	0.36	0.33
42	MC	1	76,004	0.47	0.43
43	MC	1	76,004	0.39	0.36
44	MC	1	76,004	0.42	0.30
45	MC	1	76,004	0.46	0.42
46	MX	2	76,004	0.56	0.54
47	MX	2	76,004	0.54	0.42
48	OE	2	37,998	0.54	0.56
49	OE	2	37,998	0.57	0.61
50	OE	2	38,006	0.56	0.58
51	OE	2	38,006	0.55	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.2. Item-Level CTT Statistics, ELA Grade 4

Item Number	Item Type	Max. Points	N	<i>P</i> -Value	Point-Biserial
1	OE	4	36,638	0.27	0.73
2	OE	4	36,638	0.29	0.74
3	OE	2	36,638	0.75	0.68
4	OE	4	35,373	0.34	0.78
5	OE	4	35,373	0.32	0.75
6	OE	2	35,373	0.75	0.69
7	MC	1	74,857	0.77	0.56
8	MC	1	74,857	0.54	0.57
9	MC	1	74,857	0.43	0.43
10	MC	1	74,857	0.72	0.48
11	MX	1	74,857	0.56	0.60
12	MC	1	74,857	0.41	0.36
13	MC	1	74,857	0.55	0.46
14	MC	1	74,857	0.58	0.53
15	MC	1	74,857	0.55	0.46
16	MC	1	74,857	0.45	0.44
17	MX	1	74,857	0.26	0.38
18	MC	1	74,857	0.42	0.26
19	MC	1	74,857	0.58	0.36
20	MC	1	74,857	0.49	0.48
21	MX	2	74,857	0.56	0.54
22	MX	2	74,857	0.50	0.51
23	MC	1	74,857	0.63	0.38

Item Number	Item Type	Max. Points	N	<i>P</i> -Value	Point-Biserial
24	MC	1	74,857	0.78	0.49
25	MX	1	74,857	0.28	0.40
26	MC	1	74,857	0.65	0.45
27	MC	1	74,857	0.33	0.46
28	MC	1	74,857	0.77	0.51
29	MC	1	74,857	0.71	0.38
30	MC	1	74,857	0.68	0.40
31	XI	1	74,857	0.39	0.41
32	MC	1	74,857	0.58	0.55
33	MC	1	74,857	0.71	0.58
34	MC	1	74,857	0.47	0.62
35	MC	1	74,857	0.60	0.48
36	MC	1	74,857	0.43	0.33
37	MC	1	74,857	0.29	0.48
38	MC	1	74,857	0.72	0.57
39	MC	1	74,857	0.32	0.20
40	MC	1	74,857	0.54	0.46
41	MX	1	74,857	0.35	0.41
42	XI	1	74,857	0.20	0.40
43	MC	1	74,857	0.40	0.36
44	MC	1	74,857	0.45	0.37
45	MC	1	74,857	0.76	0.53
46	MX	2	74,857	0.53	0.53
47	MX	2	74,857	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	<i>P</i> -Value	Point-Biserial
1	OE	4	38,945	0.26	0.74
2	OE	4	38,945	0.23	0.75
3	OE	2	38,945	0.72	0.68
4	OE	4	38,606	0.30	0.71
5	OE	4	38,606	0.29	0.72
6	OE	2	38,606	0.71	0.71
7	MC	1	78,081	0.87	0.44
8	MC	1	78,081	0.67	0.49
9	MC	1	78,081	0.63	0.46
10	MX	1	78,081	0.48	0.56
11	MC	1	78,081	0.30	0.37
12	XI	1	78,081	0.34	0.46
13	MC	1	78,081	0.56	0.41
14	MC	1	78,081	0.66	0.32
15	MX	1	78,081	0.59	0.62
16	MX	1	78,081	0.45	0.45
17	MC	1	78,081	0.57	0.60

Item Number	Item Type	Max. Points	N	<i>P</i> -Value	Point-Biserial
18	MC	1	78,081	0.35	0.43
19	MC	1	78,081	0.44	0.38
20	MC	1	78,081	0.65	0.36
21	MX	2	78,081	0.48	0.50
22	MX	2	78,081	0.51	0.50
23	MC	1	78,081	0.63	0.41
24	MC	1	78,081	0.24	0.48
25	MC	1	78,081	0.49	0.46
26	MC	1	78,081	0.66	0.38
27	MC	1	78,081	0.66	0.44
28	MC	1	78,081	0.65	0.50
29	MC	1	78,081	0.48	0.35
30	MC	1	78,081	0.47	0.41
31	MC	1	78,081	0.62	0.57
32	MC	1	78,081	0.50	0.38
33	MC	1	78,081	0.76	0.55
34	MC	1	78,081	0.51	0.56
35	MC	1	78,081	0.36	0.48
36	MC	1	78,081	0.61	0.64
37	MX	1	78,081	0.30	0.53
38	MC	1	78,081	0.52	0.34
39	OE	1	66,102	0.38	0.56
40	MC	1	78,081	0.67	0.52
41	MC	1	78,081	0.40	0.34
42	MX	1	78,081	0.38	0.64
43	MC	1	78,081	0.54	0.46
44	XI	1	78,081	0.18	0.50
45	MC	1	78,081	0.72	0.40
46	MX	2	78,081	0.60	0.59
47	MX	2	78,081	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	<i>P</i> -Value	Point-Biserial
1	OE	4	38,432	0.45	0.75
2	OE	4	38,432	0.39	0.72
3	OE	2	38,432	0.75	0.72
4	OE	4	36,714	0.36	0.72
5	OE	4	36,714	0.33	0.71
6	OE	2	36,714	0.69	0.73
7	MC	1	77,352	0.63	0.45
8	MC	1	77,352	0.48	0.38
9	MC	1	77,352	0.81	0.42
10	XI	1	77,352	0.25	0.40
11	MC	1	77,352	0.66	0.39

Item Number	Item Type	Max. Points	N	<i>P</i> -Value	Point-Biserial
12	MC	1	77,352	0.65	0.50
13	MC	1	77,352	0.62	0.44
14	MX	1	77,352	0.44	0.52
15	MC	1	77,352	0.58	0.43
16	MC	1	77,352	0.58	0.47
17	MC	1	77,352	0.46	0.30
18	MC	1	77,352	0.61	0.49
19	MC	1	77,352	0.41	0.32
20	MC	1	77,352	0.72	0.42
21	MX	2	77,352	0.65	0.62
22	MX	2	77,352	0.32	0.38
23	MC	1	77,352	0.66	0.50
24	MC	1	77,352	0.66	0.50
25	MC	1	77,352	0.61	0.43
26	MC	1	77,352	0.55	0.37
27	MC	1	77,352	0.50	0.37
28	MC	1	77,352	0.57	0.47
29	MC	1	77,352	0.63	0.46
30	MC	1	77,352	0.41	0.44
31	MC	1	77,352	0.55	0.44
32	MC	1	77,352	0.55	0.50
33	MC	1	77,352	0.83	0.49
34	MX	1	77,352	0.63	0.40
35	MC	1	77,352	0.72	0.56
36	MX	1	77,352	0.11	0.33
37	MC	1	77,352	0.46	0.52
38	MC	1	77,352	0.22	0.35
39	MC	1	77,352	0.50	0.36
40	MC	1	77,352	0.43	0.41
41	MX	1	77,352	0.16	0.32
42	MC	1	77,352	0.62	0.47
43	MX	1	77,352	0.38	0.53
44	MC	1	77,352	0.37	0.28
45	MC	1	77,352	0.66	0.62
46	MX	2	77,352	0.51	0.51
47	MX	2	77,352	0.54	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	<i>P</i> -Value	Point-Biserial
1	OE	4	38,502	0.34	0.72
2	OE	4	38,502	0.32	0.72
3	OE	2	38,502	0.77	0.69
4	OE	4	35,483	0.43	0.79
5	OE	4	35,483	0.37	0.77

Item Number	Item Type	Max. Points	N	P-Value	Point-Biserial
6	OE	2	35,483	0.74	0.71
7	MC	1	77,204	0.44	0.38
8	MC	1	77,204	0.74	0.46
9	XI	1	77,204	0.40	0.36
10	MC	1	77,204	0.53	0.34
11	MC	1	77,204	0.60	0.48
12	XI	1	77,204	0.41	0.49
13	MC	1	77,204	0.60	0.36
14	MC	1	77,204	0.57	0.47
15	MC	1	77,204	0.63	0.50
16	MC	1	77,204	0.28	0.38
17	MC	1	77,204	0.51	0.26
18	MC	1	77,204	0.50	0.29
19	MC	1	77,204	0.41	0.27
20	MC	1	77,204	0.62	0.40
21	MX	2	77,204	0.72	0.61
22	MX	2	77,204	0.47	0.57
23	MC	1	77,204	0.77	0.49
24	MC	1	77,204	0.69	0.36
25	MC	1	77,204	0.81	0.49
26	MC	1	77,204	0.46	0.34
27	MC	1	77,204	0.71	0.51
28	MC	1	77,204	0.53	0.48
29	XI	1	77,204	0.23	0.35
30	MC	1	77,204	0.74	0.51
31	MC	1	77,204	0.59	0.35
32	MC	1	77,204	0.76	0.49
33	MC	1	77,204	0.58	0.53
34	MC	1	77,204	0.51	0.33
35	MC	1	77,204	0.64	0.45
36	MC	1	77,204	0.65	0.58
37	MC	1	77,204	0.61	0.44
38	MC	1	77,204	0.64	0.58
39	XI	1	77,204	0.27	0.35
40	MC	1	77,204	0.60	0.44
41	MC	1	77,204	0.42	0.38
42	MX	1	77,204	0.42	0.63
43	MC	1	77,204	0.54	0.47
44	XI	1	77,204	0.25	0.55
45	MC	1	77,204	0.48	0.33
46	MX	2	77,204	0.44	0.45
47	MX	2	77,204	0.51	0.38

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	39,551	0.41	0.74
2	OE	4	39,551	0.34	0.75
3	OE	2	39,551	0.76	0.70
4	OE	4	37,675	0.40	0.76
5	OE	4	37,675	0.31	0.69
6	OE	2	37,675	0.76	0.70
7	MC	1	78,671	0.73	0.45
8	XI	1	78,671	0.71	0.51
9	MC	1	78,671	0.54	0.50
10	MC	1	78,671	0.26	0.34
11	MC	1	78,671	0.55	0.47
12	MC	1	78,671	0.53	0.27
13	MC	1	78,671	0.45	0.41
14	MX	1	78,671	0.28	0.31
15	MC	1	78,671	0.53	0.47
16	MC	1	78,671	0.46	0.40
17	MC	1	78,671	0.38	0.38
18	XI	1	78,671	0.29	0.39
19	MX	1	78,671	0.37	0.41
20	MC	1	78,671	0.75	0.47
21	MX	2	78,671	0.65	0.42
22	MX	2	78,671	0.71	0.60
23	MC	1	78,671	0.67	0.45
24	MC	1	78,671	0.63	0.50
25	MC	1	78,671	0.69	0.54
26	MC	1	78,671	0.68	0.49
27	MC	1	78,671	0.68	0.56
28	MC	1	78,671	0.54	0.37
29	MC	1	78,671	0.80	0.54
30	MC	1	78,671	0.57	0.36
31	MC	1	78,671	0.72	0.41
32	MX	1	78,671	0.32	0.51
33	MC	1	78,671	0.67	0.55
34	MC	1	78,671	0.49	0.39
35	MC	1	78,671	0.39	0.38
36	MX	1	78,671	0.41	0.62
37	MC	1	78,671	0.50	0.42
38	MC	1	78,671	0.28	0.26
39	MC	1	78,671	0.62	0.55
40	MX	1	78,671	0.32	0.46
41	MC	1	78,671	0.62	0.57
42	MC	1	78,671	0.56	0.47
43	MC	1	78,671	0.83	0.52
44	MC	1	78,671	0.27	0.45
45	MC	1	78,671	0.69	0.53

Item Number	Item Type	Max. Points	N	P-Value	Point-Biserial
46	MX	2	78,671	0.42	0.42
47	MX	2	78,671	0.50	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	77,564	0.65	0.62
2	MC	1	77,564	0.83	0.55
3	XI	1	77,564	0.57	0.55
4	MC	1	77,564	0.41	0.34
5	XI	1	77,564	0.34	0.61
6	MC	1	77,564	0.80	0.54
7	MC	1	77,564	0.54	0.61
8	XI	1	77,564	0.78	0.60
9	XI	1	77,564	0.18	0.46
10	MC	1	77,564	0.68	0.62
11	XI	1	77,564	0.35	0.51
12	MC	1	77,564	0.36	0.28
13	MC	1	77,564	0.61	0.39
14	MC	1	77,564	0.83	0.52
15	MC	1	77,564	0.41	0.49
16	XI	1	77,564	0.28	0.55
17	MC	1	77,564	0.57	0.53
18	XI	1	77,564	0.62	0.53
19	MC	1	77,564	0.57	0.40
20	MC	1	77,564	0.72	0.61
21	XI	1	77,564	0.74	0.51
22	MC	1	77,564	0.81	0.53
23	MC	1	77,564	0.62	0.44
24	MC	1	77,564	0.79	0.42
25	MC	1	77,564	0.83	0.50
26	XI	1	77,564	0.62	0.60
27	XI	1	77,564	0.75	0.42
28	XI	1	77,564	0.62	0.46
29	MC	1	77,564	0.46	0.28
30	MC	1	77,564	0.59	0.51
31	MC	1	77,564	0.41	0.55
32	MC	1	77,564	0.68	0.44
33	MC	1	77,564	0.69	0.62
34	MX	1	77,564	0.21	0.43
35	MC	1	77,564	0.44	0.47
36	XI	1	77,564	0.58	0.52
37	MC	1	77,564	0.73	0.57
38	MC	1	77,564	0.48	0.46
39	XI	1	77,564	0.70	0.66

Item Number	Item Type	Max. Points	N	P-Value	Point-Biserial
40	XI	1	77,564	0.45	0.62
41	MC	1	77,564	0.73	0.52
42	XI	1	77,564	0.32	0.56
43	XI	1	77,564	0.60	0.59
44	MC	1	77,564	0.60	0.66
45	MC	1	77,564	0.79	0.50

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	75,510	0.89	0.45
2	XI	1	75,510	0.53	0.61
3	XI	1	75,510	0.39	0.56
4	MX	1	75,510	0.34	0.57
5	MC	1	75,510	0.41	0.43
6	MC	1	75,510	0.42	0.37
7	MC	1	75,510	0.77	0.48
8	XI	1	75,510	0.62	0.59
9	XI	1	75,510	0.19	0.47
10	MC	1	75,510	0.70	0.49
11	MC	1	75,510	0.56	0.57
12	XI	1	75,510	0.68	0.62
13	MC	1	75,510	0.32	0.62
14	XI	1	75,510	0.19	0.42
15	MC	1	75,510	0.48	0.61
16	MC	1	75,510	0.44	0.49
17	MC	1	75,510	0.70	0.53
18	MC	1	75,510	0.43	0.60
19	MX	1	75,510	0.33	0.45
20	MC	1	75,510	0.64	0.57
21	XI	1	75,510	0.69	0.57
22	MC	1	75,510	0.45	0.57
23	XI	1	75,510	0.71	0.60
24	MC	1	75,510	0.81	0.46
25	MC	1	75,510	0.65	0.50
26	MC	1	75,510	0.60	0.50
27	XI	1	75,510	0.26	0.58
28	MC	1	75,510	0.45	0.71
29	XI	1	75,510	0.44	0.66
30	MC	1	75,510	0.70	0.51
31	MX	1	75,510	0.19	0.53
32	XI	1	75,510	0.63	0.63
33	MC	1	75,510	0.62	0.62
34	MC	1	75,510	0.83	0.44
35	MC	1	75,510	0.40	0.39

Item Number	Item Type	Max. Points	N	<i>P</i> -Value	Point-Biserial
36	XI	1	75,510	0.21	0.47
37	MC	1	75,510	0.62	0.40
38	XI	1	75,510	0.68	0.56
39	MC	1	75,510	0.45	0.50
40	MC	1	75,510	0.62	0.50
41	XI	1	75,510	0.53	0.53
42	MC	1	75,510	0.34	0.41
43	XI	1	75,510	0.36	0.69
44	MC	1	75,510	0.69	0.56
45	XI	1	75,510	0.45	0.61

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	<i>P</i> -Value	Point-Biserial
1	XI	1	78,380	0.74	0.30
2	XI	1	78,380	0.74	0.46
3	MC	1	78,380	0.55	0.48
4	XI	1	78,380	0.42	0.70
5	MC	1	78,380	0.54	0.48
6	MC	1	78,380	0.62	0.50
7	MX	1	78,380	0.33	0.53
8	MC	1	78,380	0.79	0.50
9	XI	1	78,380	0.16	0.46
10	MC	1	78,380	0.33	0.50
11	MC	1	78,380	0.62	0.53
12	MC	1	78,380	0.47	0.45
13	XI	1	78,380	0.49	0.60
14	XI	1	78,380	0.36	0.47
15	MC	1	78,380	0.23	0.31
16	XI	1	78,380	0.33	0.61
17	MC	1	78,380	0.69	0.55
18	XI	1	78,380	0.34	0.64
19	XI	1	78,380	0.56	0.51
20	MC	1	78,380	0.61	0.39
21	MC	1	78,380	0.38	0.46
22	XI	1	78,380	0.29	0.55
23	MC	1	78,380	0.55	0.47
24	MC	1	78,380	0.70	0.50
25	XI	1	78,380	0.45	0.58
26	MC	1	78,380	0.21	0.41
27	MC	1	78,380	0.61	0.36
28	MC	1	78,380	0.59	0.57
29	MC	1	78,380	0.45	0.39
30	XI	1	78,380	0.52	0.56
31	XI	1	78,380	0.20	0.50

Item Number	Item Type	Max. Points	N	<i>P</i> -Value	Point-Biserial
32	XI	1	78,380	0.67	0.50
33	XI	1	78,380	0.48	0.40
34	XI	1	78,380	0.39	0.52
35	MC	1	78,380	0.66	0.52
36	MC	1	78,380	0.58	0.49
37	MC	1	78,380	0.58	0.44
38	XI	1	78,380	0.23	0.62
39	MC	1	78,380	0.56	0.41
40	MC	1	78,380	0.46	0.36
41	MC	1	78,380	0.54	0.46
42	XI	1	78,380	0.30	0.63
43	MC	1	78,380	0.30	0.34
44	XI	1	78,380	0.28	0.52
45	XI	1	78,380	0.69	0.54

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	<i>P</i> -Value	Point-Biserial
1	MC	1	77,676	0.84	0.41
2	XI	1	77,676	0.52	0.64
3	MC	1	77,676	0.66	0.28
4	XI	1	77,676	0.43	0.55
5	MC	1	77,676	0.65	0.43
6	MC	1	77,676	0.19	0.39
7	MC	1	77,676	0.50	0.68
8	XI	1	77,676	0.37	0.66
9	XI	1	77,676	0.38	0.58
10	MC	1	77,676	0.37	0.55
11	MC	1	77,676	0.32	0.32
12	XI	1	77,676	0.16	0.52
13	MC	1	77,676	0.29	0.34
14	XI	1	77,676	0.24	0.58
15	MC	1	77,676	0.41	0.29
16	XI	1	77,676	0.28	0.56
17	MC	1	77,676	0.40	0.35
18	XI	1	77,676	0.35	0.61
19	MC	1	77,676	0.36	0.56
20	MC	1	77,676	0.39	0.38
21	XI	1	77,676	0.51	0.55
22	MC	1	77,676	0.46	0.47
23	MC	1	77,676	0.48	0.37
24	MC	1	77,676	0.65	0.40
25	MC	1	77,676	0.79	0.40
26	XI	1	77,676	0.49	0.56
27	XI	1	77,676	0.31	0.63

Item Number	Item Type	Max. Points	N	P-Value	Point-Biserial
28	MC	1	77,676	0.53	0.44
29	XI	1	77,676	0.51	0.55
30	MC	1	77,676	0.49	0.59
31	XI	1	77,676	0.36	0.45
32	MC	1	77,676	0.18	0.33
33	MC	1	77,676	0.73	0.38
34	MX	1	77,676	0.38	0.36
35	MC	1	77,676	0.56	0.42
36	XI	1	77,676	0.38	0.66
37	XI	1	77,676	0.54	0.63
38	MC	1	77,676	0.31	0.51
39	XI	1	77,676	0.66	0.48
40	MC	1	77,676	0.23	0.43
41	XI	1	77,676	0.16	0.46
42	MC	1	77,676	0.57	0.48
43	XI	1	77,676	0.35	0.61
44	MC	1	77,676	0.65	0.59
45	MC	1	77,676	0.32	0.57
46	XI	1	77,676	0.27	0.63
47	MC	1	77,676	0.67	0.50

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	77,670	0.78	0.36
2	MC	1	77,670	0.74	0.47
3	XI	1	77,670	0.27	0.54
4	XI	1	77,670	0.40	0.58
5	XI	1	77,670	0.41	0.52
6	MC	1	77,670	0.25	0.55
7	MC	1	77,670	0.42	0.36
8	MX	1	77,670	0.22	0.37
9	MC	1	77,670	0.20	0.43
10	XI	1	77,670	0.50	0.59
11	MC	1	77,670	0.36	0.37
12	MC	1	77,670	0.53	0.45
13	MC	1	77,670	0.63	0.51
14	MC	1	77,670	0.56	0.42
15	XI	1	77,670	0.31	0.65
16	MC	1	77,670	0.34	0.46
17	XI	1	77,670	0.26	0.64
18	MC	1	77,670	0.56	0.43
19	XI	1	77,670	0.33	0.64
20	MC	1	77,670	0.68	0.48
21	XI	1	77,670	0.14	0.56

Item Number	Item Type	Max. Points	N	P-Value	Point-Biserial
22	MC	1	77,670	0.35	0.47
23	MC	1	77,670	0.66	0.49
24	MC	1	77,670	0.67	0.44
25	MC	1	77,670	0.53	0.38
26	XI	1	77,670	0.25	0.58
27	MC	1	77,670	0.48	0.58
28	MC	1	77,670	0.43	0.49
29	MC	1	77,670	0.51	0.46
30	XI	1	77,670	0.33	0.53
31	MC	1	77,670	0.44	0.47
32	XI	1	77,670	0.48	0.45
33	MC	1	77,670	0.50	0.46
34	XI	1	77,670	0.53	0.47
35	MC	1	77,670	0.19	0.39
36	MC	1	77,670	0.43	0.44
37	XI	1	77,670	0.29	0.52
38	MC	1	77,670	0.24	0.35
39	MC	1	77,670	0.54	0.40
40	XI	1	77,670	0.20	0.59
41	MC	1	77,670	0.42	0.53
42	MC	1	77,670	0.39	0.64
43	MC	1	77,670	0.74	0.28
44	MC	1	77,670	0.37	0.51
45	XI	1	77,670	0.46	0.57
46	MC	1	77,670	0.44	0.39
47	MC	1	77,670	0.54	0.50

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	79,153	0.64	0.41
2	MC	1	79,153	0.57	0.33
3	XI	1	79,153	0.35	0.64
4	XI	1	79,153	0.19	0.57
5	MC	1	79,153	0.24	0.44
6	MC	1	79,153	0.48	0.54
7	MC	1	79,153	0.52	0.52
8	XI	1	79,153	0.26	0.67
9	MX	1	79,153	0.33	0.53
10	MC	1	79,153	0.81	0.35
11	MC	1	79,153	0.42	0.40
12	MC	1	79,153	0.63	0.34
13	XI	1	79,153	0.47	0.56
14	XI	1	79,153	0.17	0.47
15	MC	1	79,153	0.37	0.49

Appendix A: Item-Level CTT Statistics

Item Number	Item Type	Max. Points	N	<i>P</i> -Value	Point-Biserial
16	MC	1	79,153	0.42	0.41
17	MX	1	79,153	0.23	0.37
18	MC	1	79,153	0.48	0.57
19	MC	1	79,153	0.70	0.45
20	XI	1	79,153	0.34	0.40
21	MC	1	79,153	0.22	0.42
22	MC	1	79,153	0.34	0.34
23	MC	1	79,153	0.51	0.41
24	MC	1	79,153	0.66	0.34
25	XI	1	79,153	0.50	0.34
26	MC	1	79,153	0.44	0.35
27	XI	1	79,153	0.17	0.57
28	MC	1	79,153	0.48	0.53
29	MC	1	79,153	0.55	0.35
30	XI	1	79,153	0.22	0.36
31	MC	1	79,153	0.53	0.43
32	MC	1	79,153	0.31	0.45
33	XI	1	79,153	0.50	0.56
34	MC	1	79,153	0.42	0.54
35	XI	1	79,153	0.47	0.53
36	MC	1	79,153	0.48	0.48
37	XI	1	79,153	0.20	0.49
38	MC	1	79,153	0.48	0.35
39	MC	1	79,153	0.51	0.45
40	MC	1	79,153	0.40	0.48
41	MC	1	79,153	0.45	0.49
42	MC	1	79,153	0.22	0.40
43	XI	1	79,153	0.33	0.64
44	MC	1	79,153	0.36	0.50
45	MX	1	79,153	0.35	0.46
46	MC	1	79,153	0.28	0.54
47	MC	1	79,153	0.41	0.41

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
7	51.8	0.50	9.0	-0.18	21.1	-0.25	18.0	-0.25	0.1
10	77.9	0.52	3.2	-0.22	9.5	-0.25	9.3	-0.35	0.1
11	50.3	0.42	19.7	-0.23	14.8	-0.18	15.2	-0.14	0.1
12	46.1	0.40	25.4	-0.11	15.3	-0.24	13.1	-0.20	0.1
13	54.7	0.46	19.0	-0.30	17.9	-0.17	8.2	-0.16	0.1
14	69.2	0.57	8.7	-0.29	12.8	-0.32	9.2	-0.25	0.1
16	44.8	0.39	13.2	-0.30	18.2	-0.13	23.7	-0.10	0.1
17	35.9	0.44	15.2	-0.20	35.3	-0.14	13.5	-0.19	0.2
19	49.2	0.46	17.9	-0.17	17.6	-0.30	15.2	-0.14	0.2
20	50.1	0.52	23.0	-0.27	12.0	-0.31	14.7	-0.12	0.2
23	70.8	0.45	11.7	-0.30	10.3	-0.21	7.1	-0.18	0.1
24	43.7	0.46	22.0	-0.15	20.3	-0.21	14.0	-0.23	0.1
25	46.9	0.40	25.2	-0.20	18.8	-0.23	9.0	-0.07	0.2
26	45.2	0.45	20.0	-0.15	22.7	-0.22	12.0	-0.21	0.1
27	43.6	0.46	17.8	-0.23	18.0	-0.22	20.5	-0.13	0.1
28	70.9	0.50	9.5	-0.25	13.1	-0.26	6.5	-0.27	0.1
29	54.0	0.33	11.7	-0.33	18.5	-0.14	15.7	-0.01	0.1
30	49.0	0.39	14.0	-0.25	21.8	-0.13	15.1	-0.15	0.1
32	47.2	0.45	7.0	-0.26	8.8	-0.28	36.8	-0.16	0.1
33	49.1	0.42	15.9	-0.20	21.6	-0.13	13.3	-0.24	0.1
34	32.3	0.30	24.6	-0.16	25.1	-0.09	17.8	-0.08	0.1
35	47.6	0.41	15.3	-0.15	14.3	-0.28	22.7	-0.12	0.1
36	42.1	0.44	14.4	-0.22	21.0	-0.23	22.4	-0.12	0.1
37	37.5	0.33	17.9	-0.14	25.2	-0.17	19.3	-0.07	0.1
39	51.6	0.47	31.8	-0.18	9.0	-0.30	7.5	-0.23	0.1
41	36.4	0.33	29.6	-0.02	15.0	-0.29	18.8	-0.11	0.2
42	46.6	0.43	22.8	-0.13	18.5	-0.27	12.0	-0.17	0.2
43	38.9	0.36	12.4	-0.26	37.2	-0.04	11.4	-0.21	0.2
44	42.1	0.30	16.2	-0.19	20.6	-0.11	21.0	-0.07	0.2
45	46.4	0.42	12.7	-0.26	25.5	-0.14	15.2	-0.16	0.2

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
7	77.4	0.56	8.3	-0.34	10.3	-0.34	4.0	-0.18	0.0
9	42.8	0.43	34.5	-0.16	9.4	-0.30	13.2	-0.15	0.1
10	72.3	0.48	4.0	-0.23	15.5	-0.27	8.1	-0.26	0.1
12	41.5	0.36	32.5	-0.21	11.9	-0.07	14.1	-0.16	0.0
13	55.2	0.46	25.5	-0.17	13.2	-0.29	6.1	-0.22	0.1
14	58.1	0.53	11.4	-0.29	25.4	-0.27	5.0	-0.22	0.1
15	55.3	0.46	6.5	-0.27	29.4	-0.22	8.7	-0.21	0.1
16	45.1	0.44	24.7	-0.10	13.8	-0.19	16.3	-0.30	0.1
18	41.7	0.26	35.2	0.03	13.4	-0.26	9.7	-0.19	0.1
19	58.0	0.36	11.5	-0.20	11.4	-0.29	19.0	-0.05	0.1
20	49.1	0.48	23.4	-0.20	8.3	-0.32	19.1	-0.16	0.1
23	62.9	0.38	14.5	-0.23	15.3	-0.18	7.3	-0.15	0.0
24	77.8	0.49	5.8	-0.21	11.5	-0.29	4.9	-0.30	0.0
26	65.2	0.45	6.6	-0.25	18.1	-0.25	10.0	-0.20	0.1
28	76.9	0.51	7.7	-0.27	10.3	-0.31	5.1	-0.22	0.0
29	70.6	0.38	5.5	-0.22	10.6	-0.24	13.2	-0.14	0.0
30	67.6	0.40	9.3	-0.27	18.5	-0.14	4.6	-0.26	0.0
33	70.5	0.58	9.0	-0.31	13.9	-0.34	6.6	-0.24	0.1
35	59.6	0.48	19.3	-0.16	14.1	-0.31	7.0	-0.26	0.1
36	43.4	0.33	13.5	-0.15	27.0	-0.15	16.0	-0.12	0.1
38	72.0	0.57	9.5	-0.30	11.0	-0.36	7.5	-0.22	0.1
39	31.8	0.20	32.9	-0.10	18.8	-0.13	16.3	0.02	0.1
40	54.4	0.46	20.7	-0.16	8.7	-0.33	16.1	-0.19	0.1
43	40.3	0.36	20.2	-0.13	28.6	-0.22	10.8	-0.09	0.1
44	44.9	0.37	20.7	-0.29	18.4	-0.18	15.9	0.01	0.1
45	75.9	0.53	9.2	-0.32	7.6	-0.31	7.3	-0.20	0.1

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
7	87.3	0.44	3.4	-0.25	7.8	-0.30	1.6	-0.16	0.0
8	67.4	0.49	11.4	-0.14	5.6	-0.28	15.5	-0.33	0.0
9	62.5	0.46	12.1	-0.29	9.5	-0.25	15.9	-0.15	0.0
11	29.5	0.37	23.8	-0.13	14.7	-0.22	31.9	-0.08	0.0
13	55.6	0.41	32.8	-0.18	5.8	-0.30	5.8	-0.23	0.0
14	66.0	0.32	13.7	-0.09	12.6	-0.22	7.6	-0.18	0.0
17	57.2	0.60	15.3	-0.26	13.8	-0.32	13.7	-0.27	0.0
19	43.6	0.38	11.0	-0.35	27.0	-0.10	18.4	-0.09	0.0
20	64.6	0.36	10.2	-0.22	14.9	-0.17	10.2	-0.15	0.1
23	63.3	0.41	16.7	-0.26	13.8	-0.16	6.2	-0.19	0.0
25	49.0	0.46	12.2	-0.26	17.8	-0.15	21.0	-0.22	0.0
26	66.1	0.38	7.0	-0.19	11.1	-0.27	15.8	-0.12	0.0
27	65.8	0.44	20.7	-0.17	7.9	-0.31	5.6	-0.25	0.0
28	65.2	0.50	9.6	-0.25	12.6	-0.27	12.6	-0.22	0.0
29	47.7	0.35	9.9	-0.30	33.3	-0.07	9.1	-0.19	0.0
30	47.4	0.41	24.3	-0.17	11.0	-0.28	17.3	-0.12	0.0
31	62.0	0.57	15.9	-0.28	15.0	-0.33	7.1	-0.23	0.0
32	50.2	0.38	18.3	-0.06	17.1	-0.22	14.5	-0.23	0.0
33	75.5	0.55	6.2	-0.25	12.6	-0.35	5.6	-0.26	0.0
35	36.1	0.48	22.3	-0.16	16.8	-0.28	24.9	-0.14	0.0
38	51.9	0.34	12.9	-0.13	22.9	-0.20	12.3	-0.14	0.0
40	66.7	0.52	10.7	-0.19	17.9	-0.40	4.7	-0.16	0.0
41	40.5	0.34	19.1	-0.12	25.8	-0.20	14.6	-0.09	0.0
43	53.8	0.46	17.3	-0.15	20.9	-0.26	8.0	-0.24	0.1
45	71.8	0.40	11.2	-0.20	4.9	-0.29	12.1	-0.17	0.0

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	62.9	0.45	7.8	-0.18	18.5	-0.20	10.8	-0.31	0.0
8	47.6	0.38	20.8	-0.25	26.6	-0.12	4.9	-0.15	0.0
9	80.6	0.42	10.7	-0.24	4.1	-0.25	4.6	-0.20	0.0
11	66.4	0.39	6.6	-0.16	15.5	-0.22	11.4	-0.20	0.0
12	65.0	0.50	6.8	-0.23	8.2	-0.32	20.0	-0.24	0.0
13	62.0	0.44	20.7	-0.18	13.5	-0.28	3.8	-0.23	0.0
15	58.4	0.43	18.2	-0.13	16.6	-0.29	6.7	-0.22	0.1
16	57.6	0.47	14.7	-0.13	9.1	-0.30	18.6	-0.25	0.1
17	46.0	0.30	27.9	-0.12	14.4	-0.18	11.6	-0.09	0.1
18	60.9	0.49	9.7	-0.29	17.3	-0.24	12.1	-0.19	0.1
19	41.0	0.32	16.5	-0.23	10.9	-0.20	31.5	-0.02	0.1
20	71.5	0.42	14.6	-0.21	6.1	-0.28	7.8	-0.19	0.1
23	65.9	0.50	20.4	-0.26	11.2	-0.34	2.5	-0.16	0.0
24	66.2	0.50	16.7	-0.27	8.8	-0.27	8.3	-0.21	0.0
25	60.9	0.43	8.6	-0.22	23.0	-0.23	7.5	-0.19	0.0
26	54.6	0.37	18.7	-0.17	13.0	-0.21	13.7	-0.14	0.0
27	50.2	0.37	19.4	-0.09	16.3	-0.20	14.2	-0.21	0.0
28	56.6	0.47	10.1	-0.27	23.1	-0.21	10.2	-0.21	0.0
29	63.3	0.46	14.6	-0.33	11.4	-0.23	10.7	-0.09	0.0
31	55.0	0.44	11.4	-0.25	22.6	-0.23	11.1	-0.14	0.0
32	55.1	0.50	8.4	-0.23	22.4	-0.34	14.0	-0.12	0.0
33	82.7	0.49	6.3	-0.28	6.0	-0.29	5.0	-0.22	0.0
35	71.9	0.56	8.7	-0.27	10.7	-0.33	8.8	-0.26	0.0
37	45.7	0.52	14.0	-0.27	16.5	-0.34	23.7	-0.10	0.0
39	49.8	0.36	18.7	-0.37	19.9	-0.09	11.5	0.00	0.1
40	43.4	0.42	17.2	-0.14	25.3	-0.22	14.1	-0.15	0.1
42	62.4	0.47	10.4	-0.25	20.4	-0.21	6.8	-0.26	0.1
44	36.7	0.28	7.9	-0.08	41.0	-0.20	14.5	-0.05	0.1
45	66.1	0.62	13.9	-0.39	7.1	-0.31	12.9	-0.23	0.1

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
7	44.5	0.38	30.2	-0.14	16.4	-0.25	8.9	-0.10	0.0
8	73.6	0.46	11.4	-0.31	9.2	-0.25	5.8	-0.14	0.0
10	53.1	0.34	24.7	-0.03	15.6	-0.26	6.6	-0.24	0.0
11	60.3	0.48	18.3	-0.26	12.4	-0.22	8.9	-0.22	0.0
13	60.0	0.36	7.9	-0.13	25.2	-0.19	6.9	-0.24	0.0
14	56.6	0.47	26.8	-0.19	12.5	-0.31	4.1	-0.24	0.1
15	62.8	0.50	13.8	-0.22	14.9	-0.31	8.4	-0.21	0.0
17	51.4	0.26	15.1	-0.20	24.3	-0.03	9.3	-0.17	0.1
18	49.7	0.29	31.5	-0.13	9.0	-0.23	9.8	-0.06	0.1
19	40.9	0.27	16.6	-0.10	17.8	-0.18	24.7	-0.05	0.1
20	61.5	0.40	23.4	-0.15	8.8	-0.28	6.2	-0.21	0.1
23	77.4	0.49	8.1	-0.27	7.2	-0.32	7.3	-0.18	0.0
24	68.8	0.36	5.0	-0.26	4.8	-0.26	21.4	-0.13	0.0
25	81.5	0.49	6.8	-0.25	7.0	-0.28	4.7	-0.25	0.0
26	46.3	0.34	20.5	-0.21	8.3	-0.22	24.9	-0.05	0.0
27	70.7	0.51	8.6	-0.30	15.6	-0.26	5.1	-0.24	0.0
28	53.2	0.48	15.0	-0.24	10.6	-0.21	21.1	-0.22	0.0
30	74.0	0.51	6.5	-0.25	14.1	-0.31	5.4	-0.25	0.0
31	59.2	0.35	6.9	-0.30	5.7	-0.29	28.2	-0.07	0.0
32	75.8	0.49	6.4	-0.30	8.5	-0.29	9.3	-0.18	0.0
33	58.5	0.53	13.2	-0.18	13.1	-0.30	15.2	-0.27	0.0
34	51.4	0.33	16.4	-0.12	7.1	-0.31	25.1	-0.09	0.0
35	64.1	0.45	14.2	-0.15	9.4	-0.36	12.3	-0.18	0.1
36	65.0	0.58	19.1	-0.26	8.4	-0.31	7.4	-0.32	0.1
37	61.1	0.44	12.9	-0.31	21.8	-0.16	4.1	-0.25	0.0
38	64.0	0.58	16.0	-0.25	10.3	-0.33	9.6	-0.28	0.1
40	59.9	0.44	13.7	-0.17	16.0	-0.30	10.4	-0.15	0.1
41	41.9	0.38	27.4	-0.07	24.0	-0.24	6.6	-0.22	0.1
43	54.1	0.47	20.2	-0.14	16.3	-0.28	9.3	-0.25	0.1
45	48.3	0.33	20.3	-0.17	9.5	-0.30	21.8	-0.03	0.1

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
7	72.6	0.45	12.4	-0.26	7.6	-0.27	7.4	-0.17	0.0
9	53.9	0.50	13.1	-0.27	23.6	-0.18	9.5	-0.28	0.0
11	55.3	0.47	10.4	-0.25	22.1	-0.16	12.2	-0.28	0.0
12	52.6	0.27	15.2	-0.10	19.3	-0.14	12.8	-0.14	0.0
13	45.1	0.41	28.5	-0.16	13.6	-0.17	12.9	-0.22	0.0
15	52.7	0.47	12.2	-0.18	15.6	-0.22	19.5	-0.23	0.1
16	46.1	0.40	17.7	-0.17	21.7	-0.11	14.5	-0.24	0.0
17	38.2	0.38	16.2	-0.25	23.7	-0.13	21.9	-0.09	0.0
20	74.9	0.47	16.4	-0.25	5.9	-0.33	2.7	-0.20	0.0
23	67.3	0.45	9.9	-0.29	11.5	-0.28	11.3	-0.12	0.0
24	62.9	0.50	16.4	-0.20	13.3	-0.31	7.4	-0.24	0.0
25	69.1	0.54	7.7	-0.21	17.2	-0.36	6.0	-0.25	0.0
26	67.9	0.49	8.6	-0.23	15.7	-0.30	7.7	-0.20	0.0
27	68.0	0.56	8.2	-0.26	8.9	-0.32	14.9	-0.27	0.0
28	53.7	0.37	17.0	-0.23	10.3	-0.31	19.0	-0.01	0.0
29	80.3	0.54	7.2	-0.29	5.7	-0.30	6.7	-0.29	0.0
30	56.5	0.36	7.8	-0.30	11.3	-0.25	24.3	-0.04	0.0
31	72.1	0.41	7.0	-0.28	17.5	-0.20	3.4	-0.21	0.0
33	67.0	0.55	6.7	-0.26	20.6	-0.32	5.7	-0.27	0.0
34	48.8	0.39	20.8	-0.09	22.6	-0.26	7.9	-0.19	0.0
35	38.7	0.38	35.4	-0.13	14.7	-0.23	11.2	-0.14	0.0
37	50.0	0.42	22.3	-0.01	13.1	-0.32	14.6	-0.27	0.0
38	28.4	0.26	17.4	-0.15	38.5	0.07	15.8	-0.26	0.0
39	62.2	0.55	12.8	-0.28	17.2	-0.29	7.9	-0.23	0.0
41	61.6	0.57	12.3	-0.24	14.7	-0.35	11.3	-0.23	0.1
42	55.8	0.47	13.4	-0.16	18.9	-0.21	11.9	-0.29	0.0
43	83.5	0.52	7.6	-0.31	4.4	-0.28	4.5	-0.26	0.0
45	69.2	0.53	8.2	-0.30	12.8	-0.34	9.8	-0.16	0.0

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
2	83.4	0.55	3.3	-0.25	10.8	-0.40	2.4	-0.24	0.1
4	40.8	0.34	7.1	-0.12	18.3	-0.08	33.8	-0.22	0.1
6	80.2	0.54	12.7	-0.42	1.6	-0.14	5.4	-0.25	0.1
10	67.5	0.62	8.2	-0.37	8.0	-0.26	16.2	-0.32	0.1
12	36.3	0.28	24.7	-0.14	28.4	-0.08	10.5	-0.12	0.2
13	60.9	0.39	23.0	-0.24	6.1	-0.17	9.9	-0.16	0.1
14	82.6	0.52	4.5	-0.33	7.5	-0.30	5.3	-0.22	0.2
15	40.9	0.49	20.4	-0.27	23.2	-0.16	15.3	-0.17	0.2
17	57.0	0.53	28.3	-0.34	8.9	-0.27	5.7	-0.13	0.2
19	56.7	0.40	19.9	-0.22	14.9	-0.21	8.2	-0.12	0.3
20	71.5	0.61	12.7	-0.28	10.4	-0.37	5.3	-0.29	0.2
22	80.8	0.53	7.8	-0.30	6.5	-0.31	4.7	-0.23	0.2
23	62.4	0.44	7.5	-0.29	8.0	-0.28	21.9	-0.13	0.2
24	78.7	0.42	10.8	-0.31	5.7	-0.16	4.7	-0.19	0.1
25	82.9	0.50	4.6	-0.29	6.9	-0.30	5.5	-0.21	0.1
29	45.9	0.28	24.7	-0.12	13.7	-0.17	15.7	-0.07	0.1
30	59.3	0.51	19.4	-0.30	13.1	-0.19	8.0	-0.24	0.1
32	68.2	0.44	10.2	-0.32	16.3	-0.18	5.3	-0.17	0.1
33	69.2	0.62	11.4	-0.33	8.5	-0.28	10.8	-0.32	0.1
35	43.9	0.47	18.8	-0.22	15.6	-0.23	21.6	-0.15	0.2
37	72.5	0.57	9.6	-0.35	9.5	-0.34	8.2	-0.17	0.2
45	79.0	0.50	10.5	-0.36	5.6	-0.23	4.8	-0.19	0.1

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
1	88.7	0.45	7.1	-0.35	2.8	-0.21	1.4	-0.15	0.0
5	40.9	0.43	34.6	-0.07	12.0	-0.23	12.5	-0.31	0.1
6	42.4	0.37	27.2	-0.19	14.0	-0.17	16.4	-0.10	0.1
7	77.2	0.48	12.1	-0.30	7.2	-0.24	3.6	-0.23	0.1
11	56.4	0.57	14.3	-0.15	21.8	-0.40	7.4	-0.25	0.1
15	48.0	0.61	24.8	-0.36	18.3	-0.26	8.8	-0.17	0.1
16	44.4	0.49	18.7	-0.25	26.8	-0.21	10.0	-0.19	0.1
17	70.1	0.53	14.1	-0.32	10.2	-0.27	5.5	-0.20	0.1
20	64.2	0.57	15.7	-0.22	10.0	-0.32	10.0	-0.32	0.1
22	45.0	0.57	38.3	-0.35	12.3	-0.28	4.3	-0.10	0.1
24	80.6	0.46	15.7	-0.38	2.6	-0.18	1.1	-0.12	0.0
25	65.5	0.50	15.8	-0.32	10.2	-0.28	8.5	-0.13	0.1
26	59.9	0.50	12.2	-0.23	16.4	-0.32	11.4	-0.16	0.0
30	70.4	0.51	7.4	-0.29	14.1	-0.27	8.0	-0.24	0.1
33	62.2	0.62	10.4	-0.30	16.7	-0.33	10.6	-0.27	0.1
34	82.7	0.44	7.0	-0.29	7.7	-0.23	2.6	-0.20	0.1
35	39.5	0.39	15.7	-0.07	19.7	-0.19	25.0	-0.21	0.1
37	61.6	0.40	10.5	-0.17	22.0	-0.25	5.9	-0.16	0.1
39	44.5	0.50	23.3	-0.14	15.7	-0.29	16.4	-0.22	0.1
40	61.9	0.50	10.4	-0.25	20.0	-0.30	7.6	-0.18	0.1
42	34.0	0.41	7.3	-0.31	11.7	-0.40	46.9	0.03	0.1
44	68.6	0.56	9.2	-0.28	9.4	-0.29	12.7	-0.29	0.1

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
3	54.8	0.48	36.1	-0.36	5.6	-0.21	3.6	-0.11	0.0
5	53.8	0.48	15.5	-0.19	22.5	-0.28	8.1	-0.19	0.0
6	62.3	0.50	15.2	-0.21	18.0	-0.36	4.6	-0.15	0.0
8	78.8	0.50	7.0	-0.29	10.4	-0.30	3.7	-0.18	0.0
11	61.7	0.53	12.8	-0.21	12.1	-0.33	13.3	-0.22	0.0
12	46.6	0.45	17.0	-0.20	25.0	-0.24	11.5	-0.13	0.1
17	68.8	0.55	3.3	-0.20	15.5	-0.32	12.5	-0.31	0.1
20	60.9	0.39	23.1	-0.23	10.3	-0.28	5.7	-0.04	0.1
21	38.1	0.46	15.4	-0.28	14.9	-0.25	31.6	-0.07	0.1
23	55.2	0.47	14.7	-0.26	20.5	-0.30	9.6	-0.08	0.1
24	69.6	0.50	10.6	-0.28	14.6	-0.31	5.2	-0.16	0.0
27	60.8	0.36	5.6	-0.21	11.9	-0.19	21.7	-0.16	0.0
29	45.0	0.39	14.6	-0.28	34.4	-0.11	6.0	-0.16	0.1
35	65.5	0.52	12.7	-0.26	8.5	-0.27	13.2	-0.25	0.1
36	57.7	0.49	15.2	-0.19	12.6	-0.31	14.5	-0.21	0.1
37	58.5	0.44	31.0	-0.23	5.8	-0.26	4.7	-0.22	0.1
39	56.3	0.41	25.6	-0.21	9.5	-0.27	8.5	-0.10	0.1
40	45.7	0.36	35.8	-0.08	10.1	-0.29	8.3	-0.20	0.1
41	54.4	0.46	20.9	-0.27	16.3	-0.23	8.2	-0.12	0.1
43	30.2	0.34	29.9	-0.09	28.2	-0.19	11.6	-0.09	0.1

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
1	83.9	0.41	9.7	-0.26	3.5	-0.22	2.9	-0.18	0.0
3	65.8	0.28	18.3	-0.17	5.5	-0.17	10.4	-0.09	0.0
5	65.3	0.43	12.7	-0.24	14.7	-0.27	7.2	-0.11	0.0
6	19.4	0.39	24.2	-0.19	39.7	-0.09	16.6	-0.08	0.0
10	36.6	0.55	15.0	-0.26	18.6	-0.31	29.8	-0.11	0.0
11	32.4	0.32	26.7	-0.21	22.1	-0.04	18.8	-0.10	0.1
13	29.3	0.34	7.9	-0.09	30.8	-0.17	32.0	-0.11	0.1
15	41.4	0.29	17.6	-0.09	24.2	-0.20	16.7	-0.06	0.1
17	40.4	0.35	22.7	-0.09	19.8	-0.31	17.0	-0.02	0.1
19	36.1	0.56	19.5	-0.29	17.7	-0.26	26.6	-0.12	0.1
20	38.6	0.38	22.7	-0.25	28.5	-0.11	10.1	-0.10	0.1
22	46.1	0.47	14.1	-0.12	27.8	-0.31	11.9	-0.16	0.1
23	47.8	0.37	25.2	-0.20	15.5	-0.14	11.4	-0.16	0.1
24	65.0	0.40	18.0	-0.31	13.5	-0.16	3.4	-0.11	0.0
25	79.3	0.40	11.4	-0.20	5.1	-0.24	4.2	-0.22	0.0
28	53.4	0.44	12.8	-0.24	18.5	-0.17	15.3	-0.21	0.0
30	49.2	0.59	20.6	-0.31	18.7	-0.27	11.6	-0.20	0.0
33	72.7	0.38	8.1	-0.18	12.6	-0.29	6.6	-0.08	0.0
35	55.6	0.42	13.0	-0.23	20.0	-0.22	11.4	-0.12	0.0
40	22.8	0.43	23.3	-0.13	26.4	-0.24	27.4	-0.05	0.1
42	57.4	0.48	19.1	-0.24	16.4	-0.32	7.0	-0.08	0.1
44	64.8	0.59	11.0	-0.27	19.2	-0.39	4.9	-0.20	0.1
47	66.9	0.50	7.8	-0.20	13.2	-0.29	12.2	-0.25	0.1

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
1	78.4	0.36	10.1	-0.24	5.6	-0.22	5.9	-0.10	0.0
2	74.4	0.47	12.5	-0.30	8.6	-0.25	4.5	-0.19	0.0
7	41.9	0.36	25.3	-0.19	23.5	-0.20	9.3	-0.04	0.1
9	20.2	0.43	23.9	0.06	36.2	-0.24	19.6	-0.20	0.1
11	36.3	0.37	10.8	-0.13	35.3	-0.29	17.6	0.00	0.1
12	53.3	0.45	14.2	-0.12	19.7	-0.34	12.8	-0.14	0.1
13	63.5	0.51	8.4	-0.23	18.2	-0.34	9.8	-0.16	0.1
14	56.0	0.42	9.5	-0.17	27.8	-0.31	6.7	-0.08	0.1
16	33.8	0.46	23.7	-0.08	26.6	-0.32	15.8	-0.11	0.1
18	55.7	0.43	11.9	-0.22	21.1	-0.21	11.3	-0.18	0.1
20	67.6	0.48	3.3	-0.14	23.9	-0.37	5.2	-0.20	0.1
22	34.5	0.47	18.8	-0.27	36.8	-0.23	9.8	-0.02	0.1
23	66.2	0.49	17.8	-0.28	10.3	-0.28	5.7	-0.17	0.1
24	67.0	0.44	13.2	-0.26	13.4	-0.24	6.5	-0.15	0.0
25	53.4	0.38	23.2	-0.07	18.5	-0.33	4.8	-0.15	0.1
27	47.7	0.58	20.2	-0.27	27.5	-0.35	4.6	-0.12	0.0
28	43.1	0.49	22.4	-0.23	15.3	-0.20	19.1	-0.19	0.0
29	51.3	0.46	26.1	-0.22	15.5	-0.26	7.0	-0.16	0.0
31	44.0	0.47	21.5	-0.15	23.8	-0.27	10.7	-0.19	0.1
33	50.3	0.46	16.5	-0.19	18.1	-0.25	15.1	-0.17	0.1
36	42.9	0.44	26.2	-0.22	22.3	-0.20	8.5	-0.12	0.1
38	24.4	0.35	26.0	-0.15	29.7	-0.18	19.8	0.00	0.1
39	54.4	0.40	9.9	-0.24	16.5	-0.29	19.2	-0.05	0.1
41	41.7	0.53	24.3	-0.10	16.4	-0.32	17.6	-0.25	0.1
43	74.3	0.28	1.7	-0.10	6.5	-0.15	17.6	-0.19	0.1
44	36.7	0.51	12.7	-0.21	29.1	-0.21	21.4	-0.19	0.1
46	43.6	0.39	22.1	-0.25	18.4	-0.24	15.8	0.01	0.1
47	53.7	0.50	10.7	-0.17	28.8	-0.30	6.7	-0.22	0.1

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		%Omit
	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	%	Pt. Bis.	
1	64.1	0.41	6.0	-0.16	15.0	-0.22	14.9	-0.23	0.0
2	56.6	0.33	5.8	-0.10	17.4	-0.23	20.2	-0.13	0.0
5	24.0	0.44	23.2	-0.05	24.2	-0.07	28.6	-0.30	0.0
6	48.3	0.54	22.7	-0.25	16.1	-0.24	12.8	-0.22	0.0
10	81.4	0.35	7.4	-0.21	6.0	-0.18	5.1	-0.17	0.0
11	42.2	0.40	24.0	-0.14	16.4	-0.22	17.4	-0.14	0.0
12	63.4	0.34	22.5	-0.15	8.4	-0.25	5.7	-0.14	0.0
15	37.1	0.49	27.8	-0.22	25.2	-0.23	9.9	-0.13	0.1
16	42.1	0.41	20.9	-0.13	21.2	-0.13	15.8	-0.26	0.1
18	47.9	0.57	23.2	-0.30	17.7	-0.25	11.1	-0.19	0.1
19	70.2	0.45	12.9	-0.25	10.8	-0.24	6.1	-0.19	0.1
21	22.3	0.42	16.9	-0.15	25.3	-0.28	35.5	0.00	0.1
22	34.1	0.34	25.8	-0.09	20.4	-0.20	19.6	-0.11	0.1
23	51.4	0.41	17.6	-0.14	21.0	-0.22	9.9	-0.20	0.1
24	66.4	0.34	13.4	-0.21	12.4	-0.11	7.8	-0.19	0.0
26	43.6	0.35	12.9	-0.19	29.8	-0.19	13.7	-0.07	0.0
28	48.1	0.53	17.7	-0.24	14.3	-0.25	19.9	-0.21	0.0
31	52.9	0.43	16.0	-0.15	21.5	-0.26	9.6	-0.18	0.0
32	31.2	0.45	19.8	-0.21	25.5	-0.13	23.5	-0.15	0.0
34	42.2	0.54	8.8	-0.18	33.7	-0.26	15.2	-0.26	0.0
36	47.7	0.48	19.6	-0.23	25.7	-0.25	7.1	-0.16	0.0
38	47.6	0.35	19.6	-0.17	22.0	-0.16	10.8	-0.15	0.0
39	51.1	0.45	17.6	-0.25	10.9	-0.23	20.3	-0.15	0.0
40	40.0	0.48	25.6	-0.24	22.7	-0.25	11.7	-0.08	0.0
41	44.6	0.49	17.5	-0.17	25.0	-0.29	12.8	-0.16	0.0
44	36.1	0.50	20.8	-0.19	21.4	-0.21	21.7	-0.20	0.0
46	28.3	0.54	24.3	-0.19	29.9	-0.25	17.5	-0.13	0.0
47	41.5	0.41	19.9	-0.10	25.2	-0.30	13.4	-0.10	0.0

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), test information function (TIF), 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	1.4688	0.0088	0.75
2	OE	2.1137	0.0094	0.77
3	OE	-0.8527	0.0083	0.76
4	OE	2.5036	0.0089	0.76
5	OE	2.7680	0.0097	0.84
6	OE	-0.6730	0.0080	0.80
7	MC	-0.1954	0.0082	0.95
8	MC	1.1028	0.0091	1.07
9	MX	-0.1786	0.0082	0.86
10	MC	-1.6895	0.0095	0.82
11	MC	-0.1179	0.0082	1.06
12	MC	0.0521	0.0082	1.08
13	MC	-0.3805	0.0082	1.00
14	MC	-1.1550	0.0087	0.82
15	XI	-0.7663	0.0084	0.82
16	MC	0.3002	0.0083	1.12
17	MC	0.6049	0.0086	1.02
18	XI	1.6063	0.0100	1.01
19	MC	-0.0982	0.0082	1.01
20	MC	-0.1028	0.0082	0.93
21	MX	0.0580	0.0061	1.21
22	MX	-1.0578	0.0059	0.95
23	MC	-1.2433	0.0088	0.94
24	MC	0.1852	0.0083	1.01
25	MC	0.0204	0.0082	1.08
26	MC	0.1529	0.0083	1.03
27	MC	0.2144	0.0083	1.01
28	MC	-1.2499	0.0088	0.88
29	MC	-0.3443	0.0082	1.14
30	MC	-0.0872	0.0082	1.09

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
31	XI	0.6072	0.0086	0.88
32	MC	0.0040	0.0082	1.02
33	MC	-0.0944	0.0082	1.06
34	MC	0.8174	0.0088	1.17
35	MC	0.1488	0.0083	1.08
36	MC	0.2703	0.0083	1.03
37	MC	0.5187	0.0085	1.16
38	MX	0.5364	0.0085	0.93
39	MC	-0.2187	0.0082	1.00
40	MX	1.2530	0.0094	0.80
41	MC	0.5751	0.0085	1.13
42	MC	0.0352	0.0082	1.04
43	MC	0.2964	0.0083	1.11
44	MC	0.2692	0.0083	1.21
45	MC	-0.1251	0.0082	1.06
46	MX	-0.5748	0.0060	1.07
47	MX	-0.4032	0.0065	1.18
48	OE	-0.4532	0.0109	0.88
49	OE	-0.6821	0.0102	0.85
50	OE	-0.6199	0.0105	0.88
51	OE	-0.5483	0.0106	0.84

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	1.7692	0.0081	0.76
2	OE	1.7348	0.0083	0.74
3	OE	-0.8768	0.0088	0.75
4	OE	1.5622	0.0080	0.68
5	OE	1.7429	0.0083	0.73
6	OE	-0.8724	0.0091	0.74
7	MC	-1.1279	0.0095	0.79
8	MC	0.0219	0.0084	0.89
9	MC	0.7500	0.0084	1.04
10	MC	-0.8981	0.0092	0.95
11	MX	0.0108	0.0084	0.85
12	MC	0.8220	0.0085	1.14
13	MC	0.1003	0.0084	1.02
14	MC	-0.0525	0.0084	0.94
15	MC	0.0942	0.0084	1.02
16	MC	0.6272	0.0084	1.04
17	MX	1.7067	0.0093	1.06
18	MC	0.8101	0.0084	1.27
19	MC	-0.0490	0.0084	1.14
20	MC	0.4170	0.0083	1.00

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
21	MX	0.0309	0.0060	1.09
22	MX	0.3502	0.0061	1.13
23	MC	-0.3075	0.0086	1.10
24	MC	-1.2136	0.0097	0.88
25	MX	1.6200	0.0092	1.05
26	MC	-0.4407	0.0087	1.00
27	MC	1.3689	0.0089	1.00
28	MC	-1.1581	0.0096	0.87
29	MC	-0.7538	0.0090	1.06
30	MC	-0.5753	0.0088	1.04
31	XI	0.9378	0.0085	1.06
32	MC	-0.0410	0.0084	0.90
33	MC	-0.7470	0.0090	0.83
34	MC	0.4611	0.0083	0.81
35	MC	-0.0609	0.0084	0.98
36	MC	0.7066	0.0084	1.18
37	MC	1.6629	0.0093	0.97
38	MC	-0.8388	0.0091	0.83
39	MC	1.3580	0.0089	1.26
40	MC	0.1436	0.0084	1.03
41	MX	1.1998	0.0087	1.05
42	XI	2.1747	0.0102	0.99
43	MC	0.8816	0.0085	1.11
44	MC	0.6392	0.0084	1.13
45	MC	-1.0882	0.0095	0.86
46	MX	0.2945	0.0059	1.14
47	MX	-0.8483	0.0064	1.06

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	1.8996	0.0081	0.77
2	OE	2.1237	0.0085	0.72
3	OE	-0.9807	0.0086	0.77
4	OE	1.5410	0.0079	0.85
5	OE	1.7158	0.0083	0.80
6	OE	-0.9567	0.0087	0.70
7	MC	-2.2262	0.0115	0.86
8	MC	-0.7847	0.0086	0.93
9	MC	-0.5117	0.0084	1.00
10	MX	0.2449	0.0082	0.90
11	MC	1.2925	0.0089	1.09
12	XI	1.0317	0.0086	1.03
13	MC	-0.1382	0.0082	1.08
14	MC	-0.7056	0.0085	1.16

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
15	MX	-0.1997	0.0082	0.81
16	MX	0.6657	0.0084	1.08
17	MC	-0.2249	0.0082	0.85
18	MC	1.2005	0.0088	1.11
19	MC	0.5174	0.0083	1.15
20	MC	-0.6254	0.0085	1.12
21	MX	0.1654	0.0061	1.16
22	MX	0.1703	0.0062	1.19
23	MC	-0.5521	0.0084	1.06
24	MC	1.6251	0.0094	0.93
25	MC	0.2156	0.0082	1.03
26	MC	-0.7090	0.0085	1.10
27	MC	-0.6957	0.0085	1.00
28	MC	-0.6615	0.0085	0.94
29	MC	0.1600	0.0082	1.18
30	MC	0.1335	0.0082	1.10
31	MC	-0.4039	0.0083	0.86
32	MC	0.2012	0.0082	1.13
33	MC	-1.4313	0.0095	0.89
34	MC	0.0917	0.0082	0.90
35	MC	0.9064	0.0085	0.97
36	MC	-0.4510	0.0083	0.77
37	MX	1.2401	0.0089	0.90
38	MC	0.0548	0.0082	1.18
39	OE	0.7985	0.0092	0.91
40	MC	-0.7453	0.0086	0.92
41	MC	0.6604	0.0084	1.18
42	MX	0.7928	0.0084	0.78
43	MC	-0.0467	0.0082	1.03
44	XI	2.0783	0.0103	0.86
45	MC	-1.0498	0.0089	1.03
46	MX	-0.3805	0.0058	1.04
47	MX	0.6550	0.0062	1.38

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	0.5417	0.0070	0.79
2	OE	0.8560	0.0073	0.83
3	OE	-1.1609	0.0091	0.66
4	OE	1.0120	0.0077	0.80
5	OE	1.2673	0.0078	0.83
6	OE	-0.8189	0.0088	0.69
7	MC	-0.4863	0.0084	0.99
8	MC	0.3599	0.0082	1.10

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
9	MC	-1.5869	0.0099	0.93
10	XI	1.5617	0.0092	1.02
11	MC	-0.5633	0.0084	1.04
12	MC	-0.6006	0.0085	0.93
13	MC	-0.4348	0.0083	1.02
14	MX	0.5037	0.0082	0.93
15	MC	-0.2484	0.0082	1.04
16	MC	-0.2038	0.0082	0.99
17	MC	0.3942	0.0082	1.20
18	MC	-0.3768	0.0083	0.96
19	MC	0.6553	0.0083	1.17
20	MC	-1.0511	0.0090	1.01
21	MX	-0.5847	0.0059	0.91
22	MX	1.2331	0.0065	1.24
23	MC	-0.6522	0.0085	0.94
24	MC	-0.6652	0.0085	0.93
25	MC	-0.3788	0.0083	1.03
26	MC	-0.0477	0.0082	1.11
27	MC	0.1806	0.0081	1.13
28	MC	-0.1535	0.0082	0.99
29	MC	-0.5114	0.0084	0.99
30	MC	0.8249	0.0084	1.05
31	MC	0.0233	0.0082	1.03
32	MC	0.0223	0.0082	0.95
33	MC	-1.7496	0.0103	0.85
34	MX	-0.3626	0.0083	1.04
35	MC	-0.9971	0.0089	0.85
36	MX	2.7716	0.0123	0.98
37	MC	0.4094	0.0082	0.93
38	MC	1.8099	0.0096	1.05
39	MC	0.1967	0.0082	1.13
40	MC	0.3218	0.0082	1.04
41	MX	2.2736	0.0107	1.03
42	MC	-0.4650	0.0084	0.98
43	MX	0.5972	0.0082	0.88
44	MC	0.9370	0.0085	1.19
45	MC	-0.6626	0.0085	0.79
46	MX	0.1710	0.0059	1.13
47	MX	-0.0494	0.0058	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	1.3714	0.0081	0.76
2	OE	1.5055	0.0082	0.75
3	OE	-1.1187	0.0090	0.70
4	OE	0.6575	0.0072	0.68
5	OE	1.0457	0.0076	0.70
6	OE	-0.8547	0.0088	0.69
7	MC	0.5485	0.0082	1.08
8	MC	-1.0099	0.0090	0.93
9	XI	0.7866	0.0083	1.10
10	MC	0.1100	0.0081	1.14
11	MC	-0.2621	0.0082	0.96
12	XI	0.6267	0.0082	0.96
13	MC	-0.2437	0.0082	1.09
14	MC	-0.0514	0.0081	0.98
15	MC	-0.3720	0.0083	0.93
16	MC	1.4534	0.0089	1.04
17	MC	0.1955	0.0081	1.22
18	MC	0.3786	0.0081	1.19
19	MC	0.6684	0.0082	1.19
20	MC	-0.3270	0.0083	1.05
21	MX	-0.8380	0.0061	0.88
22	MX	0.4338	0.0058	1.02
23	MC	-1.2584	0.0094	0.89
24	MC	-0.7259	0.0086	1.07
25	MC	-1.5530	0.0100	0.86
26	MC	0.4526	0.0081	1.13
27	MC	-0.8311	0.0087	0.90
28	MC	0.1038	0.0081	0.97
29	XI	1.7473	0.0094	1.02
30	MC	-1.0368	0.0090	0.88
31	MC	-0.2022	0.0082	1.11
32	MC	-1.1476	0.0092	0.88
33	MC	-0.1659	0.0082	0.91
34	MC	0.1949	0.0081	1.15
35	MC	-0.5328	0.0084	1.00
36	MC	-0.3840	0.0083	0.81
37	MC	-0.2956	0.0083	1.00
38	MC	-0.4977	0.0084	0.85
39	XI	1.6608	0.0092	1.13
40	MC	-0.1890	0.0082	1.00
41	MC	0.6796	0.0082	1.07
42	MX	0.6734	0.0082	0.78
43	MC	0.0571	0.0081	0.99
44	XI	1.6298	0.0092	0.84
45	MC	0.3506	0.0081	1.14

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
46	MX	0.4899	0.0059	1.18
47	MX	0.2932	0.0060	1.30

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	0.8278	0.0072	0.80
2	OE	1.1054	0.0072	0.75
3	OE	-1.2319	0.0091	0.71
4	OE	0.8212	0.0074	0.74
5	OE	1.2152	0.0072	0.90
6	OE	-1.1895	0.0091	0.69
7	MC	-1.0482	0.0089	0.99
8	XI	-0.9560	0.0088	0.89
9	MC	-0.0021	0.0081	0.96
10	MC	1.5655	0.0091	1.09
11	MC	-0.0750	0.0082	1.00
12	MC	0.0618	0.0081	1.25
13	MC	0.3584	0.0081	1.07
14	MX	1.4367	0.0089	1.15
15	MC	0.0608	0.0081	1.01
16	MC	0.5562	0.0082	1.11
17	MC	0.5266	0.0082	1.08
18	XI	1.3475	0.0088	1.04
19	MX	0.8866	0.0084	1.09
20	MC	-1.1918	0.0091	0.93
21	MX	-0.6283	0.0065	1.15
22	MX	-0.9164	0.0060	0.95
23	MC	-0.7307	0.0086	1.00
24	MC	-0.4802	0.0083	0.95
25	MC	-0.8329	0.0087	0.88
26	MC	-0.7633	0.0086	0.96
27	MC	-0.7706	0.0086	0.87
28	MC	0.0069	0.0081	1.13
29	MC	-1.5743	0.0098	0.81
30	MC	-0.1398	0.0082	1.15
31	MC	-1.0130	0.0089	1.02
32	MX	1.1480	0.0086	0.92
33	MC	-0.7099	0.0085	0.88
34	MC	0.2209	0.0081	1.10
35	MC	0.7054	0.0082	1.07
36	MX	0.6699	0.0082	0.81
37	MC	0.3753	0.0081	1.08
38	MC	1.3048	0.0087	1.17
39	MC	-0.2514	0.0082	0.88

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
40	MX	1.1873	0.0086	0.99
41	MC	-0.4114	0.0083	0.87
42	MC	-0.1014	0.0082	1.00
43	MC	-1.7565	0.0102	0.79
44	MC	1.4613	0.0089	0.94
45	MC	-0.8377	0.0087	0.90
46	MX	0.6409	0.0060	1.28
47	MX	0.1700	0.0058	1.01

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	-0.2639	0.0089	0.86
2	MC	-1.6472	0.0110	0.84
3	XI	0.1795	0.0087	1.00
4	MC	1.1192	0.0087	1.29
5	XI	1.5809	0.0090	0.85
6	MC	-1.3577	0.0104	0.89
7	MC	0.6355	0.0086	0.90
8	XI	-1.1718	0.0100	0.81
9	XI	2.7204	0.0108	0.91
10	MC	-0.4119	0.0090	0.85
11	XI	1.3005	0.0088	1.00
12	MC	1.3892	0.0088	1.31
13	MC	-0.0458	0.0088	1.24
14	MC	-1.7818	0.0113	0.99
15	MC	1.0418	0.0087	1.03
16	XI	1.8944	0.0093	0.88
17	MC	0.1403	0.0087	1.03
18	XI	-0.0905	0.0088	1.02
19	MC	0.1213	0.0087	1.23
20	MC	-0.8415	0.0095	0.91
21	XI	-0.9194	0.0097	0.97
22	MC	-1.4722	0.0106	0.92
23	MC	-0.2121	0.0089	1.18
24	MC	-1.1807	0.0100	1.05
25	MC	-1.5958	0.0109	0.93
26	XI	-0.0977	0.0088	0.90
27	XI	-0.9228	0.0096	1.11
28	XI	-0.0985	0.0088	1.13
29	MC	0.8246	0.0086	1.40
30	MC	0.0505	0.0087	1.05
31	MC	1.1328	0.0087	0.96
32	MC	-0.4904	0.0091	1.14
33	MC	-0.5571	0.0092	0.86

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
34	MX	2.4438	0.0102	0.98
35	MC	0.9383	0.0086	1.05
36	XI	0.1386	0.0087	1.05
37	MC	-0.7664	0.0094	0.93
38	MC	0.8026	0.0086	1.15
39	XI	-0.5956	0.0092	0.79
40	XI	0.8927	0.0086	0.87
41	MC	-0.7203	0.0094	0.96
42	XI	1.6688	0.0091	0.89
43	XI	0.0258	0.0087	0.93
44	MC	0.0749	0.0087	0.80
45	MC	-1.2638	0.0102	0.97

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	-2.5830	0.0126	0.84
2	XI	0.0166	0.0089	0.93
3	XI	0.8473	0.0091	1.02
4	MX	1.2425	0.0094	0.97
5	MC	0.7612	0.0090	1.22
6	MC	0.7896	0.0090	1.32
7	MC	-1.4709	0.0099	0.97
8	XI	-0.4122	0.0090	0.92
9	XI	2.3166	0.0110	0.97
10	MC	-1.0142	0.0094	1.02
11	MC	-0.2174	0.0089	0.98
12	XI	-0.8413	0.0092	0.82
13	MC	1.2190	0.0093	0.85
14	XI	2.3683	0.0111	1.09
15	MC	0.3325	0.0089	0.92
16	MC	0.4229	0.0089	1.11
17	MC	-1.0134	0.0094	0.99
18	MC	0.6378	0.0090	0.96
19	MX	1.2709	0.0094	1.14
20	MC	-0.4650	0.0090	0.93
21	XI	-0.9408	0.0093	0.90
22	MC	0.5118	0.0089	0.99
23	XI	-1.1063	0.0095	0.82
24	MC	-1.8328	0.0106	0.99
25	MC	-0.7164	0.0091	1.06
26	MC	-0.3345	0.0089	1.10
27	XI	1.5507	0.0097	0.88
28	MC	0.4931	0.0089	0.75
29	XI	0.5519	0.0089	0.84

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
30	MC	-1.0363	0.0094	1.01
31	MX	2.3051	0.0109	0.90
32	XI	-0.5353	0.0090	0.84
33	MC	-0.5146	0.0090	0.88
34	MC	-1.9653	0.0109	0.94
35	MC	0.8444	0.0091	1.27
36	XI	1.9322	0.0102	0.97
37	MC	-0.3783	0.0089	1.26
38	XI	-0.8734	0.0092	0.92
39	MC	0.5401	0.0089	1.12
40	MC	-0.4970	0.0090	1.08
41	XI	0.1922	0.0089	1.07
42	MC	1.1943	0.0093	1.22
43	XI	1.0134	0.0092	0.74
44	MC	-0.9157	0.0093	0.93
45	XI	0.5007	0.0089	0.93

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	-1.4127	0.0091	1.13
2	XI	-1.4512	0.0092	0.95
3	MC	-0.2939	0.0083	1.04
4	XI	0.3865	0.0085	0.74
5	MC	-0.1505	0.0083	1.04
6	MC	-0.7043	0.0084	0.98
7	MX	0.8991	0.0089	0.97
8	MC	-1.7325	0.0096	0.86
9	XI	2.1702	0.0110	0.96
10	MC	0.9218	0.0089	1.00
11	MC	-0.6665	0.0084	0.94
12	MC	0.1520	0.0084	1.09
13	XI	-0.0237	0.0083	0.88
14	XI	0.7644	0.0087	1.06
15	MC	1.6376	0.0099	1.28
16	XI	1.0037	0.0090	0.88
17	MC	-0.9310	0.0086	0.85
18	XI	0.8420	0.0088	0.83
19	XI	-0.3399	0.0083	0.99
20	MC	-0.6270	0.0084	1.14
21	MC	0.7844	0.0087	1.11
22	XI	1.1739	0.0092	0.95
23	MC	-0.2750	0.0083	1.05
24	MC	-1.1279	0.0088	0.92
25	XI	0.2134	0.0084	0.91

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
26	MC	1.7428	0.0101	1.10
27	MC	-0.8369	0.0085	1.22
28	MC	-0.3931	0.0083	0.89
29	MC	0.2411	0.0084	1.19
30	XI	-0.1651	0.0083	0.93
31	XI	1.8458	0.0103	0.94
32	XI	-0.9479	0.0086	0.94
33	XI	-0.0006	0.0083	1.16
34	XI	0.5781	0.0086	1.00
35	MC	-0.8912	0.0086	0.94
36	MC	-0.4555	0.0083	1.01
37	MC	-0.4936	0.0084	1.06
38	XI	1.5369	0.0097	0.77
39	MC	-0.3789	0.0083	1.14
40	MC	0.1392	0.0084	1.22
41	MC	-0.2759	0.0083	1.06
42	XI	1.1418	0.0091	0.81
43	MC	0.9331	0.0089	1.17
44	XI	1.2276	0.0092	1.00
45	XI	-1.0945	0.0087	0.87

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	-2.5380	0.0106	0.89
2	XI	-0.5565	0.0084	0.80
3	MC	-1.3975	0.0087	1.26
4	XI	-0.0515	0.0085	0.96
5	MC	-1.3801	0.0087	1.06
6	MC	1.4946	0.0104	1.08
7	MC	-0.5048	0.0084	0.74
8	XI	0.2963	0.0087	0.79
9	XI	0.1530	0.0086	0.91
10	MC	0.3106	0.0087	0.95
11	MC	0.5102	0.0089	1.26
12	XI	1.8102	0.0112	0.93
13	MC	0.7981	0.0093	1.23
14	XI	1.1415	0.0098	0.87
15	MC	0.0521	0.0085	1.33
16	XI	0.8510	0.0093	0.95
17	MC	0.3665	0.0088	1.32
18	XI	0.3752	0.0088	0.88
19	MC	0.3428	0.0088	0.93
20	MC	0.0743	0.0086	1.19
21	XI	-0.4724	0.0084	0.94

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
22	MC	-0.2308	0.0084	1.07
23	MC	-0.3239	0.0084	1.21
24	MC	-1.2664	0.0086	1.09
25	MC	-2.1729	0.0097	0.95
26	XI	-0.4093	0.0084	0.92
27	XI	0.6351	0.0091	0.83
28	MC	-0.6260	0.0083	1.09
29	XI	-0.4709	0.0084	0.95
30	MC	-0.1403	0.0084	0.92
31	XI	0.4728	0.0089	1.14
32	MC	1.6405	0.0108	1.18
33	MC	-1.7254	0.0090	1.06
34	MX	0.2060	0.0087	1.23
35	MC	-0.7462	0.0084	1.12
36	XI	0.2496	0.0087	0.80
37	XI	-0.6890	0.0084	0.81
38	MC	0.4962	0.0089	0.97
39	XI	-1.3310	0.0086	0.94
40	MC	1.2218	0.0099	1.05
41	XI	1.6132	0.0107	0.91
42	MC	-0.8449	0.0084	1.03
43	XI	0.3913	0.0088	0.88
44	MC	-1.2546	0.0086	0.81
45	MC	0.5808	0.0090	0.91
46	XI	0.9057	0.0094	0.82
47	MC	-1.3719	0.0086	0.95

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	-1.9141	0.0094	0.97
2	MC	-1.6556	0.0090	0.87
3	XI	1.0077	0.0094	0.97
4	XI	0.2559	0.0086	0.90
5	XI	0.1972	0.0085	0.99
6	MC	1.1595	0.0097	0.94
7	MC	0.1296	0.0085	1.20
8	MX	1.4380	0.0102	1.16
9	MC	1.5711	0.0105	1.07
10	XI	-0.1984	0.0083	0.86
11	MC	0.4638	0.0087	1.21
12	MC	-0.4811	0.0082	1.05
13	MC	-1.0690	0.0084	0.92
14	MC	-0.4913	0.0082	1.07
15	XI	0.7670	0.0091	0.81

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
16	MC	0.4523	0.0087	1.05
17	XI	1.1041	0.0096	0.83
18	MC	-0.6091	0.0082	1.06
19	XI	0.6388	0.0089	0.83
20	MC	-1.2652	0.0085	0.92
21	XI	2.0962	0.0118	0.84
22	MC	0.5512	0.0088	1.05
23	MC	-1.1736	0.0085	0.91
24	MC	-1.2477	0.0085	0.98
25	MC	-0.5221	0.0082	1.13
26	XI	1.1444	0.0096	0.92
27	MC	-0.1871	0.0083	0.89
28	MC	0.1140	0.0084	1.04
29	MC	-0.3800	0.0082	1.03
30	XI	0.6330	0.0089	0.98
31	MC	0.0133	0.0084	1.04
32	XI	-0.2158	0.0083	1.06
33	MC	-0.4477	0.0082	1.04
34	XI	-0.4507	0.0082	1.02
35	MC	1.6997	0.0107	1.12
36	MC	0.0701	0.0084	1.10
37	XI	1.0104	0.0094	1.01
38	MC	1.2220	0.0098	1.20
39	MC	-0.5408	0.0082	1.09
40	XI	1.5495	0.0104	0.84
41	MC	0.1385	0.0085	0.98
42	MC	0.2225	0.0085	0.81
43	MC	-1.6474	0.0090	1.10
44	MC	0.4226	0.0087	1.01
45	XI	-0.0928	0.0083	0.91
46	MC	0.0306	0.0084	1.16
47	MC	-0.5062	0.0082	0.98

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	-1.3503	0.0081	0.98
2	MC	-0.9719	0.0080	1.12
3	XI	0.1888	0.0086	0.81
4	XI	1.2775	0.0105	0.86
5	MC	0.8711	0.0096	1.06
6	MC	-0.5165	0.0080	0.91
7	MC	-0.8414	0.0080	0.90
8	XI	0.6984	0.0093	0.78
9	MX	0.2859	0.0087	0.94

Item Number	Item Type	Rasch Difficulty	SE	MNSQ Infit
10	MC	-2.2891	0.0094	0.89
11	MC	-0.2369	0.0082	1.10
12	MC	-1.3151	0.0081	1.06
13	XI	-0.5013	0.0080	0.88
14	XI	1.4017	0.0108	1.03
15	MC	0.0801	0.0085	1.00
16	MC	-0.2307	0.0082	1.08
17	MX	0.9553	0.0098	1.16
18	MC	-0.5309	0.0080	0.87
19	MC	-1.6808	0.0084	0.90
20	XI	0.2188	0.0086	1.11
21	MC	0.9956	0.0098	1.05
22	MC	0.2109	0.0086	1.18
23	MC	-0.6997	0.0080	1.05
24	MC	-1.4719	0.0082	1.06
25	XI	-0.5994	0.0080	1.14
26	MC	-0.3731	0.0081	1.14
27	XI	1.2197	0.0103	0.78
28	MC	-0.5404	0.0080	0.92
29	MC	-0.9019	0.0080	1.09
30	XI	1.0153	0.0099	1.12
31	MC	-0.7845	0.0080	1.01
32	MC	0.3889	0.0088	1.05
33	XI	-0.6190	0.0080	0.86
34	MC	-0.2171	0.0082	0.92
35	XI	-0.2976	0.0081	0.95
36	MC	-0.5192	0.0080	0.98
37	XI	1.1450	0.0102	0.96
38	MC	-0.5141	0.0080	1.13
39	MC	-0.6959	0.0080	1.00
40	MC	-0.3360	0.0081	0.97
41	MC	-0.3631	0.0081	0.98
42	MC	1.0026	0.0099	1.14
43	XI	0.3671	0.0088	0.85
44	MC	0.0964	0.0085	0.98
45	MX	0.1438	0.0085	1.03
46	MC	0.5740	0.0091	0.92
47	MC	-0.1966	0.0082	1.08

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	20	1
1	3	2395	20	1
1	4	2395	20	1
1	5	2400	18	1
1	6	2410	16	1
1	7	2418	15	1
1	8	2425	14	1
1	9	2431	13	1
1	10	2436	12	1
1	11	2441	12	1
1	12	2445	11	1
1	13	2449	11	1
1	14	2453	11	1
1	15	2457	10	1
1	16	2460	10	1
1	17	2464	10	1
1	18	2467	10	1
1	19	2470	9	1
1	20	2473	9	1
1	21	2476	9	1
1	22	2478	9	1
1	23	2481	9	1
1	24	2484	9	1
1	25	2486	9	1
1	26	2489	9	1
1	27	2492	9	1
1	28	2494	9	1
1	29	2497	9	2
1	30	2500	9	2
1	31	2502	9	2
1	32	2505	9	2
1	33	2507	9	2
1	34	2510	9	3
1	35	2513	9	3
1	36	2515	9	3
1	37	2518	9	3
1	38	2521	9	3
1	39	2524	9	3
1	40	2527	9	3
1	41	2530	10	3
1	42	2533	10	3
1	43	2536	10	3
1	44	2539	10	3
1	45	2543	10	4
1	46	2546	11	4
1	47	2550	11	4
1	48	2554	11	4
1	49	2559	12	4

Form	Raw Score	Scale Score	CSEM	Performance Level
1	50	2564	12	4
1	51	2569	13	4
1	52	2575	14	4
1	53	2581	15	4
1	54	2589	16	4
1	55	2598	17	4
1	56	2605	19	4
1	57	2605	19	4
1	58	2605	19	4
1	59	2605	19	4
2	2	2395	20	1
2	3	2395	20	1
2	4	2395	20	1
2	5	2401	19	1
2	6	2411	16	1
2	7	2419	15	1
2	8	2426	14	1
2	9	2432	13	1
2	10	2437	12	1
2	11	2442	12	1
2	12	2446	11	1
2	13	2450	11	1
2	14	2454	11	1
2	15	2458	10	1
2	16	2461	10	1
2	17	2464	10	1
2	18	2468	10	1
2	19	2471	9	1
2	20	2473	9	1
2	21	2476	9	1
2	22	2479	9	1
2	23	2482	9	1
2	24	2484	9	1
2	25	2487	9	1
2	26	2490	9	1
2	27	2492	9	1
2	28	2495	9	1
2	29	2497	9	2
2	30	2500	9	2
2	31	2503	9	2
2	32	2505	9	2
2	33	2508	9	2
2	34	2510	9	3
2	35	2513	9	3
2	36	2516	9	3
2	37	2518	9	3
2	38	2521	9	3
2	39	2524	9	3
2	40	2527	9	3

Form	Raw Score	Scale Score	CSEM	Performance Level
2	41	2530	10	3
2	42	2533	10	3
2	43	2536	10	3
2	44	2541	10	4
2	45	2543	10	4
2	46	2547	11	4
2	47	2551	11	4
2	48	2555	12	4
2	49	2560	12	4
2	50	2565	13	4
2	51	2571	13	4
2	52	2577	14	4
2	53	2585	16	4
2	54	2594	18	4
2	55	2605	20	4
2	56	2605	20	4
2	57	2605	20	4
2	58	2605	20	4
2	59	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	2407	22	1
1	5	2420	18	1
1	6	2430	16	1
1	7	2437	14	1
1	8	2444	13	1
1	9	2449	13	1
1	10	2454	12	1
1	11	2459	11	1
1	12	2463	11	1
1	13	2467	11	1
1	14	2471	10	1
1	15	2474	10	1
1	16	2477	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	2501	9	1
1	25	2504	9	1
1	26	2507	9	1
1	27	2510	9	2
1	28	2512	9	2

Form	Raw Score	Scale Score	CSEM	Performance Level
1	29	2515	9	2
1	30	2518	9	2
1	31	2521	9	2
1	32	2524	9	3
1	33	2527	9	3
1	34	2530	9	3
1	35	2533	9	3
1	36	2536	10	3
1	37	2539	10	3
1	38	2542	10	3
1	39	2545	10	3
1	40	2549	10	3
1	41	2552	10	3
1	42	2556	11	3
1	43	2560	11	4
1	44	2564	11	4
1	45	2568	12	4
1	46	2573	12	4
1	47	2578	13	4
1	48	2583	13	4
1	49	2589	14	4
1	50	2596	15	4
1	51	2604	16	4
1	52	2610	18	4
1	53	2610	18	4
1	54	2610	18	4
1	55	2610	18	4
2	2	2400	24	1
2	3	2400	24	1
2	4	2407	22	1
2	5	2420	18	1
2	6	2429	16	1
2	7	2437	14	1
2	8	2443	13	1
2	9	2449	13	1
2	10	2454	12	1
2	11	2458	11	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	2483	10	1
2	19	2486	9	1
2	20	2489	9	1
2	21	2492	9	1
2	22	2495	9	1
2	23	2498	9	1

Form	Raw Score	Scale Score	CSEM	Performance Level
2	24	2501	9	1
2	25	2503	9	1
2	26	2506	9	1
2	27	2509	9	1
2	28	2511	9	2
2	29	2514	9	2
2	30	2517	9	2
2	31	2520	9	2
2	32	2523	9	3
2	33	2525	9	3
2	34	2528	9	3
2	35	2531	9	3
2	36	2534	10	3
2	37	2537	10	3
2	38	2540	10	3
2	39	2544	10	3
2	40	2547	10	3
2	41	2551	10	3
2	42	2554	11	3
2	43	2559	11	4
2	44	2562	11	4
2	45	2567	12	4
2	46	2572	12	4
2	47	2577	13	4
2	48	2583	14	4
2	49	2589	15	4
2	50	2597	16	4
2	51	2606	17	4
2	52	2610	18	4
2	53	2610	18	4
2	54	2610	18	4
2	55	2610	18	4

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

Form	Raw Score	Scale Score	CSEM	Performance Level
1	2	2419	23	1
1	3	2419	23	1
1	4	2421	22	1
1	5	2434	18	1
1	6	2444	16	1
1	7	2452	15	1
1	8	2459	14	1
1	9	2464	13	1
1	10	2469	12	1
1	11	2474	12	1
1	12	2478	11	1
1	13	2482	11	1
1	14	2486	10	1
1	15	2490	10	1

Appendix B: Item-Level IRT Statistics

Form	Raw Score	Scale Score	CSEM	Performance Level
1	16	2493	10	1
1	17	2496	10	1
1	18	2499	10	1
1	19	2503	10	1
1	20	2506	9	1
1	21	2508	9	1
1	22	2511	9	1
1	23	2514	9	1
1	24	2517	9	1
1	25	2520	9	2
1	26	2523	9	2
1	27	2525	9	2
1	28	2528	9	2
1	29	2531	9	2
1	30	2534	9	2
1	31	2537	9	2
1	32	2539	9	2
1	33	2543	9	3
1	34	2545	9	3
1	35	2548	9	3
1	36	2551	10	3
1	37	2554	10	3
1	38	2558	10	3
1	39	2561	10	3
1	40	2564	10	3
1	41	2568	10	3
1	42	2572	11	3
1	43	2576	11	3
1	44	2580	11	4
1	45	2584	12	4
1	46	2589	12	4
1	47	2595	13	4
1	48	2601	14	4
1	49	2607	15	4
1	50	2615	16	4
1	51	2625	18	4
1	52	2629	19	4
1	53	2629	19	4
1	54	2629	19	4
1	55	2629	19	4
2	2	2419	22	1
2	3	2419	22	1
2	4	2421	22	1
2	5	2434	18	1
2	6	2444	16	1
2	7	2451	15	1
2	8	2458	14	1
2	9	2464	13	1
2	10	2469	12	1

Form	Raw Score	Scale Score	CSEM	Performance Level
2	11	2473	12	1
2	12	2478	11	1
2	13	2482	11	1
2	14	2485	10	1
2	15	2489	10	1
2	16	2492	10	1
2	17	2496	10	1
2	18	2499	10	1
2	19	2502	10	1
2	20	2505	9	1
2	21	2508	9	1
2	22	2511	9	1
2	23	2513	9	1
2	24	2516	9	1
2	25	2520	9	2
2	26	2522	9	2
2	27	2525	9	2
2	28	2527	9	2
2	29	2530	9	2
2	30	2533	9	2
2	31	2536	9	2
2	32	2538	9	2
2	33	2541	9	2
2	34	2544	9	3
2	35	2547	9	3
2	36	2550	10	3
2	37	2553	10	3
2	38	2557	10	3
2	39	2560	10	3
2	40	2563	10	3
2	41	2567	10	3
2	42	2571	11	3
2	43	2575	11	3
2	44	2579	11	4
2	45	2583	12	4
2	46	2588	12	4
2	47	2593	13	4
2	48	2599	14	4
2	49	2605	14	4
2	50	2613	16	4
2	51	2622	17	4
2	52	2629	19	4
2	53	2629	19	4
2	54	2629	19	4
2	55	2629	19	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	2444	18	1
1	6	2454	16	1
1	7	2462	14	1
1	8	2468	13	1
1	9	2474	13	1
1	10	2479	12	1
1	11	2483	11	1
1	12	2487	11	1
1	13	2491	11	1
1	14	2495	10	1
1	15	2498	10	1
1	16	2502	10	1
1	17	2505	10	1
1	18	2508	10	1
1	19	2511	9	1
1	20	2514	9	1
1	21	2517	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	2535	9	2
1	29	2538	9	2
1	30	2541	9	2
1	31	2544	9	2
1	32	2546	9	2
1	33	2549	9	2
1	34	2553	9	3
1	35	2555	9	3
1	36	2558	9	3
1	37	2561	10	3
1	38	2564	10	3
1	39	2567	10	3
1	40	2571	10	3
1	41	2574	10	3
1	42	2578	11	3
1	43	2582	11	3
1	44	2586	11	3
1	45	2590	12	3
1	46	2597	12	4
1	47	2600	13	4
1	48	2605	13	4
1	49	2612	14	4

Form	Raw Score	Scale Score	CSEM	Performance Level
1	50	2619	15	4
1	51	2627	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
2	4	2432	22	1
2	5	2445	18	1
2	6	2455	16	1
2	7	2463	14	1
2	8	2469	13	1
2	9	2475	13	1
2	10	2480	12	1
2	11	2484	11	1
2	12	2488	11	1
2	13	2492	11	1
2	14	2496	10	1
2	15	2500	10	1
2	16	2503	10	1
2	17	2506	10	1
2	18	2509	10	1
2	19	2512	9	1
2	20	2515	9	1
2	21	2518	9	1
2	22	2521	9	1
2	23	2524	9	1
2	24	2526	9	1
2	25	2529	9	1
2	26	2532	9	2
2	27	2535	9	2
2	28	2537	9	2
2	29	2540	9	2
2	30	2543	9	2
2	31	2546	9	2
2	32	2548	9	2
2	33	2551	9	2
2	34	2554	9	3
2	35	2557	9	3
2	36	2560	10	3
2	37	2563	10	3
2	38	2566	10	3
2	39	2570	10	3
2	40	2573	10	3
2	41	2577	10	3
2	42	2580	11	3
2	43	2584	11	3
2	44	2588	11	3

Form	Raw Score	Scale Score	CSEM	Performance Level
2	45	2593	12	3
2	46	2598	12	4
2	47	2603	13	4
2	48	2608	13	4
2	49	2615	14	4
2	50	2622	15	4
2	51	2630	17	4
2	52	2640	19	4
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	2440	22	1
1	5	2453	18	1
1	6	2463	16	1
1	7	2470	14	1
1	8	2477	13	1
1	9	2482	13	1
1	10	2487	12	1
1	11	2492	11	1
1	12	2496	11	1
1	13	2500	11	1
1	14	2504	10	1
1	15	2507	10	1
1	16	2510	10	1
1	17	2513	10	1
1	18	2517	10	1
1	19	2520	9	1
1	20	2523	9	1
1	21	2525	9	1
1	22	2528	9	1
1	23	2531	9	1
1	24	2534	9	1
1	25	2536	9	1
1	26	2539	9	1
1	27	2543	9	2
1	28	2545	9	2
1	29	2547	9	2
1	30	2550	9	2
1	31	2553	9	2
1	32	2556	9	2
1	33	2558	9	2
1	34	2561	9	3
1	35	2564	9	3
1	36	2567	10	3

Form	Raw Score	Scale Score	CSEM	Performance Level
1	37	2570	10	3
1	38	2574	10	3
1	39	2577	10	3
1	40	2580	10	3
1	41	2584	10	3
1	42	2587	11	3
1	43	2591	11	3
1	44	2595	11	3
1	45	2600	12	4
1	46	2605	12	4
1	47	2610	13	4
1	48	2615	13	4
1	49	2622	14	4
1	50	2629	15	4
1	51	2638	17	4
1	52	2648	19	4
1	53	2648	19	4
1	54	2648	19	4
1	55	2648	19	4
2	2	2438	23	1
2	3	2438	23	1
2	4	2442	22	1
2	5	2455	18	1
2	6	2464	16	1
2	7	2472	14	1
2	8	2478	13	1
2	9	2483	12	1
2	10	2488	12	1
2	11	2493	11	1
2	12	2497	11	1
2	13	2501	11	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	2523	9	1
2	21	2525	9	1
2	22	2528	9	1
2	23	2531	9	1
2	24	2534	9	1
2	25	2536	9	1
2	26	2539	9	1
2	27	2541	9	1
2	28	2544	9	2
2	29	2547	9	2
2	30	2549	9	2
2	31	2552	9	2

Form	Raw Score	Scale Score	CSEM	Performance Level
2	32	2555	9	2
2	33	2557	9	2
2	34	2561	9	3
2	35	2563	9	3
2	36	2566	9	3
2	37	2568	9	3
2	38	2571	10	3
2	39	2575	10	3
2	40	2578	10	3
2	41	2581	10	3
2	42	2584	10	3
2	43	2588	11	3
2	44	2592	11	3
2	45	2596	11	3
2	46	2600	12	4
2	47	2605	12	4
2	48	2610	13	4
2	49	2616	14	4
2	50	2623	15	4
2	51	2630	16	4
2	52	2640	18	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	2458	18	1
1	6	2468	16	1
1	7	2476	15	1
1	8	2482	13	1
1	9	2488	13	1
1	10	2493	12	1
1	11	2498	12	1
1	12	2502	11	1
1	13	2506	11	1
1	14	2510	10	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

Form	Raw Score	Scale Score	CSEM	Performance Level
1	24	2541	9	1
1	25	2544	9	1
1	26	2546	9	1
1	27	2549	9	1
1	28	2552	9	2
1	29	2555	9	2
1	30	2557	9	2
1	31	2560	9	2
1	32	2563	9	2
1	33	2566	9	2
1	34	2569	9	2
1	35	2572	9	3
1	36	2575	9	3
1	37	2578	10	3
1	38	2581	10	3
1	39	2584	10	3
1	40	2587	10	3
1	41	2591	10	3
1	42	2594	10	3
1	43	2598	11	3
1	44	2602	11	3
1	45	2606	11	4
1	46	2610	12	4
1	47	2615	12	4
1	48	2621	13	4
1	49	2626	14	4
1	50	2633	15	4
1	51	2641	16	4
1	52	2651	18	4
1	53	2658	20	4
1	54	2658	20	4
1	55	2658	20	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	13	1
2	9	2489	13	1
2	10	2494	12	1
2	11	2498	12	1
2	12	2503	11	1
2	13	2507	11	1
2	14	2510	10	1
2	15	2514	10	1
2	16	2517	10	1
2	17	2521	10	1
2	18	2524	10	1

Form	Raw Score	Scale Score	CSEM	Performance Level
2	19	2527	10	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
2	27	2550	9	1
2	28	2552	9	2
2	29	2555	9	2
2	30	2558	9	2
2	31	2561	9	2
2	32	2563	9	2
2	33	2566	9	2
2	34	2569	9	2
2	35	2572	9	3
2	36	2575	9	3
2	37	2578	10	3
2	38	2581	10	3
2	39	2584	10	3
2	40	2587	10	3
2	41	2591	10	3
2	42	2594	10	3
2	43	2598	11	3
2	44	2602	11	3
2	45	2606	11	4
2	46	2610	12	4
2	47	2615	12	4
2	48	2620	13	4
2	49	2626	14	4
2	50	2633	15	4
2	51	2641	16	4
2	52	2651	18	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	3396	22	1
3	3409	18	1
4	3419	16	1
5	3427	15	1
6	3434	14	1
7	3440	13	1
8	3446	12	1

Raw Score	Scale Score	CSEM	Performance Level
9	3451	12	1
10	3455	12	1
11	3460	11	1
12	3464	11	1
13	3468	11	1
14	3471	11	1
15	3475	10	1
16	3479	10	1
17	3482	10	1
18	3486	10	1
19	3489	10	1
20	3492	10	1
21	3496	10	2
22	3499	10	2
23	3503	10	2
24	3506	10	2
25	3509	10	2
26	3513	10	2
27	3516	10	2
28	3520	10	2
29	3523	11	2
30	3527	11	2
31	3531	11	3
32	3535	11	3
33	3539	11	3
34	3543	12	3
35	3548	12	3
36	3553	12	3
37	3558	13	3
38	3564	13	3
39	3573	14	4
40	3578	15	4
41	3586	17	4
42	3597	19	4
43	3605	21	4
44	3605	21	4
45	3605	21	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	19	1
4	3458	17	1
5	3466	15	1
6	3473	14	1
7	3480	13	1
8	3485	13	1

Raw Score	Scale Score	CSEM	Performance Level
9	3491	12	1
10	3495	12	1
11	3500	11	1
12	3504	11	1
13	3508	11	1
14	3512	11	1
15	3516	11	1
16	3520	10	1
17	3523	10	1
18	3527	10	1
19	3530	10	2
20	3534	10	2
21	3537	10	2
22	3541	10	2
23	3544	10	2
24	3548	10	2
25	3551	10	2
26	3554	10	2
27	3558	10	2
28	3562	10	3
29	3565	11	3
30	3569	11	3
31	3573	11	3
32	3577	11	3
33	3581	11	3
34	3585	12	3
35	3590	12	3
36	3595	12	3
37	3600	13	3
38	3606	13	4
39	3612	14	4
40	3619	15	4
41	3627	17	4
42	3638	19	4
43	3645	20	4
44	3645	20	4
45	3645	20	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	3483	22	1
3	3496	18	1
4	3506	16	1
5	3514	15	1
6	3520	14	1
7	3526	13	1
8	3532	12	1

Raw Score	Scale Score	CSEM	Performance Level
9	3537	12	1
10	3541	11	1
11	3545	11	1
12	3549	11	1
13	3553	11	1
14	3557	10	1
15	3560	10	1
16	3564	10	2
17	3567	10	2
18	3571	10	2
19	3574	10	2
20	3577	10	2
21	3580	10	2
22	3584	10	2
23	3587	10	2
24	3590	10	2
25	3593	10	2
26	3597	10	3
27	3600	10	3
28	3603	10	3
29	3607	10	3
30	3610	10	3
31	3614	11	3
32	3618	11	3
33	3622	11	3
34	3626	11	3
35	3630	12	3
36	3635	12	4
37	3640	13	4
38	3645	13	4
39	3652	14	4
40	3658	15	4
41	3667	16	4
42	3677	19	4
43	3688	21	4
44	3688	21	4
45	3688	21	4

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

Raw Score	Scale Score	CSEM	Performance Level
0	3512	21	1
1	3512	21	1
2	3512	21	1
3	3521	19	1
4	3531	16	1
5	3540	15	1
6	3547	14	1
7	3553	13	1
8	3558	13	1

Raw Score	Scale Score	CSEM	Performance Level
9	3563	12	1
10	3568	12	1
11	3572	11	1
12	3576	11	1
13	3580	11	1
14	3584	10	1
15	3588	10	1
16	3591	10	1
17	3595	10	1
18	3598	10	1
19	3602	10	2
20	3604	10	2
21	3607	10	2
22	3611	10	2
23	3614	10	2
24	3617	10	2
25	3620	10	2
26	3623	10	2
27	3626	10	2
28	3629	10	3
29	3632	10	3
30	3636	10	3
31	3639	10	3
32	3642	10	3
33	3646	10	3
34	3649	11	3
35	3653	11	3
36	3657	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	3707	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	3533	22	1
3	3546	18	1
4	3556	16	1
5	3564	15	1
6	3571	14	1
7	3577	13	1

Raw Score	Scale Score	CSEM	Performance Level
8	3582	12	1
9	3587	12	1
10	3592	11	1
11	3596	11	1
12	3600	11	1
13	3604	11	1
14	3607	10	1
15	3611	10	1
16	3614	10	1
17	3617	10	1
18	3621	10	1
19	3624	10	1
20	3627	10	1
21	3630	10	2
22	3633	10	2
23	3636	10	2
24	3639	10	2
25	3642	10	2
26	3645	10	2
27	3648	10	2
28	3652	10	3
29	3655	10	3
30	3658	10	3
31	3661	10	3
32	3665	10	3
33	3668	10	3
34	3672	11	3
35	3675	11	3
36	3680	11	4
37	3684	11	4
38	3688	12	4
39	3693	12	4
40	3698	13	4
41	3704	14	4
42	3711	15	4
43	3719	16	4
44	3729	18	4
45	3739	21	4
46	3739	21	4
47	3739	21	4

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

Raw Score	Scale Score	CSEM	Performance Level
0	3566	22	1
1	3566	22	1
2	3566	22	1
3	3577	18	1
4	3587	16	1
5	3595	15	1

Raw Score	Scale Score	CSEM	Performance Level
6	3602	14	1
7	3608	13	1
8	3613	12	1
9	3618	12	1
10	3622	11	1
11	3626	11	1
12	3630	11	1
13	3634	10	1
14	3637	10	1
15	3641	10	1
16	3644	10	1
17	3647	10	1
18	3650	10	2
19	3653	10	2
20	3656	9	2
21	3659	9	2
22	3662	9	2
23	3665	9	2
24	3668	9	2
25	3671	9	2
26	3674	9	3
27	3677	10	3
28	3680	10	3
29	3683	10	3
30	3686	10	3
31	3690	10	3
32	3693	10	3
33	3696	10	3
34	3700	10	3
35	3703	11	3
36	3707	11	4
37	3711	11	4
38	3716	12	4
39	3721	12	4
40	3726	13	4
41	3732	14	4
42	3738	15	4
43	3746	16	4
44	3756	18	4
45	3769	22	4
46	3776	24	4
47	3776	24	4

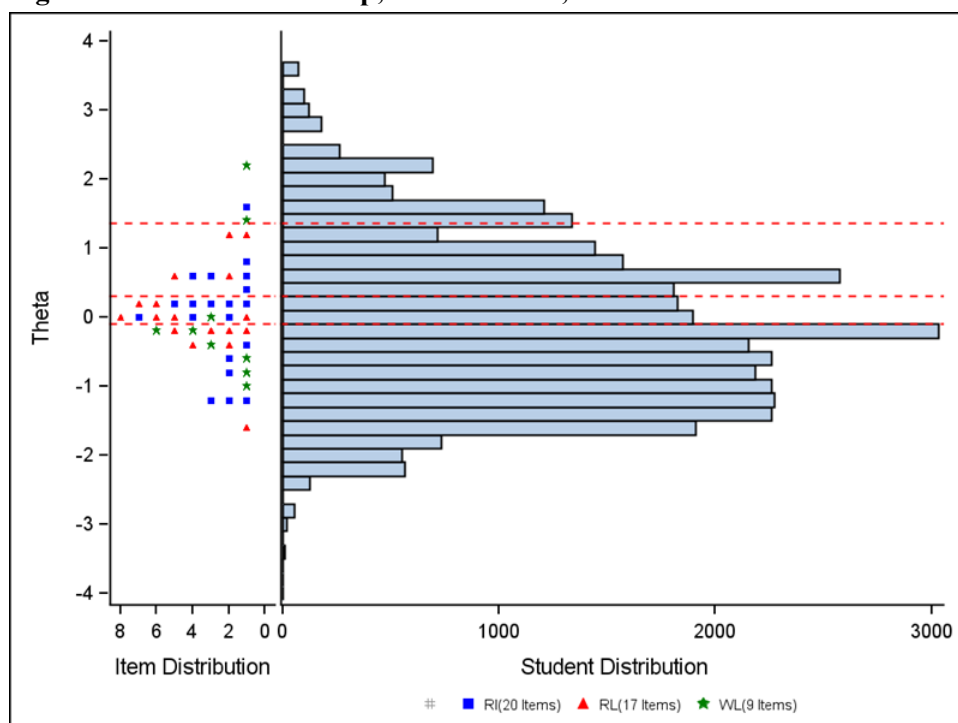
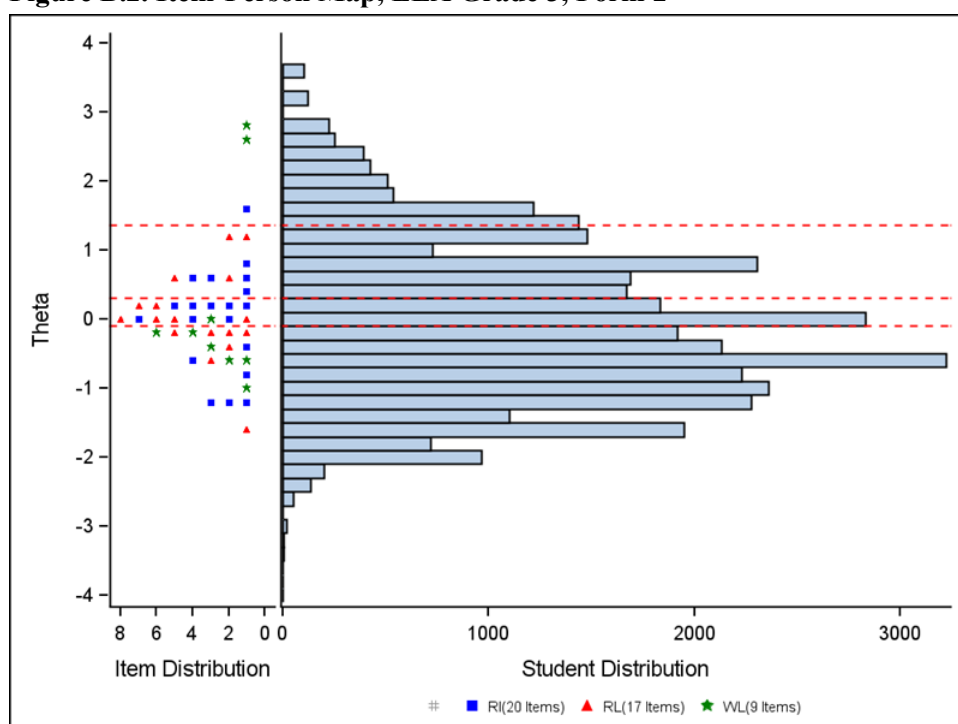
Figure B.1. Item-Person Map, ELA Grade 3, Form 1**Figure B.2. Item-Person Map, ELA Grade 3, Form 2**

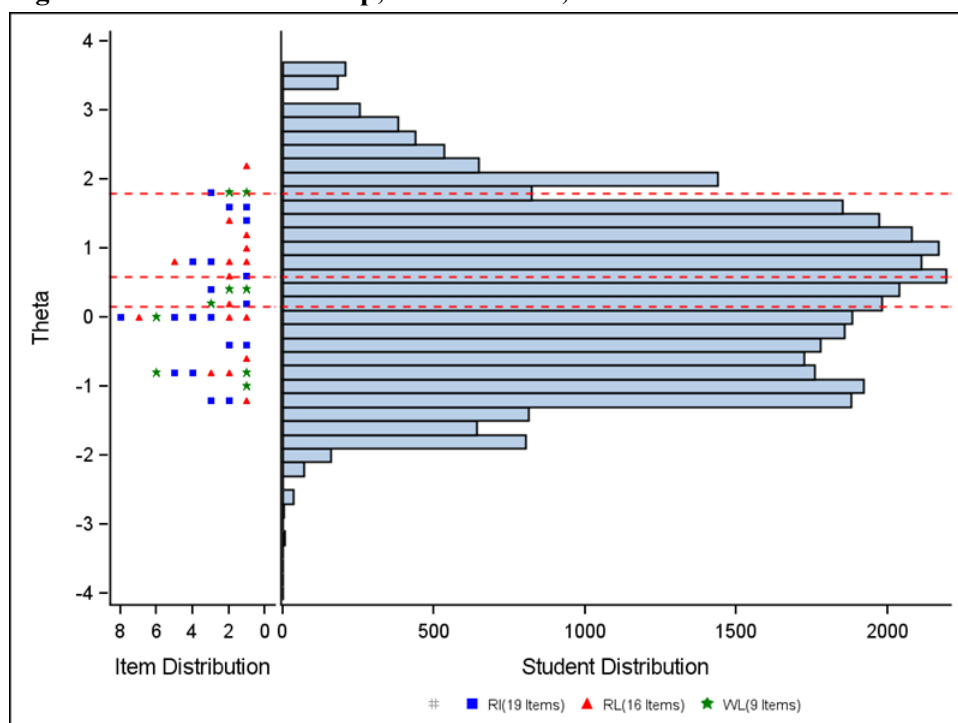
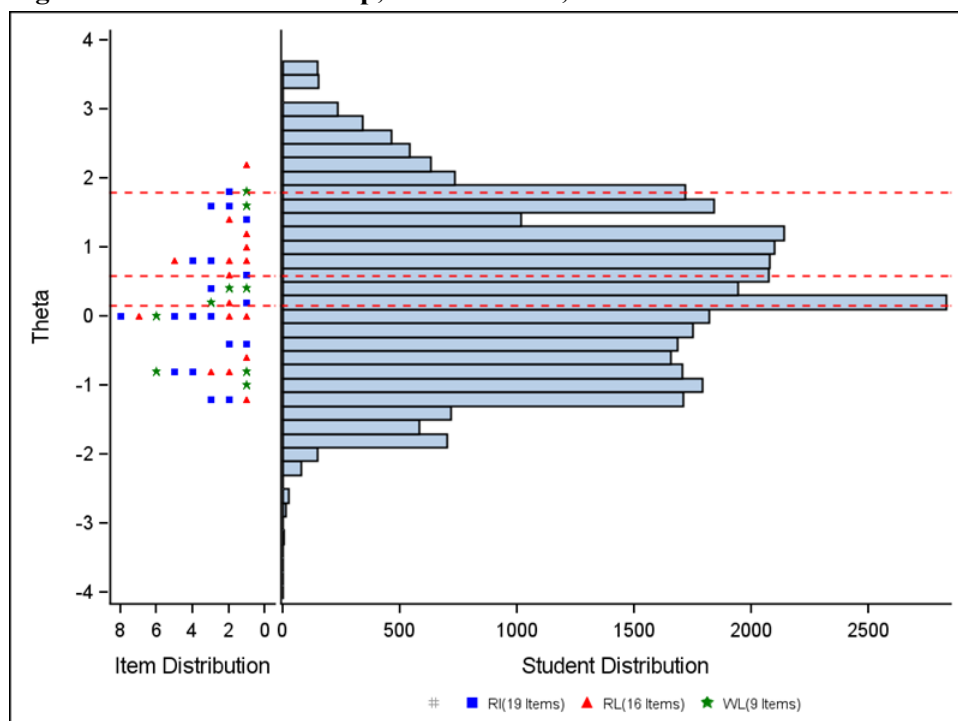
Figure B.3. Item-Person Map, ELA Grade 4, Form 1**Figure B.4. Item-Person Map, ELA Grade 4, Form 2**

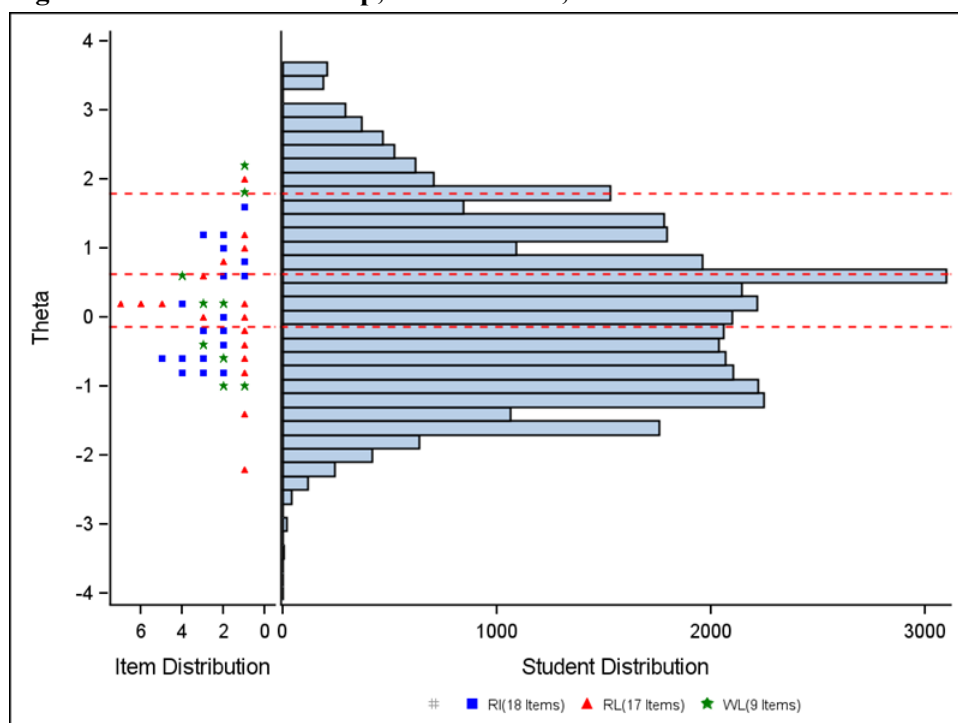
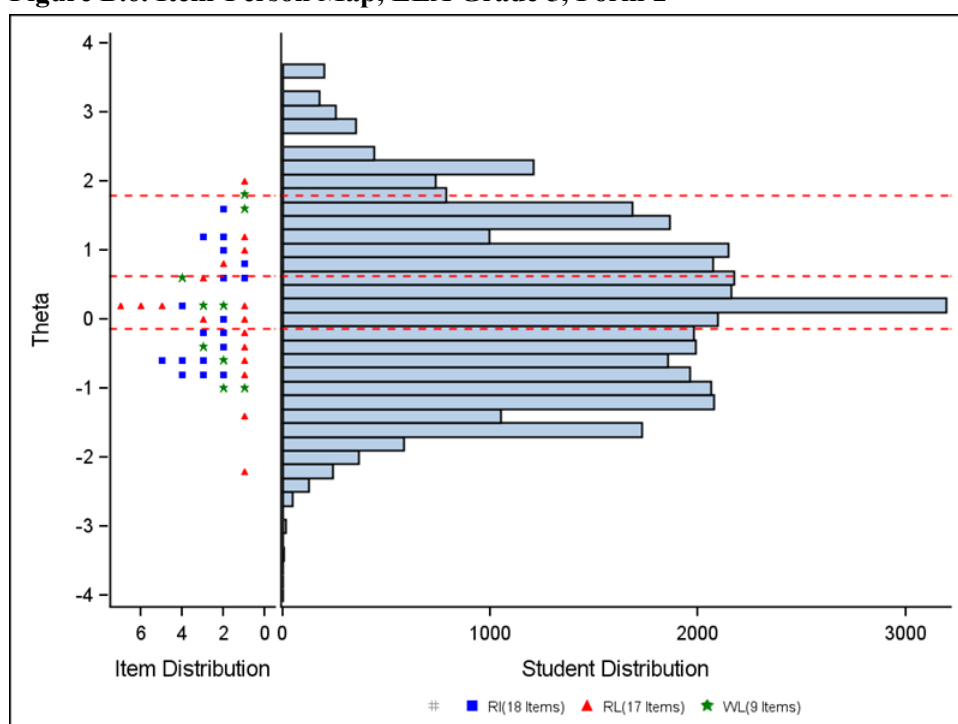
Figure B.5. Item-Person Map, ELA Grade 5, Form 1**Figure B.6. Item-Person Map, ELA Grade 5, Form 2**

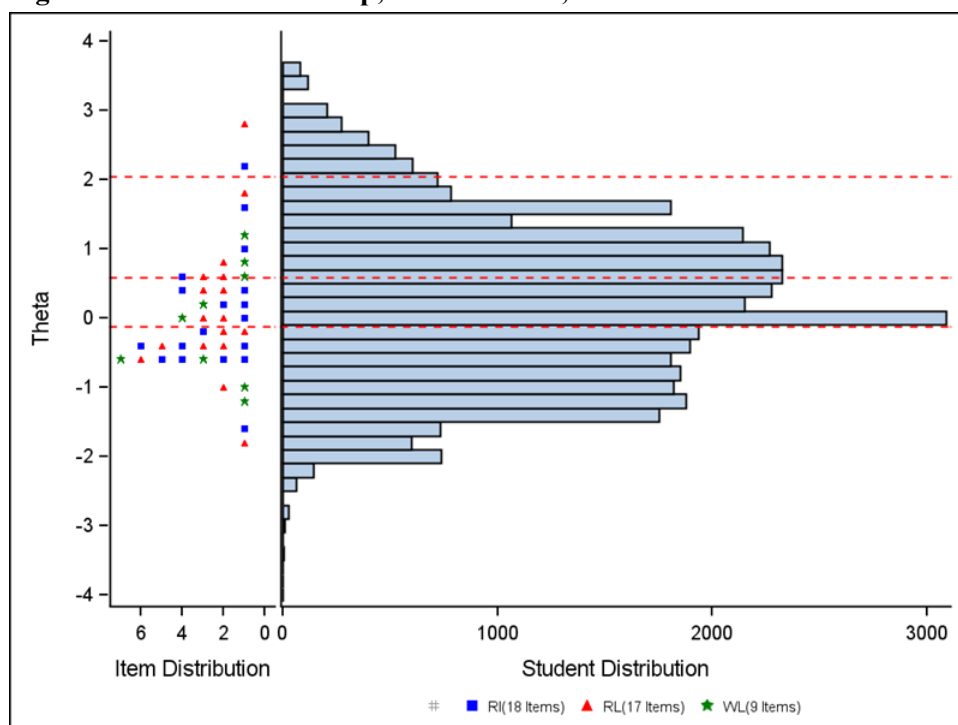
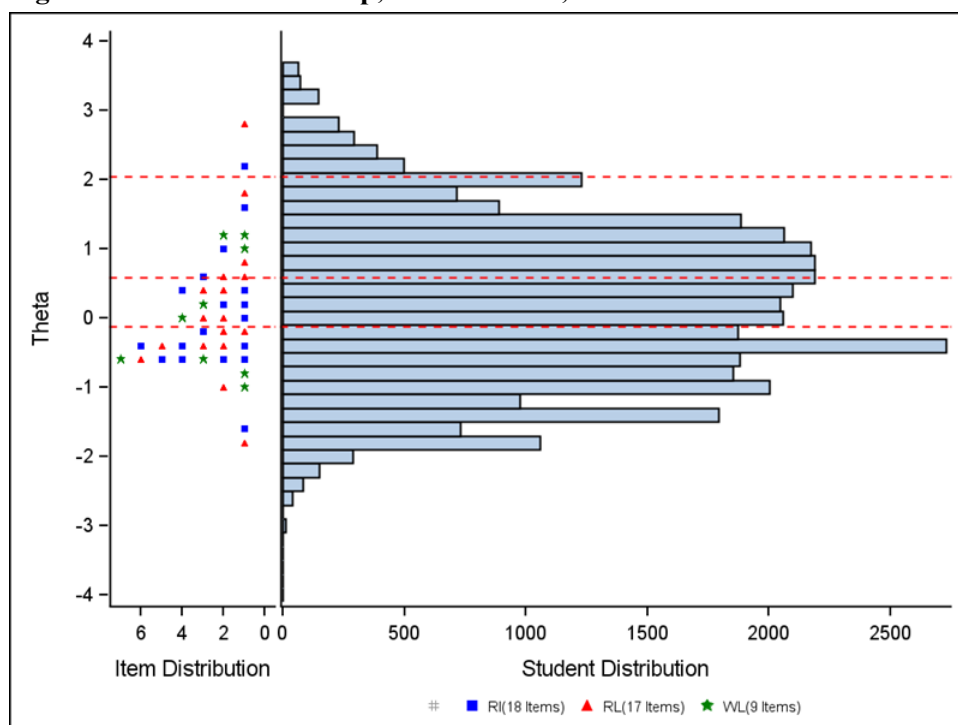
Figure B.7. Item-Person Map, ELA Grade 6, Form 1**Figure B.8. Item-Person Map, ELA Grade 6, Form 2**

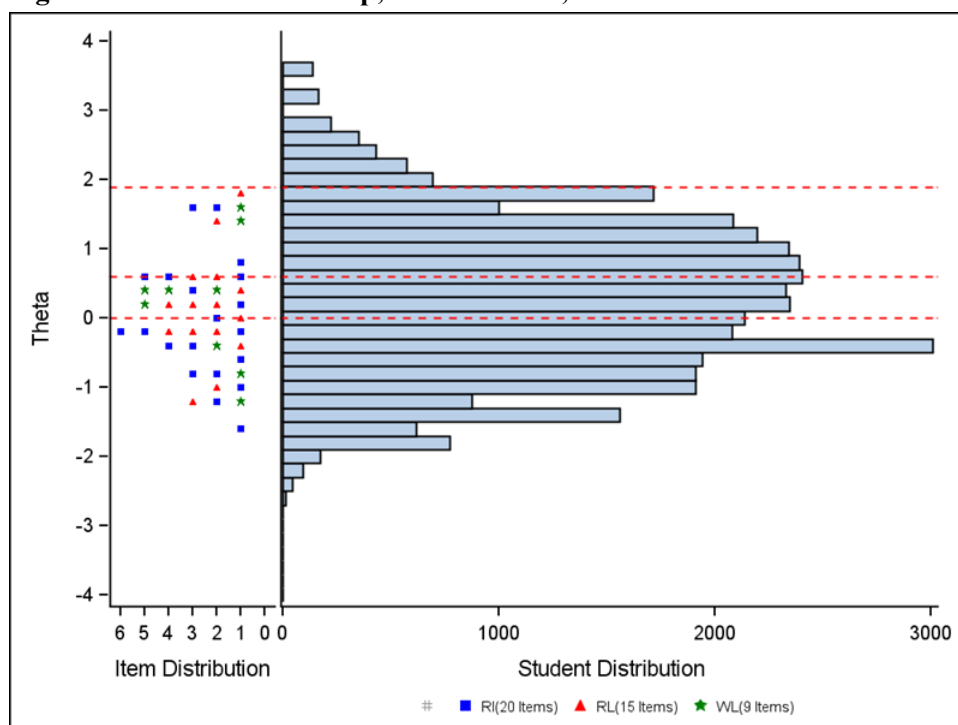
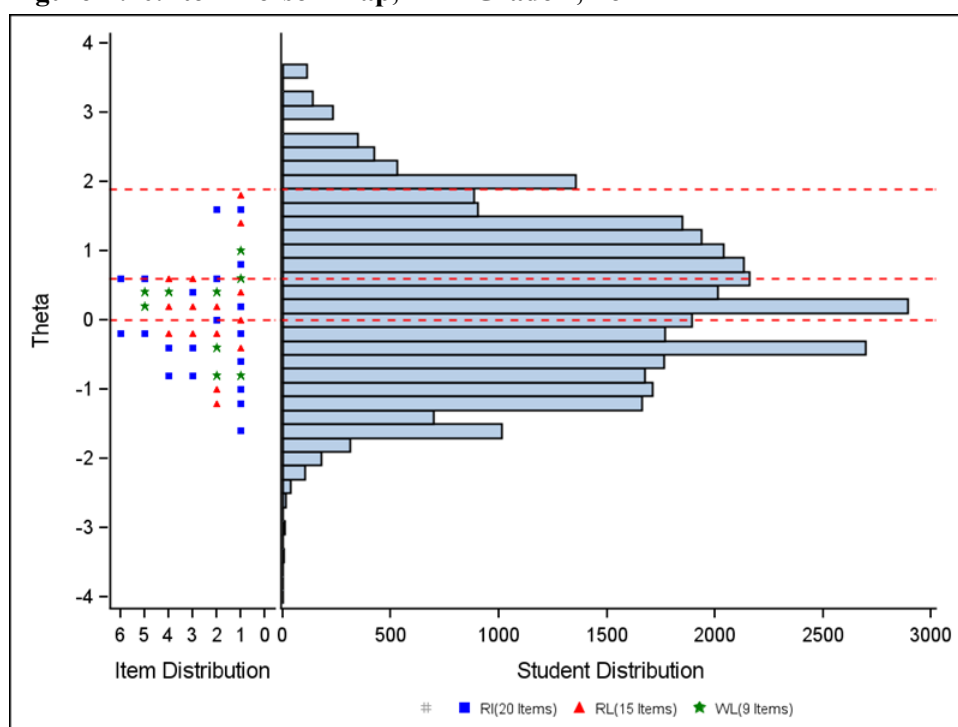
Figure B.9. Item-Person Map, ELA Grade 7, Form 1**Figure B.10. Item-Person Map, ELA Grade 7, Form 2**

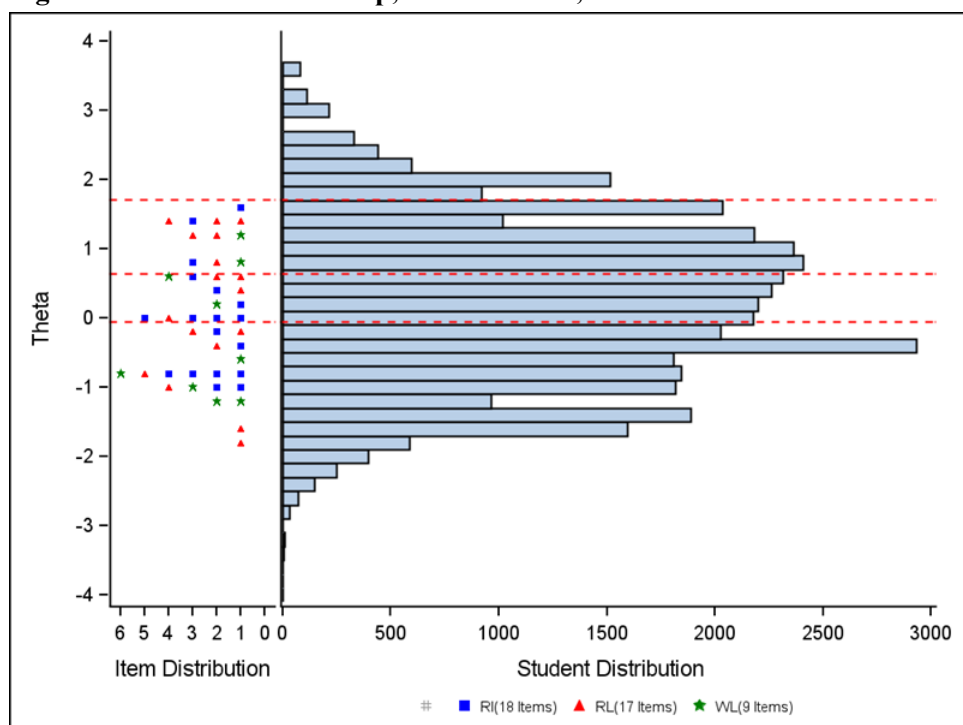
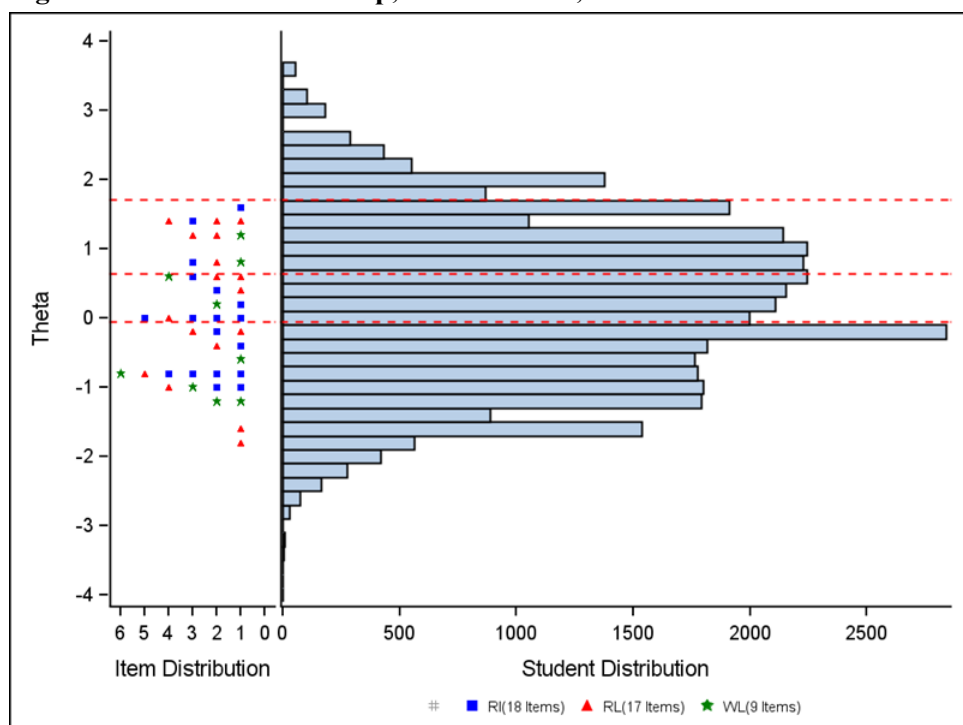
Figure B.11. Item-Person Map, ELA Grade 8, Form 1**Figure B.12. Item-Person Map, ELA Grade 8, Form 2**

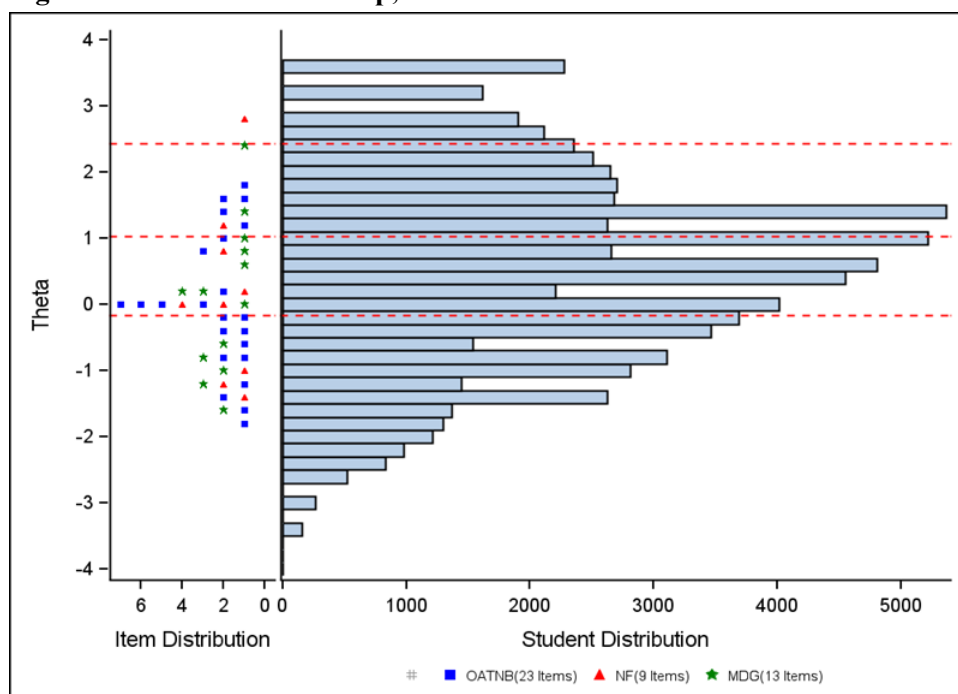
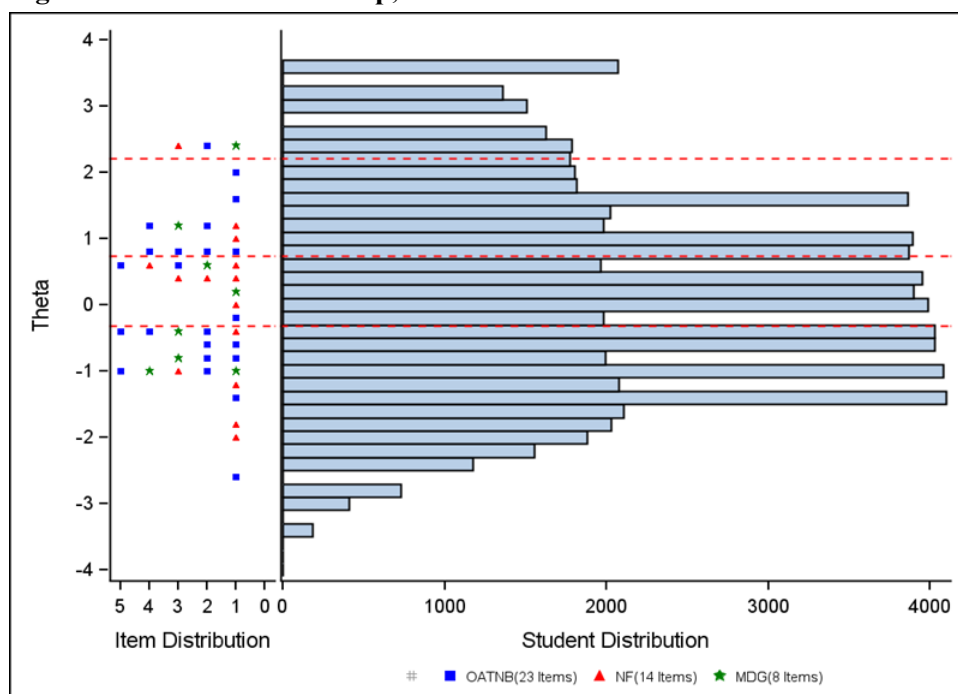
Figure B.13. Item-Person Map, Mathematics Grade 3**Figure B.14. Item-Person Map, Mathematics Grade 4**

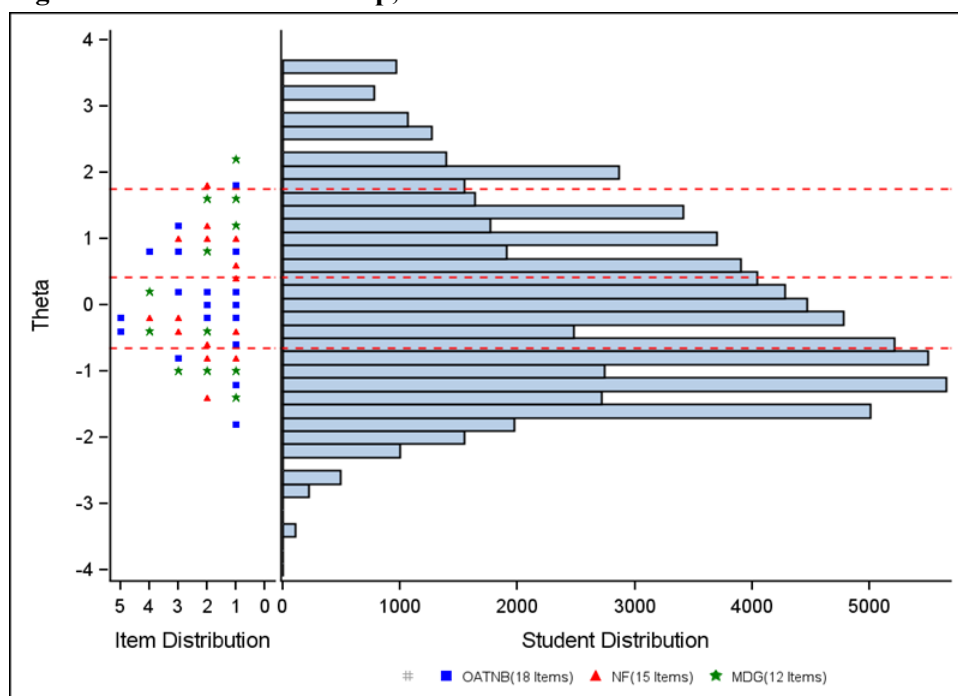
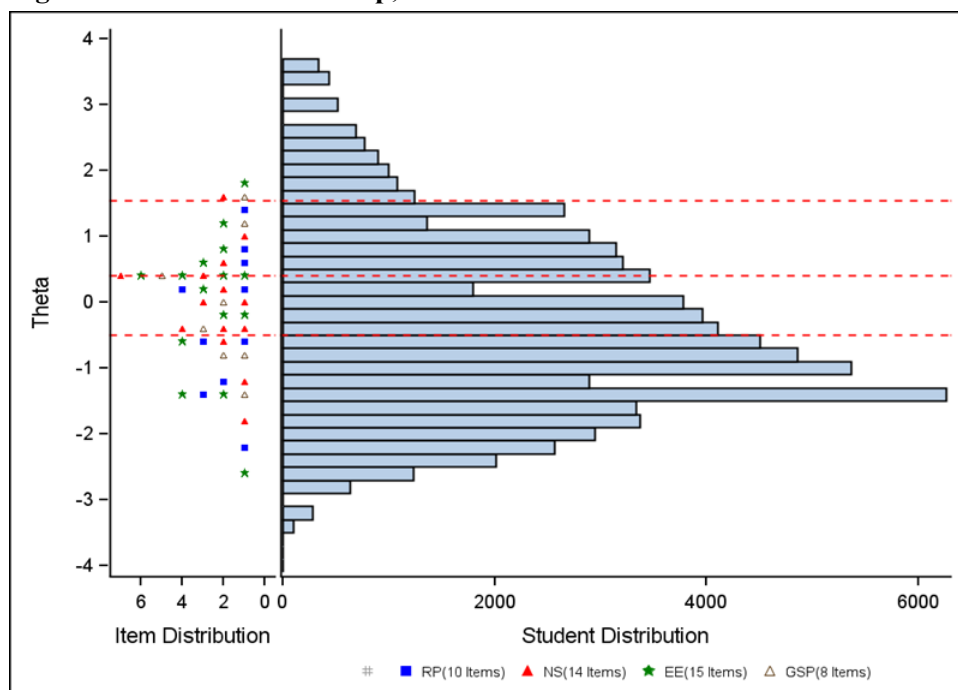
Figure B.15. Item-Person Map, Mathematics Grade 5**Figure B.16. Item-Person Map, Mathematics Grade 6**

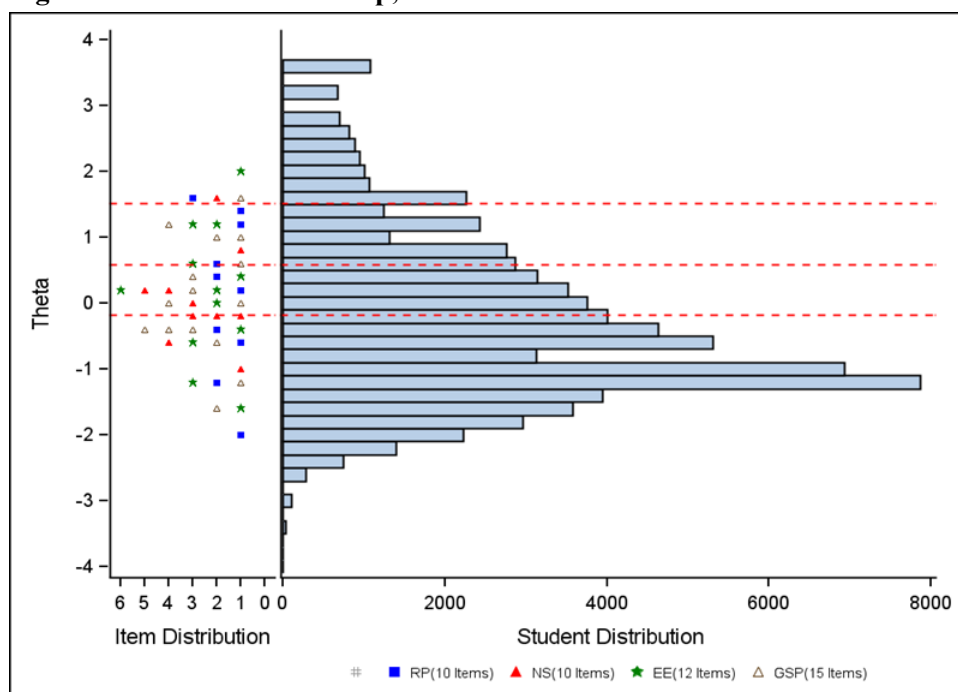
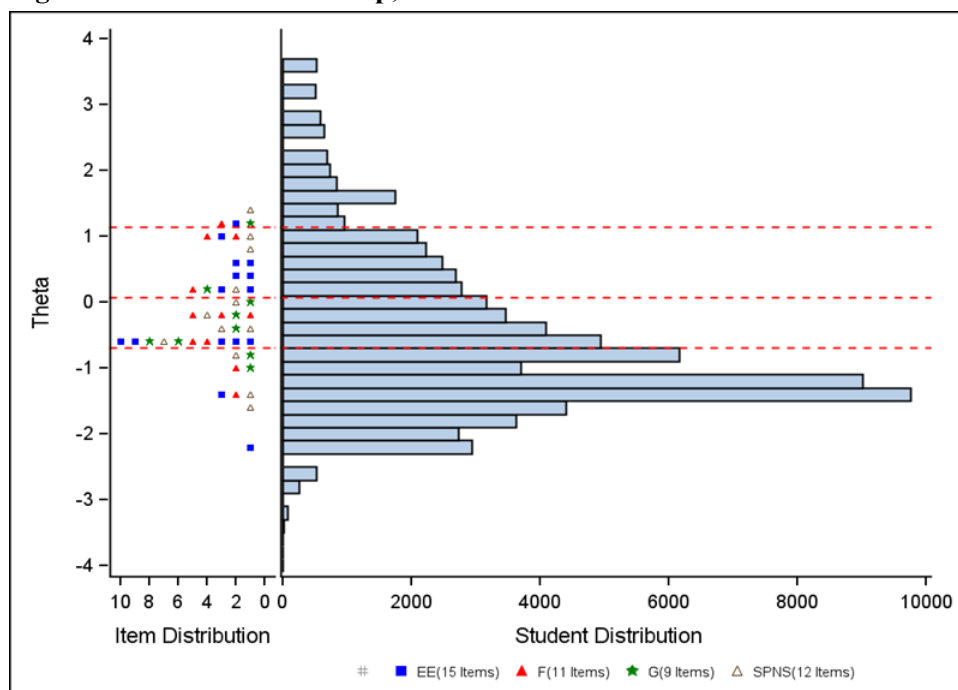
Figure B.17. Item-Person Map, Mathematics Grade 7**Figure B.18. Item-Person Map, Mathematics Grade 8**

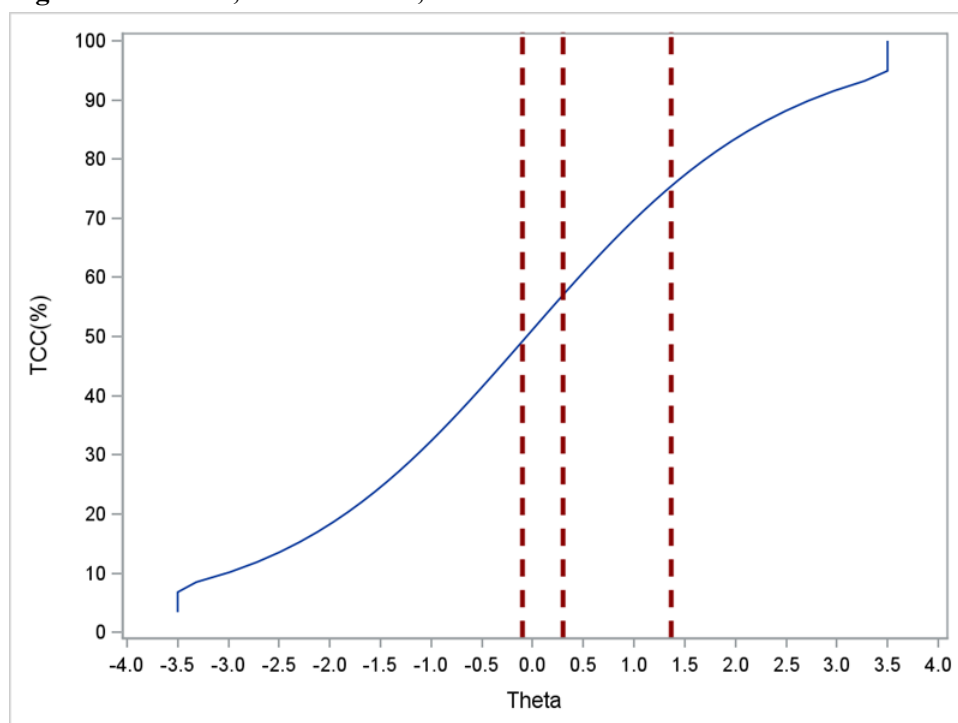
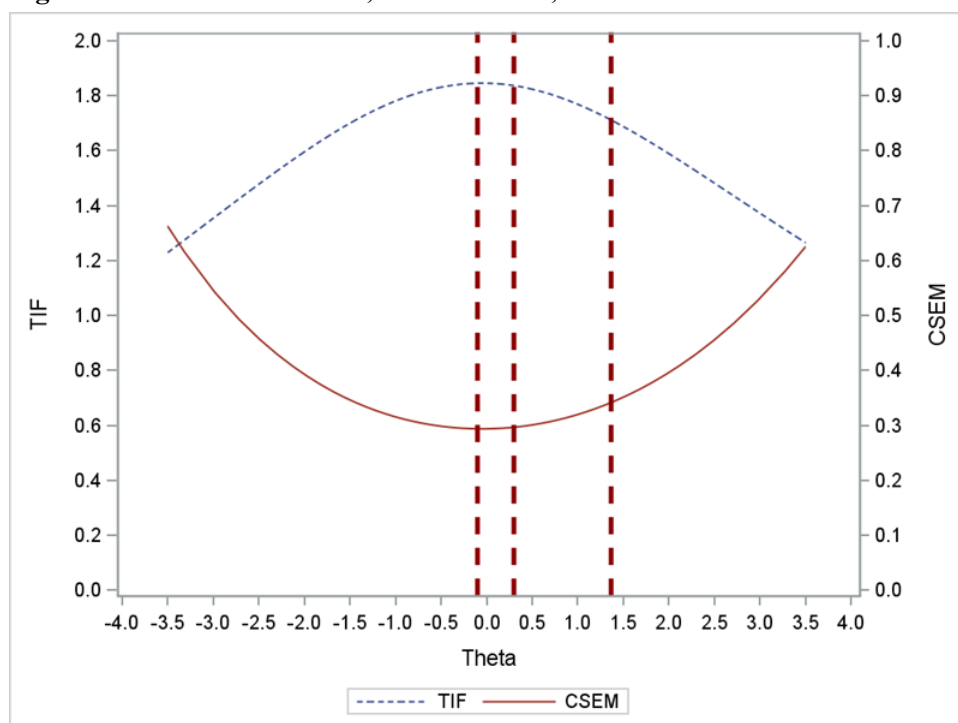
Figure B.19. TCC, ELA Grade 3, Form 1**Figure B.20. TIF and CSEM, ELA Grade 3, Form 1**

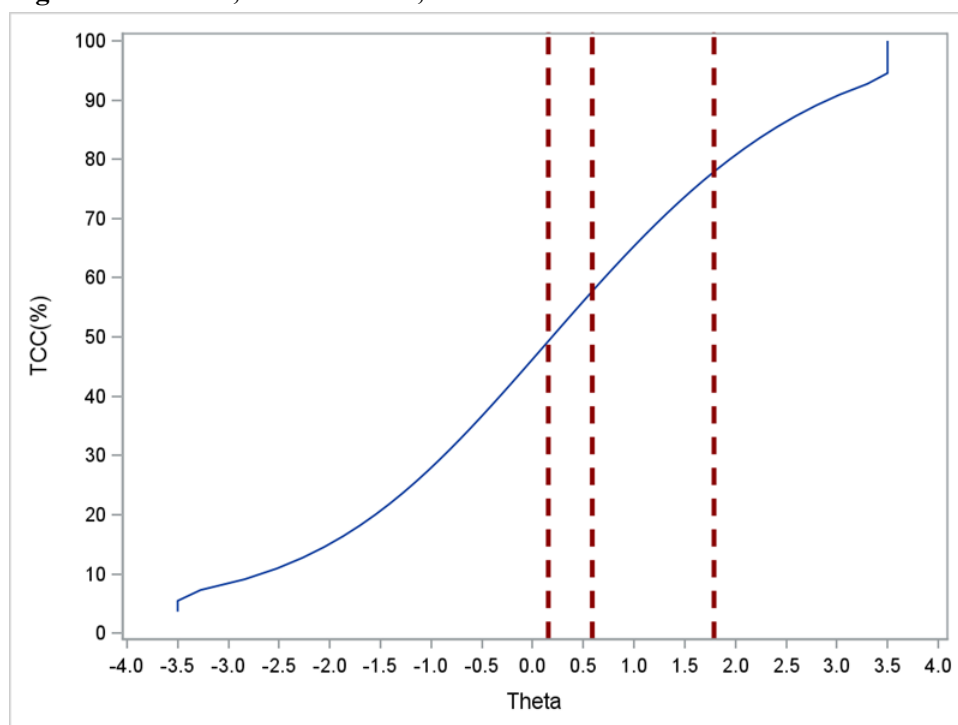
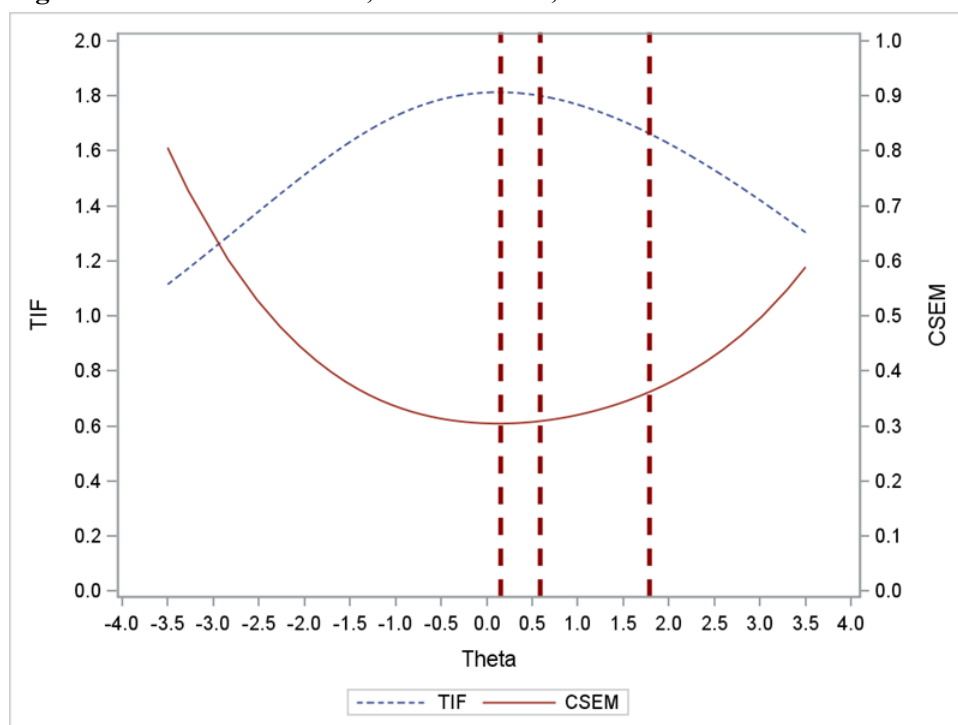
Figure B.21. TCC, ELA Grade 4, Form 1**Figure B.22. TIF and CSEM, ELA Grade 4, Form 1**

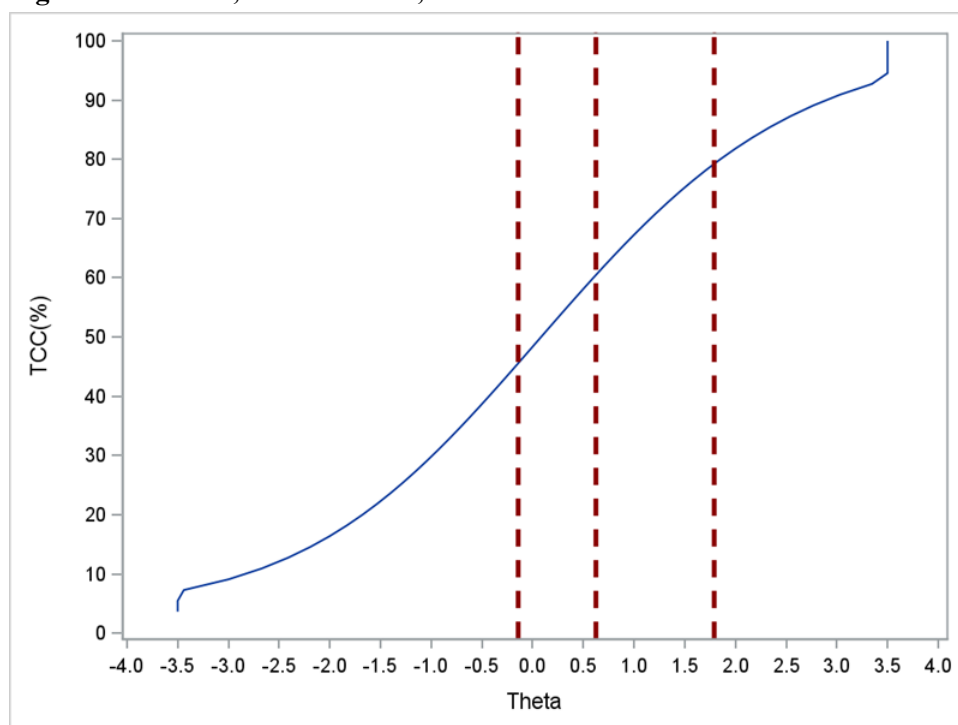
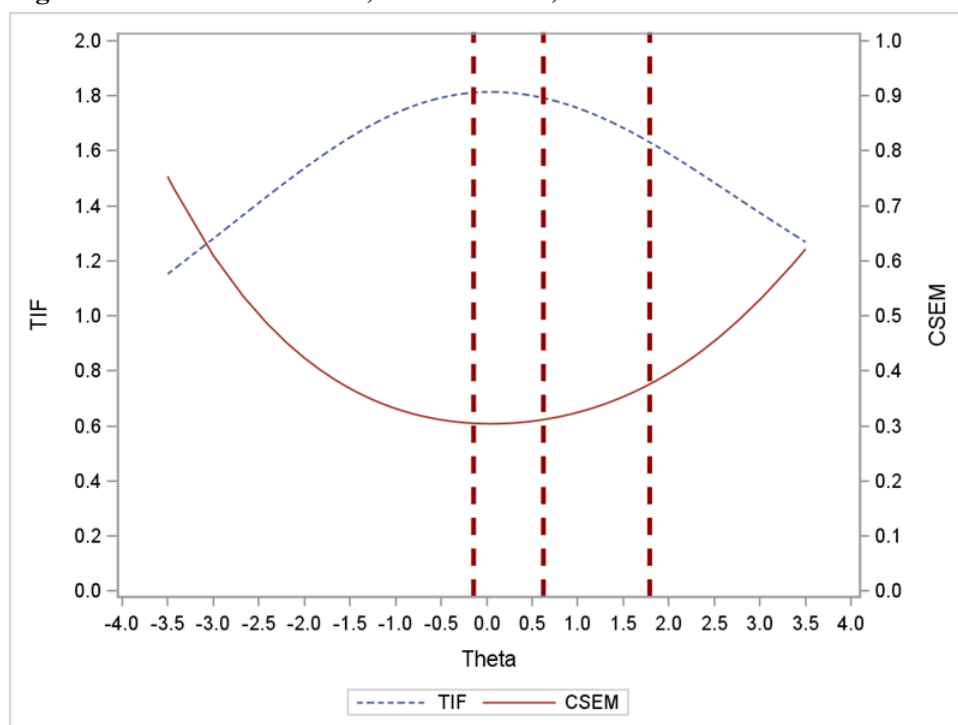
Figure B.23. TCC, ELA Grade 5, Form 1**Figure B.24. TIF and CSEM, ELA Grade 5, Form 1**

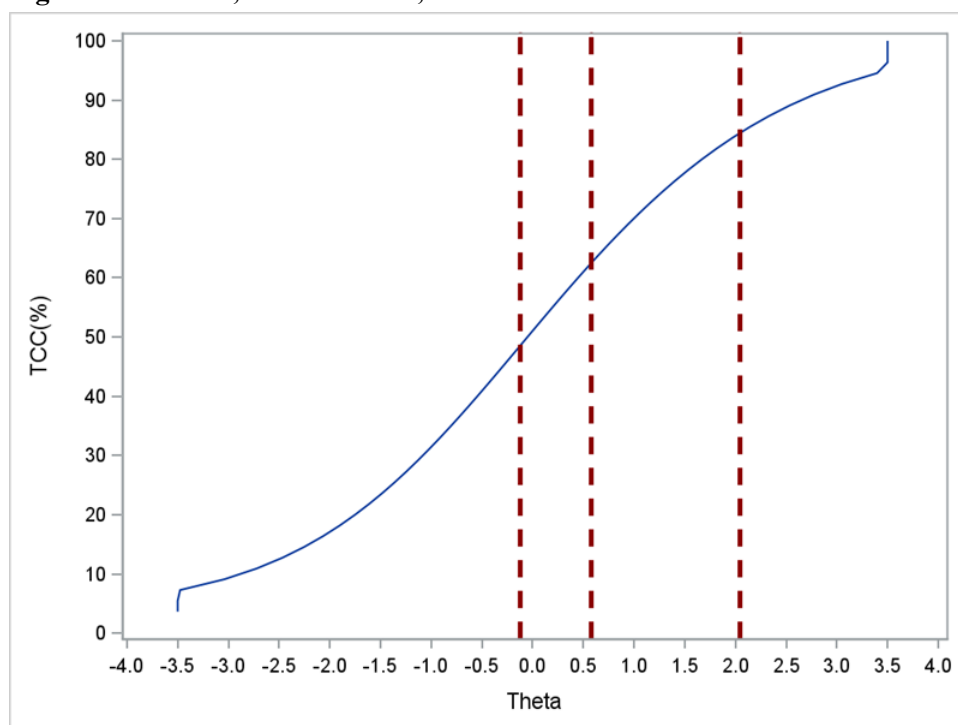
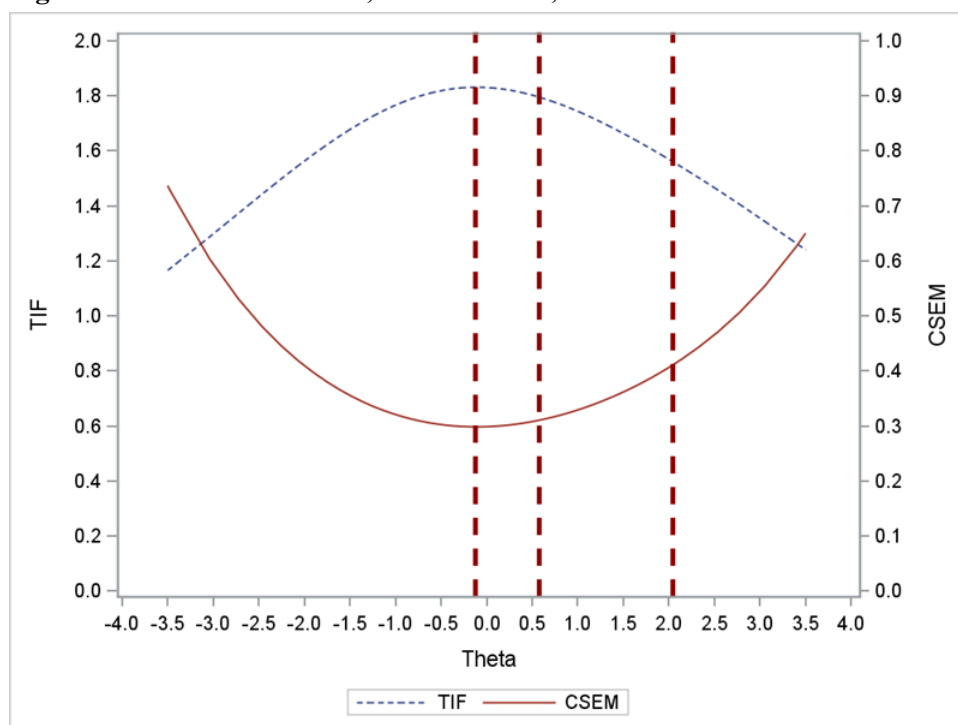
Figure B.25. TCC, ELA Grade 6, Form 1**Figure B.26. TIF and CSEM, ELA Grade 6, Form 1**

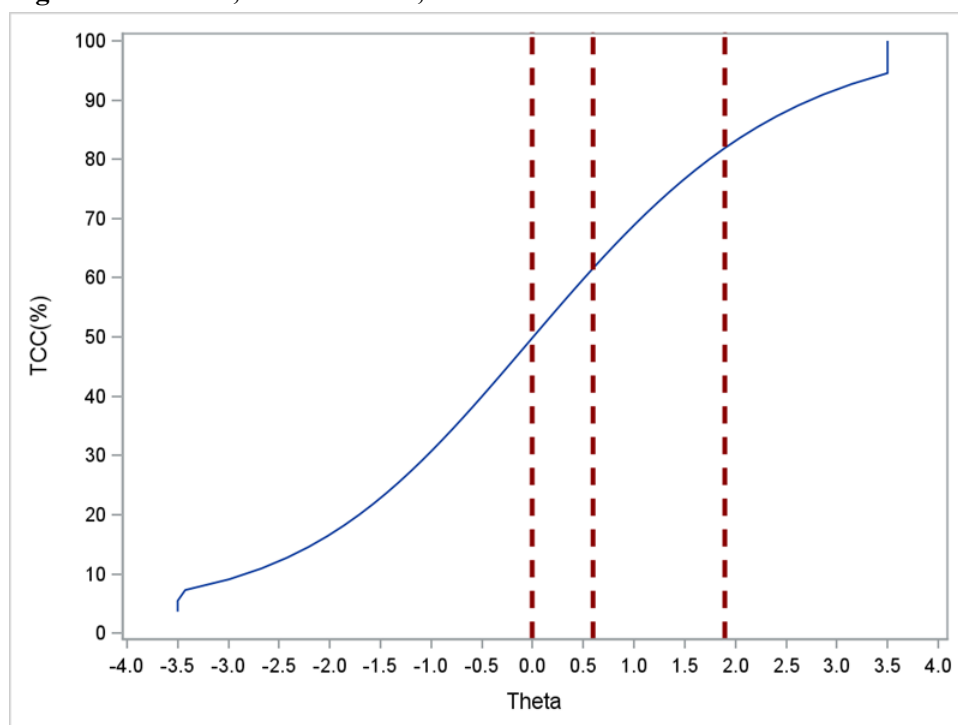
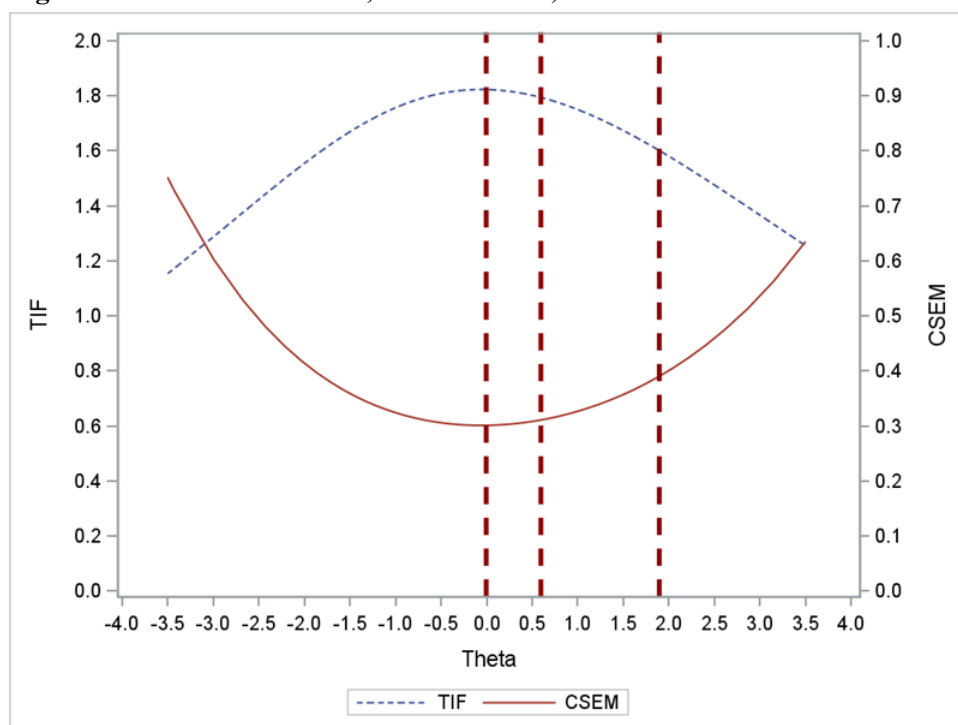
Figure B.27. TCC, ELA Grade 7, Form 1**Figure B.28. TIF and CSEM, ELA Grade 7, Form 1**

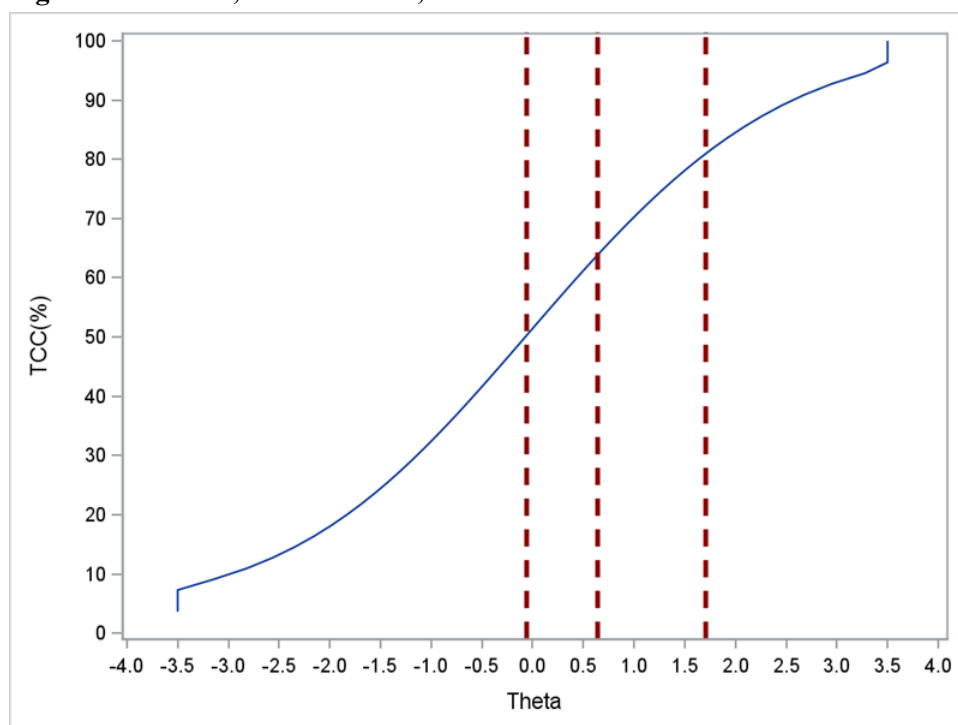
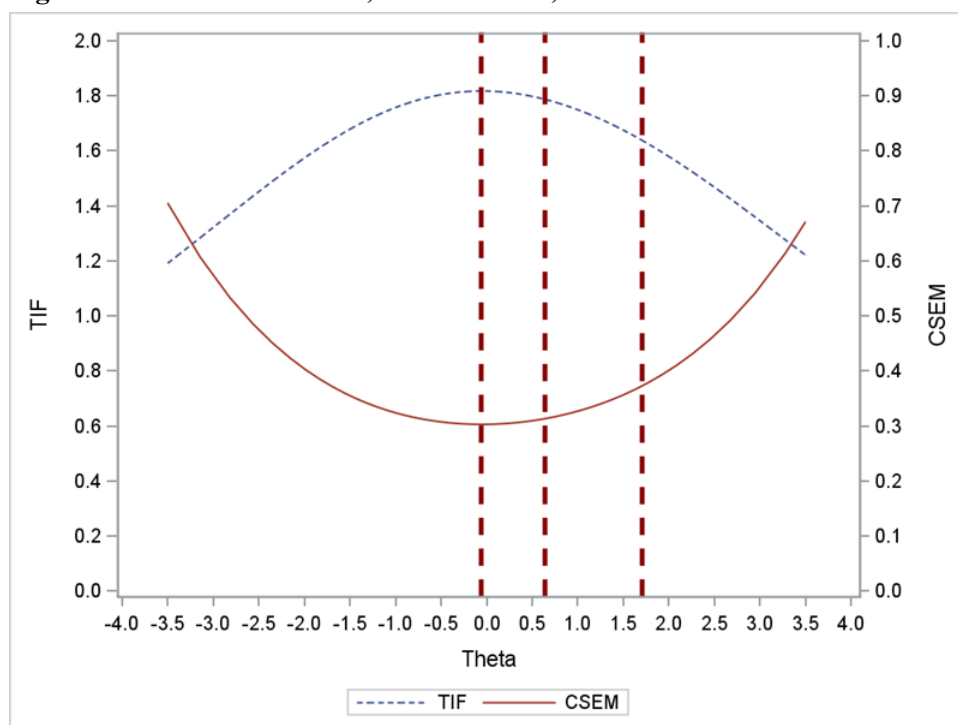
Figure B.29. TCC, ELA Grade 8, Form 1**Figure B.30. TIF and CSEM, ELA Grade 8, Form 1**

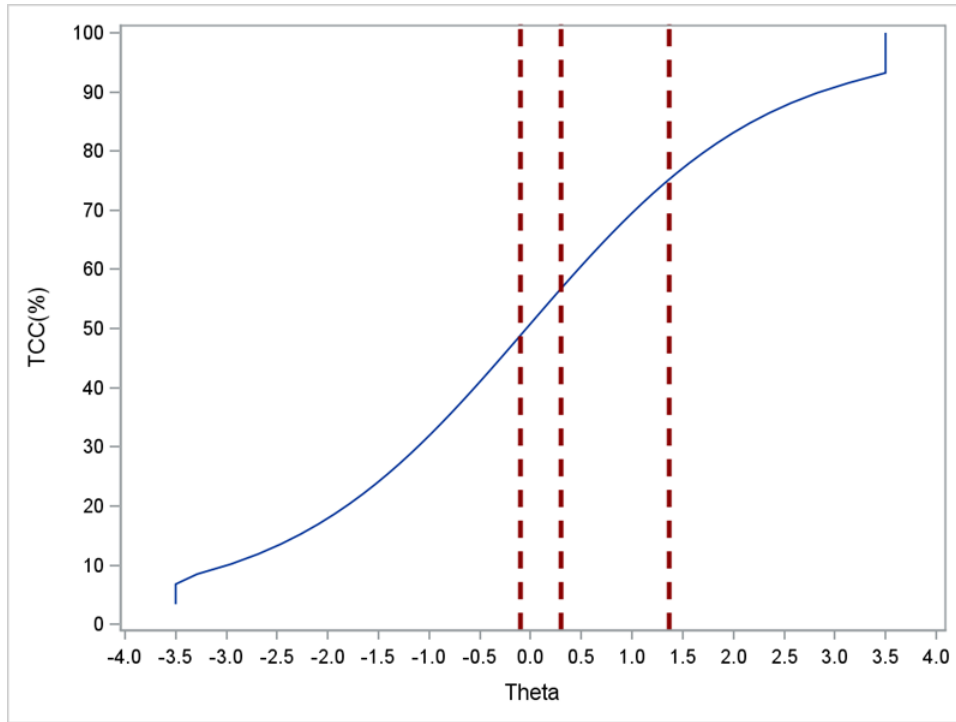
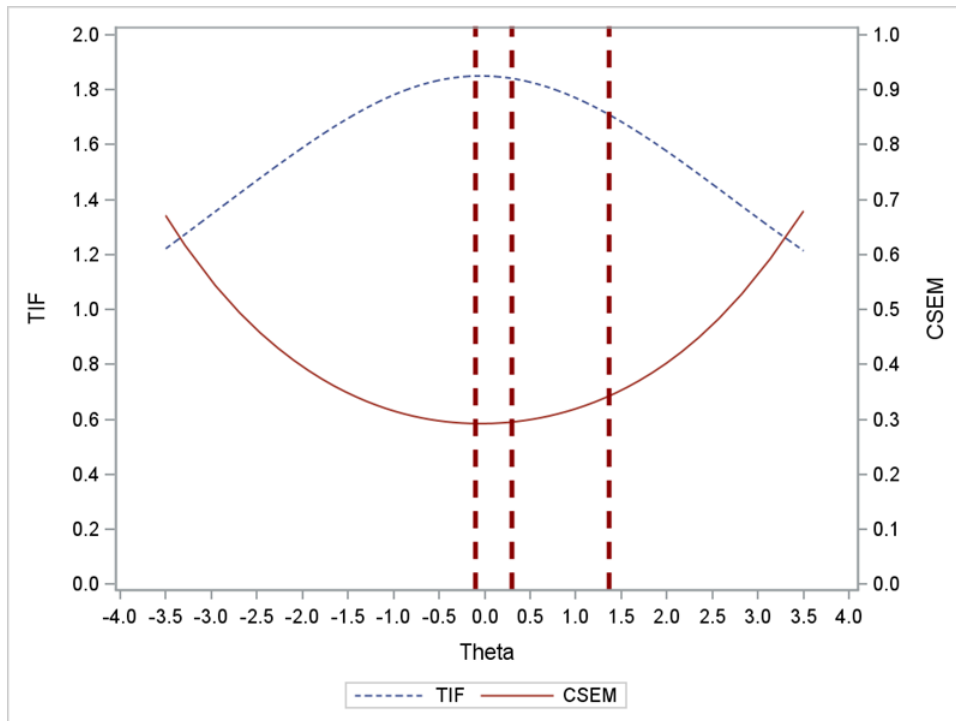
Figure B.31. TCC, ELA Grade 3, Form 2**Figure B.32. TIF and CSEM, ELA Grade 3, Form 2**

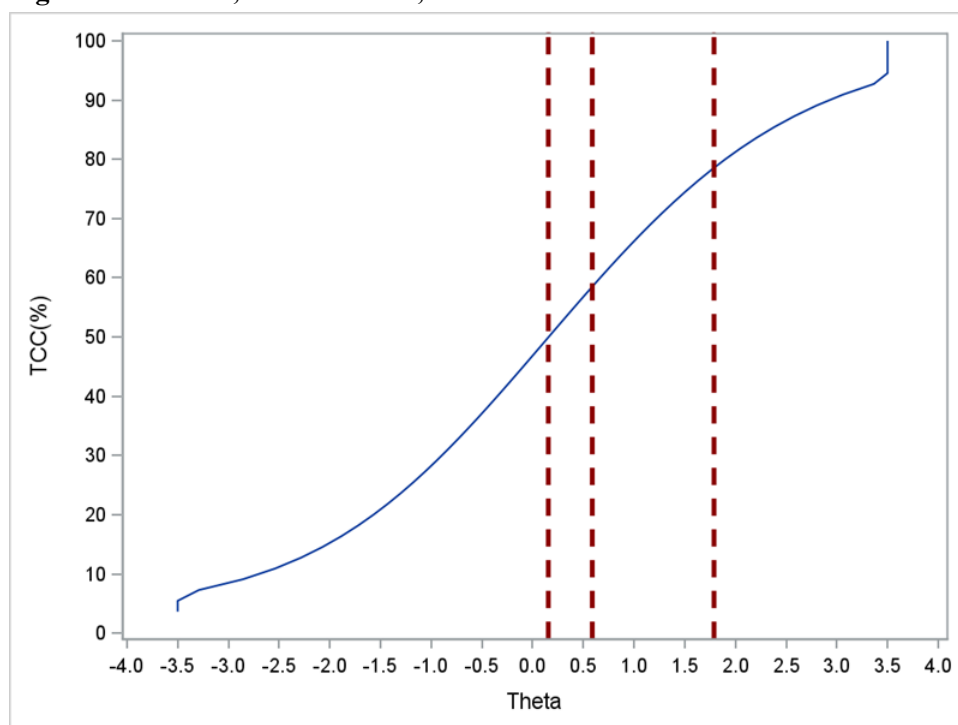
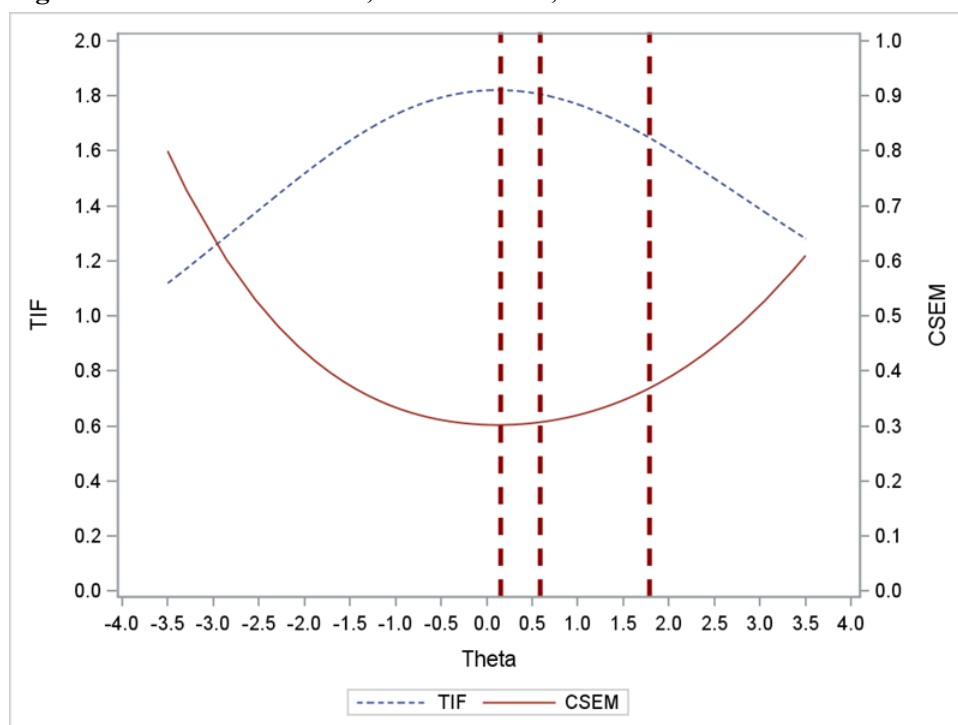
Figure B.33. TCC, ELA Grade 4, Form 2**Figure B.34. TIF and CSEM, ELA Grade 4, Form 2**

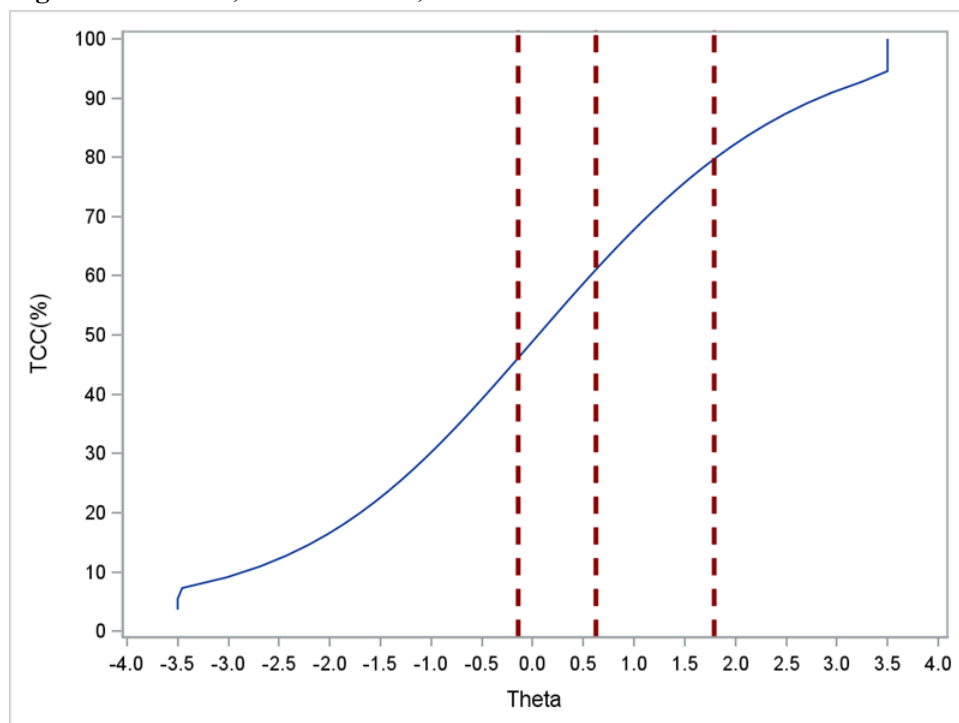
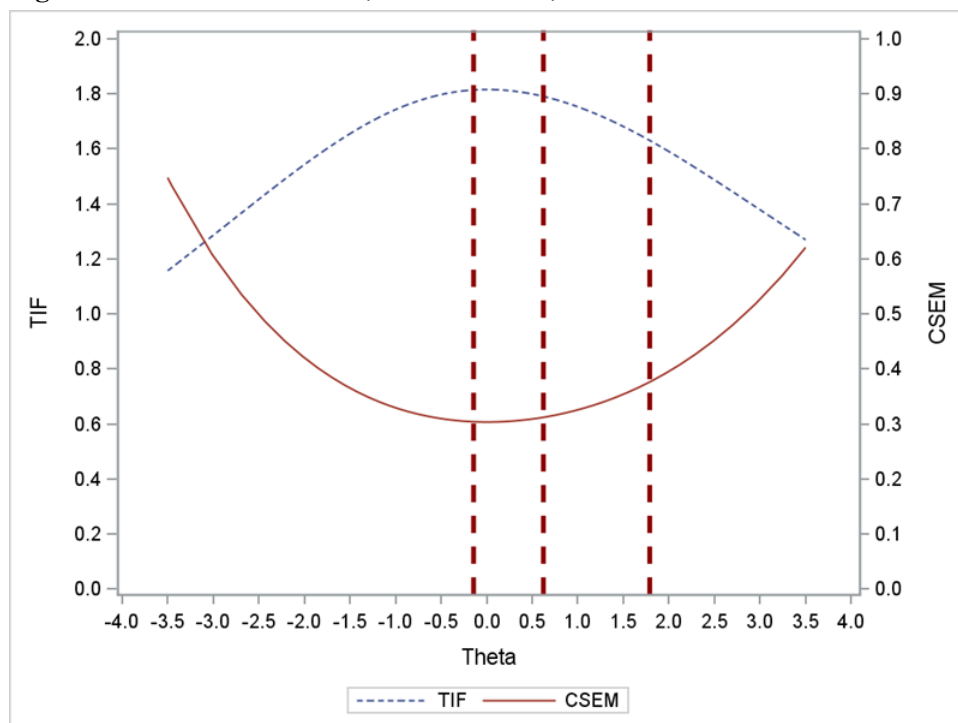
Figure B.35. TCC, ELA Grade 5, Form 2**Figure B.36. TIF and CSEM, ELA Grade 5, Form 2**

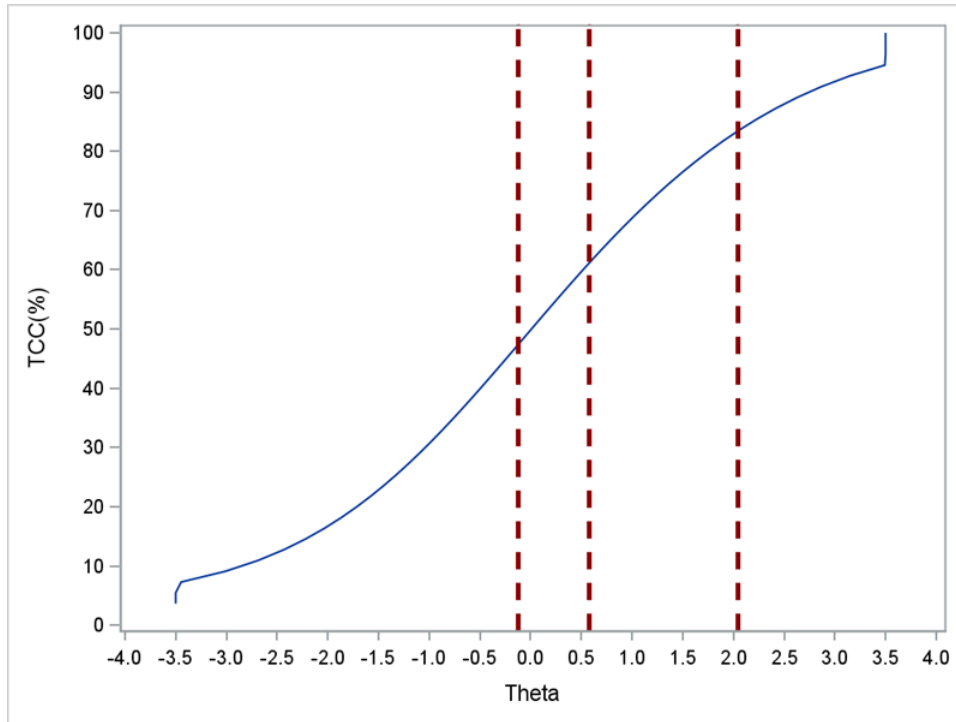
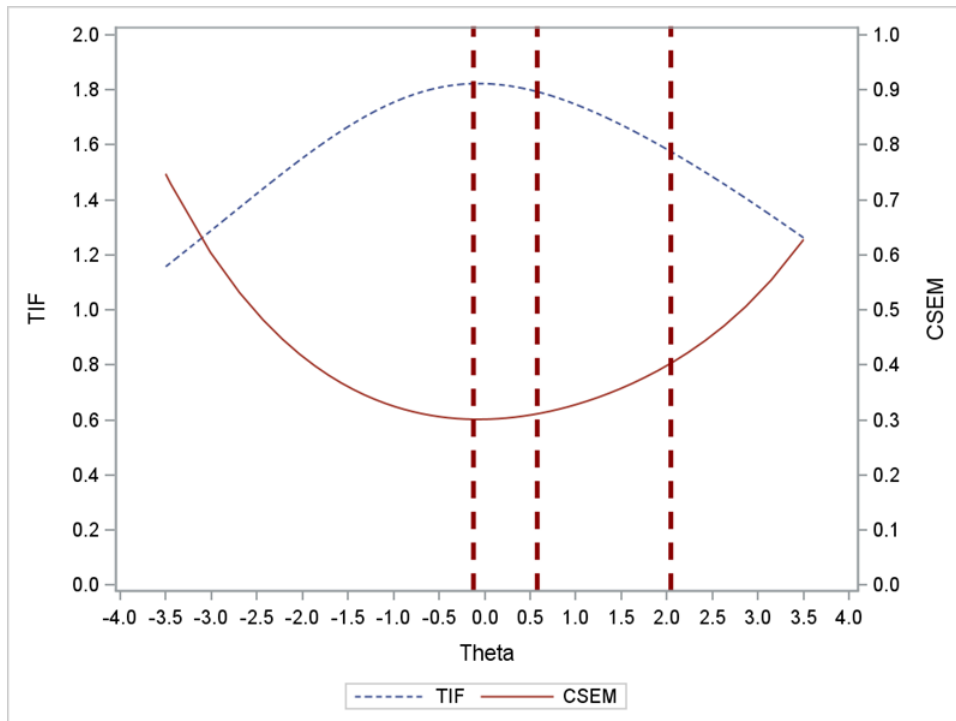
Figure B.37. TCC, ELA Grade 6, Form 2**Figure B.38. TIF and CSEM, ELA Grade 6, Form 2**

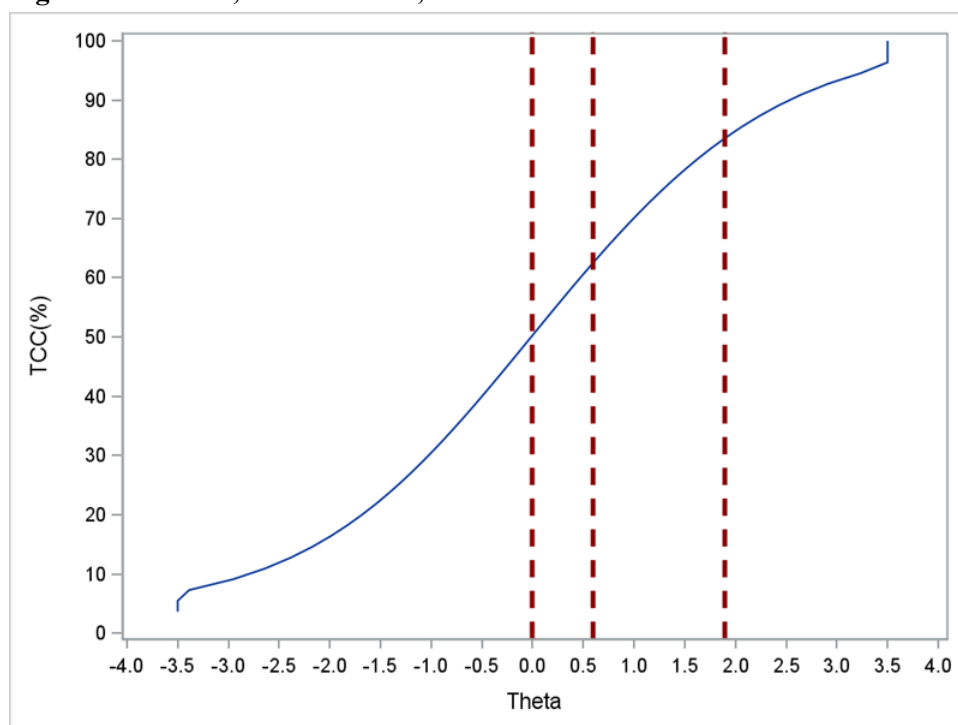
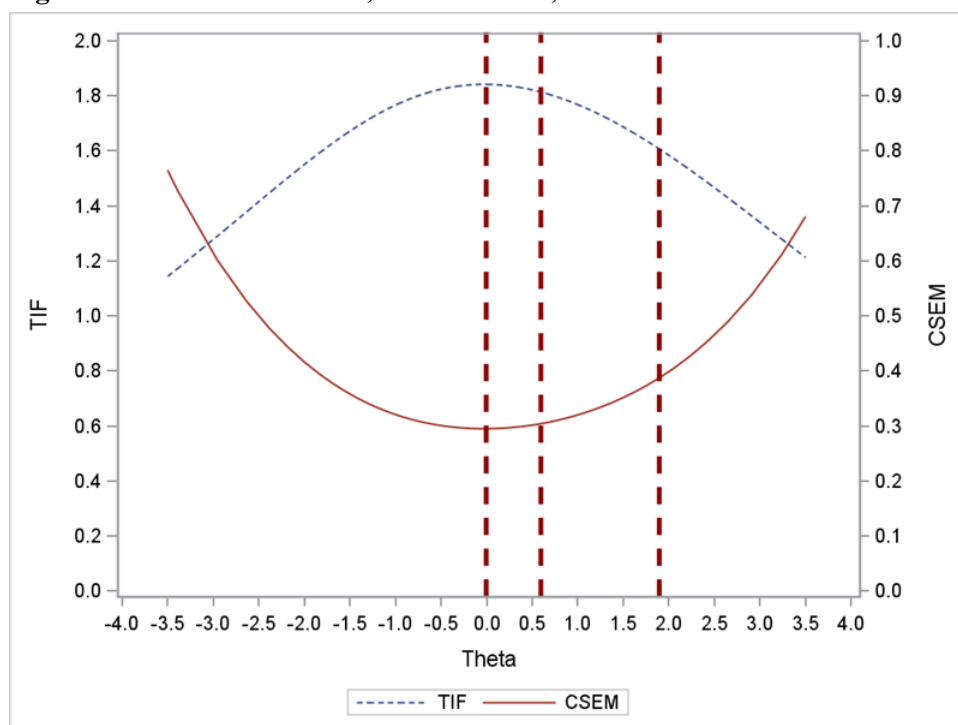
Figure B.39. TCC, ELA Grade 7, Form 2**Figure B.40. TIF and CSEM, ELA Grade 7, Form 2**

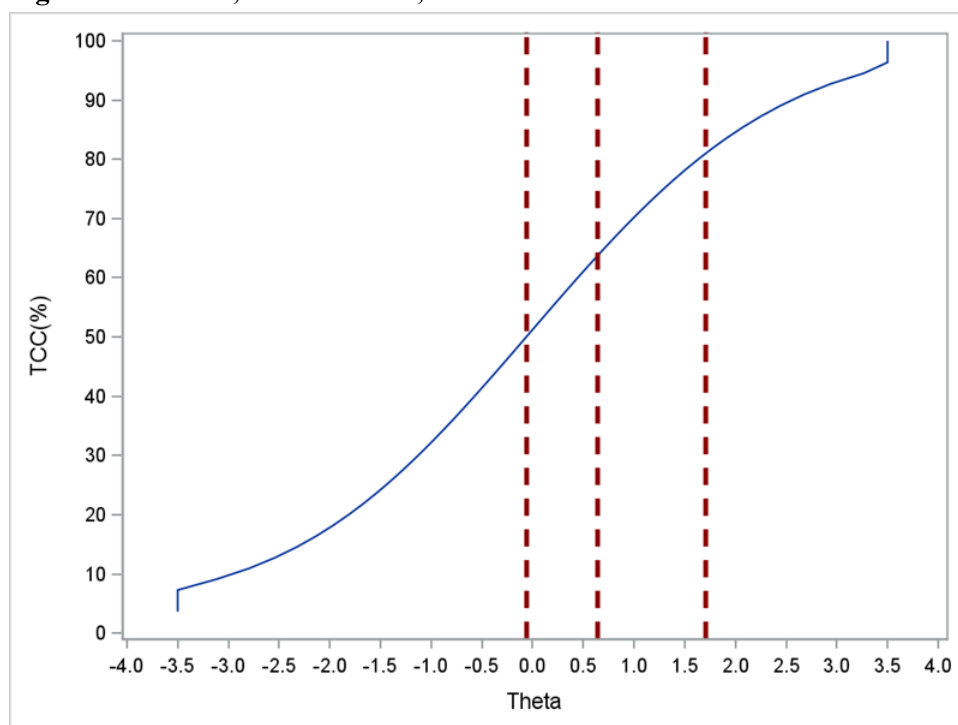
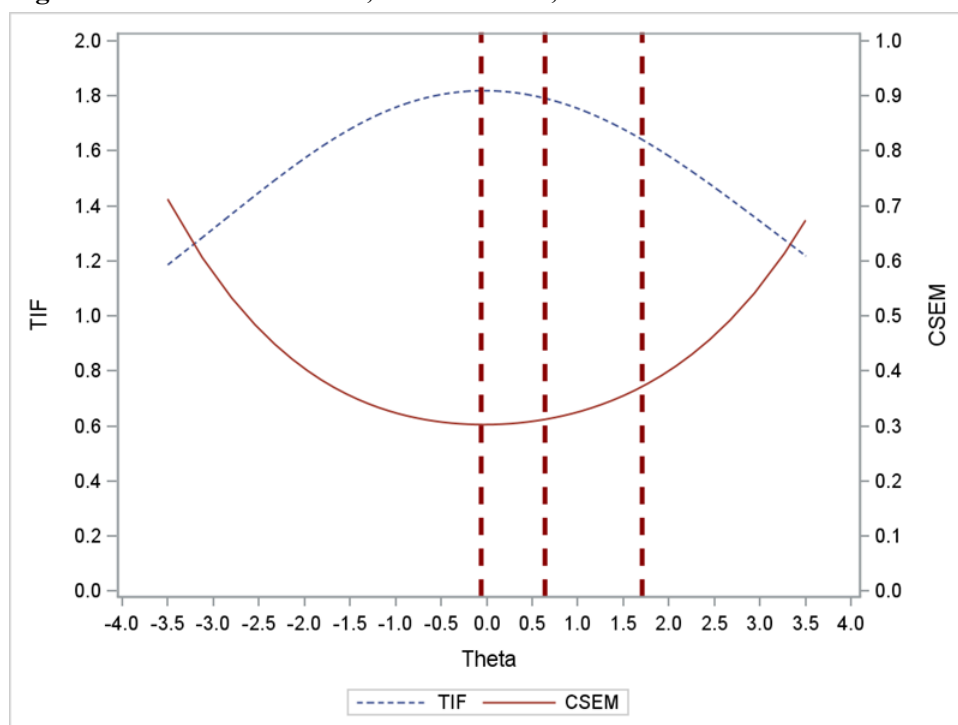
Figure B.41. TCC, ELA Grade 8, Form 2**Figure B.42. TIF and CSEM, ELA Grade 8, Form 2**

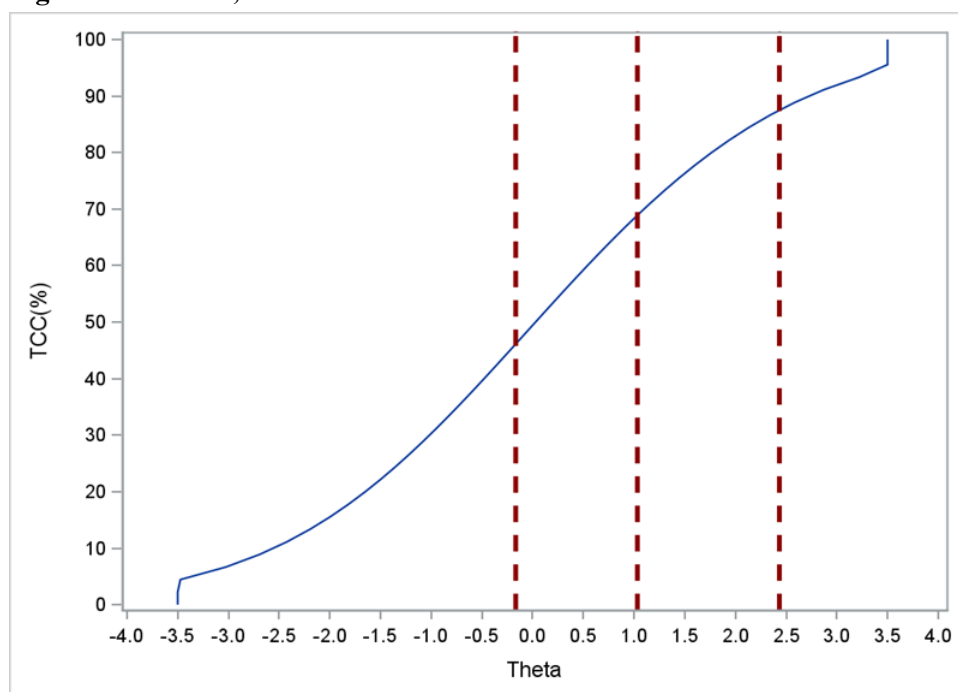
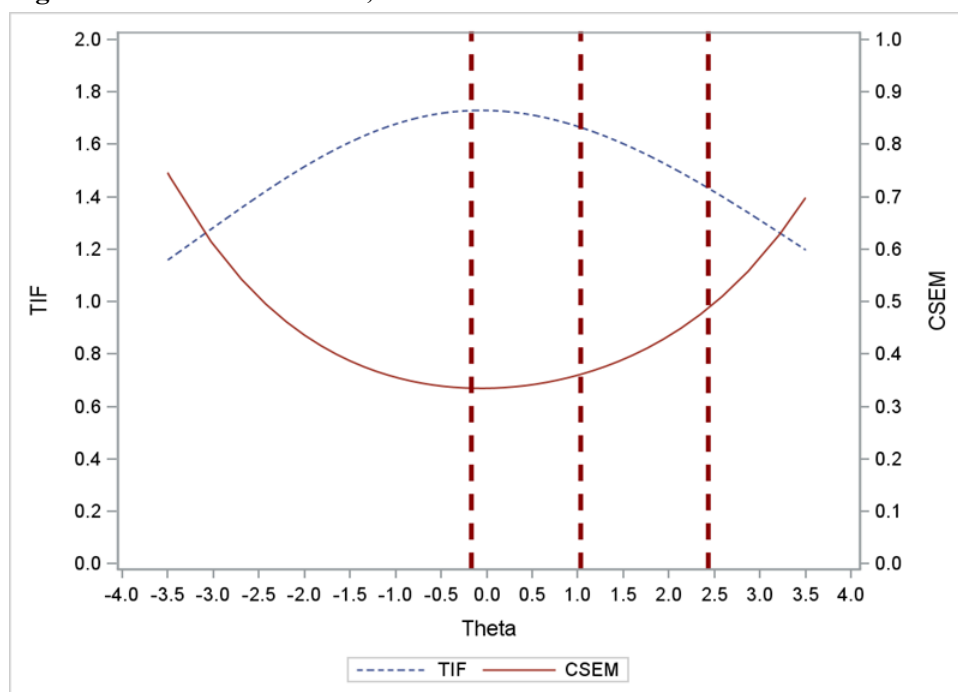
Figure B.43. TCC, Mathematics Grade 3**Figure B.44. TIF and CSEM, Mathematics Grade 3**

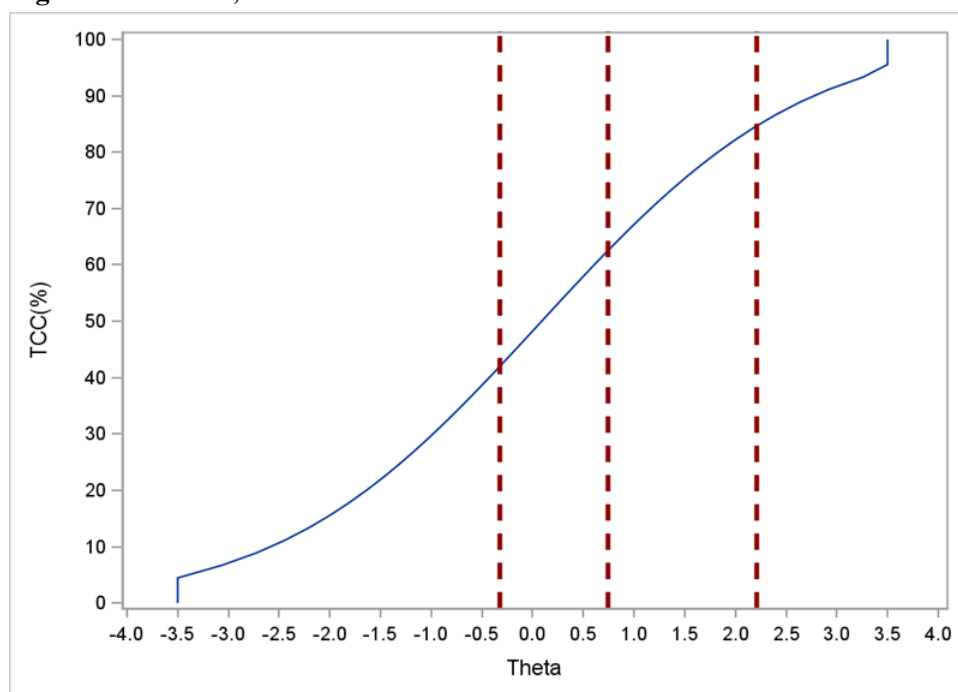
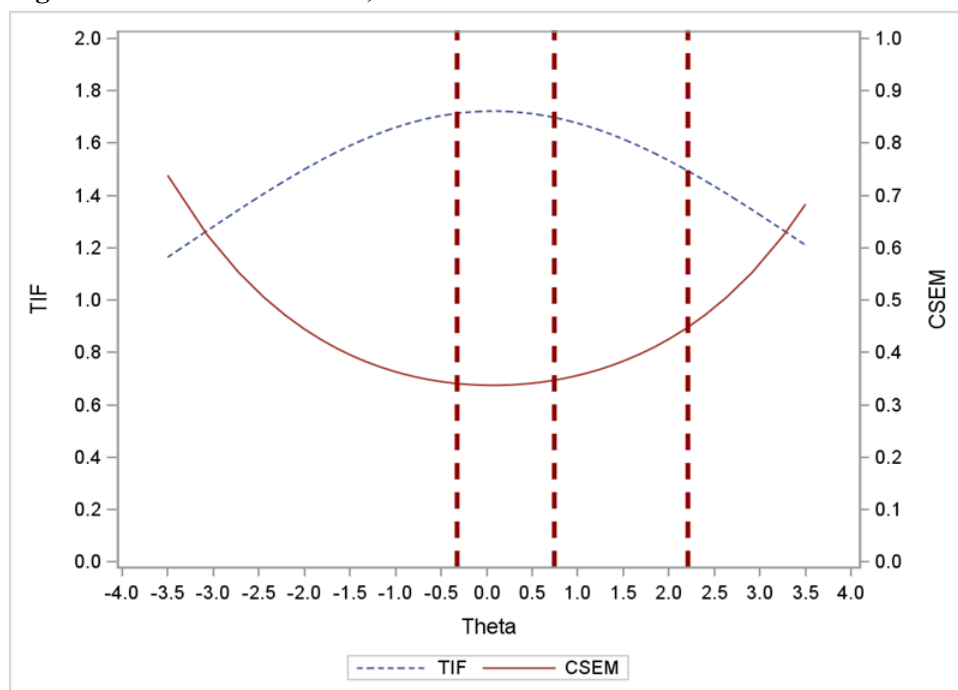
Figure B.45. TCC, Mathematics Grade 4**Figure B.46. TIF and CSEM, Mathematics Grade 4**

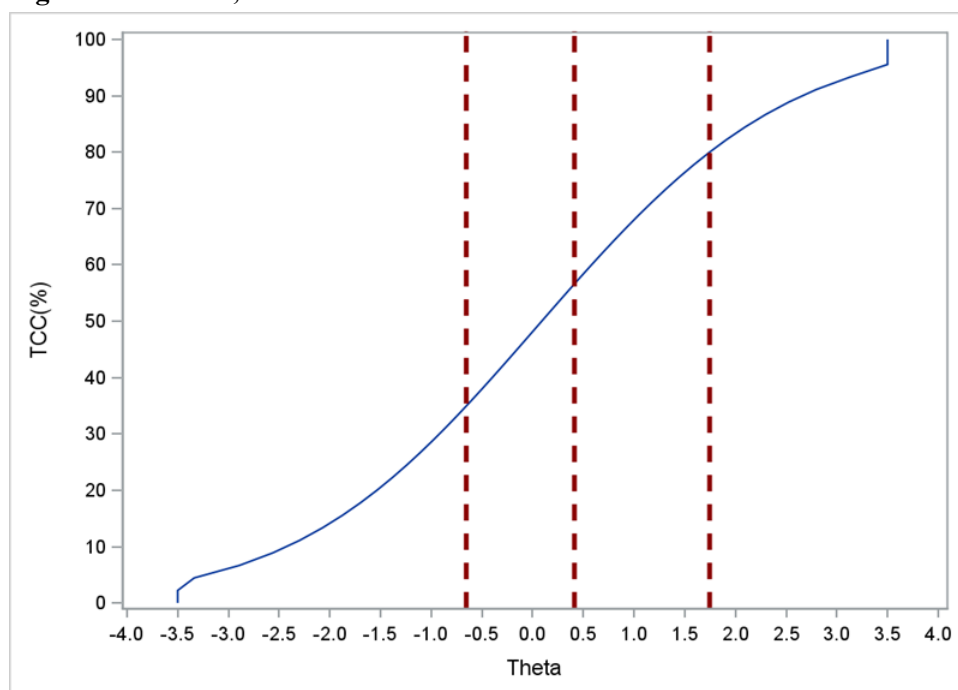
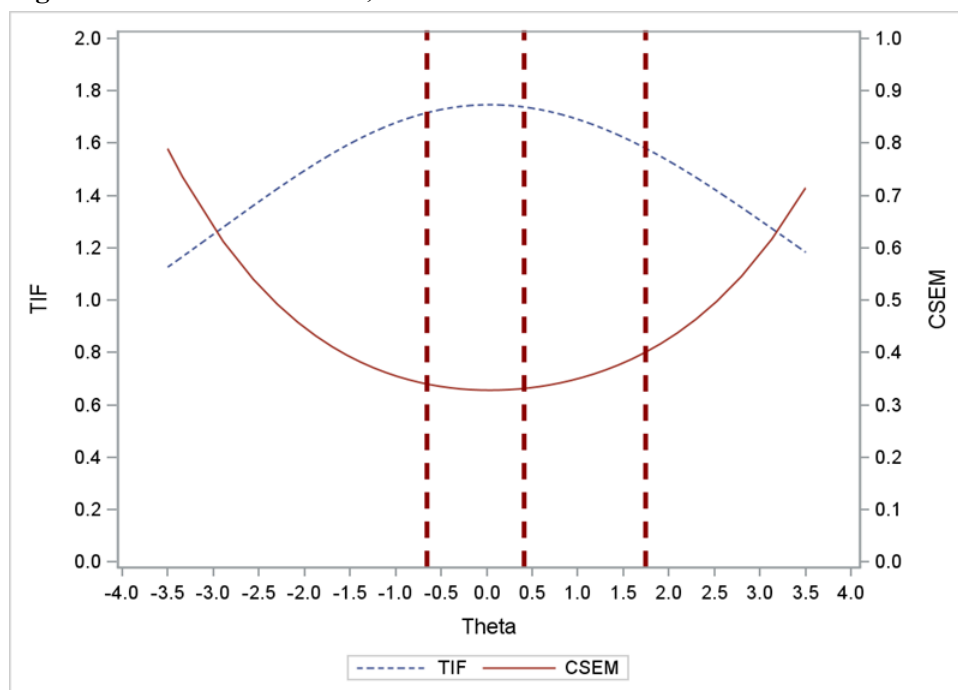
Figure B.47. TCC, Mathematics Grade 5**Figure B.48. TIF and CSEM, Mathematics Grade 5**

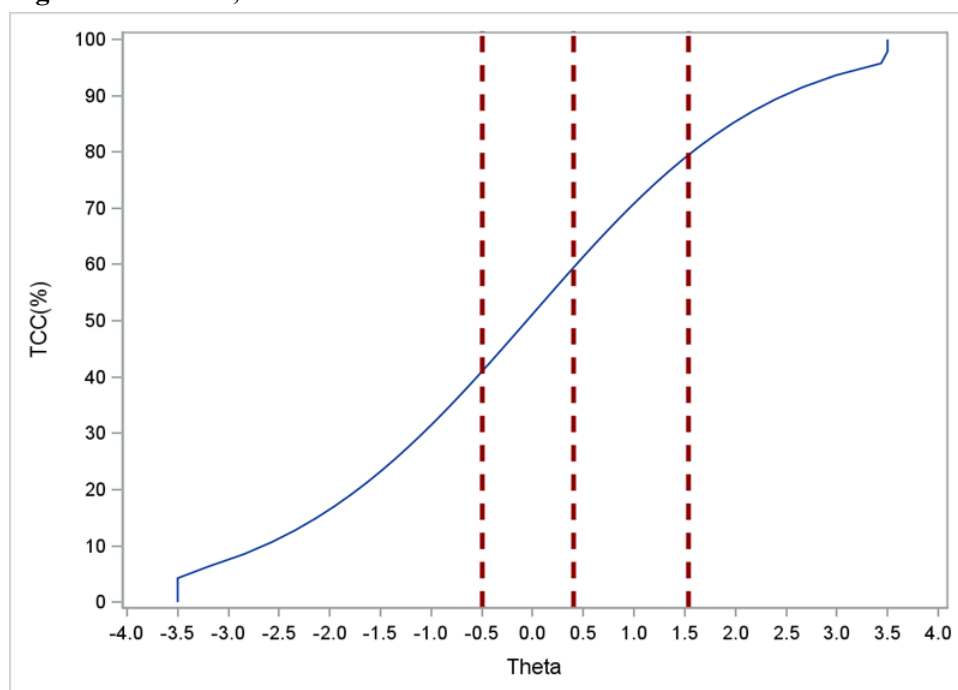
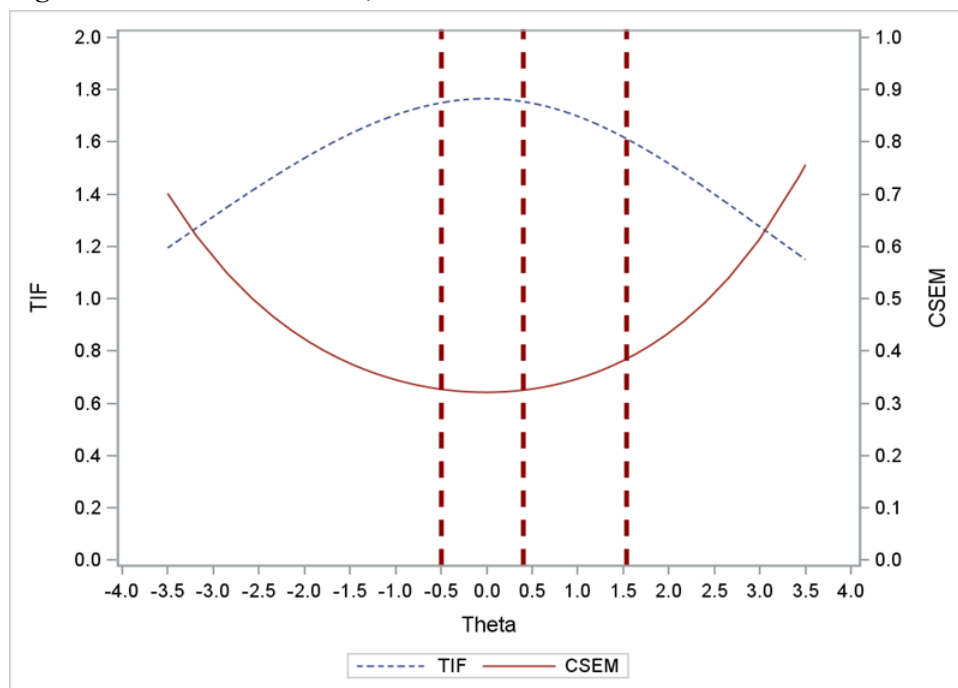
Figure B.49. TCC, Mathematics Grade 6**Figure B.50. TIF and CSEM, Mathematics Grade 6**

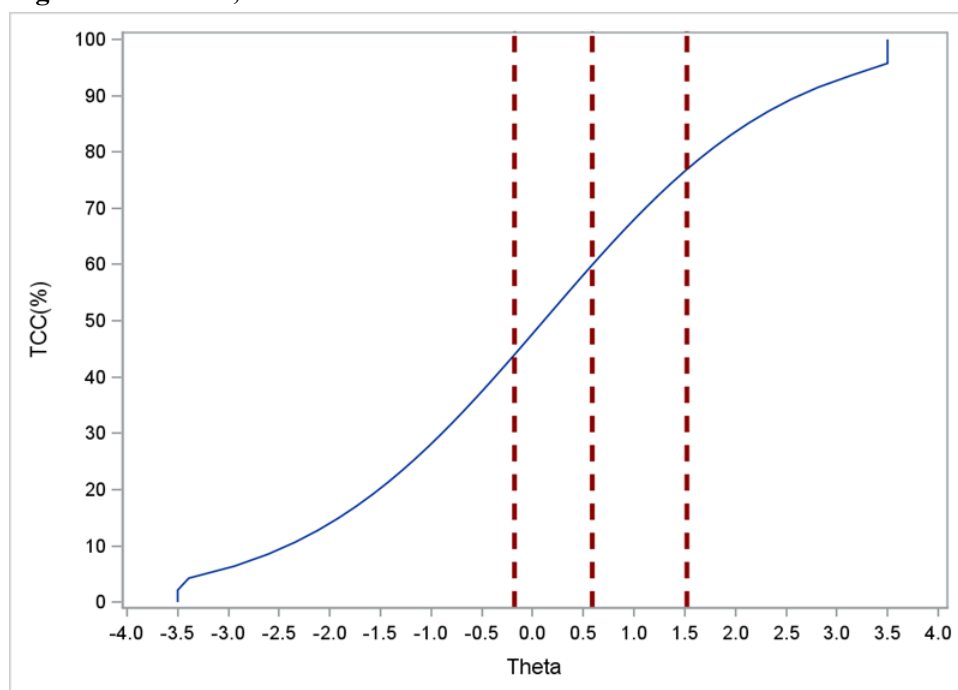
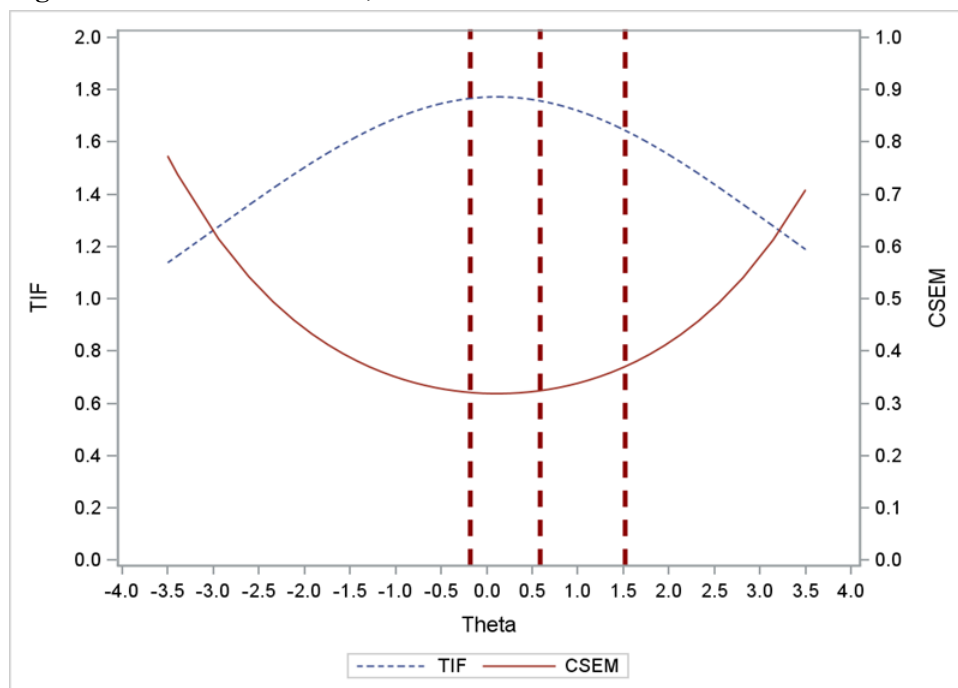
Figure B.51. TCC, Mathematics Grade 7**Figure B.52. TIF and CSEM, Mathematics Grade 7**

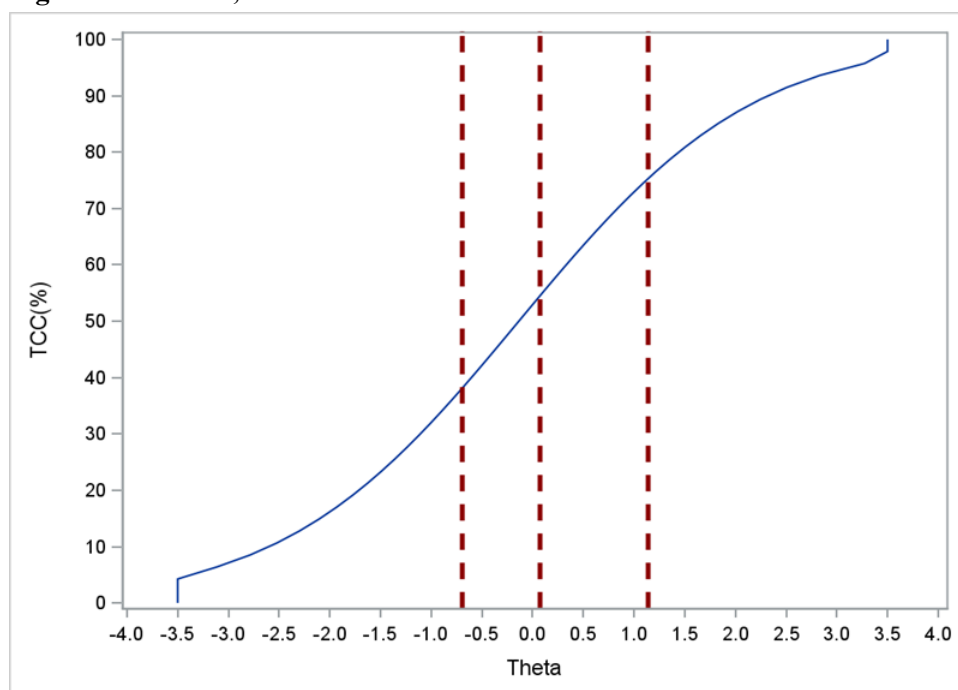
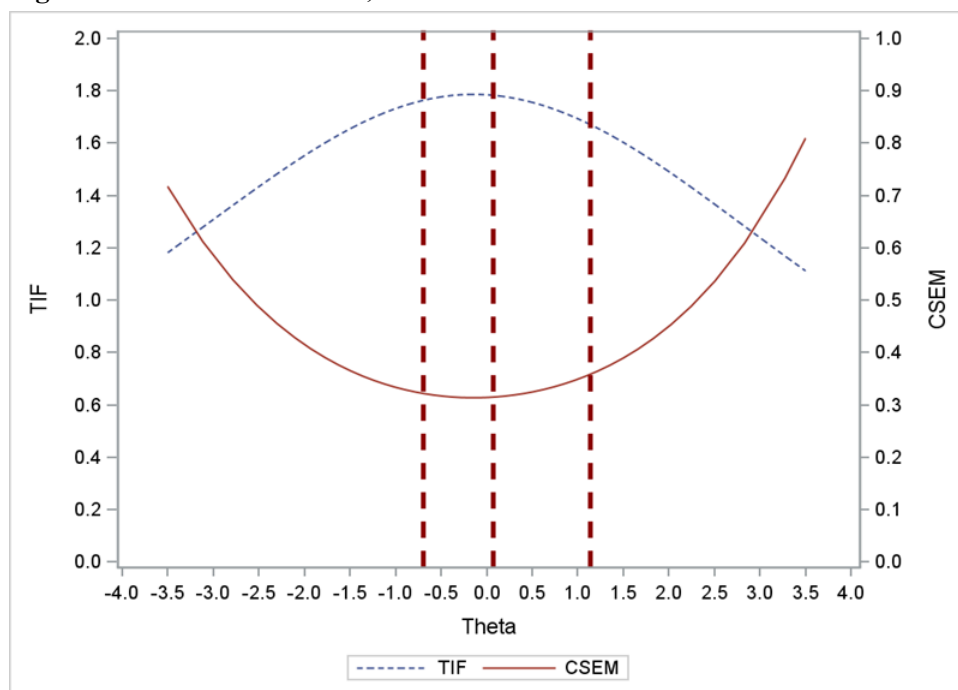
Figure B.53. TCC, Mathematics Grade 8**Figure B.54. TIF and CSEM, Mathematics Grade 8**

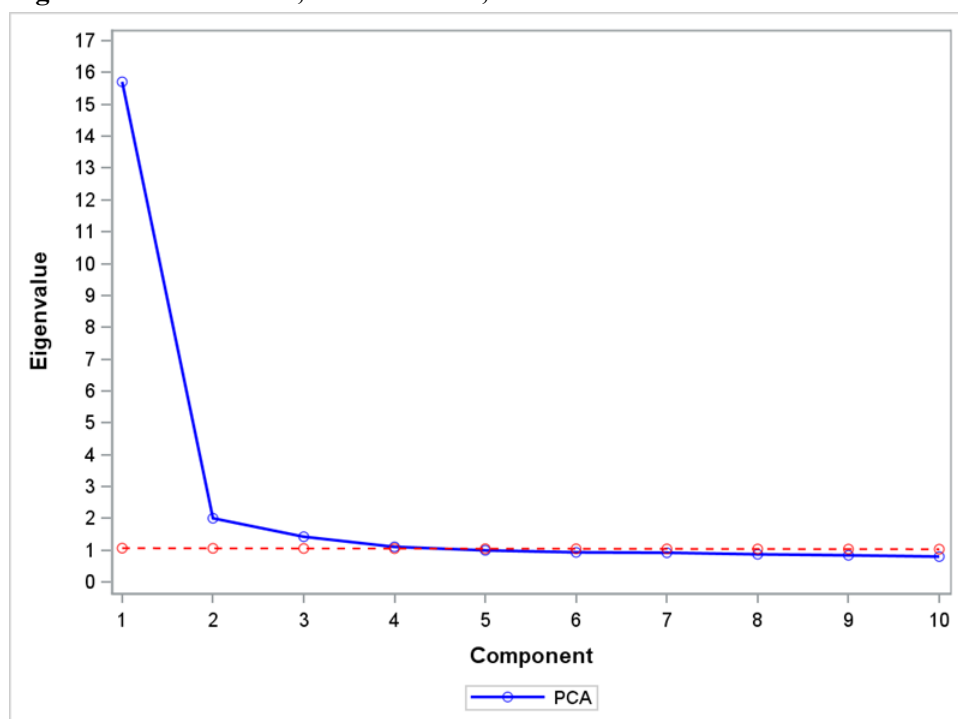
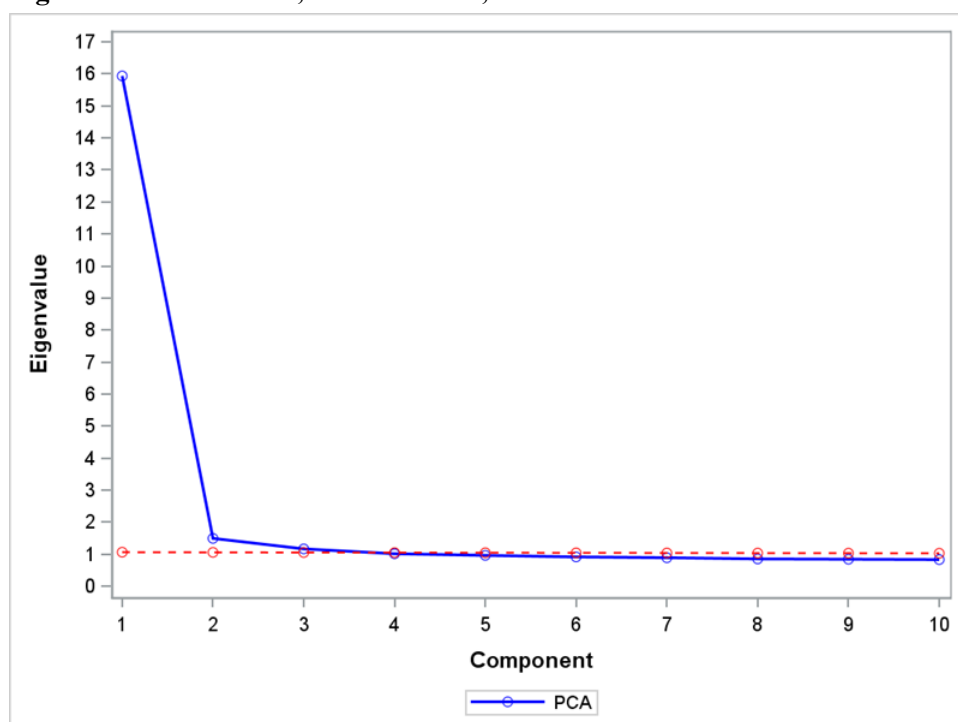
Figure B.55. Scree Plot, ELA Grade 3, Form 1**Figure B.56. Scree Plot, ELA Grade 4, Form 1**

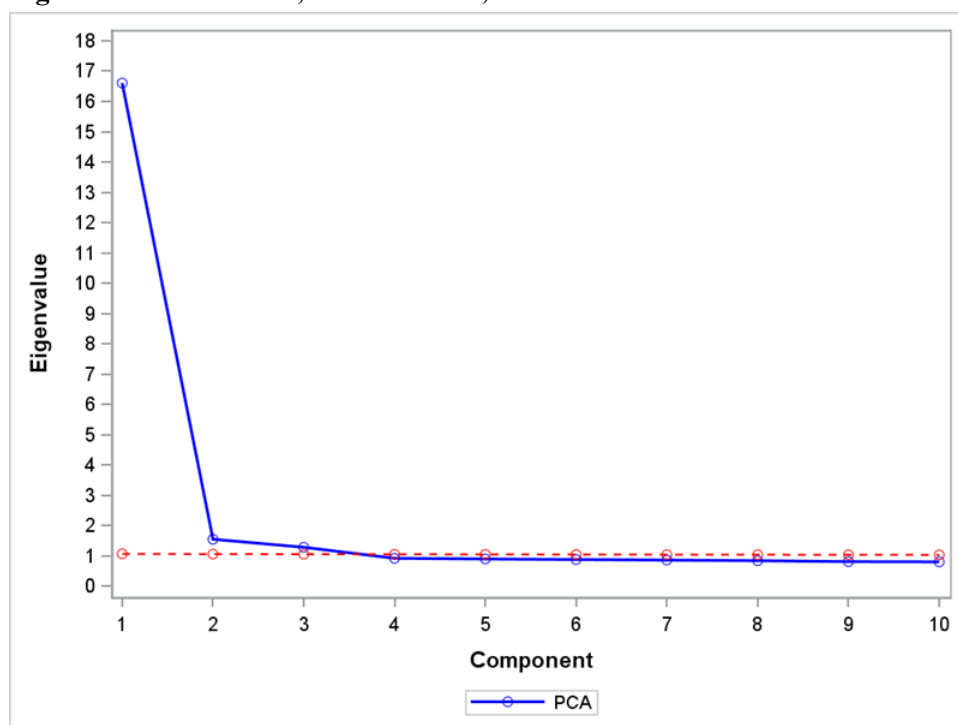
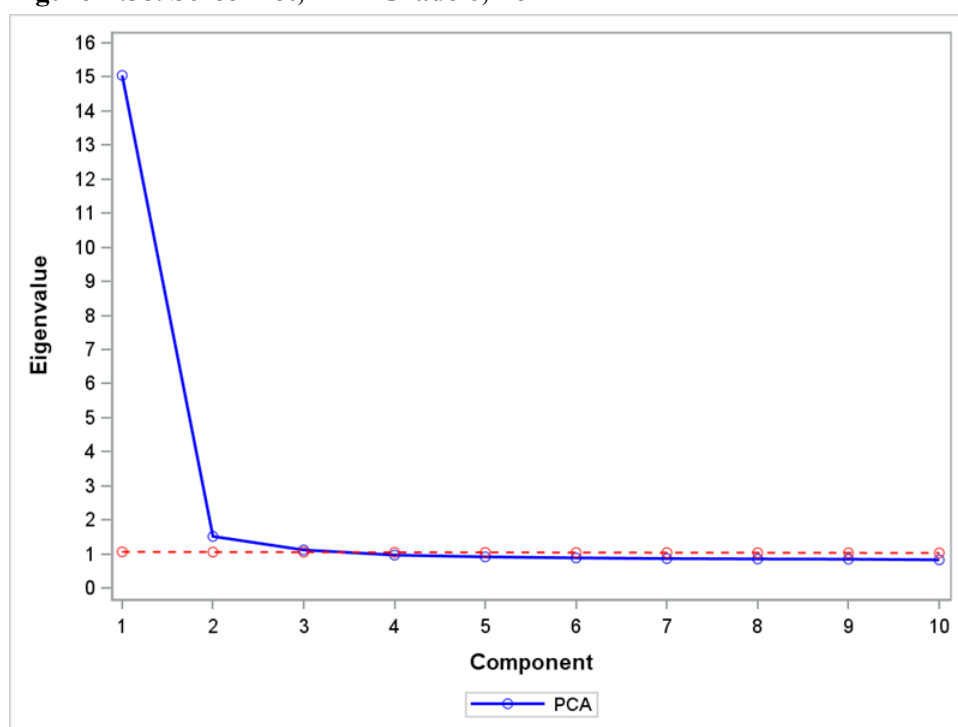
Figure B.57. Scree Plot, ELA Grade 5, Form 1**Figure B.58. Scree Plot, ELA Grade 6, Form 1**

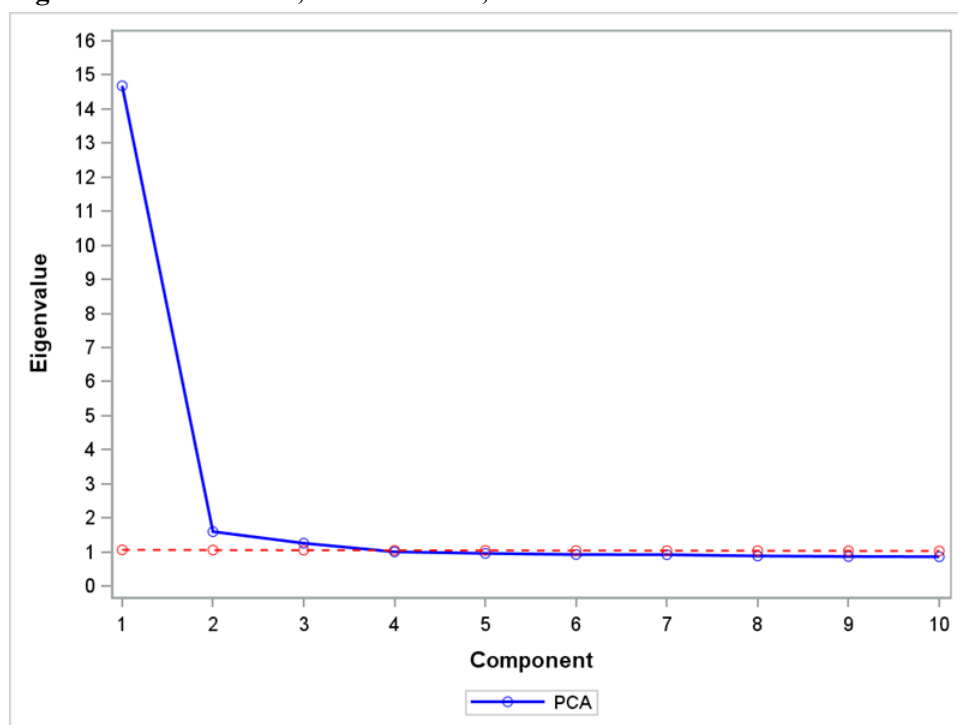
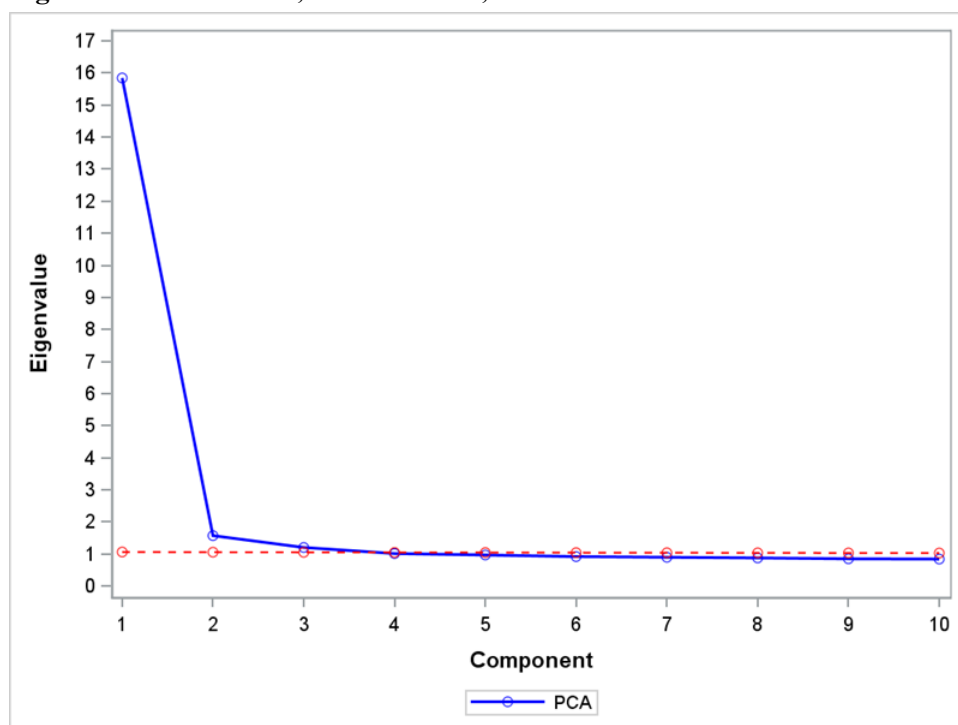
Figure B.59. Scree Plot, ELA Grade 7, Form 1**Figure B.60. Scree Plot, ELA Grade 8, Form 1**

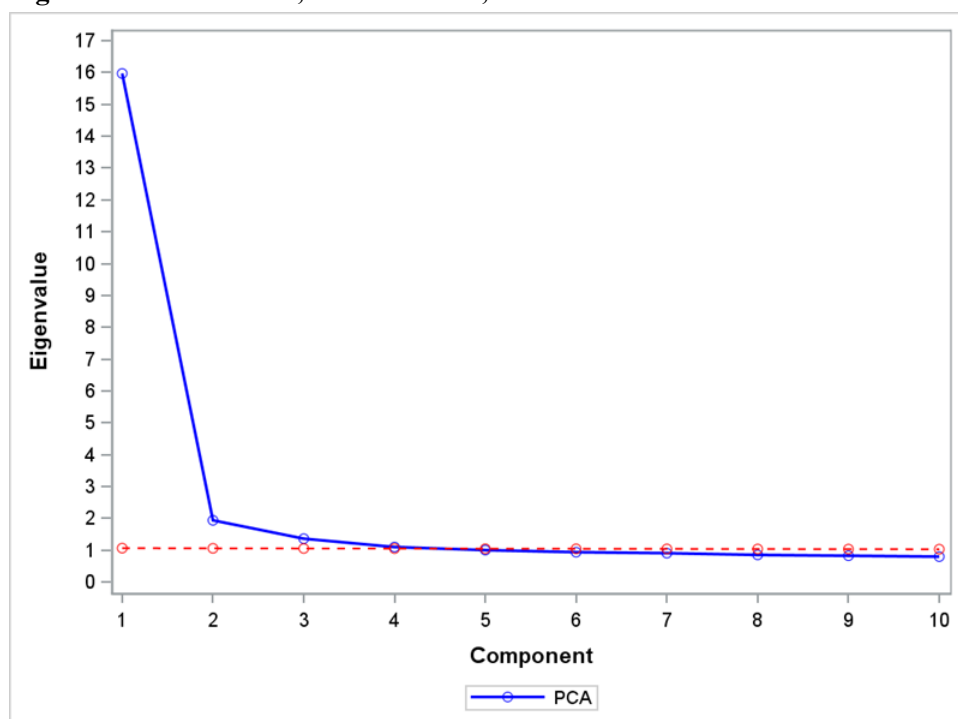
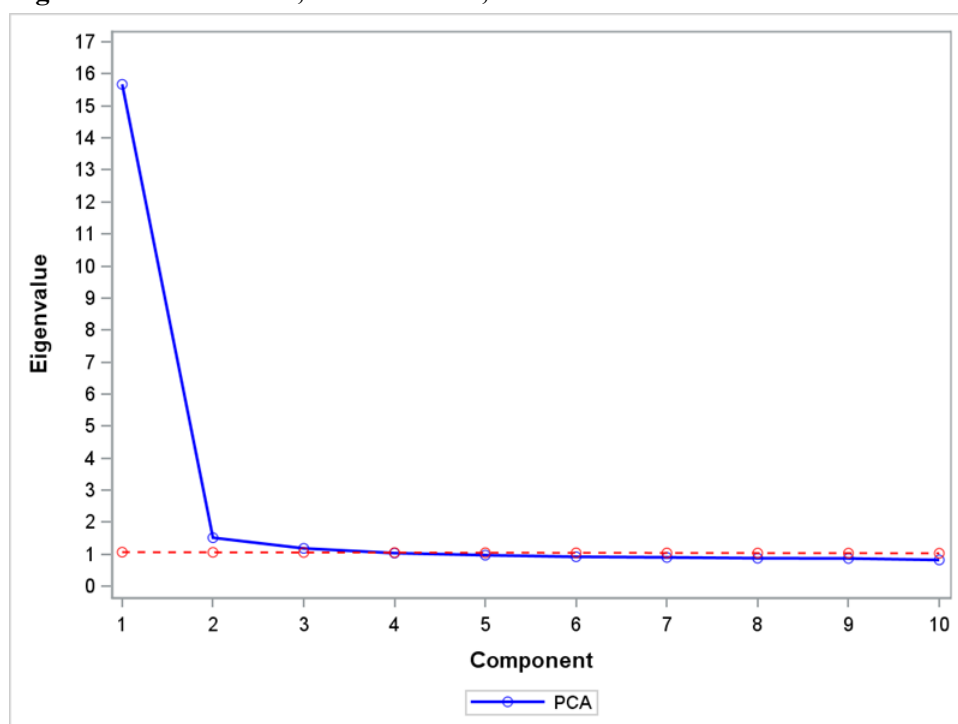
Figure B.61. Scree Plot, ELA Grade 3, Form 2**Figure B.62. Scree Plot, ELA Grade 4, Form 2**

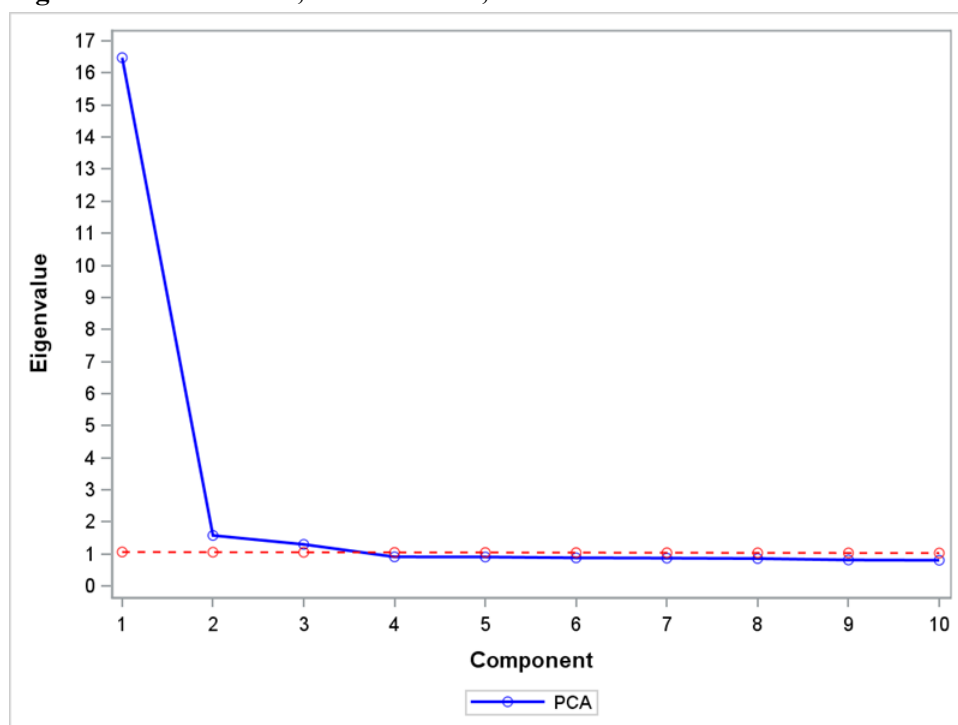
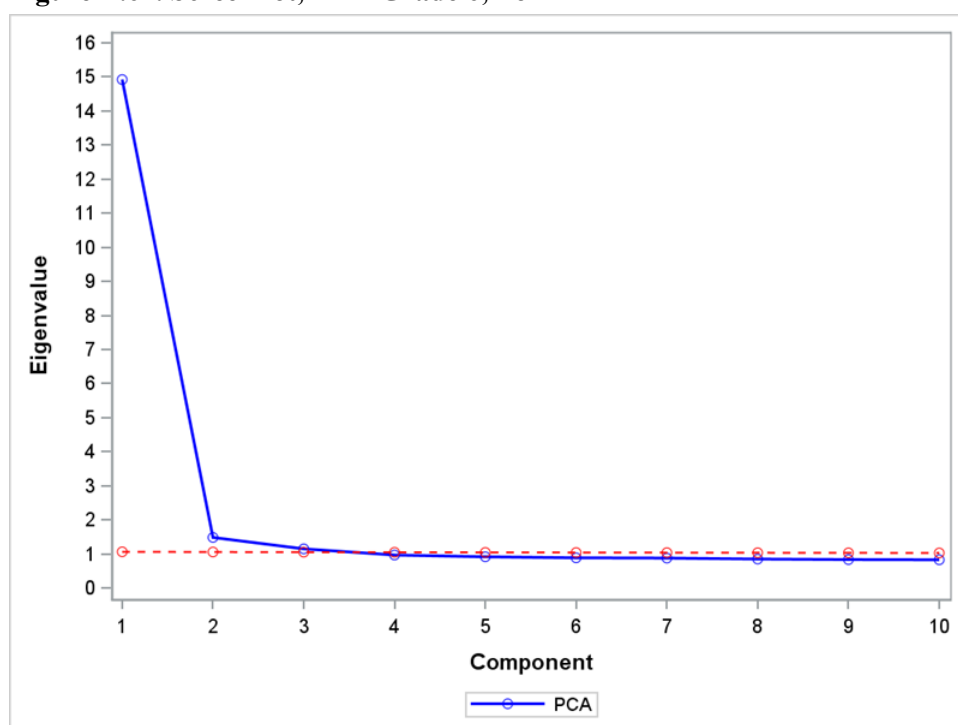
Figure B.63. Scree Plot, ELA Grade 5, Form 2**Figure B.64. Scree Plot, ELA Grade 6, Form 2**

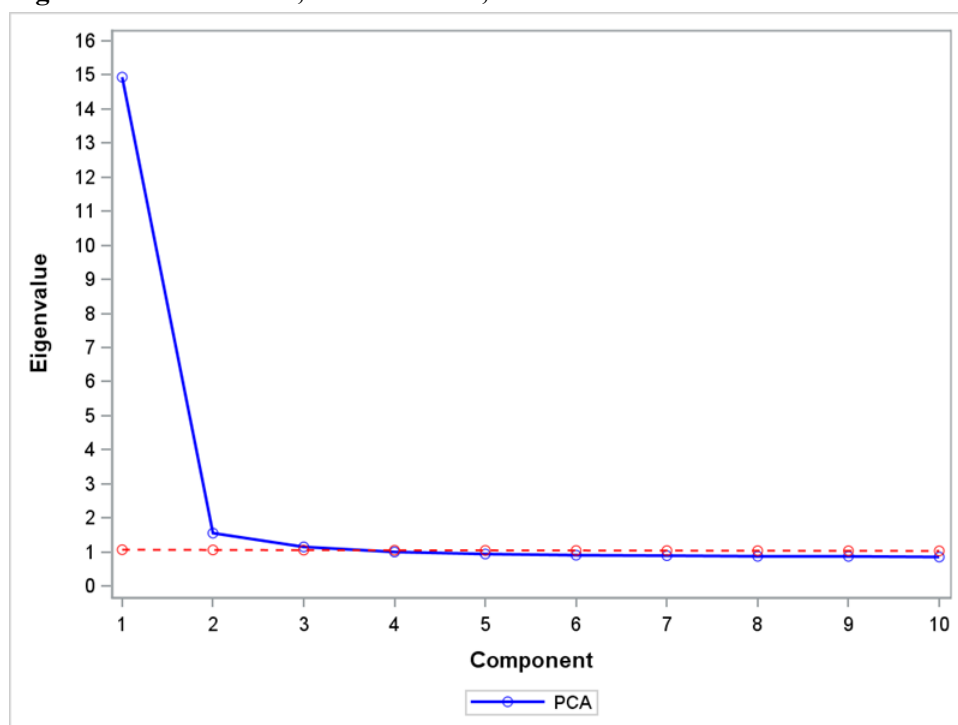
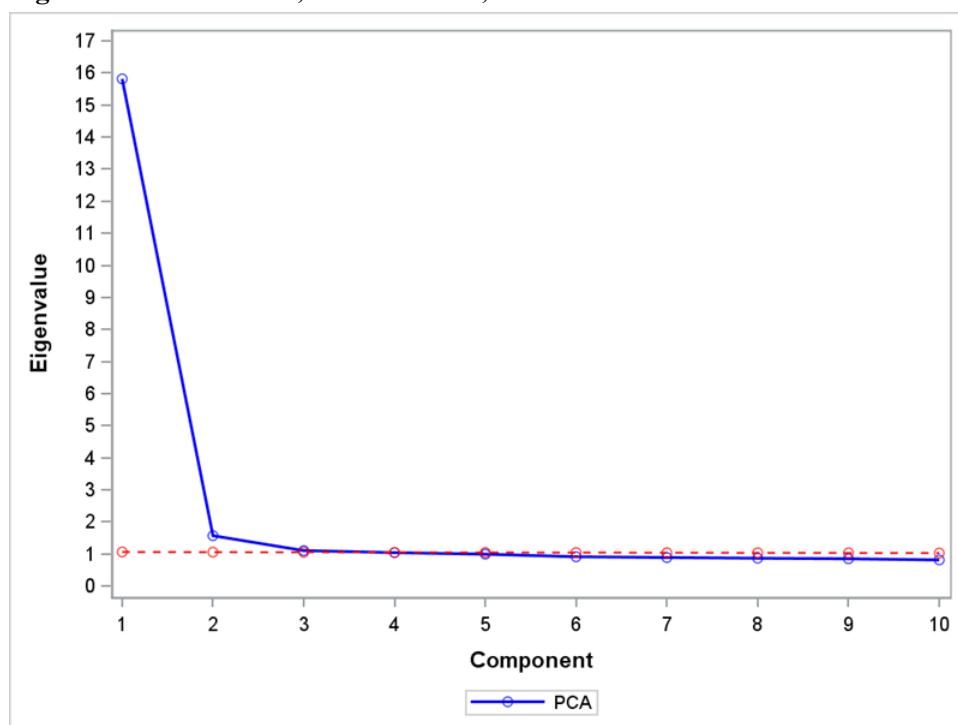
Figure B.65. Scree Plot, ELA Grade 7, Form 2**Figure B.66. Scree Plot, ELA Grade 8, Form 2**

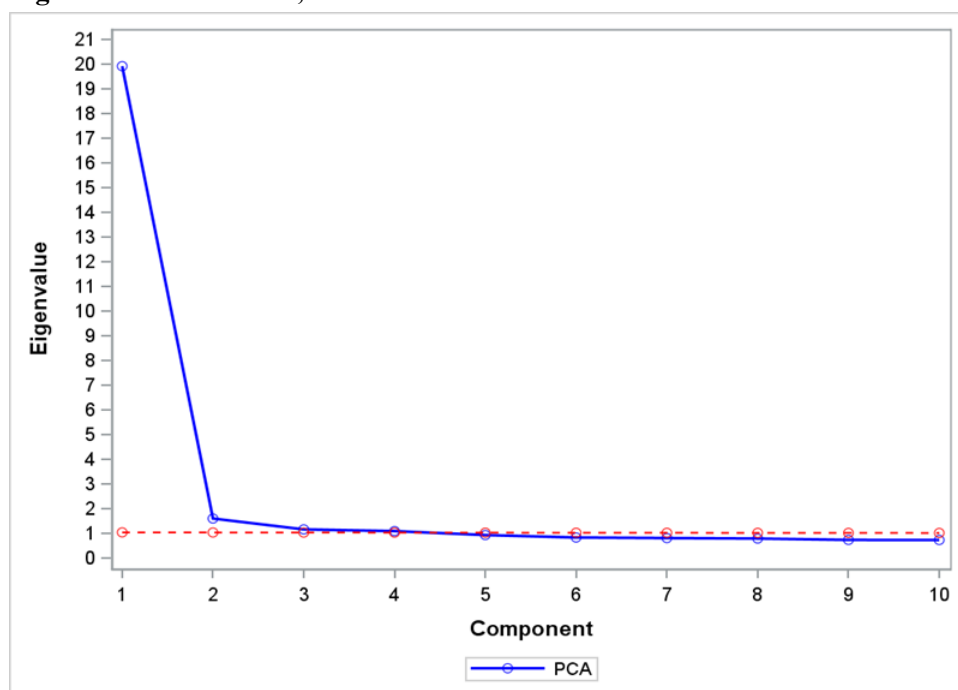
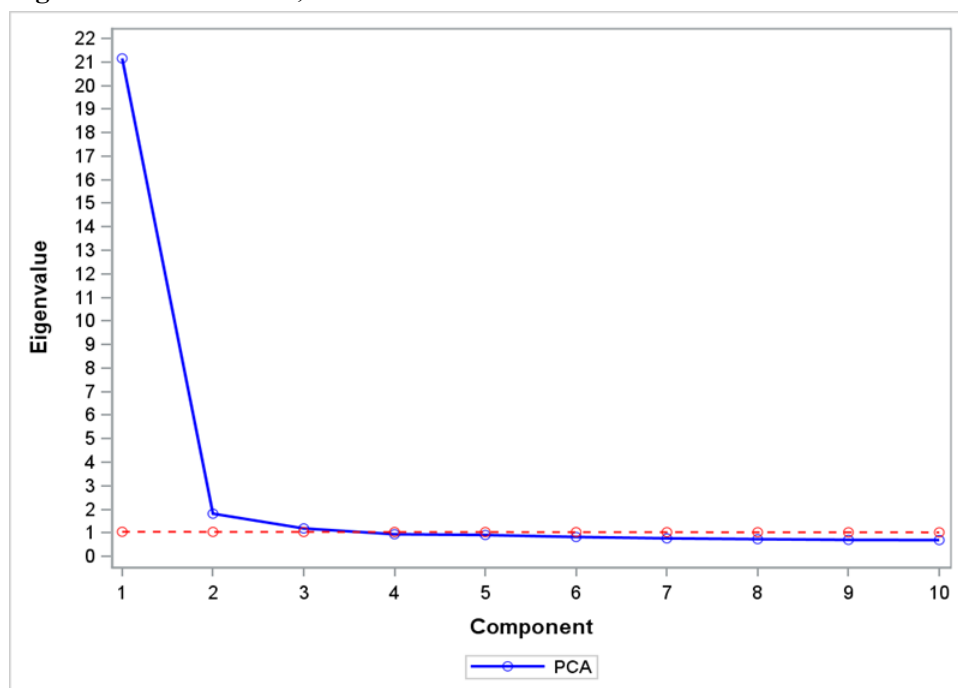
Figure B.67. Scree Plot, Mathematics Grade 3**Figure B.68. Scree Plot, Mathematics Grade 4**

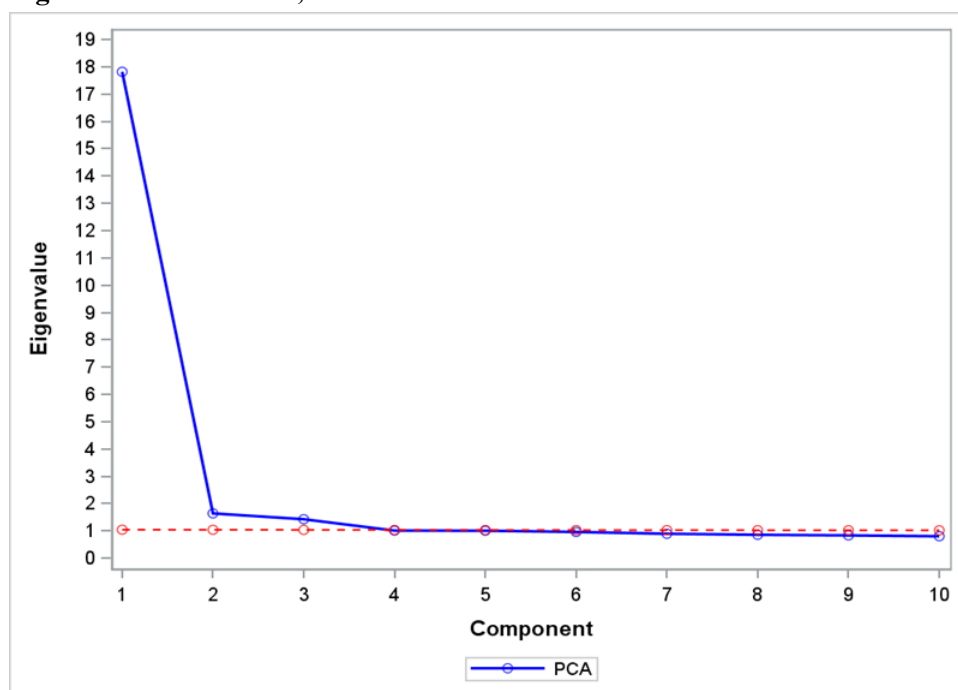
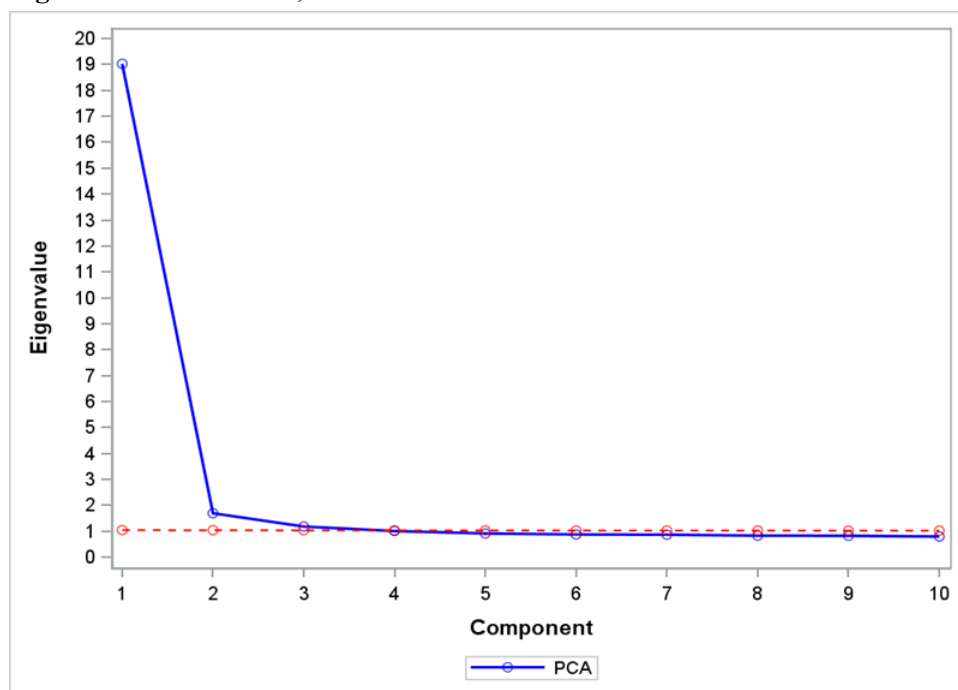
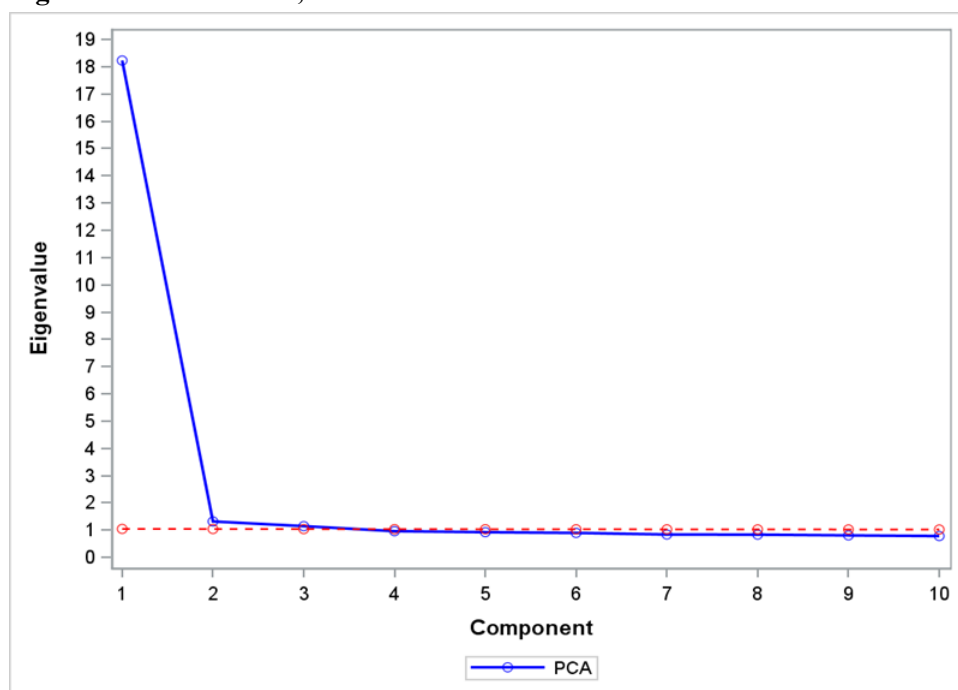
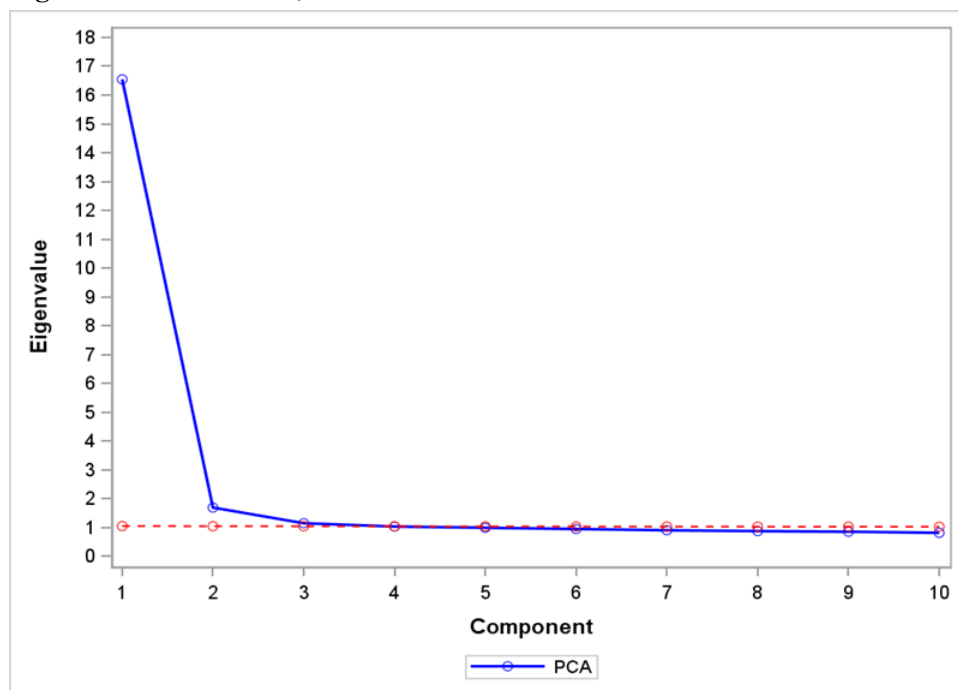
Figure B.69. Scree Plot, Mathematics Grade 5**Figure B.70. Scree Plot, Mathematics Grade 6**

Figure B.71. Scree Plot, Mathematics Grade 7**Figure B.72. Scree Plot, Mathematics Grade 8**

Appendix C: SPRING 2025 ADMINISTRATION RESULTS

This appendix presents the spring 2025 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.
- Table C.13 and Table C.14 present the number of state mean raw score points by reporting category and cluster for the online assessments.
- Figure C.1 – Figure C.12 present histograms of the total scale score distribution.
- Figure C.13 – C.24 present graphs displaying the trend in the percent of students on-grade.

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4	%MOWR Met
All	79,299	2497.40	34.73	51.7	12.5	22.8	13.1	96.4
Male	40,066	2495.72	34.44	53.6	12.2	22.2	12.1	96.2
Female	39,232	2499.11	34.94	49.7	12.8	23.4	14.1	96.7
Missing	1	*	*	*	*	*	*	*
Hispanic	39,009	2488.82	31.61	62.1	12.3	18.3	7.3	95.3
Non-Hispanic	40,290	2505.70	35.59	41.5	12.6	27.2	18.7	97.5
American Indian	4,437	2482.67	28.75	70.9	11.0	14.0	4.1	94.3
Asian	2,915	2518.25	35.56	26.8	11.8	31.9	29.5	98.6
Black or African American	5,929	2487.38	31.36	64.9	11.3	16.7	7.1	94.8
Multi-racial	5,661	2502.98	34.55	44.5	14.3	25.0	16.2	97.2
Native Hawaiian or Other Pacific Islander	533	2492.76	31.69	58.2	14.4	18.4	9.0	96.8
White	59,785	2497.99	34.73	50.7	12.5	23.4	13.3	96.6
Missing	39	2479.49	24.92	74.4	10.3	12.8	2.6	97.4
Special Education	12,744	2473.06	29.81	80.8	6.3	9.1	3.8	89.7
English Learner (EL)	11,151	2469.48	21.87	88.5	6.6	4.5	0.4	90.6
Low Socioeconomic Status (SES)	44,844	2487.46	31.01	64.0	12.0	17.5	6.4	95.1
Migrant	504	2475.52	26.96	79.6	7.7	9.9	2.8	92.1

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*. Statistics for subgroups with less than 11 students are omitted in compliance with FERPA regulations and replaced with an asterisk (*).

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	78,164	2517.61	35.84	41.7	12.3	31.5	14.5
Male	39,447	2515.18	36.05	44.6	12.1	29.8	13.4
Female	38,717	2520.08	35.45	38.7	12.5	33.2	15.6
Hispanic	38,365	2509.38	33.58	50.7	13.1	27.8	8.5
Non-Hispanic	39,799	2525.54	36.15	33.0	11.6	35.1	20.3
American Indian	4,082	2502.15	31.07	59.5	13.2	22.6	4.7
Asian	3,016	2538.87	35.49	19.9	9.8	37.5	32.8
Black or African American	5,904	2507.79	33.24	52.3	13.4	26.7	7.6
Multi-racial	5,241	2523.92	35.68	34.0	12.1	35.2	18.6
Native Hawaiian or Other Pacific Islander	443	2513.41	31.91	47.2	13.3	29.8	9.7
White	59,441	2518.05	35.71	41.1	12.3	32.0	14.6
Missing	37	2503.11	35.57	62.2	5.4	24.3	8.1
Special Education	13,127	2490.29	31.41	75.4	8.3	12.4	3.9
English Learner (EL)	10,172	2484.28	23.45	84.6	8.4	6.6	0.4
Low Socioeconomic Status (SES)	43,763	2507.46	32.96	53.2	13.0	26.4	7.4
Migrant	483	2493.88	31.57	68.9	10.8	16.8	3.5

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	81,330	2530.45	36.80	39.2	23.1	26.3	11.3
Male	41,215	2527.25	36.82	43.0	22.1	25.0	9.9
Female	40,115	2533.73	36.49	35.4	24.1	27.7	12.7
Hispanic	39,858	2521.56	33.79	48.3	24.1	21.5	6.0
Non-Hispanic	41,472	2538.99	37.55	30.5	22.1	31.0	16.4
American Indian	4,317	2513.03	30.80	58.7	23.3	14.9	3.0
Asian	3,103	2553.38	37.87	18.3	18.1	35.6	28.0
Black or African American	6,069	2520.64	33.56	49.4	23.9	21.3	5.4
Multi-racial	5,285	2535.54	35.84	33.0	23.6	30.3	13.2
Native Hawaiian or Other Pacific Islander	484	2526.38	33.26	41.7	26.0	24.6	7.6
White	62,039	2531.08	36.73	38.4	23.2	26.9	11.5
Missing	33	2511.48	40.62	60.6	9.1	24.2	6.1
Special Education	13,090	2501.24	30.14	75.6	13.5	8.6	2.3
English Learner (EL)	9,358	2495.49	22.34	83.8	13.2	2.9	0.2
Low Socioeconomic Status (SES)	44,787	2519.82	33.22	50.5	24.0	20.2	5.3
Migrant	470	2509.13	31.68	64.3	18.3	15.1	2.3

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	81,086	2542.75	34.88	38.2	20.8	33.8	7.2
Male	41,073	2539.94	34.69	41.4	20.4	32.3	6.0
Female	40,013	2545.63	34.85	35.0	21.2	35.3	8.5
Hispanic	39,316	2534.45	32.72	47.5	21.7	26.9	3.8
Non-Hispanic	41,770	2550.56	35.05	29.5	19.9	40.2	10.5
American Indian	4,414	2526.02	30.42	58.1	20.8	19.6	1.5
Asian	3,168	2563.77	34.84	17.3	15.4	49.2	18.1
Black or African American	6,123	2532.79	32.39	49.9	21.1	25.6	3.4
Multi-racial	5,202	2548.12	34.50	31.9	21.3	37.8	9.0
Native Hawaiian or Other Pacific Islander	463	2537.03	32.11	43.4	27.6	24.8	4.1
White	61,688	2543.45	34.67	37.2	20.9	34.5	7.4
Missing	28	2518.68	32.20	64.3	21.4	10.7	3.6
Special Education	11,881	2512.80	28.18	77.0	12.8	9.0	1.2
English Learner (EL)	8,808	2507.26	21.86	85.2	11.4	3.4	0.0
Low Socioeconomic Status (SES)	44,131	2532.41	32.08	50.1	21.7	25.0	3.2
Migrant	499	2522.12	31.36	63.9	17.2	16.8	2.0

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	80,645	2552.37	33.64	39.1	18.5	32.9	9.4
Male	41,020	2548.92	33.70	43.0	18.1	31.2	7.7
Female	39,625	2555.93	33.20	35.2	18.9	34.8	11.1
Hispanic	38,826	2544.60	31.55	48.3	19.2	27.3	5.2
Non-Hispanic	41,819	2559.58	33.91	30.6	17.9	38.2	13.3
American Indian	4,457	2536.66	28.79	58.9	18.9	19.9	2.3
Asian	3,127	2576.03	33.45	15.7	12.7	44.5	27.1
Black or African American	5,874	2542.88	30.89	50.8	20.0	24.9	4.4
Multi-racial	5,099	2557.94	32.87	31.4	19.7	37.2	11.6
Native Hawaiian or Other Pacific Islander	468	2547.95	32.08	42.9	20.5	31.0	5.6
White	61,577	2552.79	33.44	38.4	18.6	33.7	9.3
Missing	43	2527.91	28.81	69.8	14.0	14.0	2.3
Special Education	10,612	2523.35	26.64	77.8	11.2	9.9	1.1
English Learner (EL)	8,421	2518.43	21.59	85.5	10.1	4.3	0.1
Low Socioeconomic Status (SES)	42,717	2542.84	31.06	50.5	19.2	25.8	4.4
Migrant	493	2534.44	31.53	60.9	17.4	17.8	3.9

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	82,002	2559.85	35.32	40.4	19.4	29.0	11.3
Male	41,576	2554.65	35.00	46.0	19.1	26.3	8.6
Female	40,426	2565.20	34.85	34.5	19.6	31.8	14.0
Hispanic	39,507	2552.14	33.64	48.9	20.3	24.2	6.7
Non-Hispanic	42,495	2567.03	35.33	32.4	18.6	33.5	15.5
American Indian	4,601	2543.76	31.30	60.0	18.7	18.0	3.2
Asian	3,039	2584.57	33.66	15.1	13.8	39.8	31.3
Black or African American	6,039	2551.57	33.25	49.3	20.6	23.6	6.4
Multi-racial	5,074	2565.81	34.59	33.8	19.6	32.3	14.3
Native Hawaiian or Other Pacific Islander	500	2556.89	33.29	42.6	21.4	28.6	7.4
White	62,702	2560.19	35.12	39.8	19.6	29.5	11.1
Missing	47	2548.26	38.01	61.7	10.6	23.4	4.3
Special Education	10,284	2528.46	28.22	79.4	11.3	8.0	1.3
English Learner (EL)	7,429	2522.07	22.90	88.0	9.0	2.9	0.1
Low Socioeconomic Status (SES)	42,906	2550.44	33.22	51.3	20.0	22.8	5.9
Migrant	498	2536.86	32.66	69.9	13.9	12.7	3.6

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	80,802	3517.85	43.94	29.6	29.2	27.6	13.6
Male	41,008	3519.97	45.38	28.7	27.3	28.4	15.6
Female	39,793	3515.67	42.29	30.6	31.2	26.7	11.4
Missing	1	*	*	*	*	*	*
Hispanic	39,839	3507.70	41.45	37.1	32.0	23.3	7.6
Non-Hispanic	40,963	3527.73	44.04	22.3	26.6	31.7	19.4
American Indian	4,536	3499.83	40.49	44.9	31.3	18.7	5.1
Asian	2,970	3548.41	41.25	10.9	17.4	36.7	35.1
Black or African American	6,095	3504.04	42.06	40.9	31.0	21.3	6.9
Multi-racial	5,744	3523.70	43.19	24.4	28.8	30.4	16.3
Native Hawaiian or Other Pacific Islander	546	3509.41	41.36	34.1	33.7	23.8	8.4
White	60,873	3518.63	43.53	28.7	29.5	28.2	13.6
Missing	38	3488.66	40.18	57.9	31.6	5.3	5.3
Special Education	13,272	3487.67	44.64	58.2	23.2	13.6	5.0
English Learner (EL)	11,530	3489.04	37.35	55.7	30.2	12.0	2.1
Low Socioeconomic Status (SES)	45,873	3506.08	41.33	38.5	32.0	22.4	7.0
Migrant	514	3498.00	37.50	45.9	31.3	20.6	2.1

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*. Statistics for subgroups with less than 11 students are omitted in compliance with FERPA regulations and replaced with an asterisk (*).

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	78,836	3547.84	46.54	36.8	23.6	25.9	13.7
Male	39,884	3551.81	48.40	34.1	22.2	26.9	16.7
Female	38,952	3543.77	44.19	39.5	24.9	24.9	10.7
Hispanic	38,694	3536.77	43.04	45.5	25.1	21.8	7.6
Non-Hispanic	40,142	3558.51	47.29	28.4	22.1	29.9	19.6
American Indian	4,119	3526.67	39.27	54.8	25.2	16.1	3.8
Asian	3,030	3580.09	44.70	14.3	16.7	34.4	34.6
Black or African American	5,995	3530.65	41.79	51.3	24.5	18.9	5.3
Multi-racial	5,283	3554.61	46.54	31.1	23.7	28.2	17.0
Native Hawaiian or Other Pacific Islander	449	3540.31	43.71	38.3	29.8	23.2	8.7
White	59,926	3548.85	46.29	35.7	23.7	26.7	13.9
Missing	34	3526.68	54.04	58.8	11.8	17.6	11.8
Special Education	13,392	3517.19	43.08	65.9	17.0	12.5	4.7
English Learner (EL)	10,332	3512.34	34.42	70.3	20.4	8.1	1.2
Low Socioeconomic Status (SES)	44,181	3534.36	42.50	47.7	25.2	20.3	6.7
Migrant	487	3523.53	41.85	57.9	23.6	14.6	3.9

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	81,638	3583.69	40.17	33.6	29.5	23.9	13.0
Male	41,412	3585.74	42.17	33.2	27.3	24.2	15.2
Female	40,226	3581.59	37.88	34.0	31.8	23.6	10.6
Hispanic	39,984	3574.00	36.07	42.0	31.4	19.7	6.9
Non-Hispanic	41,654	3593.00	41.67	25.6	27.7	27.9	18.8
American Indian	4,358	3564.48	32.18	51.9	30.9	14.0	3.1
Asian	3,107	3615.35	41.61	11.0	20.3	32.3	36.4
Black or African American	6,115	3569.88	35.63	46.8	30.0	17.6	5.6
Multi-racial	5,304	3588.13	39.69	29.2	29.4	26.8	14.6
Native Hawaiian or Other Pacific Islander	487	3579.53	36.00	32.4	36.8	23.8	7.0
White	62,235	3584.48	39.88	32.6	29.8	24.5	13.1
Missing	32	3561.19	44.27	71.9	9.4	12.5	6.3
Special Education	13,213	3555.93	34.00	65.3	21.6	9.6	3.6
English Learner (EL)	9,413	3552.92	28.09	68.5	23.4	6.6	1.4
Low Socioeconomic Status (SES)	45,001	3572.04	35.63	44.0	31.4	18.4	6.3
Migrant	472	3569.19	35.35	46.6	30.5	18.2	4.7

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	81,413	3606.43	40.56	47.9	22.8	18.1	11.2
Male	41,262	3608.25	42.02	46.4	22.2	18.6	12.9
Female	40,151	3604.57	38.92	49.4	23.5	17.6	9.5
Hispanic	39,455	3596.35	36.22	58.2	22.5	13.4	5.8
Non-Hispanic	41,958	3615.91	42.11	38.2	23.1	22.4	16.3
American Indian	4,437	3589.02	34.24	66.6	19.7	9.7	4.0
Asian	3,170	3638.76	42.40	18.6	20.9	28.7	31.9
Black or African American	6,185	3590.96	34.77	64.1	20.4	11.2	4.3
Multi-racial	5,226	3610.45	40.96	43.1	24.6	19.9	12.5
Native Hawaiian or Other Pacific Islander	469	3598.83	36.41	56.1	22.0	15.6	6.4
White	61,895	3607.31	40.16	46.7	23.3	18.7	11.3
Missing	31	3575.84	31.76	80.6	9.7	9.7	0.0
Special Education	11,991	3576.93	31.85	80.3	12.1	5.3	2.3
English Learner (EL)	8,896	3574.92	26.96	84.4	10.9	3.9	0.9
Low Socioeconomic Status (SES)	44,295	3594.30	35.52	60.7	21.8	12.3	5.2
Migrant	505	3592.53	35.69	63.0	19.0	13.5	4.6

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	81,135	3628.84	39.59	57.0	15.8	13.9	13.2
Male	41,317	3631.32	40.79	54.6	15.7	14.5	15.2
Female	39,818	3626.27	38.13	59.5	16.0	13.4	11.1
Hispanic	39,064	3618.48	34.01	68.5	14.4	10.3	6.8
Non-Hispanic	42,071	3638.46	41.91	46.3	17.2	17.3	19.2
American Indian	4,484	3609.86	30.30	78.4	10.9	7.2	3.5
Asian	3,144	3667.25	44.65	22.3	13.1	19.8	44.8
Black or African American	5,910	3614.17	32.18	73.6	12.8	8.3	5.3
Multi-racial	5,121	3633.19	39.70	51.0	17.6	16.4	15.0
Native Hawaiian or Other Pacific Islander	470	3621.36	34.63	64.5	16.6	12.1	6.8
White	61,960	3629.38	39.01	56.1	16.5	14.5	12.9
Missing	46	3602.65	25.54	82.6	15.2	0.0	2.2
Special Education	10,728	3601.79	27.58	86.8	6.7	3.9	2.6
English Learner (EL)	8,531	3598.40	22.40	91.7	5.1	2.2	1.0
Low Socioeconomic Status (SES)	42,951	3616.64	33.12	70.5	14.0	9.7	5.9
Migrant	499	3612.53	35.64	76.6	10.8	4.6	8.0

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*

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

Subgroup	N	SS Mean	SS SD	%L1	%L2	%L3	%L4
All	82,495	3655.33	36.72	53.4	19.8	15.9	10.9
Male	41,864	3656.04	38.59	53.1	18.6	16.0	12.2
Female	40,631	3654.61	34.69	53.6	20.9	15.9	9.6
Hispanic	39,728	3646.27	30.98	64.0	18.7	11.7	5.6
Non-Hispanic	42,767	3663.75	39.54	43.5	20.7	19.9	15.9
American Indian	4,632	3639.27	26.35	73.9	15.4	7.7	3.0
Asian	3,059	3692.52	44.59	19.3	16.6	25.6	38.6
Black or African American	6,104	3643.97	29.98	67.1	17.2	11.2	4.5
Multi-racial	5,102	3659.99	37.01	47.0	21.7	18.4	12.8
Native Hawaiian or Other Pacific Islander	510	3649.69	31.91	59.0	21.8	12.4	6.9
White	63,040	3655.49	36.19	52.6	20.3	16.4	10.7
Missing	48	3638.42	24.15	68.8	20.8	10.4	0.0
Special Education	10,389	3632.00	23.40	85.2	9.3	3.7	1.8
English Learner (EL)	7,503	3630.06	20.10	87.8	8.5	2.8	0.9
Low Socioeconomic Status (SES)	43,156	3644.84	30.26	66.1	18.1	10.7	5.1
Migrant	500	3637.72	28.92	78.2	11.8	5.8	4.2

Note. SS = scale score, SD = standard deviation, L1 = Level 1: *Minimally Proficient*, L2 = Level 2: *Partially Proficient*, L3 = Level 3: *Proficient*, L4 = Level 4: *Highly Proficient*

Table C.13. State Mean Raw Score Points by Reporting Category and Cluster—ELA Online

Grade	Reporting Category/Cluster	N	#Points Possible	State Mean Points
3	Reading for Information	76,012	21	10.2
	Key Ideas and Details	76,012	5	2.5
	Craft and Structure	76,012	***	***
	Integration of Knowledge and Ideas	76,012	10	4.5
	Oral Reading Fluency	76,012	***	***
	Reading for Literature	76,012	18	8.4
	Key Ideas and Details	76,012	7	3.1
	Craft and Structure	76,012	6	3.2
	Integration of Knowledge and Ideas	76,012	***	***
	Oral Reading Fluency	76,012	***	***
	Writing and Language	76,012	20	10.1
	Writing Essay	76,012	10	4.7
	Language Items	76,012	10	5.5
4	Reading for Information	74,849	19	10.0
	Key Ideas and Details	74,849	7	4.0
	Craft and Structure	74,849	***	***
	Integration of Knowledge and Ideas	74,849	8	4.0
	Reading for Literature	74,849	16	8.0
	Key Ideas and Details	74,849	9	4.3
	Craft and Structure	74,849	***	***
	Integration of Knowledge and Ideas	74,849	***	***
	Writing and Language	74,849	20	11.2
	Writing Essay	74,849	10	5.3
	Language Items	74,849	10	5.9
5	Reading for Information	78,067	18	9.2
	Key Ideas and Details	78,067	7	3.2
	Craft and Structure	78,067	***	***
	Integration of Knowledge and Ideas	78,067	7	4.1
	Reading for Literature	78,067	17	8.7
	Key Ideas and Details	78,067	8	4.4
	Craft and Structure	78,067	8	3.9
	Integration of Knowledge and Ideas	78,067	***	***
	Writing and Language	78,067	20	10.4
	Writing Essay	78,067	10	5.0
	Language Items	78,067	10	5.4

Appendix C: Spring 2025 Administration Results

Grade	Reporting Category/Cluster	N	#Points Possible	State Mean Points
6	Reading for Information	77,349	18	9.5
	Key Ideas and Details	77,349	***	***
	Craft and Structure	77,349	7	3.7
	Integration of Knowledge and Ideas	77,349	8	3.8
	Reading for Literature	77,349	17	8.8
	Key Ideas and Details	77,349	10	4.8
	Craft and Structure	77,349	6	3.6
	Integration of Knowledge and Ideas	77,349	***	***
	Writing and Language	77,349	20	11.2
	Writing Essay	77,349	10	5.7
	Language Items	77,349	10	5.4
7	Reading for Information	77,194	20	11.2
	Key Ideas and Details	77,194	***	***
	Craft and Structure	77,194	8	4.6
	Integration of Knowledge and Ideas	77,194	8	4.6
	Reading for Literature	77,194	15	7.9
	Key Ideas and Details	77,194	8	4.2
	Craft and Structure	77,194	5	2.7
	Integration of Knowledge and Ideas	77,194	***	***
	Writing and Language	77,194	20	11.0
	Writing Essay	77,194	10	5.7
	Language Items	77,194	10	5.4
8	Reading for Information	78,664	18	9.9
	Key Ideas and Details	78,664	6	3.0
	Craft and Structure	78,664	6	3.4
	Integration of Knowledge and Ideas	78,664	6	3.5
	Reading for Literature	78,664	17	8.4
	Key Ideas and Details	78,664	11	5.4
	Craft and Structure	78,664	***	***
	Integration of Knowledge and Ideas	78,664	***	***
	Writing and Language	78,664	20	11.7
	Writing Essay	78,664	10	5.7
	Language Items	78,664	10	6.0

Table C.14. State Mean Raw Score Points by Reporting Category and Cluster—Mathematics Online

Grade	Reporting Category/Cluster	N	#Points Possible	State Mean Points
3	Operations, Algebraic Thinking, and Numbers in Base Ten	77,537	23	13.6
	Represent and solve problems involving whole number multiplication and division	77,537	9	5.3
	Understand properties of multiplication and the relationship between multiplication and division	77,537	***	***
	Multiply and divide within 100	77,537	***	***
	Solve problems involving the four operations, and identify and explain patterns in arithmetic	77,537	***	***
	Use place value understanding and properties of operations to perform multi-digit arithmetic	77,537	5	3.0
	Numbers and Operations – Fractions	77,537	9	5.2
	Understand fractions as numbers	77,537	9	5.2
	Measurement, Data, and Geometry	77,537	13	7.5
	Solve problems involving measurement	77,537	***	***
	Represent and interpret data	77,537	***	***
	Geometric measurement: Understand concepts of area and perimeter	77,537	7	4.0
	Reason with shapes and their attributes	77,537	***	***
4	Operations, Algebraic Thinking, and Numbers in Base Ten	75,503	23	11.9
	Use the four operations with whole numbers to solve problems	75,503	9	4.7
	Gain familiarity with factors and multiples	75,503	***	***
	Generate and analyze patterns	75,503	***	***
	Generalize place value understanding for multi-digit whole numbers	75,503	***	***
	Use place value understanding and properties of operations to perform multi-digit arithmetic	75,503	9	4.9
	Numbers and Operations – Fractions	75,503	14	7.3
	Extend understanding of fraction equivalence and ordering	75,503	***	***
	Build fractions from unit fractions by applying and extending previous understanding of operations on whole numbers	75,503	7	4.4
	Understand decimal notation for fractions, and compare decimal fractions	75,503	***	***
	Measurement, Data, and Geometry	75,503	8	4.2
	Solve problems involving measurement and conversion of measurements from a larger unit to a smaller unit	75,503	***	***
	Represent and interpret data	75,503	***	***
	Geometric measurement: Understand concepts of angle and measure angles	75,503	***	***
	Draw and identify lines and angles, and classify shapes by properties of their lines and angles	75,503	***	***

Appendix C: Spring 2025 Administration Results

Grade	Reporting Category/Cluster	N	#Points Possible	State Mean Points
5	Operations, Algebraic Thinking, and Numbers in Base Ten	78,373	18	8.8
	Write and interpret numerical expressions	78,373	***	***
	Analyze patterns and relationships	78,373	***	***
	Understand the place value system	78,373	***	***
	Perform operations with multi-digit whole numbers and with decimals to hundredths	78,373	11	5.6
	Numbers and Operations – Fractions	78,373	15	7.1
	Use equivalent fractions to add and subtract fractions	78,373	5	2.2
	Use previous understandings of multiplication and division to multiply and divide fractions	78,373	10	4.9
	Measurement, Data, and Geometry	78,373	12	5.7
	Convert like measurement units within a given measurement system	78,373	***	***
	Represent and interpret data	78,373	***	***
	Geometric measurement: Understand concepts of volume and relate volume to multiplication and to addition	78,373	6	3.2
	Graph points on the coordinate plane to solve mathematical problems as well as problems in real-world context	78,373	***	***
	Classify two-dimensional figures into categories based on their properties	78,373	***	***
6	Ratio and Proportional Relationships	77,673	10	4.7
	Understand ratio concepts and use ratio reasoning to solve problems	77,673	10	4.7
	The Number System	77,673	14	6.0
	Apply and extend previous understanding of multiplication and division to divide fractions by fractions	77,673	5	1.8
	Compute fluently with multi-digit numbers and find common factors and multiples	77,673	5	2.1
	Apply and extend previous understanding of numbers to the system of rational numbers	77,673	***	***
	Expressions and Equations	77,673	15	6.5
	Apply and extend previous understanding of arithmetic to algebraic expressions	77,673	***	***
	Reason about and solve one-variable equations and inequalities	77,673	8	3.7
	Represent and analyze quantitative relationships between dependent and independent events	77,673	***	***
	Geometry, Statistics and Probability	77,673	8	3.5
	Solve mathematical problems and problems in real-world context involving area, surface area, and volume	77,673	5	2.4
	Develop understanding of statistical variability	77,673	***	***
	Summarize and describe distributions	77,673	***	***

Appendix C: Spring 2025 Administration Results

Grade	Reporting Category/Cluster	N	#Points Possible	State Mean Points
7	Ratio and Proportional Relationships	77,671	10	4.3
	Analyze proportional relationships and use them to solve mathematical problems and problems in real-world context	77,671	10	4.3
	The Number System	77,671	10	4.4
	Apply and extend previous understanding of operations with fractions to add, subtract, multiply and divide rational numbers except division by zero	77,671	10	4.4
	Expressions and Equations	77,671	12	5.1
	Use properties of operations to generate equivalent expressions	77,671	***	***
	Solve mathematical problems and problems in real-world context using numerical and algebraic expressions and equations	77,671	8	3.3
	Geometry, Statistics and Probability	77,671	15	6.5
	Draw, construct, and describe geometrical figures, and describe the relationships between them	77,671	***	***
	Solve mathematical problems and problems in real-world context involving angle measure, area, surface area, and volume	77,671	6	2.1
	Use random sampling to draw inferences about a population	77,671	***	***
	Draw informal comparative inferences about two populations	77,671	***	***
	Investigate chance processes and develop, use, and evaluate probability models	77,671	***	***
8	Expressions and Equations	79,148	15	6.2
	Work with radicals and integer exponents	79,148	6	2.4
	Understand the connections between proportional relationships, lines, and linear equations	79,148	***	***
	Analyze and solve linear equations, inequalities, and pairs of simultaneous linear equations	79,148	6	2.5
	Functions	79,148	11	4.5
	Define, evaluate, and compare functions	79,148	7	2.5
	Use functions to model relationships between quantities	79,148	***	***
	Geometry	79,148	9	3.8
	Understand congruence and similarity	79,148	***	***
	Understand and apply the Pythagorean Theorem	79,148	***	***
	Solve real-world and mathematical problems involving volume of cylinders, cones, and spheres	79,148	***	***
	Statistics and Probability and The Number System	79,148	12	5.1
	Understand that there are irrational numbers, and approximate them using rational numbers	79,148	8	3.0
	Investigate patterns of association in bivariate data	79,148	***	***
	Investigate chance processes and develop, use, and evaluate probability models	79,148	***	***

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

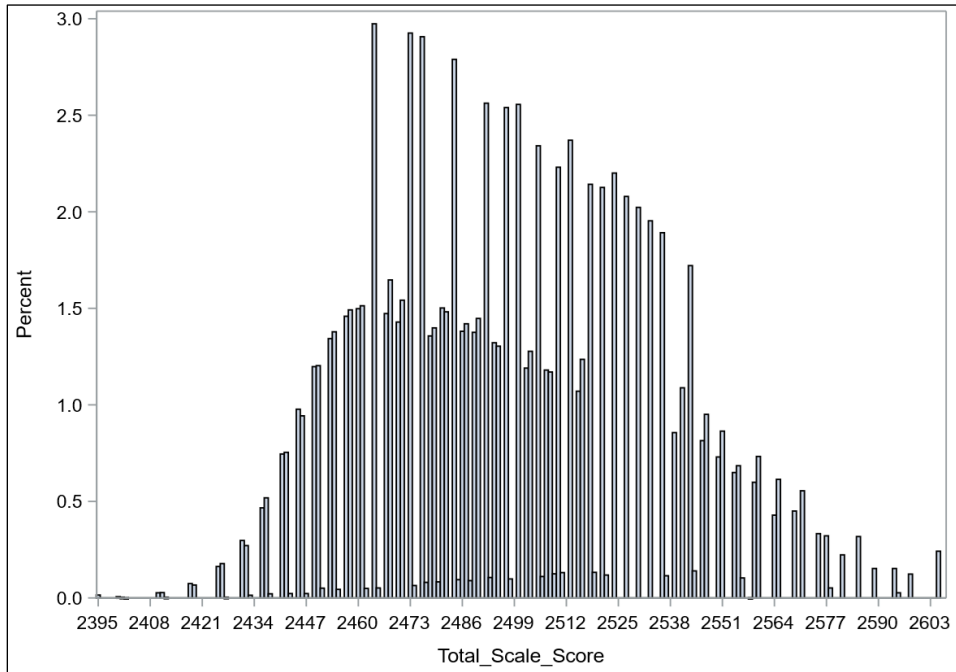


Figure C.2. Total Scale Score Distribution, ELA Grade 4

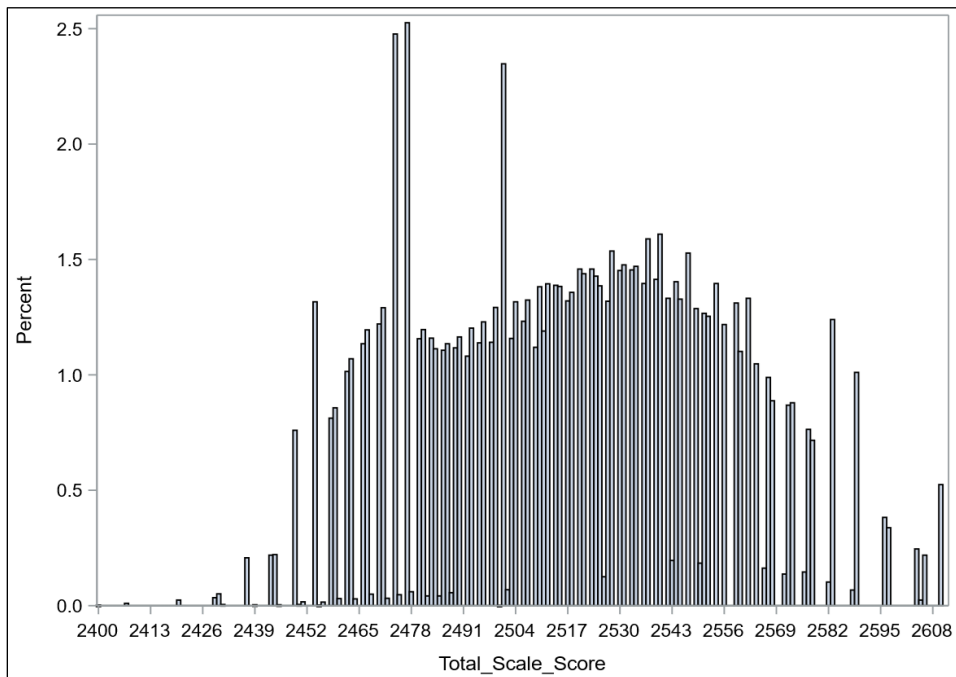


Figure C.3. Total Scale Score Distribution, ELA Grade 5

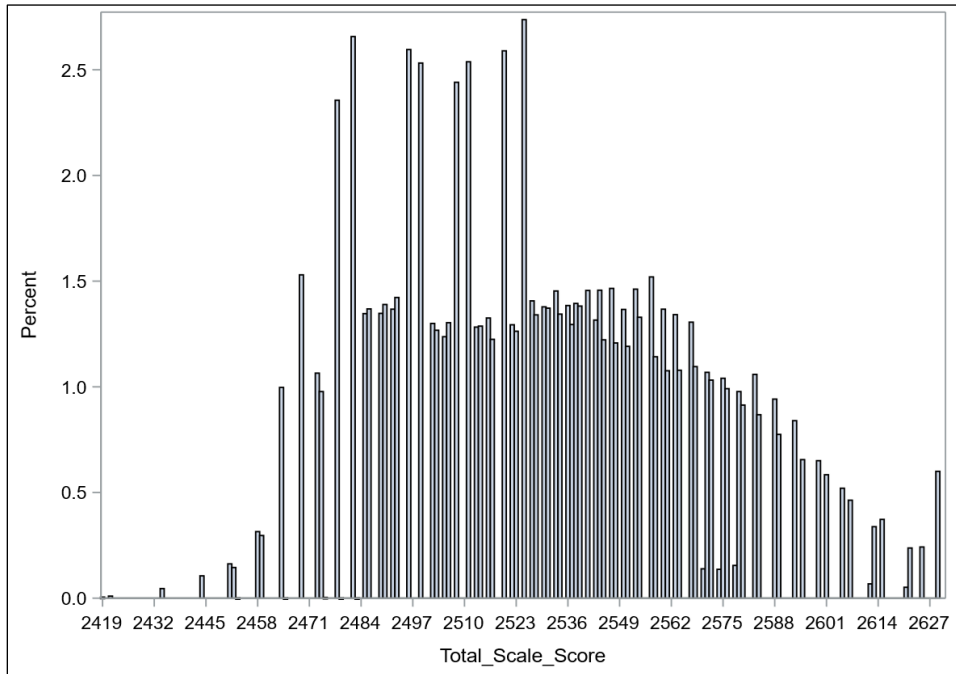


Figure C.4. Total Scale Score Distribution, ELA Grade 6

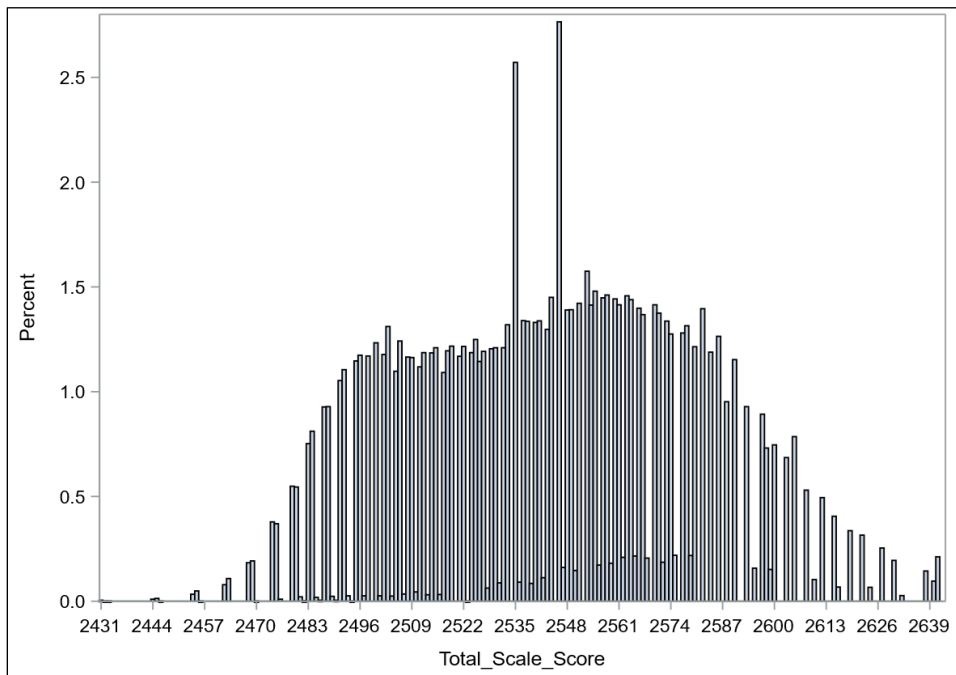


Figure C.5. Total Scale Score Distribution, ELA Grade 7

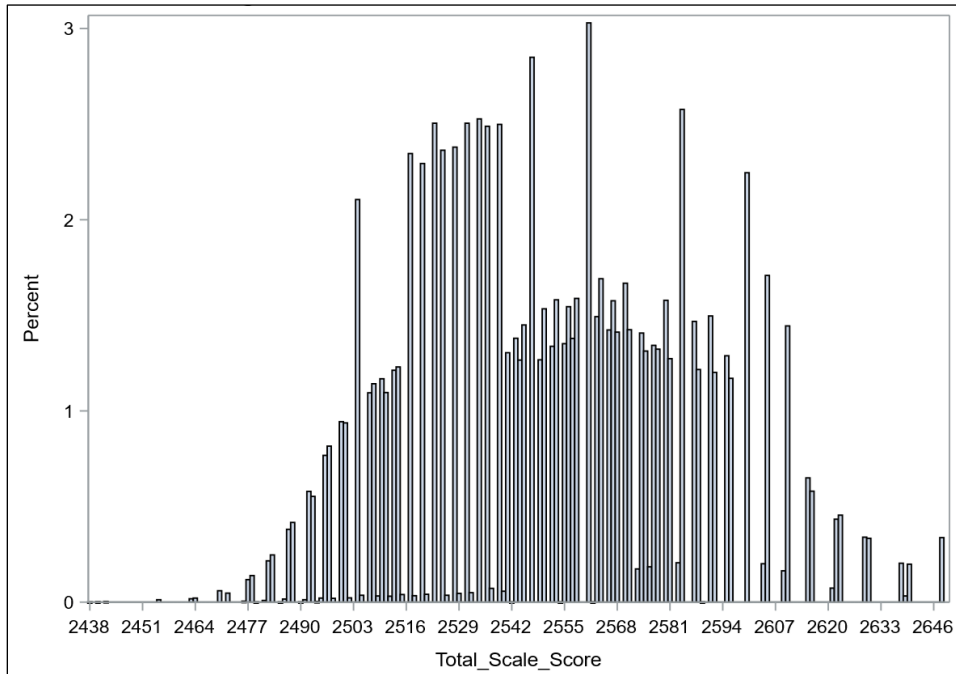


Figure C.6. Total Scale Score Distribution, ELA Grade 8

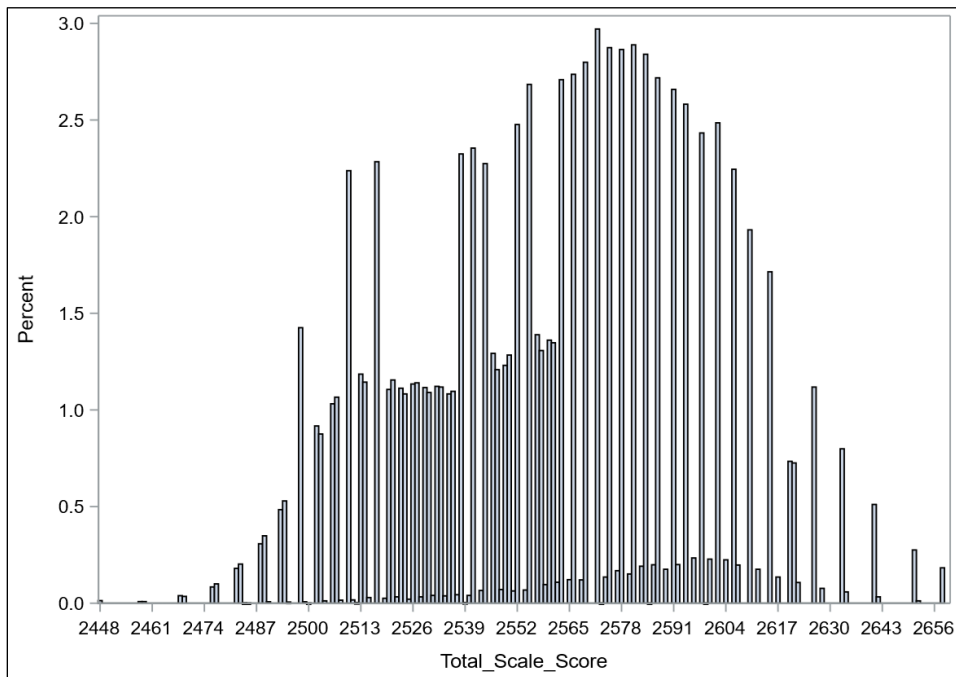


Figure C.7. Total Scale Score Distribution, Mathematics Grade 3

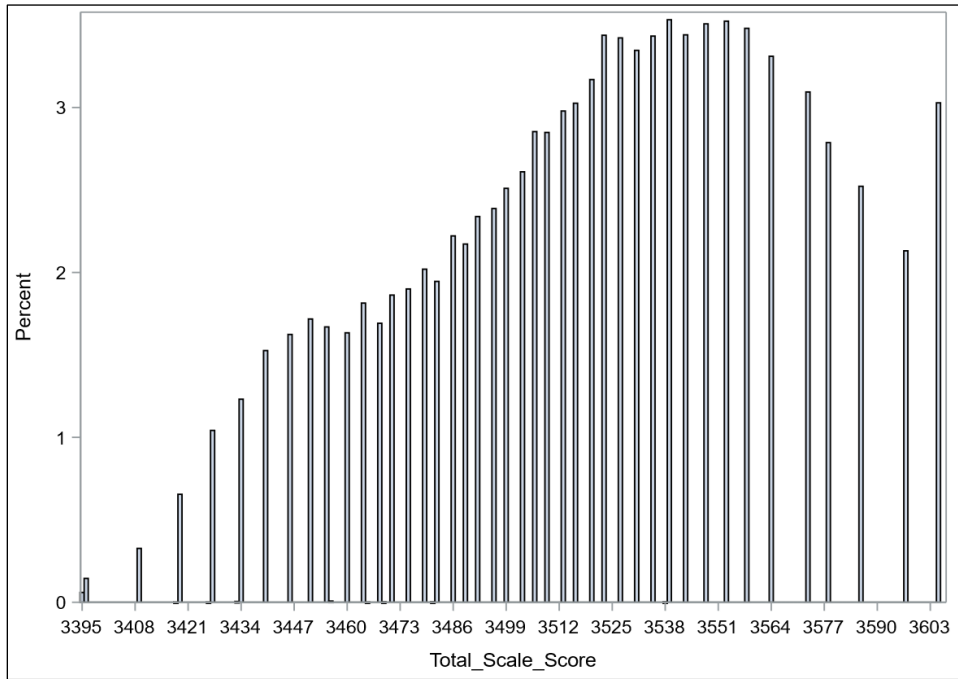


Figure C.8. Total Scale Score Distribution, Mathematics Grade 4

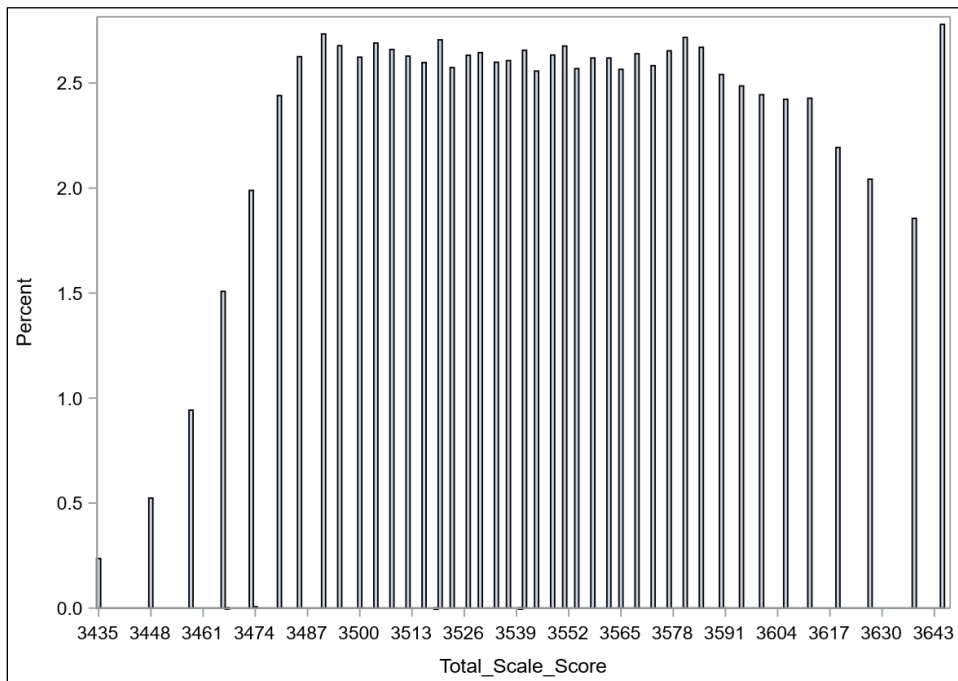


Figure C.9. Total Scale Score Distribution, Mathematics Grade 5

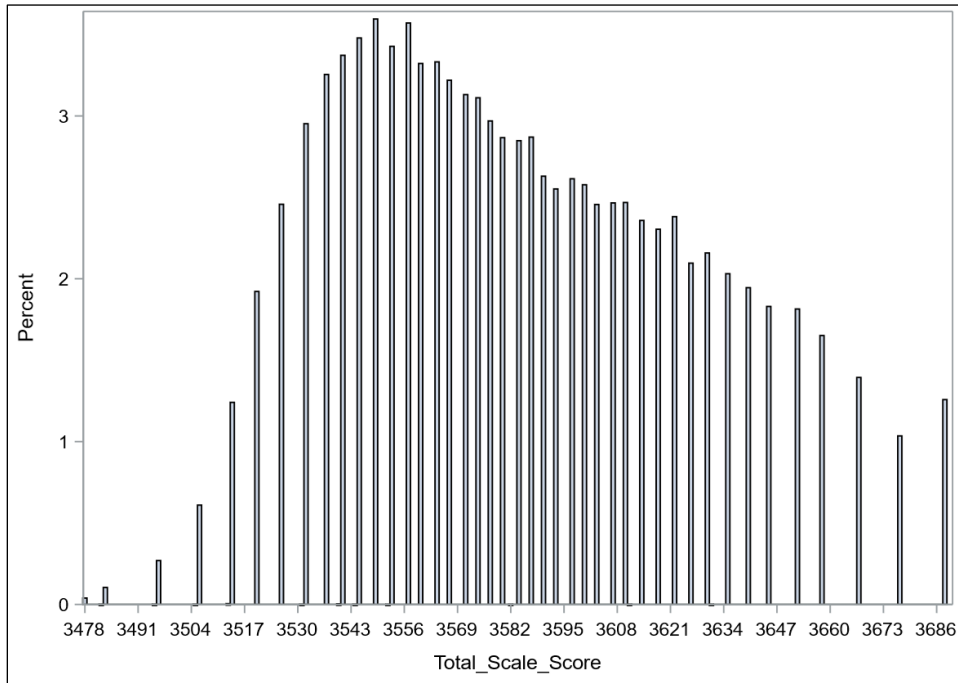


Figure C.10. Total Scale Score Distribution, Mathematics Grade 6

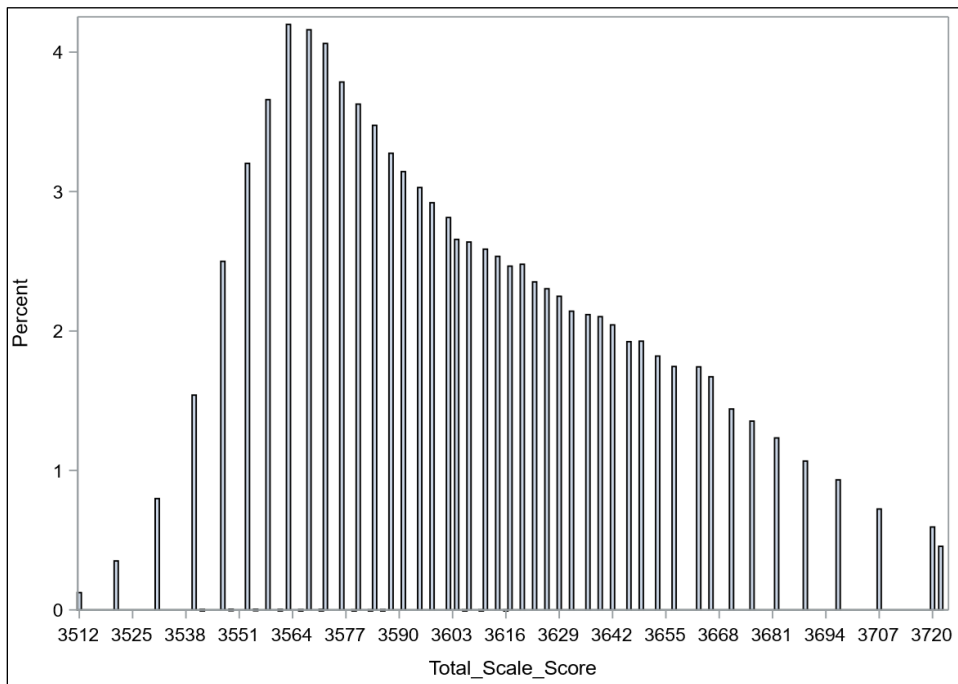


Figure C.11. Total Scale Score Distribution, Mathematics Grade 7

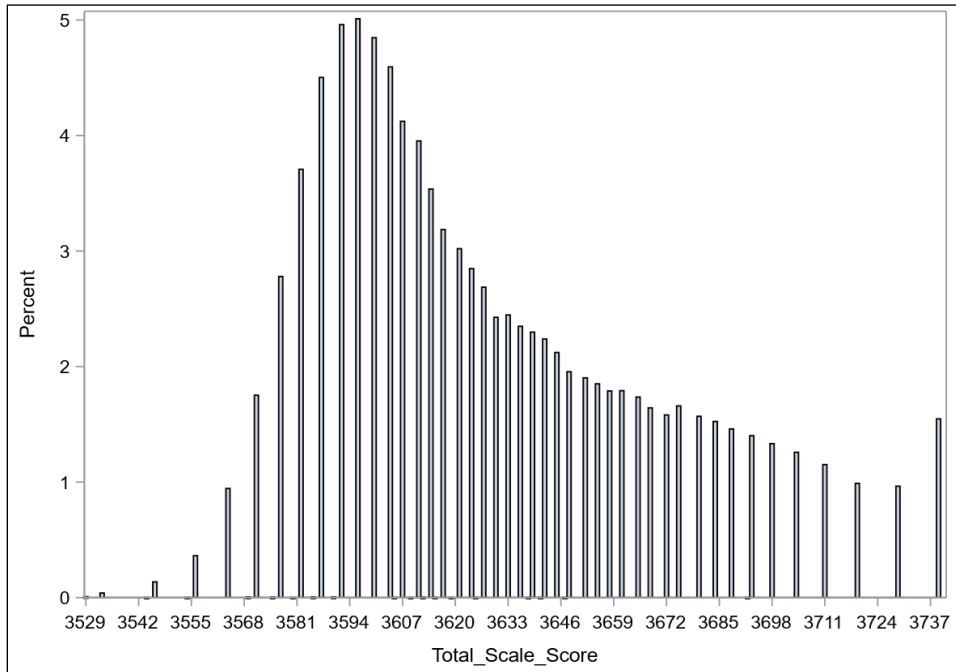


Figure C.12. Total Scale Score Distribution, Mathematics Grade 8

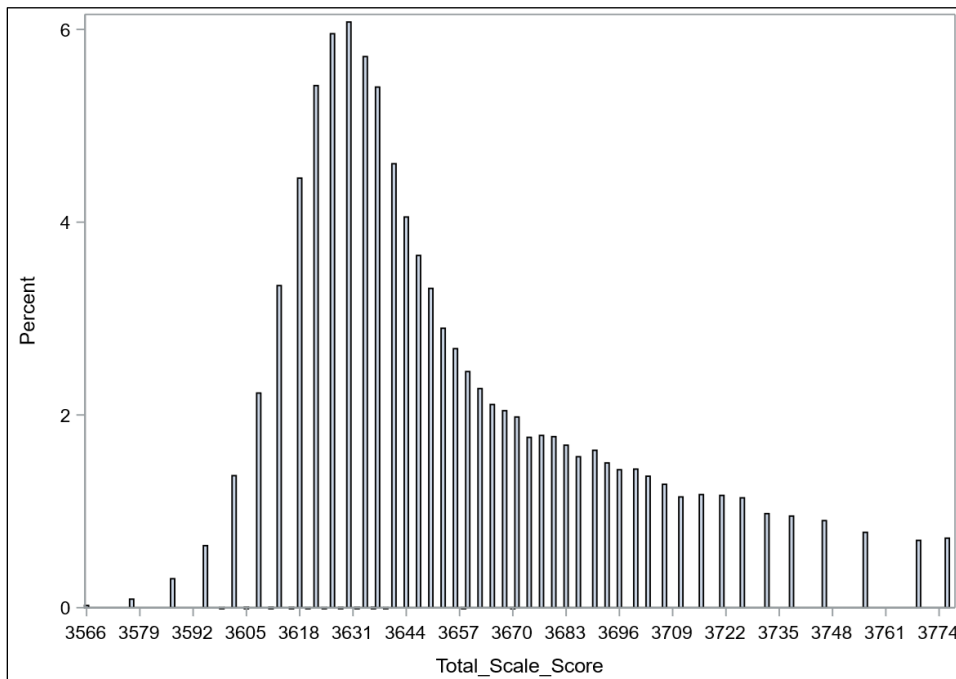


Figure C.13. Trend in % of Students On-Grade, ELA Grade 3

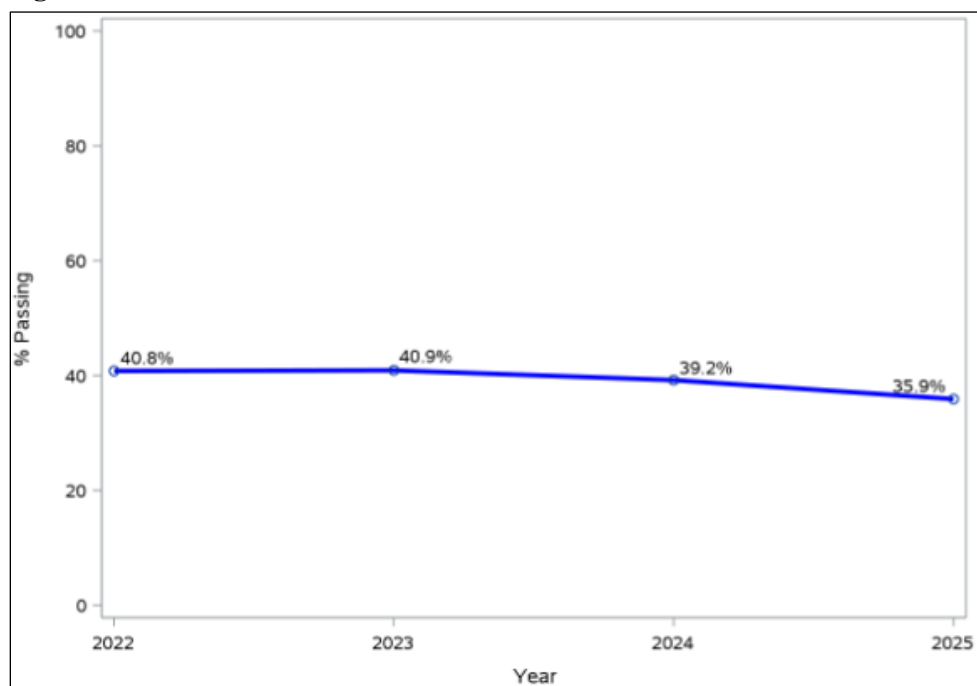


Figure C.14. Trend in % of Students On-Grade, ELA Grade 4

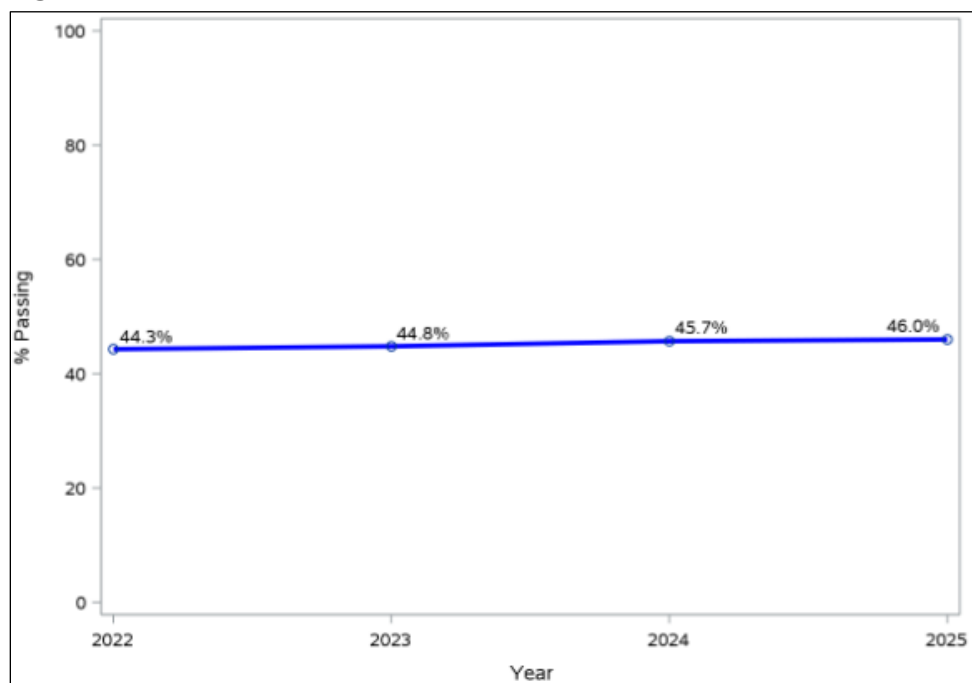


Figure C.15. Trend in % of Students On-Grade, ELA Grade 5

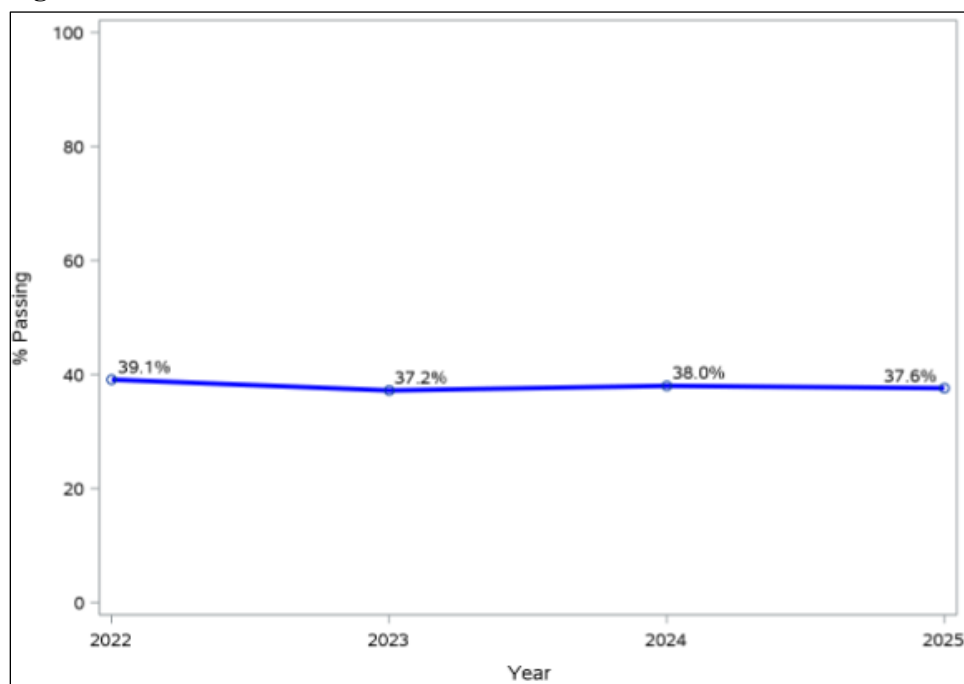


Figure C.16. Trend in % of Students On-Grade, ELA Grade 6

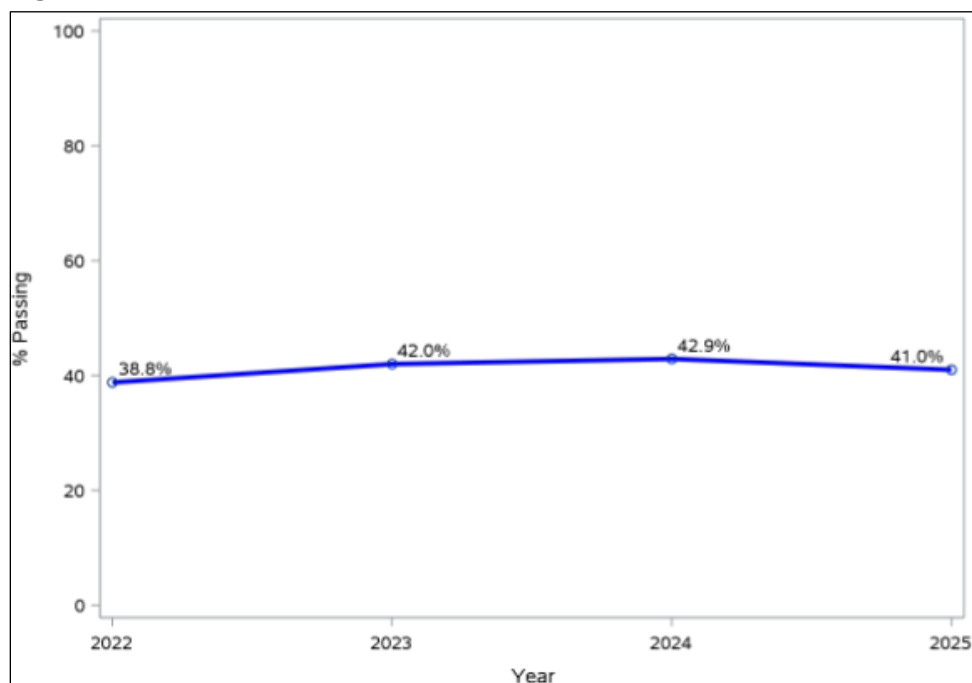


Figure C.17. Trend in % of Students On-Grade, ELA Grade 7

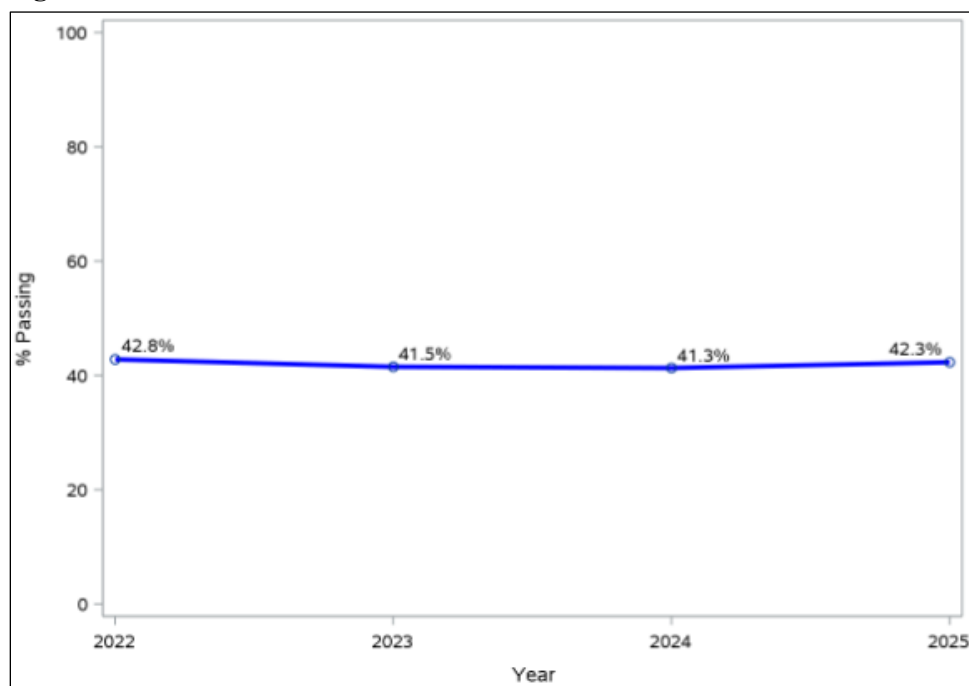


Figure C.18. Trend in % of Students On-Grade, ELA Grade 8

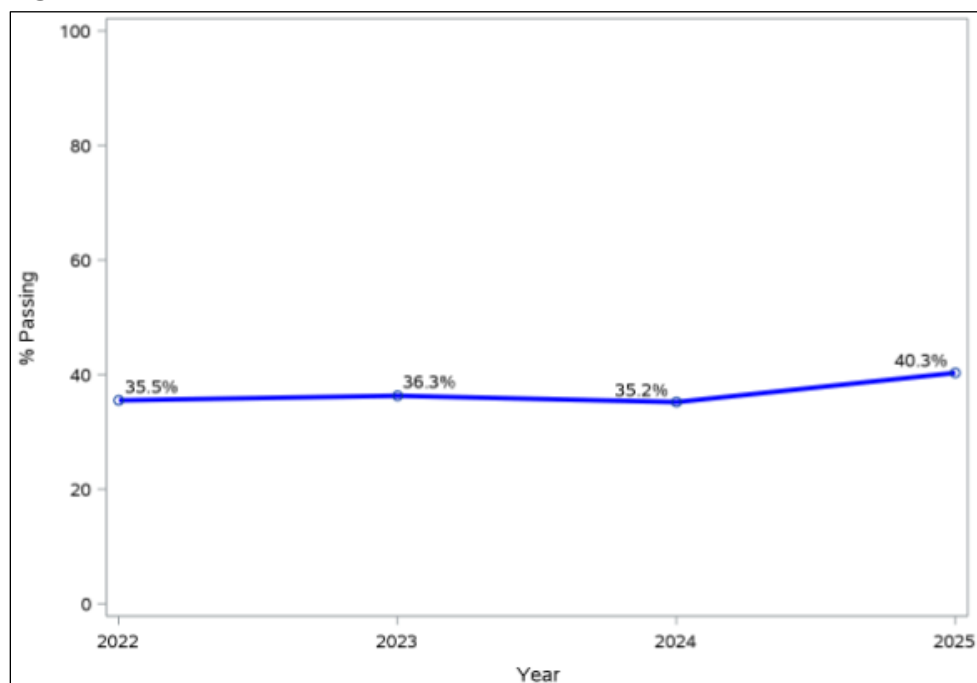


Figure C.19. Trend in % of Students On-Grade, Mathematics Grade 3

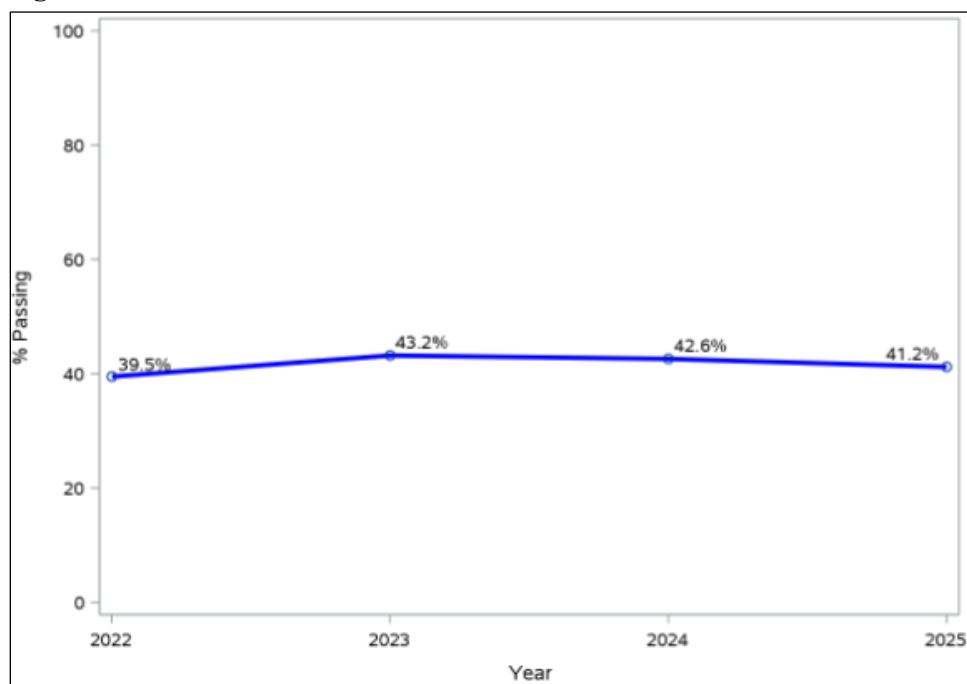


Figure C.20. Trend in % of Students On-Grade, Mathematics Grade 4

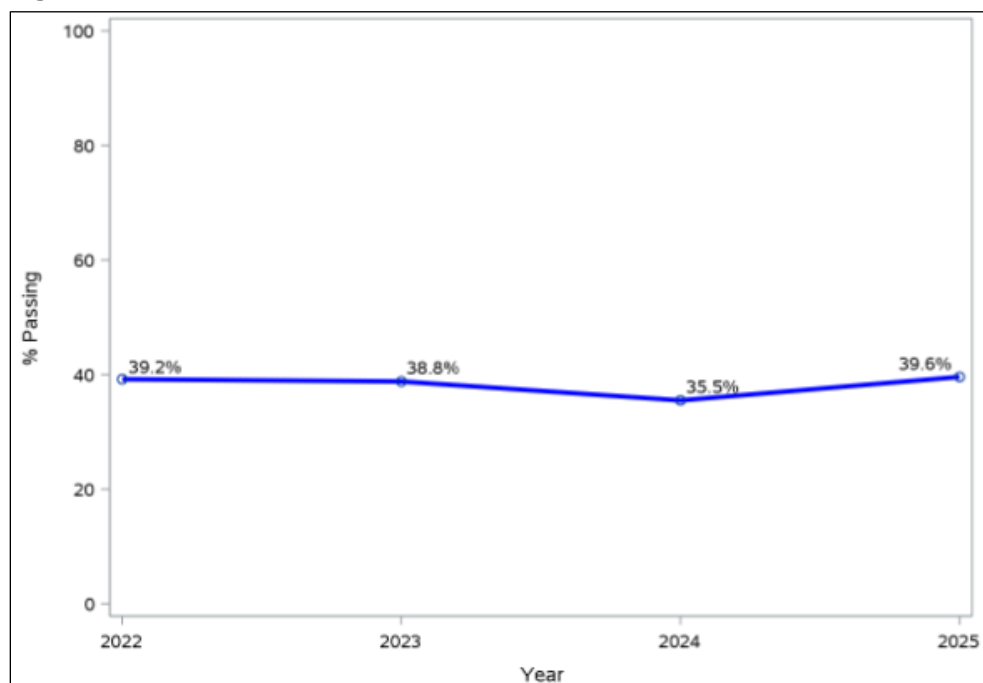


Figure C.21. Trend in % of Students On-Grade, Mathematics Grade 5

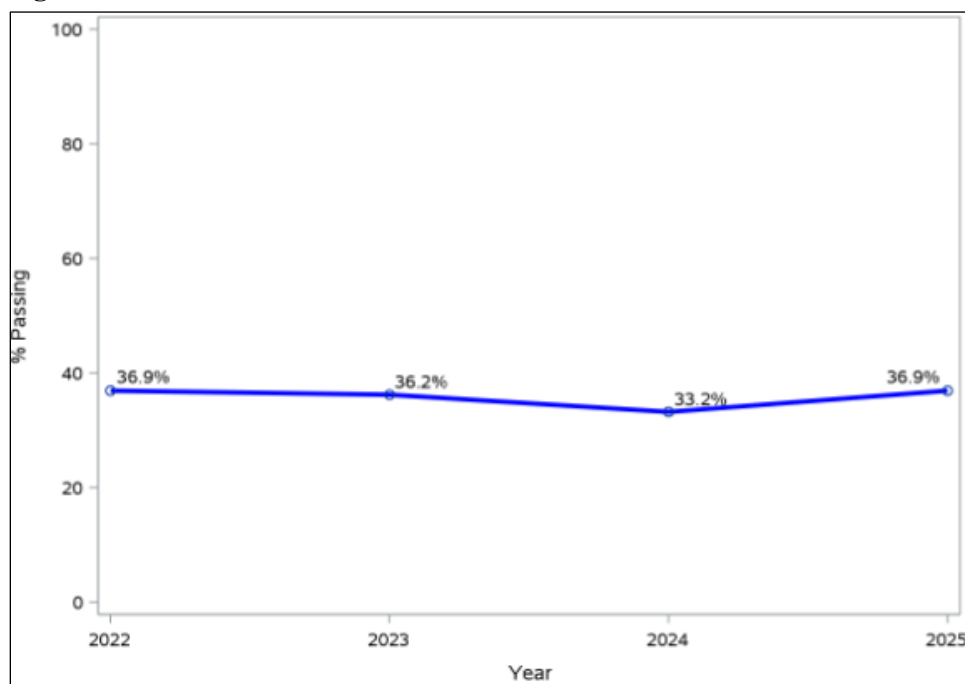


Figure C.22. Trend in % of Students On-Grade, Mathematics Grade 6

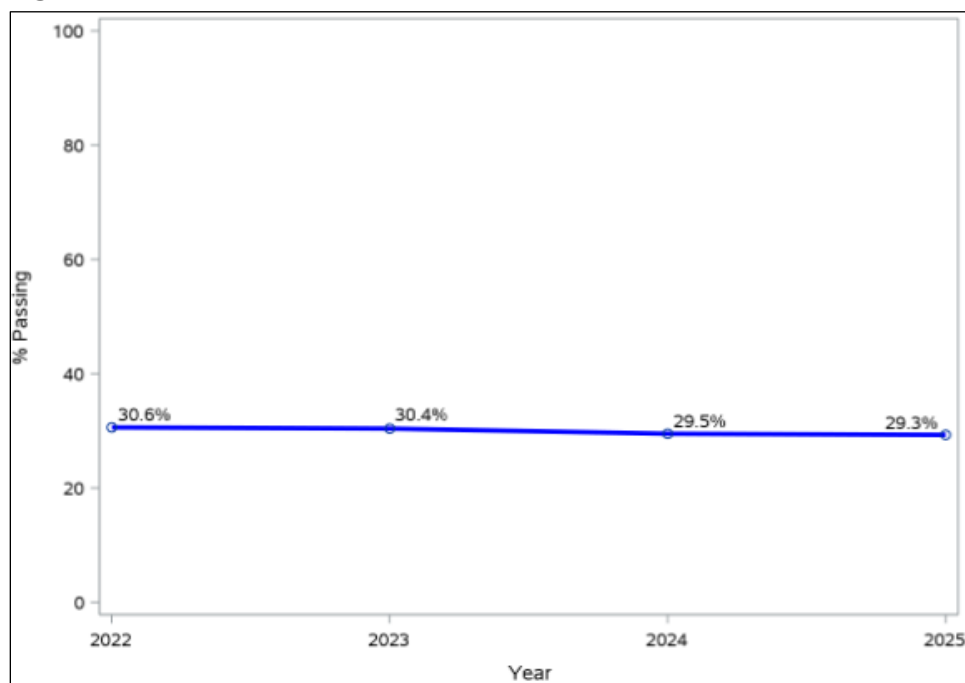


Figure C.23. Trend in % of Students On-Grade, Mathematics Grade 7

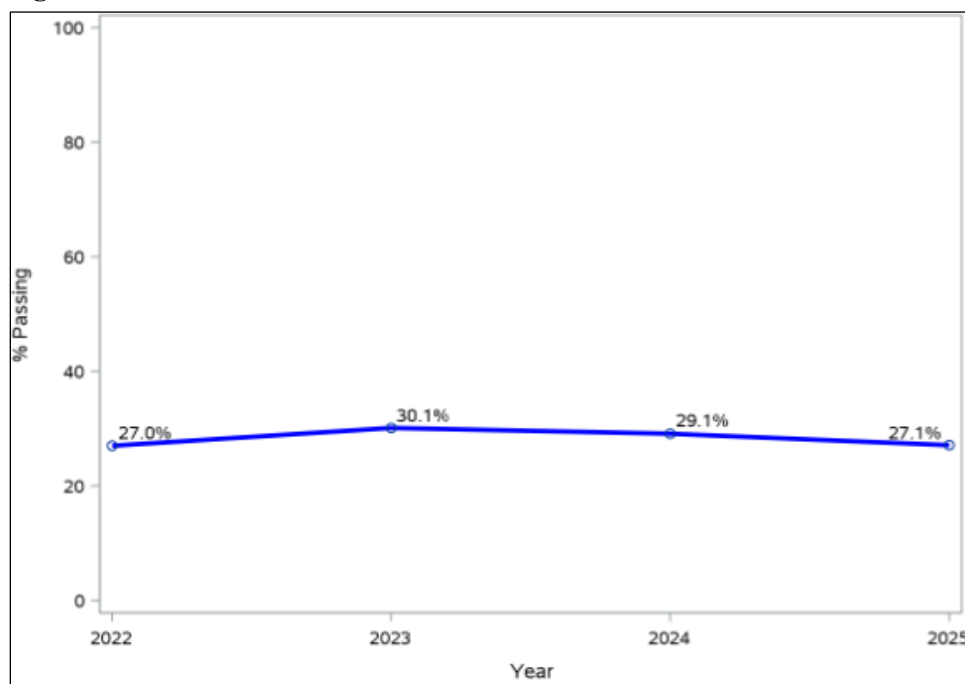


Figure C.24. Trend in % of Students On-Grade, Mathematics Grade 8

