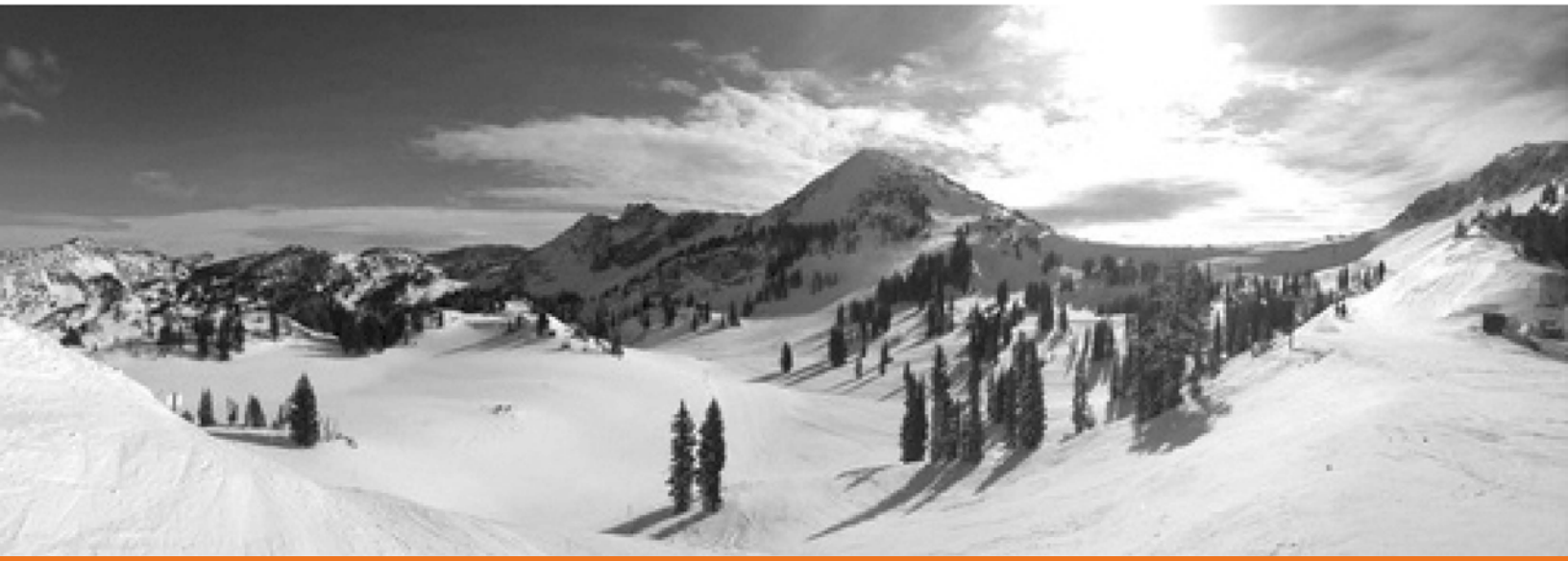
Colorado Measures of Academic Success



Technical Report

Math, English Language Arts (ELA), and Science

2023

Copyright © 2023 by the Colorado Department of Education. All rights reserved.

Table of Contents

Chapter 1: Introduction 13
1.1. Testing Requirements 13
1.2. Intended Population 13
1.3. CMAS Background 14
1.4. Purpose of CMAS 15
1.5. Assessment Development Partners 15
1.5.1. Colorado Department of Education 15
1.5.2. Colorado Educator Community 16
1.5.3. Pearson 17
1.5.4. Tri-Lin Integrated Services, Inc. 17
1.5.5. Colorado Technical Advisory Committee 17
Chapter 2: Test Design 18
2.1. Colorado Academic Standards 18
2.2. Test Frameworks and Blueprints 20
2.2.1. Mathematics and ELA 21
2.2.2. Science 21
2.3. Claims and Subclaims 21
2.4. Performance Levels 23
2.5. Cognitive Complexity 24
2.6. Test Design 25
2.7. Timing of Tests 26
Chapter 3: Item Development 27
3.1. Item Banking System 27
3.2. Item Development Plan 28
3.3. ELA Passage Development 28
3.4. Science Scenario Development 28
3.5. Item Writing 29
3.6. Item Review 29
3.6.1. Internal Review 29
3.6.2. CDE Review 30
3.6.3. External Content and Bias Review 30
3.7. Data Review 31
Chapter 4: Test Construction 34
4.1. Test Form Construction 34
4.2. Accommodated Test Forms 35
4.2.1. Paper 35
4.2.2. Braille 36
4.2.3. Large Print 36
Chapter 5: Test Administration 37
5.1. Manuals 37
5.2. Administration Training 37
5.3. Practice Resources 38

5.4. Onsite Preparation 38
5.5. Accessibility Features and Accommodations 38
5.6. Test Security 39
Chapter 6: Scoring 41
6.1. Machine Scoring 41
6.2. Human Scoring 41
6.2.1. Operational Scoring 41
6.2.2. Field Test Scoring 42
6.2.3. Rangefinding 43
6.2.4. Backreading 44
6.2.5. Calibration 44
6.2.6. Validity Papers 44
6.3. MRE Scoring 44
6.4. Automated Scoring 45
6.4.1. Quality Criteria for Evaluating Automated Scoring 45
6.4.1.1. Primary Criterion 46
6.4.1.2. Secondary Criterion 46
6.4.2. Hierarchy of Assigned Scores for Reporting 47
Chapter 7: Standard Setting 48
Chapter 8: Reporting 50
8.1. Description of Scores 50
8.1.1. Scale Scores 50
8.1.2. Performance Levels 50
8.1.3. Percentile Ranking 50
8.1.4. Percent Earned 50
8.2. Score Reports 51
8.2.1. Student Performance Reports 51
8.2.2. Aggregate Reports 51
Chapter 9: Test Results and Analysis 52
9.1. Student Participation 52
9.2. Performance Results 52
9.3. Classical Item Analysis 54
9.4. Subclaim Correlations 55
Chapter 10: Calibration, Equating, and Scaling 58
10.1. IRT Models 60
10.2. Item Response Curves (IRCs) and Item Characteristic Curves (ICCs) 60
10.3. Data Preparation 62
10.4. Checking Model Assumptions 64
10.4.1. Unidimensionality (Factor Analysis) 64
10.4.2. Item Fit 64
10.5. Calibration 65
10.6. Equating 65
10.6.1. Mathematics 66
10.6.1.1. Pre-Equating Design 66

10.6.1.2. Post-Equating Check 66
10.6.2. ELA 67
10.6.2.1. Consistency of Constructed-Response Scoring Check 67
10.6.2.2. Stability Check 67
10.6.2.3. Calibration and Anchor Set Evaluation 68
10.6.2.4. Final Anchor Sets 68
10.6.2.5. Equating Method 68
10.6.2.6. Paper Forms 68
10.6.3. CSLA 68
10.6.3.1. Stability Check 69
10.6.3.2. Final Anchor Sets 69
10.6.3.3. Equating Method 69
10.6.4. Science 69
10.6.4.1. Consistency of Constructed-Response Scoring Check 69
10.6.4.2. Stability Check 70
10.6.4.3. Calibration and Anchor Set Evaluation 70
10.6.4.4. Final Anchor Sets 70
10.6.4.5. Equating Method 70
10.6.4.6. Paper Forms 71
10.7. Field Test Equating 71
10.8. Ability Estimates 71
10.8.1. Mathematics, ELA, and Science 71
10.8.2. CSLA 72
10.9. Overall and Subscale Scale Scores 72
10.10. Item-Level IRT Statistics 72
10.11. IRT Curves 73
10.12. Comparability of Online and Paper Forms 73
Chapter 11: Reliability 75
11.1. Internal Consistency (Coefficient Alpha) 75
11.2. Standard Error of Measurement (SEM) 77
11.3. Conditional Standard Error of Measurement (CSEM) 78
11.4. Decision Consistency and Accuracy 79
11.5. Inter-Rater Agreement 81
Chapter 12: Validity 82
12.1. Evidence Based on Test Content 82
12.1.1. Test Development Process 82
12.1.2. Alignment Study 83
12.2. Evidence Based on Internal Structure 84
12.3. Evidence Based on Relationships to Other Variables 84
12.4. Evidence Based on Response Processes 85
12.5. Evidence Based on the Consequences of Testing 85
12.6. Fairness 88
References 89
Appendix A: Test Blueprints 92

Appendix B: Science Cognitive Complexity Framework 95
Appendix C: ELA and CSLA Scoring Rubrics 98
Appendix D: Sample Student Performance Reports 104
Appendix E: Student Participation by Demographic Group 110
Appendix F: Scale Score Distributions 114
Appendix G: Scale Score Distribution Graphs 146
Appendix H: Scale Score Summary Statistics by Demographic Group 155
Appendix I: Summary Statistics for Points Earned by Subclaim 166
Appendix J: Classical Item-Level Statistics 169
Appendix K: Scree Plots 187
Appendix L: Mathematics 2022 Post-Equating Check TCCs 196
Appendix M: IRT Item-Level Statistics 199
Appendix N: TCC, TIC, and CSEM Curves 210
Appendix O: Inter-Rater Agreement 235

List of Tables

Table 1.1. Schedule of Major Events 16
Table 2.1. Subclaims 22
Table 2.2. Performance Levels and Policy Claims—Mathematics and ELA 23
Table 2.3. Performance Levels and Policy Claims—Science 24
Table 2.4. Mathematics and ELA Cognitive Complexity Levels 24
Table 2.5. Science Cognitive Complexity Criteria 25
Table 2.6. Testing Times 26
Table 3.1. Refresh Rates—Mathematics and ELA 27
Table 3.2. Item Statistical Flagging Criteria 32
Table 3.3. Data Review Results 33
Table 7.1. Performance Level Cut Scores 48
Table 7.2. ELA Reading *Met Expectations* Cut Score 49
Table 7.3. 2023 CMAS Science Performance Indicator Cut Scores 49
Table 9.1. Student Participation N-Count by Form 52
Table 9.2. Scale Score Performance Summary 53
Table 9.3. Performance Level Distribution: Percent of Students in Each Performance Level 53
Table 9.4. Correlations Between Subclaims—Mathematics 55
Table 9.5. Correlations Between Subclaims—ELA 56
Table 9.6. Correlations Between Subclaims—CSLA 57
Table 9.7. Correlations between Subclaims—Reading vs. Writing 57
Table 9.8. Correlations Between Subclaims—Science 57
Table 10.1. Online vs. Paper Comparability Mode Effect Sizes 74
Table 11.1. Coefficient Alpha by Subclaim—Mathematics 76
Table 11.2. Coefficient Alpha by Subclaim—ELA 76
Table 11.3. Coefficient Alpha by Subclaim—CSLA 77
Table 11.4. Coefficient Alpha by Subclaim—Science 77
Table 11.5. SEM by Subclaim—Mathematics 77

Table 11.6. SEM by Subclaim—ELA 78
Table 11.7. SEM by Subclaim—CSLA 78
Table 11.8. SEM by Subclaim—Science 78
Table 11.9. Kappa Values 79
Table 11.10. Decision Consistency and Accuracy Estimates 80
Table 11.11. Accuracy of Cut Scores 80
Table 11.12. Consistency of Cut Scores 81
Table 12.1. Correlations Between CMAS Scale Scores 84
Table 12.2. Student Performance Over Time 87
Table A.1. Test Blueprint—Mathematics Grades 3–5 92
Table A.2. Test Blueprint—Mathematics Grades 6–8 92
Table A.3. Test Blueprint—ELA Grades 3–5 93
Table A.4. Test Blueprint—ELA Grades 6–8 93
Table A.5. Test Blueprint—Science Grade 5 93
Table A.6. Test Blueprint—Science Grade 8 (MS) 94
Table A.7. Test Blueprint—Science Grade 11 (HS) 94
Table E.1. Student Participation N-Count Demographic Distribution—Mathematics 110
Table E.2. Student Participation N-Count Demographic Distribution—ELA 111
Table E.3. Student Participation N-Count Demographic Distribution—CSLA 112
Table E.4. Student Participation N-Count Demographic Distribution—Science 113
Table F.1. Scale Score Distribution—Mathematics Grade 3 114
Table F.2. Scale Score Distribution—Mathematics Grade 4 116
Table F.3. Scale Score Distribution—Mathematics Grade 5 118
Table F.4. Scale Score Distribution—Mathematics Grade 6 120
Table F.5. Scale Score Distribution—Mathematics Grade 7 122
Table F.6. Scale Score Distribution—Mathematics Grade 8 124
Table F.7. Scale Score Distribution—ELA Grade 3 126
Table F.8. Scale Score Distribution—ELA Grade 4 128
Table F.9. Scale Score Distribution—ELA Grade 5 130
Table F.10. Scale Score Distribution—ELA Grade 6 132
Table F.11. Scale Score Distribution—ELA Grade 7 134
Table F.12. Scale Score Distribution—ELA Grade 8 136
Table F.13. Scale Score Distribution—CSLA Grade 3 138
Table F.14. Scale Score Distribution—CSLA Grade 4 139
Table F.15. Scale Score Distribution—Science Grade 5 140
Table F.16. Scale Score Distribution—Science Grade 8 142
Table F.17. Scale Score Distribution—Science Grade 11 144
Table H.1. Performance by Subgroup—Mathematics Grade 3 155
Table H.2. Performance by Subgroup—Mathematics Grade 4 155
Table H.3. Performance by Subgroup—Mathematics Grade 5 156
Table H.4. Performance by Subgroup—Mathematics Grade 6 157
Table H.5. Performance by Subgroup—Mathematics Grade 7 157
Table H.6. Performance by Subgroup—Mathematics Grade 8 158
Table H.7. Performance by Subgroup—ELA Grade 3 159
Table H.8. Performance by Subgroup—ELA Grade 4 159
Table H.9. Performance by Subgroup—ELA Grade 5 160
Table H.10. Performance by Subgroup—ELA Grade 6 160

Table H.11. Performance by Subgroup—ELA Grade 7 161
Table H.12. Performance by Subgroup—ELA Grade 8 162
Table H.13. Performance by Subgroup—CSLA Grade 3 162
Table H.14. Performance by Subgroup—CSLA Grade 4 163
Table H.15. Performance by Subgroup—Science Grade 5 164
Table H.16. Performance by Subgroup—Science Grade 8 164
Table H.17. Performance by Subgroup—Science Grade 11 165
Table I.1. Points Earned Summary by Subclaim—Mathematics 166
Table I.2. Points Earned Summary by Subclaim—ELA 167
Table I.3. Points Earned Summary by Subclaim—CSLA 168
Table I.4. Points Earned Summary by Subclaim—Science 168
Table J.1. SR Item Classical Statistics—Mathematics Grade 3 169
Table J.2. CR Item Classical Statistics—Mathematics Grade 3 169
Table J.3. SR Item Classical Statistics—Mathematics Grade 4 170
Table J.4. CR Item Classical Statistics—Mathematics Grade 4 170
Table J.5. SR Item Classical Statistics—Mathematics Grade 5 171
Table J.6. CR Item Classical Statistics—Mathematics Grade 5 171
Table J.7. SR Item Classical Statistics—Mathematics Grade 6 172
Table J.8. CR Item Classical Statistics—Mathematics Grade 6 172
Table J.9. SR Item Classical Statistics—Mathematics Grade 7 173
Table J.10. CR Item Classical Statistics—Mathematics Grade 7 173
Table J.11. SR Item Classical Statistics—Mathematics Grade 8 174
Table J.12. CR Item Classical Statistics—Mathematics Grade 8 174
Table J.13. SR Item Classical Statistics—ELA Grade 3 175
Table J.14. CR Item Classical Statistics—ELA Grade 3 175
Table J.15. SR Item Classical Statistics—ELA Grade 4 176
Table J.16. CR Item Classical Statistics—ELA Grade 4 176
Table J.17. SR Item Classical Statistics—ELA Grade 5 177
Table J.18. CR Item Classical Statistics—ELA Grade 5 177
Table J.19. SR Item Classical Statistics—ELA Grade 6 178
Table J.20. CR Item Classical Statistics—ELA Grade 6 178
Table J.21. SR Item Classical Statistics—ELA Grade 7 179
Table J.22. CR Item Classical Statistics—ELA Grade 7 179
Table J.23. SR Item Classical Statistics—ELA Grade 8 180
Table J.24. CR Item Classical Statistics—ELA Grade 8 180
Table J.25. SR Item Classical Statistics—CSLA Grade 3 181
Table J.26. CR Item Classical Statistics—CSLA Grade 3 181
Table J.27. SR Item Classical Statistics—CSLA Grade 4 182
Table J.28. CR Item Classical Statistics—CSLA Grade 4 182
Table J.29. SR Item Classical Statistics—Science Grade 5 183
Table J.30. CR Item Classical Statistics—Science Grade 5 183
Table J.31. SR Item Classical Statistics—Science Grade 8 184
Table J.32. CR Item Classical Statistics—Science Grade 8 185
Table J.33. SR Item Classical Statistics—Science Grade 11 185
Table J.34. CR Item Classical Statistics—Science Grade 11 186
Table M.1. Operational Item Parameter Estimates—ELA Grade 3 199
Table M.2. Operational Item Parameter Estimates—ELA Grade 4 200

Table M.3. Operational Item Parameter Estimates—ELA Grade 5 201
Table M.4. Operational Item Parameter Estimates—ELA Grade 6 202
Table M.5. Operational Item Parameter Estimates—ELA Grade 7 203
Table M.6. Operational Item Parameter Estimates—ELA Grade 8 204
Table M.7. Operational Item Parameter Estimates—CSLA Grade 3 205
Table M.8. Operational Item Parameter Estimates—CSLA Grade 4 206
Table M.9. Operational Item Parameter Estimates—Science Grade 5 207
Table M.10. Operational Item Parameter Estimates—Science Grade 8 208
Table M.11. Operational Item Parameter Estimates—Science Grade 11 209
Table O.1. Operational Rater Agreement Statistics—Mathematics Grade 3 235
Table O.2. Operational Rater Agreement Statistics—Mathematics Grade 4 235
Table O.3. Operational Rater Agreement Statistics—Mathematics Grade 5 235
Table O.4. Operational Rater Agreement Statistics—Mathematics Grade 6 235
Table O.5. Operational Rater Agreement Statistics—Mathematics Grade 7 235
Table O.6. Operational Rater Agreement Statistics—Mathematics Grade 8 236
Table O.7. Field Test Rater Agreement Statistics—Mathematics Grade 3 236
Table O.8. Field Test Rater Agreement Statistics—Mathematics Grade 4 237
Table O.9. Field Test Rater Agreement Statistics—Mathematics Grade 5 238
Table O.10. Field Test Rater Agreement Statistics—Mathematics Grade 6 239
Table O.11. Field Test Rater Agreement Statistics—Mathematics Grade 7 240
Table O.12. Field Test Rater Agreement Statistics—Mathematics Grade 8 241
Table O.13. Operational Rater Agreement Statistics—ELA Grade 3 241
Table O.14. Operational Rater Agreement Statistics—ELA Grade 4 241
Table O.15. Operational Rater Agreement Statistics—ELA Grade 5 241
Table O.16. Operational Rater Agreement Statistics—ELA Grade 6 242
Table O.17. Operational Rater Agreement Statistics—ELA Grade 7 242
Table O.18. Operational Rater Agreement Statistics—ELA Grade 8 242
Table O.19. Field Test Rater Agreement Statistics—ELA Grade 3 242
Table O.20. Field Test Rater Agreement Statistics—ELA Grade 4 243
Table O.21. Field Test Rater Agreement Statistics—ELA Grade 5 243
Table O.22. Field Test Rater Agreement Statistics—ELA Grade 6 244
Table O.23. Field Test Rater Agreement Statistics—ELA Grade 7 244
Table O.24. Field Test Rater Agreement Statistics—ELA Grade 8 245
Table O.25. Operational Rater Agreement Statistics—CSLA Grade 3 245
Table O.26. Operational Rater Agreement Statistics—CSLA Grade 4 245
Table O.27. Field Test Rater Agreement Statistics—CSLA Grade 3 246
Table O.28. Field Test Rater Agreement Statistics—CSLA Grade 4 246
Table O.29. Operational Rater Agreement Statistics—Science Grade 5 246
Table O.30. Operational Rater Agreement Statistics—Science Grade 8 246
Table O.31. Operational Rater Agreement Statistics—Science Grade 11 247
Table O.32. Field Test Rater Agreement Statistics—Science Grade 5 247
Table O.33. Field Test Rater Agreement Statistics—Science Grade 8 248
Table O.34. Field Test Rater Agreement Statistics—Science Grade 11 248

List of Figures

Figure 2.1. How to Read the Colorado Academic Standards 18
Figure 10.1. Sample 2PL Item Characteristic Curve (ICC) 61
Figure 10.2. Sample 2PL Item Response Curves (IRCs) 61
Figure 10.3. Item Response Curves (IRCs) for a 2-point Polytomous Item 62
Figure 10.4. Sample Key Check (TRIAN) Report 63
Figure 10.5. Sample Adjudication Spreadsheet 63
Figure G.1. Scale Score Distribution—Mathematics Grade 3 146
Figure G.2. Scale Score Distribution—Mathematics Grade 4 146
Figure G.3. Scale Score Distribution—Mathematics Grade 5 147
Figure G.4. Scale Score Distribution—Mathematics Grade 6 147
Figure G.5. Scale Score Distribution—Mathematics Grade 7 148
Figure G.6. Scale Score Distribution—Mathematics Grade 8 148
Figure G.7. Scale Score Distribution—ELA Grade 3 149
Figure G.8. Scale Score Distribution—ELA Grade 4 149
Figure G.9. Scale Score Distribution—ELA Grade 5 150
Figure G.10. Scale Score Distribution—ELA Grade 6 150
Figure G.11. Scale Score Distribution—ELA Grade 7 151
Figure G.12. Scale Score Distribution—ELA Grade 8 151
Figure G.13. Scale Score Distribution—CSLA Grade 3 152
Figure G.14. Scale Score Distribution—CSLA Grade 4 152
Figure G.15. Scale Score Distribution—Science Grade 5 153
Figure G.16. Scale Score Distribution—Science Grade 8 153
Figure G.17. Scale Score Distribution—Science Grade 11 154
Figure K.1. Scree Plot—Mathematics Grade 3 187
Figure K.2. Scree Plot—Mathematics Grade 4 187
Figure K.3. Scree Plot—Mathematics Grade 5 188
Figure K.4. Scree Plot—Mathematics Grade 6 188
Figure K.5. Scree Plot—Mathematics Grade 7 189
Figure K.6. Scree Plot—Mathematics Grade 8 189
Figure K.7. Scree Plot—ELA Grade 3 190
Figure K.8. Scree Plot—ELA Grade 4 190
Figure K.9. Scree Plot—ELA Grade 5 191
Figure K.10. Scree Plot—ELA Grade 6 191
Figure K.11. Scree Plot—ELA Grade 7 192
Figure K.12. Scree Plot—ELA Grade 8 192
Figure K.13. Scree Plot—CSLA Grade 3 193
Figure K.14. Scree Plot—CSLA Grade 4 193
Figure K.15. Scree Plot—Science Grade 5 194
Figure K.16. Scree Plot—Science Grade 8 194
Figure K.17. Scree Plot—Science Grade 11 195
Figure L.1. 2022 Post-Equating Check TCCs—Mathematics Grade 3 196
Figure L.2. 2022 Post-Equating Check TCCs—Mathematics Grade 4 196
Figure L.3. 2022 Post-Equating Check TCCs—Mathematics Grade 5 197
Figure L.4. 2022 Post-Equating Check TCCs—Mathematics Grade 6 197
Figure L.5. 2022 Post-Equating Check TCCs—Mathematics Grade 7 198

Figure L.6. 2022 Post-Equating Check TCCs—Mathematics Grade 8 198
Figure N.1. Mathematics Grade 3 TCC 210
Figure N.2. Mathematics Grade 3 TIC 210
Figure N.3. Mathematics Grade 3 CSEM Curve 210
Figure N.4. Mathematics Grade 4 TCC 211
Figure N.5. Mathematics Grade 4 TIC 211
Figure N.6. Mathematics Grade 4 CSEM Curve 211
Figure N.7. Mathematics Grade 5 TCC 212
Figure N.8. Mathematics Grade 5 TIC 212
Figure N.9. Mathematics Grade 5 CSEM Curve 212
Figure N.10. Mathematics Grade 6 TCC 213
Figure N.11. Mathematics Grade 6 TIC 213
Figure N.12. Mathematics Grade 6 CSEM Curve 213
Figure N.13. Mathematics Grade 7 TCC 214
Figure N.14. Mathematics Grade 7 TIC 214
Figure N.15. Mathematics Grade 7 CSEM Curve 214
Figure N.16. Mathematics Grade 8 TCC 215
Figure N.17. Mathematics Grade 8 TIC 215
Figure N.18. Mathematics Grade 8 CSEM Curve 215
Figure N.19. ELA Grade 3 TCC 216
Figure N.20. ELA Grade 3 TIC 216
Figure N.21. ELA Grade 3 CSEM Curve 216
Figure N.22. ELA Grade 4 TCC 217
Figure N.23. ELA Grade 4 TIC 217
Figure N.24. ELA Grade 4 CSEM Curve 217
Figure N.25. ELA Grade 5 TCC 218
Figure N.26. ELA Grade 5 TIC 218
Figure N.27. ELA Grade 5 CSEM Curve 218
Figure N.28. ELA Grade 6 TCC 219
Figure N.29. ELA Grade 6 TIC 219
Figure N.30. ELA Grade 6 CSEM Curve 219
Figure N.31. ELA Grade 7 TCC 220
Figure N.32. ELA Grade 7 TIC 220
Figure N.33. ELA Grade 7 CSEM Curve 220
Figure N.34. ELA Grade 8 TCC 221
Figure N.35. ELA Grade 8 TIC 221
Figure N.36. ELA Grade 8 CSEM Curve 221
Figure N.37. CSLA Grade 3 TCC 222
Figure N.38. CSLA Grade 3 TIC 222
Figure N.39. CSLA Grade 3 CSEM Curve 222
Figure N.40. CSLA Grade 4 TCC 223
Figure N.41. CSLA Grade 4 TIC 223
Figure N.42. CSLA Grade 4 CSEM Curve 223
Figure N.43. ELA Reading Grade 3 TCC 224
Figure N.44. ELA Reading Grade 3 TIC 224
Figure N.45. ELA Reading Grade 3 CSEM Curve 224
Figure N.46. ELA Reading Grade 4 TCC 225

Figure N.47. ELA Reading Grade 4 TIC 225
Figure N.48. ELA Reading Grade 4 CSEM Curve 225
Figure N.49. ELA Reading Grade 5 TCC 226
Figure N.50. ELA Reading Grade 5 TIC 226
Figure N.51. ELA Reading Grade 5 CSEM Curve 226
Figure N.52. ELA Reading Grade 6 TCC 227
Figure N.53. ELA Reading Grade 6 TIC 227
Figure N.54. ELA Reading Grade 6 CSEM Curve 227
Figure N.55. ELA Reading Grade 7 TCC 228
Figure N.56. ELA Reading Grade 7 TIC 228
Figure N.57. ELA Reading Grade 7 CSEM Curve 228
Figure N.58. ELA Reading Grade 8 TCC 229
Figure N.59. ELA Reading Grade 8 TIC 229
Figure N.60. ELA Reading Grade 8 CSEM Curve 229
Figure N.61. CSLA Reading Grade 3 TCC 230
Figure N.62. CSLA Reading Grade 3 TIC 230
Figure N.63. CSLA Reading Grade 3 CSEM Curve 230
Figure N.64. CSLA Reading Grade 4 TCC 231
Figure N.65. CSLA Reading Grade 4 TIC 231
Figure N.66. CSLA Reading Grade 4 CSEM Curve 231
Figure N.67. Science Grade 5 TCC 232
Figure N.68. Science Grade 5 TIC 232
Figure N.69. Science Grade 5 CSEM 232
Figure N.70. Science Grade 8 TCC 233
Figure N.71. Science Grade 8 TIC 233
Figure N.72. Science Grade 8 CSEM 233
Figure N.73. Science Grade 11 TCC 234
Figure N.74. Science Grade 11 TIC 234
Figure N.75. Science Grade 11 CSEM 234

Chapter 1: Introduction

The purpose of this technical report is to inform users and other interested parties about the development, content, administration, and technical characteristics of the Spring 2023 Colorado Measures of Academic Success (CMAS) assessments for mathematics and English language arts (ELA) in Grades 3–8; for science in Grades 5, 8, and 11; and for the Colorado Spanish Language Arts (CSLA) assessment in Grades 3 and 4. The report includes an overview and summary of the components of the program, including information regarding the planning and administration of the assessments and details regarding item development, test construction, administration procedures, scoring, reporting, reliability, and validity, as well as a statistical summary of the Spring 2023 operational and field test items.

1.1. Testing Requirements

All public schools in Colorado are required by state law to administer a standards-based summative assessment each year in specified content areas and grade levels. Every student, regardless of ability or language background, must be provided the opportunity to demonstrate their content knowledge through the state assessments. The CMAS assessments in mathematics, ELA, science, and social studies are Colorado's end-of-year standards-based assessments designed to measure students' achievement of the Colorado Academic Standards (CAS).

As a requirement of Colorado School Law C.R.S. §22-7-1006.3 (4) (II)(b), students with Spanish as their home language in Grades 3 and 4 who meet established eligibility criteria may take the CSLA forms of the ELA assessment. The CSLA forms serve as accommodated versions of the CMAS ELA assessments and are parallel and comparable to ELA in test design, scoring, and reporting.

Colorado legislation (C.R.S. §22-7-1006.3 (1) (d)) also requires that a paper-based version be available for all online assessments that may be used by local educational providers for their students. The comparable paper-based forms may also be administered to students with disabilities and multilingual learner (ML) students as appropriate. Multilingual learners for assessment purposes are students with a home language other than English who are designated as not English proficient or limited English proficient (NEP/LEP) by an English language proficiency assessment or screener.

In 2015, Colorado passed legislation (C.R.S. §22-7-1013 (8) (a-c)) that allows for parents/guardians to excuse their child(ren) from testing.

1.2. Intended Population

The CMAS assessments are intended to be taken by all students enrolled in public schools, except for some students with the most significant cognitive disabilities who take the Colorado Alternate (CoAlt) assessment as determined by the student's Individualized Education Plan (IEP) team. ML students in their first year in the U.S. are exempt from the ELA assessment. However, ML students in Grades 3 and 4 designated as NEP whose native language is Spanish and who have received language arts instruction in Spanish during the current school year are required to take the CSLA assessment. Students with disabilities and ML students may take the CMAS assessments with or without accommodations that do not change the construct of the assessment. Accommodations are determined based on classroom experience and educational team decisions.

1.3. CMAS Background

The CMAS Science assessments were first administered in 2013–2014, the CMAS Mathematics and ELA assessments were first administered in 2014–2015, and the CSLA assessments were first administered in 2015-2016. Colorado developed the CMAS Mathematics and ELA assessments in collaboration with the Partnership for Assessment of Readiness for College and Careers (PARCC) consortium, with Pearson taking over as the testing contractor for mathematics and ELA in 2017–2018.[1] Pearson has been the testing contractor for the CMAS Science and Social Studies assessments and the CSLA assessments since their inception.

In 2017, the Colorado State Board of Education provided direction to the Colorado Department of Education (CDE) to decrease testing time. CDE began exploring the use of abbreviated versions of the prior years' test blueprints with the goal of decreasing testing time while retaining comparability to the CMAS Mathematics and ELA/CSLA assessments previously administered in Colorado to maintain longitudinal trend data. Test forms based on the abbreviated blueprints were developed in Fall 2017 and administered beginning in Spring 2018.

In 2021, Colorado received a partial waiver of the federal assessment requirements from the U.S. Department of Education (USED) due to COVID-19 conditions in Colorado. The number of tests students were required to take was reduced, with alternating grades for mathematics and ELA. Students were required to take one test in either mathematics or ELA, depending on their grade, although parents/guardians could choose to have their children take both tests. With the exception of students with a parent/guardian excusal, students in Grades 4, 6, and 8 were required to take the mathematics assessments; students in Grades 3, 5, and 7 were required to take the ELA assessments; and students in Grade 8 also took the science assessment. The Grade 5 and high school science and Grades 4 and 7 social studies assessments were not administered.

In 2022, newly revised standards were implemented for mathematics, ELA, and science. In 2008, Colorado passed Senate Bill 212 (also known as CAP4K) that required the State Board of Education to adopt content standards that prepare students for the 21st century workforce and for active citizenship upon receiving a high school diploma. It also required a revision to the CAS by July 1, 2018, and every six years thereafter. As such, the 2009/2010 CAS were reviewed and revised, resulting in the 2020 CAS. While minimal changes were made to the mathematics and ELA standards, the science standards underwent a substantial update to keep up with the shift to the Next Generation Science Standards (NGSS; NGSS Lead States, 2013).

Full implementation of the new three-dimensional science standards took place in 2021–2022. The new CMAS Science test was administered to all tested students for the first time in Spring 2022, which made it possible to test enough new content to allow for a robust item bank and to obtain a sufficient sample of students to conduct field test analyses. Standard setting was conducted in Fall 2022 so that full results with scale scores and proficiency levels could be reported for the Spring 2023 administration. While the Spring 2022 CMAS Science assessment reported percentile ranks only, the Spring 2023 science assessment reported scale scores and performance levels.

[1] For information on the background of the consortium and the development and administration of the assessments, see prior years' technical reports at https://www.cde.state.co.us/assessment/cmas_coalt_techreport.

Regular testing procedures resumed in Spring 2022 for mathematics and ELA/CSLA. Social studies was not administered in Spring 2022 or Spring 2023.

1.4. Purpose of CMAS

CMAS assessments were designed to be used for a variety of purposes, including informing parents/guardians and educators about individual student achievement of the grade-level CAS and allowing comparisons to other students across the state. Results are intended to provide one measure of a student's academic progress relative to the CAS. Results should be taken into consideration alongside other achievement information available locally. Results are also used as a piece of information in the evaluation of educator, school, and district performance. State assessment data typically help inform the state's school and district accountability system, including assigning performance ratings to schools and districts. State assessment results are also typically a component of educator evaluation. CMAS is a source of data that

- may be used as a prompt for further investigation at the student, classroom, school, and district levels;
- supports districts/schools in reviewing and developing goals for the performance of their students, including subgroups;
- may indicate that a review of programs, curricula, materials, and/or scope and sequence may be appropriate; and
- may inform the evaluation of district/school approaches.

Assessment results also support a range of data-driven stakeholder conversations, activities, and decisions such as school selection, program evaluation, investigative research, and policy/legislation formation and review. For example, educators can use the test scores to plan for further instruction and curriculum development and to report progress to parents/guardians. The results can also be used as one factor in making administrative decisions about program effectiveness, teacher effectiveness, class grouping, and needs assessment. CMAS results can also be used for research purposes and for informing community and organization efforts.

1.5. Assessment Development Partners

Activities specific to the CMAS assessments were conducted collaboratively by CDE, the Colorado educator community, and Pearson, the assessment contractor. Input and advice were provided by the Colorado Technical Advisory Committee (TAC).

1.5.1. Colorado Department of Education

As the administrative arm of the State Board of Education, CDE is responsible for implementing state and federal education laws. CDE's Assessment Unit works closely with Colorado school districts, educators, community stakeholders, and test development partners to develop and administer the state assessments. CDE focuses on creating assessments that serve students, schools, districts, and the community while complying with state and federal legal requirements. CDE also works closely with Pearson on each facet of the assessment, with CDE serving as the ultimate approver of services and products provided.

1.5.2. Colorado Educator Community

Educator participation in the CMAS development process is critical to ensuring that the assessments are aligned to the CAS, are appropriate for Colorado students at the assessed grade level, and are free from potential bias and sensitivity issues. Throughout the test development process, educators participate in the following development activities, as shown in Table 1.1:

- Item writing: After receiving item writing assignments based on the CAS, educators create assessment items. Items that successfully move through the entire item development process will eventually appear on the operational assessments.
- Content and bias review: Educators review items to ensure content alignment and identify potential bias and sensitivity concerns before items are field tested.
- Rangefinding: Educators review student responses to field tested constructed-response items and define the score point ranges for the scoring rubrics that are used to score student responses.
- Data review: Before field tested items are included on operational assessments, educators review items with statistical parameters outside of normal ranges to determine if the item is acceptable for inclusion in the operational item bank.
- Standard setting: Colorado educators participated in the CMAS Science standard setting in September 2022 to recommend cut scores for the new three-dimensional science assessment aligned to the 2020 CAS.

Table 1.1. Schedule of Major Events

Event	Date(s)
ELA Passage Reviews	January 5–11, 2022
	January 14–21, 2022
	February 3–9, 2022
	February 21–25, 2022
	March 8–14, 2022
IWW Training (ELA)	February 16, 2022
IWW Trainings (Math)	February 17–18, 2022
	March 9, 2022
IWW Training (Science)	March 21–24, 2022
Sim Storyboard Review (Science)	March 1, 2022
Content and Bias Review (Science)	Aug 2–5, 2022
Content and Bias Review (Math)	July 19–20, 2022
Content and Bias Review (ELA)	July 25–29, 2022
CMAS Science Standard Setting	September 27–28, 2022
Braille Review (Math & Science)	December 8–9, 2022
DAC Administration Training	November 9–15, 2022
Spring 2023 Administration Window	April 10–28, 2023
Rangefinding	June 5–8, 2023
Data Review (ELA)	August 17–18, 2023
Data Review (Science)	August 4, 2023
Data Review (CSLA)	September 13, 2023
Data Review (Math)	September 22, 2023
CMAS Reports Available	July 10, 2023

Note. The ELA passage review included five windows where educators reviewed batches of passages independently.

1.5.3. Pearson

As the primary contractor responsible for the end-to-end assessment cycle services and products, Pearson works closely with CDE throughout the CMAS and CoAlt Science assessment development and administration processes. This includes item and test development, forms creation, enrollment, packaging and distribution, test delivery, scoring, customer service, standard setting, scoring, score reporting, and psychometric services.

1.5.4. Tri-Lin Integrated Services, Inc.

As a subcontractor to Pearson, Tri-Lin is responsible for CSLA content and test development, including passage development, item development, and test form construction.

1.5.5. Colorado Technical Advisory Committee

The Colorado TAC is comprised of psychometric, assessment, and special populations experts tasked with providing high-level consulting and expert advice regarding validity and reliability issues. Topics for which the TAC has provided input include the blueprint design, scaling and equating, mode comparability, scoring, reporting, alignment study feedback, peer review, and standard setting. The TAC included the following members during the 2023 assessment cycle:

- Dr. Elliot Asp, Senior Partner, The Colorado Education Initiative
- Dr. Jonathan Dings, Executive Director of Student Assessment and Program Evaluation, Boulder Valley School District
- Dr. Michael Kolen, Psychometric Consultant
- Dr. Suzanne Lane, Professor, University of Pittsburgh
- Dr. Martha Thurlow, Director, National Center on Educational Outcomes

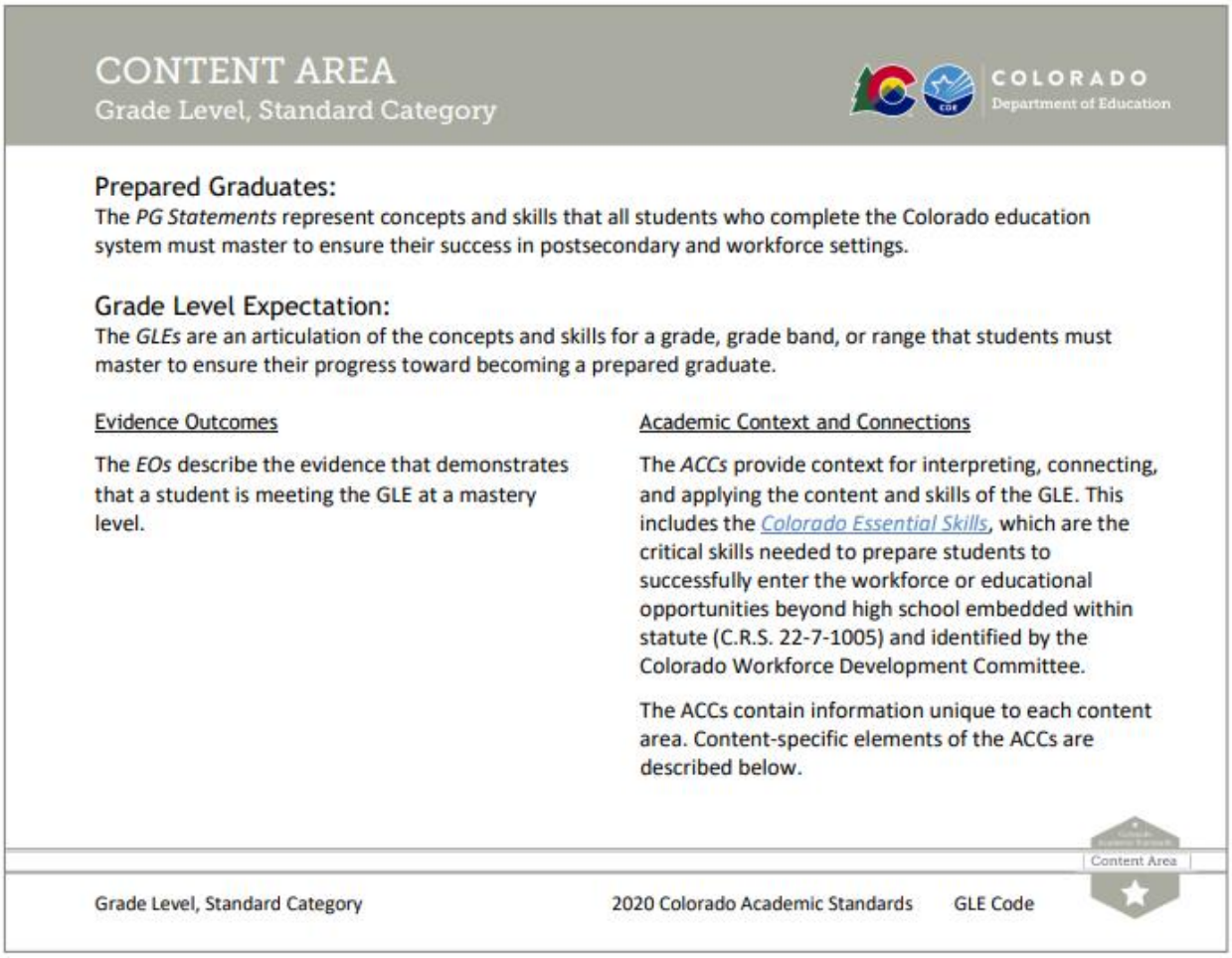
Chapter 2: Test Design

2.1. Colorado Academic Standards

The CMAS assessments are standards-based tests designed to measure what students should know and be able to demonstrate at the end of each grade or grade band based on the 2020 CAS located at the following links for each content area. The CAS for all content areas include the components in Figure 2.1.

- 2020 Mathematics Standards: http://www.cde.state.co.us/comath/statestandards
- 2020 Reading, Writing, and Communicating Standards: http://www.cde.state.co.us/coreadingwriting/statestandards
- 2020 Science Standards: https://www.cde.state.co.us/coscience/statestandards

Figure 2.1. How to Read the Colorado Academic Standards



The 2020 CAS for Mathematics and ELA had minimal changes compared to the previous 2009/2010 standards, whereas the 2020 CAS for Science underwent significant changes to be based on the NGSS.[2] The NGSS were guided by *A Framework for K–12 Science Education* (National Research Council, 2012) and designed to reflect more recent research and thinking in science education. The 2020 CAS for Science represent what all Colorado students should know and be able to do in science based on their PreK–Grade 12 science education.

The new science content standards are considered three-dimensional in that they incorporate Disciplinary Core Ideas (DCIs), Science and Engineering Practices (SEPs), and Crosscutting Concepts (CCCs). The DCIs encompass the content that occurs at each grade and provides the background knowledge for students to develop sense-making around phenomena in the three standards of Physical Science, Life Science, and Earth and Space Science. The DCIs are as follows[3]:

- Physical Science: Students know and understand common properties, forms, and changes in matter and energy.
    - PS1: Matter and its interactions
    - PS2: Motion and stability: Forces and interactions
    - PS3: Energy
    - PS4: Waves and their applications in technologies for information transfer

- Life Science: Students know and understand the characteristics and structure of living things, the processes of life, and how living things interact with each other and their environment.
    - LS1: From molecules to organisms: Structures and processes
    - LS2: Ecosystems: Interactions, energy, and dynamics
    - LS3: Heredity: Inheritance and variation of traits
    - LS4: Biological evolution: Unity and diversity

- Earth and Space Science: Students know and understand the processes and interactions of Earth's systems and the structure and dynamics of Earth and other objects in space.
    - ESS1: Earth’s place in the universe
    - ESS2: Earth’s systems
    - ESS3: Earth and human activity

The SEPs describe how scientists investigate and build models and theories of the natural world or how engineers design and build systems. They reflect science and engineering as they are practiced and experienced. There are eight SEPs:

1. Asking questions (for science) and defining problems (for engineering)
2. Developing and using models
3. Planning and carrying out investigations

[2]A summary of all the changes made to the standards are available on the CDE website for mathematics at https://www.cde.state.co.us/comath/2020cas-ma-changes, for ELA at https://www.cde.state.co.us/coreadingwriting/2020cas-rw-changes, and for science at https://www.cde.state.co.us/coscience/2020cas-sc-changes.

[3]Adaptation of the NGSS occurred by not adopting the fourth standard of Engineering, Technology, and Applications of Science (although engineering is still incorporated within the SEPs).

4. Analyzing and interpreting data
5. Using mathematics and computational thinking
6. Constructing explanations (for science) and designing solutions (for engineering)
7. Engaging in argument from evidence
8. Obtaining, evaluating, and communicating information

CCCs cross boundaries between science disciplines and provide an organizational framework to connect knowledge from various disciplines into a coherent and scientifically based view of the world. They build bridges between science and other disciplines and connect the DCIs and SEPs throughout the fields of science and engineering. There are seven CCCs:

1. Patterns
2. Cause and Effect
3. Scale, Proportion, and Quantity
4. Systems and System Models
5. Energy and Matter
6. Structure and Function
7. Stability and Change

The CMAS Science assessment is given in Grades 5, 8, and 11. Consistent with the standards, the Grade 5 assessment assesses the grade-level standards. Because the science standards are articulated by grade band at the middle school and high school levels rather than grade levels, the Grade 8 assessment assesses all middle school science standards, and the Grade 11 assessment assesses all high school science standards.

2.2. Test Frameworks and Blueprints

Concepts and skills identified in the CAS are the basis for the CMAS assessments. The CMAS frameworks list the percent representation and number of score points for each subclaim and standard area that appear on the assessments and specify the Evidence Outcomes (EOs) from the CAS that are included on the assessments. The mathematics and ELA frameworks continue to use Evidence Statements (ES) developed in collaboration with PARCC that describe the knowledge and skills an assessment item/task elicits from students. Together, the CMAS frameworks and ES provide the foundation for ensuring that the full range and depth of the standards are assessed. CDE incorporated feedback from content experts and educators throughout the state to create the final versions of the frameworks. The frameworks and ES are both available on the CDE website at https://www.cde.state.co.us/assessment/cmas_testdesign.

The test blueprints take the frameworks a step further by specifying the number of test items by Prepared Graduate (PG) Statement, Grade-Level Expectation (GLE), EO, item type, and cognitive complexity. The specificity of the test blueprints ensures that the assessments cover the breadth of the content indicated by the CAS within the associated grade or grade band. Appendix A presents the high-level test blueprints that summarize the percentage of score points on each test for each claim and subclaim on each assessment as shown in the frameworks. The most recent versions of the mathematics and ELA blueprints were developed in 2017–2018, while new test blueprints were created for CMAS Science in 2021–2022.

2.2.1. Mathematics and ELA

In 2017, the State Board of Education provided direction to CDE to decrease testing time. CDE began exploring the use of abbreviated versions of the prior years' test blueprints with the goal of decreasing testing time while retaining comparability to the CMAS Mathematics and ELA assessments previously administered in Colorado to maintain longitudinal trend data. Therefore, with the intent to reduce testing time, the 2018 blueprints were a proportionate abbreviation of the 2017 forms. CDE and Pearson collaborated in designing the CMAS subject- and grade-specific blueprints for mathematics and ELA in 2017–2018. The blueprints were designed to measure the same constructs as, and provide content comparability to, the previous year's assessments. Eligible content continued to reflect the CAS and ES used in prior years.[4]

2.2.2. Science

Pearson worked with Achieve, a nonprofit education organization that leads the effort to help states make college and career readiness a priority for all students, during the initial development of the new science assessment. Achieve provided background on how other states were approaching the new three-dimensional science standards and assessments and advice on how to proceed with cognitive complexity, blueprints, and reporting. With guidance from Achieve, Pearson, CDE, and Colorado educators collaborated in designing the science blueprints in a workshop held from November 6–7, 2019, in Denver. An effort was made to involve educators who were from areas representative of the entire state of Colorado (in terms of geographic location, gender, and race) and familiar with the 2020 CAS, related three-dimensional science instruction, and the assessment interaction and demonstration of achievement of the CAS of different groups of students, including students with disabilities and ML students. The blueprints were reviewed on October 14, 2021, by the TAC.

Results from the Spring 2022 test administration showed that items at the end of the test units were often left unanswered on the Grade 11 CMAS Science assessment, indicating that high school students had difficulty finishing the test in the time given. Therefore, after review by CDE and Pearson psychometrics, a proportional reduction was made to the Grade 11 blueprint to prevent speededness. The proposed blueprint reduction and test timing data was reviewed by the TAC on August 30, 2022. Grades 5 and 8 did not show a similar concern with speededness, so the length of these tests was not changed.

2.3. Claims and Subclaims

Student performance on the CMAS assessments is reported at the overall content area level as a scale score and performance level. Their performance is broken down even further at the claim and subclaim levels. The mathematics subclaims provide information on a student's achievement on grade-level math skills and concepts, as well as reasoning and modeling based on both grade-level and securely held knowledge of the skills and concepts from the previous grade level. The Reading and Writing claims for ELA provide information on a student's achievement in reading and comprehending a range of sufficiently complex texts independently. The subclaims are intended to provide more granular information about student demonstration of the knowledge and skills within the content area as reflected in the CAS.

[4] For more information about the transition and abbreviated assessments, see the 2017–2018 CMAS Mathematics and ELA technical report on the CDE website at https://www.cde.state.co.us/assessment/cmas_coalt_techreport.

Table 2.1 presents the content reflected in each subclaim by content area. The mathematics score is a composite of the four subclaims (Major Content, Supporting Content, Mathematical Reasoning, and Modeling and Application). The Reading score is a composite of the three reading subclaims (Reading: Literary Text, Reading: Informational Text, and Reading: Vocabulary, and the Written Expression subclaim that measures reading), and the Writing claim is a composite of the two writing subclaims (Writing: Written Expression and Writing: Knowledge and Use of Language Conventions). The science score is a composite of the three standards (Physical, Life, and Earth and Space Science), as well as an SEP score.

Table 2.1. Subclaims

Content Area	Subclaim	Description
Mathematics	Subclaim A: Major Content	Students solve problems involving the Major Content of the grade level with connections to the Standards for Mathematical Practice.
	Subclaim B: Additional & Supporting Content	Students solve problems involving the Additional and Supporting Content of the grade level with connections to the Standards for Mathematical Practice.
	Subclaim C: Expressing Mathematical Reasoning	In connection with content, the student expresses grade/course-level appropriate mathematical reasoning by constructing viable arguments, critiquing the reasoning of others and/or attending to precision when making mathematical statements.
	Subclaim D: Modeling & Application	In connection with content, the student solves real-world problems with a degree of difficulty appropriate to the grade/course by applying knowledge and skills articulated in the standards for the current grade/course (or for more complex problems, knowledge and skills articulated in the standards for previous grades/courses), engaging particularly in the Modeling practice, and where helpful making sense of problems and persevering to solve them, reasoning abstractly and quantitatively, using appropriate tools strategically, looking for the making use of structure, and/or looking for and expressing regularity in repeated reasoning.
ELA	Reading: Literary Text	Students read and analyze fiction, drama, and poetry.
	Reading: Informational Text	Students read and analyze nonfiction, history, science, and the arts.
	Reading: Vocabulary	Students use context to determine what words and phrases mean.
	Writing: Written Expression	Students compose well-developed writing using details from what they have read.
	Writing: Knowledge and Use of Language Conventions	Students demonstrate knowledge of conventions and other important elements of language.
Science	Physical Science	Students know and understand common properties, forms, and changes in matter and energy.
	Life Science	Students know and understand the characteristics and structure of living things, the processes of life, and how living things interact with each other and their environment.
	Earth and Space Science	Students know and understand the processes and interactions of Earth's systems and the structure and dynamics of Earth and other objects in space.
	Science and Engineering Practices (SEPs)	The SEPs describe how scientists investigate and build models and theories of the natural world or how engineers design and build systems. They reflect science and engineering as they are practiced and experienced.

2.4. Performance Levels

Student performance on the CMAS Mathematics and ELA assessments is categorized into five performance levels (*Did Not Yet Meet Expectations, Partially Met Expectations, Approached Expectations, Met Expectations,* and *Exceeded Expectations*), whereas student performance on CMAS Science is categorized into four performance levels (*Partially Met Expectations, Approached Expectations, Met Expectations,* and *Exceeded Expectations*). The performance levels are based on the overall scale score, and cut scores divide the score scale for a grade and content area into the performance levels (see Chapter 7 for more information on the cut scores). Students in the *Met Expectations* and *Exceeded Expectations* levels are considered on track to being college and career ready.

The performance levels are accompanied by performance level descriptors (PLDs) that articulate what a student should know and be able to do in a particular performance level (e.g., the set of statements describing what it means for a Grade 8 student to reach *Met Expectations* in mathematics). The CMAS assessments use two types of PLDs: (1) policy PLDs (also known as policy claims) that provide a general idea of what is expected of a student at each level regardless of their grade level, as shown in Table 2.2 and Table 2.3, and (2) grade-level PLDs that provide detailed descriptions of performance levels by grade level and content area, available online at https://www.cde.state.co.us/assessment/cmas_plds and included on the Individual Student Performance Report and in the *CMAS and CoAlt Interpretive Guide to Assessment Reports*.

Table 2.2. Performance Levels and Policy Claims—Mathematics and ELA

Performance Level	*Did Not Yet Meet Expectations*	*Partially Met Expectations*	*Approached Expectations*	*Met Expectations*	*Exceeded Expectations*
Policy Claim	Students who do not yet meet academic expectations for the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They will *need extensive academic support* to engage successfully in further studies in this content area.	Students who demonstrate a limited command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They will *need additional academic support* to engage successfully in further studies in this content area.	Students who demonstrate a moderate command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They will *likely need additional academic support* to engage successfully in further studies in this content area.	Students who demonstrate a strong command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They are *academically prepared* to engage successfully in further studies in this content area.	Students who demonstrate a distinguished command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They are *academically well prepared* to engage successfully in further studies in this content area.
Scale Score	650–699	700–724	725–749	750–varies*	varies*–850

*Varies by grade and content area

Table 2.3. Performance Levels and Policy Claims—Science

Performance Level	*Partially Met Expectations*	*Approached Expectations*	*Met Expectations*	*Exceeded Expectations*
Policy Claim	Students who demonstrate a limited command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They will *need additional academic support* to engage successfully in further studies in this content area.	Students who demonstrate a moderate command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They will *likely need additional academic support* to engage successfully in further studies in this content area.	Students who demonstrate a strong command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They are *academically prepared* to engage successfully in further studies in this content area.	Students who demonstrate a distinguished command of the concepts, skills, and practices embodied by the Colorado Academic Standards assessed at their grade level. They are *academically well prepared* to engage successfully in further studies in this content area.
Scale Score	650–724	725–749	750–varies*	varies*–850

*Varies by grade

2.5. Cognitive Complexity

All mathematics and ELA items are tagged with a cognitive complexity level of high, moderate/medium, or low, as described in Table 2.4.

Table 2.4. Mathematics and ELA Cognitive Complexity Levels

Content Area	High	Moderate/Medium	Low
Mathematics	• Significant shift from previous content • Open ended, sophisticated reasoning, critiquing, modeling • Single/multi-part that requires more evidence from the student	• Moderate shift into new content • Moderately scaffolded, some choice in approach • Single/multi-part, multi-step, moderate reading load	• Low shift from previous content • Very scaffolded, rote, recall, recognize • Single part, one step with low reading load
ELA	Items require synthesis of ideas and details **across multiple texts or ideas** (can be single passage). For example, items may require students to construct the main idea or theme that is common across multiple texts, especially multiple texts that are not closely related in theme and/or genre.	Items require **analysis of ideas and details** across multiple sections **in a single text**. It requires more close analytic reading than low complexity items. For example, identifying the main idea or theme of a text may require inferring the main or theme or integrating ideas and details from several locations in the text.	Items require students to identify **a single idea or detail in a text** (e.g., identifying a term or phrase using context). It requires students to recall, observe, question, or represent facts or simple skills or abilities.

Science transitioned away from Depth of Knowledge (DOK) in 2021–2022 with the adoption of the new science standards. From Achieve:

> *As states and districts develop new assessment systems, they need support for developing assessments that balance the vision and integrity of multi-dimensional standards with ensuring that they are sensitive to varying levels of student performance. This... (requires a) ...new approach to capturing and communicating the complexity of summative assessment items and tasks designed for three dimensional standards that can be used to ensure that all learners can make their thinking and abilities visible without compromising the rigor and expectations of the standards* (Achieve, 2019, p. 1).

The CMAS Science assessment now uses a cognitive complexity framework that examines items via three criteria, as summarized in Table 2.5 and presented in Appendix B. Phenomenon in the stimulus material is examined separately for its own cognitive complexity.

Table 2.5. Science Cognitive Complexity Criteria

Criterion	Description
Item Dimensionality	Item alignment to one, two, or three dimensions: • Content of EO (Disciplinary Core Idea DCI) • Science and Engineering Practice (SEP) of EO • Cross Cutting Concept (CCC) of EO Items aligned to a single dimension only are **not** acceptable for CMAS Science.
Scaffolding/Support	The more guidance and structure the item provides the student, the lower the cognitive load required. The matrix categorizes scaffolding/support into three levels: heavy, moderate, and minimal. Heavy refers to a specific, step-by-step process is given, and the student merely needs to follow that process to supply the answer. Moderate and minimal provide increasing degrees of freedom to make choices on the part of the student and require an increasing degree of initiative to make those choices.
Sensemaking	Fundamental to the approach of three-dimensional standards is student use of the dimensions to make sense of scientific phenomena. Some degree of sensemaking is required for all CMAS Science items. A sensemaking situation is one in which students (1) are provided material without obvious ties/connections to content (e.g., language of the standard) and (2) use their knowledge of the standard to explain what they see in the material.

2.6. Test Design

CMAS Mathematics and Science contain selected-response (SR), technology-enhanced (TE), and constructed-response (CR) items. Mathematics also contains fill-in-the-blank (FIB) items. The CMAS ELA/CSLA assessments are passage-based with a combination of literary and informational passages and contain SR, TE, and prose constructed-response (PCR) items. Multiple passages may be used to respond to some items.

For the ELA PCRs, students receive a prompt, respond to reading items, and write an extended response. It is then scored on a multi-trait rubric, as provided in Appendix C and on the CDE website at https://www.cde.state.co.us/assessment/cmas_testdesign. The ELA PCRs include three task types: literary analysis, research simulation, and narrative writing. Because it is administered on paper, CSLA forms contain SR, paper-based TE, and PCR items. The CSLA paper-based TE items are developed to have similar item formats and scoring rules to the paper-based versions of TE items developed for CMAS ELA.

All mathematics items are aligned to both an ES and an EO. The ES are grouped into three types to ensure that the full range and depth of the standards are assessed:

- Type I items:
  - Assess a specific EO, a specific part of an EO, or multiple EOs
  - Subclaims A and B
  - 1- or 2-point items (Grades 3–8) and 4-point items (Grades 6–8)
  - SR, TE, and FIB items
  - Calculator (Grades 6–8) and non-calculator (Grades 3–8)

- Type II items (reasoning):
  - Assess a specific type of mathematical reasoning and a specific scope in the EOs to reason about
  - Subclaim C
  - 3- or 4-point items
  - SR, TE, FIB, and CR parts; all items have at least one CR part
  - Calculator (Grades 6–8) and non-calculator (Grades 3–5)

- Type III items (modeling):
  - Assess a specific type of mathematical modeling and a specific scope in the EOs to model about
  - Subclaim D
  - 3- or 6-point items
  - SR, TE, FIB, and CR parts; all items have at least one CR part
  - Calculator (Grades 6–8), non-calculator (Grades 3–5)

The CMAS Science assessment is divided into item sets that present phenomenon-based scenarios as either interactive science simulations or static stimuli, followed by associated standalone items or clusters of items related to the simulation or scenario. A phenomenon is an observable event that students can use the three dimensions (DCI, SEP, and CCC) to explain or make sense of. Separate standalone items are also included that are not associated with a stimulus to target a small number of CAS not represented in the scenarios. The items are either 1-point SR, 1-point TE, or 2-point CR item types.

2.7. Timing of Tests

Each assessment was composed of three units with field test items embedded to allow the assessments to be administered in a reasonable timeframe, as shown in Table 2.6.

Table 2.6. Testing Times

Grades	Mathematics	ELA	Science
3–5	Units 1–3: 65 minutes Total time: 195 minutes	Units 1–3: 90 minutes Total time: 270 minutes	Units 1–3: 80 minutes Total time: 240 minutes
6–8	Units 1–3: 65 minutes Total time: 195 minutes	Units 1–3: 110 minutes Total time: 330 minutes	Units 1–3: 80 minutes Total time: 240 minutes
High School	N/A	N/A	Units 1–3: 50 minutes Total time: 150 minutes

Chapter 3: Item Development

The CMAS item development process results in a diverse bank of items that align to the CAS. All items are developed with the intention of being administered on multiple testing platforms, including online, online-accommodated, and paper-based assessments. The item writing process is a tiered, inter-related process that began with the development of the test blueprint for each grade level within each content area, followed by creating the item development plan (IDP) used to forecast the targeted number of items and associated stimuli across ESs or EOs needed to create a robust item bank. Once written, all newly developed items go through multiple rounds of review, including contractor, CDE, and Colorado educator content, bias, and data reviews.

As part of the test construction process, a selection of the proposed set of operational items are refreshed, as illustrated in Table 3.1. Therefore, a portion of the operational items have been used operationally on a previous CMAS form, while the remaining items are refreshed using Colorado-developed field test items. All items were reviewed by Colorado educators. (Please note that the Spring 2023 CMAS Grade 11 Science assessment included a set of core items held constant from 2022 to 2023 with a proportional reduction in length of the form.)

Table 3.1. Refresh Rates—Mathematics and ELA

Content Area	Item Type	Refresh Rate Minimum Targets	Refresh Rate Maximums
Mathematics	Type I: 1-point	25%	50%
	Type I: 2- and 4-point	40%	60%
	Type II	33%	67%
	Type III	50%	50%
ELA	Selected Response	50%	–
	Short Constructed Response	50%	–
	Extended Constructed Response	50%	–
	Passage Sets	50%	–
Science	Regular Cluster	33%	–
	Mini cluster	50%	–
	Overall items	33%	66%

3.1. Item Banking System

Pearson's proprietary software, ABBI (Assessment Banking and Building solutions for Interoperable assessments), is used to support the test development process from initial content authoring through the review cycles. ABBI is the authoritative source for all content, data, and functionality for all CMAS system components. It serves as the repository where the item bank is housed, item revisions are catalogued, and items and item metadata are uploaded and revised by assessment specialists. Items can be moved into various statuses, each representing a step in the item development process. The items and associated stimuli are tracked, and revisions are recorded from creation through retirement in a secure environment.

Custom development reports can be generated out of ABBI, which allows users to generate Excel reports that capture metadata (e.g., unique item number, ES, task type, cognitive complexity, associated stimulus, item status, item statistics, and comments) useful for analyzing the item bank. ABBI is the source of reference for how and when changes to the item and the metadata have been implemented.

3.2. Item Development Plan

An IDP for each content area and grade is created at the beginning of each item development cycle to determine the number of items, passages, and science cluster stimuli needed to construct the assessments based on the blueprint requirements, with development targets that address any task model, passage type, ES, EO, item/task type, and cognitive complexity shortages. To accomplish this, the item bank is analyzed, and the ES, EO, task type, and cognitive complexity gaps are identified so a variety of item types aligning to the ES, EOs, and the corresponding CAS can be created.

3.3. ELA Passage Development

Item development for ELA begins with the research and selection of high-quality literary and informational texts. Due to the availability of appropriate passages and challenges with acquiring permissions, passages to be used on the CSLA forms are commissioned by Tri-Lin, either in-house or by professional passage writers. The number and types of needed passages are determined by the test construction specifications, a gap analysis of the pool of available passages, and the IDP. The passage selection (and writing) guidelines, task model descriptions, and cognitive complexity framework define the number of texts/passages by text type, genre, length, and complexity. Contractor assessment specialists train passage searchers to find (or write for the CSLA items) relevant and rich texts that permit a range of content to be developed.

Passage searchers and writers submit the passages for the contractor assessment specialists to review and evaluate using approved criteria, including adherence to the cognitive demand, relevance, and purpose of the test and the appropriate use of graphics to improve text comprehension. Test passages are analyzed and rated for text complexity. The assessment specialists check the passages for clarity, correctness of language, appropriateness of language for the grade level, and adherence to the style guidelines.

Accepted passages are presented to CDE for review. Once the passages are accepted by CDE, committees of educators review them for content and bias. The committees are comprised of educators from throughout the state representing a variety of student populations, including students with disabilities and ML students. Passages accepted by both CDE and the educator committees are then used for item writing.

3.4. Science Scenario Development

Item development for science begins with the preliminary conception and composition of the interactive science simulations and cluster stimuli. The number and types of needed simulations and cluster stimuli are determined by the test construction specifications, a gap analysis of the pool of available SIMs and stimuli, and the IDP, and the topics are researched for suitability of science content, alignment to the standards, and grade-level appropriateness.

Simulations and cluster stimuli follow slightly different paths through the development process, but both include multiple steps of review and evaluation by assessment specialists using approved criteria such as adherence to cognitive complexity requirements, relevance to standards, purpose of the test, and the appropriate use of graphics and or animations. Pearson checks all stimulus text for scientific accuracy, clarity, correctness of language, appropriateness of language and science concepts for the grade level, and adherence to the style guidelines.

Simulation ideas are presented to CDE for review and feedback in the form of storyboards illustrating the intended virtual interaction, along with suggested EOs that the simulations address. CDE provides feedback on how to move forward with the development of the simulations. Revised storyboards are then reviewed by committees of educators from throughout the state representing a variety of student populations, including students with disabilities and EL/ML students. The simulations are then developed into animated interactions and reviewed by CDE, after which items are written to a variety of EOs, either internally or by educators.

Cluster stimuli are proposed as topics to CDE and then developed into drafts based on CDE feedback. Drafts are refined by Pearson with CDE input and presented to educators for review and item writing using the same criteria used for the simulations.

3.5. Item Writing

Item writer workshops (IWWs) with Colorado educators were conducted for the development of new items for the 2022–2023 cycle for mathematics, ELA, and science. IWW participants are educators from across the state representing a variety of student populations, including students with disabilities and students with limited English proficiency. CSLA item writers are proficient in written academic Spanish and begin developing CSLA items after receiving training. The educators are given item writing assignments and develop a variety of items across task types, ES, and EOs. The item writers work with Pearson and/or Tri-Lin assessment specialists when clarification is needed for CSLA items. Content specialists from CDE are also present to assist as needed. Item writers use the ESs and EOs; the CAS; secure item specification documents, including item-writing guidelines (universal design guidelines, bias and sensitivity guidelines, and editorial guidelines); and an item writing checklist to guide them in completing their assignments.

All item writers author the items in ABBI, where Pearson or Tri-Lin assessment specialists complete their initial review. The assessment specialists review and suggest revisions to the items and metadata for the item authors, who then make the revisions and resubmit the items within ABBI.

3.6. Item Review

3.6.1. Internal Review

Pearson and Tri-Lin assessment specialists evaluate each newly developed item for content correctness; grade appropriateness; and ES, EO, CAS, and cognitive complexity alignment, focusing on the quality of the items, adherence to the principles of universal design, cognitive demand, relevance to the purpose of the test, and appropriateness of graphics. Research librarians perform additional fact checking to ensure accuracy. Pearson and Tri-Lin copy editors check items for clarity, correctness of language, appropriateness of language for the grade level, adherence to style guidelines, and conformity with acceptable item-writing practices.

When appropriate, human-scored CR items are reviewed for their scorability by a Scoring Services director, and items and/or scoring rubrics with score points deemed “difficult to score” are revised in collaboration with the assessment specialist(s). Equation editor/CR items scored by an automated engine are reviewed for their scorability by a mathematics product analyst, and items and/or scoring rubrics are revised in collaboration with the assessment specialist(s).

Pearson and Tri-Lin assessment specialists also perform a universal design review to assess item accessibility irrespective of diversity of background, cultural tradition, and viewpoints; to evaluate changing roles and attitudes toward various groups; to review the role of language in setting and changing attitudes toward various groups; to appraise contributions of diverse groups (including ethnic and minority groups, individuals with disabilities, and women) to the history and culture of the United States and the achievements of individuals within these groups; and to edit for inappropriate language usage or stereotyping with regard to sex, race, culture, ethnicity, class, disability, or geographic region. The universal design review also includes reviewing items for potential bias to ensure that all items are fair and all students would have an equal opportunity to demonstrate achievement regardless of their gender, ethnic background, religion, socioeconomic status, disability, or geographic region. Items are also reviewed for visual bias, accessibility for students with disabilities, and convertibility to braille and text-to-speech.

Once the internal reviews are complete, each item's status is updated in ABBI and a lead assessment specialist conducts a final content review. Item statuses are updated in ABBI upon approval, and items are presented to CDE for review. Adhering to these processes ensures that each Colorado item measures the ES or EO and standard, is content- and grade-appropriate, is factually accurate, has appropriate answers and distractors, is accessible to all populations required to take the assessments, is free from any bias, and follows the Colorado style guidelines.

3.6.2. CDE Review

CDE reviews items in ABBI to ensure that the content is correct, the alignment is sound, the cognitive complexity is appropriate, the language and content are grade-appropriate, the graphics are clear and relevant to the item, and the content is free of bias/sensitivity issues. Once complete, CDE alerts Pearson or Tri-Lin assessment specialists. CDE's comments and determinations regarding the status of the items are recorded in ABBI, as indicated below:

- Items marked "Accept" need no more revisions and are ready for external Colorado educator content and bias reviews.
- Items marked "Accept with Edits" are revised per CDE's feedback and re-reviewed by the internal review team if necessary. These items are then reviewed by CDE again, reconciled with the assessment specialists, and deemed either "Accept" or "Reject."
- Items marked "Reject" are rejected and given a status of "Do Not Use" in ABBI. These items are either rewritten or replaced with items written by an assessment specialist. In either case, the items go through the same rigorous review process as newly developed items.

3.6.3. External Content and Bias Review

All items that pass the internal and CDE reviews are brought to external content and bias committees comprised of Colorado educators. The purposes of these educator reviews are to (1) ensure that the items are properly aligned to the CAS, accurately measure the intended content, and are grade-appropriate; and (2) identify any potential bias or stereotypes in the items. Separate committees are convened for each content area, as well as for the accommodated CSLA items. The meetings are conducted either in person or virtually and include group training on the expectations and processes of each meeting, followed by breakout groups by content area and grade where additional training is provided.

The Colorado educators are selected from across the state with diverse backgrounds and experience working with diverse learners (e.g., based on gender, race/ethnicity, income, and geography), standards and content expertise, and special population expertise (i.e., students with disabilities and EL/ML students). For science, educators are also selected based on their experience in the domain they are reviewing. For the accommodated CSLA items, an effort is made to involve educators who teach EL/ML students, are familiar with the instruction and needs of the students in an English language development program that uses native language instruction, and are proficient in written Spanish.

The committee members are trained and instructed to verify that each item and stimulus

- displays and functions correctly in TestNav 8 Preview (i.e., Pearson's online testing platform students use to access the assessment);
- aligns to the ES and/or EO;
- uses clear, unambiguous, and grade-appropriate language;
- avoids construct-irrelevant complex sentence structure;
- uses everyday words to convey meaning when vocabulary is not part of the tested construct;
- has one correct answer (depending on the item type);
- contains plausible distractors that represent feasible misunderstandings of the content (depending on the item type);
- represents the range of cognitive complexities and includes challenging items for students performing at all levels;
- is appropriate for students in the assigned grade in terms of reading level, vocabulary, interest, and experience;
- has scoring guidelines that capture exemplar responses at each score point for CR items;
- includes appropriate and clear graphics/art/photos that are relevant to the item and accessible to all testing populations;
- is free of ethnic, gender, political, and religious bias;
- avoids construct-irrelevant content that may unfairly advantage or disadvantage any student subgroup; and
- considers access issues at the time of item writing (e.g., determine how students with visual disabilities would access items with needed visuals/graphics/animation).

The committees make one of three recommendations on every item: "Accept," "Accept with Edits," or "Reject." Following the educator meetings, CDE, Pearson, and Tri-Lin review committee comments, reconcile proposed edits, and finalize item outcomes. ABBI is updated to reflect the edits and outcomes. The approved items, passages, and simulations/clusters are then made ready for inclusion on the spring operational forms as embedded field test items.

3.7. Data Review

After item development is complete, selected items are placed on the operational assessments in embedded field test positions. The goal of field testing is to allow for the evaluation of the quality of the items through a review of item performance data to determine their inclusion in the operational item pool. To accomplish this, psychometricians perform statistical analyses on the field tested items following their administration in a field test environment to evaluate their quality.

Classical statistics include item means (*p*-values), item-total correlations/point biserials, and distribution of responses across answer options or score points, depending on the item type. Differential item functioning (DIF) analyses are conducted on various subgroups (gender, ethnicity, free and reduced lunch, IEP, and MLs) using Mantel–Haenszel Delta DIF statistics (Dorans & Holland, 1992). The same analysis methods are used for CSLA items, but the DIF analyses are conducted by gender only due to the population of students taking the form. Classification rules derived from National Assessment of Educational Progress (NAEP) guidelines (Allen et al., 1999) were used to classify items as having either negligible, moderate, or significant DIF. Items are then flagged based on the criteria in Table 3.2, and flagged items are taken to a data review meeting where a committee of educators reviews the flagged items and their statistics along with student performance data.

Table 3.2 presents the statistical flags applied to the field tested items. Classical statistics include item means (*p*-values), item-total correlations/point biserials, and distribution of responses across answer options or score points, depending on the item type. Differential item functioning (DIF) analyses are conducted on various subgroups (gender, ethnicity, free and reduced lunch, IEP, and MLs) using Mantel–Haenszel Delta DIF statistics (Dorans & Holland, 1992). The same analysis methods are used for CSLA items, but the DIF analyses are conducted by gender only due to the population of students taking the form. Classification rules derived from National Assessment of Educational Progress (NAEP) guidelines (Allen et al., 1999) were used to classify items as having either negligible, moderate, or significant DIF. Items are then flagged based on the criteria in Table 3.2, and flagged items are taken to a data review meeting where a committee of educators reviews the flagged items and their statistics along with student performance data.

Table 3.2. Item Statistical Flagging Criteria

Statistic	Criterion	Possible Indication
P-value	< 0.1 or > 0.9	Very difficult or easy item
Item-total correlation	< 0.15	Poorly discriminating item
Distractor item-total correlation (SR only)	> 0.0	Possible miskey*
Score point percentage (multi-point items only)**	<1%, >50%, or >60%	Very few students or many 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 item-total correlation
**If a multi-point item has less than 1% for a score point or more than 50% 0s, the item is flagged. The rule is 50%+ 0s for mathematics, ELA, and CSLA and 60%+ for science.
***B DIF indicates moderate DIF, whereas C DIF indicates significant DIF.

Separate data review committees are convened for each content area, including the accommodated CSLA items. Participants are provided item images and metadata, along with the classical and DIF statistics. During the data review meetings, educators are trained to interpret the statistical information and judge the appropriateness of the flagged items. The committee members use the data as a tool to direct them toward potential flaws in an item and discuss whether there are construct-irrelevant reasons for a data flag. A data flag, by itself, is not the sole reason an item is rejected. Committee members are instructed that their final judgments about the appropriateness or fairness of an item for any individual and subgroup encompassed by the data flag should be based on their expertise with their content area and experience as Colorado educators.

Committee members review each item and recommend whether to accept or reject it. An accepted item indicates that the educators, through their varying expertise, determined that there is not a construct-irrelevant reason for the data flag within the item, whereas a rejected item indicates that the educators determined there is a construct-irrelevant reason for the data flag. Construct-irrelevant reasons for data flags could include issues such as language that is above grade-level or content that is biased against a particular group. In contrast, construct-relevant explanations could be difficult content that is part of the standards or distractors that reflect a very common misunderstanding of the concept covered by the item, which would not be a reason to reject the item.

Following the data review meetings, CDE reviews the committees' recommendations and makes final decisions. All accepted items are moved into "Ready for Operational" status. Table 3.3 presents the final results following the data review based on Spring 2023 data (i.e., the number of field tested items that were either accepted, accepted for revision and re-field test, or rejected as a result of the data review).

Table 3.3. Data Review Results

Content Area	Grade	#Accepted	#Accepted for Revision and Re-Field Test	#Rejected
Mathematics	3	17	2	0
	4	9	0	0
	5	15	2	0
	6	28	3	0
	7	24	3	0
	8	18	0	0
ELA	3	7	0	4
	4	29	0	1
	5	31	0	3
	6	28	0	1
	7	42	0	3
	8	47	0	5
CSLA	3	4	0	1
	4	2	0	3
Science	5	87	2	2
	8	87	1	2
	11	84	5	15

Chapter 4: Test Construction

The Spring 2023 ELA and science grades 5 and 8 operational test forms were newly developed test forms developed by Pearson, whereas the Spring 2023 mathematics operational test forms were either intended for use in the postponed 2019–2020 test administration or were newly developed test forms for use in the 2021–2022 test administration. The Spring 2023 CSLA newly constructed forms were constructed through an iterative process between Pearson and Tri-Lin. The Spring 2023 CMAS Science grade 11 test form was a subset of the 2022 core form based on the proportional blueprint reduction due to speededness. Items were selected as anchors only if they were found in the first part of a unit that showed any speededness concerns. Once the test forms were constructed, CDE reviewed the forms, provided feedback, and gave final approval.

4.1. Test Form Construction

Most students take the CMAS assessments online, which allows for the use of innovative item types and for accessibility features such as text-to-speech and color contrast to be available to all students in both English and Spanish for mathematics and science and in English for the online ELA forms. When building the test forms, assessment specialists select a set of operational items in accordance with the test blueprint and test construction specifications. Items selected for operational use must meet the blueprint requirements and should include a variety of topics and contexts with specified psychometric targets. The following guidelines were used during the Spring 2023 form construction:

- Adherence to the test blueprints and test construction specification targets
    - Exact match to blueprint for subclaims
    - Same distribution of cognitive complexity
    - Same percentage of TEs
- Review of the item statistics and adherence to the statistical criteria in the test construction specifications
    - Evaluation of item means, point biserial correlations, and score point distributions
    - Evaluation of item response theory (IRT) item parameter estimates
    - Evaluation of item fit statistics
    - Mirroring of 2018 test characteristic curves (TCCs) and conditional standard error of measurement (CSEM) curves (mathematics and ELA only[5])
    - Minimization of CSEM curves around the cut scores
- Balance in the representation of gender, ethnicity, geographic regions, and relevant demographic factors
- Thorough review of individual items to establish that the content within items is up-to-date and relevant
- Selection of items with various stimulus types throughout the test form to enhance the test-taking experience by providing variation in the appearance of item types presented
- Efficient and deliberate use of varied content representative of the knowledge and skills in the ESs or EOs
- Review of the full form, including field test items, for clueing and/or content overlap

[5] There was no mirroring of the CSEM curves for the Spring 2023 science assessments because the 2023 curves serve as the base to which future forms will need to mirror.

After the initial operational item pull is complete, assessment specialists verify that the test forms meet the blueprint and test construction specifications (i.e., the required ES or EO coverage, claim and subclaim coverage, cognitive complexity allocation, and task type). The form is then presented to a Pearson psychometrician who verifies that the form falls within the established psychometric and blueprint parameters and identifies the anchor item set within each operational form. (See Chapter 9 for details about the anchor sets.) Once the form is vetted internally, the form is presented to CDE for review. If needed, the assessment specialists, Pearson psychometricians, and CDE collaborate to finalize the form. This can be an iterative process, with the result being CDE's approval of the form.

After the operational form is approved, field test items are selected from the item bank. Items chosen for field testing are placed on a form in a designated section and sequence. Pearson and Tri-Lin assessment specialists assemble field test sets of items so that they comprise the appropriate distribution of standards, subclaims, task types, topic coverage, cognitive levels, and key distributions to meet the required item refresh rates in following years.

4.2. Accommodated Test Forms

Accommodated test forms are available for students who need them and include paper, large print, and braille forms, as well as auditory/signed presentation scripts and online forms designed to work with assistive technology such as screen readers. Auditory/signed presentation scripts are available for the paper forms in both English and Spanish for mathematics and science. English auditory/signed presentation scripts are available for both online and paper forms for local translation into languages other than Spanish including sign language. Due to the effort involved in creating an approved accommodated form, these forms are not refreshed at the same rate as the online forms.

4.2.1. Paper

Paper-based versions of the CMAS assessments are available as an accommodation or for schools that choose not to test online as allowed by state law. CSLA is the accommodated version of CMAS ELA for eligible Spanish-speaking students in Grades 3 and 4 and is administered on paper. A Spanish transadaptation is also available on paper for CMAS Mathematics and Science.

The paper form is parallel to the online form, meaning the paper and online forms include the same operational items. To support this, parallel paper-based items were developed for TE items in a way that was comparable in terms of student interaction. In some cases, this was achieved with traditional SR items, and in others it required an item that had to be human-scored. For example, a drag-and-drop TE item may have been converted to an item in which the student had to draw lines from the draggers to the drop bays. During equating, the TE item statistics are compared to the paper-based version to confirm equivalence. CSLA also has paper-based versions of TE items that were developed to be similar to the ELA paper-based versions of TE items, although the CSLA paper-based TE items were all machine scored.

The Spring 2023 operational items on the ELA and science paper-based forms were the same as the operational items on the online forms. The Spanish version of the science paper form was the same as the operational items on the English paper form. The mathematics paper-based forms in grades 3, 5 and 7 were the same as the 2020 administration, and the mathematics paper-based forms in grades 4, 6 and 8 were the same as the 2022 administration.

4.2.2. Braille

After approval of the paper test materials, a braille version of the assessments is created according to the process outlined below:

1. Pearson Braille Services uses constructed test forms to review the items and clusters for identifying potential modifications related to spacing constraints, visual bias in response expectations, and illustration complexity. Recommendations are documented for modifications to text and images.
2. The modifications document is provided to Pearson assessment specialists to ensure compliance with item constructs and assessed standards.
3. Pearson assessment specialists and CDE review the recommendations and provide feedback regarding any modification concerns.
4. Pearson Braille Services translates the test form into braille and designs print images as tactile graphics.
5. The braille form is proofread by a two-person proof team consisting of a native braille reader, certified as a braille proofreader by the National Library Service, and a sighted copyholder.
6. Edits to text and graphics are made based on the proof team's feedback.
7. The braille form is reviewed by a committee of Pearson staff, CDE staff, and Colorado Teachers of the Visually Impaired (TVI).
8. The braille form is finalized, and hardcopy test books are produced.

The Spring 2023 mathematics assessments in grades 3, 5, and 7 used the same braille form intended for use in the 2020 administration, whereas the Spring 2023 mathematics assessments in grades 4, 6, and 8 used the same braille form intended for use in the 2022 administration.

4.2.3. Large Print

Large print versions of the CMAS assessments are a 50% enlargement of the regular paper form and are printed on 14" × 18" paper. When needed, the large print version includes a visual description booklet that contains a description of artwork (maps, photographs) for which it may be difficult for a student with visual impairments to see the subtleties within the art. CDE reviews the paper form and identifies which pieces of art need to be described in the visual description test booklet.

Chapter 5: Test Administration

The CMAS assessments are administered in TestNav, Pearson's online testing platform. PearsonAccess^{next} is the student test management portal Assessment Coordinators and Test Administrators use to manage student tests and registrations and order materials if needed. Prior to the administration of the assessments, districts, schools, and teachers are to ensure that their students and systems are prepared for the assessments. Such information is communicated to the appropriate individuals via manuals, virtual trainings, and recorded modules.

5.1. Manuals

The following manuals are available online at https://coassessments.com/manuals/ to support the CMAS administration:

- The *CMAS Test Administrator Manual* for both online and paper-based testing describes the procedures Test Administrators are to follow when administering the assessments. Test administration policies and procedures are to be followed as written so all testing conditions are uniform statewide. The guidelines and test administration scripts in these manuals are provided to ensure that every student in Colorado receives the same standard directions during the test administration by content area, grade level, and accommodation.
- The *CMAS and CoAlt Procedures Manual* provides instructions for coordination of the CMAS assessments. Instructions include the protocols all school staff are to follow related to test security, test administration, and providing accommodations to students with disabilities and ML students and accessibility features to all students. The manual also includes the tasks to be completed by District Assessment Coordinators (DACs), School Assessment Coordinators (SACs), and District Technology Coordinators (DTCs) before, during, and after the test administration.
- The *PearsonAccess^{next} Online User Guide* provides guidance for DACs, SACs, DTCs, Test Administrators, and student enrollment/sensitive data personnel who use PearsonAccess^{next}.

5.2. Administration Training

Administration training is intended to make sure all individuals involved in CMAS assessment activities at the school and district levels are prepared to follow administration processes and procedures with fidelity, as well as to support adherence to security procedures. Fidelity to standardized test administration processes and procedures helps ensure the comparability of resulting scores and accurate interpretation of results.

Live virtual trainings were conducted by CDE for groups of DACs, during which the DACs independently accessed CDE- and Pearson-developed lessons through an interactive training platform. The lessons contained information regarding proper procedures for administration, security requirements, receiving and returning materials to Pearson, and the use of PearsonAccess^{next} with TestNav. Upon completion of each training lesson, CDE provided additional details pertaining to the covered information and an opportunity for questions and answers. After CDE trained the DACs, the DACs trained the SACs, Test Administrators, and any other individuals within the district who planned to participate in the CMAS administration.

Pearson customer service center staff were also trained to answer questions about the administration and to escalate inquiries as necessary. A knowledge base of common questions was created by CDE and Pearson based on information covered in the training materials and manuals to ensure accurate and consistent responses to school and district personnel, with revisions and additions made as needed. CDE met with Pearson daily during the administration window to review questions from districts and ensure that appropriate answers were provided. Policy questions received by the Pearson customer service center were referred to CDE.

Live webinar accommodations and accessibility features training was also conducted by CDE for district-level personnel to ensure that all individuals providing these supports across the state follow the procedures associated with each accommodation and accessibility feature. Providing accessibility features and accommodations in a standardized manner helps to ensure the comparability of resulting scores and accurate interpretation of results.

Resources used during the live trainings are posted on the CDE website at http://www.cde.state.co.us/assessment/trainings-archive. Administration training materials such as slide decks, manuals, and how-to guides were also available on the CDE Assessment Unit website for training SACs and Test Administrators.

5.3. Practice Resources

Colorado Practice Resources (CPRs) are available online at https://coassessments.com/practice-resources/ to help students become familiar with the CMAS item types. The CPRs are updated as needed to reflect current accessibility features and any updates to TestNav that may impact student interactions with the assessment. Accommodated versions of the CPRs are also available so students can practice using accommodations and accessibility features such as English text-to-speech, color contrast, and Spanish text-to-speech. Paper sample items for students taking the paper versions of the assessments are available in PDF format for download. CPRs are accompanied by scoring guides that include performance metrics and alignment to the CAS.

5.4. Onsite Preparation

Districts were instructed in site readiness preparations, TestNav, proctor caching, and use of the SystemCheck tool to configure their testing technology environments and evaluate their configuration for district readiness. Districts were also provided tools and resources to test their environment readiness status and infrastructure systems.

5.5. Accessibility Features and Accommodations

Accessibility is considered from the beginning of the test development process and is inherent within the CMAS assessment and administration. For example, TestNav includes tools and accessibility features that are available to all students to increase the accessibility of the assessments (e.g., highlighter, online color contrast). Also included is the text-to-speech accessibility feature for mathematics and science that allows for text to be read to students by the embedded software audio feature. Although this feature is available to all students, only students who need text-to-speech are assigned to it in advance of testing. Similarly, the CSLA assessments were developed to be linguistically accommodated Spanish tests and, as such, are designed to be linguistically accessible for eligible Spanish-speaking students.

Accommodations are also available to the population of students with IEP or 504 plans or ML students. For example, students may have extended time as required by their IEP or as allowed for students classified as ML. The test is also available with Spanish text-to-speech (mathematics and science only) and paper transadaptations or auditory presentation scripts that can be translated into other languages. Accommodations are intended to provide a student with an opportunity to access the assessment without impacting the measured construct. Accommodations can be adjustments to the test presentation, materials, environment, or response mode of the student and are based on individual student need. They should not provide an unfair advantage to any student. Providing an accommodation for the sole purpose of increasing test scores is not ethical.

Accommodations must be documented and used regularly during classroom instruction and assessments prior to the testing window to ensure that the student can successfully use the accommodation. However, although accommodations are used for classroom instruction and assessments, some may not be appropriate for use on statewide assessments. As a result, it is important that educators become familiar with the state assessment policies about the appropriate use of accommodations and that districts have a plan in place to ensure and monitor the appropriate use of accommodations.

Certain accommodations are allowed only in special cases with CDE approval due to being an inherent violation of the intended construct. For example, the accommodations of calculator on non-calculator sections of mathematics and a scribe for CR items for ELA/CSLA require approval to preserve the intended constructs of mathematics and writing according to the CAS.

Some of the available accommodations for CMAS include CSLA in place of ELA (other linguistic accommodations do not apply as CSLA is the linguistic accommodation), English auditory/signed presentation scripts (mathematics and science), Spanish auditory/signed presentation scripts (mathematics and science), auditory/signed presentation scripts for signed presentation and local translation into languages other than English and Spanish, braille forms, large print forms, assistive technology forms for screen readers (mathematics and ELA only), and Spanish forms with and without text-to-speech for mathematics and science.

5.6. Test Security

Test security procedures are put in place to enhance the likelihood that security is maintained before, during, and after the assessment administration. For example, materials used during the paper administration of the assessment are to be kept in locked storage locations when not under the direct supervision of Pearson or approved testing coordinators and administrators. All district and school personnel involved in the CMAS test administration are required to participate in annual local training. DACs are responsible for overseeing training for the district, including verifying that the DTC and SACs are trained. SACs are responsible for ensuring that Test Administrators and all other individuals involved in test administration at the school level are trained and subsequently act in accordance with all security requirements.

A chain of custody plan for materials is required to be written and implemented to ensure that materials are securely distributed from DACs to SACs to Test Administrators and securely returned from Test Administrators to SACs and then to DACs. SACs are required to distribute materials to and collect materials from the Test Administrators each day of testing and to securely store and deliver materials to DACs after testing is completed in accordance with the instructions in the *CMAS and CoAlt Procedures Manual*.

All individuals involved in the test administration are required to sign a security agreement prior to handling test materials, which requires them to follow all procedures set forth in the aforementioned manuals and prevents them from divulging the contents of the assessment, copying any part of the assessment, reviewing test items with the students, allowing students to remove test materials from the testing room, or interfering with the independent work of any student taking the assessment. During online testing, all computer functions not necessary to complete the test are disabled, and access is restricted to disallow activities in all applications outside the testing program.

PearsonAccess^{next}, the assessment management system used during the administration, includes permissions-based user role access to all information within the system, including accessing student information, setting up and delivering test sessions (preparing, starting, and stopping sessions), administering tests (unlocking, resuming, and locking units), and accessing reports. Access to the online assessments through the student testing system, TestNav, is tightly controlled before, during, and after test administration, requiring a login ID and password to enter the system for each unit. Test content is locked and cannot be accessed by students or district/school-level users after the students submit their answers. Each unit of the paper test requires students to break the unit seal before accessing the test content. To enhance security during test administration, test forms are spiraled, decreasing the likelihood that a student would be working on the same items as their peers at the same time.

After all test sessions are completed at a school, used and unused materials are required to be securely stored and returned to the DAC by the district deadline for shipment to Pearson. DACs are required to report any missing test materials or test irregularities and to complete the appropriate documentation.

Chapter 6: Scoring

The CMAS assessments use a combination of machine, human, and automated scoring. All SR and TE items are machine-scored, with point values varying by item type and assessment. Most mathematics and all science CR items are handscored, with a small number of CR mathematics items scored by the math reasoning engine (MRE). The ELA PCR items are scored on two trait dimensions using a combination of human scoring and automated scoring. Pearson's Scoring Services team conducted the handscoring for the CR, human PCR, and parallel paper-based versions of the TE items for CMAS. Appendix C presents the holistic rubrics used to score both the CMAS ELA/CSLA PCR items. To maintain comparability, scoring rules for the machine-scored items and rubrics, anchor papers, rules and scoring methods for the handscored items were preserved from previous years.

6.1. Machine Scoring

Machine-scored items include key-based and rule-based items. Key-based items tend to be a version of multiple-choice and multiple-select (i.e., students select more than one correct answer) items. Rule-based items are machine-scored TE items. Initial scoring expectations are developed during item development and are included in the item review process. The scoring rules and correct responses are included in the items' XML coding. Prior to scoring, key checks and adjudication are completed for all machine-scored items to verify that the machine is correctly identifying correct and incorrect responses. If there is a discrepancy in the scoring, content experts review the item and adjustments are made as needed. During testing, actual distribution of scores is compared to expected distribution. Further evaluation is completed if a discrepancy is identified.

6.2. Human Scoring

6.2.1. Operational Scoring

Human-scored operational items are scored using either a distributed or synchronous scoring model depending on the content area. Items on the CSLA form and paper-based TE items are scored synchronously, while scoring for all other human-scored items is completed through distributed scoring. At times, distributed scorers are leveraged to score paper-based TE items. Scoring includes several components that together provide a comprehensive performance scoring model. For example:

- All scorers are required to pass a background check and sign a nondisclosure agreement, agreeing to adhere to all security and confidentiality requirements.
- All scorers have a four-year degree at a minimum. Scorers are assigned to content areas based on their educational backgrounds, related fields of work, and their demonstrated knowledge in the content area.
- Scorers of CSLA items must be proficient in written Spanish and English languages.
- Scorers are trained using comprehensive training materials developed by scoring experts that rely on student responses scored at the rangefinding meetings. Prior to qualifying for an item, scorers review an online training module that includes an overview of scoring; information specific to the item such as the prompt and rubric; and anchor sets. Scorers then score multiple practice sets prior to attempting qualification. After successful qualification, scorers begin scoring the item.

- For CSLA items, training is led by a Pearson scoring director who presents item-specific materials, including the prompt and rubric. The scoring team then receives training on anchor sets prior to moving into the online portion of training where scorers apply scores on multiple practice sets within the electronic scoring system. After each practice set, the scoring director reviews the practice set results with the scorers prior to scorers taking the qualification sets. After successful qualification, scorers begin scoring the item.
- Scorers must pass a qualifying test for the item types that they score. Qualification sets are designed to test scorer accuracy across the range of score points for a given item.
- Student responses are converted to electronic images at Pearson facilities and are then transmitted for computer-based scoring.
- Distributed scorers are located across the United States and work from their homes. Their computers are set up for image-based scoring. A comprehensive set of scoring and monitoring tools are integrated into the scoring system, and content supervisory staff are available by phone to help answer any training or scoring questions. With distributed scoring, scorers are able to score seven days per week with extended evening hours.
- Synchronous scorers are located across the United States and also work from their homes; however, they are only permitted to score while attending daily Microsoft Teams meetings with content supervisory staff. As with distributed scoring, synchronous scoring uses a comprehensive set of scoring and monitoring tools integrated into the scoring system, with content supervisory staff available within the Microsoft Teams interface to help answer any training or scoring questions. Unlike distributed scoring, synchronous scoring is typically only completed Monday through Friday during normal business hours. Synchronous scorers are used for CSLA forms and paper-based TE items.
- Additional security procedures are in place for distributed scoring. Data are securely transmitted through HTTPS and SSL technology using secure protocols for system authentication. Student responses are randomly routed through the scoring platform to prevent scorer knowledge of student information, unless a student self-identified in the response. Scorers agree not to use shared, institutional, or public computers to score and not to save student responses or test materials. Scorer printing capabilities of materials, such as anchor papers, are only approved for printing after they have undergone and passed a personally identifiable information review by CDE. Scorers agree to securely destroy or return printed materials to Pearson at the conclusion of scoring.

Pearson's processes and tools provide a replicable quality system that strengthens consistency across projects and locations within Pearson's Scoring Services operations. Pearson's Scoring Services team uses a comprehensive system for continually monitoring and maintaining the accuracy of scoring at both the group and individual levels. This system includes daily analysis of a comprehensive set of statistical monitoring reports, as well as regular "backreading" of scorers. Reliability statistics are monitored during scoring, and interventions are applied if a scorer or item is not meeting the minimum requirements.

6.2.2. Field Test Scoring

Embedded field test scoring is completed using synchronous scoring that took place within daily Microsoft Teams meetings. All scorers are required to have a four-year college degree. Field test scorers receive stand-up training led by a Pearson scoring director who presents item-specific materials, including the prompt and rubric. Scorers then review the anchor sets in a group setting prior to scoring practice sets on paper.

6.2.3. Rangefinding

Scoring rubrics are generated for each unique item for mathematics and science, while ELA/CSLA use holistic rubrics for each item type (as shown in Appendix C:). Rubrics are finalized during rangefinding and are maintained, along with the training materials for each item, by Pearson's Scoring Services group.

Rangefinding meetings take place following the administration in which an item was field tested. The purpose of rangefinding is to define the range of performance levels within the score points of the rubrics using student responses. Each rangefinding committee includes Pearson's Scoring Services and content staff, state content representatives, and educators with relevant grade-level and content expertise and experience with special populations. Participants create consensus scores for a sample set of student responses that are subsequently used to develop effective training materials for scoring of the CR items.

Pearson's scoring directors construct one rangefinding set per item, which includes approximately 30 responses. For multi-point items, pre-constructed sets with additional responses are brought to the meeting. Responses included in these sets represent the full spectrum of scores to the greatest extent possible. The responses for each item are randomly ordered to provide committee members an opportunity to determine the spectrum of scores without bias, although actual scores are not revealed to committee members. Each set includes responses clearly earning each available score point for each item type. The set also includes sample responses that may have been challenging to score (i.e., the score points earned were not necessarily clear).

Following an introductory session presented by a member of the Scoring Services group, the rangefinding committee is divided into several breakout groups based on educator expertise. Each group is assigned a range of field test items to be reviewed based on the following process:

1. The scoring director introduces each item. The committee reviews the item and corresponding rubric.
2. The committee reads student responses—individually or as a group—and then discusses and decides the most appropriate score for each response.
3. The scoring director records committee members' comments and the final consensus score for each student response. Consensus is reached when a majority of committee members agree on a particular score point for a response and all members agree to accept the score of the majority.
4. A designated committee member records consensus scores. After reviewing responses for each item, the committee member compares their notes with those kept by the scoring director and provides sign-off to indicate agreement with the recorded scores.

Following the rangefinding meetings, Scoring Services creates training materials with an anchor set that is used for initial training (up to 15 responses) and a full practice set (up to 10 responses). For ELA, two anchor sets are used per item, one for content and one for conventions. Each CR item is then scored with the associated training materials.

6.2.4. Backreading

Backreading is the method of immediately monitoring a scorer's performance and is an important tool for Pearson's scoring supervisors. Backreading is performed in conjunction with the statistics provided by reader performance reports and as indicated by scoring directors, allowing scoring supervisors to target particular readers and areas of concern. Scorers showing low inter-rater agreement or those showing anomalous frequency distributions are given immediate, constructive feedback and monitored closely until sufficient improvement is demonstrated. Scorers who demonstrate through their agreement rates and frequency distributions that they are scoring accurately are continued to be spot-checked as an added confirmation of their accuracy. The agreement rate requirements are as follows:

- 1-point item: 90% perfect and 95% perfect plus adjacent agreement
- 2-point item: 90% perfect and 95% perfect plus adjacent agreement
- 3-point item: 80% perfect and 95% perfect plus adjacent agreement
- 4-point item: 70% perfect and 95% perfect plus adjacent agreement
- 5+-point item: 65% perfect and 95% perfect plus adjacent agreement

Section 11.5 of this technical report presents an explanation of the rater agreement statistics and the results from the Spring 2023 administration.

6.2.5. Calibration

Calibration sets are responses selected as examples that help clarify scoring issues, define more clearly the lines between certain score points, and reinforce the scoring guidelines as presented in the original training sets. They can be applied to groups, a subset of groups, or individual scorers as needed. These sets are used to proactively promote accuracy by exploring project-specific issues, score boundaries, or types of responses that are particularly challenging to score consistently. Scoring directors administer calibration sets as needed, particularly for more difficult items.

6.2.6. Validity Papers

As a quality monitoring tool used during scoring, validity papers are student responses chosen by Pearson scoring directors to measure the accuracy of a scorer when applying the scoring rubric. Validity papers are blind to scorers, which means a scorer is not aware when they are scoring a validity paper. Scoring directors may choose to include an annotation with a validity paper so that a scorer will receive immediate feedback if a validity paper is scored incorrectly. Validity statistics are monitored by scoring directors throughout the life of a scoring project.

6.3. MRE Scoring

The math reasoning engine (MRE) evaluates responses based on rubric criteria specific to the expected item responses and does not require human scoring to train the engine. Because the engine can reason about the student's math, there is no need to enumerate all possible correct responses (as one would with other more traditional machine scoring). The MRE scoring criteria are defined in terms of the mathematical characteristics of a response the engine uses to evaluate the extent to which a student response satisfies a scoring rubric (e.g., equivalence, form, precision, and constraints).

Before a mathematics item is scored operationally, it goes through a rubric refinement and validation process. After field testing, online student responses are aggregated into a set of individual unique responses that are back-read by human scorers. When the human scorers disagree with the score assigned by the scoring engine, they flag the response for further review. The flagged responses are then reviewed by Pearson content experts to determine if scoring rules should be adjusted to better fit the rubric.

If scoring rule adjustments are made, MRE can rescore the entire item based on the new scoring rules configuration. During rescores, every changed response score is identified and reviewed by content experts to evaluate the impact of the rule change on all responses to ensure that the change had the intended effect and nothing more. The data provided to committee members is based on the final scoring rules reflecting how the responses are scored operationally and includes the most common field test item responses providing committee members with additional information to evaluate item quality.

6.4. Automated Scoring

Pearson's Intelligent Essay Assessor (IEA) was used for scoring one of the 12 operational PCRs in 2023 for the CMAS ELA assessment. The prompt had an automated scoring model based on training from prior operational years and was also checked for quality using second scores by human scorers. Ten percent of responses were randomly selected and given a second reliability score to provide data for evaluating the consistency of scoring.

Some responses were not scored by the engine at all and received a first human score based on Smart Routing. The use of Smart Routing during operational scoring increases the quality of automated scoring by routing responses that are more likely to disagree with a human score to receive an additional human score. Because the cases that result from "in between" scores are based on modeling human scores, it follows that human scores may be less certain as well. Therefore, responses are more likely to be double-scored and resolved if the IEA and human scores are non-adjacent. Smart routing is used as needed to achieve targeted quality metrics (e.g., validity agreement or agreement with human scorers).

6.4.1. Quality Criteria for Evaluating Automated Scoring

The primary evaluation criteria for IEA are based on responses to validity papers with "known" scores assigned by experts. A set of validity papers seeded into human scoring throughout the administration is used to monitor the human-scoring process over time. The expectation is that IEA can score validity papers at least as accurately as humans can score the papers. Additional measures for evaluating automated scoring include Pearson correlation, kappa, quadratic-weighted kappa (QWK), exact agreement, and standardized mean difference (SMD) that are computed between pairs of human scores and between IEA and humans to evaluate how performance is the same or different. Criteria for evaluating the training of IEA given these measures include the following:

- Pearson correlation between IEA and human scores should be within 0.1 of human–human correlation.
- Kappa between IEA and human scores should be within 0.1 of human–human kappa.
- QWK between IEA and human scores should be within 0.1 of human–human QWK.

- Exact agreement rate for IEA and human scores should be within 3.0% of the human–human exact agreement rate.
- SMD between IEA and human scores should be less than 0.15.

The specific criteria for evaluating IEA included both primary and secondary criteria, as described below.

6.4.1.1. Primary Criterion

The performance of IEA was evaluated by comparing IEA scores with human scores for the set of validity papers. The primary criterion is stated as follows: *With smart routing applied as needed, IEA agreement is as good as or better than human agreement for each trait score.* For a given prompt, this criterion is operationalized as follows:

1. Determine agreement of the human scores with the validity papers for each trait.
2. Calculate agreement of the IEA scores with the validity papers for each trait.
3. Compare the IEA-human agreement on the validity papers.
4. Deploy IEA operationally if the IEA validity agreement is greater than or equal to the human agreement for each trait.

When it is not possible to use human-scored validity responses in evaluating IEA performance, IEA is evaluated based on IEA–human exact agreement for each trait score and compared to agreement based on responses that are double-scored by humans. IEA-human agreement is evaluated on a portion of the data according to the following steps:

1. Determine exact agreement of the two human scores with each other for each trait.
2. Calculate agreement of the IEA scores with the human scores for each trait.
3. Compare the IEA-human agreement with the human-human agreement.
4. Deploy IEA operationally if the IEA-human agreement is within 5.25% of the human-human agreement.

In addition to the overall comparison, the following performance thresholds are targeted in the test dataset: (1) at least 65% overall IEA-human agreement and (2) 50% IEA-human agreement by score point (i.e., conditioned on the human score). These targets go beyond the contingent primary criteria approved by the consortium state leads.

6.4.1.2. Secondary Criterion

The secondary criterion involves comparing agreement indices for IEA-human scoring for various demographic subgroups and is stated as follows: *With smart routing applied as needed, IEA-human differences on statistical measures for each trait score are within the Williamson et al. (2012) tolerances for subgroups with at least 50 responses.* IEA-human agreement is evaluated according to the following steps:

1. Determine exact agreement of the two human scores with each other for each trait.
2. Calculate agreement of the IEA scores with the human scores for each trait.
3. Compare the IEA-human agreement with the human–human agreement.
4. For subgroups with at least 50 IEA-human scores and at least 50 human-human scores, compare agreement indices to the following criteria:

a. Pearson correlation between IEA-human should be within 0.1 of human-human.
b. Kappa between IEA-human should be within 0.1 of human–human.
c. Quadratic-weighted kappa between IEA-human should be within 0.1 of human-human.
d. Exact agreement between IEA-human should be within 5.25% of human-human.
e. SMD between IEA-human should be less than ±0.15 (this criterion was applied to subgroups with at least 50 IEA-human scores).

Although it is not expected that these criteria will be met for all subgroups for all prompts, if results of the evaluation between IEA and human scoring for subgroups for any prompt indicate that IEA performance persistently fails on the criteria listed above, considerations would be given to resetting the responses scored by IEA and reverting to human scoring until such time that an alternate IEA model could be established with improved subgroup performance.

In addition to the secondary criterion, the performance of IEA is also compared with the following targets on the various measures for subgroups with at least 50 responses:

- Pearson correlation between IEA-human should be 0.70 or above.
- Kappa between IEA-human should be 0.40 or above.
- Quadratic-weighted kappa between IEA-human should be 0.70 or above.
- Exact agreement between IEA-human should be 65% or above.

6.4.2. Hierarchy of Assigned Scores for Reporting

When multiple scores are assigned for a given response, the following hierarchy determines which score is reported operationally:

- The IEA score is reported if it is the only score assigned.
- If an IEA score and a human score are assigned, the human score is reported.
- If two human scores are assigned, the first human score is reported.
- If a backread score and human and/or IEA scores are assigned, the last backread score is reported.
- If a resolution score is assigned, the resolution score is reported. If nonadjacent scores are encountered, responses are automatically routed to resolution.

Chapter 7: Standard Setting

To support the interpretation of student results, student performance on the CMAS assessments is described in terms of performance levels as presented in Table 2.2. Standard setting is the process of translating those policy-driven performance standards into scores on the assessment. The purpose of standard setting is to determine the boundaries—or cut scores—along the score scale that differentiate student performance among those levels (e.g., Cizek et al., 2004; Kane, 1994).

Table 7.1 presents the cut scores for each content area and grade. The mathematics and ELA cut scores were set in 2015 in collaboration with the PARCC consortium using the Evidence-Based Standard Setting (EBSS) method (Beimers et al., 2012), as detailed in the 2015 PARCC *Performance Level Setting Technical Report* (Davis & Moyer, 2015). CSLA cut scores were set in 2016 using the Modified Extended Angoff method, as detailed in the *CSLA Colorado Spanish Language Arts Standard Setting Report* (CDE, 2016). Standard setting for the new science assessment took place from September 27–28, 2022, using a modified version of the Item Descriptor (ID) Matching method (Ferrara et al., 2008), as detailed in the *CMAS Science 2022 Standard Setting Report* (Pearson, 2024).

Table 7.1. Performance Level Cut Scores

Content Area	Grade	*Did Not Yet Meet Expectations*	*Partially Met Expectations*	*Approached Expectations*	*Met Expectations*	*Exceeded Expectations*
Mathematics	3	650–699	700–724	725–749	750–789	790–850
	4	650–699	700–724	725–749	750–795	796–850
	5	650–699	700–724	725–749	750–789	790–850
	6	650–699	700–724	725–749	750–787	788–850
	7	650–699	700–724	725–749	750–785	786–850
	8	650–699	700–724	725–749	750–800	801–850
ELA	3	650–699	700–724	725–749	750–809	810–850
	4	650–699	700–724	725–749	750–789	790–850
	5	650–699	700–724	725–749	750–798	799–850
	6	650–699	700–724	725–749	750–789	790–850
	7	650–699	700–724	725–749	750–784	785–850
	8	650–699	700–724	725–749	750–793	794–850
Science	5	–	650–724	725–749	750–788	789–850
	8	–	650–724	725–749	750–796	797–850
	11	–	650–724	725–749	750–786	787–850
CSLA	3	650–699	700–724	725–749	750–778	779–850
	4	650–699	700–724	725–749	750–771	772–850

The ELA assessment also includes a Reading score that has the same range and cut score for all grades. There is only one cut score that corresponds to the *Met Expectations* overall performance level, as shown in Table 7.2, that was determined using the cut information from setting the standards on the overall ELA test (i.e., it was not set separately at the standard setting meeting).

Table 7.2. ELA Reading *Met Expectations* Cut Score

Scale Range	Cut Score
110–190	150

Science also includes performance indicator cut scores that indicate average performance in each reporting category compared to the state, as shown in Table 7.3. These cuts are not used for accountability and change from year to year. Students with scores below this range scored "lower than average" in the reporting category, and students above the range scored "higher than average."

Table 7.3. 2023 CMAS Science Performance Indicator Cut Scores

Grade	Physical Science	Life Science	Earth and Space Science	SEP
5	441–519	440–522	446–519	448–517
8	444–515	440–516	442–516	447–514
11	442–511	441–512	440–512	447–509

Chapter 8: Reporting

8.1. Description of Scores

The CMAS reports provide information on student performance in terms of scale scores, performance levels, percentile ranks, and percent earned scores, as described below.

8.1.1. Scale Scores

A scale score is a conversion of a student's response pattern to a common scale that allows for a numerical comparison between students. Scale scores are particularly useful for comparing test scores over time and creating comparable scores when a test has multiple forms. Students taking CMAS mathematics and science receive overall scale scores, whereas students taking the CMAS ELA/CSLA assessments receive scale scores at the overall test and Reading claim levels. Science reports also provide separate scale scores for content standards and SEPs (referred to as reporting categories). The overall scale for each content area assessment ranges from 650 to 850, and the ELA Reading scale ranges from 110 to 190, as shown in Table 7.1 and Table 7.2 in the previous chapter. The science content standards scale score ranges from 400 to 600, although the graph displayed on the student reports ends at 550. Any student who earned a score greater than 550 still had their score written on the report but the diamond representing this performance would appear at the end of the graph at 550.

8.1.2. Performance Levels

Performance levels and their accompanying PLDs are reported at the overall assessment level. Students are classified into performance levels based on their scale score and the cut scores obtained from standard setting. The CMAS Mathematics and ELA assessments have five performance levels (*Did Not Yet Meet Expectations*, *Partially Met Expectations*, *Approached Expectations*, *Met Expectations*, and *Exceeded Expectations*), whereas CMAS Science has four performance levels (*Partially Met Expectations*, *Approached Expectations*, *Met Expectations*, and *Exceeded Expectations*). Students in the top two performance levels met or exceeded the expectations of the CAS and are considered on track for the next grade level in that content area.

8.1.3. Percentile Ranking

Percentile rankings are provided on student performance reports to indicate how the student performed compared with other students in the state. For example, a student with a percentile ranking of 70 performed better than 70% of students in Colorado. The percentile rankings are based on the overall scale score.

8.1.4. Percent Earned

To prevent incorrect interpretations and provide a metric that is more generally understood, student performance for subclaims and the Writing claim (ELA/CSLA) are reported as the percentage of points earned (i.e., the number of points a student earned out of the total number of points possible within a claim or subclaim). Unlike scale scores, the percent of points possible scores cannot be compared across years because individual items change from year to year and are not constructed to be comparable in difficulty at the claim, subclaim, or subscale level. Performance on different subclaims or subscales also cannot be compared within an administration because the number of items and the difficulty of the items within each claim, subclaim or subscale may not be the same.

The percent of points possible can be compared to aggregated state, district, and school performance. The student performance reports also include an indicator of how students who scored just above the *Met Expectations* cut score on the overall assessment performed on each category. This indicator gives similar information to the *Met Expectations* cuts.

8.2. Score Reports

Two types of score reports are provided: (1) the student-level Student Performance Report and (2) the aggregate reports at the school and district levels. Appendix D presents sample Student Performance Reports, and examples of each type of aggregate report are provided in the *CMAS and CoAlt Interpretive Guide to Assessment Reports*. For a detailed explanation of the information provided in all reports, refer to the *CMAS and CoAlt Interpretive Guide to Assessment Reports* located online at https://www.cde.state.co.us/assessment/cmas_coalt_interpretiveguide_2023. CSLA assessments are parallel and comparable to the CMAS ELA assessments in scoring and reporting. Therefore, separate CSLA reports are not included (please refer to the CMAS ELA examples).

8.2.1. Student Performance Reports

The Student Performance Report provides information about the performance of a particular student. The student's scale score, performance level, percentile ranking, and percent of points possible scores are displayed on a two-page report, along with comparative information related to the student's school, district, and state performance. PLDs are also provided. In addition to the electronic versions made available to districts and schools, two copies of the Student Performance Report are printed and shipped to districts for distributing to parents/guardians and for maintaining locally.

8.2.2. Aggregate Reports

The following aggregate reports are produced at the school and/or district levels and provide summary information for a given school or district. They are provided electronically through PearsonAccess^{Next}, with access limited to authorized users. The participation report provides a comparison of the demographic characteristics of the tested students compared to all students eligible for testing. This information can assist districts and schools in determining how to interpret their aggregated results.

- Performance Level Summaries
- Content Standards Rosters
- Evidence Statement Analysis Reports (mathematics and ELA only)
- Item Analysis Report (science only)
- District Summary of Schools (district level only)
- District and School Participation Reports

Chapter 9: Test Results and Analysis

This chapter presents the test results and statistical analyses for the Spring 2023 CMAS assessments in mathematics, ELA/CSLA, and science.

9.1. Student Participation

Table 9.1 presents a breakdown of the number of students who took the assessment online compared with those who took accommodated forms, and Appendix E presents n-counts for various demographic characteristics for the students who took the CMAS assessments. Most students took the assessments online.

Table 9.1. Student Participation N-Count by Form

Content Area	Form	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
Mathematics	Online	51,956	51,828	52,685	52,174	50,536	48,825	–
	Spanish Online	1,188	1,084	617	346	372	335	–
	Paper	4,127	3,810	3,549	3,370	3,207	2,843	–
	Spanish Paper	93	39	14	5	12	12	–
	Text-to-Speech	12,430	11,426	10,405	5,211	4,089	3,620	–
	Spanish Text-to-Speech	649	530	355	138	125	130	
	Assistive Technology	7	11	6	7	11	6	–
	Total	57,371	56,772	56,871	55,902	54,138	52,021	–
ELA	Online	51,525	51,652	53,041	52,552	51,105	49,260	–
	Paper	4,201	3,854	3,604	3,038	2,771	2,489	–
	Assistive Technology	11	13	12	12	19	11	–
	Total	55,737	55,519	56,657	55,602	53,895	51,760	–
CSLA	Paper	1,440	1,180	–	–	–	–	–
Science	Online	–	–	43,155	–	–	45,087	29,886
	Spanish Online	–	–	203	–	–	189	131
	Paper	–	–	2,586	–	–	2,037	816
	Spanish Paper	–	–	10	–	–	12	1
	Text-to-Speech	–	–	10,175	–	–	3,499	855
	Spanish Text-to-Speech	–	–	299	–	–	123	78
	Total	–	–	56,428	–	–	50,947	31,767

9.2. Performance Results

Table 9.2 presents summary statistics for the overall scale scores, including the mean, standard deviation (SD), and median. The previous year's results are also included for comparison. The tables also include the coefficient alpha (see Section 11.1 for more information). Table 9.3 presents the performance level distributions and includes the distributions from last year's administration for comparison.

Appendix F presents the cumulative scale score distributions by grade, Appendix G displays the same information in graphical form, and Appendix H presents the summary statistics for the overall scale scores by demographic subgroup.

Table 9.2. Scale Score Performance Summary

Content Area	Grade	2023				2022			
		N	Mean	SD	Median	N	Mean	SD	Median
Mathematics	3	57,382	738	39.5	739	56,482	737	39.3	738
	4	56,789	733	33.5	733	56,886	732	32.9	732
	5	56,896	737	35.2	736	57,423	736	35.1	734
	6	55,913	730	33.3	729	55,936	728	32.5	728
	7	54,148	731	28.1	729	55,283	730	27.7	728
	8	52,036	732	40.9	728	52,804	731	40.1	728
ELA	3	55,737	737	43.6	738	55,081	737	43.9	739
	4	55,519	742	36.9	744	55,746	740	36.4	744
	5	56,657	747	33.8	747	57,362	745	32.8	745
	6	55,602	743	33.0	744	55,960	742	34.1	743
	7	53,895	744	38.0	745	55,265	741	37.2	742
	8	51,760	741	40.7	742	52,727	742	40.9	743
ELA – Reading	3	55,737	145	17.7	145	55,081	145	17.7	145
	4	55,519	147	14.8	147	55,746	146	14.6	148
	5	56,657	149	13.6	149	57,362	148	13.2	148
	6	55,602	147	13.2	147	55,960	147	13.7	147
	7	53,895	148	15.1	148	55,265	146	14.8	146
	8	51,760	147	16.2	147	52,727	147	16.2	147
CSLA	3	1,440	724	27.3	725	1,301	726	26.9	726
	4	1,180	725	23.8	725	1,151	726	22.0	727
CSLA – Reading	3	1,440	140	10.2	139	1,301	140	10.5	139
	4	1,180	140	9.3	140	1,151	141	8.3	140
Science	5	56,428	733	33.9	737	–	–	–	–
	8	50,947	731	33.0	735	–	–	–	–
	11	31,767	729	29.5	732	–	–	–	–

Note. The Spring 2022 CMAS Science assessment reported percentile ranks only.

Table 9.3. Performance Level Distribution: Percent of Students in Each Performance Level

Content Area	Grade	2023					2022				
		1	2	3	4	5	1	2	3	4	5
Mathematics	3	17.76	18.62	23.25	31.02	9.34	18.68	18.95	23.00	30.73	8.64
	4	17.38	23.48	26.46	29.75	2.93	18.05	24.05	27.20	28.28	2.43
	5	15.17	23.48	24.85	29.05	7.44	16.17	24.31	24.59	28.06	6.86
	6	19.93	24.86	26.97	24.17	4.07	20.33	26.02	27.37	22.84	3.44
	7	12.10	31.95	29.69	23.46	2.79	12.01	32.73	30.21	22.49	2.56
	8	23.74	23.27	20.32	26.98	5.68	23.60	23.05	20.97	27.42	4.96
ELA	3	22.19	17.03	20.91	34.92	4.95	22.14	16.56	20.55	35.86	4.88
	4	14.52	16.47	25.22	35.05	8.72	15.11	15.59	25.18	36.85	7.27
	5	7.93	18.51	25.80	41.11	6.65	8.21	19.34	27.05	40.62	4.79
	6	10.31	20.16	26.16	35.75	7.62	11.49	18.52	26.95	35.64	7.40
	7	13.23	17.52	24.24	30.35	14.67	15.09	18.93	24.21	29.46	12.30
	8	16.54	17.95	23.14	32.12	10.26	16.65	17.08	22.33	33.29	10.65
CSLA	3	19.93	29.65	31.74	16.81	1.88	18.45	27.98	33.74	17.60	2.23
	4	13.98	34.07	37.71	12.20	2.03	11.56	33.54	41.18	12.51	1.22

Content Area	Grade	2023					2022				
		1	2	3	4	5	1	2	3	4	5
Science	5	–	35.33	30.79	31.15	2.73	–	–	–	–	–
	8	–	38.53	30.18	30.76	0.53	–	–	–	–	–
	11	–	38.79	36.63	23.84	0.75	–	–	–	–	–

Note. The Spring 2022 CMAS Science assessment reported percentile ranks only. 1 = *Did Not Yet Meet Expectations*, 2 = *Partially Met Expectations*, 3 = *Approached Expectations*, 4 = *Met Expectations*, 5 = *Exceeded Expectations*. Percentages may not sum to 100 due to rounding.

Appendix I presents the summary statistics for points earned by subclaim. While the overall scale scores and Reading scale scores are comparable to results from previous administrations, the assessments are not designed to permit meaningful comparisons across percent earned scores, either within an assessment or across administration years. The difficulty of the items that make up each subscore can vary across subscores and from year to year, making it inappropriate to make inferences based on percent earned performance across subscores or based on subscore performance across years. The only percent earned subscore comparisons supported by the CMAS assessments are those comparing individual or group performance within one subclaim with the performance of other students or groups within the same subclaim and administration.

9.3. Classical Item Analysis

Appendix J presents the item-level classical statistics for each CMAS assessment, including the omit rate, *p*-value, item-total correlation, and the percentage of students earning each score point (CR items only).

Item difficulty is measured by the *p*-value, which is bounded by 0.0 and 1.0. The *p*-value for 1-point items is the proportion of students who answered an item correctly and is calculated 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) that is then put on a 0 to 1 scale by dividing the average item score by the maximum number of points for 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. Easy and hard items are both necessary to include on an assessment to balance the test difficulty.

Item discrimination is represented by the item-total correlation (also known as the point-biserial correlation), is bounded by -1.0 and 1.0, and indicates how well an item discriminates, or distinguishes, between low-performing and high-performing students. The item-total 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 do well on a given item, and students who do not do well on a test are expected to not do well on a given item. 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 item-total correlation discriminates between low-performing and high-performing students better than an item with an item-total correlation near zero. A negative item-total correlation indicates that low-performing students did better on that item than high-performing students.

9.4. Subclaim Correlations

The ELA/CSLA tests include Reading and Writing claim scores and five subclaim scores: Reading: Literary Text (RL), Reading: Informational Text (RI), Reading: Vocabulary (RV), Writing: Written Expression (WE), and Writing: Knowledge and Use of Language Conventions (WKL). The Reading score is a composite of RL, RI, and RV, and the Writing score is a composite of WE and WKL reported as a percentage of points earned. It comprises PCR items only. The operational test analyses were performed by evaluating the separate trait scores of WE and WKL. Some PCR items also include RL or RI points, but the reading points for those items were a duplicate of the WE score and were not included in calibrations.

The mathematics tests have four subclaim scores: Subclaim A: Major Content, Subclaim B: Additional & Supporting Content, Subclaim C: Expressing Mathematical Reasoning, and Subclaim D: Modeling & Application. The science test has four subclaim scores: Physical Science, Life Science, Earth Systems Science, and Science and Engineering Practices (SEPs).

One way to assess the internal structure of a test is through the evaluation of correlations among subscores, as presented in Table 9.4 – Table 9.8. For CMAS ELA/CSLA, these analyses were conducted between the Reading and Writing claim scores and the subclaims (RL, RI, RV, WE, and WKL). For CMAS Mathematics and Science, the analyses were conducted between the subclaim scores. There is evidence of unidimensionality if the components within a content area are strongly related to each other.

The intercorrelations for the mathematics and science subclaims were higher overall than the ELA/CSLA intercorrelations. Correlations between subclaims for mathematics and science ranged from 0.61 to 0.83, while for ELA/CSLA they ranged from 0.27 to 0.91. For CMAS ELA/CSLA, the two writing subclaims tended to have higher correlations with one another than they did with any of the reading subclaims. Correlations between the subclaims and the total test ranged from 0.63 to 0.96.

Table 9.4. Correlations Between Subclaims—Mathematics

Grade	Subclaim	Subclaim B	Subclaim C	Subclaim D	Total Test
3	A	0.831	0.736	0.756	0.957
	B	–	0.697	0.722	0.901
	C	–	–	0.692	0.855
	D	–	–	–	0.860
4	A	0.744	0.791	0.732	0.960
	B	–	0.666	0.625	0.815
	C	–	–	0.702	0.894
	D	–	–	–	0.843
5	A	0.781	0.784	0.765	0.960
	B	–	0.699	0.714	0.866
	C	–	–	0.747	0.883
	D	–	–	–	0.873
6	A	0.712	0.768	0.748	0.943
	B	–	0.662	0.632	0.829
	C	–	–	0.729	0.891
	D	–	–	–	0.861

Grade	Subclaim	Subclaim B	Subclaim C	Subclaim D	Total Test
7	A	0.682	0.779	0.740	0.946
	B	–	0.648	0.610	0.797
	C	–	–	0.744	0.900
	D	–	–	–	0.860
8	A	0.715	0.794	0.768	0.957
	B	–	0.667	0.674	0.820
	C	–	–	0.745	0.889
	D	–	–	. –	0.877

Table 9.5. Correlations Between Subclaims—ELA

Grade	Subclaim	RI	RV	WE	WKL	Total Test
3	RL	0.724	0.674	0.659	0.551	0.904
	RI	–	0.703	0.577	0.504	0.876
	RV	–	–	0.535	0.478	0.826
	WE	–	–	–	0.603	0.811
	WKL	–	–	–	–	0.673
4	RL	0.684	0.683	0.691	0.615	0.898
	RI	–	0.646	0.543	0.525	0.845
	RV	–	–	0.547	0.505	0.800
	WE	–	–	–	0.783	0.844
	WKL	–	–	–	–	0.766
5	RL	0.679	0.595	0.616	0.582	0.871
	RI	–	0.538	0.685	0.631	0.862
	RV	–	–	0.496	0.459	0.706
	WE	–	–	–	0.817	0.877
	WKL	–	–	–	–	0.806
6	RL	0.738	0.619	0.644	0.632	0.881
	RI	–	0.621	0.702	0.677	0.899
	RV	–	–	0.504	0.489	0.729
	WE	–	–	–	0.885	0.877
	WKL	–	–	–	–	0.841
7	RL	0.729	0.722	0.731	0.700	0.901
	RI	–	0.672	0.616	0.607	0.852
	RV	–	–	0.589	0.581	0.802
	WE	–	–	–	0.906	0.896
	WKL	–	–	–	–	0.863
8	RL	0.759	0.692	0.665	0.661	0.867
	RI	–	0.692	0.741	0.725	0.909
	RV	–	–	0.580	0.578	0.781
	WE	–	–	–	0.943	0.909
	WKL	–	–	–	–	0.890

Note. RL = Reading: Literary Text, RI = Reading: Informational Text, RV = Reading: Vocabulary, WE = Writing: Written Expression, WKL = Writing: Knowledge and Use of Language Conventions.

Table 9.6. Correlations Between Subclaims—CSLA

Grade	Subclaim	RI	RV	WE	WKL	Total Test
3	RL	0.564	0.735	0.689	0.500	0.895
	RI	–	0.584	0.463	0.334	0.722
	RV	–	–	0.561	0.442	0.824
	WE	–	–	–	0.626	0.866
	WKL	–	–	–	–	0.668
4	RL	0.534	0.649	0.706	0.490	0.898
	RI	–	0.499	0.424	0.268	0.685
	RV	–	–	0.503	0.294	0.730
	WE	–	–	–	0.565	0.880
	WKL	–	–	–	–	0.626

Note. RL = Reading: Literary Text, RI = Reading: Informational Text, RV = Reading: Vocabulary, WE = Writing: Written Expression, WKL = Writing: Knowledge and Use of Language Conventions.

Table 9.7. Correlations between Subclaims—Reading vs. Writing

Content Area	Grade	Correlation
ELA	3	0.697
	4	0.691
	5	0.718
	6	0.727
	7	0.730
	8	0.754
CSLA	3	0.678
	4	0.671

Table 9.8. Correlations Between Subclaims—Science

Grade	Subclaim	Life Science	Earth and Space Science	Total Test
5	Physical Science	0.693	0.784	0.928
	Life Science	–	0.698	0.844
	Earth and Space Science	–	–	0.931
8	Physical Science	0.812	0.766	0.933
	Life Science	–	0.768	0.939
	Earth and Space Science	–	–	0.900
11	Physical Science	0.760	0.751	0.930
	Life Science	–	0.726	0.906
	Earth and Space Science	–	–	0.896

Note. For Grade 5, the subclaim is Physical Science/Life Science.

Chapter 10: Calibration, Equating, and Scaling

Item response theory (IRT) was used to develop, calibrate, equate, and scale the CMAS assessments. All test analyses including calibration, scaling, and item model fit were accomplished within the IRT framework. The CMAS Mathematics and ELA scales were equated to the previous CMAS (i.e., PARCC) base scale. The calibration of the first operational administration determined the base scale for CSLA and CMAS Science.

Calibration is the process of estimating the parameters (such as item difficulty) for each item on an assessment so that all items are placed on a common scale. To maintain the same performance standards across different administrations of a particular test, it is necessary for each administration of the test to be of comparable difficulty. It is not fair to compare students to a common standard if the overall difficulty of the forms changes from year to year. Maintaining test form difficulty across administrations is achieved through equating. Equating adjusts for differences in overall test difficulty of test forms so that the scores resulting from two different administrations can be considered interchangeable.

Equating and scaling typically occur in sequence. First, equating is used to adjust for differences in test difficulty so resulting estimates of student proficiency (i.e., equated raw scores, theta estimates) are on a common metric. The equated estimates of proficiency are then converted to scale scores for reporting purposes.

Calibration, equating, and scaling analyses for the operational and field test items were as follows for the Spring 2023 administration. The entire process was completed for each CMAS assessment, and all steps were independently replicated by at least two members of the Pearson psychometrics team to ensure accuracy.

- CMAS Mathematics
    - Operational items
        - All items had parameter estimates already equated to the base scale
        - Estimated student abilities using ISE (Chien & Shin, 2012)
    - Embedded field test items
        - Used IRTPRO control files and IDM to obtain item parameter estimates of the operational and field test items
        - Used STUIRT to scale the field test items to the operational scale using the online operational items as the anchor set
        - Calculated item fit statistics and plotted expected vs. observed IRFs for each field test item

- CMAS ELA
    - Operational items
        - Obtained the online operational item parameter estimates using IRTPRO (Vector Psychometric Group, 2022) control files and incomplete data matrix (IDM)
        - Evaluated the consistency of scoring and stability of the anchor items
        - Scaled the 2023 operational items to the operational scale using STUIRT (Kim & Kolen, 2004)

        - Calculated item fit statistics and plotted expected vs. observed item response functions (IRFs) for each operational item
        - Estimated student abilities using ISE (Chien & Shin, 2012)
    - Embedded field test items
        - Obtained item parameter estimates of the operational and field test items using IRTPRO control files and IDM
        - Scaled the field test items to the operational scale using STUIRT and the online operational items as the anchor set
        - Calculated item fit statistics and plotted expected vs. observed IRFs for each field test item

- CSLA
    - Operational items
        - Obtained the non-anchor operational item parameter estimates using Winsteps (Linacre, 2011) control files and IDM
        - Evaluated the stability of the anchor items to obtain the final anchor set
        - Scale the 2023 non-anchor items to the operational scale using the final anchor set in Winsteps
        - Obtained item difficulty values, step deviation values, and item fit values for all items
        - Estimated student abilities using Winsteps
    - Embedded field test items
        - Scale the field test item parameter estimates to the operational scale by fixing the item parameter estimates of the operational items using Winsteps control files and IDM
        - Obtained field test item difficulty values, step deviation values, and item fit values for each field test item

- CMAS Science
    - Operational items
        - Obtained the online operational item parameter estimates using IRTPRO (Vector Psychometric Group, 2022) control files and IDM
        - Evaluated the consistency of scoring and stability of the anchor items
        - Scaled the 2022 items and cut scores to the operational scale using STUIRT (Kim & Kolen, 2004)
        - Calculated item fit statistics and plotted expected vs. observed item response functions (IRFs) for each operational item
        - Estimated student abilities using ISE (Chien & Shin, 2012)
    - Embedded field test items
        - Obtained item parameter estimates of the operational and field test items using IRTPRO control files and IDM
        - Scale the field test items to the operational scale using STUIRT and the online operational items as the anchor set
        - Calculated item fit statistics and plotted expected vs. observed IRFs for each field test item

10.1. IRT Models

The two-parameter logistic (2PL; Birnbaum, 1968) and generalized partial credit (GPC; Muraki, 1992) models were applied to CMAS Mathematics and ELA; the 2PL, three-parameter logistic (3PL; Birnbaum, 1968), and GPC models were applied to CMAS Science; and the Rasch partial credit model (RPCM) was applied to CSLA. The 2PL model uses two item parameters to relate the probability of person *i* correctly answering a dichotomously scored item *j*:

$$P_{ij}(\theta) = \frac{1}{1+\exp\left[-Da_j(\theta_i - b_j)\right]}$$

where *D* is set equal to 1 when defined on the logistic scale, as IRTPRO parameterizes all models. The item discrimination parameter is a_j, and the item difficulty parameter is b_j. The 3PL model adds an item parameter:

$$P_{ij}(\theta) = c_j + \frac{1-c_j}{1+\exp\left[-Da_j(\theta_i - b_j)\right]}$$

where c_j is the item pseudo-guessing parameter. The GPC model has three item parameters to relate the probability of person *i* responding in the *x*-th category to a polytomous scored item *j*:

$$P_{ij}(\theta) = \frac{\exp\left[\sum_{v=0}^{x} Da_j\left(\theta_i - b_j + d_{jv}\right)\right]}{\sum_{k=0}^{M_i}\exp\left[\sum_{v=0}^{k} Da_j\left(\theta_i - b_j + d_{jv}\right)\right]}, x = 0, 1, \ldots, M_j$$

where all parameters are as they were before, and d_{jv} is the category parameter for category *v* of item *j* and M_j is the maximum score on item *j*. To put the parameters on the normal ogive metric, the a_j is then divided by 1.7.

The RPCM used for CSLA is an extension of the Rasch one-parameter IRT model attributed to Georg Rasch (1966), as extended by Wright and Stone (1979), Masters (1982), and Wright and Masters (1982). The RPCM is a mathematical measurement model with a single item parameter relating a student's performance on a given item involving *m*+1 score categories. The probability of student *n* scoring *x* on *m* steps of item *i* is a function of the student's proficiency level, θ_n (also referred to as "ability"), and the step difficulties, δ_{ij}, of the *m* steps in question *i* as follows:

$$P_{xni} = \frac{exp\sum_{j=0}^{x}\left(\theta_n - \delta_{ij}\right)}{\sum_{k=0}^{m_i} exp\sum_{j=0}^{k}\left(\theta_n - \delta_{ij}\right)}, x = 0, 1, \ldots m_i$$

10.2. Item Response Curves (IRCs) and Item Characteristic Curves (ICCs)

The IRFs of the 2PL, 3PL, and GPC IRT models relate student ability to the probability of observing a particular item response given the item's characteristics, whereas the item characteristic function (ICF) relates student ability to the expected student score. The graphical representation of the IRF and ICF are the item response curves (IRCs) and item characteristic curves (ICCs), respectively. The IRF and ICF for dichotomous items are equal, but the IRC and ICF are different for polytomous items.

For example, consider Figure 10.1 that depicts a 2PL item that falls at approximately 0.85 on the ability (horizontal) scale. When a student answers an item at the same level as their ability, they have a roughly 50% probability of answering the item correctly. Another way of expressing this is that in a group of 100 students, all of whom have an ability of 0.85, about 50% of them would be expected to answer the item correctly. A student whose ability is above 0.85 would have a higher probability of getting the item right, while a student whose ability is below 0.85 would have a lower probability of getting the item right.

Figure 10.1. Sample 2PL Item Characteristic Curve (ICC)

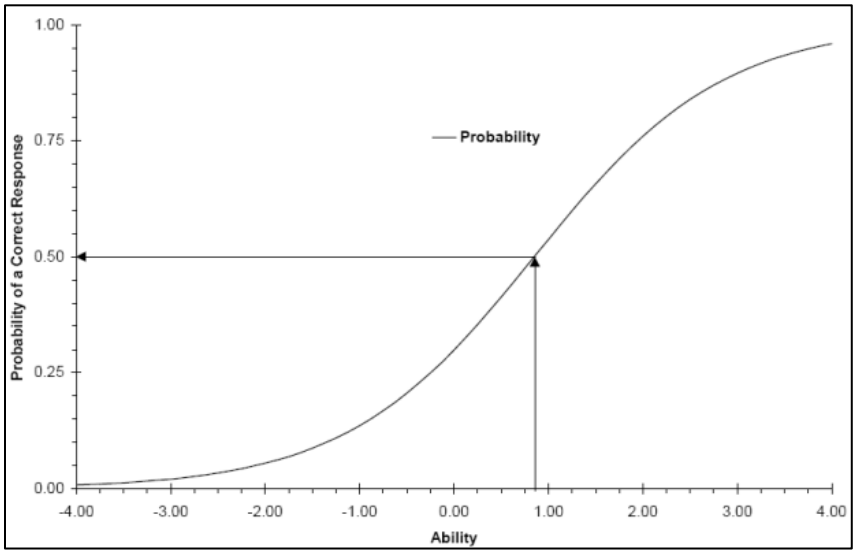


Figure 10.2 shows IRCs of obtaining a wrong answer or a right answer. The dotted-line curve (j=0) shows the probability of getting a score of 0, while the solid-line curve (j=1) shows the probability of getting a score of 1. The point at which the two curves cross indicates the transition point on the ability scale where the most likely response changes from a 0 to a 1. At this intersection, the probability of answering the item correctly is 50%.

Figure 10.2. Sample 2PL Item Response Curves (IRCs)

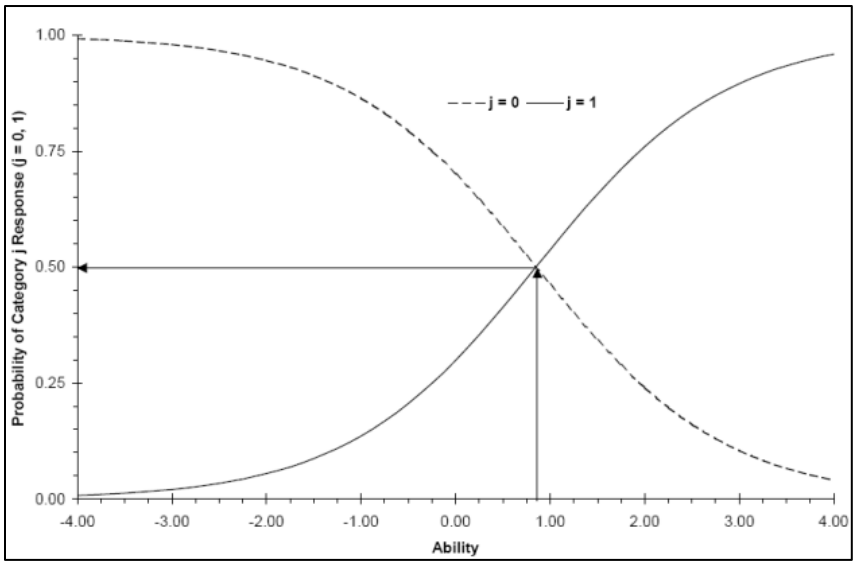
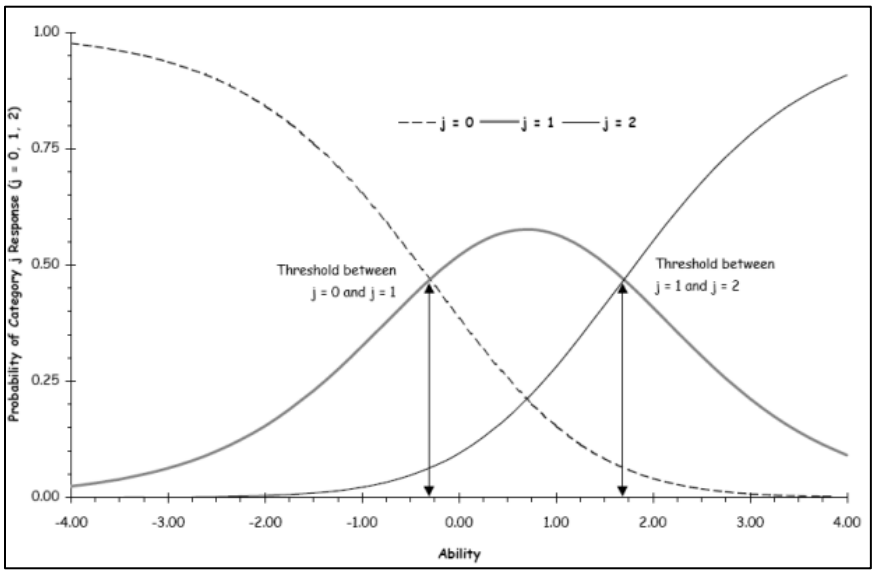


Figure 10.3 shows IRCs of obtaining each score category for a polytomously scored item. The dotted-line curve (*j*=0) shows the probability of getting a score of 0. Those of very low ability (e.g., below -2) are likely to be in this category. Those receiving a 1 (partial credit) tend to fall in the middle range of abilities (the thick, solid-line curve, *j*=1). The final, thin, solid-line curve (*j*=2) represents the probability for those receiving scores of 2 (completely correct). Very high-ability students are more likely to be in this category, but there are still some of average and low ability who can get full credit for the item.

The points at which the lines cross have a similar interpretation as that for dichotomous items. For abilities to the left of (or less than) the point at which the *j*=0 line crosses the *j*=1 line, indicated by the left arrow, the probability is greatest for a 0 response. To the right of (or above) this point and up to the point at which the *j*=1 and *j*=2 lines cross (marked by the right arrow), the most likely response is a 1. For abilities to the right of this point, the most likely response is a 2. The probability of scoring a 1 response (*j*=1) declines in both directions as ability decreases to the low extreme and increases to the high extreme. These points may be thought of as the difficulties of crossing the *thresholds* between categories.

Figure 10.3. Item Response Curves (IRCs) for a 2-point Polytomous Item



10.3. Data Preparation

Prior to any analyses, several steps were completed as preparation:

- Verify the data file containing student responses and apply the exclusion rules.
- Complete a traditional item analysis (TRIAN) and adjudication, where applicable, on all items.
- Create incomplete data matrices (IDMs).

A TRIAN of all SR items was conducted prior to calibration. The purpose of this review is to use classical statistics to identify potential test administration and score issues. Items with one or more of the following characteristics are flagged:

- *P*-value <0.15
- Item-total score correlation < 0.10
- Incorrect option selected by more high-performing students (top 33%) than the keyed response
- Distractor *p*-value ≥ 40%
- Distractor-total score correlation > 0
- One or more score points earned by less than 5% of students

A list of flagged items is communicated to the assessment specialists for review and confirmation that the correct key has been applied. Figure 10.4 presents a sample TRIAN report.

Figure 10.4. Sample Key Check (TRIAN) Report

Item	Form	Key	Corr.	*	PV<15	A%	*	B%	*	C%	*	D%	*	Omit%	Ncount
1	ALL	B	0.49			11		**46**		24		17		3	6578
2	ALL	D	0.46			17		12		9		**59**		2	6560
3	ALL	B	0.40			16		**50**		16		12		**6**	6572
4	ALL	D	0.47			5		9		21		**63**		2	6605
5	ALL	C	0.40			3		19		**51**		26		2	6643
6	ALL	C	0.46			12		5		**78**		4		2	6614
7	ALL	A	0.30			**33**		36		15		13		3	6643
8	ALL	C	0.43			21		35		**35**		6		3	6646

All TE items and ELA SR items are put through an adjudication process. For each item, the frequency distribution of responses that are scored correctly is created, along with the frequency distribution of responses that are scored as incorrect. Assessment specialists review each response in the frequency reports and indicate whether the response should be scored as correct. The assessment specialists' indications are then cross-referenced with how the responses are scored to confirm that scoring is accurate. Figure 10.5 presents a sample adjudication spreadsheet.

Figure 10.5. Sample Adjudication Spreadsheet

Item ID	Func.	Item Response	Scored Response	Freq. Count	% of Total Freq.	Date 1st Reviewer	1st Reviewer Initials	Issue? (Y/N)	Description of Issue	Date 2nd Reviewer	2nd Reviewer Initials	Issue? (Y/N)	Description of Issue
Item1		A_A1:B_B2	2	28339	59								
Item1		A_A1	1	35	0								
Item1		A_A1:A_A2	1	3782	8								
Item1		A_A1:C_C2	1	4803	10								
Item1		A_A1:D_D2	1	970	2								
Item1			0	56	0								
Item1		A_A2	0	1	0								
Item1		B_B1	0	12	0								
Item1		B_B1:A_A2	0	464	1								
Item1		B_B1:B_B2	0	1038	2								
Item1		B_B1:C_C2	0	844	2								
Item1		B_B1:D_D2	0	405	1								
Item1		B_B2	0	4	0								

Item ID	Func.	Item Response	Scored Response	Freq. Count	% of Total Freq.	Date 1st Reviewer	1st Reviewer Initials	Issue? (Y/N)	Description of Issue	Date 2nd Reviewer	2nd Reviewer Initials	Issue? (Y/N)	Description of Issue
Item1		C_C1	0	10	0								
Item1		C_C1:A_A2	0	501	1								
Item1		C_C1:B_B2	0	841	2								
Item1		C_C1:C_C2	0	582	1								
Item1		C_C1:D_D2	0	1510	3								
Item1		C_C2	0	1	0								
Item1		D_D1	0	10	0								
Item1		D_D1:A_A2	0	652	1								

10.4. Checking Model Assumptions

It is important to evaluate how the IRT models applied for CMAS fit the data because reported scale scores are derived from theta estimated under the IRT models. Two major assumptions are investigated: unidimensionality and item fit.

10.4.1. Unidimensionality (Factor Analysis)

An assumption under the IRT 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). A factor analysis was performed on the item response data for the CMAS assessments to analyze the number of dimensions the assessments appear to be measuring. Given that unidimensional IRT models are used for calibration and scaling, it is important that there be evidence to support their use.

Appendix K presents the scree plots for the Spring 2023 administration. For most of the assessments, one factor explained most of the variance, which supports the use of a unidimensional IRT model, although the ELA/CSLA scree plots do suggest that Reading and Writing are distinct subscores. The loadings for Factor 2 for ELA were all much higher for the PCR trait items than any other items. This may indicate the influence of a writing construct that is separate from what is measured by the reading items.

10.4.2. Item Fit

Appendix M presents the item fit results. (The results are not included if a test is pre-equated.) Item fit refers to how well the data fit the IRT calibration model, and it is evaluated using Yen's (1981) Q_1statistic that allows for the evaluation of an item's IRT model fit to observed student performance. In the calculations of Q_1, the observed and expected (based on the model) frequencies were compared at 10 intervals, or deciles, along the scale. Yen's Q_1fit statistic was computed for each item using the following formula:

$$Q_{1_i} = \sum_{j=1}^{10} \frac{N_{ij}\left(O_{ij} - E_{ij}\right)^2}{E_{ij}(1 - E_{ij})}$$

where N_{ij} is the number of students in interval j for item I, and O_{ij} and E_{ij} are the observed and expected proportions of students in interval j for item i. The Q_1statistic was then transformed so that items with different degrees of freedom can have comparable fit statistics:

$$Z_{Q_{1_i}} = \frac{Q_{1_i} - df}{\sqrt{2df}}$$

where *df* is the degree of freedom for the statistic (*df* = 10–the number of parameters estimated; *df* = 7 for SR items in a 3PL model). If $Z_{Q_{1_i}}$ is greater than Z_{crit}, the item is flagged for poor model fit:

$$Z_{crit} = \frac{N_i * 4}{1500}$$

where N_i is the sample size.

10.5. Calibration

Calibration refers to the estimation of item parameters in the IRT framework, which places items and students on a common scale. To obtain item parameter estimates for CMAS ELA, the GPC model was applied to the items. IRTPRO was used for all calibrations, and all operational item parameters were estimated in a single calibration (i.e., concurrent calibration) for each assessment. For CSLA, the RPCM was applied to all items to obtain item parameter estimates. All operational items within a grade were also calibrated concurrently. Winsteps was used for all CSLA calibrations.

PCR items were calibrated at the (unweighted) trait score level rather than as aggregated scores. To account for potential local dependence between the two trait scores, the item response matrix was modified before operational calibrations. For each PCR item, one of the two trait scores for each student was randomly selected, and the non-selected trait score was then removed from the dataset and treated as missing for calibration. The resulting item response dataset, known as a "Moulder" matrix, contained roughly half as many observations for each PCR trait score as for the non-PCR items. However, the datasets still contained an adequate number of student responses to conduct the calibrations. Due to the small population of students taking the CSLA assessment, trait scores were not removed from the data when conducting calibrations for CSLA.

For each CMAS Science assessment, a concurrent calibration was conducted to obtain item parameters for all the operational items.

10.6. Equating

Equating is used to place new forms onto the operational base scale. Equating of the operational test forms involves adjusting for differences in the difficulty of forms, both within and across assessment administrations, to ensure that students taking one form of a test are neither advantaged nor disadvantaged when compared to students taking a different form. Each time a new form is constructed, equating is used to allow scores on the new form to be comparable to scores on the previous form. If the IRT models fit the data and the model assumptions are met, calibration of test items places both items and students on a scale that is independent of any sample of students up to a linear transformation. Equating is used to determine and apply a scale transformation that allows for meaningful comparisons of student performance across different forms or administrations of the test.

To maintain comparability with prior administrations, CMAS Mathematics and ELA item parameter estimates were equated to the established base scale used in 2017, and the CSLA item parameter estimates were equated to the Spring 2016 CSLA base scale. The 2023 scale was chosen to be the base scale for CMAS Science due to concerns about the validity of the 2022 administration since it was the first time students had seen items aligned to the new standards. The Spring 2022 items were equated to the 2023 scale so that all items from the bank would be available for future test construction.

10.6.1. Mathematics

10.6.1.1. Pre-Equating Design

The Spring 2023 CMAS Mathematics assessments were equated to the base scales using an item pre-equating design, meaning all items had already been administered, with item parameters already estimated and placed onto the base scale. Students were scored based on these previously banked item parameter estimates. All operational items on these forms had been previously calibrated and equated to the base scale.[6] The forms were subsequently scored using these existing item parameters rather than performing a new calibration and equating. To help ensure the stability of item parameter estimates across administration, items were positioned as closely as possible to their positions when they were calibrated. To ensure that the assumptions of pre-equating were met, a post-equating check was performed using anchor sets identified during test construction. The results of this check were compared with the pre-equated results during a post-equating check.

10.6.1.2. Post-Equating Check

Because pre-equating relies on stronger assumptions than post-equating, an additional post-equating analysis was conducted and compared with the pre-equated results for the CMAS mathematics assessments in 2022. (A post-equating check was not conducted in 2023 because the 2023 forms were identical.) Large discrepancies between the two could suggest that pre-equating assumptions have not been met. Conversely, similarity between pre- and post-equated item parameters suggests that the pre-equated item parameters are appropriate for students taking the current form. The post-equating check followed the same procedures as those of the other post-equated assessments, using an anchor set for each assessment that was identified during test construction and that met the operational anchor test specifications.

Results of the post-equating check from 2022 suggested that pre- and post-equated item parameters were quite similar. Appendix L compares the 2022 pre-equated and post-equated test characteristic curves (TCCs) for each assessment. The results show that post-equated scores would have been highly similar to the pre-equated results. The high degree of similarity across the entire scale score range for each grade suggests that pre-equating assumptions were met and that the pre-equated item parameters were appropriate for this administration.

[6] Please refer to previous versions of the CMAS technical reports for information on these calibration and equating procedures, located on the CDE website at https://www.cde.state.co.us/assessment/cmas_coalt_techreport.

10.6.2. ELA

The Spring 2023 CMAS ELA assessments were calibrated and post-equated to the base scale following the procedures described below. The ELA assessments have historically been post-equated. All post-equating analyses were conducted using a representative sample of students that was evaluated based on the following demographics to ensure that the expected population demographic distributions were met: gender, ethnicity/race, economic disadvantage, language proficiency, students with disabilities, and district setting.

A common items approach was used for equating the operational forms. Forms from adjacent administrations contain a set of items that are the same across the two administrations. This set of items represents the blueprint in terms of content and represents roughly 40% of a full form.

10.6.2.1. Consistency of Constructed-Response Scoring Check

Because the ELA assessments include a high percentage of CR items, the anchor sets include CR items to be more reflective of the construct being measured. For accurate equating, it is important that the items in the anchor sets be consistently scored across administrations. With SR items, scoring is the same each time the item is administered (e.g., 'A' is always scored as the correct answer) such that changes in item performance across administrations can be solely attributed to changes in student performance. With CR items, scoring is done by human raters, so it is important that scoring be monitored both within an administration and across administrations to maintain consistent scoring throughout. Such procedures were in place, including consistency in training and the use of validity papers throughout scoring.

As an additional check, the consistency of the CR scoring was examined prior to equating via the rescoring of a subset of the previous year's papers to remove any items that exhibited statistical drift in scoring characteristics so that the accuracy of the equating was not jeopardized. If a CR item appeared to lack consistency across the administrations, considerations were given to removing the item from the anchor set.

10.6.2.2. Stability Check

The item parameter stability check for the anchor items was conducted using classical item analyses, scatterplots of item parameter estimates, and ICC comparison. For the ICC comparison, old and new ICCs were compared using the *z*-score approach based on D^2 (Wells et al., 2014), as outlined below:

1. Obtain the theoretically weighted estimated posterior theta distribution using 31 quadrature points (-5 to 5).
2. Compute the slope and intercept constants using the Stocking and Lord (1993) method with all anchor items in the linking set.
3. Place the original anchor item parameter estimates onto the baseline scale by applying the constants obtained in Step 2.
4. For each anchor item, calculate D^2 between the ICCs based on old (x) and new (y) parameters at each point in this theta distribution:

$$D_i^2 = \sum^{k} \left[P_{ix}(\theta_k) - P_{iy}(\theta_k) \right]^2 \bullet g(\theta_k)$$

where i = item, x = old form, y = new form, k = theta quadrature point, and g = theoretically weighted posterior theta distribution.

5. Compute the mean and standard deviation of the D^2 values.
6. Flag the items with a D^2 more than 2 standard deviations above the mean.

10.6.2.3. Calibration and Anchor Set Evaluation

The initial calibration results were reviewed for problematic item parameter estimates, and fit plots were examined to detect items with poor model–data fit. Review of anchor item stability analyses resulted in dropping one to four items from the anchor set, depending on grade. The final anchor sets for ELA represented between 39% and 47% of the unweighted total test points. The online and paper versions were constructed to be parallel, and item parameter estimates were assumed to be the same. The information provided for the item statistics and IRT curves are based on the online estimates.

10.6.2.4. Final Anchor Sets

Items flagged from the stability check and consistency of CR scoring check were examined, and consideration was given to the impact of flagged item(s) on the content representativeness of the resulting anchor set. A flag alone was not the sole criteria for removing an item from the linking set; it was important to also make sure that the remaining anchor set continued to be representative of the overall content and structure of the test.

10.6.2.5. Equating Method

Using the item parameter estimates for the anchor set from the item bank and the current administration, the computer program STUIRT was used to obtain the transformation constants to place the current administration's items on the operational scale using the Stocking and Lord (1983) method. The scale transformation constants, Slope A and Intercept B, were applied to the item parameter estimates to place the new test items (new, N) on the operational scale (old, O) (Kolen & Brennan, 2004), as follows:

$$\alpha_{jO} = \alpha_{jN}/A$$

$$b_{jO} = A * b_{jN} + B$$

$$d_{jvO} = A * d_{jvN}$$

10.6.2.6. Paper Forms

Online and paper items were developed to be parallel to the online items. Operational paper items deemed identical to the operational online items were assumed to have the same item parameter estimates. Paper items were fixed to their online counterparts' item parameter estimates. This process produced item parameter estimates for all paper items.

10.6.3. CSLA

A common items approach was used to equate the CSLA operational forms. Forms from adjacent administrations contained a set of items that were the same across the two administrations (i.e., anchor items). Anchor items were operational items already equated to the base scale. The anchor items were placed in the same positions across all test forms within a grade and anchored the scale between the new test form and the base scale. This set of items represents the blueprint in terms of content and represents roughly 30% of a full form.

10.6.3.1. Stability Check

The stability check for the CSLA anchor items was conducted using classical item analysis, scatter plots of item difficulty, and displacement estimates from Winsteps. Items were flagged if the absolute value of the displacement estimate was greater than or equal to 0.30.

10.6.3.2. Final Anchor Sets

Items flagged from the stability check were examined, and consideration was given to the impact of flagged item(s) on the content representativeness of the resulting anchor set. A flag alone was not the sole criteria for removing an item from the linking set. It was important to also make sure that the remaining anchor set continues to be representative of the overall content and structure of the test. The final anchor sets for Grades 3 and 4 represented 42% and 36%, respectively, of the unweighted total test points.

10.6.3.3. Equating Method

To obtain equated Rasch parameter estimates for the Spring 2023 assessments, anchor item parameter estimates for each grade-level assessment were fixed to their previously equated item parameter estimates before calibrating the remaining non-anchor operational items on that assessment. This method placed the non-anchor operational items on the same scale as the anchor items.

10.6.4. Science

The Spring 2023 CMAS Science assessments were calibrated and the 2022 item parameters and cut scores were post-equated to the 2023 scale using a representative sample of students that was evaluated based on the following demographics to ensure that the expected population demographic distributions were met: gender, ethnicity/race, economic disadvantage, language proficiency, students with disabilities, and district setting. A common items approach was used for equating the operational forms. Forms from adjacent administrations contain a set of items that are the same across the two administrations. This set of items represents the blueprint in terms of content and represents roughly 30% of a full form.

10.6.4.1. Consistency of Constructed-Response Scoring Check

Because the Science assessments include a high percentage of CR items, the anchor sets include CR items to be more reflective of the construct being measured. For accurate equating, it is important that the items in the anchor sets be consistently scored across administrations. With SR items, scoring is the same each time the item is administered (e.g., ‘A’ is always scored as the correct answer) such that changes in item performance across administrations can be solely attributed to changes in student performance. With CR items, scoring is done by human raters, so it is important that scoring be monitored both within an administration and across administrations to maintain consistent scoring throughout. Such procedures were in place, including consistency in training and the use of validity papers throughout scoring.

As an additional check, the consistency of the CR scoring was examined prior to equating via the rescoring of a subset of the previous year’s papers to remove any items that exhibited statistical drift in scoring characteristics so that the accuracy of the equating was not jeopardized. If a CR item appeared to lack consistency across the administrations, considerations were given to removing the item from the anchor set.

10.6.4.2. Stability Check

The item parameter stability check for the anchor items was conducted using classical item analyses, scatterplots of item parameter estimates, and ICC comparison. For the ICC comparison, old and new ICCs were compared using the *z*-score approach based on D^2 (Wells et al., 2014), as outlined below:

1. Obtain the theoretically weighted estimated posterior theta distribution using 31 quadrature points (-5 to 5).
2. Compute the slope and intercept constants using the Stocking and Lord (1993) method with all anchor items in the linking set.
3. Place the original anchor item parameter estimates onto the baseline scale by applying the constants obtained in Step 2.
4. For each anchor item, calculate D^2 between the ICCs based on old (x) and new (y) parameters at each point in this theta distribution:

$$D_i^2 = \sum^{k} \left[P_{ix}(\theta_k) - P_{iy}(\theta_k) \right]^2 \bullet g(\theta_k)$$

   where i = item, x = old form, y = new form, k = theta quadrature point, and g = theoretically weighted posterior theta distribution.

5. Compute the mean and standard deviation of the D^2 values.
6. Flag the items with a D^2 more than 2 standard deviations above the mean.

10.6.4.3. Calibration and Anchor Set Evaluation

The initial calibration results were reviewed for problematic item parameter estimates, and fit plots were examined to detect items with poor model–data fit. Review of anchor item stability analyses resulted in dropping one to four items from the anchor set, depending on grade. The final anchor sets for Science represented between 25% and 45% of the unweighted total test points. The online and paper versions were constructed to be parallel, and item parameter estimates were assumed to be the same. The information provided for the item statistics and IRT curves are based on the online estimates.

10.6.4.4. Final Anchor Sets

Items flagged from the stability check and consistency of CR scoring check were examined, and consideration was given to the impact of flagged item(s) on the content representativeness of the resulting anchor set. A flag alone was not the sole criteria for removing an item from the linking set; it was important to also make sure that the remaining anchor set continued to be representative of the overall content and structure of the test.

10.6.4.5. Equating Method

Using the item parameter estimates for the anchor set from the item bank and the current administration, the computer program STUIRT was used to obtain the transformation constants to place the current administration's items on the operational scale using the Stocking and Lord (1983) method. The scale transformation constants, Slope A and Intercept B, were applied to the item parameter estimates to place the 2022 test items (new, N) on the operational (2023) scale (old, O) (Kolen & Brennan, 2004), as follows:

$$\alpha_{jO} = \alpha_{jN}/A$$

$$b_{jO} = A * b_{jN} + B$$

$$d_{jvO} = A * d_{jvN}$$

10.6.4.6. Paper Forms

Paper items were developed to be parallel to the online items. Operational paper items deemed identical to the operational online items were assumed to have the same item parameter estimates. Paper items were fixed to their online counterparts' item parameter estimates. This process produced item parameter estimates for all paper items.

10.7. Field Test Equating

The field test equating process is similar to that of operational equating, except that the anchor items are the operational items. This process places the field test item parameter estimates onto the operational base scale. All field test items are calibrated concurrently, with the exception of the ELA PCR items.

A minimum of 3,000 student responses for each field tested PCR item per trait is sampled for scoring and calibration. Due to possible dependency between the two trait scores for each PCR item, the field test items on each ELA assessment went through two calibrations. The first calibration included all field test items except the Writing Knowledge Language and Conventions (WKL) trait scores, and the second calibration included all field test items except the Writing Written Expression (WE) trait scores (with all operational items serving as anchor items in both cases).

The estimates from each calibration were then equated to the base scale separately following the same procedures as the operational equating. Finally, the two sets of equated field test parameters were combined by adding the equated field test WKL trait estimates to the equated estimates from the first calibration. This "double-calibration" method allowed for separate calibration of the field test trait scores while reducing the number of field test responses that needed to be scored per trait. Using a "Moulder" calibration method (as in the operational item calibration) would have meant using scoring resources to score traits that were never actually used for calibration or scoring.

10.8. Ability Estimates

10.8.1. Mathematics, ELA, and Science

Student ability was estimated using IRT pattern scoring based on student responses and the operational item parameter estimates for all students who met the relevant attemptedness criterion. Student ability was estimated at the overall test level, as well as for Reading on the ELA assessment. Estimates were obtained via the maximum likelihood method (MLE) applied within the ISE software program. Pattern scores use the student's individual item response pattern (overall or Reading claim) to determine their ability estimate, which may lead to different ability estimates for the same raw score.

10.8.2. CSLA

After the item parameter estimates were obtained for the CSLA operational items, student abilities were estimated for each grade-level assessment by conducting an anchored calibration of the operational items' item parameter estimates. Student abilities were calculated for the overall test and for Reading. To obtain student ability estimates for the overall test, all the operational items were included in the anchored calibration. To obtain student ability estimates for Reading, only the operational items representing the specific claim were included in the anchored calibration. The calibrations included the weighting of the PCR WE trait score. Student ability estimates were obtained via the joint maximum likelihood method (JMLE) applied within Winsteps.

10.9. Overall and Subscale Scale Scores

For CMAS Mathematics and ELA/CSLA, student ability estimates for the overall test were transformed to scale scores ranging from 650 to 850 using the same scaling transformations as the prior year's administrations. For CMAS Science, the student ability estimates for the overall test were transformed to scale scores ranging from 650 to 850 using the scaling transformations determined from standard setting. The student ability estimates for the subscores for CMAS Science were transformed to scale scores ranging from 400 to 600. For ELA/CSLA, the student ability estimates for Reading were transformed to scale scores ranging from 110 to 190. The following linear transformation was used to convert examinee theta estimates into scale scores where *A* and *B* are unique scaling constants for each subject/grade:

$$SS = A^{*}\theta + B$$

After the scale scores were calculated, the lowest obtainable scale score (LOSS) and highest obtainable scale score (HOSS) were applied. LOSS and HOSS were set to 650 and 850, respectively, for the overall test scale. For the Reading scale, LOSS and HOSS were set to 110 and 190. For the subscores of CMAS Science, LOSS and HOSS were set to 400 and 600.

10.10. Item-Level IRT Statistics

Appendix M presents the item parameter estimates for each grade. (The results are not included if a test is pre-equated.) The item numbers are merely identifiers and do not reflect the sequence of items as they were presented to students. The "Item Type" uses the coding of SR for selected-response, XI for technology-enhanced, and CR for constructed-response items. The "Model" refers to the IRT model under which the item was estimated (2PL, 3PL, GPC, or RPCM). The "A" column shows the item parameter estimate for discrimination, "B" for difficulty, and "D1" through "D7" for GPC or RPCM category threshold estimates. Not all item parameters apply to each item. For example, there are no category threshold estimates for 2PL items.

The last column of the ELA and science tables reflects whether an item was flagged for misfit based on Q1 for those calibrated assessments. Several items in each grade were flagged for misfit. Misfit plots for all items were reviewed, and misfit statistics were compared with data from the previous administration. Based on these reviews, no additional items were removed due to misfit flags. The last two columns for CSLA reflect the infit and outfit statistics generated from Winsteps. Fit values were reviewed, and no items were removed due to misfit.

10.11. IRT Curves

Appendix N presents the test characteristic curves (TCCs), test information curves (TICs), and CSEM curves for both the overall scale scores and the Reading scale scores. The curves for CMAS Mathematics are from 2022 because the forms were reused for 2023 and item parameters were held the same. The 2023 CMAS ELA TCCs matched those from 2018 in terms of shape and position. The 2023 TCCs were reviewed across the distribution and at the cuts to ensure the match between years. Colorado's established maximum TCC difference of 0.05 was also maintained between the 2018 and 2023 forms. The TCCs are provided in terms of expected percent correct rather than expected raw score. Along with the curves, each cut score for a given grade is indicated with a red vertical line, as are the cut scores for Reading. On the overall scale score TCCs for science, mathematics, and ELA, the vertical line at a scale score of 750 corresponds to the cut for *Met Expectations* for each assessment.

10.12. Comparability of Online and Paper Forms

The scale score distributions for students taking the online and paper CMAS Mathematics, Science, and ELA assessments were examined using a matched samples approach to investigate the extent to which the online and paper forms produced comparable scores. Multiple variables were used for determining the matched groups to result in "equal" groups of online and paper students. The matching variables included sex, race/ethnicity, free and reduced lunch status, language proficiency, IEP, and district setting, plus the prior year's overall test score.

Because science is not assessed in consecutive grade levels, the prior year's score did not come from science. Rather, the Grade 4 mathematics score was used for Grade 5 science and the Grade 7 mathematics score was used for the Grade 8 science assessment. There were an insufficient (<1,000) number of students who took the Grade 11 CMAS Science assessment on paper to complete a comparability and mode analysis.

Scale score distributions of CMAS scores between the matched samples were compared to estimate the mode effect. To quantify the differences between the two distributions, the effect size of the differences between the two distributions was calculated as Cohen's *d* (Cohen, 1977) using the mean scale score from each group and the pooled standard deviation:

$$d = \frac{M_{group1} - M_{group2}}{SD_{pooled}}$$

Suggested interpretations of Cohen's *d* are as follows:

- 0.2 = a small effect size
- 0.5 = a medium effect size
- 0.8 = a large effect size

A threshold for a possible mode effect was set to an effect size of 0.1 or greater and a matched sample size of at least 1,000 students. The effect size was calculated for the mathematics and ELA assessments in each grade, and for science in Grades 5 and 8. The results were presented to CDE who made the final decision on whether to make an adjustment for mode differences for each assessment. Table 10.1 presents the mode effect sizes from the Spring 2023 administration.

Based on evaluation of the effect sizes, mode adjustments were made for ELA Grades 3 and 5–7 and mathematics Grades 7 and 8. For assessments where an adjustment was deemed necessary, scores from the paper form were adjusted using a linear transformation to match the mean and standard deviation of the online form. The conversion was applied to the overall scores. For ELA, the conversion was also applied to the Reading score. For the paper-based mathematics assessments from the prior administration, mode adjustments from that prior administration were applied to those forms.

Table 10.1. Online vs. Paper Comparability Mode Effect Sizes

	Mathematics		ELA		Science	
Grade	N	Effect Size	N	Effect Size	N	Effect Size
3	4,484	0.00	4,186	-0.15	–	–
4	3,316	0.02	3,478	-0.01	–	–
5	3,185	0.06	3,255	0.10	1,788	0.04
6	3,000	-0.08	2,751	-0.19	–	–
7	2,873	-0.10	2,481	-0.10	–	–
8	2,691	-0.12	2,270	-0.08	1,454	-0.02
11	–	–	–	–	N/A	N/A

Note. N/A = not applicable. Comparability analyses were not conducted for science Grade 11 because the n-count was less than 1,000.

Chapter 11: 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; 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 other words, 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. Reliability for the CMAS assessments is evaluated with the following analyses:

- Internal consistency (coefficient alpha)
- Standard error of measurement (SEM)
- Conditional standard error of measurement (CSEM)
- Decision consistent and accuracy
- Inter-rater agreement

11.1. Internal Consistency (Coefficient Alpha)

Within the framework of classical test theory, an observed test score is defined as the sum of a student's true score and error ($X = T + E$, where X = the observed score, T = the true score, and E = error). A true score is considered the student's true standing on the measure, while the error score reflects a random error component. Thus, error is the discrepancy between a student's observed and true score. Internal consistency is typically measured via correlations among the items on an assessment and provides an indication of how much the items measure the same general construct. High reliability of test scores implies that the test items within a subclaim are measuring a single construct, which is a necessary condition for validity when the intention is to measure a single construct.

The reliability coefficient of a measure is the proportion of variance in observed scores accounted for by the variance in true scores. The coefficient can be interpreted as the degree to which scores remain consistent over parallel forms of an assessment (Ferguson & Takane, 1989; Crocker & Algina, 1986). In the internal consistency method used to estimate reliability for the CMAS assessments, a single form is administered to the same group of students to determine whether students respond consistently across the items within a test. A basic estimate of internal consistency reliability is Cronbach's coefficient alpha statistic (Cronbach, 1951). Coefficient alpha is equivalent to the average split-half correlation based on all possible divisions of a test into two halves. Coefficient alpha can be used on any combination of dichotomous and polytomous test items and is computed as follows:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{j=1}^{n} S_j^2}{S_X^2}\right)$$

where n is the number of items, S_j^2 is the variance of students' scores on item j, and S_X^2 is the variance of the total-test scores.

Coefficient alpha ranges from 0.0 to 1.0, where higher values indicate a greater proportion of observed score variance. Two factors affect estimates of internal consistency: test length and homogeneity of items. The longer the test, the more observed score variance is likely to be true score variance. The more similar the items, the more likely students will respond consistently across items within the test.

Table 11.1 – Table 11.4 present the coefficient alpha results overall and by subclaim for each content area. Appendix H presents the coefficient alpha estimates by demographic subgroup. The internal consistency values for the overall test ranged from 0.89 to 0.93. Given the differences in length, it is expected that the coefficient alpha for the overall test will be higher than that of the subscales.

The overall test reliability does not correspond directly with the overall student scale scores, as those are based on IRT pattern scoring. However, the overall estimates do provide evidence of unidimensionality of the assessments. Furthermore, the subgroup reliabilities were consistent for the various demographic subgroups, except for those based on language proficiency. The reliability of the tests tended to be lower for students identified as non-English proficient or limited English proficient.

Table 11.1. Coefficient Alpha by Subclaim—Mathematics

Grade	Overall	Subclaim A	Subclaim B	Subclaim C	Subclaim D
3	0.92	0.87	0.79	0.67	0.67
4	0.91	0.88	0.63	0.73	0.51
5	0.93	0.88	0.72	0.74	0.66
6	0.91	0.84	0.63	0.75	0.63
7	0.90	0.82	0.57	0.76	0.56
8	0.91	0.85	0.43	0.78	0.68

Table 11.2. Coefficient Alpha by Subclaim—ELA

Grade	Overall	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Reading	Writing
3	0.90	0.80	0.76	0.68	0.64	0.75	0.90	0.67
4	0.89	0.72	0.72	0.64	0.48	0.75	0.87	0.74
5	0.88	0.78	0.69	0.44	0.54	0.75	0.86	0.75
6	0.90	0.80	0.70	0.55	0.81	0.80	0.87	0.80
7	0.90	0.78	0.73	0.66	0.81	0.83	0.89	0.82
8	0.90	0.78	0.79	0.60	0.72	0.85	0.90	0.84

Table 11.3. Coefficient Alpha by Subclaim—CSLA

Grade	Overall	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Reading	Writing
3	0.88	0.75	0.61	0.73	0.67	0.78	0.86	0.81
4	0.86	0.75	0.57	0.63	0.58	0.74	0.84	0.76

Table 11.4. Coefficient Alpha by Subclaim—Science

Grade	Overall	Physical Science	Life Science	Earth and Space Science
5	0.89	0.79	0.60	0.76
8	0.92	0.81	0.82	0.75
11	0.90	0.77	0.74	0.72

Note. For Grade 5, the subclaim is Physical Science/Life Science.

11.2. Standard Error of Measurement (SEM)

The SEM is another measure of reliability. This statistic uses the standard deviation of test scores along with a reliability coefficient (e.g., coefficient alpha) to estimate the number of score points that a student's test score would be expected to vary if the student was tested multiple times with equivalent forms of the assessment. It is calculated as follows:

$$SEM = s_x\sqrt{1-p_{XX}}$$

where s_x is the standard deviation of test scores, and p_{XX} is the reliability coefficient.

There is an inverse relationship between the reliability coefficient and SEM: the higher the reliability, the lower the SEM. Table 11.5 – Table 11.8 present the SEM results by subclaim for each content area. The classical SEM estimate is not reported for the overall test scale scores and the Reading subscore, as those scores are based on IRT pattern scoring rather than the sum of item scores.

Table 11.5. SEM by Subclaim—Mathematics

Grade	Subclaim A	Subclaim B	Subclaim C	Subclaim D
3	1.91	1.15	1.53	1.32
4	2.15	1.10	1.70	1.87
5	2.08	1.23	1.37	1.47
6	1.93	1.48	1.53	1.46
7	1.98	1.23	1.36	1.45
8	2.04	1.42	1.12	1.34

Table 11.6. SEM by Subclaim—ELA

Grade	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Writing
3	1.87	1.82	1.61	1.99	0.54	2.33
4	2.60	2.28	1.57	2.34	0.49	2.07
5	2.12	1.87	1.51	2.56	0.57	2.38
6	2.13	2.86	1.57	1.71	1.01	2.41
7	1.91	2.39	1.63	2.47	0.76	3.12
8	2.04	2.30	1.70	2.85	0.65	2.79

Table 11.7. SEM by Subclaim—CSLA

Grade	Reading: Literary Text	Reading: Informational Text	Reading: Vocabulary	Writing: Written Expression	Writing: Knowledge and Use of Language Conventions	Writing
3	2.01	1.77	1.51	0.89	0.76	1.25
4	2.21	2.00	1.40	1.16	0.93	1.56

Table 11.8. SEM by Subclaim—Science

Grade	Physical Science	Life Science	Earth and Space Science
5	1.95	1.67	2.08
8	1.93	2.03	1.84
11	1.90	1.70	1.67

Note. For Grade 5, the subclaim is Physical Science/Life Science.

11.3. Conditional Standard Error of Measurement (CSEM)

While the SEM provides an estimate of precision for an assessment, conditional standard error of measurement (CSEM) gives an indication of how measurement error varies across the score scale. While coefficient alpha is reported as a measure of internal consistency of the items that each scale comprises, IRT-based CSEM is a more appropriate measure of the measurement error associated with these scale scores because the reported scale scores for both the overall test and Reading are determined using IRT pattern scoring.

The CSEM is defined as the standard deviation of observed scores given a particular true score and is estimated within the IRT framework as the inverse of the test information function. Plots of test information curves (TICs) and CSEM across the score scale range are provided in Appendix N for both the overall scale scores and Reading scores.

Each scale score has a CSEM estimate that indicates what the most likely range of scores would be for students receiving that score if they tested multiple times. The CMAS assessments measure more accurately at a scale score near the middle of the scale than at the ends of the scale. During test construction, CSEMs are reviewed to ensure that they are minimized around the performance level cut scores.

11.4. Decision Consistency and Accuracy

The CMAS Mathematics and ELA/CSLA scales are divided into five performance levels that a student is placed in based on their scale score: *Did Not Yet Meet Expectations*, *Partially Met Expectations*, *Approached Expectations*, *Met Expectations*, and *Exceeded Expectations*. The consistency of a decision refers to the extent to which the same classification would result if a student were to take two parallel forms of the same assessment. However, since test-retest data are not available, psychometric models can be used to estimate the decision consistency based on test scores from a single administration. The accuracy of a decision refers to the agreement between a student's observed score classification and a student's true score classification if a student's true score could be known.

Procedures developed by Livingston and Lewis (1995) were used to estimate the consistency and accuracy of performance level classifications. For the overall test, consistency and accuracy estimates, along with PChance (i.e., the probability of a consistent classification due to chance) and Cohen's Kappa (κ) coefficient (Cohen, 1960), are calculated as follows:

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

where P is the probability of consistent classification, and P_c is the probability of consistent classification by chance (Lee et al., 2000).

Table 11.9 presents the kappa interpretations. Table 11.10 presents the decision consistency and accuracy results, and Table 11.11 and Table 11.12 present the consistency and accuracy estimates at each cut score.

Table 11.9. Kappa Values

Value of Kappa	Strength of Agreement
< 0.20	Poor
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Good
0.81 – 1.00	Very Good

Table 11.10. Decision Consistency and Accuracy Estimates

Content Area	Grade	Accuracy	Consistency	PChance	Kappa
Mathematics	3	0.73	0.63	0.22	0.52
	4	0.75	0.65	0.24	0.54
	5	0.76	0.67	0.23	0.57
	6	0.74	0.64	0.23	0.53
	7	0.75	0.65	0.26	0.53
	8	0.72	0.62	0.23	0.51
ELA	3	0.72	0.63	0.25	0.50
	4	0.70	0.59	0.24	0.47
	5	0.72	0.62	0.28	0.48
	6	0.73	0.63	0.25	0.51
	7	0.70	0.59	0.22	0.48
	8	0.70	0.60	0.23	0.48
Science	5	0.77	0.69	0.32	0.54
	8	0.82	0.75	0.34	0.62
	11	0.76	0.69	0.34	0.53
CSLA	3	0.72	0.62	0.26	0.49
	4	0.72	0.62	0.29	0.46

Table 11.11. Accuracy of Cut Scores

Content Area	Grade	*Partially Met Expectations* Cut	*Approached Expectations* Cut	*Met Expectations* Cut	*Exceeded Expectations* Cut
Mathematics	3	0.95	0.92	0.91	0.89
	4	0.94	0.91	0.91	0.96
	5	0.95	0.92	0.92	0.91
	6	0.94	0.91	0.92	0.95
	7	0.95	0.90	0.92	0.97
	8	0.92	0.90	0.92	0.93
ELA	3	0.93	0.91	0.90	0.94
	4	0.95	0.91	0.89	0.89
	5	0.96	0.91	0.89	0.92
	6	0.96	0.91	0.90	0.91
	7	0.95	0.91	0.90	0.83
	8	0.94	0.91	0.90	0.88
Science	5	–	0.92	0.87	0.97
	8	–	0.93	0.90	0.99
	11	–	0.92	0.85	0.99
CSLA	3	0.93	0.89	0.92	0.98
	4	0.94	0.88	0.92	0.98

Table 11.12. Consistency of Cut Scores

Content Area	Grade	*Partially Met Expectations* Cut	*Approached Expectations* Cut	*Met Expectations* Cut	*Exceeded Expectations* Cut
Mathematics	3	0.92	0.89	0.88	0.87
	4	0.92	0.87	0.88	0.96
	5	0.93	0.89	0.89	0.90
	6	0.91	0.87	0.88	0.94
	7	0.92	0.86	0.89	0.96
	8	0.89	0.87	0.88	0.92
ELA	3	0.90	0.87	0.87	0.93
	4	0.93	0.88	0.85	0.87
	5	0.94	0.87	0.85	0.91
	6	0.94	0.88	0.86	0.89
	7	0.93	0.88	0.86	0.81
	8	0.91	0.87	0.86	0.86
Science	5	–	0.89	0.82	0.97
	8	–	0.90	0.85	0.99
	11	–	0.89	0.80	0.99
CSLA	3	0.90	0.85	0.88	0.97
	4	0.91	0.83	0.89	0.97

11.5. Inter-Rater Agreement

For CR items, inter-rater agreement examines the extent to which students would obtain the same score if scored by different scorers. For each operational item, 10% of the responses were scored by a second reader, which allowed for rater agreement statistics to be calculated. 0 presents the inter-rater agreement statistics for the CR operational items (i.e., the percentage of operational items with exact agreement, adjacent agreement, and non-adjacent agreement). The target exact plus adjacent agreement rate is 95% for all items. The following agreement rates were calculated for each CR item:

- Exact agreement, which represents exact agreement between the two raters
- Adjacent agreement, which represents adjacent agreement between the two raters (i.e., a difference of 1 score points)
- Non-adjacent agreement, which represents a difference of more than 1 score point between the two raters

For the PCR items, the following additional analyses were also conducted:

- Quadratic kappa (Kappa), $KAPPA = \frac{E([X_1 - Y_1]^2)}{E([X_1 - Y_2]^2)}$, which is a comparison between the mean square error of rating pairs that are supposed to agree (X_1, Y_1) and those that are unrelated (X_1, Y_2)
- Standardized mean differences (MD): $\bar{Z} = \frac{|\bar{X}_{R_1} - \bar{X}_{R_2}|}{\sqrt{\frac{sd^2_{R_1} + sd^2_{R_2}}{2}}}$
- Correlations (CORR): $\bar{Z} = \frac{|\bar{X}_{R_1} - \bar{X}_{R_2}|}{\sqrt{\frac{sd^2_{R_1} + sd^2_{R_2}}{2}}}$

Chapter 12: Validity

"Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (AERA et al., 2014). As such, it is not the CMAS assessments that are validated but rather the interpretations of the scores. The purpose of the CMAS assessments is to provide information about a student's level of mastery of the Colorado Academic Standards (CAS). Mastery of the standards in the elementary and middle school grades indicates that a student is on track to being college and career ready at each grade level. In support of these ends, this technical report has described processes that were implemented throughout the CMAS assessment cycle with validity and fairness considerations in mind. This chapter describes the various sources of validity evidence for CMAS as outlined in the *Standards for Educational and Psychological Testing* (AERA et al., 2014), often referencing other chapters and sections of this report.

12.1. Evidence Based on Test Content

Evidence based on the content of the assessment is supported by the degree of correspondence between test items and content standards. The degree to which the test measures what it claims to measure is known as construct validity. The CMAS assessments adhere to the principles of evidence-centered design, in which the standards to be measured (i.e., the CAS) are identified, and the performance a student needs to achieve to meet those standards is delineated in the evidence statements (ESs) or evidence outcomes (EOs). Test items are reviewed for adherence to universal design principles to maximize the participation of the widest possible range of students.

12.1.1. Test Development Process

The item development process is driven by targets at the ES or EO level. Before developing items, Pearson uses target spreadsheets to create an internal item development plan (IDP) aligned with the expectations of test design and with consideration of attrition rates at committee review and data review. The validity of a state assessment relies on the methodology that frames the development and design of the assessment. In support of that claim, Pearson upholds these considerations as the cornerstones of the CMAS item and test development:

- The item development process ensures that the mathematics and ELA/CSLA items align to the ESs and EOs and that the science items align to the EOs.
- IDPs were designed to produce and maintain a robust item bank; items are written to address the scope of measured standards, grade-level difficulties, and cognitive complexity.
- The item and test development processes promote the equivalency of the online and paper-based assessments.
- Items were developed with the intention of being administered on multiple testing platforms.
- Item and test development processes are compliant with industry standards.

Content is also aligned through the articulation of performance in the performance level descriptors (PLDs). At the policy level, the PLDs include policy claims about the educational achievement of students who attain a particular performance level, and a broad description of the grade-level knowledge, skills, and practices that students performing at a particular achievement level are able to demonstrate. Those policy-level descriptors are the foundation for the subject- and grade-specific PLDs, which, along with the ES or EO framework, guide the development of the items and tasks.

Gathering construct validity evidence for the CMAS assessments is embedded in the process by which the test content is developed and validated. At each step in the test development process, educators, assessment experts, and bias and sensitivity experts were involved in review of text, items, and tasks for accuracy, appropriateness, and freedom from bias, as described in Chapter 3: Item Development. In the early stages of development, Pearson conducted research studies to validate the item and task development approach. One such study focused on student task interaction and was designed to collect data on students' experience with the assessment tasks and technological functionalities, as well as the amount of time needed to answer each task. Pearson also conducted a rubric choice study that compared the functioning of two rubrics developed to score the ELA PCR tasks. Quantitative and qualitative evidence was collected to support the use of a condensed or expanded trait scoring rubric.

An important consideration when constructing test forms is recognition of items that may introduce construct-irrelevant variance. Such items should not be included on test forms to help ensure fairness to all student subgroups. Data reviews and content and bias reviews are held with Colorado educators to identify any issues with items before they are included on an operational test form. Accommodations were also made available based on individual need documented in the student's approved IEP or 504 Plan, as described in Section 5.5.

The CMAS operational test forms were carefully constructed to align with the test blueprints and specifications based on the CAS. Chapter 4: Test Construction provides details on the construction of the operational assessment forms, which demonstrates that all test forms for mathematics and ELA adhered to the same test design used in previous years or were previously used operationally. Science was a new assessment administered in Spring 2022; as such, validity evidence based on test content for science will continue to be documented in future iterations of this technical report as the test continues to be developed.

12.1.2. Alignment Study

An independent alignment study was conducted by the Human Resources Research Organization (HumRRO) in 2023 to provide further evidence to support the claim that the content of the CMAS Science test items matches the intended content as specified in the 2020 CAS (Revivo et al., 2023). For the study, three panels (one per grade) of Colorado educators were convened to review the alignment between the CMAS Science items and the CAS. Every effort was made to recruit panels consisting of teachers reflecting the various demographic subgroups and regions across Colorado. HumRRO applied alignment criteria they developed that was approved by CDE. This procedure required the panelists to (a) provide cognitive complexity ratings for each item, (b) indicate the CAS best aligned to each item, and (c) indicate if each item aligned to an SEP or CCC.

Overall, the results of the study provide validity evidence to support the claim that the content of the CMAS Science test items matches the intended content as specified in the CAS and test blueprint. Across all grades, items represented the intended content and reflected the multidimensional nature of the CAS, although only Grade 5 items reflected appropriate levels of cognitive complexity whereas Grades 8 and 11 items narrowly missed the cognitive complexity criterion requirements. The results of the alignment study have been considered during the item development process for subsequent administrations.

12.2. Evidence Based on Internal Structure

Analyses of the internal structure of a test typically involve studies of the relationships among test items and/or test components (i.e., subclaims) in the interest of establishing the degree to which the items or components appear to reflect the construct on which a test score interpretation is based (AERA et al., 2014, p. 16). The term *construct* refers to the characteristics that a test is intended to measure; in the case of the CMAS assessments, the characteristics of interest are the knowledge and skills defined by the test blueprints.

The CMAS assessments provide a full summative test score and a Reading score, as well as percent of points earned scores for Writing and mathematics, ELA, and science subclaims. The goal of reporting at this level is to provide criterion-referenced data to assess the strengths and weaknesses of a student's achievement in specific components of each content area compared with other students taking the same assessment (for overall and subclaim scores) and students who took the assessment in prior years (for overall scores). This information can then be used for a variety of purposes as indicated in Section 1.4. Evidence based on internal structure is provided in the following sections of this technical report:

- Subclaim correlations (Section 9.4)
- Internal consistency (Section 11.1)
- Factor analysis (Section 10.4.1)

12.3. Evidence Based on Relationships to Other Variables

Correlations were calculated between the mathematics, ELA, and science assessments, as shown in Table 12.1. (The samples include only students with valid scores on both assessments.) These scores may be expected to have lower correlations if the tests are measuring distinct constructs. The correlations between the scale scores of the CMAS assessments ranged from 0.75 to 0.82; these values are also very close to the 2018 values.

Table 12.1. Correlations Between CMAS Scale Scores

	ELA & Mathematics		ELA & Science		Mathematics & Science	
Grade	N	Correlation	N	Correlation	N	Correlation
3	55,537	0.78	–	–	–	–
4	55,290	0.77	–	–	–	–
5	56,377	0.75	55,619	0.81	55,771	0.78
6	55,413	0.77	–	–	–	–
7	53,606	0.77	–	–	–	–
8	51,376	0.77	50,050	0.82	50,236	0.78

12.4. Evidence Based on Response Processes

As noted in the *Standards for Educational and Psychological Testing* (AERA et al., 2014), additional support for a particular score interpretation or use can be provided by theoretical and empirical evidence indicating that students are using the intended response processes when responding to the items in a test. This type of evidence may be gathered from interacting with students to understand what processes underlie their item responses. Evidence may also be derived from feedback provided by test proctors/teachers involved in the administration of the test and raters involved in the scoring of CR items. Evidence may also be gathered by evaluating the correct and incorrect responses to short CR items (e.g., items requiring a few words to respond) or by evaluating the response patterns to multi-part items.

Prior to the 2016 administration, the PARCC consortium undertook research investigating the quality of the items, tasks, and stimuli, focusing on whether students interact with the online items/tasks as intended through cognitive labs. In these studies, students were asked to narrate how they interact with an item and answer questions about their experience with the item and online platform.

Cognitive labs were conducted for CMAS Science with Colorado students in May 2013. Students attempted a variety of item types on the TestNav platform and were asked to "think-aloud" as they worked through each item. Students showed a high degree of facility in responding to the items, and only a small bit of supplemental training was speculated to be needed to acquaint them with the tools and navigation of the TestNav interface. Surveys were given to the students after completion of the assessment, which included a question that asked them to indicate whether they preferred paper or computer-based tests. Most students indicated that they preferred the computer-based version, and many commented that it had been an enjoyable experience. For a full report on the cognitive labs, see the *2013–2014 CMAS Technical Report*.

As new items are developed, the field test responses are reviewed. Sample responses to the CR items are also reviewed by educator committees during rangefinding to ensure that the rubrics make sense and provide example scored responses. During the data review meeting, item statistics are reviewed to ensure that the students are responding to items in the expected way. Low item item-total correlations and aberrant response distributions can all indicate that there are unexpected issues with either the correct or incorrect responses. Items where the correct response is not accurate or there are distractor responses that are technically correct can be identified and rejected at this step. During the adjudication step, incorrect responses to fill-in-the-blank items are also reviewed to make sure that no technically correct responses are excluded. These include entry issues such as extra spaces or unexpected responses such as adding an unnecessary decimal (e.g., 3.0 rather than 3).

12.5. Evidence Based on the Consequences of Testing

Because state tests are administered "with the expectation that some benefit will be realized from the intended use of the scores" (AERA et al., 2014), validity evidence supporting the use and interpretation of CMAS results may be investigated as a consequence of testing. One intended consequence of testing is that more students will demonstrate mastery over the CAS over time, as evidenced by more students achieving in the top performance levels, if the data are used appropriately to make improvements in programming at the school and district levels.

Table 12.2 presents the percentage of students who have reached proficiency on the CMAS assessments over the years. The CMAS Mathematics and ELA assessments have been administered to Colorado students since Spring 2015, and CSLA has been administered since Spring 2016. While CMAS Science has been administered since Spring 2014, it is not included in the table because a new CMAS science assessment based on new standards was administered for the first time in Spring 2022.

As shown in the table, student performance has improved since the first administration with the exception of Grade 6 mathematics and Grade 3 CSLA. The decrease in 2021 of the percent of students meeting or exceeding in the required grades was expected given the learning disruptions caused by COVID-19 in 2020 and 2021. The pandemic's continued impact on student learning opportunities should be considered when interpreting the 2022 performance results. There have also been changes in the available assessments by grade for Grades 7 and 8 mathematics across administrations, so comparisons across years for those grades are not included.

Table 12.2. Student Performance Over Time

Content Area	Grade	First Administration *%Met or Exceeded*	2019 *%Met or Exceeded*	2021 *%Met or Exceeded*	2022 *%Met or Exceeded*	2023 *%Met or Exceeded*	% Change, First Administration to 2019	% Change, First Administration to 2021	% Change, First Administration to 2022	% Change, First Administration to 2023
Mathematics	3	36.7	41.0	–	39.4	40.4	4.3	–	2.7	3.7
	4	30.2	33.6	28.5	30.7	32.7	3.4	-1.7	0.5	2.5
	5	30.1	35.7	–	34.9	36.5	5.6	–	4.8	6.4
	6	31.7	29.5	24.1	26.3	28.2	-2.2	-7.6	-5.4	-3.5
ELA	3	38.2	41.3	39.1	40.7	39.9	3.1	0.9	2.5	1.7
	4	41.7	48.0	–	44.1	43.8	6.3	–	2.4	2.1
	5	40.5	48.4	47.2	45.4	47.8	7.9	6.7	4.9	7.3
	6	39.1	43.6	–	43.0	43.4	4.5	–	3.9	4.3
	7	41.0	46.5	42.6	41.8	45.0	5.5	1.6	0.8	4.0
	8	40.9	46.9	–	43.9	42.4	5.9	–	3.0	1.5
CSLA	3	22.0	27.5	15.4	19.8	18.7	5.5	-6.6	-2.2	-3.3
	4	13.9	19.1	–	13.7	14.2	5.2	–	-0.2	0.3
Science	5	33.9	–	–	–	33.9	–	–	–	–
	8	31.3	–	–	–	31.3	–	–	–	–
	11	24.6	–	–	–	24.6	–	–	–	–

Note. The first administration for mathematics and ELA was Spring 2015, the first administration for which scale scores and performance levels were generated for science was Spring 2023, and the first administration for CSLA was Spring 2016. Performance results are not included for the Spring 2021 mathematics, ELA, and CSLA opt-in grades.

12.6. Fairness

Fairness is an important aspect of validity, as it is critical that an assessment provide accurate measurements for **all** students. To that end, the following fairness considerations were woven into the development and administration of the CMAS assessments:

- Sample items that provide the opportunity for teachers and students to become familiar with the test design and scoring of the assessments before experiencing the items on an operational test (Section 5.3)
- Universal design principles that are adhered to during the test development process with the goal of avoiding construct-irrelevant aspects of the assessment that could impact student performance (Chapter 3: Item Development)
- Items are reviewed by educators for potential issues which could impact the performance of student groups prior to field testing (Chapter 3: Item Development).
- Differential item functioning (DIF) analyses to identify any items that appear to be unfairly favoring one subgroup over another. All items which show DIF are reviewed by educators for potential bias in the item. (Section 3.7)
- Accessibility tools and accommodations to allow students to fully demonstrate their content knowledge without being hindered by non-construct related elements (Sections 4.2 and 5.5)

Participation information must also be reviewed and taken into consideration thoughtfully when interpreting the district and school results. As participation rates vary across student, school, and district groups, challenges with interpreting results increase. Depending on the specific school or district, some student groups may have been overrepresented in the results and others may have been underrepresented. Students may have also experienced ongoing reduced, disrupted, and/or adjusted learning opportunities during the school year. Due to these factors and many more challenges experienced due to COVID-19, districts and schools should be cautious when interpreting results because the data may not support all cross-state comparisons and historical uses when participation rates are low and/or representativeness is limited.

References

Achieve. (2019). *A framework to evaluate cognitive complexity in science assessments*. https://files.eric.ed.gov/fulltext/ED603610.pdf

Allen, N. L., Carlson, J. E., & Zelenak, C. A. (1999). *The NAEP 1996 technical report (NCES 1999–452)*. National Center for Education Statistics, US Department of Education.

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

Beimers, J. N., Way, W. D., McClarty, K. L., & Miles, J. A. (2012, January). Evidence based standard setting: Establishing cut scores by integrating research evidence with expert content judgments. Pearson.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 392–479). Addison-Wesley.

Chien, M., & Shin, D. (2012). *IRT score estimation program*, V1.3 [computer program]. Pearson.

Cizek, G. J., Bunch, M. B., & Koons, H. (2004). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, *23*(4), 31–31.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–47.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Academic Press.

Colorado Department of Education (CDE). (2016). *Colorado Spanish Language Arts (CSLA) standard setting report*. CDE and Pearson.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich College Publishers.

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

Davis, L. L., & Moyer, E. L. (2015). *Performance level setting technical report*. https://parcc-assessment.org/wp-content/uploads/2017/12/PARCC_PLS_TechReport_011316_toPARCC-final.pdf

Dorans, N. J., & Holland, P. W. (1992). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning: Theory and practice* (pp. 35–66). Erlbaum.

Ferguson, G. A., & Takane, Y. (1989). *Statistical analysis in psychology and education* (6th ed.). McGraw-Hill.

Ferrara, S., Perie, M., & Johnson, E. (2008). Matching the judgmental task with standard setting panelist expertise: The item-descriptor (ID) matching method. *Journal of Applied Testing Technology, 9*(1), 1–22.

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, *64*(3), 425–461.

Kim, S., & Kolen, M. (2004). *STUIRT [computer program]*. The University of Iowa.

Kolen, M. J. & Brennan, R. L. (2004). *Test equating: Methods and practices.* (2nd ed.). Springer-Verlag.

Lee, W., Hanson, B. A., & Brennan, R. L. (2000, October). *Procedures for computing classification consistency and accuracy indices with multiple categories.* (ACT Research Report Series 2000–10). ACT, Inc.

Linacre, J. M. (2011). *Winsteps® Rasch measurement computer program* [computer program]. 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.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174.

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

National Research Council (NRC). (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. https://doi.org/10.17226/13165

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academic Press. https://www.nextgenscience.org/search-standards

Pearson. (2024). *Colorado Measures of Academic Success (CMAS) Science 2022 standard setting report*. Report developed under contract for the Colorado Department of Education (CDE).

Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology, 19*, 49–57.

Revivo, R. Z., Dickinson, E. R., & Borawski, E. A. (2023, October). *Colorado Measures of Academic Success (CMAS) Science alignment study report*. Human Resources Research Organization (HumRRO).

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7(2), 201–210.

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

Vector Psychometric Group. (2022). *IRTPRO* [computer program].

Wells, C. S., Hambleton, R. K., Kirkpatrick, R., & Meng, Y. (2014). An examination of two procedures for identifying consequential item parameter drift. *Applied Measurement in Education, 27,* 214–231.

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, *31*(1), 2–13.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. MESA Press.

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement, 5*(2), 245–262.

Appendix A: Test Blueprints

The following tables present the percentage targets for each content area and grade-level assessment.

Table A.1. Test Blueprint—Mathematics Grades 3–5

Item Type/Subclaim/Calculator Use	Grade 3	Grade 4	Grade 5
Total #Points	50–51	50–51	50–51
Type I	61–62%	61–61%	61–62%
1.1	49–50%	37–38%	37–38%
1.2	12%	24%	24%
1.4	–	–	–
Subclaim A: Major Content	43–44%	47–48%	45–46%
Subclaim B: Supporting Content	18%	14%	16%
Type II	20–22%	20–22%	20–22%
2.3	6–12%	6–12%	6–12%
2.4	8–16%	8–16%	8–16%
Subclaim C: Expressing Mathematical Reasoning	20–22%	20–22%	20–22%
Type III	18%	18%	18%
3.3	6%	6%	6%
3.6	12%	12%	12%
Subclaim D: Modeling and Application	18%	18%	18%
Calculator	–	–	–
Non-Calculator	100%	100%	100%

Table A.2. Test Blueprint—Mathematics Grades 6–8

Item Type/Subclaim/Calculator Use	Grade 6	Grade 7	Grade 8
Total #Points	50–51	50–51	50–51
Type I	61–62%	61–62%	61–62%
1.1	37–38%	37–38%	33–34%
1.2	16%	16%	12–20%
1.4	8%	8%	8–16%
Subclaim A: Major Content	39–40%	45–46%	41–42%
Subclaim B: Supporting Content	22%	16%	20%
Type II	20–22%	20–22%	20–22%
2.3	6–12%	6–12%	6–12%
2.4	8–16%	8–16%	8–16%
Subclaim C: Expressing Mathematical Reasoning	20–22%	20–22%	20–22%
Type III	18%	18%	18%
3.3	6%	6%	6%
3.6	12%	12%	12%
Subclaim D: Modeling and Application	18%	18%	18%
Calculator	72–73%	76%	72–73%
Non-Calculator	27–28%	24%	27–28%

Table A.3. Test Blueprint—ELA Grades 3–5

Subclaim	Grade 3 (includes CSLA)	Grade 4 (includes CSLA)	Grade 5
Total #Points	53 (65)	59 (73)	57 (71)
Reading	77% (63%)	78% (63%)	77% (62%)
Literary Text	32% (26%)	31–34% (25–27%)	32% (25%)
Informational Text	26% (22%)	31% (25%)	32% (25%)
Vocabulary	19% (15%)	14–17% (11–14%)	14% (11%)
Writing	23% (37%)	22% (37%)	23% (38%)
Written Expression	11% (28%)	12% (29%)	12% (30%)
Knowledge and Use of Language Conventions	11% (9%)	10% (8%)	10% (8%)

Note. The numbers in parantheses bare based on weighted Written Expression scores. Scores may not add up as expected due to rounding.

Table A.4. Test Blueprint—ELA Grades 6–8

Subclaim	Grade 6	Grade 7	Grade 8
Total #Points	62 (78)	64 (80)	64 (80)
Reading	77% (62%)	78% (63%)	78% (63%)
Literary Text	26–29% (21–23%)	28% (23%)	28% (23%)
Informational Text	35% (28%)	34% (28%)	34% (28%)
Vocabulary	13–16% (10–13%)	16% (13%)	16% (13%)
Writing	23% (38%)	22% (38%)	22% (38%)
Written Expression	13% (31%)	13% (30%)	13% (30%)
Knowledge and Use of Language Conventions	10% (8%)	9% (8%)	9% (8%)

Note. The numbers in parantheses bare based on weighted Written Expression scores. Scores may not add up as expected due to rounding.

Table A.5. Test Blueprint—Science Grade 5

Standard/Item Type	%Total Test Score Points	#Points		
		Cluster	Mini Cluster	Standalone
Physical	35	7-9	0-6	3-7
Physical/Life	24	7-9	0-6	0-7
Earth and Space	41	7-9	0-6	4-10
Science and Engineering Practices	65–75	–	–	–
SR and TE	53	–	–	–
CR	47	–	–	–
Total	100	24-26	10-12	15

Table A.6. Test Blueprint—Science Grade 8 (MS)

Standard/Item Type	%Total Test Score Points	#Points		
		Cluster	Mini Cluster	Standalone
Physical	34	7	6-12	0-6
Life	36	7	6-12	1-7
Earth and Space	30	7	6	3
Science and Engineering Practices	65–74	–	–	–
SR and TE	53	–	–	–
CR	47	–	–	–
Total	100	21	24	16

Table A.7. Test Blueprint—Science Grade 11 (HS)

Standard/Item Type	%Total Test Score Points	#Points		
		Cluster	Mini Cluster	Standalone
Physical	46	6	5-9	4-7
Life	32	6	4-8	2-6
Earth and Space	31	9	5-8	2-5
Science and Engineering Practices	65–74	–	–	–
SR and TE	53	–	–	–
CR	47	–	–	–
Total	100	18	19	13

Appendix B: Science Cognitive Complexity Framework

Colorado Measures of Academic Success (CMAS) Science

Final Cognitive Complexity Framework, January 2023

COLORADO
Department of Education

Item Cognitive Complexity

Note: Examples provided are not intended to be comprehensive of all items meeting that descriptor.

Item Cognitive Complexity			
	Low - Single Dimension	**Medium - Two Dimensions**	**High - All Three Dimensions**
Item Dimensionality[1]	Item requires demonstration of only one dimension.	Item requires integration of two dimensions, described in the CAS Learning Progressions documents.	Item requires integration of three dimensions, described in the CAS Learning Progressions documents.
	Low - Heavy	**Medium - Moderate**	**High - Minimal**
Scaffolding/Support	The task prescribes a *fully specified* approach for responding. - All components are provided and commonplace; student does not need to infer or select from them: o formulas o Punnett squares o components of energy in a system o unbalanced chemical equations o labels to be applied to familiar models - Problem-solving steps are provided in the stimulus; student only executes them.	The task focuses student to *apply* an approach that is only partially specified. - Student is provided either a partial set of components for a routine task or an excess of components for a more complex task; student must either infer what is missing or discard what is irrelevant: o suggest original components to incorporate into an incomplete model o apply the correct formula(s) out of several available options o explain an observed instance using a specified concept o describe the missing steps from an incomplete investigational procedure - There is some uncertainty associated with the outcome of the scenario. - Some portion of the task is accomplished by way of provided problem-solving steps, but the student must choose or devise some portion of the process.	The task frames a situation that the student must *interpret to select or develop multiple steps* of an approach. - Student is presented with multiple informational inputs between which the relationship is not immediately obvious. - A high degree of uncertainty is associated with the outcome of the scenario. - Student selects or develops multiple problem-solving steps to complete a specific, structured, defined task or goal: o design an investigation o use listed materials to develop an original model o explain an observed instance using one or more unspecified concepts o determine and apply a sequence of formulas to solve a problem

Final CMAS Cognitive Complexity Framework 2023
Adapted from Achieve Cognitive Complexity in Science Assessments and Task Analysis Guide for Science

	Low - Minimal	Medium - Surface	High - Intensive
Sensemaking Sensemaking situation: students are provided material without obvious ties/connections to content (e.g., language of the standard); they must use their knowledge of the standard to explain what they see in the material.	- Task is answerable via rote knowledge connected to the phenomenon solely by context. - Item requires no engagement with the stimulus. - The student can correctly answer without addressing the central concept [mystery/puzzle] of the phenomenon. - Focused on identifying an answer, not on explaining phenomena. o identify the components of a familiar system without explaining their importance to the system. o identify the trend in a graph without using it to explain or predict anything.	- Making sense of a phenomenon or addressing a problem is necessary to accomplish at least a portion of the item, or answering the item is a strategic step toward a sensemaking goal. - Answer requires: o use of information, data, or a model to develop an explanation or argument o connection of multiple pieces of information. - Task asks for standards-based explanation of observations, but not the detailed relationships behind those observations: o determine which of several data sets correlates with the trend under observation o determine which portion of a system is most directly connected to the phenomenon o given data and a proposed cause for a phenomenon, provide support for that cause from the data	- Making sense of a phenomenon or addressing a problem is the fundamental source of challenge in the item. - Meaningful (valid, accurate, causal, etc.) information must be distinguished from other information through reasoning: o speculate a cause for an unusual observation and provide support for that cause from given data o determine corresponding trends between multiple data sets and evaluate for causation o notice patterns within data and connect them to the phenomenon under consideration o predict how a change to one part of a system will impact another part - Task requires use of pertinent standard knowledge to explain both observations and the detailed relationships behind those observations.

[1] - Disciplinary Core Ideas (DCI), Science and Engineering Practices (SEP), Crosscutting Concepts (CCC)

Overall Item Cognitive Complexity Rating

The overall cognitive complexity rating for the item follows a majority rules approach when comparing the ratings for dimensionality, scaffolding/support, and sensemaking.

- Item complexity is **High** if at least two of the three categories are rated at the highest level
- Item complexity is **Low** if two are rated at the lowest level.
- Item complexity is **Medium** in all other cases.

Final CMAS Cognitive Complexity Framework 2023
Adapted from Achieve Cognitive Complexity in Science Assessments and Task Analysis Guide for Science

Cluster Stimulus Cognitive Complexity

Note: Use the following **only for** the purpose of rating cluster and simulation stimuli.

Cluster Stimulus Complexity			
	1 - Low	**2 - Medium**	**3 - High**
Phenomenon/ Stimulus Material	- The task provides a problem or a phenomenon that students are already familiar with how to explain or solve. - Student is presented a simple, probably familiar situation/scenario and selects the appropriate, direct scientific explanation for the phenomenon. - Context is rudimentary or taken directly from the EO, Clarification Statement, or DCI. - Information is limited to that specifically needed to address the task.	- The scenario presents a relatively new phenomenon that students might have some familiarity with, which contains some specific uncertainty for tasks to focus on. - The scenario has multiple facets of information for students to interpret at a grade-appropriate level of sophistication. - Within the scenario there are explicit cues and/or scaffolding to focus students toward related tasks. - The provided components are sufficient for students to arrive at the appropriate scientific explanation for the phenomenon. - Context is substantial and goes beyond examples listed in the standards text.	- The scenario presents a new phenomenon or problem that o is at a level that "figuring out" would be real and authentic for students o is not immediately explainable by the student o likely involves multiple appropriate ways to engage and pursue the task - Connection of context to the standards is indirect or unobvious.

Final CMAS Cognitive Complexity Framework 2023
Adapted from Achieve Cognitive Complexity in Science Assessments and Task Analysis Guide for Science

Appendix C: ELA and CSLA Scoring Rubrics

Grade 3

CMAS Scoring Rubric for Prose Constructed Response Items

COLORADO
Department of Education

Research Simulation Task (RST) and Literary Analysis Task (LAT)

Construct Measured	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Reading Comprehension and Written Expression	The student response • demonstrates **full comprehension** by providing an **accurate** explanation/description/comparison; • addresses the prompt and provides **effective** development of the topic that is **consistently appropriate** to task, purpose, and audience; • uses **clear** reasoning supported by **relevant**, text- based evidence in the development of the topic; • is **effectively** organized with **clear and coherent** writing; • uses language **effectively** to clarify ideas.	The student response • demonstrates **comprehension** by providing a **mostly accurate** explanation/description/comparison; • addresses the prompt and provides **some** development of the topic that is **generally appropriate** to task, purpose, and audience; • uses reasoning and **relevant**, text-based evidence in the development of the topic; • is organized with **mostly clear and coherent** writing; • uses language in a way that is **mostly effective** to darify ideas.	The student response • demonstrates **limited comprehension**; • addresses the prompt and provides **minimal** development of the topic that is **limited in its appropriateness** to task, purpose, and audience; • uses **limited** reasoning and text-based evidence; • demonstrates **limited** organization and coherence; • uses language to express ideas with **limited** clarity.	The student response • **does not** demonstrate comprehension; • is **undeveloped and/or inappropriate** to the task, purpose, and audience; • includes **little to no** text-based evidence; • **lacks** organization and coherence; • **does not** use language to express ideas with clarity.
Knowledge of Language and Conventions	The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There **may be a few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **often impede understanding.**	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

Grade 3 CMAS Scoring Rubric for Prose Constructed Response Items

Narrative Task (NT)

Construct Measured	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Written Expression	The student response • is **effectively** developed with narrative elements and is **consistently appropriate** to the task; • is **effectively** organized with **clear and coherent** writing • uses language **effectively** to clarify ideas.	The student response • is developed with **some** narrative elements and is **generally appropriate** to the task; • is organized with **mostly coherent** writing; • uses language in a way that is **mostly effective** to clarify ideas.	The student response • is **minimally** developed with **few** narrative elements and is **limited in its appropriateness** to the task; • demonstrates **limited** organization and coherence; • uses language to express ideas with **limited** darity.	The student response • is **undeveloped** and/or **inappropriate** to the task; • **lacks** organization and coherence; • **does not** use language to express ideas with clarity.
Knowledge of Language and Conventions	The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There may be a **few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **often impede understanding**.	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

NOTE:

- The reading dimension is not scored for elicited narrative stories.
- Per the CCSS, narrative elements in grades 3-5 may include: establishing a situation; organizing a logical event sequence; describing scenes, objects, or people; developing characters' personalities; and using dialogue as appropriate.
- The elements of organization to be assessed are expressed in the grade-level standards W1-W3.

Developed collaboratively with PARCC.

Grades 4 and 5

CMAS Scoring Rubric for Prose Constructed Response Items

COLORADO
Department of Education

Research Simulation Task (RST) and Literary Analysis Task (LAT)

Construct Measured	Score Point 4	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Reading Comprehension and Written Expression	The student response • demonstrates **full comprehension** of ideas stated explicitly and/or inferentially by providing an **accurate** analysis; • addresses the prompt and provides **effective** development of the topic that is **consistently appropriate** to task, purpose, and audience; • uses **clear** reasoning supported by **relevant**, text-based evidence in the development of the topic; • is **effectively** organized with **clear and coherent** writing; • uses language **effectively** to clarify ideas.	The student response • demonstrates **comprehension** of ideas stated explicitly and/or inferentially by providing a **mostly accurate** analysis; • addresses the prompt and provides **mostly effective** development of the topic that is **appropriate** to task, purpose, and audience; • uses **mostly clear** reasoning supported by **relevant** text-based evidence in the development of the topic; • is organized with **mostly clear and coherent** writing • uses language that is **mostly effective** to clarify ideas.	The student response • demonstrates **basic comprehension** of ideas stated explicitly and/or inferentially by providing a **generally accurate** analysis; • addresses the prompt and provides **some** development of the topic that is **somewhat appropriate** to task, purpose, and audience; • uses **some** reasoning and text-based evidence in the development of the topic; • demonstrates **some** organization with **somewhat** coherent writing; • uses language to express ideas with **some** clarity.	The student response • demonstrates **limited comprehension** of ideas by providing a **minimally accurate** analysis; • addresses the prompt and provides **minimal** development of the topic that is **limited in its appropriateness** to task, purpose, and audience; • uses **limited** reasoning and text-based evidence; • demonstrates **limited** organization and coherence; • uses language to express ideas with **limited** clarity.	The student response • demonstrates **no comprehension** of ideas by providing an **inaccurate or no** analysis. • is **undeveloped and/or inappropriate** to the task, purpose, and audience; • includes **little to no** text- based evidence; • **lacks** organization and coherence; • **does not** use language to express ideas with clarity.
Knowledge of Language and Conventions		The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There **may be a few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **often impede understanding.**	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

Grades 4 and 5 CMAS Scoring Rubric for Prose Constructed Response Items

Narrative Task (NT)

Construct Measured	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Written Expression	The student response • is **effectively** developed with narrative elements and is **consistently appropriate** to the task; • is **effectively** organized with **clear and coherent** writing • uses language **effectively** to clarify ideas.	The student response • is developed with **some** narrative elements and is **generally appropriate** to the task; • is organized with **mostly coherent** writing; • uses language in a way that is **mostly effective** to clarify ideas.	The student response • is **minimally** developed with **few** narrative elements and is **limited in its appropriateness** to the task; • demonstrates **limited** organization and coherence; • uses language to express ideas with **limited** darity.	The student response • is **undeveloped** and/or **inappropriate** to the task; • **lacks** organization and coherence; • **does not** use language to express ideas with clarity.
Knowledge of Language and Conventions	The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There may be a **few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **often impede understanding**.	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

NOTE:

- The reading dimension is not scored for elicited narrative stories.
- Per the CCSS, narrative elements in grades 3-5 may include: establishing a situation; organizing a logical event sequence; describing scenes, objects, or people; developing characters' personalities; and using dialogue as appropriate.
- The elements of organization to be assessed are expressed in the grade-level standards W1-W3.

Developed collaboratively with PARCC.

Grades 6 through 8

CMAS Scoring Rubric for Prose Constructed Response Items

COLORADO
Department of Education

Research Simulation Task (RST) and Literary Analysis Task (LAT)

Construct Measured	Score Point 4	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Reading Comprehension and Written Expression	The student response • demonstrates **full** comprehension of ideas stated explicitly and inferentially by providing an **accurate** analysis; • addresses the prompt and provides **effective and comprehensive** development of the claim or topic that is **consistently appropriate** to task, purpose, and audience; • uses **clear** reasoning supported by **relevant** text-based evidence in the development of the claim or topic; • is **effectively** organized with **clear and coherent** writing; • establishes and maintains an **effective** style.	The student response • demonstrates comprehension of ideas stated explicitly and/or inferentially by providing a **mostly accurate** analysis; • addresses the prompt and provides **mostly effective** development of claim or topic that is **mostly appropriate** to task, purpose, and audience; • uses **mostly clear** reasoning supported by **relevant** text-based evidence in the development of the claim or topic; • is organized with **mostly clear and coherent** writing; • establishes and maintains a **mostly effective** style.	The student response • demonstrates **basic** comprehension of ideas stated explicitly and/or inferentially by providing a **generally accurate** analysis; • addresses the prompt and provides **some** development of claim or topic that is **somewhat appropriate** to task, purpose, and audience; • uses **some** reasoning and text-based evidence in the development of the claim or topic; • demonstrates **some** organization with **somewhat** coherent writing; • has a style that is **somewhat** effective.	The student response • demonstrates **limited** comprehension of ideas stated explicitly and/or inferentially by providing a **minimally accurate** analysis; • addresses the prompt and provides **minimal** development of claim or topic that is **limited in its appropriateness** to task, purpose, and audience; • uses **limited** reasoning and text-based evidence; • demonstrates **limited** organization and coherence; • has a style that is **minimally** effective.	The student response • demonstrates **no** comprehension of ideas by providing an **inaccurate or no** analysis; • is **undeveloped and/or inappropriate** to task, purpose, and audience; • includes **little to no** text-based evidence; • **lacks** organization and coherence; • has an **inappropriate** style.
Knowledge of Language and Conventions		The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There **may be a few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may be errors** in mechanics, grammar, and usage that **often impede understanding.**	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

Grades 6 through 8 CMAS Scoring Rubric for Prose Constructed Response Items

Narrative Task (NT)

Construct Measured	Score Point 4	Score Point 3	Score Point 2	Score Point 1	Score Point 0
Written Expression	The student response • is **effectively** developed with narrative elements and is **consistently appropriate** to the task; • is **effectively** organized with **clear and coherent** writing • establishes and maintains an **effective** style.	The student response • is **mostly effectively** developed with narrative elements and is **mostly appropriate** to the task; • is organized with **mostly clear and coherent** writing • establishes and maintains a **mostly effective** style.	The student response • is developed with **some** narrative elements and is **generally appropriate** to the task; • demonstrates **some** organization with **somewhat coherent** writing; • has a style that is **somewhat** effective.	The student response • is **minimally** developed with **few** narrative elements and is **limited in its appropriateness** to the task; • demonstrates **limited** organization and coherence; • has a style that has **limited** effectiveness.	The student response • is **undeveloped and/or inappropriate** to the task; • **lacks** organization and coherence; • has an **inappropriate** style.
Knowledge of Language and Conventions		The student response to the prompt demonstrates **full command** of the conventions of standard English at an appropriate level of complexity. There may be a **few minor errors** in mechanics, grammar, and usage, but **meaning is clear**.	The student response to the prompt demonstrates **some command** of the conventions of standard English at an appropriate level of complexity. There **may** be errors in mechanics, grammar, and usage that **occasionally impede understanding**, but the **meaning is generally clear**.	The student response to the prompt demonstrates **limited command** of the conventions of standard English at an appropriate level of complexity. There **may** be errors in mechanics, grammar, and usage that **often impede understanding**.	The student response to the prompt **does not demonstrate command** of the conventions of standard English at the appropriate level of complexity. **Frequent and varied errors** in mechanics, grammar, and usage **impede understanding**.

NOTE:

- The reading dimension is not scored for elicited narrative stories.
- The elements of coherence, clarity, and cohesion to be assessed are expressed in the grade-level standards 1-4 for writing.
- Tone is not assessed in grade 6.
- Per the CCSS, narrative elements in grades 3-5 may include: establishing a situation; organizing a logical event sequence; describing scenes, objects, or people; developing characters' personalities; and using dialogue as appropriate. In grades 6-8, narrative elements may include, in addition to the grades 3-5 elements: establishing a context, situating events in a time and place, developing a point of view, and developing characters' motives. The elements to be assessed are expressed in grade-level standards 3 for writing.

Developed collaboratively with PARCC.

Appendix D: Sample Student Performance Reports

Confidential Student Performance Report

Colorado Measures of Academic Success

Student: FIRSTNAME M. (PREFERRED) LASTNAME
SASID: 9999999999 Birthdate: MM/DD/YYYY
School: ELEMENTARY SCHOOL (9999)
District: DISTRICT NAME (9999)

Spring 2023

Mathematics — Grade 3

CMAS is the only assessment given to all Colorado students that measures what students should know and be able to do at the end of each grade. This report describes your student's understanding of Colorado's grade 3 Mathematics expectations. Scan the QR code at the bottom of this page to see a video that will talk you through your student's report.

Your student's performance is shown as:

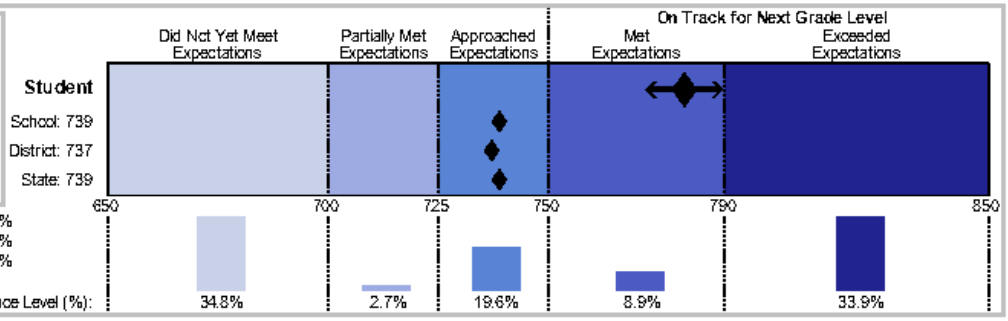
- A scale score: A numerical score based on Colorado's grade 3 Mathematics expectations
- A performance level: Your student's performance level is described at the bottom of this page
- A percentile: How your student performed in comparison to other Colorado students

Consider as you review this report:

- Arrows around your student's diamond show where your student may have scored if the assessment was taken multiple times.
- Make school, district, and state comparisons with caution if participation is low.
- Talk with your student's teacher about your student's progress in Mathematics.

Your Student's Score
781
Met Expectations
62nd Percentile

School Participation: 64.2%
District Participation: 64.8%
State Participation: 73.7%



Performance Level Description* - Met Expectations

3

PREFERRED **Met Expectations** and is on track for the next grade level. Students in this level typically demonstrate the following:

Major, Additional & Supporting Content

- Determine unknown numbers in problems with one factor greater than or equal to 5.
- Justify comparisons of two fractions with the same numerator or denominator with a visual model. Demonstrate understanding of the quantity a/b on a number line and its relationship to $1/b$.
- Solve one-step and two-step word problems involving addition or subtraction of time intervals. Measure and estimate liquid volumes and masses using any of the four operations. Solve one-step word problems using estimated measurements. Represent data on a scaled picture graph, a scaled bar graph, or a line plot with appropriate units.
- Represent area of a plane figure as square units. Solve mathematical problems with unknown side lengths in perimeters of polygons. Understand properties of quadrilaterals and subcategories and draw examples of quadrilaterals with stated attributes

Expressing Mathematical Reasoning

- Communicate reasoning with no calculation errors. Interpret and critique the reasoning of others. Use precision in grade-appropriate communication.

Modeling & Application

- Use approximations to apply mathematics to a real-world situation. Analyze relationships between values to draw conclusions. Create a model by selecting appropriate tools, then improve the model based upon results.

Performance level descriptors (PLDs) are organized in a manner that assumes students demonstrating higher levels of command have mastered the concepts and skills within lower levels. To view a video report and the full version of the performance level descriptor, visit https://coassessments.com/parentsandguardians or access the QR code.

*Adapted from ilClassroom in Action's Performance Level Summaries

Watch a video about this report!

Information about the Colorado Academic Standards measured by this assessment:
http://www.cde.state.co.us/comath/statestandards.

Page 1 of 2

06062023-Z9999999-0130-0016 - 0000000

FIRSTNAME M. (PREFERRED) LASTNAME

Mathematics

Confidential

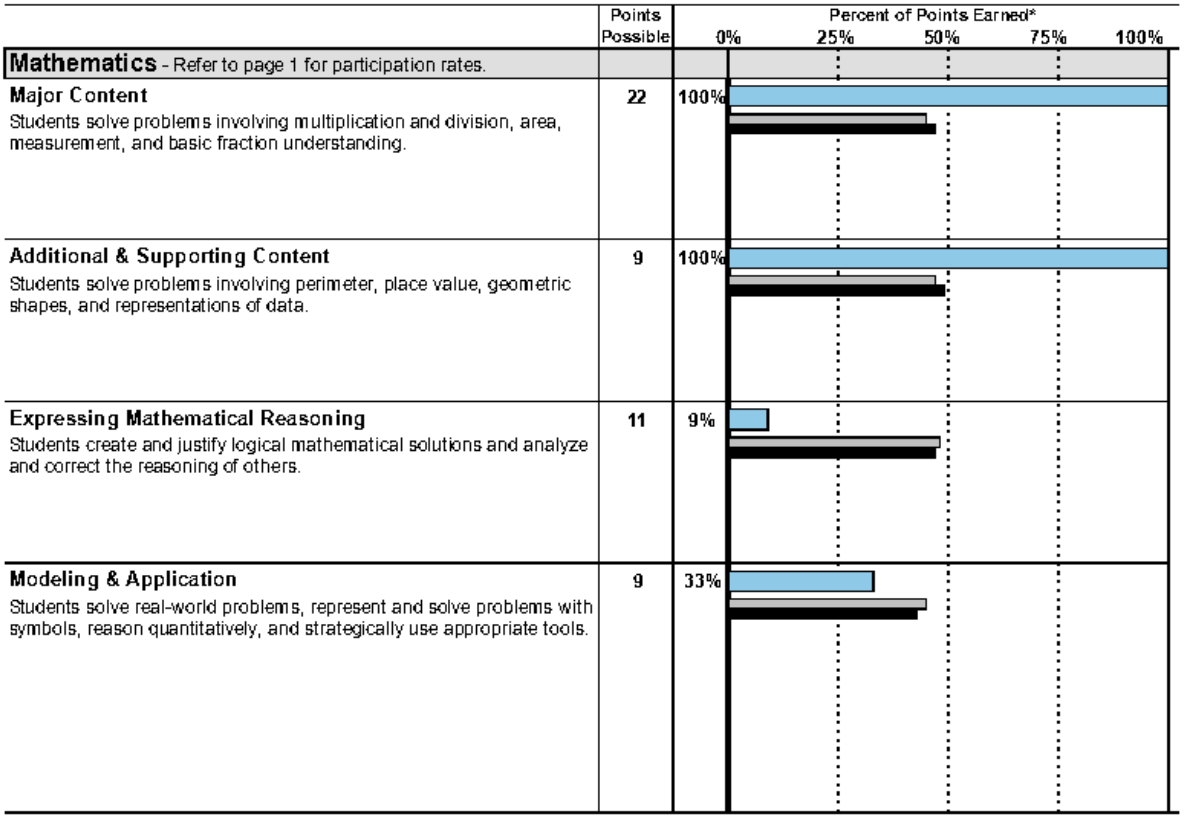
Subclaim Performance

The top bar in each of the other graphs shows the percent of points your student earned for each of the four mathematics assessment subclaims.

District Averages are provided for comparison.

State Averages are provided for comparison.

Average of students at the Met Expectations performance level starting point.



*Percent of points earned cannot be compared across years because individual items change from year to year. They also cannot be compared across subclaims because the number of items and the difficulty of items may not be the same.

For information about the CMAS assessment program, visit http://www.cde.state.co.us/assessment/cmas.

Page 2 of 2

Confidential Student Performance Report

Colorado Measures of Academic Success

Student: **FIRSTNAME M. LASTNAME**
SASID: 999999999 Birthdate: MM/DD/YYYY
School: ELEMENTARY SCHOOL (9999)
District: DISTRICT NAME (9999)

Spring 2023

English Language Arts/Literacy — Grade 3

CMAS is the only assessment given to all Colorado students that measures what students should know and be able to do at the end of each grade. This report describes your student's understanding of Colorado's grade 3 English Language Arts/Literacy expectations. Scan the QR code at the bottom of this page to see a video that will talk you through your student's report.

Your student's performance is shown as:

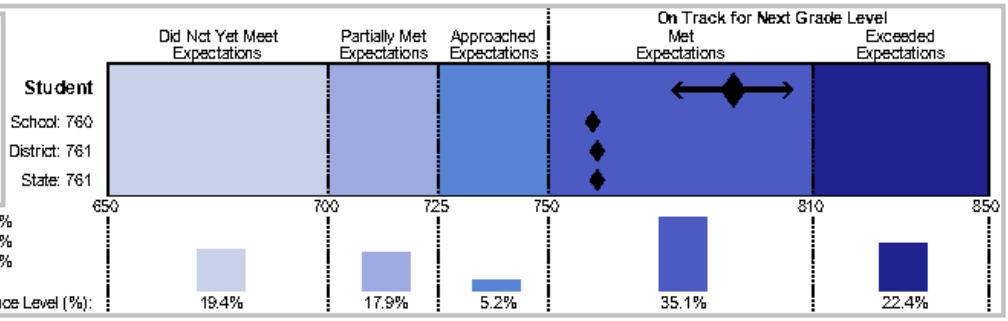
- A scale score: A numerical score based on Colorado's grade 3 English Language Arts/Literacy expectations
- A performance level: Your student's performance level is described at the bottom of this page
- A percentile: How your student performed in comparison to other Colorado students

Consider as you review this report:

- Arrows around your student's diamond show where your student may have scored if the assessment was taken multiple times.
- Make school, district, and state comparisons with caution if participation is low.
- Talk with your student's teacher about your student's progress in English Language Arts/Literacy.

Your Student's Score
792
Met Expectations
66th Percentile

School Participation: 81.5%
District Participation: 81.8%
State Participation: 88.2%



Performance Level Description - Met Expectations

3

FIRSTNAME **Met Expectations** and is on track for the next grade level. Students in this level typically demonstrate the following:

Reading

- With very complex text: the ability to be generally accurate when asking and/or answering questions, showing general understanding of the text when referring to explicit details and examples in the text.
- With moderately complex text: the ability to be generally accurate when asking and/or answering questions, showing general understanding of the text when referring to explicit details and examples in the text.
- With readily accessible text: the ability to be mostly accurate when asking and/or answering questions, showing understanding of the text when referring to explicit details and examples in the text.

Writing

Written Expression: students typically address the prompts and provide development of ideas, including when drawing evidence from multiple sources, while in the majority of instances demonstrating purposeful and mostly controlled organization. Students typically:

- Develop the topic and/or narrative elements using reasoning, details, text-based evidence, and/or description.
- Develop topic and/or narrative elements in a manner that is mostly appropriate to the task and purpose.
- Demonstrate purposeful organization that is mostly controlled and may include an introduction and/or conclusion.
- Use linking words and phrases, descriptive words, and/or temporal words to express ideas with clarity.

Knowledge and use of Language and Conventions: students typically demonstrate command of the conventions of Standard English consistent with edited writing. There are errors in grammar and usage that may occasionally impede understanding.

To view a video report and the full version of the performance level descriptor, visit **https://coassessments.com/parentsandguardians** or access the QR code.

Watch a video about this report!

Information about the Colorado Academic Standards measured by this assessment:
http://www.cde.state.co.us/coreadingwriting/statestandards.

Page 1 of 2

06062023-Z9999999-9999-9999- 0000000

COPVABCFN S. COABCPVLNME

English Language Arts/Literacy

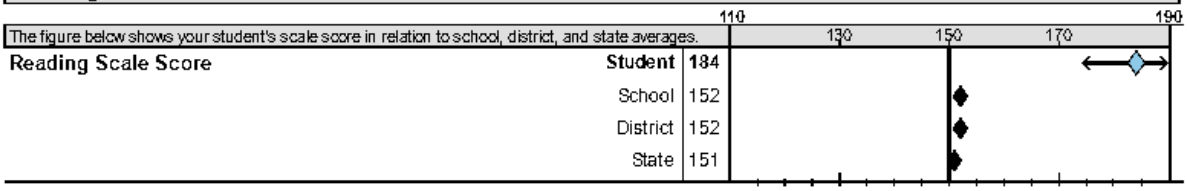
Confidential

Subclaim Performance

The top diamond in the figure below shows your student's performance in Reading.
The top bar in each of the other graphs shows the percent of points your student earned for writing and specific areas of reading and writing.
District Averages are provided for comparison.
State Averages are provided for comparison.
Average of students at the Met Expectations performance level starting point.

Reading - Refer to page 1 for participation rates.

The figure below shows your student's scale score in relation to school, district, and state averages.



	Points Possible	Percent of Points Earned* (0% 25% 50% 75% 100%)
Literary Text Students read and analyze fiction, drama, and poetry.	17	82%
Informational Text Students read and analyze nonfiction, history, science, and the arts.	14	100%
Vocabulary Students use context to determine what words and phrases mean.	10	100%

	Points Possible	Percent of Points Earned* (0% 25% 50% 75% 100%)
Writing - Refer to page 1 for participation rates.		
Overall Writing Overall is calculated by multiplying Written Expression points by three and adding Language and Conventions points.	24	0%
Written Expression Students compose well-developed writing using details from what they have read.	6	0%
Language and Conventions Students demonstrate knowledge of conventions and other important elements of language.	6	0%

*Percent of points earned cannot be compared across years because individual items change from year to year. They also cannot be compared across subclaims because the number of items and the difficulty of items may not be the same.

For information about the CMAS assessment program, visit
http://www.cde.state.co.us/assessment/cmas.

Page 2 of 2

Confidential Student Performance Report

Colorado Measures of Academic Success
Based on the 2020 Colorado Academic Standards
Student: **FIRSTNAME M. LASTNAME**
SASID: 9999999999 Birthdate: MM/DD/YYYY
School: ELEMENTARY SCHOOL (9999)
District: DISTRICT NAME (9999)
Spring 2023

Science — Grade 5

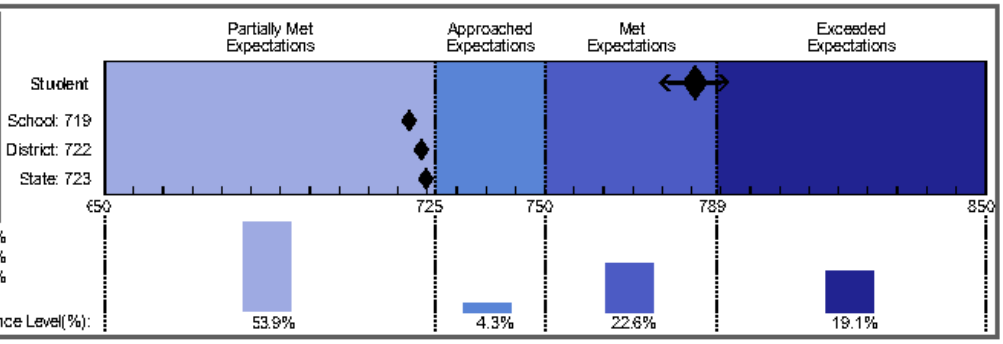
CMAS is the only assessment given to all Colorado students that measures what students should know and be able to do at the end of each grade. This report describes your student's understanding of Colorado's grade 5 science expectations. Scan the QR code at the bottom of this page to see a video that will talk you through your student's report.

Your student's performance is shown as:
- A scale score: A numerical score based on Colorado's grade 5 science expectations
- A performance level: Your student's performance level is described at the bottom of this page
- A percentile: How your student performed in comparison to other Colorado students

Consider as you review this report:
- Arrows around your student's diamond show where your student may have scored if the assessment was taken multiple times.
- Make school, district, and state comparisons with caution if participation is low.
- Talk with your student's teacher about your student's progress in science.

Your Student's Score
784
Met Expectations
70th Percentile

School Participation: 94.4%
District Participation: 94.6%
State Participation: 89.8%



Performance Level Description - Met Expectations

FIRSTNAME showed a strong understanding of the Colorado Academic Standards' grade 5 science expectations and is ready for the next grade level. Students in the Met Expectations level typically:

- Describe matter (particles too small to be seen) as always conserved, and mixing can result in new substances.
- Make observations and measurements of properties used to identify materials.
- Describe evidence that demonstrates Earth's gravity as the cause of objects being pulled down toward its center.
- Demonstrate that all energy in food on Earth was once energy from the Sun.
- Explain matter and energy cycles in an ecosystem, and explain that plants get materials to grow from air and water.
- Describe that a star's distance from Earth affects its apparent brightness.
- Demonstrate patterns caused by Earth's orbit and rotation and the orbit of the Moon around Earth.
- Model the interactions between Earth's major systems and their impact on shaping Earth's surface.
- Describe the relative proportions of salt water and fresh water in different reservoirs on Earth.
- Communicate ways that communities use scientific ideas to protect Earth's environment and resources.

5

To view a video report and the full version of the performance level descriptor, visit https://coassessments.com/parentsandguardians/ or access the QR code.

Watch a video about this report!

Information about the Colorado Academic Standards measured by this assessment:
http://www.cde.state.co.us/coscience/statestandards.

Page 1 of 2 (6062023-Z9999999-0130-0016 - 0000000)

FIRSTNAME M. LASTNAME

Science

Confidential

Subscale Performance

- The shaded areas below represent about 70% of student scores across the state.
- Diamonds outside of the shaded area indicate a potential weakness or strength compared to the state.

| Average of students at the Met Expectations performance level starting point. |

Reporting Category Description	Subscale Score		Lower than Average / Average / Higher than Average (400 – 450 – 500 – 550)
Physical Science			
Common properties, forms, and changes in matter and energy	**519**	**Student**	
	460	School	
	463	District	
	464	State	
Physical/Life Science			448 – 498
Characteristics of living things, processes of life, and how living things interact with each other and their environment	**538**	**Student**	
	457	School	
	461	District	
	461	State	
Earth and Space Science			452 – 500
Processes and interactions of Earth's systems, and the structure and dynamics of Earth and other objects in space	**550**	**Student**	
	458	School	
	461	District	
	463	State	
Science and Engineering Practices			453 – 502
Making sense of the natural world through investigation and problem solving	**544**	**Student**	
	459	School	
	463	District	
	464	State	

Performance by Prepared Graduate Statements (PGs) and Grade Level Expectations (GLEs)

- PGs and GLEs identify what students need to master to be ready for the next grade level.
- The figure below shows the percent of points your student earned for each grade 5 science GLE.

Student's performance
District average
State average

Standard, PG and GLE	Points Possible	Percent of Points Earned* (0% 25% 50% 75% 100%)
Physical Science		
PG 1: Structure, properties, and interactions of matter		
GLE 1: Matter exists as particles too small to be seen; Properties can be used to identify materials	6	67%
GLE 2: Chemical reactions and the Law of Conservation of Mass	6	67%
GLE 3: Gravity	6	50%
Physical/Life Science		
PG 1: Structure, properties, and interactions of matter		
GLE 4: Energy from food was once energy from the sun	6	100%
PG 6: How living systems interact with the environment		
GLE 2: Plants get most of their material for growth from air and water		
PG 6: How living systems interact with the environment		
GLE 1: Matter cycles between air and soil; Organisms live and die	6	50%
Earth and Space Science		
PG 9: The universe and Earth's place in it		
GLE 1: Earth's major systems interact in multiple ways	8	88%
GLE 2: Interactions between Earth's orbit and the moon's orbit		
PG 10: How and why Earth is constantly changing		
GLE 3: Earth's major systems interact in multiple ways	7	100%
GLE 4: Earth's major water is in the ocean and much of Earth's freshwater is in glaciers or underground		
GLE 5: Societal activities have major effects on land, ocean, atmosphere and even outer space	6	83%

*Percent of points earned cannot be compared across years because individual items change from year to year. They also cannot be compared across PGs because the number of items and the difficulty of items may not be the same.

Page 2 of 2

Appendix E: Student Participation by Demographic Group

Table E.1. Student Participation N-Count Demographic Distribution—Mathematics

Subgroup	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
No IEP	49,964	49,488	49,824	49,244	48,095	46,428
IEP	7,418	7,301	7,072	6,669	6,053	5,608
No Accommodation	53,110	51,979	52,062	51,042	49,474	47,522
Accommodation	4,272	4,810	4,834	4,871	4,674	4,514
Am. Indian/Alaska Native	354	360	358	342	351	361
Asian	1,982	2,061	1,976	1,970	1,737	1,700
Black	2,646	2,617	2,508	2,412	2,349	2,285
Hispanic	19,134	19,426	19,695	19,778	19,719	19,525
White	29,847	29,091	29,224	28,353	27,182	25,671
Hawaiian/Pacific Islander	191	191	202	192	161	155
Two or More Races	3,225	3,040	2,930	2,862	2,640	2,333
Missing	*	*	*	*	*	*
No Economic Disadvantage	22,504	22,021	21,945	21,173	20,649	19,955
Economic Disadvantage	34,878	34,768	34,951	34,740	33,499	32,081
Female	28,318	27,931	27,700	27,266	26,065	24,738
Male	29,064	28,858	29,196	28,647	28,083	27,298
Language Proficiency NA	47,355	46,522	46,354	45,451	44,035	42,810
Language Proficiency NEP	2,969	2,387	1,456	1,364	1,470	1,433
Language Proficiency LEP	5,985	5,674	5,196	4,066	4,000	3,637
Language Proficiency FEP	1,073	2,206	3,890	5,032	4,643	4,156
Not Migrant	57,188	56,583	56,696	55,735	53,973	51,865
Migrant	194	206	200	178	175	171

*n-count less than 16

Table E.2. Student Participation N-Count Demographic Distribution—ELA

Subgroup	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
No IEP	48,525	48,349	49,580	48,929	47,822	46,148
IEP	7,212	7,170	7,077	6,673	6,073	5,612
No Accommodation	51,771	50,988	51,842	50,912	49,420	47,482
Accommodation	3,966	4,531	4,815	4,690	4,475	4,278
Am. Indian/Alaska Native	352	361	362	345	349	358
Asian	1,945	2,033	1,954	1,937	1,721	1,672
Black	2,644	2,621	2,504	2,397	2,349	2,268
Hispanic	17,620	18,202	19,484	19,559	19,487	19,315
White	29,752	29,056	29,208	28,323	27,185	25,646
Hawaiian/Pacific Islander	184	193	201	191	158	155
Two or More Races	3,237	3,050	2,941	2,846	2,637	2,340
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,560	34,522	34,895	34,649	33,423	32,002
Economic Disadvantage	21,177	20,997	21,762	20,953	20,472	19,758
Female	27,479	27,290	27,614	27,126	25,982	24,638
Male	28,258	28,229	29,043	28,476	27,913	27,122
Language Proficiency NA	47,322	46,561	46,414	45,433	44,086	42,820
Language Proficiency NEP	2,050	1,647	1,159	1,067	1,167	1,168
Language Proficiency LEP	5,293	5,101	5,185	4,060	3,992	3,632
Language Proficiency FEP	1,072	2,210	3,899	5,042	4,650	4,140
Not Migrant	55,564	55,337	56,468	55,431	53,731	51,594
Migrant	173	182	189	171	164	166

*n-count less than 16

Table E.3. Student Participation N-Count Demographic Distribution—CSLA

Subgroup	Grade 3	Grade 4
No IEP	1,298	1,070
IEP	142	110
No Accommodation	1,239	976
Accommodation	201	204
Am. Indian/Alaska Native	*	*
Asian	*	*
Black	*	*
Hispanic	1,430	1,170
White	*	*
Hawaiian/Pacific Islander	*	*
Two or More Races	*	*
Missing	*	*
No Economic Disadvantage	241	191
Economic Disadvantage	1,199	989
Female	758	604
Male	682	576
Language Proficiency NA	*	*
Language Proficiency NEP	747	604
Language Proficiency LEP	693	576
Language Proficiency FEP	*	*
Not Migrant	1,426	1,163
Migrant	*	17

*n-count less than 16

Table E.4. Student Participation N-Count Demographic Distribution—Science

Subgroup	Grade 5	Grade 8	Grade 11
No IEP	49,442	45,490	28,972
IEP	6,986	5,457	2,795
No Accommodation	52,265	47,278	29,510
Accommodation	4,163	3,669	2,257
Am. Indian/Alaska Native	359	352	223
Asian	1,967	1,671	969
Black	2,481	2,199	1,399
Hispanic	19,509	19,164	13,088
White	29,002	25,117	14,679
Hawaiian/Pacific Islander	201	153	93
Two or More Races	2,906	2,286	1,309
Missing	*	*	*
No Economic Disadvantage	34,695	31,412	20,014
Economic Disadvantage	21,733	19,535	11,753
Female	27,472	24,168	14,935
Male	28,956	26,779	16,832
Language Proficiency NA	45,968	41,893	27,032
Language Proficiency NEP	1,433	1,397	814
Language Proficiency LEP	5,151	3,568	1,985
Language Proficiency FEP	3,876	4,089	1,936
Not Migrant	56,232	50,777	31,674
Migrant	196	170	93

*n-count less than 16

Appendix F: Scale Score Distributions

Table F.1. Scale Score Distribution—Mathematics Grade 3

SS	Freq.	%	Cum. Freq.	Cum. %
650	845	1.47	845	1.47
651	51	0.09	896	1.56
652	83	0.14	979	1.71
653	70	0.12	1049	1.83
654	50	0.09	1099	1.92
655	61	0.11	1160	2.02
656	75	0.13	1235	2.15
657	54	0.09	1289	2.25
658	86	0.15	1375	2.40
659	113	0.20	1488	2.59
660	107	0.19	1595	2.78
661	103	0.18	1698	2.96
662	129	0.22	1827	3.18
663	101	0.18	1928	3.36
664	99	0.17	2027	3.53
665	109	0.19	2136	3.72
666	126	0.22	2262	3.94
667	115	0.20	2377	4.14
668	158	0.28	2535	4.42
669	124	0.22	2659	4.63
670	152	0.26	2811	4.90
671	165	0.29	2976	5.19
672	150	0.26	3126	5.45
673	192	0.33	3318	5.78
674	172	0.30	3490	6.08
675	175	0.30	3665	6.39
676	183	0.32	3848	6.71
677	182	0.32	4030	7.02
678	192	0.33	4222	7.36
679	222	0.39	4444	7.74
680	197	0.34	4641	8.09
681	221	0.39	4862	8.47
682	227	0.40	5089	8.87
683	240	0.42	5329	9.29
684	236	0.41	5565	9.70
685	238	0.41	5803	10.11
686	275	0.48	6078	10.59
687	262	0.46	6340	11.05
688	309	0.54	6649	11.59
689	319	0.56	6968	12.14
690	260	0.45	7228	12.60
691	284	0.49	7512	13.09
692	274	0.48	7786	13.57
693	307	0.54	8093	14.10
694	321	0.56	8414	14.66
695	329	0.57	8743	15.24
696	343	0.60	9086	15.83
697	374	0.65	9460	16.49
698	384	0.67	9844	17.16
699	349	0.61	10193	17.76
700	358	0.62	10551	18.39
701	381	0.66	10932	19.05
702	357	0.62	11289	19.67
703	393	0.68	11682	20.36
704	395	0.69	12077	21.05
705	366	0.64	12443	21.68
706	411	0.72	12854	22.40
707	399	0.70	13253	23.10
708	400	0.70	13653	23.79
709	451	0.79	14104	24.58
710	401	0.70	14505	25.28
711	418	0.73	14923	26.01
712	442	0.77	15365	26.78
713	392	0.68	15757	27.46
714	426	0.74	16183	28.20
715	440	0.77	16623	28.97
716	444	0.77	17067	29.74
717	465	0.81	17532	30.55
718	467	0.81	17999	31.37
719	498	0.87	18497	32.23
720	465	0.81	18962	33.05
721	462	0.81	19424	33.85
722	476	0.83	19900	34.68
723	513	0.89	20413	35.57
724	466	0.81	20879	36.39
725	516	0.90	21395	37.29
726	477	0.83	21872	38.12
727	495	0.86	22367	38.98
728	540	0.94	22907	39.92
729	534	0.93	23441	40.85
730	494	0.86	23935	41.71
731	508	0.89	24443	42.60
732	548	0.96	24991	43.55
733	480	0.84	25471	44.39
734	560	0.98	26031	45.36
735	574	1.00	26605	46.36
736	520	0.91	27125	47.27
737	529	0.92	27654	48.19
738	531	0.93	28185	49.12
739	584	1.02	28769	50.14
740	547	0.95	29316	51.09
741	558	0.97	29874	52.06
742	543	0.95	30417	53.01
743	565	0.98	30982	53.99
744	523	0.91	31505	54.90
745	584	1.02	32089	55.92

SS	Freq.	%	Cum. Freq.	Cum. %
746	567	0.99	32656	56.91
747	551	0.96	33207	57.87
748	498	0.87	33705	58.74
749	514	0.90	34219	59.63
750	546	0.95	34765	60.59
751	553	0.96	35318	61.55
752	528	0.92	35846	62.47
753	508	0.89	36354	63.35
754	545	0.95	36899	64.30
755	509	0.89	37408	65.19
756	522	0.91	37930	66.10
757	550	0.96	38480	67.06
758	543	0.95	39023	68.01
759	497	0.87	39520	68.87
760	533	0.93	40053	69.80
761	510	0.89	40563	70.69
762	496	0.86	41059	71.55
763	539	0.94	41598	72.49
764	468	0.82	42066	73.31
765	520	0.91	42586	74.21
766	480	0.84	43066	75.05
767	503	0.88	43569	75.93
768	479	0.83	44048	76.76
769	470	0.82	44518	77.58
770	451	0.79	44969	78.37
771	438	0.76	45407	79.13
772	454	0.79	45861	79.92
773	420	0.73	46281	80.65
774	398	0.69	46679	81.35
775	420	0.73	47099	82.08
776	398	0.69	47497	82.77
777	388	0.68	47885	83.45
778	392	0.68	48277	84.13
779	394	0.69	48671	84.82
780	383	0.67	49054	85.49
781	393	0.68	49447	86.17
782	384	0.67	49831	86.84
783	340	0.59	50171	87.43
784	355	0.62	50526	88.05
785	301	0.52	50827	88.58
786	327	0.57	51154	89.15
787	325	0.57	51479	89.71
788	280	0.49	51759	90.20
789	262	0.46	52021	90.66
790	266	0.46	52287	91.12
791	276	0.48	52563	91.60
792	247	0.43	52810	92.03
793	190	0.33	53000	92.36
794	235	0.41	53235	92.77
795	228	0.40	53463	93.17
796	201	0.35	53664	93.52
797	222	0.39	53886	93.91
798	151	0.26	54037	94.17
799	199	0.35	54236	94.52
800	187	0.33	54423	94.84
801	194	0.34	54617	95.18
802	157	0.27	54774	95.46
803	108	0.19	54882	95.64
804	173	0.30	55055	95.94
805	120	0.21	55175	96.15
806	149	0.26	55324	96.41
807	110	0.19	55434	96.61
808	139	0.24	55573	96.85
809	112	0.20	55685	97.04
810	108	0.19	55793	97.23
811	72	0.13	55865	97.36
812	122	0.21	55987	97.57
813	85	0.15	56072	97.72
814	88	0.15	56160	97.87
815	50	0.09	56210	97.96
816	73	0.13	56283	98.08
817	80	0.14	56363	98.22
818	44	0.08	56407	98.30
819	78	0.14	56485	98.44
820	36	0.06	56521	98.50
821	55	0.10	56576	98.60
822	59	0.10	56635	98.70
823	40	0.07	56675	98.77
824	57	0.10	56732	98.87
825	37	0.06	56769	98.93
826	22	0.04	56791	98.97
827	38	0.07	56829	99.04
828	46	0.08	56875	99.12
829	19	0.03	56894	99.15
830	32	0.06	56926	99.21
831	37	0.06	56963	99.27
832	22	0.04	56985	99.31
833	24	0.04	57009	99.35
834	33	0.06	57042	99.41
835	30	0.05	57072	99.46
836	14	0.02	57086	99.48
837	9	0.02	57095	99.50
838	35	0.06	57130	99.56
839	27	0.05	57157	99.61
840	11	0.02	57168	99.63
841	5	0.01	57173	99.64
842	10	0.02	57183	99.65
843	10	0.02	57193	99.67
844	10	0.02	57203	99.69
845	9	0.02	57212	99.70
846	3	0.01	57215	99.71
847	8	0.01	57223	99.72
848	29	0.05	57252	99.77
849	9	0.02	57261	99.79
850	121	0.21	57382	100.00

Table F.2. Scale Score Distribution—Mathematics Grade 4

SS	Freq.	%	Cum. Freq.	Cum. %
650	293	0.52	293	0.52
651	37	0.07	330	0.58
652	14	0.02	344	0.61
653	2	0.00	346	0.61
654	33	0.06	379	0.67
655	33	0.06	412	0.73
656	30	0.05	442	0.78
657	24	0.04	466	0.82
658	15	0.03	481	0.85
659	27	0.05	508	0.89
660	90	0.16	598	1.05
661	31	0.05	629	1.11
662	27	0.05	656	1.16
663	92	0.16	748	1.32
664	47	0.08	795	1.40
665	56	0.10	851	1.50
666	49	0.09	900	1.58
667	74	0.13	974	1.72
668	58	0.10	1032	1.82
669	105	0.18	1137	2.00
670	39	0.07	1176	2.07
671	111	0.20	1287	2.27
672	129	0.23	1416	2.49
673	99	0.17	1515	2.67
674	168	0.30	1683	2.96
675	117	0.21	1800	3.17
676	169	0.30	1969	3.47
677	166	0.29	2135	3.76
678	184	0.32	2319	4.08
679	196	0.35	2515	4.43
680	213	0.38	2728	4.80
681	190	0.33	2918	5.14
682	281	0.49	3199	5.63
683	271	0.48	3470	6.11
684	278	0.49	3748	6.60
685	292	0.51	4040	7.11
686	324	0.57	4364	7.68
687	285	0.50	4649	8.19
688	360	0.63	5009	8.82
689	363	0.64	5372	9.46
690	361	0.64	5733	10.10
691	370	0.65	6103	10.75
692	452	0.80	6555	11.54
693	431	0.76	6986	12.30
694	476	0.84	7462	13.14
695	460	0.81	7922	13.95
696	467	0.82	8389	14.77
697	465	0.82	8854	15.59
698	511	0.90	9365	16.49
699	504	0.89	9869	17.38
700	508	0.89	10377	18.27
701	528	0.93	10905	19.20
702	496	0.87	11401	20.08
703	540	0.95	11941	21.03
704	528	0.93	12469	21.96
705	524	0.92	12993	22.88
706	491	0.86	13484	23.74
707	465	0.82	13949	24.56
708	506	0.89	14455	25.45
709	508	0.89	14963	26.35
710	531	0.94	15494	27.28
711	512	0.90	16006	28.19
712	523	0.92	16529	29.11
713	572	1.01	17101	30.11
714	545	0.96	17646	31.07
715	514	0.91	18160	31.98
716	551	0.97	18711	32.95
717	560	0.99	19271	33.93
718	559	0.98	19830	34.92
719	551	0.97	20381	35.89
720	570	1.00	20951	36.89
721	564	0.99	21515	37.89
722	581	1.02	22096	38.91
723	552	0.97	22648	39.88
724	555	0.98	23203	40.86
725	593	1.04	23796	41.90
726	562	0.99	24358	42.89
727	548	0.96	24906	43.86
728	556	0.98	25462	44.84
729	546	0.96	26008	45.80
730	594	1.05	26602	46.84
731	591	1.04	27193	47.88
732	614	1.08	27807	48.97
733	614	1.08	28421	50.05
734	607	1.07	29028	51.12
735	642	1.13	29670	52.25
736	609	1.07	30279	53.32
737	621	1.09	30900	54.41
738	574	1.01	31474	55.42
739	641	1.13	32115	56.55
740	639	1.13	32754	57.68
741	602	1.06	33356	58.74
742	557	0.98	33913	59.72
743	629	1.11	34542	60.83
744	604	1.06	35146	61.89
745	581	1.02	35727	62.91
746	627	1.10	36354	64.02
747	623	1.10	36977	65.11
748	628	1.11	37605	66.22
749	625	1.10	38230	67.32

SS	Freq.	%	Cum. Freq.	Cum. %
750	577	1.02	38807	68.34
751	612	1.08	39419	69.41
752	609	1.07	40028	70.49
753	598	1.05	40626	71.54
754	594	1.05	41220	72.58
755	583	1.03	41803	73.61
756	584	1.03	42387	74.64
757	565	0.99	42952	75.63
758	545	0.96	43497	76.59
759	516	0.91	44013	77.50
760	512	0.90	44525	78.40
761	536	0.94	45061	79.35
762	503	0.89	45564	80.23
763	498	0.88	46062	81.11
764	450	0.79	46512	81.90
765	419	0.74	46931	82.64
766	450	0.79	47381	83.43
767	440	0.77	47821	84.21
768	403	0.71	48224	84.92
769	411	0.72	48635	85.64
770	386	0.68	49021	86.32
771	350	0.62	49371	86.94
772	365	0.64	49736	87.58
773	366	0.64	50102	88.22
774	347	0.61	50449	88.84
775	350	0.62	50799	89.45
776	340	0.60	51139	90.05
777	311	0.55	51450	90.60
778	310	0.55	51760	91.14
779	295	0.52	52055	91.66
780	256	0.45	52311	92.11
781	276	0.49	52587	92.60
782	253	0.45	52840	93.05
783	232	0.41	53072	93.45
784	221	0.39	53293	93.84
785	205	0.36	53498	94.20
786	208	0.37	53706	94.57
787	201	0.35	53907	94.93
788	204	0.36	54111	95.28
789	179	0.32	54290	95.60
790	146	0.26	54436	95.86
791	151	0.27	54587	96.12
792	150	0.26	54737	96.39
793	136	0.24	54873	96.63
794	131	0.23	55004	96.86
795	119	0.21	55123	97.07
796	118	0.21	55241	97.27
797	115	0.20	55356	97.48
798	82	0.14	55438	97.62
799	92	0.16	55530	97.78
800	92	0.16	55622	97.95
801	82	0.14	55704	98.09
802	73	0.13	55777	98.22
803	77	0.14	55854	98.35
804	66	0.12	55920	98.47
805	57	0.10	55977	98.57
806	57	0.10	56034	98.67
807	50	0.09	56084	98.76
808	58	0.10	56142	98.86
809	43	0.08	56185	98.94
810	34	0.06	56219	99.00
811	44	0.08	56263	99.07
812	30	0.05	56293	99.13
813	30	0.05	56323	99.18
814	33	0.06	56356	99.24
815	20	0.04	56376	99.27
816	19	0.03	56395	99.31
817	28	0.05	56423	99.36
818	25	0.04	56448	99.40
819	16	0.03	56464	99.43
820	19	0.03	56483	99.46
821	12	0.02	56495	99.48
822	24	0.04	56519	99.52
823	7	0.01	56526	99.54
824	12	0.02	56538	99.56
825	21	0.04	56559	99.59
826	12	0.02	56571	99.62
827	10	0.02	56581	99.63
828	9	0.02	56590	99.65
829	7	0.01	56597	99.66
830	3	0.01	56600	99.67
831	7	0.01	56607	99.68
832	11	0.02	56618	99.70
833	8	0.01	56626	99.71
834	17	0.03	56643	99.74
835	3	0.01	56646	99.75
836	7	0.01	56653	99.76
837	3	0.01	56656	99.77
838	4	0.01	56660	99.77
839	16	0.03	56676	99.80
840	1	0.00	56677	99.80
842	14	0.02	56691	99.83
844	2	0.00	56693	99.83
846	1	0.00	56694	99.83
847	5	0.01	56699	99.84
849	5	0.01	56704	99.85
850	85	0.15	56789	100.00

Table F.3. Scale Score Distribution—Mathematics Grade 5

SS	Freq.	%	Cum. Freq.	Cum. %
650	218	0.38	218	0.38
651	6	0.01	224	0.39
652	27	0.05	251	0.44
653	34	0.06	285	0.50
654	35	0.06	320	0.56
655	55	0.10	375	0.66
656	49	0.09	424	0.75
657	26	0.05	450	0.79
658	33	0.06	483	0.85
659	51	0.09	534	0.94
660	13	0.02	547	0.96
661	38	0.07	585	1.03
662	27	0.05	612	1.08
663	16	0.03	628	1.10
664	45	0.08	673	1.18
665	27	0.05	700	1.23
666	64	0.11	764	1.34
667	88	0.15	852	1.50
668	55	0.10	907	1.59
669	120	0.21	1027	1.81
670	87	0.15	1114	1.96
671	83	0.15	1197	2.10
672	89	0.16	1286	2.26
673	76	0.13	1362	2.39
674	103	0.18	1465	2.57
675	106	0.19	1571	2.76
676	165	0.29	1736	3.05
677	160	0.28	1896	3.33
678	175	0.31	2071	3.64
679	179	0.31	2250	3.95
680	149	0.26	2399	4.22
681	173	0.30	2572	4.52
682	190	0.33	2762	4.85
683	233	0.41	2995	5.26
684	226	0.40	3221	5.66
685	241	0.42	3462	6.08
686	271	0.48	3733	6.56
687	254	0.45	3987	7.01
688	310	0.54	4297	7.55
689	323	0.57	4620	8.12
690	323	0.57	4943	8.69
691	343	0.60	5286	9.29
692	381	0.67	5667	9.96
693	410	0.72	6077	10.68
694	352	0.62	6429	11.30
695	397	0.70	6826	12.00
696	441	0.78	7267	12.77
697	436	0.77	7703	13.54
698	473	0.83	8176	14.37
699	457	0.80	8633	15.17
700	442	0.78	9075	15.95
701	500	0.88	9575	16.83
702	513	0.90	10088	17.73
703	483	0.85	10571	18.58
704	516	0.91	11087	19.49
705	503	0.88	11590	20.37
706	512	0.90	12102	21.27
707	501	0.88	12603	22.15
708	528	0.93	13131	23.08
709	577	1.01	13708	24.09
710	528	0.93	14236	25.02
711	520	0.91	14756	25.94
712	531	0.93	15287	26.87
713	546	0.96	15833	27.83
714	558	0.98	16391	28.81
715	553	0.97	16944	29.78
716	544	0.96	17488	30.74
717	599	1.05	18087	31.79
718	526	0.92	18613	32.71
719	546	0.96	19159	33.67
720	601	1.06	19760	34.73
721	558	0.98	20318	35.71
722	562	0.99	20880	36.70
723	554	0.97	21434	37.67
724	561	0.99	21995	38.66
725	571	1.00	22566	39.66
726	557	0.98	23123	40.64
727	562	0.99	23685	41.63
728	590	1.04	24275	42.67
729	589	1.04	24864	43.70
730	598	1.05	25462	44.75
731	599	1.05	26061	45.80
732	572	1.01	26633	46.81
733	566	0.99	27199	47.80
734	637	1.12	27836	48.92
735	565	0.99	28401	49.92
736	587	1.03	28988	50.95
737	571	1.00	29559	51.95
738	577	1.01	30136	52.97
739	560	0.98	30696	53.95
740	548	0.96	31244	54.91
741	539	0.95	31783	55.86
742	549	0.96	32332	56.83
743	539	0.95	32871	57.77
744	572	1.01	33443	58.78
745	530	0.93	33973	59.71
746	577	1.01	34550	60.72
747	531	0.93	35081	61.66
748	519	0.91	35600	62.57
749	536	0.94	36136	63.51

SS	Freq.	%	Cum. Freq.	Cum. %
750	535	0.94	36671	64.45
751	502	0.88	37173	65.33
752	523	0.92	37696	66.25
753	528	0.93	38224	67.18
754	510	0.90	38734	68.08
755	490	0.86	39224	68.94
756	496	0.87	39720	69.81
757	488	0.86	40208	70.67
758	495	0.87	40703	71.54
759	484	0.85	41187	72.39
760	445	0.78	41632	73.17
761	454	0.80	42086	73.97
762	476	0.84	42562	74.81
763	506	0.89	43068	75.70
764	477	0.84	43545	76.53
765	443	0.78	43988	77.31
766	465	0.82	44453	78.13
767	437	0.77	44890	78.90
768	452	0.79	45342	79.69
769	469	0.82	45811	80.52
770	449	0.79	46260	81.31
771	390	0.69	46650	81.99
772	413	0.73	47063	82.72
773	398	0.70	47461	83.42
774	366	0.64	47827	84.06
775	367	0.65	48194	84.71
776	344	0.60	48538	85.31
777	391	0.69	48929	86.00
778	365	0.64	49294	86.64
779	357	0.63	49651	87.27
780	350	0.62	50001	87.88
781	356	0.63	50357	88.51
782	310	0.54	50667	89.05
783	297	0.52	50964	89.57
784	332	0.58	51296	90.16
785	290	0.51	51586	90.67
786	281	0.49	51867	91.16
787	281	0.49	52148	91.65
788	277	0.49	52425	92.14
789	237	0.42	52662	92.56
790	247	0.43	52909	92.99
791	231	0.41	53140	93.40
792	226	0.40	53366	93.80
793	202	0.36	53568	94.15
794	234	0.41	53802	94.56
795	200	0.35	54002	94.91
796	168	0.30	54170	95.21
797	197	0.35	54367	95.56
798	185	0.33	54552	95.88
799	166	0.29	54718	96.17
800	140	0.25	54858	96.42
801	161	0.28	55019	96.70
802	136	0.24	55155	96.94
803	124	0.22	55279	97.16
804	117	0.21	55396	97.36
805	112	0.20	55508	97.56
806	119	0.21	55627	97.77
807	111	0.20	55738	97.96
808	77	0.14	55815	98.10
809	78	0.14	55893	98.24
810	84	0.15	55977	98.38
811	61	0.11	56038	98.49
812	91	0.16	56129	98.65
813	72	0.13	56201	98.78
814	48	0.08	56249	98.86
815	58	0.10	56307	98.96
816	68	0.12	56375	99.08
817	30	0.05	56405	99.14
818	38	0.07	56443	99.20
819	29	0.05	56472	99.25
820	32	0.06	56504	99.31
821	23	0.04	56527	99.35
822	22	0.04	56549	99.39
823	30	0.05	56579	99.44
824	30	0.05	56609	99.50
825	24	0.04	56633	99.54
826	16	0.03	56649	99.57
827	24	0.04	56673	99.61
828	17	0.03	56690	99.64
829	16	0.03	56706	99.67
830	6	0.01	56712	99.68
831	14	0.02	56726	99.70
832	12	0.02	56738	99.72
833	7	0.01	56745	99.73
834	11	0.02	56756	99.75
835	10	0.02	56766	99.77
836	14	0.02	56780	99.80
837	17	0.03	56797	99.83
838	17	0.03	56814	99.86
839	2	0.00	56816	99.86
840	3	0.01	56819	99.86
841	7	0.01	56826	99.88
842	3	0.01	56829	99.88
843	4	0.01	56833	99.89
844	4	0.01	56837	99.90
846	1	0.00	56838	99.90
847	2	0.00	56840	99.90
850	56	0.10	56896	100.00

Table F.4. Scale Score Distribution—Mathematics Grade 6

SS	Freq.	%	Cum. Freq.	Cum. %
650	444	0.79	444	0.79
651	23	0.04	467	0.84
652	2	0.00	469	0.84
653	35	0.06	504	0.90
654	3	0.01	507	0.91
655	132	0.24	639	1.14
656	25	0.04	664	1.19
657	26	0.05	690	1.23
658	55	0.10	745	1.33
659	21	0.04	766	1.37
660	9	0.02	775	1.39
661	53	0.09	828	1.48
662	142	0.25	970	1.73
663	37	0.07	1007	1.80
664	94	0.17	1101	1.97
665	48	0.09	1149	2.05
666	69	0.12	1218	2.18
667	170	0.30	1388	2.48
668	66	0.12	1454	2.60
669	125	0.22	1579	2.82
670	111	0.20	1690	3.02
671	81	0.14	1771	3.17
672	167	0.30	1938	3.47
673	85	0.15	2023	3.62
674	223	0.40	2246	4.02
675	155	0.28	2401	4.29
676	181	0.32	2582	4.62
677	138	0.25	2720	4.86
678	268	0.48	2988	5.34
679	210	0.38	3198	5.72
680	223	0.40	3421	6.12
681	353	0.63	3774	6.75
682	272	0.49	4046	7.24
683	307	0.55	4353	7.79
684	336	0.60	4689	8.39
685	335	0.60	5024	8.99
686	362	0.65	5386	9.63
687	339	0.61	5725	10.24
688	384	0.69	6109	10.93
689	427	0.76	6536	11.69
690	394	0.70	6930	12.39
691	419	0.75	7349	13.14
692	412	0.74	7761	13.88
693	426	0.76	8187	14.64
694	472	0.84	8659	15.49
695	465	0.83	9124	16.32
696	487	0.87	9611	17.19
697	492	0.88	10103	18.07
698	507	0.91	10610	18.98
699	535	0.96	11145	19.93
700	527	0.94	11672	20.88
701	494	0.88	12166	21.76
702	520	0.93	12686	22.69
703	547	0.98	13233	23.67
704	562	1.01	13795	24.67
705	553	0.99	14348	25.66
706	534	0.96	14882	26.62
707	555	0.99	15437	27.61
708	567	1.01	16004	28.62
709	514	0.92	16518	29.54
710	572	1.02	17090	30.57
711	579	1.04	17669	31.60
712	528	0.94	18197	32.55
713	553	0.99	18750	33.53
714	573	1.02	19323	34.56
715	555	0.99	19878	35.55
716	537	0.96	20415	36.51
717	589	1.05	21004	37.57
718	605	1.08	21609	38.65
719	549	0.98	22158	39.63
720	538	0.96	22696	40.59
721	568	1.02	23264	41.61
722	583	1.04	23847	42.65
723	604	1.08	24451	43.73
724	592	1.06	25043	44.79
725	666	1.19	25709	45.98
726	569	1.02	26278	47.00
727	581	1.04	26859	48.04
728	589	1.05	27448	49.09
729	599	1.07	28047	50.16
730	625	1.12	28672	51.28
731	600	1.07	29272	52.35
732	626	1.12	29898	53.47
733	623	1.11	30521	54.59
734	621	1.11	31142	55.70
735	593	1.06	31735	56.76
736	656	1.17	32391	57.93
737	606	1.08	32997	59.01
738	616	1.10	33613	60.12
739	636	1.14	34249	61.25
740	629	1.12	34878	62.38
741	613	1.10	35491	63.48
742	571	1.02	36062	64.50
743	608	1.09	36670	65.58
744	625	1.12	37295	66.70
745	585	1.05	37880	67.75
746	602	1.08	38482	68.82
747	553	0.99	39035	69.81
748	536	0.96	39571	70.77
749	554	0.99	40125	71.76

SS	Freq.	%	Cum. Freq.	Cum. %
750	577	1.03	40702	72.80
751	519	0.93	41221	73.72
752	547	0.98	41768	74.70
753	514	0.92	42282	75.62
754	512	0.92	42794	76.54
755	513	0.92	43307	77.45
756	500	0.89	43807	78.35
757	519	0.93	44326	79.28
758	478	0.85	44804	80.13
759	504	0.90	45308	81.03
760	493	0.88	45801	81.91
761	451	0.81	46252	82.72
762	426	0.76	46678	83.48
763	397	0.71	47075	84.19
764	441	0.79	47516	84.98
765	396	0.71	47912	85.69
766	385	0.69	48297	86.38
767	380	0.68	48677	87.06
768	358	0.64	49035	87.70
769	349	0.62	49384	88.32
770	308	0.55	49692	88.87
771	289	0.52	49981	89.39
772	313	0.56	50294	89.95
773	324	0.58	50618	90.53
774	312	0.56	50930	91.09
775	291	0.52	51221	91.61
776	245	0.44	51466	92.05
777	228	0.41	51694	92.45
778	251	0.45	51945	92.90
779	224	0.40	52169	93.30
780	214	0.38	52383	93.69
781	224	0.40	52607	94.09
782	164	0.29	52771	94.38
783	188	0.34	52959	94.72
784	168	0.30	53127	95.02
785	158	0.28	53285	95.30
786	185	0.33	53470	95.63
787	168	0.30	53638	95.93
788	144	0.26	53782	96.19
789	126	0.23	53908	96.41
790	130	0.23	54038	96.65
791	109	0.19	54147	96.84
792	116	0.21	54263	97.05
793	111	0.20	54374	97.25
794	88	0.16	54462	97.40
795	100	0.18	54562	97.58
796	81	0.14	54643	97.73
797	96	0.17	54739	97.90
798	84	0.15	54823	98.05
799	76	0.14	54899	98.19
800	72	0.13	54971	98.32
801	64	0.11	55035	98.43
802	61	0.11	55096	98.54
803	55	0.10	55151	98.64
804	59	0.11	55210	98.74
805	45	0.08	55255	98.82
806	44	0.08	55299	98.90
807	33	0.06	55332	98.96
808	30	0.05	55362	99.01
809	31	0.06	55393	99.07
810	32	0.06	55425	99.13
811	37	0.07	55462	99.19
812	35	0.06	55497	99.26
813	22	0.04	55519	99.30
814	23	0.04	55542	99.34
815	35	0.06	55577	99.40
816	27	0.05	55604	99.45
817	20	0.04	55624	99.48
818	19	0.03	55643	99.52
819	27	0.05	55670	99.57
820	14	0.03	55684	99.59
821	15	0.03	55699	99.62
822	20	0.04	55719	99.65
823	6	0.01	55725	99.66
824	7	0.01	55732	99.68
825	11	0.02	55743	99.70
826	9	0.02	55752	99.71
827	7	0.01	55759	99.72
828	7	0.01	55766	99.74
829	9	0.02	55775	99.75
830	10	0.02	55785	99.77
831	1	0.00	55786	99.77
832	10	0.02	55796	99.79
833	4	0.01	55800	99.80
834	4	0.01	55804	99.81
835	2	0.00	55806	99.81
836	9	0.02	55815	99.82
837	3	0.01	55818	99.83
838	3	0.01	55821	99.84
839	5	0.01	55826	99.84
840	5	0.01	55831	99.85
842	9	0.02	55840	99.87
843	9	0.02	55849	99.89
844	2	0.00	55851	99.89
846	4	0.01	55855	99.90
847	2	0.00	55857	99.90
848	1	0.00	55858	99.90
849	1	0.00	55859	99.90
850	54	0.10	55913	100.00

Table F.5. Scale Score Distribution—Mathematics Grade 7

SS	Freq.	%	Cum. Freq.	Cum. %
650	151	0.28	151	0.28
651	19	0.04	170	0.31
653	3	0.01	173	0.32
654	17	0.03	190	0.35
655	30	0.06	220	0.41
656	5	0.01	225	0.42
657	21	0.04	246	0.45
658	6	0.01	252	0.47
659	30	0.06	282	0.52
660	20	0.04	302	0.56
661	4	0.01	306	0.57
662	36	0.07	342	0.63
663	8	0.01	350	0.65
664	6	0.01	356	0.66
665	18	0.03	374	0.69
666	36	0.07	410	0.76
667	47	0.09	457	0.84
668	26	0.05	483	0.89
669	40	0.07	523	0.97
670	32	0.06	555	1.02
671	63	0.12	618	1.14
672	73	0.13	691	1.28
673	61	0.11	752	1.39
674	27	0.05	779	1.44
675	69	0.13	848	1.57
676	78	0.14	926	1.71
677	71	0.13	997	1.84
678	69	0.13	1066	1.97
679	104	0.19	1170	2.16
680	98	0.18	1268	2.34
681	107	0.20	1375	2.54
682	140	0.26	1515	2.80
683	138	0.25	1653	3.05
684	143	0.26	1796	3.32
685	188	0.35	1984	3.66
686	202	0.37	2186	4.04
687	205	0.38	2391	4.42
688	238	0.44	2629	4.86
689	244	0.45	2873	5.31
690	242	0.45	3115	5.75
691	281	0.52	3396	6.27
692	314	0.58	3710	6.85
693	336	0.62	4046	7.47
694	370	0.68	4416	8.16
695	387	0.71	4803	8.87
696	391	0.72	5194	9.59
697	422	0.78	5616	10.37
698	435	0.80	6051	11.17
699	502	0.93	6553	12.10
700	530	0.98	7083	13.08
701	574	1.06	7657	14.14
702	580	1.07	8237	15.21
703	616	1.14	8853	16.35
704	667	1.23	9520	17.58
705	611	1.13	10131	18.71
706	694	1.28	10825	19.99
707	670	1.24	11495	21.23
708	675	1.25	12170	22.48
709	688	1.27	12858	23.75
710	762	1.41	13620	25.15
711	759	1.40	14379	26.55
712	728	1.34	15107	27.90
713	737	1.36	15844	29.26
714	745	1.38	16589	30.64
715	712	1.31	17301	31.95
716	731	1.35	18032	33.30
717	744	1.37	18776	34.68
718	702	1.30	19478	35.97
719	754	1.39	20232	37.36
720	735	1.36	20967	38.72
721	744	1.37	21711	40.10
722	705	1.30	22416	41.40
723	701	1.29	23117	42.69
724	737	1.36	23854	44.05
725	722	1.33	24576	45.39
726	718	1.33	25294	46.71
727	667	1.23	25961	47.94
728	676	1.25	26637	49.19
729	656	1.21	27293	50.40
730	688	1.27	27981	51.68
731	688	1.27	28669	52.95
732	700	1.29	29369	54.24
733	632	1.17	30001	55.41
734	618	1.14	30619	56.55
735	654	1.21	31273	57.75
736	595	1.10	31868	58.85
737	661	1.22	32529	60.07
738	645	1.19	33174	61.27
739	670	1.24	33844	62.50
740	632	1.17	34476	63.67
741	632	1.17	35108	64.84
742	615	1.14	35723	65.97
743	618	1.14	36341	67.11
744	612	1.13	36953	68.24
745	594	1.10	37547	69.34
746	595	1.10	38142	70.44
747	605	1.12	38747	71.56
748	595	1.10	39342	72.66
749	589	1.09	39931	73.74
750	565	1.04	40496	74.79

SS	Freq.	%	Cum. Freq.	Cum. %
751	541	1.00	41037	75.79
752	559	1.03	41596	76.82
753	534	0.99	42130	77.81
754	535	0.99	42665	78.79
755	558	1.03	43223	79.82
756	468	0.86	43691	80.69
757	490	0.90	44181	81.59
758	495	0.91	44676	82.51
759	438	0.81	45114	83.32
760	469	0.87	45583	84.18
761	447	0.83	46030	85.01
762	447	0.83	46477	85.83
763	443	0.82	46920	86.65
764	401	0.74	47321	87.39
765	404	0.75	47725	88.14
766	347	0.64	48072	88.78
767	356	0.66	48428	89.44
768	363	0.67	48791	90.11
769	312	0.58	49103	90.68
770	307	0.57	49410	91.25
771	281	0.52	49691	91.77
772	268	0.49	49959	92.26
773	279	0.52	50238	92.78
774	235	0.43	50473	93.21
775	245	0.45	50718	93.67
776	248	0.46	50966	94.12
777	231	0.43	51197	94.55
778	196	0.36	51393	94.91
779	218	0.40	51611	95.31
780	199	0.37	51810	95.68
781	204	0.38	52014	96.06
782	167	0.31	52181	96.37
783	180	0.33	52361	96.70
784	143	0.26	52504	96.96
785	131	0.24	52635	97.21
786	122	0.23	52757	97.43
787	108	0.20	52865	97.63
788	112	0.21	52977	97.84
789	98	0.18	53075	98.02
790	101	0.19	53176	98.20
791	75	0.14	53251	98.34
792	87	0.16	53338	98.50
793	81	0.15	53419	98.65
794	58	0.11	53477	98.76
795	57	0.11	53534	98.87
796	48	0.09	53582	98.95
797	71	0.13	53653	99.09
798	47	0.09	53700	99.17
799	49	0.09	53749	99.26
800	30	0.06	53779	99.32
801	27	0.05	53806	99.37
802	33	0.06	53839	99.43
803	38	0.07	53877	99.50
804	25	0.05	53902	99.55
805	22	0.04	53924	99.59
806	17	0.03	53941	99.62
807	25	0.05	53966	99.66
808	21	0.04	53987	99.70
809	6	0.01	53993	99.71
810	9	0.02	54002	99.73
811	18	0.03	54020	99.76
812	12	0.02	54032	99.79
813	7	0.01	54039	99.80
814	6	0.01	54045	99.81
815	5	0.01	54050	99.82
816	5	0.01	54055	99.83
817	9	0.02	54064	99.84
818	7	0.01	54071	99.86
819	9	0.02	54080	99.87
820	3	0.01	54083	99.88
821	4	0.01	54087	99.89
822	3	0.01	54090	99.89
823	7	0.01	54097	99.91
824	2	0.00	54099	99.91
825	4	0.01	54103	99.92
826	6	0.01	54109	99.93
827	2	0.00	54111	99.93
828	4	0.01	54115	99.94
830	1	0.00	54116	99.94
831	7	0.01	54123	99.95
832	1	0.00	54124	99.96
833	4	0.01	54128	99.96
834	1	0.00	54129	99.96
838	1	0.00	54130	99.97
839	1	0.00	54131	99.97
841	1	0.00	54132	99.97
843	1	0.00	54133	99.97
845	1	0.00	54134	99.97
846	1	0.00	54135	99.98
847	1	0.00	54136	99.98
850	12	0.02	54148	100.00

Table F.6. Scale Score Distribution—Mathematics Grade 8

SS	Freq.	%	Cum. Freq.	Cum. %
650	645	1.24	645	1.24
651	40	0.08	685	1.32
652	31	0.06	716	1.38
653	52	0.10	768	1.48
654	77	0.15	845	1.62
655	52	0.10	897	1.72
656	75	0.14	972	1.87
657	97	0.19	1069	2.05
658	64	0.12	1133	2.18
659	54	0.10	1187	2.28
660	79	0.15	1266	2.43
661	86	0.17	1352	2.60
662	88	0.17	1440	2.77
663	102	0.20	1542	2.96
664	134	0.26	1676	3.22
665	139	0.27	1815	3.49
666	115	0.22	1930	3.71
667	146	0.28	2076	3.99
668	135	0.26	2211	4.25
669	152	0.29	2363	4.54
670	164	0.32	2527	4.86
671	168	0.32	2695	5.18
672	173	0.33	2868	5.51
673	184	0.35	3052	5.87
674	238	0.46	3290	6.32
675	216	0.42	3506	6.74
676	199	0.38	3705	7.12
677	257	0.49	3962	7.61
678	217	0.42	4179	8.03
679	253	0.49	4432	8.52
680	319	0.61	4751	9.13
681	288	0.55	5039	9.68
682	247	0.47	5286	10.16
683	371	0.71	5657	10.87
684	346	0.66	6003	11.54
685	353	0.68	6356	12.21
686	338	0.65	6694	12.86
687	370	0.71	7064	13.58
688	381	0.73	7445	14.31
689	415	0.80	7860	15.10
690	423	0.81	8283	15.92
691	412	0.79	8695	16.71
692	453	0.87	9148	17.58
693	437	0.84	9585	18.42
694	468	0.90	10053	19.32
695	466	0.90	10519	20.21
696	433	0.83	10952	21.05
697	465	0.89	11417	21.94
698	438	0.84	11855	22.78
699	500	0.96	12355	23.74
700	494	0.95	12849	24.69
701	501	0.96	13350	25.66
702	475	0.91	13825	26.57
703	478	0.92	14303	27.49
704	526	1.01	14829	28.50
705	482	0.93	15311	29.42
706	492	0.95	15803	30.37
707	471	0.91	16274	31.27
708	551	1.06	16825	32.33
709	482	0.93	17307	33.26
710	524	1.01	17831	34.27
711	524	1.01	18355	35.27
712	468	0.90	18823	36.17
713	483	0.93	19306	37.10
714	521	1.00	19827	38.10
715	500	0.96	20327	39.06
716	464	0.89	20791	39.96
717	491	0.94	21282	40.90
718	430	0.83	21712	41.72
719	451	0.87	22163	42.59
720	474	0.91	22637	43.50
721	471	0.91	23108	44.41
722	475	0.91	23583	45.32
723	454	0.87	24037	46.19
724	426	0.82	24463	47.01
725	450	0.86	24913	47.88
726	436	0.84	25349	48.71
727	452	0.87	25801	49.58
728	412	0.79	26213	50.37
729	460	0.88	26673	51.26
730	450	0.86	27123	52.12
731	447	0.86	27570	52.98
732	415	0.80	27985	53.78
733	426	0.82	28411	54.60
734	398	0.76	28809	55.36
735	420	0.81	29229	56.17
736	434	0.83	29663	57.00
737	431	0.83	30094	57.83
738	416	0.80	30510	58.63
739	426	0.82	30936	59.45
740	432	0.83	31368	60.28
741	424	0.81	31792	61.10
742	406	0.78	32198	61.88
743	454	0.87	32652	62.75
744	402	0.77	33054	63.52
745	395	0.76	33449	64.28
746	382	0.73	33831	65.01
747	407	0.78	34238	65.80
748	409	0.79	34647	66.58
749	391	0.75	35038	67.33

SS	Freq.	%	Cum. Freq.	Cum. %
750	366	0.70	35404	68.04
751	407	0.78	35811	68.82
752	365	0.70	36176	69.52
753	389	0.75	36565	70.27
754	371	0.71	36936	70.98
755	388	0.75	37324	71.73
756	372	0.71	37696	72.44
757	367	0.71	38063	73.15
758	387	0.74	38450	73.89
759	350	0.67	38800	74.56
760	390	0.75	39190	75.31
761	374	0.72	39564	76.03
762	339	0.65	39903	76.68
763	345	0.66	40248	77.35
764	345	0.66	40593	78.01
765	297	0.57	40890	78.58
766	333	0.64	41223	79.22
767	299	0.57	41522	79.79
768	322	0.62	41844	80.41
769	310	0.60	42154	81.01
770	325	0.62	42479	81.63
771	306	0.59	42785	82.22
772	271	0.52	43056	82.74
773	287	0.55	43343	83.29
774	271	0.52	43614	83.82
775	322	0.62	43936	84.43
776	277	0.53	44213	84.97
777	265	0.51	44478	85.48
778	258	0.50	44736	85.97
779	244	0.47	44980	86.44
780	261	0.50	45241	86.94
781	243	0.47	45484	87.41
782	248	0.48	45732	87.89
783	212	0.41	45944	88.29
784	211	0.41	46155	88.70
785	246	0.47	46401	89.17
786	208	0.40	46609	89.57
787	209	0.40	46818	89.97
788	203	0.39	47021	90.36
789	221	0.42	47242	90.79
790	206	0.40	47448	91.18
791	172	0.33	47620	91.51
792	180	0.35	47800	91.86
793	188	0.36	47988	92.22
794	170	0.33	48158	92.55
795	160	0.31	48318	92.85
796	171	0.33	48489	93.18
797	146	0.28	48635	93.46
798	161	0.31	48796	93.77
799	147	0.28	48943	94.06
800	136	0.26	49079	94.32
801	132	0.25	49211	94.57
802	130	0.25	49341	94.82
803	115	0.22	49456	95.04
804	111	0.21	49567	95.26
805	121	0.23	49688	95.49
806	124	0.24	49812	95.73
807	90	0.17	49902	95.90
808	116	0.22	50018	96.12
809	91	0.17	50109	96.30
810	86	0.17	50195	96.46
811	95	0.18	50290	96.64
812	77	0.15	50367	96.79
813	84	0.16	50451	96.95
814	82	0.16	50533	97.11
815	63	0.12	50596	97.23
816	67	0.13	50663	97.36
817	61	0.12	50724	97.48
818	64	0.12	50788	97.60
819	65	0.12	50853	97.73
820	65	0.12	50918	97.85
821	64	0.12	50982	97.97
822	54	0.10	51036	98.08
823	44	0.08	51080	98.16
824	69	0.13	51149	98.30
825	48	0.09	51197	98.39
826	45	0.09	51242	98.47
827	48	0.09	51290	98.57
828	24	0.05	51314	98.61
829	36	0.07	51350	98.68
830	40	0.08	51390	98.76
831	42	0.08	51432	98.84
832	29	0.06	51461	98.89
833	27	0.05	51488	98.95
834	34	0.07	51522	99.01
835	24	0.05	51546	99.06
836	30	0.06	51576	99.12
837	14	0.03	51590	99.14
838	18	0.03	51608	99.18
839	34	0.07	51642	99.24
840	20	0.04	51662	99.28
841	19	0.04	51681	99.32
842	18	0.03	51699	99.35
843	27	0.05	51726	99.40
844	10	0.02	51736	99.42
845	15	0.03	51751	99.45
846	23	0.04	51774	99.50
847	10	0.02	51784	99.52
848	11	0.02	51795	99.54
849	15	0.03	51810	99.57
850	226	0.43	52036	100.00

Table F.7. Scale Score Distribution—ELA Grade 3

SS	Freq.	%	Cum. Freq.	Cum. %
650	882	1.58	882	1.58
651	82	0.15	964	1.73
652	78	0.14	1042	1.87
653	72	0.13	1114	2.00
654	81	0.15	1195	2.14
655	71	0.13	1266	2.27
656	88	0.16	1354	2.43
657	96	0.17	1450	2.60
658	108	0.19	1558	2.80
659	110	0.20	1668	2.99
660	127	0.23	1795	3.22
661	131	0.24	1926	3.46
662	140	0.25	2066	3.71
663	151	0.27	2217	3.98
664	154	0.28	2371	4.25
665	167	0.30	2538	4.55
666	170	0.31	2708	4.86
667	174	0.31	2882	5.17
668	225	0.40	3107	5.57
669	197	0.35	3304	5.93
670	199	0.36	3503	6.28
671	233	0.42	3736	6.70
672	224	0.40	3960	7.10
673	246	0.44	4206	7.55
674	241	0.43	4447	7.98
675	276	0.50	4723	8.47
676	260	0.47	4983	8.94
677	283	0.51	5266	9.45
678	294	0.53	5560	9.98
679	296	0.53	5856	10.51
680	292	0.52	6148	11.03
681	314	0.56	6462	11.59
682	310	0.56	6772	12.15
683	304	0.55	7076	12.70
684	318	0.57	7394	13.27
685	324	0.58	7718	13.85
686	323	0.58	8041	14.43
687	317	0.57	8358	15.00
688	302	0.54	8660	15.54
689	345	0.62	9005	16.16
690	314	0.56	9319	16.72
691	305	0.55	9624	17.27
692	331	0.59	9955	17.86
693	330	0.59	10285	18.45
694	353	0.63	10638	19.09
695	351	0.63	10989	19.72
696	371	0.67	11360	20.38
697	310	0.56	11670	20.94
698	327	0.59	11997	21.52
699	373	0.67	12370	22.19
700	326	0.58	12696	22.78
701	343	0.62	13039	23.39
702	324	0.58	13363	23.98
703	369	0.66	13732	24.64
704	383	0.69	14115	25.32
705	336	0.60	14451	25.93
706	338	0.61	14789	26.53
707	338	0.61	15127	27.14
708	362	0.65	15489	27.79
709	368	0.66	15857	28.45
710	369	0.66	16226	29.11
711	365	0.65	16591	29.77
712	387	0.69	16978	30.46
713	381	0.68	17359	31.14
714	381	0.68	17740	31.83
715	390	0.70	18130	32.53
716	369	0.66	18499	33.19
717	406	0.73	18905	33.92
718	431	0.77	19336	34.69
719	444	0.80	19780	35.49
720	405	0.73	20185	36.21
721	415	0.74	20600	36.96
722	403	0.72	21003	37.68
723	446	0.80	21449	38.48
724	411	0.74	21860	39.22
725	447	0.80	22307	40.02
726	435	0.78	22742	40.80
727	471	0.85	23213	41.65
728	455	0.82	23668	42.46
729	444	0.80	24112	43.26
730	462	0.83	24574	44.09
731	464	0.83	25038	44.92
732	439	0.79	25477	45.71
733	459	0.82	25936	46.53
734	481	0.86	26417	47.40
735	466	0.84	26883	48.23
736	442	0.79	27325	49.02
737	494	0.89	27819	49.91
738	476	0.85	28295	50.77
739	455	0.82	28750	51.58
740	440	0.79	29190	52.37
741	481	0.86	29671	53.23
742	505	0.91	30176	54.14
743	485	0.87	30661	55.01
744	481	0.86	31142	55.87
745	463	0.83	31605	56.70
746	510	0.92	32115	57.62
747	482	0.86	32597	58.48
748	462	0.83	33059	59.31
749	453	0.81	33512	60.13

SS	Freq.	%	Cum. Freq.	Cum. %
750	474	0.85	33986	60.98
751	466	0.84	34452	61.81
752	432	0.78	34884	62.59
753	458	0.82	35342	63.41
754	461	0.83	35803	64.24
755	466	0.84	36269	65.07
756	449	0.81	36718	65.88
757	420	0.75	37138	66.63
758	445	0.80	37583	67.43
759	450	0.81	38033	68.24
760	402	0.72	38435	68.96
761	418	0.75	38853	69.71
762	457	0.82	39310	70.53
763	431	0.77	39741	71.30
764	412	0.74	40153	72.04
765	400	0.72	40553	72.76
766	404	0.72	40957	73.48
767	424	0.76	41381	74.24
768	378	0.68	41759	74.92
769	394	0.71	42153	75.63
770	396	0.71	42549	76.34
771	367	0.66	42916	77.00
772	393	0.71	43309	77.70
773	410	0.74	43719	78.44
774	421	0.76	44140	79.19
775	362	0.65	44502	79.84
776	356	0.64	44858	80.48
777	374	0.67	45232	81.15
778	309	0.55	45541	81.71
779	329	0.59	45870	82.30
780	364	0.65	46234	82.95
781	337	0.60	46571	83.55
782	320	0.57	46891	84.13
783	329	0.59	47220	84.72
784	303	0.54	47523	85.26
785	306	0.55	47829	85.81
786	275	0.49	48104	86.31
787	286	0.51	48390	86.82
788	278	0.50	48668	87.32
789	301	0.54	48969	87.86
790	272	0.49	49241	88.35
791	254	0.46	49495	88.80
792	268	0.48	49763	89.28
793	286	0.51	50049	89.79
794	256	0.46	50305	90.25
795	220	0.39	50525	90.65
796	246	0.44	50771	91.09
797	206	0.37	50977	91.46
798	238	0.43	51215	91.89
799	191	0.34	51406	92.23
800	204	0.37	51610	92.60
801	162	0.29	51772	92.89
802	171	0.31	51943	93.19
803	156	0.28	52099	93.47
804	167	0.30	52266	93.77
805	147	0.26	52413	94.04
806	145	0.26	52558	94.30
807	131	0.24	52689	94.53
808	147	0.26	52836	94.80
809	141	0.25	52977	95.05
810	121	0.22	53098	95.27
811	156	0.28	53254	95.55
812	103	0.18	53357	95.73
813	116	0.21	53473	95.94
814	147	0.26	53620	96.20
815	83	0.15	53703	96.35
816	106	0.19	53809	96.54
817	73	0.13	53882	96.67
818	111	0.20	53993	96.87
819	91	0.16	54084	97.03
820	59	0.11	54143	97.14
821	120	0.22	54263	97.36
822	71	0.13	54334	97.48
823	73	0.13	54407	97.61
824	78	0.14	54485	97.75
825	65	0.12	54550	97.87
826	57	0.10	54607	97.97
827	54	0.10	54661	98.07
828	41	0.07	54702	98.14
829	60	0.11	54762	98.25
830	42	0.08	54804	98.33
831	41	0.07	54845	98.40
832	64	0.11	54909	98.51
833	56	0.10	54965	98.61
834	33	0.06	54998	98.67
835	51	0.09	55049	98.77
836	36	0.06	55085	98.83
837	23	0.04	55108	98.87
838	17	0.03	55125	98.90
839	36	0.06	55161	98.97
840	27	0.05	55188	99.02
841	39	0.07	55227	99.08
842	35	0.06	55262	99.15
843	18	0.03	55280	99.18
844	23	0.04	55303	99.22
845	31	0.06	55334	99.28
846	34	0.06	55368	99.34
847	14	0.03	55382	99.36
848	21	0.04	55403	99.40
849	12	0.02	55415	99.42
850	322	0.58	55737	100.00

Table F.8. Scale Score Distribution—ELA Grade 4

SS	Freq.	%	Cum. Freq.	Cum. %
650	455	0.82	455	0.82
651	34	0.06	489	0.88
652	39	0.07	528	0.95
653	25	0.05	553	1.00
654	26	0.05	579	1.04
655	30	0.05	609	1.10
656	46	0.08	655	1.18
657	62	0.11	717	1.29
658	57	0.10	774	1.39
659	58	0.10	832	1.50
660	64	0.12	896	1.61
661	68	0.12	964	1.74
662	72	0.13	1036	1.87
663	69	0.12	1105	1.99
664	77	0.14	1182	2.13
665	75	0.14	1257	2.26
666	85	0.15	1342	2.42
667	118	0.21	1460	2.63
668	108	0.19	1568	2.82
669	105	0.19	1673	3.01
670	115	0.21	1788	3.22
671	135	0.24	1923	3.46
672	125	0.23	2048	3.69
673	151	0.27	2199	3.96
674	131	0.24	2330	4.20
675	159	0.29	2489	4.48
676	157	0.28	2646	4.77
677	157	0.28	2803	5.05
678	183	0.33	2986	5.38
679	168	0.30	3154	5.68
680	223	0.40	3377	6.08
681	217	0.39	3594	6.47
682	185	0.33	3779	6.81
683	213	0.38	3992	7.19
684	206	0.37	4198	7.56
685	255	0.46	4453	8.02
686	219	0.39	4672	8.42
687	241	0.43	4913	8.85
688	244	0.44	5157	9.29
689	210	0.38	5367	9.67
690	260	0.47	5627	10.14
691	253	0.46	5880	10.59
692	255	0.46	6135	11.05
693	239	0.43	6374	11.48
694	265	0.48	6639	11.96
695	285	0.51	6924	12.47
696	314	0.57	7238	13.04
697	282	0.51	7520	13.54
698	273	0.49	7793	14.04
699	270	0.49	8063	14.52
700	260	0.47	8323	14.99
701	278	0.50	8601	15.49
702	275	0.50	8876	15.99
703	322	0.58	9198	16.57
704	277	0.50	9475	17.07
705	304	0.55	9779	17.61
706	363	0.65	10142	18.27
707	318	0.57	10460	18.84
708	328	0.59	10788	19.43
709	362	0.65	11150	20.08
710	382	0.69	11532	20.77
711	358	0.64	11890	21.42
712	355	0.64	12245	22.06
713	368	0.66	12613	22.72
714	362	0.65	12975	23.37
715	380	0.68	13355	24.05
716	384	0.69	13739	24.75
717	385	0.69	14124	25.44
718	397	0.72	14521	26.16
719	440	0.79	14961	26.95
720	421	0.76	15382	27.71
721	465	0.84	15847	28.54
722	454	0.82	16301	29.36
723	432	0.78	16733	30.14
724	476	0.86	17209	31.00
725	459	0.83	17668	31.82
726	476	0.86	18144	32.68
727	474	0.85	18618	33.53
728	504	0.91	19122	34.44
729	494	0.89	19616	35.33
730	531	0.96	20147	36.29
731	549	0.99	20696	37.28
732	540	0.97	21236	38.25
733	538	0.97	21774	39.22
734	572	1.03	22346	40.25
735	520	0.94	22866	41.19
736	536	0.97	23402	42.15
737	558	1.01	23960	43.16
738	569	1.02	24529	44.18
739	575	1.04	25104	45.22
740	598	1.08	25702	46.29
741	621	1.12	26323	47.41
742	569	1.02	26892	48.44
743	604	1.09	27496	49.53
744	625	1.13	28121	50.65
745	587	1.06	28708	51.71
746	630	1.13	29338	52.84
747	615	1.11	29953	53.95
748	619	1.11	30572	55.07
749	641	1.15	31213	56.22

SS	Freq.	%	Cum. Freq.	Cum. %
750	541	0.97	31754	57.19
751	615	1.11	32369	58.30
752	635	1.14	33004	59.45
753	647	1.17	33651	60.61
754	585	1.05	34236	61.67
755	614	1.11	34850	62.77
756	575	1.04	35425	63.81
757	597	1.08	36022	64.88
758	631	1.14	36653	66.02
759	609	1.10	37262	67.12
760	562	1.01	37824	68.13
761	619	1.11	38443	69.24
762	568	1.02	39011	70.27
763	573	1.03	39584	71.30
764	511	0.92	40095	72.22
765	555	1.00	40650	73.22
766	509	0.92	41159	74.13
767	541	0.97	41700	75.11
768	499	0.90	42199	76.01
769	586	1.06	42785	77.06
770	516	0.93	43301	77.99
771	504	0.91	43805	78.90
772	497	0.90	44302	79.80
773	509	0.92	44811	80.71
774	428	0.77	45239	81.48
775	425	0.77	45664	82.25
776	478	0.86	46142	83.11
777	451	0.81	46593	83.92
778	387	0.70	46980	84.62
779	412	0.74	47392	85.36
780	389	0.70	47781	86.06
781	354	0.64	48135	86.70
782	369	0.66	48504	87.36
783	331	0.60	48835	87.96
784	331	0.60	49166	88.56
785	315	0.57	49481	89.12
786	318	0.57	49799	89.70
787	284	0.51	50083	90.21
788	300	0.54	50383	90.75
789	292	0.53	50675	91.28
790	273	0.49	50948	91.77
791	286	0.52	51234	92.28
792	242	0.44	51476	92.72
793	248	0.45	51724	93.16
794	229	0.41	51953	93.58
795	211	0.38	52164	93.96
796	181	0.33	52345	94.28
797	179	0.32	52524	94.61
798	152	0.27	52676	94.88
799	175	0.32	52851	95.19
800	150	0.27	53001	95.46
801	139	0.25	53140	95.71
802	139	0.25	53279	95.97
803	149	0.27	53428	96.23
804	133	0.24	53561	96.47
805	122	0.22	53683	96.69
806	118	0.21	53801	96.91
807	110	0.20	53911	97.10
808	114	0.21	54025	97.31
809	94	0.17	54119	97.48
810	96	0.17	54215	97.65
811	82	0.15	54297	97.80
812	87	0.16	54384	97.96
813	64	0.12	54448	98.07
814	71	0.13	54519	98.20
815	85	0.15	54604	98.35
816	47	0.08	54651	98.44
817	56	0.10	54707	98.54
818	65	0.12	54772	98.65
819	40	0.07	54812	98.73
820	35	0.06	54847	98.79
821	43	0.08	54890	98.87
822	37	0.07	54927	98.93
823	37	0.07	54964	99.00
824	37	0.07	55001	99.07
825	37	0.07	55038	99.13
826	31	0.06	55069	99.19
827	35	0.06	55104	99.25
828	26	0.05	55130	99.30
829	17	0.03	55147	99.33
830	35	0.06	55182	99.39
831	25	0.05	55207	99.44
832	27	0.05	55234	99.49
833	19	0.03	55253	99.52
834	22	0.04	55275	99.56
835	15	0.03	55290	99.59
836	21	0.04	55311	99.63
837	16	0.03	55327	99.65
838	15	0.03	55342	99.68
839	15	0.03	55357	99.71
840	12	0.02	55369	99.73
841	7	0.01	55376	99.74
842	7	0.01	55383	99.76
843	8	0.01	55391	99.77
844	8	0.01	55399	99.78
845	17	0.03	55416	99.81
846	7	0.01	55423	99.83
847	6	0.01	55429	99.84
848	7	0.01	55436	99.85
849	5	0.01	55441	99.86
850	78	0.14	55519	100.00

Table F.9. Scale Score Distribution—ELA Grade 5

SS	Freq.	%	Cum. Freq.	Cum. %
650	84	0.15	84	0.15
651	7	0.01	91	0.16
652	3	0.01	94	0.17
653	11	0.02	105	0.19
654	4	0.01	109	0.19
655	1	0.00	110	0.19
656	7	0.01	117	0.21
657	14	0.02	131	0.23
658	11	0.02	142	0.25
659	10	0.02	152	0.27
660	14	0.02	166	0.29
661	13	0.02	179	0.32
662	12	0.02	191	0.34
663	17	0.03	208	0.37
664	19	0.03	227	0.40
665	23	0.04	250	0.44
666	20	0.04	270	0.48
667	26	0.05	296	0.52
668	26	0.05	322	0.57
669	26	0.05	348	0.61
670	32	0.06	380	0.67
671	50	0.09	430	0.76
672	36	0.06	466	0.82
673	45	0.08	511	0.90
674	54	0.10	565	1.00
675	45	0.08	610	1.08
676	64	0.11	674	1.19
677	77	0.14	751	1.33
678	81	0.14	832	1.47
679	91	0.16	923	1.63
680	67	0.12	990	1.75
681	83	0.15	1073	1.89
682	115	0.20	1188	2.10
683	99	0.17	1287	2.27
684	112	0.20	1399	2.47
685	124	0.22	1523	2.69
686	118	0.21	1641	2.90
687	167	0.29	1808	3.19
688	149	0.26	1957	3.45
689	160	0.28	2117	3.74
690	193	0.34	2310	4.08
691	174	0.31	2484	4.38
692	184	0.32	2668	4.71
693	227	0.40	2895	5.11
694	255	0.45	3150	5.56
695	267	0.47	3417	6.03
696	279	0.49	3696	6.52
697	249	0.44	3945	6.96
698	262	0.46	4207	7.43
699	287	0.51	4494	7.93
700	300	0.53	4794	8.46
701	338	0.60	5132	9.06
702	334	0.59	5466	9.65
703	336	0.59	5802	10.24
704	361	0.64	6163	10.88
705	351	0.62	6514	11.50
706	399	0.70	6913	12.20
707	375	0.66	7288	12.86
708	406	0.72	7694	13.58
709	372	0.66	8066	14.24
710	393	0.69	8459	14.93
711	428	0.76	8887	15.69
712	406	0.72	9293	16.40
713	419	0.74	9712	17.14
714	419	0.74	10131	17.88
715	438	0.77	10569	18.65
716	470	0.83	11039	19.48
717	482	0.85	11521	20.33
718	470	0.83	11991	21.16
719	472	0.83	12463	22.00
720	530	0.94	12993	22.93
721	478	0.84	13471	23.78
722	484	0.85	13955	24.63
723	475	0.84	14430	25.47
724	550	0.97	14980	26.44
725	502	0.89	15482	27.33
726	517	0.91	15999	28.24
727	556	0.98	16555	29.22
728	503	0.89	17058	30.11
729	535	0.94	17593	31.05
730	548	0.97	18141	32.02
731	560	0.99	18701	33.01
732	608	1.07	19309	34.08
733	559	0.99	19868	35.07
734	593	1.05	20461	36.11
735	581	1.03	21042	37.14
736	569	1.00	21611	38.14
737	640	1.13	22251	39.27
738	629	1.11	22880	40.38
739	582	1.03	23462	41.41
740	604	1.07	24066	42.48
741	632	1.12	24698	43.59
742	608	1.07	25306	44.67
743	615	1.09	25921	45.75
744	622	1.10	26543	46.85
745	622	1.10	27165	47.95
746	620	1.09	27785	49.04
747	646	1.14	28431	50.18
748	580	1.02	29011	51.20
749	588	1.04	29599	52.24

SS	Freq.	%	Cum. Freq.	Cum. %
750	644	1.14	30243	53.38
751	630	1.11	30873	54.49
752	600	1.06	31473	55.55
753	652	1.15	32125	56.70
754	654	1.15	32779	57.86
755	606	1.07	33385	58.92
756	610	1.08	33995	60.00
757	631	1.11	34626	61.12
758	589	1.04	35215	62.15
759	605	1.07	35820	63.22
760	595	1.05	36415	64.27
761	594	1.05	37009	65.32
762	555	0.98	37564	66.30
763	600	1.06	38164	67.36
764	561	0.99	38725	68.35
765	610	1.08	39335	69.43
766	588	1.04	39923	70.46
767	547	0.97	40470	71.43
768	570	1.01	41040	72.44
769	544	0.96	41584	73.40
770	534	0.94	42118	74.34
771	547	0.97	42665	75.30
772	488	0.86	43153	76.17
773	506	0.89	43659	77.06
774	510	0.90	44169	77.96
775	488	0.86	44657	78.82
776	466	0.82	45123	79.64
777	480	0.85	45603	80.49
778	411	0.73	46014	81.22
779	431	0.76	46445	81.98
780	444	0.78	46889	82.76
781	454	0.80	47343	83.56
782	442	0.78	47785	84.34
783	416	0.73	48201	85.08
784	425	0.75	48626	85.83
785	393	0.69	49019	86.52
786	353	0.62	49372	87.14
787	344	0.61	49716	87.75
788	332	0.59	50048	88.34
789	333	0.59	50381	88.92
790	331	0.58	50712	89.51
791	298	0.53	51010	90.03
792	291	0.51	51301	90.55
793	293	0.52	51594	91.06
794	277	0.49	51871	91.55
795	290	0.51	52161	92.06
796	250	0.44	52411	92.51
797	224	0.40	52635	92.90
798	253	0.45	52888	93.35
799	236	0.42	53124	93.76
800	201	0.35	53325	94.12
801	209	0.37	53534	94.49
802	219	0.39	53753	94.87
803	198	0.35	53951	95.22
804	163	0.29	54114	95.51
805	150	0.26	54264	95.78
806	150	0.26	54414	96.04
807	158	0.28	54572	96.32
808	143	0.25	54715	96.57
809	123	0.22	54838	96.79
810	121	0.21	54959	97.00
811	117	0.21	55076	97.21
812	122	0.22	55198	97.42
813	100	0.18	55298	97.60
814	87	0.15	55385	97.75
815	105	0.19	55490	97.94
816	99	0.17	55589	98.11
817	68	0.12	55657	98.23
818	81	0.14	55738	98.38
819	68	0.12	55806	98.50
820	69	0.12	55875	98.62
821	74	0.13	55949	98.75
822	52	0.09	56001	98.84
823	57	0.10	56058	98.94
824	25	0.04	56083	98.99
825	49	0.09	56132	99.07
826	39	0.07	56171	99.14
827	43	0.08	56214	99.22
828	45	0.08	56259	99.30
829	33	0.06	56292	99.36
830	37	0.07	56329	99.42
831	41	0.07	56370	99.49
832	23	0.04	56393	99.53
833	14	0.02	56407	99.56
834	13	0.02	56420	99.58
835	19	0.03	56439	99.62
836	12	0.02	56451	99.64
837	14	0.02	56465	99.66
838	20	0.04	56485	99.70
839	16	0.03	56501	99.72
840	15	0.03	56516	99.75
841	8	0.01	56524	99.77
842	12	0.02	56536	99.79
843	12	0.02	56548	99.81
844	13	0.02	56561	99.83
845	7	0.01	56568	99.84
846	13	0.02	56581	99.87
847	8	0.01	56589	99.88
848	4	0.01	56593	99.89
849	7	0.01	56600	99.90
850	57	0.10	56657	100.00

Table F.10. Scale Score Distribution—ELA Grade 6

SS	Freq.	%	Cum. Freq.	Cum. %
650	95	0.17	95	0.17
651	7	0.01	102	0.18
652	15	0.03	117	0.21
653	4	0.01	121	0.22
654	10	0.02	131	0.24
655	13	0.02	144	0.26
656	10	0.02	154	0.28
657	12	0.02	166	0.30
658	13	0.02	179	0.32
659	17	0.03	196	0.35
660	16	0.03	212	0.38
661	9	0.02	221	0.40
662	19	0.03	240	0.43
663	20	0.04	260	0.47
664	30	0.05	290	0.52
665	20	0.04	310	0.56
666	33	0.06	343	0.62
667	30	0.05	373	0.67
668	38	0.07	411	0.74
669	42	0.08	453	0.81
670	54	0.10	507	0.91
671	55	0.10	562	1.01
672	61	0.11	623	1.12
673	62	0.11	685	1.23
674	65	0.12	750	1.35
675	87	0.16	837	1.51
676	69	0.12	906	1.63
677	87	0.16	993	1.79
678	86	0.15	1079	1.94
679	113	0.20	1192	2.14
680	110	0.20	1302	2.34
681	134	0.24	1436	2.58
682	127	0.23	1563	2.81
683	151	0.27	1714	3.08
684	144	0.26	1858	3.34
685	170	0.31	2028	3.65
686	188	0.34	2216	3.99
687	213	0.38	2429	4.37
688	194	0.35	2623	4.72
689	213	0.38	2836	5.10
690	246	0.44	3082	5.54
691	259	0.47	3341	6.01
692	250	0.45	3591	6.46
693	267	0.48	3858	6.94
694	278	0.50	4136	7.44
695	307	0.55	4443	7.99
696	309	0.56	4752	8.55
697	311	0.56	5063	9.11
698	311	0.56	5374	9.67
699	360	0.65	5734	10.31
700	364	0.65	6098	10.97
701	376	0.68	6474	11.64
702	371	0.67	6845	12.31
703	399	0.72	7244	13.03
704	394	0.71	7638	13.74
705	411	0.74	8049	14.48
706	433	0.78	8482	15.25
707	449	0.81	8931	16.06
708	453	0.81	9384	16.88
709	445	0.80	9829	17.68
710	432	0.78	10261	18.45
711	424	0.76	10685	19.22
712	428	0.77	11113	19.99
713	449	0.81	11562	20.79
714	478	0.86	12040	21.65
715	444	0.80	12484	22.45
716	430	0.77	12914	23.23
717	494	0.89	13408	24.11
718	497	0.89	13905	25.01
719	515	0.93	14420	25.93
720	494	0.89	14914	26.82
721	510	0.92	15424	27.74
722	504	0.91	15928	28.65
723	498	0.90	16426	29.54
724	517	0.93	16943	30.47
725	540	0.97	17483	31.44
726	550	0.99	18033	32.43
727	518	0.93	18551	33.36
728	560	1.01	19111	34.37
729	534	0.96	19645	35.33
730	527	0.95	20172	36.28
731	568	1.02	20740	37.30
732	553	0.99	21293	38.30
733	541	0.97	21834	39.27
734	596	1.07	22430	40.34
735	584	1.05	23014	41.39
736	607	1.09	23621	42.48
737	593	1.07	24214	43.55
738	599	1.08	24813	44.63
739	568	1.02	25381	45.65
740	627	1.13	26008	46.78
741	596	1.07	26604	47.85
742	575	1.03	27179	48.88
743	608	1.09	27787	49.97
744	621	1.12	28408	51.09
745	618	1.11	29026	52.20
746	609	1.10	29635	53.30
747	623	1.12	30258	54.42
748	612	1.10	30870	55.52
749	617	1.11	31487	56.63

SS	Freq.	%	Cum. Freq.	Cum. %
750	619	1.11	32106	57.74
751	602	1.08	32708	58.83
752	593	1.07	33301	59.89
753	598	1.08	33899	60.97
754	665	1.20	34564	62.16
755	576	1.04	35140	63.20
756	619	1.11	35759	64.31
757	605	1.09	36364	65.40
758	578	1.04	36942	66.44
759	606	1.09	37548	67.53
760	588	1.06	38136	68.59
761	631	1.13	38767	69.72
762	629	1.13	39396	70.85
763	567	1.02	39963	71.87
764	586	1.05	40549	72.93
765	602	1.08	41151	74.01
766	567	1.02	41718	75.03
767	578	1.04	42296	76.07
768	517	0.93	42813	77.00
769	551	0.99	43364	77.99
770	519	0.93	43883	78.92
771	488	0.88	44371	79.80
772	522	0.94	44893	80.74
773	460	0.83	45353	81.57
774	423	0.76	45776	82.33
775	465	0.84	46241	83.16
776	485	0.87	46726	84.04
777	416	0.75	47142	84.78
778	440	0.79	47582	85.58
779	422	0.76	48004	86.34
780	372	0.67	48376	87.00
781	395	0.71	48771	87.71
782	371	0.67	49142	88.38
783	399	0.72	49541	89.10
784	335	0.60	49876	89.70
785	333	0.60	50209	90.30
786	301	0.54	50510	90.84
787	307	0.55	50817	91.39
788	288	0.52	51105	91.91
789	258	0.46	51363	92.38
790	278	0.50	51641	92.88
791	291	0.52	51932	93.40
792	213	0.38	52145	93.78
793	216	0.39	52361	94.17
794	216	0.39	52577	94.56
795	192	0.35	52769	94.90
796	183	0.33	52952	95.23
797	174	0.31	53126	95.55
798	175	0.31	53301	95.86
799	171	0.31	53472	96.17
800	166	0.30	53638	96.47
801	129	0.23	53767	96.70
802	118	0.21	53885	96.91
803	122	0.22	54007	97.13
804	118	0.21	54125	97.34
805	112	0.20	54237	97.55
806	107	0.19	54344	97.74
807	109	0.20	54453	97.93
808	88	0.16	54541	98.09
809	86	0.15	54627	98.25
810	70	0.13	54697	98.37
811	77	0.14	54774	98.51
812	60	0.11	54834	98.62
813	51	0.09	54885	98.71
814	53	0.10	54938	98.81
815	56	0.10	54994	98.91
816	50	0.09	55044	99.00
817	46	0.08	55090	99.08
818	34	0.06	55124	99.14
819	41	0.07	55165	99.21
820	28	0.05	55193	99.26
821	47	0.08	55240	99.35
822	27	0.05	55267	99.40
823	28	0.05	55295	99.45
824	31	0.06	55326	99.50
825	18	0.03	55344	99.54
826	28	0.05	55372	99.59
827	15	0.03	55387	99.61
828	17	0.03	55404	99.64
829	15	0.03	55419	99.67
830	9	0.02	55428	99.69
831	12	0.02	55440	99.71
832	12	0.02	55452	99.73
833	17	0.03	55469	99.76
834	12	0.02	55481	99.78
835	15	0.03	55496	99.81
836	8	0.01	55504	99.82
837	8	0.01	55512	99.84
838	10	0.02	55522	99.86
839	6	0.01	55528	99.87
840	5	0.01	55533	99.88
841	8	0.01	55541	99.89
842	7	0.01	55548	99.90
843	4	0.01	55552	99.91
844	5	0.01	55557	99.92
845	7	0.01	55564	99.93
847	1	0.00	55565	99.93
848	5	0.01	55570	99.94
849	3	0.01	55573	99.95
850	29	0.05	55602	100.00

Table F.11. Scale Score Distribution—ELA Grade 7

SS	Freq.	%	Cum. Freq.	Cum. %
650	343	0.64	343	0.64
651	30	0.06	373	0.69
652	19	0.04	392	0.73
653	26	0.05	418	0.78
654	29	0.05	447	0.83
655	32	0.06	479	0.89
656	33	0.06	512	0.95
657	30	0.06	542	1.01
658	32	0.06	574	1.07
659	43	0.08	617	1.14
660	41	0.08	658	1.22
661	50	0.09	708	1.31
662	51	0.09	759	1.41
663	54	0.10	813	1.51
664	40	0.07	853	1.58
665	60	0.11	913	1.69
666	80	0.15	993	1.84
667	61	0.11	1054	1.96
668	63	0.12	1117	2.07
669	64	0.12	1181	2.19
670	94	0.17	1275	2.37
671	90	0.17	1365	2.53
672	124	0.23	1489	2.76
673	105	0.19	1594	2.96
674	118	0.22	1712	3.18
675	122	0.23	1834	3.40
676	143	0.27	1977	3.67
677	132	0.24	2109	3.91
678	134	0.25	2243	4.16
679	144	0.27	2387	4.43
680	159	0.30	2546	4.72
681	166	0.31	2712	5.03
682	180	0.33	2892	5.37
683	174	0.32	3066	5.69
684	172	0.32	3238	6.01
685	204	0.38	3442	6.39
686	206	0.38	3648	6.77
687	225	0.42	3873	7.19
688	252	0.47	4125	7.65
689	244	0.45	4369	8.11
690	255	0.47	4624	8.58
691	259	0.48	4883	9.06
692	300	0.56	5183	9.62
693	252	0.47	5435	10.08
694	274	0.51	5709	10.59
695	280	0.52	5989	11.11
696	259	0.48	6248	11.59
697	289	0.54	6537	12.13
698	305	0.57	6842	12.70
699	287	0.53	7129	13.23
700	293	0.54	7422	13.77
701	353	0.65	7775	14.43
702	317	0.59	8092	15.01
703	327	0.61	8419	15.62
704	316	0.59	8735	16.21
705	329	0.61	9064	16.82
706	347	0.64	9411	17.46
707	338	0.63	9749	18.09
708	349	0.65	10098	18.74
709	382	0.71	10480	19.45
710	341	0.63	10821	20.08
711	383	0.71	11204	20.79
712	394	0.73	11598	21.52
713	380	0.71	11978	22.22
714	400	0.74	12378	22.97
715	402	0.75	12780	23.71
716	392	0.73	13172	24.44
717	396	0.73	13568	25.17
718	435	0.81	14003	25.98
719	414	0.77	14417	26.75
720	420	0.78	14837	27.53
721	428	0.79	15265	28.32
722	453	0.84	15718	29.16
723	413	0.77	16131	29.93
724	440	0.82	16571	30.75
725	472	0.88	17043	31.62
726	443	0.82	17486	32.44
727	472	0.88	17958	33.32
728	515	0.96	18473	34.28
729	520	0.96	18993	35.24
730	479	0.89	19472	36.13
731	497	0.92	19969	37.05
732	517	0.96	20486	38.01
733	527	0.98	21013	38.99
734	521	0.97	21534	39.96
735	536	0.99	22070	40.95
736	537	1.00	22607	41.95
737	531	0.99	23138	42.93
738	589	1.09	23727	44.02
739	498	0.92	24225	44.95
740	483	0.90	24708	45.84
741	528	0.98	25236	46.82
742	535	0.99	25771	47.82
743	530	0.98	26301	48.80
744	517	0.96	26818	49.76
745	570	1.06	27388	50.82
746	580	1.08	27968	51.89
747	574	1.07	28542	52.96
748	558	1.04	29100	53.99
749	533	0.99	29633	54.98

SS	Freq.	%	Cum. Freq.	Cum. %
750	549	1.02	30182	56.00
751	576	1.07	30758	57.07
752	537	1.00	31295	58.07
753	604	1.12	31899	59.19
754	560	1.04	32459	60.23
755	553	1.03	33012	61.25
756	543	1.01	33555	62.26
757	551	1.02	34106	63.28
758	507	0.94	34613	64.22
759	519	0.96	35132	65.19
760	530	0.98	35662	66.17
761	517	0.96	36179	67.13
762	536	0.99	36715	68.12
763	527	0.98	37242	69.10
764	466	0.86	37708	69.97
765	459	0.85	38167	70.82
766	493	0.91	38660	71.73
767	486	0.90	39146	72.63
768	457	0.85	39603	73.48
769	482	0.89	40085	74.38
770	424	0.79	40509	75.16
771	446	0.83	40955	75.99
772	423	0.78	41378	76.78
773	397	0.74	41775	77.51
774	421	0.78	42196	78.29
775	401	0.74	42597	79.04
776	425	0.79	43022	79.83
777	386	0.72	43408	80.54
778	393	0.73	43801	81.27
779	378	0.70	44179	81.97
780	398	0.74	44577	82.71
781	342	0.63	44919	83.35
782	352	0.65	45271	84.00
783	365	0.68	45636	84.68
784	355	0.66	45991	85.33
785	350	0.65	46341	85.98
786	310	0.58	46651	86.56
787	335	0.62	46986	87.18
788	323	0.60	47309	87.78
789	296	0.55	47605	88.33
790	272	0.50	47877	88.83
791	280	0.52	48157	89.35
792	238	0.44	48395	89.79
793	242	0.45	48637	90.24
794	267	0.50	48904	90.74
795	251	0.47	49155	91.21
796	219	0.41	49374	91.61
797	213	0.40	49587	92.01
798	208	0.39	49795	92.39
799	204	0.38	49999	92.77
800	214	0.40	50213	93.17
801	185	0.34	50398	93.51
802	169	0.31	50567	93.83
803	172	0.32	50739	94.14
804	170	0.32	50909	94.46
805	146	0.27	51055	94.73
806	160	0.30	51215	95.03
807	144	0.27	51359	95.29
808	118	0.22	51477	95.51
809	135	0.25	51612	95.76
810	127	0.24	51739	96.00
811	145	0.27	51884	96.27
812	131	0.24	52015	96.51
813	113	0.21	52128	96.72
814	99	0.18	52227	96.91
815	98	0.18	52325	97.09
816	91	0.17	52416	97.26
817	92	0.17	52508	97.43
818	92	0.17	52600	97.60
819	76	0.14	52676	97.74
820	83	0.15	52759	97.89
821	63	0.12	52822	98.01
822	72	0.13	52894	98.14
823	75	0.14	52969	98.28
824	71	0.13	53040	98.41
825	55	0.10	53095	98.52
826	54	0.10	53149	98.62
827	49	0.09	53198	98.71
828	53	0.10	53251	98.81
829	50	0.09	53301	98.90
830	48	0.09	53349	98.99
831	33	0.06	53382	99.05
832	33	0.06	53415	99.11
833	36	0.07	53451	99.18
834	26	0.05	53477	99.22
835	29	0.05	53506	99.28
836	36	0.07	53542	99.35
837	21	0.04	53563	99.38
838	19	0.04	53582	99.42
839	17	0.03	53599	99.45
840	15	0.03	53614	99.48
841	28	0.05	53642	99.53
842	16	0.03	53658	99.56
843	17	0.03	53675	99.59
844	16	0.03	53691	99.62
845	25	0.05	53716	99.67
846	20	0.04	53736	99.70
847	14	0.03	53750	99.73
848	12	0.02	53762	99.75
849	8	0.01	53770	99.77
850	125	0.23	53895	100.00

Table F.12. Scale Score Distribution—ELA Grade 8

SS	Freq.	%	Cum. Freq.	Cum. %
650	513	0.99	513	0.99
651	30	0.06	543	1.05
652	44	0.09	587	1.13
653	48	0.09	635	1.23
654	52	0.10	687	1.33
655	72	0.14	759	1.47
656	50	0.10	809	1.56
657	64	0.12	873	1.69
658	54	0.10	927	1.79
659	52	0.10	979	1.89
660	62	0.12	1041	2.01
661	75	0.14	1116	2.16
662	91	0.18	1207	2.33
663	104	0.20	1311	2.53
664	78	0.15	1389	2.68
665	72	0.14	1461	2.82
666	90	0.17	1551	3.00
667	111	0.21	1662	3.21
668	108	0.21	1770	3.42
669	132	0.26	1902	3.67
670	130	0.25	2032	3.93
671	131	0.25	2163	4.18
672	150	0.29	2313	4.47
673	132	0.26	2445	4.72
674	158	0.31	2603	5.03
675	173	0.33	2776	5.36
676	166	0.32	2942	5.68
677	186	0.36	3128	6.04
678	190	0.37	3318	6.41
679	183	0.35	3501	6.76
680	215	0.42	3716	7.18
681	187	0.36	3903	7.54
682	210	0.41	4113	7.95
683	212	0.41	4325	8.36
684	232	0.45	4557	8.80
685	233	0.45	4790	9.25
686	243	0.47	5033	9.72
687	218	0.42	5251	10.14
688	225	0.43	5476	10.58
689	239	0.46	5715	11.04
690	246	0.48	5961	11.52
691	264	0.51	6225	12.03
692	307	0.59	6532	12.62
693	263	0.51	6795	13.13
694	271	0.52	7066	13.65
695	314	0.61	7380	14.26
696	287	0.55	7667	14.81
697	285	0.55	7952	15.36
698	305	0.59	8257	15.95
699	305	0.59	8562	16.54
700	300	0.58	8862	17.12
701	305	0.59	9167	17.71
702	326	0.63	9493	18.34
703	328	0.63	9821	18.97
704	349	0.67	10170	19.65
705	325	0.63	10495	20.28
706	338	0.65	10833	20.93
707	329	0.64	11162	21.56
708	331	0.64	11493	22.20
709	374	0.72	11867	22.93
710	391	0.76	12258	23.68
711	357	0.69	12615	24.37
712	361	0.70	12976	25.07
713	376	0.73	13352	25.80
714	392	0.76	13744	26.55
715	362	0.70	14106	27.25
716	369	0.71	14475	27.97
717	431	0.83	14906	28.80
718	389	0.75	15295	29.55
719	434	0.84	15729	30.39
720	396	0.77	16125	31.15
721	412	0.80	16537	31.95
722	438	0.85	16975	32.80
723	448	0.87	17423	33.66
724	430	0.83	17853	34.49
725	446	0.86	18299	35.35
726	463	0.89	18762	36.25
727	452	0.87	19214	37.12
728	449	0.87	19663	37.99
729	394	0.76	20057	38.75
730	448	0.87	20505	39.62
731	450	0.87	20955	40.48
732	473	0.91	21428	41.40
733	504	0.97	21932	42.37
734	463	0.89	22395	43.27
735	456	0.88	22851	44.15
736	482	0.93	23333	45.08
737	500	0.97	23833	46.05
738	520	1.00	24353	47.05
739	490	0.95	24843	48.00
740	512	0.99	25355	48.99
741	505	0.98	25860	49.96
742	473	0.91	26333	50.88
743	502	0.97	26835	51.85
744	488	0.94	27323	52.79
745	530	1.02	27853	53.81
746	483	0.93	28336	54.74
747	490	0.95	28826	55.69
748	518	1.00	29344	56.69
749	484	0.94	29828	57.63

SS	Freq.	%	Cum. Freq.	Cum. %
750	448	0.87	30276	58.49
751	468	0.90	30744	59.40
752	458	0.88	31202	60.28
753	458	0.88	31660	61.17
754	456	0.88	32116	62.05
755	482	0.93	32598	62.98
756	458	0.88	33056	63.86
757	473	0.91	33529	64.78
758	450	0.87	33979	65.65
759	454	0.88	34433	66.52
760	465	0.90	34898	67.42
761	450	0.87	35348	68.29
762	418	0.81	35766	69.10
763	467	0.90	36233	70.00
764	496	0.96	36729	70.96
765	427	0.82	37156	71.79
766	424	0.82	37580	72.60
767	421	0.81	38001	73.42
768	376	0.73	38377	74.14
769	384	0.74	38761	74.89
770	421	0.81	39182	75.70
771	384	0.74	39566	76.44
772	387	0.75	39953	77.19
773	370	0.71	40323	77.90
774	376	0.73	40699	78.63
775	360	0.70	41059	79.33
776	355	0.69	41414	80.01
777	346	0.67	41760	80.68
778	321	0.62	42081	81.30
779	347	0.67	42428	81.97
780	321	0.62	42749	82.59
781	327	0.63	43076	83.22
782	326	0.63	43402	83.85
783	285	0.55	43687	84.40
784	285	0.55	43972	84.95
785	287	0.55	44259	85.51
786	295	0.57	44554	86.08
787	262	0.51	44816	86.58
788	298	0.58	45114	87.16
789	280	0.54	45394	87.70
790	293	0.57	45687	88.27
791	285	0.55	45972	88.82
792	247	0.48	46219	89.29
793	233	0.45	46452	89.74
794	245	0.47	46697	90.22
795	223	0.43	46920	90.65
796	220	0.43	47140	91.07
797	186	0.36	47326	91.43
798	197	0.38	47523	91.81
799	180	0.35	47703	92.16
800	167	0.32	47870	92.48
801	166	0.32	48036	92.81
802	160	0.31	48196	93.11
803	160	0.31	48356	93.42
804	147	0.28	48503	93.71
805	155	0.30	48658	94.01
806	146	0.28	48804	94.29
807	137	0.26	48941	94.55
808	144	0.28	49085	94.83
809	120	0.23	49205	95.06
810	122	0.24	49327	95.30
811	132	0.26	49459	95.55
812	132	0.26	49591	95.81
813	106	0.20	49697	96.01
814	91	0.18	49788	96.19
815	119	0.23	49907	96.42
816	100	0.19	50007	96.61
817	99	0.19	50106	96.80
818	93	0.18	50199	96.98
819	87	0.17	50286	97.15
820	95	0.18	50381	97.34
821	83	0.16	50464	97.50
822	80	0.15	50544	97.65
823	76	0.15	50620	97.80
824	62	0.12	50682	97.92
825	57	0.11	50739	98.03
826	76	0.15	50815	98.17
827	62	0.12	50877	98.29
828	48	0.09	50925	98.39
829	65	0.13	50990	98.51
830	46	0.09	51036	98.60
831	52	0.10	51088	98.70
832	35	0.07	51123	98.77
833	38	0.07	51161	98.84
834	47	0.09	51208	98.93
835	41	0.08	51249	99.01
836	27	0.05	51276	99.06
837	35	0.07	51311	99.13
838	17	0.03	51328	99.17
839	28	0.05	51356	99.22
840	29	0.06	51385	99.28
841	29	0.06	51414	99.33
842	27	0.05	51441	99.38
843	24	0.05	51465	99.43
844	29	0.06	51494	99.49
845	15	0.03	51509	99.52
846	14	0.03	51523	99.54
847	7	0.01	51530	99.56
848	15	0.03	51545	99.58
849	25	0.05	51570	99.63
850	190	0.37	51760	100.00

Table F.13. Scale Score Distribution—CSLA Grade 3

SS	Freq.	%	Cum. Freq.	Cum. %
650	17	1.18	17	1.18
665	22	1.53	39	2.71
676	36	2.50	75	5.21
684	39	2.71	114	7.92
690	59	4.10	173	12.01
695	62	4.31	235	16.32
699	52	3.61	287	19.93
700	49	3.40	336	23.33
706	65	4.51	401	27.85
709	39	2.71	440	30.56
712	47	3.26	487	33.82
715	44	3.06	531	36.88
717	53	3.68	584	40.56
719	57	3.96	641	44.51
721	38	2.64	679	47.15
723	35	2.43	714	49.58
725	45	3.13	759	52.71
727	34	2.36	793	55.07
729	31	2.15	824	57.22
730	39	2.71	863	59.93
732	21	1.46	884	61.39
734	37	2.57	921	63.96
735	29	2.01	950	65.97
737	27	1.88	977	67.85
738	26	1.81	1003	69.65
740	25	1.74	1028	71.39
741	26	1.81	1054	73.19
743	21	1.46	1075	74.65
744	30	2.08	1105	76.74
746	21	1.46	1126	78.19
747	21	1.46	1147	79.65
748	24	1.67	1171	81.32
750	27	1.88	1198	83.19
751	17	1.18	1215	84.38
753	17	1.18	1232	85.56
754	9	0.63	1241	86.18
756	15	1.04	1256	87.22
757	20	1.39	1276	88.61
759	15	1.04	1291	89.65
760	14	0.97	1305	90.63
762	13	0.90	1318	91.53
763	21	1.46	1339	92.99
765	19	1.32	1358	94.31
766	13	0.90	1371	95.21
768	8	0.56	1379	95.76
770	7	0.49	1386	96.25
772	7	0.49	1393	96.74
773	7	0.49	1400	97.22
775	8	0.56	1408	97.78
777	5	0.35	1413	98.13
779	1	0.07	1414	98.19
782	8	0.56	1422	98.75
784	4	0.28	1426	99.03
787	3	0.21	1429	99.24
789	4	0.28	1433	99.51
792	1	0.07	1434	99.58
796	1	0.07	1435	99.65
799	2	0.14	1437	99.79
804	1	0.07	1438	99.86
809	1	0.07	1439	99.93
815	1	0.07	1440	100.00

Table F.14. Scale Score Distribution—CSLA Grade 4

SS	Freq.	%	Cum. Freq.	Cum. %
650	2	0.17	2	0.17
656	7	0.59	9	0.76
667	11	0.93	20	1.69
675	9	0.76	29	2.46
681	20	1.69	49	4.15
686	23	1.95	72	6.10
690	23	1.95	95	8.05
694	30	2.54	125	10.59
697	40	3.39	165	13.98
700	33	2.80	198	16.78
703	35	2.97	233	19.75
705	37	3.14	270	22.88
708	43	3.64	313	26.53
710	30	2.54	343	29.07
712	30	2.54	373	31.61
714	35	2.97	408	34.58
716	30	2.54	438	37.12
718	32	2.71	470	39.83
720	22	1.86	492	41.69
721	35	2.97	527	44.66
723	40	3.39	567	48.05
725	29	2.46	596	50.51
726	21	1.78	617	52.29
728	26	2.20	643	54.49
729	31	2.63	674	57.12
731	32	2.71	706	59.83
732	28	2.37	734	62.20
733	26	2.20	760	64.41
735	33	2.80	793	67.20
736	25	2.12	818	69.32
737	21	1.78	839	71.10
739	21	1.78	860	72.88
740	20	1.69	880	74.58
741	17	1.44	897	76.02
743	24	2.03	921	78.05
744	19	1.61	940	79.66
745	13	1.10	953	80.76
747	20	1.69	973	82.46
748	17	1.44	990	83.90
749	22	1.86	1012	85.76
750	12	1.02	1024	86.78
752	16	1.36	1040	88.14
753	18	1.53	1058	89.66
755	13	1.10	1071	90.76
756	13	1.10	1084	91.86
757	14	1.19	1098	93.05
759	6	0.51	1104	93.56
760	7	0.59	1111	94.15
762	8	0.68	1119	94.83
763	13	1.10	1132	95.93
765	5	0.42	1137	96.36
766	7	0.59	1144	96.95
768	6	0.51	1150	97.46
770	6	0.51	1156	97.97
772	3	0.25	1159	98.22
774	4	0.34	1163	98.56
776	4	0.34	1167	98.90
778	1	0.08	1168	98.98
780	3	0.25	1171	99.24
782	4	0.34	1175	99.58
785	2	0.17	1177	99.75
788	1	0.08	1178	99.83
791	1	0.08	1179	99.92
819	1	0.08	1180	100.00

Table F.15. Scale Score Distribution—Science Grade 5

SS	Freq.	%	Cum. Freq.	Cum. %
650	2413	4.28	2413	4.28
651	27	0.05	2440	4.32
652	32	0.06	2472	4.38
653	37	0.07	2509	4.45
654	24	0.04	2533	4.49
655	38	0.07	2571	4.56
656	36	0.06	2607	4.62
657	58	0.10	2665	4.72
658	57	0.10	2722	4.82
659	29	0.05	2751	4.88
660	52	0.09	2803	4.97
661	38	0.07	2841	5.03
662	54	0.10	2895	5.13
663	63	0.11	2958	5.24
664	61	0.11	3019	5.35
665	66	0.12	3085	5.47
666	57	0.10	3142	5.57
667	82	0.15	3224	5.71
668	80	0.14	3304	5.86
669	77	0.14	3381	5.99
670	95	0.17	3476	6.16
671	82	0.15	3558	6.31
672	110	0.19	3668	6.50
673	119	0.21	3787	6.71
674	102	0.18	3889	6.89
675	115	0.20	4004	7.10
676	120	0.21	4124	7.31
677	114	0.20	4238	7.51
678	124	0.22	4362	7.73
679	143	0.25	4505	7.98
680	135	0.24	4640	8.22
681	138	0.24	4778	8.47
682	159	0.28	4937	8.75
683	138	0.24	5075	8.99
684	164	0.29	5239	9.28
685	158	0.28	5397	9.56
686	158	0.28	5555	9.84
687	181	0.32	5736	10.17
688	203	0.36	5939	10.52
689	205	0.36	6144	10.89
690	202	0.36	6346	11.25
691	210	0.37	6556	11.62
692	206	0.37	6762	11.98
693	222	0.39	6984	12.38
694	233	0.41	7217	12.79
695	236	0.42	7453	13.21
696	267	0.47	7720	13.68
697	280	0.50	8000	14.18
698	279	0.49	8279	14.67
699	259	0.46	8538	15.13
700	297	0.53	8835	15.66
701	336	0.60	9171	16.25
702	312	0.55	9483	16.81
703	331	0.59	9814	17.39
704	316	0.56	10130	17.95
705	362	0.64	10492	18.59
706	378	0.67	10870	19.26
707	390	0.69	11260	19.95
708	392	0.69	11652	20.65
709	459	0.81	12111	21.46
710	467	0.83	12578	22.29
711	470	0.83	13048	23.12
712	435	0.77	13483	23.89
713	462	0.82	13945	24.71
714	502	0.89	14447	25.60
715	475	0.84	14922	26.44
716	497	0.88	15419	27.33
717	505	0.89	15924	28.22
718	561	0.99	16485	29.21
719	532	0.94	17017	30.16
720	570	1.01	17587	31.17
721	569	1.01	18156	32.18
722	564	1.00	18720	33.18
723	584	1.03	19304	34.21
724	632	1.12	19936	35.33
725	590	1.05	20526	36.38
726	646	1.14	21172	37.52
727	665	1.18	21837	38.70
728	680	1.21	22517	39.90
729	650	1.15	23167	41.06
730	681	1.21	23848	42.26
731	679	1.20	24527	43.47
732	635	1.13	25162	44.59
733	671	1.19	25833	45.78
734	730	1.29	26563	47.07
735	666	1.18	27229	48.25
736	699	1.24	27928	49.49
737	709	1.26	28637	50.75
738	754	1.34	29391	52.09
739	708	1.25	30099	53.34
740	704	1.25	30803	54.59
741	757	1.34	31560	55.93
742	691	1.22	32251	57.15
743	718	1.27	32969	58.43
744	747	1.32	33716	59.75
745	727	1.29	34443	61.04
746	760	1.35	35203	62.39
747	729	1.29	35932	63.68
748	696	1.23	36628	64.91
749	682	1.21	37310	66.12

SS	Freq.	%	Cum. Freq.	Cum. %
750	669	1.19	37979	67.31
751	683	1.21	38662	68.52
752	732	1.30	39394	69.81
753	689	1.22	40083	71.03
754	693	1.23	40776	72.26
755	638	1.13	41414	73.39
756	657	1.16	42071	74.56
757	656	1.16	42727	75.72
758	620	1.10	43347	76.82
759	614	1.09	43961	77.91
760	646	1.14	44607	79.05
761	615	1.09	45222	80.14
762	584	1.03	45806	81.18
763	568	1.01	46374	82.18
764	538	0.95	46912	83.14
765	525	0.93	47437	84.07
766	540	0.96	47977	85.02
767	469	0.83	48446	85.85
768	475	0.84	48921	86.70
769	460	0.82	49381	87.51
770	440	0.78	49821	88.29
771	418	0.74	50239	89.03
772	433	0.77	50672	89.80
773	388	0.69	51060	90.49
774	379	0.67	51439	91.16
775	361	0.64	51800	91.80
776	334	0.59	52134	92.39
777	318	0.56	52452	92.95
778	287	0.51	52739	93.46
779	300	0.53	53039	93.99
780	277	0.49	53316	94.49
781	245	0.43	53561	94.92
782	209	0.37	53770	95.29
783	230	0.41	54000	95.70
784	187	0.33	54187	96.03
785	186	0.33	54373	96.36
786	195	0.35	54568	96.70
787	158	0.28	54726	96.98
788	161	0.29	54887	97.27
789	130	0.23	55017	97.50
790	124	0.22	55141	97.72
791	120	0.21	55261	97.93
792	111	0.20	55372	98.13
793	96	0.17	55468	98.30
794	87	0.15	55555	98.45
795	79	0.14	55634	98.59
796	72	0.13	55706	98.72
797	58	0.10	55764	98.82
798	62	0.11	55826	98.93
799	60	0.11	55886	99.04
800	52	0.09	55938	99.13
801	52	0.09	55990	99.22
802	50	0.09	56040	99.31
803	38	0.07	56078	99.38
804	30	0.05	56108	99.43
805	27	0.05	56135	99.48
806	35	0.06	56170	99.54
807	33	0.06	56203	99.60
808	21	0.04	56224	99.64
809	22	0.04	56246	99.68
810	18	0.03	56264	99.71
811	21	0.04	56285	99.75
812	11	0.02	56296	99.77
813	15	0.03	56311	99.79
814	6	0.01	56317	99.80
815	13	0.02	56330	99.83
816	10	0.02	56340	99.84
817	6	0.01	56346	99.85
818	11	0.02	56357	99.87
819	8	0.01	56365	99.89
820	9	0.02	56374	99.90
821	3	0.01	56377	99.91
822	4	0.01	56381	99.92
823	5	0.01	56386	99.93
824	3	0.01	56389	99.93
825	2	0.00	56391	99.93
827	3	0.01	56394	99.94
828	6	0.01	56400	99.95
830	6	0.01	56406	99.96
833	1	0.00	56407	99.96
834	1	0.00	56408	99.96
835	1	0.00	56409	99.97
836	4	0.01	56413	99.97
840	1	0.00	56414	99.98
841	1	0.00	56415	99.98
845	1	0.00	56416	99.98
846	2	0.00	56418	99.98
848	1	0.00	56419	99.98
849	1	0.00	56420	99.99
850	8	0.01	56428	100.00

Table F.16. Scale Score Distribution—Science Grade 8

SS	Freq.	%	Cum. Freq.	Cum. %
650	1545	3.03	1545	3.03
651	26	0.05	1571	3.08
652	30	0.06	1601	3.14
653	91	0.18	1692	3.32
654	67	0.13	1759	3.45
655	54	0.11	1813	3.56
656	60	0.12	1873	3.68
657	67	0.13	1940	3.81
658	77	0.15	2017	3.96
659	70	0.14	2087	4.10
660	60	0.12	2147	4.21
661	58	0.11	2205	4.33
662	66	0.13	2271	4.46
663	66	0.13	2337	4.59
664	68	0.13	2405	4.72
665	76	0.15	2481	4.87
666	73	0.14	2554	5.01
667	102	0.20	2656	5.21
668	100	0.20	2756	5.41
669	103	0.20	2859	5.61
670	111	0.22	2970	5.83
671	108	0.21	3078	6.04
672	100	0.20	3178	6.24
673	125	0.25	3303	6.48
674	116	0.23	3419	6.71
675	141	0.28	3560	6.99
676	139	0.27	3699	7.26
677	160	0.31	3859	7.57
678	174	0.34	4033	7.92
679	161	0.32	4194	8.23
680	171	0.34	4365	8.57
681	160	0.31	4525	8.88
682	190	0.37	4715	9.25
683	194	0.38	4909	9.64
684	169	0.33	5078	9.97
685	199	0.39	5277	10.36
686	188	0.37	5465	10.73
687	215	0.42	5680	11.15
688	204	0.40	5884	11.55
689	252	0.49	6136	12.04
690	208	0.41	6344	12.45
691	261	0.51	6605	12.96
692	244	0.48	6849	13.44
693	265	0.52	7114	13.96
694	259	0.51	7373	14.47
695	252	0.49	7625	14.97
696	295	0.58	7920	15.55
697	281	0.55	8201	16.10
698	324	0.64	8525	16.73
699	290	0.57	8815	17.30
700	318	0.62	9133	17.93
701	326	0.64	9459	18.57
702	377	0.74	9836	19.31
703	333	0.65	10169	19.96
704	342	0.67	10511	20.63
705	361	0.71	10872	21.34
706	362	0.71	11234	22.05
707	364	0.71	11598	22.76
708	414	0.81	12012	23.58
709	418	0.82	12430	24.40
710	387	0.76	12817	25.16
711	428	0.84	13245	26.00
712	407	0.80	13652	26.80
713	417	0.82	14069	27.61
714	452	0.89	14521	28.50
715	451	0.89	14972	29.39
716	494	0.97	15466	30.36
717	487	0.96	15953	31.31
718	510	1.00	16463	32.31
719	492	0.97	16955	33.28
720	532	1.04	17487	34.32
721	493	0.97	17980	35.29
722	531	1.04	18511	36.33
723	530	1.04	19041	37.37
724	590	1.16	19631	38.53
725	523	1.03	20154	39.56
726	554	1.09	20708	40.65
727	540	1.06	21248	41.71
728	587	1.15	21835	42.86
729	593	1.16	22428	44.02
730	574	1.13	23002	45.15
731	580	1.14	23582	46.29
732	620	1.22	24202	47.50
733	619	1.21	24821	48.72
734	610	1.20	25431	49.92
735	606	1.19	26037	51.11
736	604	1.19	26641	52.29
737	622	1.22	27263	53.51
738	623	1.22	27886	54.74
739	632	1.24	28518	55.98
740	636	1.25	29154	57.22
741	615	1.21	29769	58.43
742	707	1.39	30476	59.82
743	666	1.31	31142	61.13
744	660	1.30	31802	62.42
745	609	1.20	32411	63.62
746	631	1.24	33042	64.86
747	702	1.38	33744	66.23
748	610	1.20	34354	67.43
749	652	1.28	35006	68.71

SS	Freq.	%	Cum. Freq.	Cum. %
750	625	1.23	35631	69.94
751	654	1.28	36285	71.22
752	631	1.24	36916	72.46
753	660	1.30	37576	73.76
754	613	1.20	38189	74.96
755	590	1.16	38779	76.12
756	575	1.13	39354	77.24
757	561	1.10	39915	78.35
758	557	1.09	40472	79.44
759	571	1.12	41043	80.56
760	527	1.03	41570	81.59
761	496	0.97	42066	82.57
762	492	0.97	42558	83.53
763	444	0.87	43002	84.41
764	488	0.96	43490	85.36
765	466	0.91	43956	86.28
766	464	0.91	44420	87.19
767	405	0.79	44825	87.98
768	380	0.75	45205	88.73
769	403	0.79	45608	89.52
770	391	0.77	45999	90.29
771	390	0.77	46389	91.05
772	350	0.69	46739	91.74
773	292	0.57	47031	92.31
774	313	0.61	47344	92.93
775	297	0.58	47641	93.51
776	301	0.59	47942	94.10
777	279	0.55	48221	94.65
778	255	0.50	48476	95.15
779	241	0.47	48717	95.62
780	244	0.48	48961	96.10
781	206	0.40	49167	96.51
782	181	0.36	49348	96.86
783	152	0.30	49500	97.16
784	153	0.30	49653	97.46
785	125	0.25	49778	97.71
786	146	0.29	49924	97.99
787	133	0.26	50057	98.25
788	123	0.24	50180	98.49
789	82	0.16	50262	98.66
790	82	0.16	50344	98.82
791	72	0.14	50416	98.96
792	59	0.12	50475	99.07
793	52	0.10	50527	99.18
794	52	0.10	50579	99.28
795	50	0.10	50629	99.38
796	48	0.09	50677	99.47
797	45	0.09	50722	99.56
798	32	0.06	50754	99.62
799	33	0.06	50787	99.69
800	15	0.03	50802	99.72
801	22	0.04	50824	99.76
802	11	0.02	50835	99.78
803	15	0.03	50850	99.81
804	15	0.03	50865	99.84
805	10	0.02	50875	99.86
806	8	0.02	50883	99.87
807	8	0.02	50891	99.89
808	12	0.02	50903	99.91
809	3	0.01	50906	99.92
810	6	0.01	50912	99.93
811	2	0.00	50914	99.94
812	2	0.00	50916	99.94
813	4	0.01	50920	99.95
815	3	0.01	50923	99.95
816	4	0.01	50927	99.96
817	2	0.00	50929	99.96
818	5	0.01	50934	99.97
819	1	0.00	50935	99.98
821	4	0.01	50939	99.98
822	2	0.00	50941	99.99
823	1	0.00	50942	99.99
824	2	0.00	50944	99.99
826	1	0.00	50945	100.00
827	1	0.00	50946	100.00
831	1	0.00	50947	100.00

Table F.17. Scale Score Distribution—Science Grade 11

SS	Freq.	%	Cum. Freq.	Cum. %
650	1304	4.10	1304	4.10
651	17	0.05	1321	4.16
652	16	0.05	1337	4.21
653	10	0.03	1347	4.24
654	12	0.04	1359	4.28
655	10	0.03	1369	4.31
656	23	0.07	1392	4.38
657	24	0.08	1416	4.46
658	12	0.04	1428	4.50
659	15	0.05	1443	4.54
660	11	0.03	1454	4.58
661	18	0.06	1472	4.63
662	9	0.03	1481	4.66
663	19	0.06	1500	4.72
664	19	0.06	1519	4.78
665	28	0.09	1547	4.87
666	19	0.06	1566	4.93
667	22	0.07	1588	5.00
668	20	0.06	1608	5.06
669	21	0.07	1629	5.13
670	22	0.07	1651	5.20
671	33	0.10	1684	5.30
672	36	0.11	1720	5.41
673	50	0.16	1770	5.57
674	34	0.11	1804	5.68
675	28	0.09	1832	5.77
676	58	0.18	1890	5.95
677	45	0.14	1935	6.09
678	48	0.15	1983	6.24
679	56	0.18	2039	6.42
680	66	0.21	2105	6.63
681	72	0.23	2177	6.85
682	99	0.31	2276	7.16
683	79	0.25	2355	7.41
684	86	0.27	2441	7.68
685	88	0.28	2529	7.96
686	96	0.30	2625	8.26
687	98	0.31	2723	8.57
688	105	0.33	2828	8.90
689	91	0.29	2919	9.19
690	105	0.33	3024	9.52
691	132	0.42	3156	9.93
692	117	0.37	3273	10.30
693	129	0.41	3402	10.71
694	132	0.42	3534	11.12
695	162	0.51	3696	11.63
696	149	0.47	3845	12.10
697	149	0.47	3994	12.57
698	168	0.53	4162	13.10
699	188	0.59	4350	13.69
700	190	0.60	4540	14.29
701	200	0.63	4740	14.92
702	190	0.60	4930	15.52
703	187	0.59	5117	16.11
704	215	0.68	5332	16.78
705	244	0.77	5576	17.55
706	243	0.76	5819	18.32
707	275	0.87	6094	19.18
708	276	0.87	6370	20.05
709	285	0.90	6655	20.95
710	268	0.84	6923	21.79
711	315	0.99	7238	22.78
712	319	1.00	7557	23.79
713	356	1.12	7913	24.91
714	348	1.10	8261	26.00
715	374	1.18	8635	27.18
716	359	1.13	8994	28.31
717	393	1.24	9387	29.55
718	410	1.29	9797	30.84
719	387	1.22	10184	32.06
720	449	1.41	10633	33.47
721	415	1.31	11048	34.78
722	435	1.37	11483	36.15
723	425	1.34	11908	37.49
724	413	1.30	12321	38.79
725	475	1.50	12796	40.28
726	466	1.47	13262	41.75
727	476	1.50	13738	43.25
728	445	1.40	14183	44.65
729	460	1.45	14643	46.10
730	464	1.46	15107	47.56
731	491	1.55	15598	49.10
732	513	1.61	16111	50.72
733	507	1.60	16618	52.31
734	488	1.54	17106	53.85
735	518	1.63	17624	55.48
736	495	1.56	18119	57.04
737	485	1.53	18604	58.56
738	468	1.47	19072	60.04
739	438	1.38	19510	61.42
740	424	1.33	19934	62.75
741	464	1.46	20398	64.21
742	484	1.52	20882	65.73
743	467	1.47	21349	67.20
744	399	1.26	21748	68.46
745	456	1.44	22204	69.90
746	449	1.41	22653	71.31
747	408	1.28	23061	72.59
748	453	1.43	23514	74.02
749	442	1.39	23956	75.41

SS	Freq.	%	Cum. Freq.	Cum. %
750	365	1.15	24321	76.56
751	391	1.23	24712	77.79
752	378	1.19	25090	78.98
753	392	1.23	25482	80.22
754	371	1.17	25853	81.38
755	340	1.07	26193	82.45
756	365	1.15	26558	83.60
757	337	1.06	26895	84.66
758	330	1.04	27225	85.70
759	289	0.91	27514	86.61
760	299	0.94	27813	87.55
761	274	0.86	28087	88.42
762	240	0.76	28327	89.17
763	285	0.90	28612	90.07
764	240	0.76	28852	90.82
765	259	0.82	29111	91.64
766	239	0.75	29350	92.39
767	209	0.66	29559	93.05
768	205	0.65	29764	93.69
769	173	0.54	29937	94.24
770	173	0.54	30110	94.78
771	167	0.53	30277	95.31
772	132	0.42	30409	95.73
773	138	0.43	30547	96.16
774	118	0.37	30665	96.53
775	114	0.36	30779	96.89
776	123	0.39	30902	97.28
777	88	0.28	30990	97.55
778	82	0.26	31072	97.81
779	92	0.29	31164	98.10
780	65	0.20	31229	98.31

SS	Freq.	%	Cum. Freq.	Cum. %
781	58	0.18	31287	98.49
782	59	0.19	31346	98.67
783	55	0.17	31401	98.85
784	45	0.14	31446	98.99
785	46	0.14	31492	99.13
786	36	0.11	31528	99.25
787	41	0.13	31569	99.38
788	30	0.09	31599	99.47
789	22	0.07	31621	99.54
790	17	0.05	31638	99.59
791	23	0.07	31661	99.67
792	24	0.08	31685	99.74
793	13	0.04	31698	99.78
794	11	0.03	31709	99.82
795	8	0.03	31717	99.84
796	8	0.03	31725	99.87
797	6	0.02	31731	99.89
798	2	0.01	31733	99.89
799	6	0.02	31739	99.91
800	7	0.02	31746	99.93
801	6	0.02	31752	99.95
802	4	0.01	31756	99.97
803	1	0.00	31757	99.97
804	2	0.01	31759	99.97
805	1	0.00	31760	99.98
807	2	0.01	31762	99.98
808	1	0.00	31763	99.99
816	1	0.00	31764	99.99
824	1	0.00	31765	99.99
836	1	0.00	31766	100.00
842	1	0.00	31767	100.00

Appendix G: Scale Score Distribution Graphs

Figure G.1. Scale Score Distribution—Mathematics Grade 3

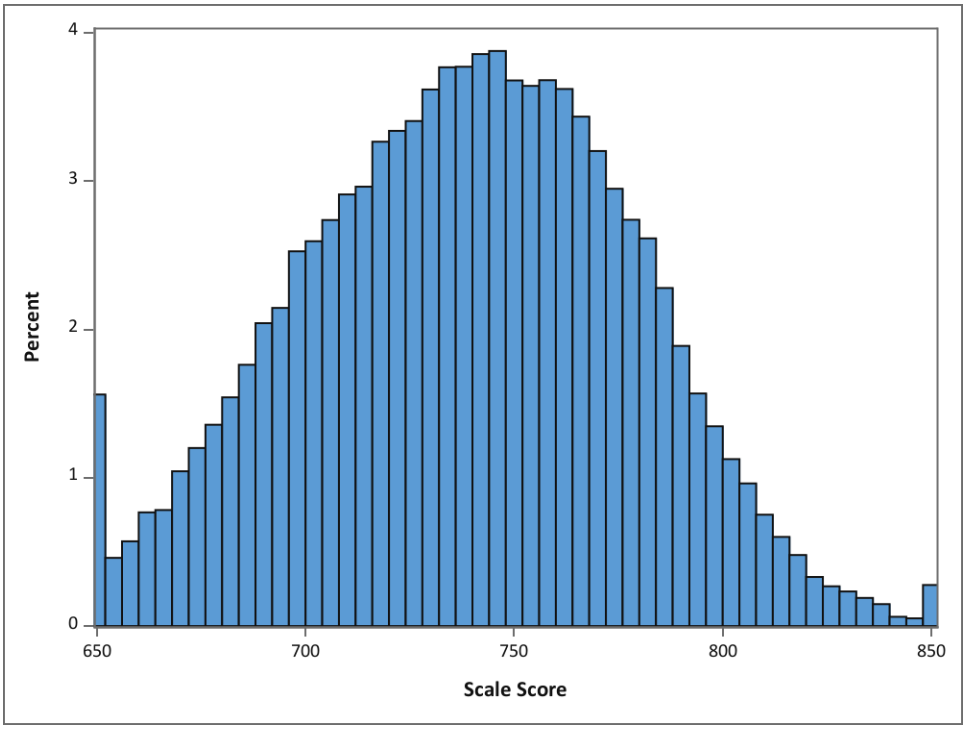


Figure G.2. Scale Score Distribution—Mathematics Grade 4

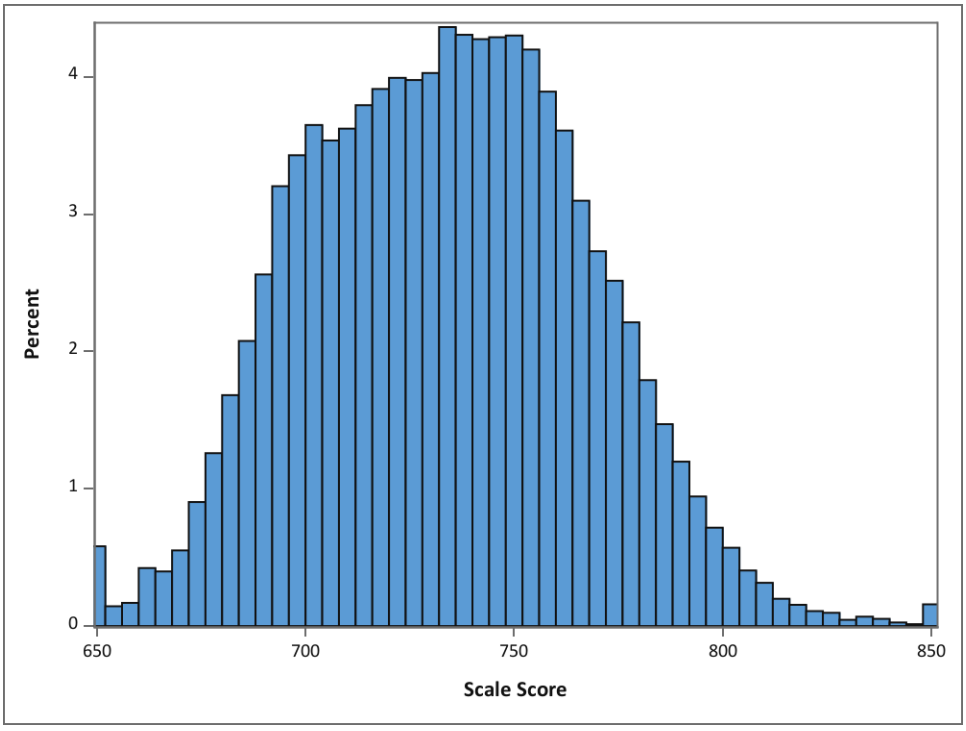


Figure G.3. Scale Score Distribution—Mathematics Grade 5

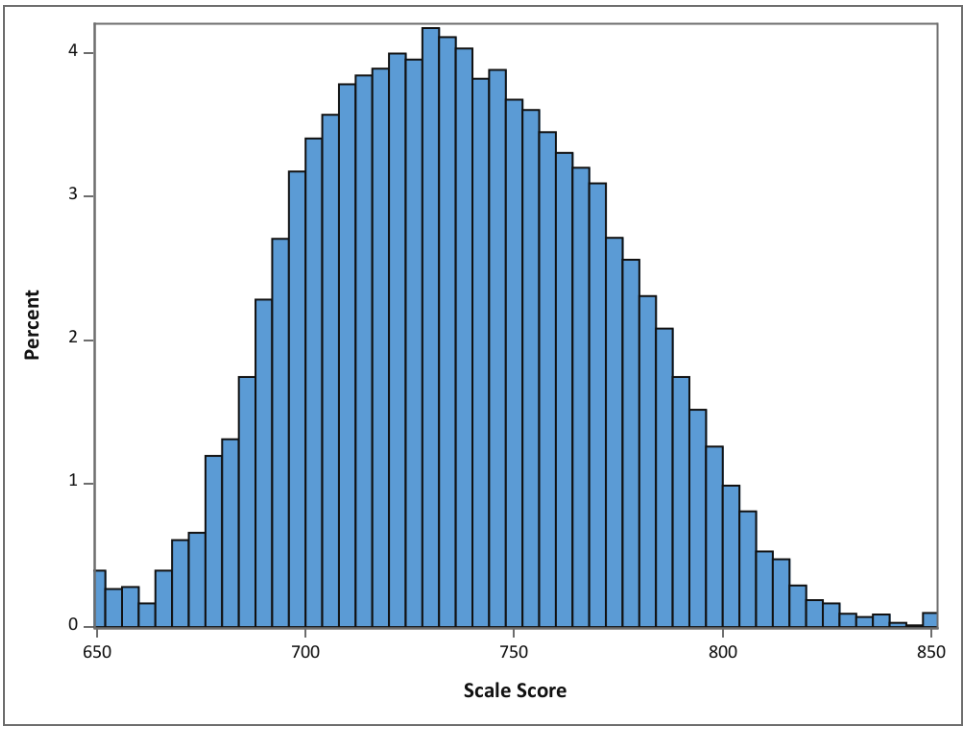


Figure G.4. Scale Score Distribution—Mathematics Grade 6

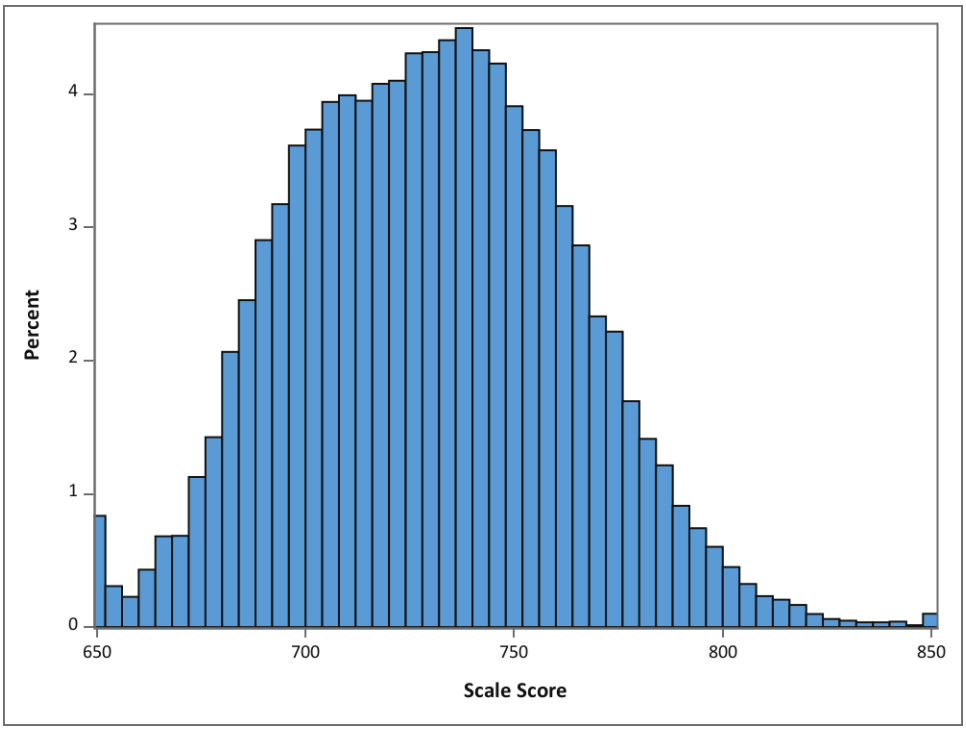


Figure G.5. Scale Score Distribution—Mathematics Grade 7

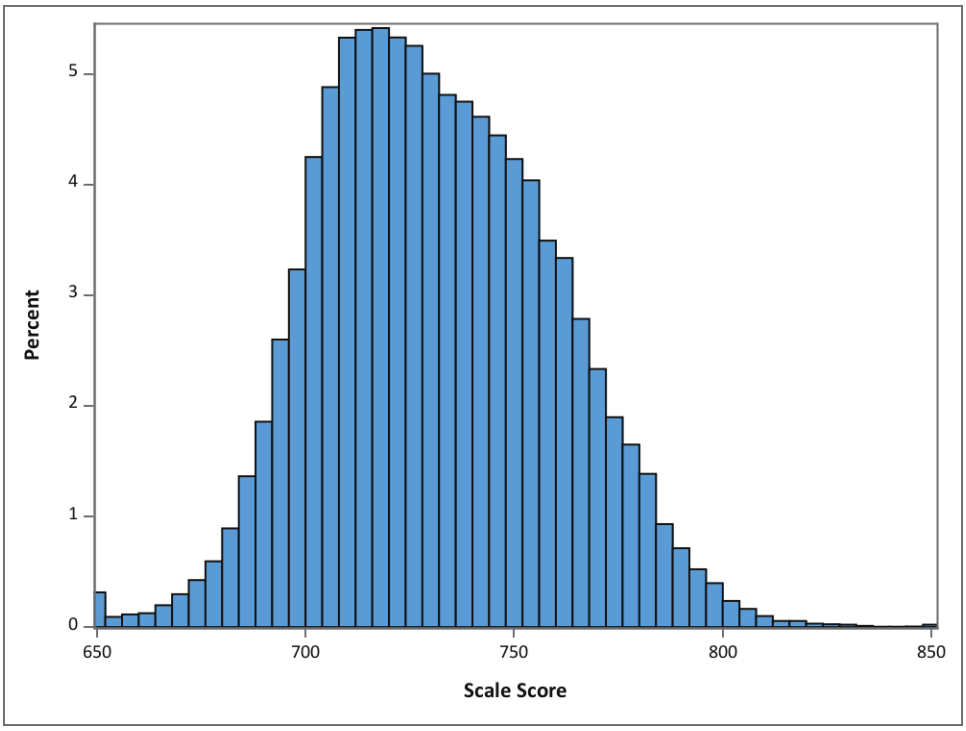


Figure G.6. Scale Score Distribution—Mathematics Grade 8

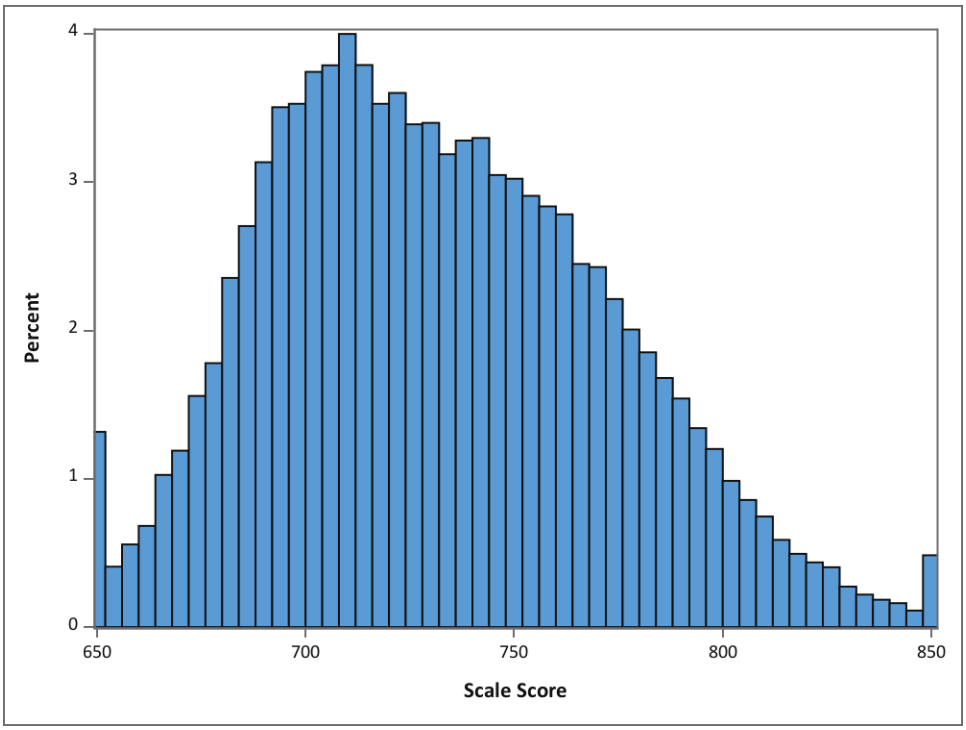


Figure G.7. Scale Score Distribution—ELA Grade 3

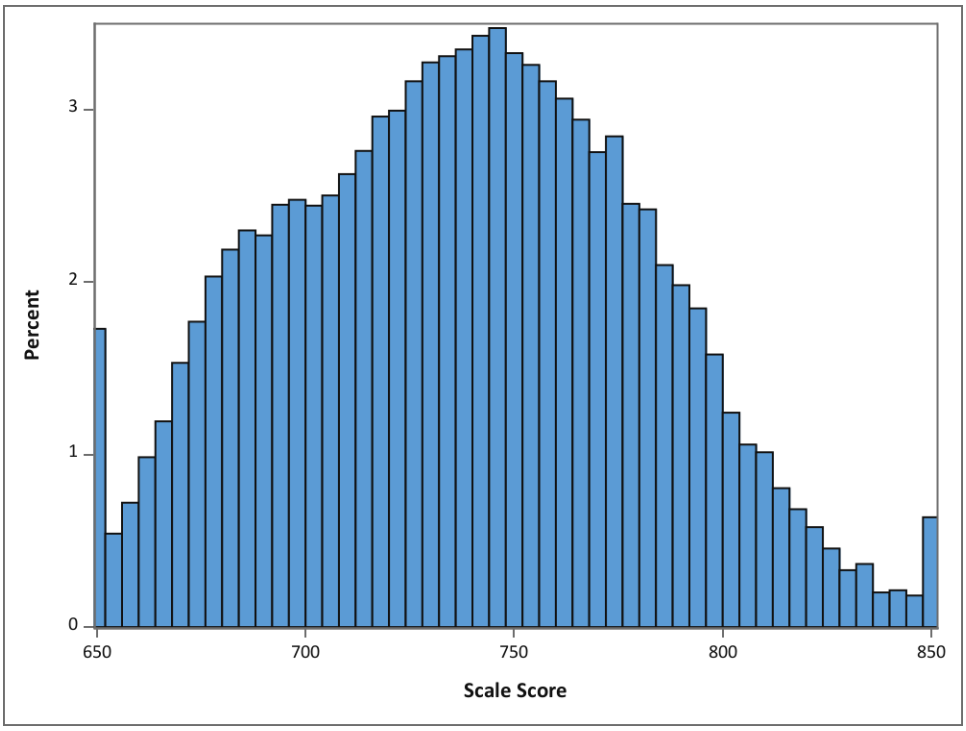


Figure G.8. Scale Score Distribution—ELA Grade 4

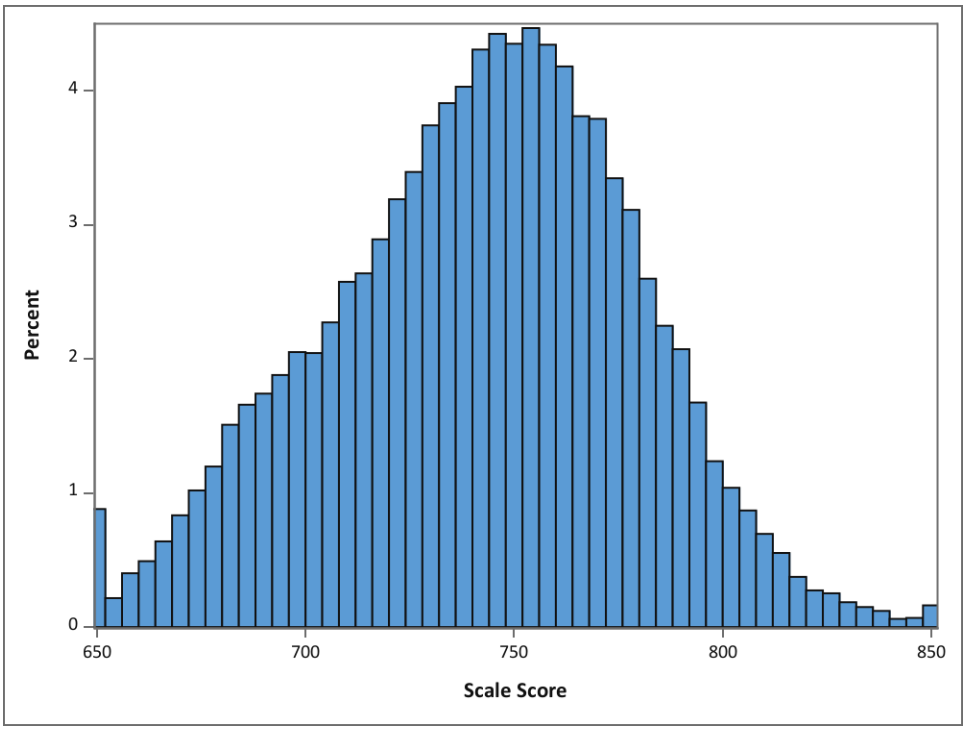


Figure G.9. Scale Score Distribution—ELA Grade 5

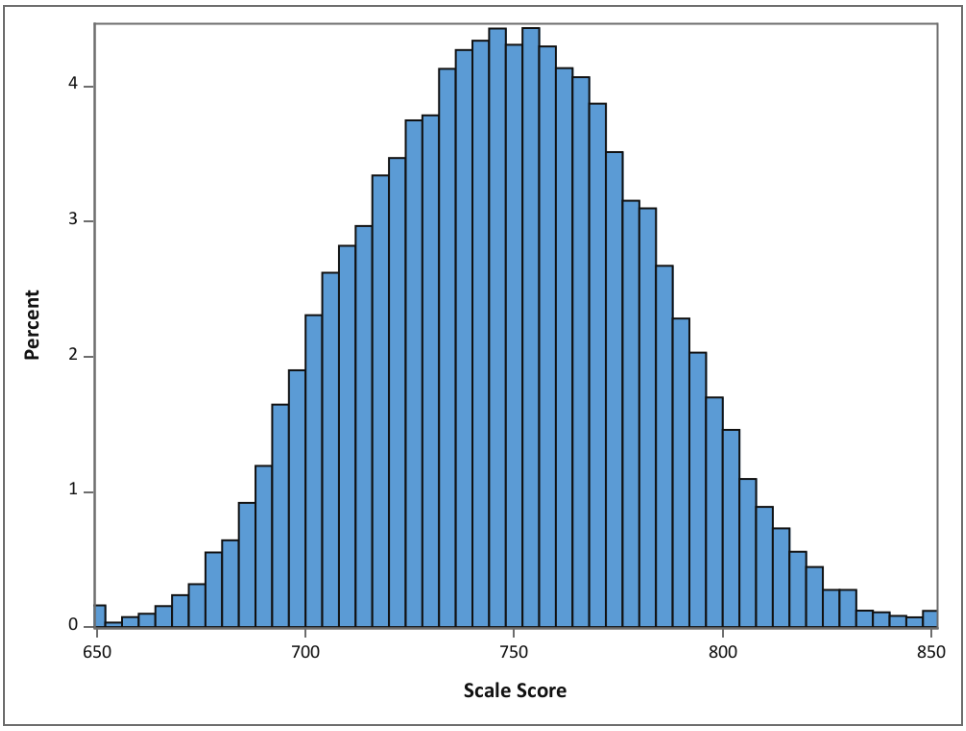


Figure G.10. Scale Score Distribution—ELA Grade 6

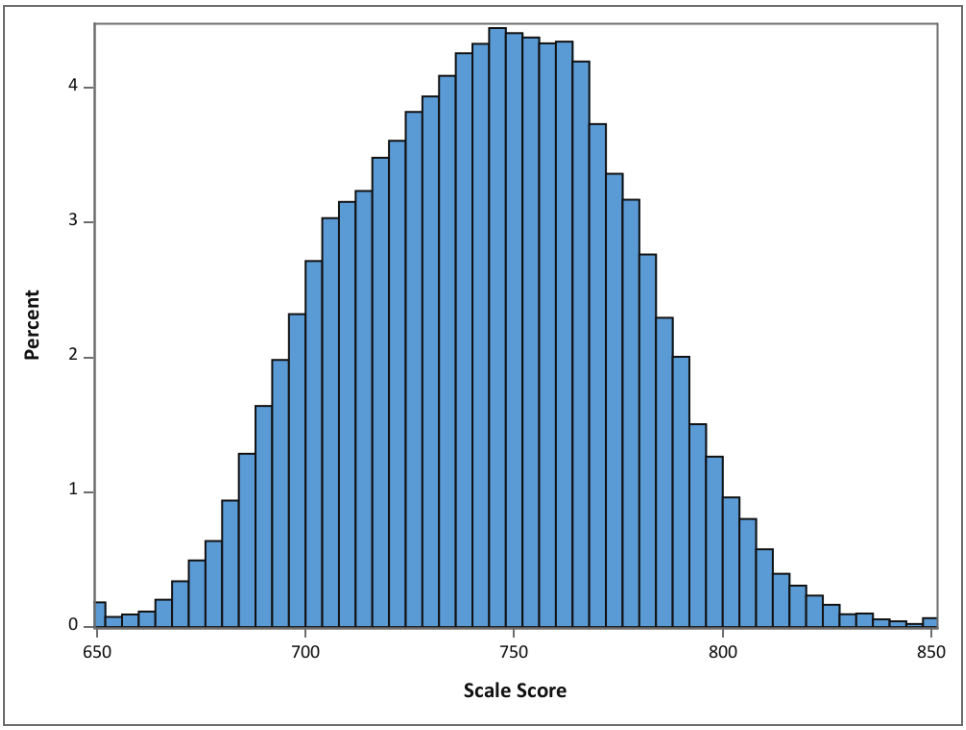


Figure G.11. Scale Score Distribution—ELA Grade 7

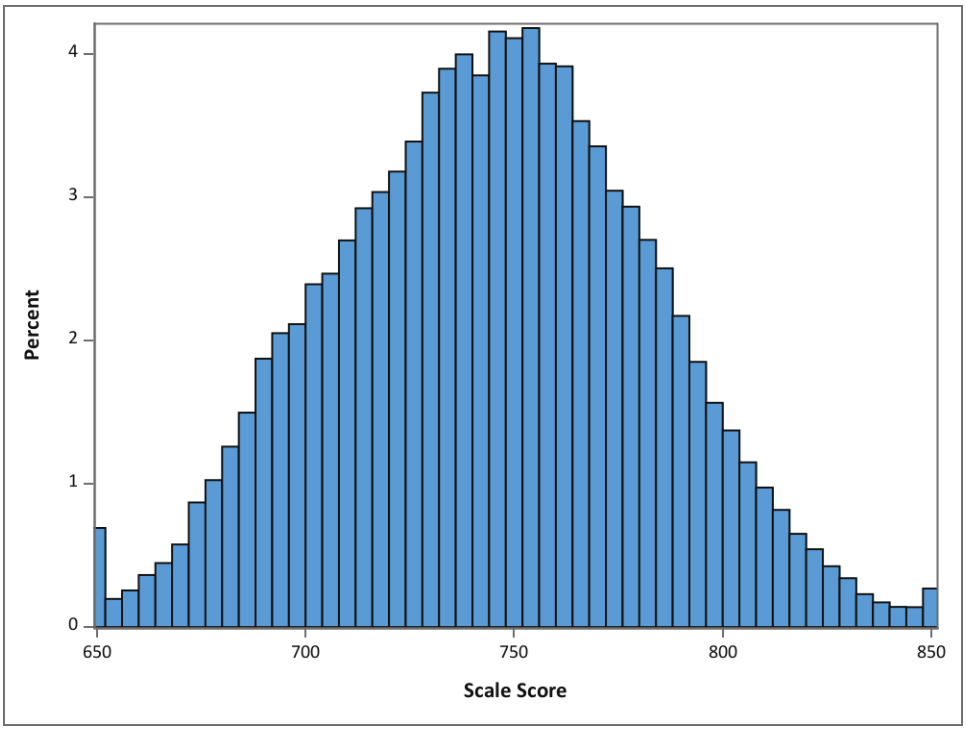


Figure G.12. Scale Score Distribution—ELA Grade 8

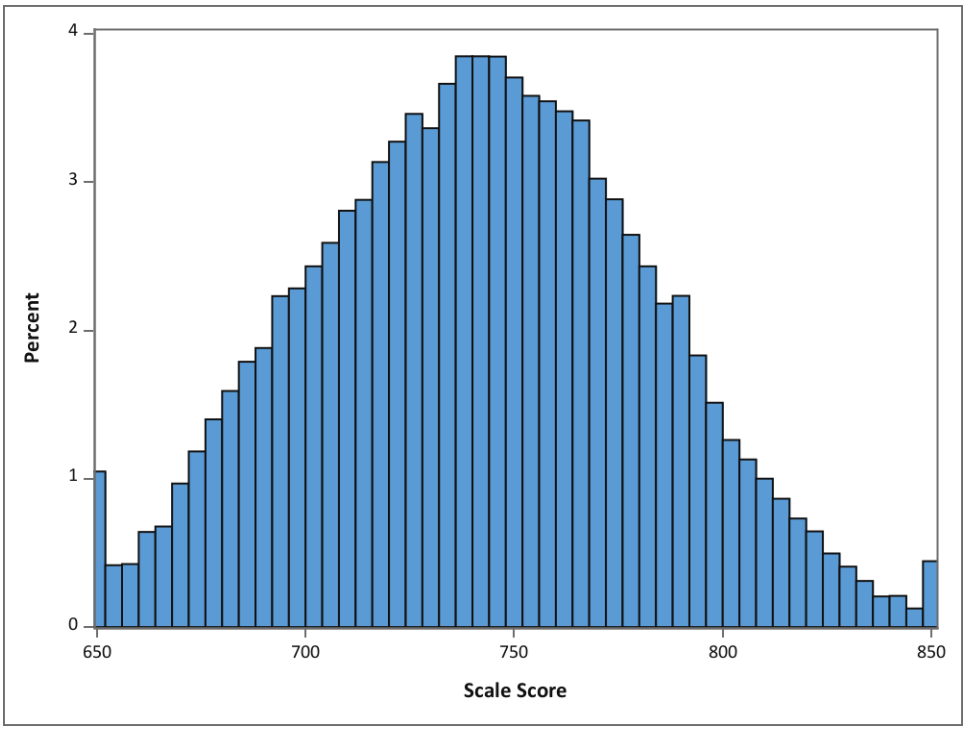


Figure G.13. Scale Score Distribution—CSLA Grade 3

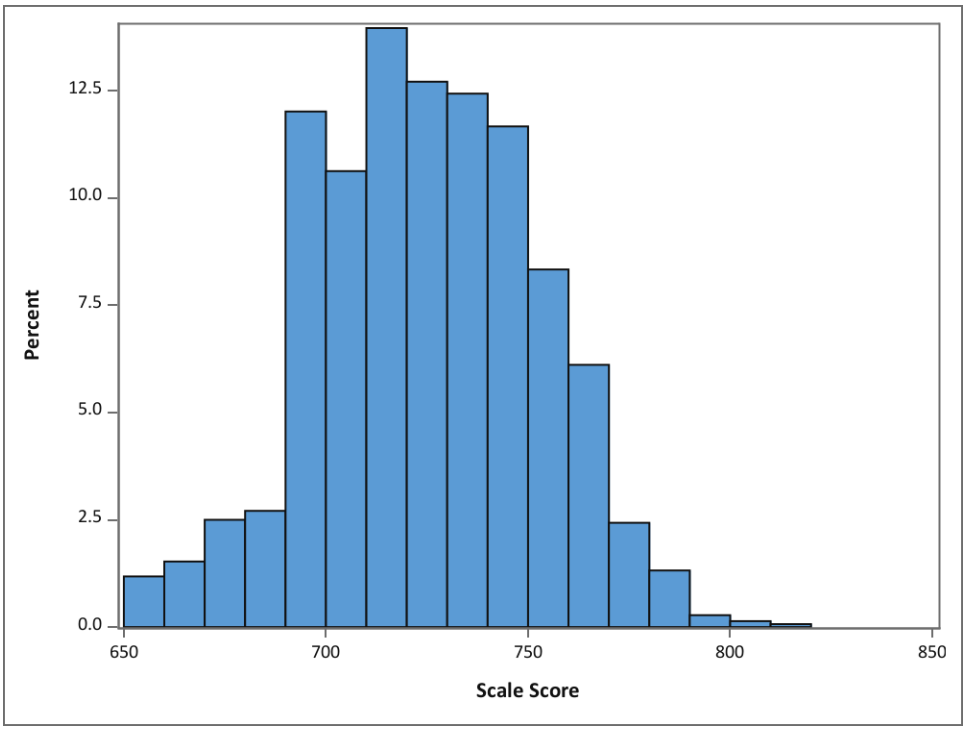


Figure G.14. Scale Score Distribution—CSLA Grade 4

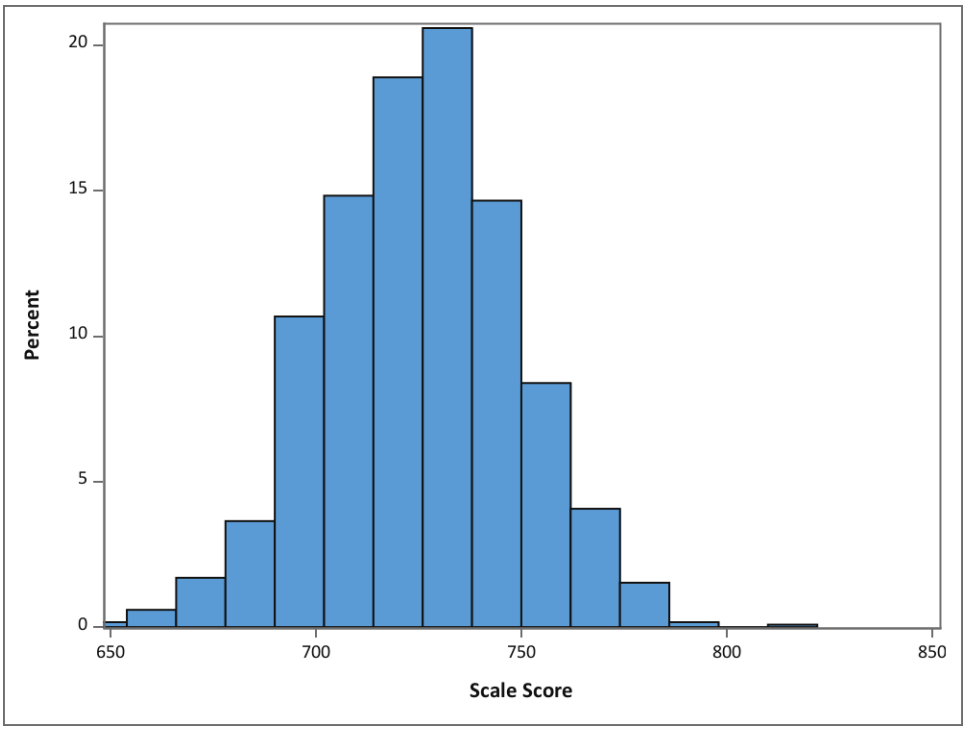


Figure G.15. Scale Score Distribution—Science Grade 5

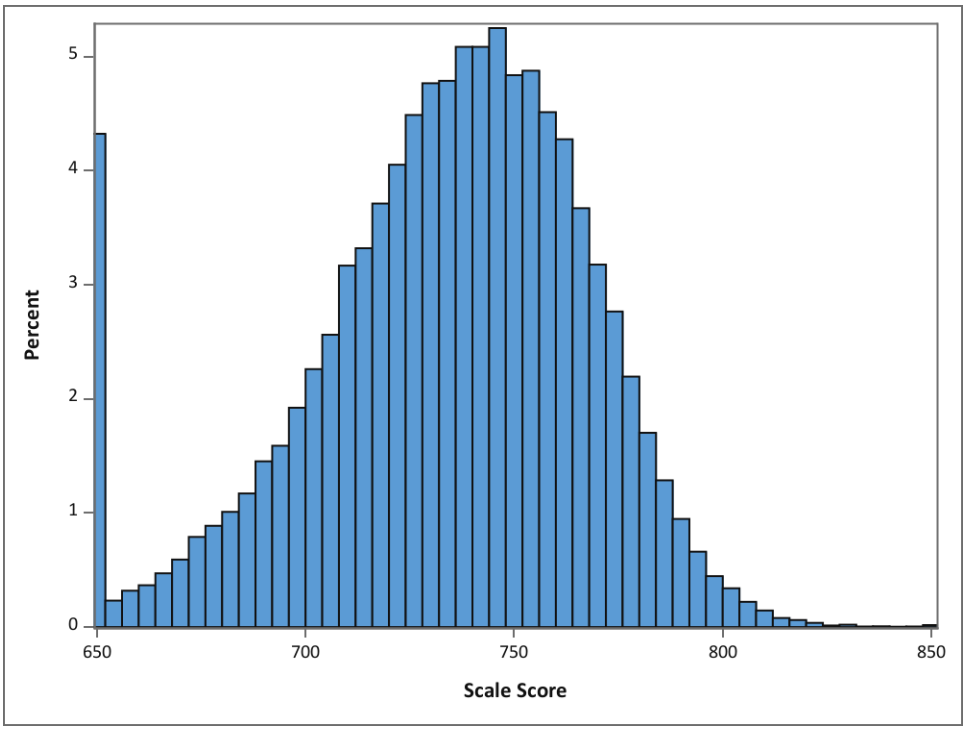


Figure G.16. Scale Score Distribution—Science Grade 8

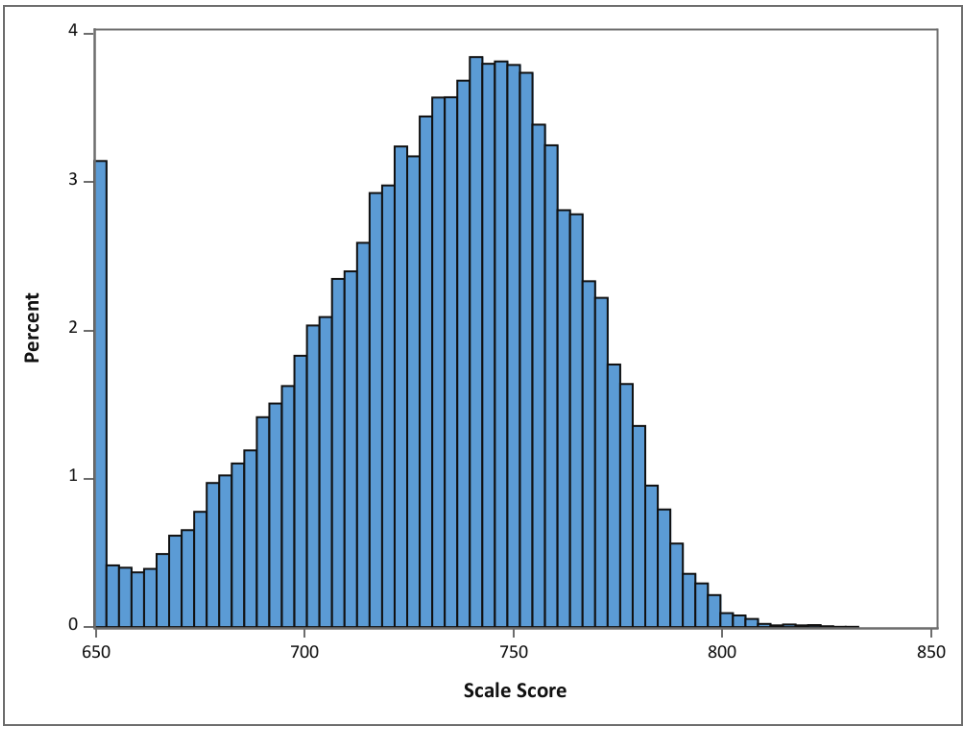
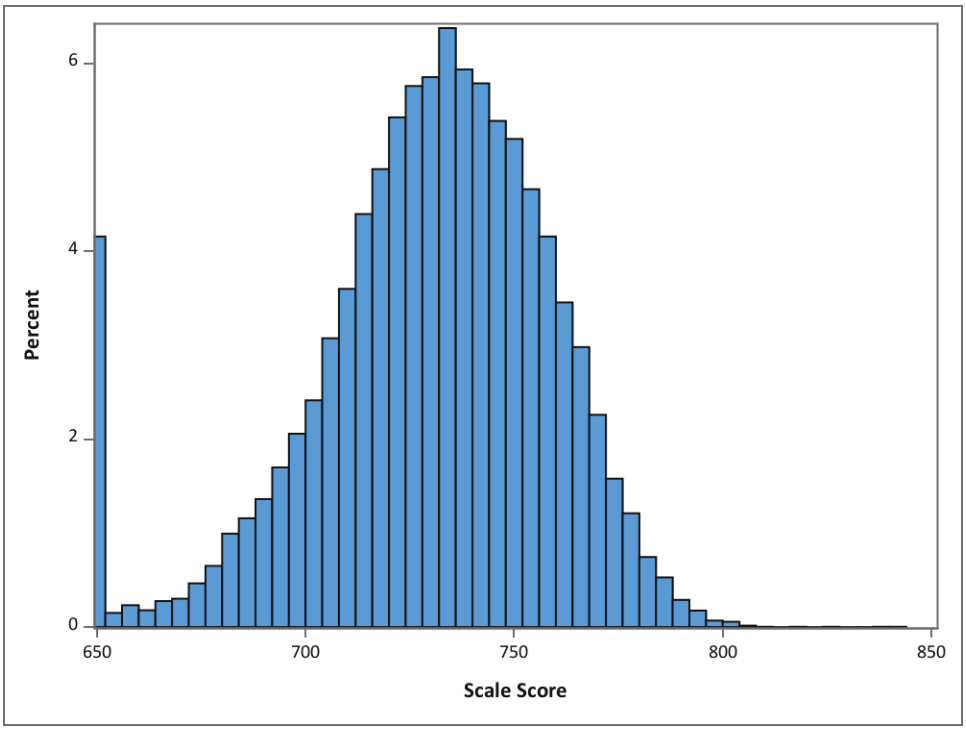


Figure G.17. Scale Score Distribution—Science Grade 11



Appendix H: Scale Score Summary Statistics by Demographic Group

Table H.1. Performance by Subgroup—Mathematics Grade 3

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,964	742.73	37.66	650	850	0.91
IEP	7,418	707.40	37.74	650	850	0.92
No Accommodation	53,110	740.61	38.70	650	850	0.92
Accommodation	4,272	707.79	36.53	650	850	0.92
Am. Indian/Alaska Native	354	714.82	39.08	650	850	0.92
Asian	1,982	756.38	41.13	650	850	0.92
Black	2,646	718.98	37.41	650	837	0.92
Hispanic	19,134	720.87	36.42	650	850	0.91
White	29,847	749.50	36.53	650	850	0.91
Hawaiian/Pacific Islander	191	718.52	38.65	650	842	0.92
Two or More Races	3,225	744.19	38.83	650	850	0.92
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,878	750.17	36.90	650	850	0.91
Economic Disadvantage	22,504	719.55	36.03	650	850	0.91
Female	28,318	735.76	38.61	650	850	0.92
Male	29,064	740.51	40.20	650	850	0.92
Language Proficiency NA	47,355	742.64	38.44	650	850	0.92
Language Proficiency NEP	2,969	692.69	30.00	650	850	0.88
Language Proficiency LEP	5,985	721.31	32.89	650	849	0.90
Language Proficiency FEP	1,073	760.69	31.74	650	850	0.88
Not Migrant	57,188	738.26	39.47	650	850	0.92
Migrant	194	709.07	36.61	650	815	0.92

*n-count less than 16

Table H.2. Performance by Subgroup—Mathematics Grade 4

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,488	737.19	32.34	650	850	0.91
IEP	7,301	706.72	28.70	650	850	0.90
No Accommodation	51,979	735.55	32.95	650	850	0.91
Accommodation	4,810	708.63	29.00	650	838	0.90
Am. Indian/Alaska Native	360	717.56	28.76	650	826	0.89
Asian	2,061	751.36	34.93	650	850	0.92
Black	2,617	718.18	30.54	650	849	0.90
Hispanic	19,426	718.47	30.07	650	850	0.90
White	29,091	742.94	31.43	650	850	0.90
Hawaiian/Pacific Islander	191	717.92	28.93	650	788	0.89
Two or More Races	3,040	738.92	33.82	650	850	0.91
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,768	743.43	31.99	650	850	0.91
Economic Disadvantage	22,021	717.24	29.25	650	850	0.90

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Female	27,931	731.09	32.55	650	850	0.91
Male	28,858	735.39	34.23	650	850	0.92
Language Proficiency NA	46,522	737.24	32.90	650	850	0.91
Language Proficiency NEP	2,387	696.96	22.20	650	830	0.83
Language Proficiency LEP	5,674	711.58	24.95	650	819	0.86
Language Proficiency FEP	2,206	744.75	28.33	650	850	0.89
Not Migrant	56,583	733.35	33.48	650	850	0.91
Migrant	206	711.98	26.10	666	792	0.90

*n-count less than 16

Table H.3. Performance by Subgroup—Mathematics Grade 5

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,824	741.12	34.04	650	850	0.92
IEP	7,072	708.26	29.17	650	833	0.91
No Accommodation	52,062	739.23	34.74	650	850	0.92
Accommodation	4,834	713.31	30.92	650	850	0.92
Am. Indian/Alaska Native	358	721.61	32.45	650	850	0.92
Asian	1,976	756.00	35.32	650	850	0.92
Black	2,508	721.61	30.88	650	850	0.91
Hispanic	19,695	721.71	30.91	650	850	0.91
White	29,224	747.25	33.81	650	850	0.92
Hawaiian/Pacific Islander	202	720.47	32.02	650	824	0.92
Two or More Races	2,930	741.56	34.98	650	850	0.92
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,951	747.49	33.90	650	850	0.92
Economic Disadvantage	21,945	720.38	30.46	650	850	0.91
Female	27,700	734.55	34.13	650	850	0.92
Male	29,196	739.38	36.00	650	850	0.93
Language Proficiency NA	46,354	741.18	34.80	650	850	0.92
Language Proficiency NEP	1,456	700.08	23.09	650	819	0.85
Language Proficiency LEP	5,196	709.95	25.17	650	850	0.87
Language Proficiency FEP	3,890	737.66	30.57	650	850	0.91
Not Migrant	56,696	737.13	35.18	650	850	0.93
Migrant	200	710.42	25.74	652	787	0.89

*n-count less than 16

Table H.4. Performance by Subgroup—Mathematics Grade 6

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,244	733.37	32.21	650	850	0.90
IEP	6,669	701.25	26.89	650	850	0.86
No Accommodation	51,042	731.87	32.76	650	850	0.90
Accommodation	4,871	705.11	28.67	650	838	0.88
Am. Indian/Alaska Native	342	715.15	29.10	650	830	0.87
Asian	1,970	749.64	36.36	650	850	0.92
Black	2,412	714.22	30.47	650	850	0.89
Hispanic	19,778	714.93	29.15	650	850	0.88
White	28,353	739.35	31.47	650	850	0.90
Hawaiian/Pacific Islander	192	713.80	32.22	650	813	0.90
Two or More Races	2,862	735.20	33.64	650	850	0.91
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,740	739.27	32.15	650	850	0.90
Economic Disadvantage	21,173	713.58	28.65	650	850	0.87
Female	27,266	727.72	32.54	650	850	0.90
Male	28,647	731.27	33.90	650	850	0.91
Language Proficiency NA	45,451	733.69	32.85	650	850	0.90
Language Proficiency NEP	1,364	692.43	21.95	650	778	0.75
Language Proficiency LEP	4,066	701.12	23.73	650	797	0.80
Language Proficiency FEP	5,032	725.10	28.01	650	850	0.87
Not Migrant	55,735	729.61	33.28	650	850	0.91
Migrant	178	706.17	27.07	650	786	0.87

*n-count less than 16

Table H.5. Performance by Subgroup—Mathematics Grade 7

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	48,095	733.85	27.50	650	850	0.90
IEP	6,053	708.20	21.10	650	805	0.81
No Accommodation	49,474	732.75	27.81	650	850	0.90
Accommodation	4,674	712.21	23.27	650	850	0.86
Am. Indian/Alaska Native	351	717.57	25.81	650	800	0.88
Asian	1,737	749.85	30.87	650	850	0.92
Black	2,349	718.12	25.35	650	825	0.87
Hispanic	19,719	718.74	24.03	650	839	0.85
White	27,182	739.56	26.68	650	850	0.89
Hawaiian/Pacific Islander	161	713.78	25.17	660	783	0.86
Two or More Races	2,640	736.01	29.09	650	850	0.91
Missing	*	*	*	*	*	*
No Economic Disadvantage	33,499	739.11	27.52	650	850	0.90
Economic Disadvantage	20,649	717.79	23.53	650	817	0.85
Female	26,065	730.07	28.20	650	850	0.90
Male	28,083	731.83	27.88	650	850	0.90

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	44,035	734.64	27.80	650	850	0.90
Language Proficiency NEP	1,470	702.13	17.78	650	773	0.60
Language Proficiency LEP	4,000	708.61	18.39	650	789	0.68
Language Proficiency FEP	4,643	724.64	24.32	650	850	0.86
Not Migrant	53,973	731.04	28.05	650	850	0.90
Migrant	175	711.90	21.01	650	770	0.75

*n-count less than 16

Table H.6. Performance by Subgroup—Mathematics Grade 8

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	46,428	735.78	40.30	650	850	0.91
IEP	5,608	698.44	28.32	650	850	0.82
No Accommodation	47,522	734.22	40.61	650	850	0.91
Accommodation	4,514	705.77	33.81	650	850	0.88
Am. Indian/Alaska Native	361	715.52	35.54	650	850	0.88
Asian	1,700	759.54	44.61	650	850	0.92
Black	2,285	715.64	35.57	650	850	0.89
Hispanic	19,525	714.06	34.32	650	850	0.87
White	25,671	744.55	39.66	650	850	0.91
Hawaiian/Pacific Islander	155	718.35	37.05	650	847	0.90
Two or More Races	2,333	738.13	42.00	650	850	0.91
Missing	*	*	*	*	*	*
No Economic Disadvantage	32,081	743.66	40.51	650	850	0.91
Economic Disadvantage	19,955	712.62	33.48	650	850	0.87
Female	24,738	731.30	40.18	650	850	0.91
Male	27,298	732.17	41.46	650	850	0.91
Language Proficiency NA	42,810	737.24	40.72	650	850	0.91
Language Proficiency NEP	1,433	690.72	23.24	650	799	0.66
Language Proficiency LEP	3,637	698.97	25.30	650	832	0.72
Language Proficiency FEP	4,156	718.09	32.59	650	850	0.86
Not Migrant	51,865	731.86	40.85	650	850	0.91
Migrant	171	701.06	31.00	650	842	0.86

*n-count less than 16

Table H.7. Performance by Subgroup—ELA Grade 3

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	48,525	742.90	41.68	650	850	0.90
IEP	7,212	699.16	36.89	650	850	0.91
No Accommodation	51,771	740.06	42.80	650	850	0.91
Accommodation	3,966	700.44	37.30	650	850	0.90
Am. Indian/Alaska Native	352	715.99	41.05	650	850	0.91
Asian	1,945	749.64	44.45	650	850	0.91
Black	2,644	719.27	40.65	650	850	0.91
Hispanic	17,620	719.42	40.03	650	850	0.91
White	29,752	748.14	41.85	650	850	0.90
Hawaiian/Pacific Islander	184	717.64	37.17	650	840	0.89
Two or More Races	3,237	744.72	43.25	650	850	0.91
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,560	749.60	41.65	650	850	0.90
Economic Disadvantage	21,177	717.08	39.06	650	850	0.90
Female	27,479	740.31	44.40	650	850	0.91
Male	28,258	734.25	42.67	650	850	0.91
Language Proficiency NA	47,322	741.55	42.92	650	850	0.91
Language Proficiency NEP	2,050	683.23	24.68	650	850	0.82
Language Proficiency LEP	5,293	714.68	34.57	650	850	0.88
Language Proficiency FEP	1,072	761.68	33.27	656	850	0.84
Not Migrant	55,564	737.33	43.62	650	850	0.91
Migrant	173	706.67	38.55	650	800	0.90

*n-count less than 16

Table H.8. Performance by Subgroup—ELA Grade 4

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	48,349	747.08	34.02	650	850	0.89
IEP	7,170	704.08	33.50	650	850	0.90
No Accommodation	50,988	744.43	35.71	650	850	0.90
Accommodation	4,531	708.79	34.03	650	850	0.90
Am. Indian/Alaska Native	361	723.35	33.95	650	812	0.90
Asian	2,033	752.83	36.03	650	850	0.89
Black	2,621	725.48	34.91	650	850	0.90
Hispanic	18,202	726.27	34.54	650	850	0.89
White	29,056	751.45	34.70	650	850	0.89
Hawaiian/Pacific Islander	193	724.24	34.64	650	813	0.87
Two or More Races	3,050	747.55	36.32	650	850	0.89
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,522	752.21	34.54	650	850	0.89
Economic Disadvantage	20,997	723.96	33.77	650	850	0.89
Female	27,290	745.24	36.61	650	850	0.90
Male	28,229	737.93	36.80	650	850	0.90

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	46,561	745.67	36.08	650	850	0.90
Language Proficiency NEP	1,647	689.26	23.09	650	799	0.78
Language Proficiency LEP	5,101	715.41	26.33	650	807	0.84
Language Proficiency FEP	2,210	753.42	25.44	650	842	0.83
Not Migrant	55,337	741.62	36.86	650	850	0.90
Migrant	182	711.40	32.69	650	806	0.77

*n-count less than 16

Table H.9. Performance by Subgroup—ELA Grade 5

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,580	751.91	32.08	650	850	0.88
IEP	7,077	715.86	28.33	650	847	0.88
No Accommodation	51,842	749.85	33.10	650	850	0.89
Accommodation	4,815	721.05	29.85	650	849	0.89
Am. Indian/Alaska Native	362	732.29	31.28	653	847	0.90
Asian	1,954	757.76	33.27	650	850	0.88
Black	2,504	733.90	30.98	650	843	0.88
Hispanic	19,484	732.90	30.93	650	850	0.88
White	29,208	757.31	32.00	650	850	0.88
Hawaiian/Pacific Islander	201	732.91	30.71	662	821	0.88
Two or More Races	2,941	752.62	33.30	650	850	0.89
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,895	757.40	32.24	650	850	0.88
Economic Disadvantage	21,762	731.38	29.86	650	850	0.88
Female	27,614	751.19	34.21	650	850	0.89
Male	29,043	743.81	33.02	650	850	0.89
Language Proficiency NA	46,414	751.94	33.05	650	850	0.89
Language Proficiency NEP	1,159	699.71	20.15	650	789	0.78
Language Proficiency LEP	5,185	717.15	22.64	650	835	0.82
Language Proficiency FEP	3,899	747.88	25.84	659	850	0.84
Not Migrant	56,468	747.50	33.79	650	850	0.89
Migrant	189	717.99	24.29	658	793	0.81

*n-count less than 16

Table H.10. Performance by Subgroup—ELA Grade 6

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	48,929	747.07	31.43	650	850	0.90
IEP	6,673	711.20	26.72	650	850	0.89
No Accommodation	50,912	745.25	32.25	650	850	0.90
Accommodation	4,690	715.83	29.13	650	850	0.90
Am. Indian/Alaska Native	345	729.36	31.32	656	817	0.91
Asian	1,937	756.54	33.15	650	850	0.90
Black	2,397	730.09	31.07	650	831	0.90

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Hispanic	19,559	728.98	30.07	650	850	0.89
White	28,323	752.05	31.35	650	850	0.90
Hawaiian/Pacific Islander	191	728.38	31.43	650	850	0.90
Two or More Races	2,846	749.07	32.62	650	850	0.90
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,649	752.09	31.50	650	850	0.90
Economic Disadvantage	20,953	727.35	29.53	650	850	0.89
Female	27,126	746.52	33.41	650	850	0.90
Male	28,476	739.19	32.25	650	850	0.91
Language Proficiency NA	45,433	747.08	32.44	650	850	0.90
Language Proficiency NEP	1,067	697.17	18.48	650	770	0.78
Language Proficiency LEP	4,060	711.62	21.74	650	800	0.83
Language Proficiency FEP	5,042	738.69	26.21	650	834	0.87
Not Migrant	55,431	742.85	33.01	650	850	0.90
Migrant	171	715.44	26.80	650	784	0.89

*n-count less than 16

Table H.11. Performance by Subgroup—ELA Grade 7

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	47,822	748.68	36.34	650	850	0.90
IEP	6,073	707.11	30.00	650	850	0.87
No Accommodation	49,420	746.78	37.16	650	850	0.90
Accommodation	4,475	713.27	33.64	650	850	0.89
Am. Indian/Alaska Native	349	726.72	35.65	650	847	0.89
Asian	1,721	762.69	39.70	650	850	0.90
Black	2,349	729.86	35.23	650	850	0.89
Hispanic	19,487	727.96	34.53	650	850	0.89
White	27,185	755.14	35.80	650	850	0.89
Hawaiian/Pacific Islander	158	722.27	33.11	650	793	0.87
Two or More Races	2,637	751.74	37.90	650	850	0.90
Missing	*	*	*	*	*	*
No Economic Disadvantage	33,423	754.68	36.45	650	850	0.90
Economic Disadvantage	20,472	726.55	33.86	650	850	0.89
Female	25,982	749.49	37.67	650	850	0.90
Male	27,913	738.88	37.63	650	850	0.90
Language Proficiency NA	44,086	749.34	37.04	650	850	0.90
Language Proficiency NEP	1,167	690.47	21.72	650	795	0.80
Language Proficiency LEP	3,992	708.57	25.75	650	812	0.83
Language Proficiency FEP	4,650	737.20	30.81	650	850	0.87
Not Migrant	53,731	744.08	38.00	650	850	0.90
Migrant	164	715.00	32.39	650	803	0.88

*n-count less than 16

Table H.12. Performance by Subgroup—ELA Grade 8

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	46,148	746.20	39.11	650	850	0.90
IEP	5,612	701.80	30.65	650	850	0.88
No Accommodation	47,482	744.23	39.90	650	850	0.90
Accommodation	4,278	709.76	35.75	650	850	0.90
Am. Indian/Alaska Native	358	727.82	37.13	650	842	0.90
Asian	1,672	760.84	41.43	650	850	0.90
Black	2,268	728.63	38.45	650	850	0.90
Hispanic	19,315	725.30	37.17	650	850	0.89
White	25,646	752.97	38.83	650	850	0.90
Hawaiian/Pacific Islander	155	728.97	40.93	650	829	0.90
Two or More Races	2,340	748.61	39.88	650	850	0.90
Missing	*	*	*	*	*	*
No Economic Disadvantage	32,002	752.40	39.20	650	850	0.90
Economic Disadvantage	19,758	723.55	36.55	650	850	0.89
Female	24,638	748.30	40.63	650	850	0.90
Male	27,122	735.10	39.73	650	850	0.90
Language Proficiency NA	42,820	747.13	39.73	650	850	0.90
Language Proficiency NEP	1,168	681.88	21.96	650	785	0.81
Language Proficiency LEP	3,632	704.83	26.56	650	803	0.84
Language Proficiency FEP	4,140	730.83	31.66	650	850	0.88
Not Migrant	51,594	741.50	40.66	650	850	0.90
Migrant	166	706.48	35.63	650	850	0.89

*n-count less than 16

Table H.13. Performance by Subgroup—CSLA Grade 3

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	1,298	726.57	26.81	650	815	0.87
IEP	142	702.92	22.67	650	768	0.83
No Accommodation	1,239	726.31	26.94	650	815	0.87
Accommodation	201	711.43	26.40	650	804	0.86
Am. Indian/Alaska Native	*	*	*	*	*	*
Asian	*	*	*	*	*	*
Black	*	*	*	*	*	*
Hispanic	1,430	724.27	27.40	650	815	0.87
White	*	*	*	*	*	*
Hawaiian/Pacific Islander	*	*	*	*	*	*
Two or More Races	*	*	*	*	*	*
Missing	*	*	*	*	*	*
No Economic Disadvantage	241	723.17	28.72	650	809	0.87
Economic Disadvantage	1,199	724.45	27.07	650	815	0.87
Female	758	728.42	27.15	650	815	0.87
Male	682	719.58	26.83	650	789	0.87

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	*	*	*	*	*	*
Language Proficiency NEP	747	716.48	26.41	650	815	0.86
Language Proficiency LEP	693	732.59	25.86	650	809	0.87
Language Proficiency FEP	*	*	*	*	*	*
Not Migrant	1,426	724.30	27.33	650	815	0.87
Migrant	*	*	*	*	*	*

*n-count less than 16

Table H.14. Performance by Subgroup—CSLA Grade 4

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	1,070	727.02	22.61	650	819	0.83
IEP	110	701.26	22.20	650	763	0.85
No Accommodation	976	727.18	23.10	656	819	0.84
Accommodation	204	712.36	23.21	650	766	0.84
Am. Indian/Alaska Native	*	*	*	*	*	*
Asian	*	*	*	*	*	*
Black	*	*	*	*	*	*
Hispanic	1,170	724.65	23.82	650	819	0.84
White	*	*	*	*	*	*
Hawaiian/Pacific Islander	*	*	*	*	*	*
Two or More Races	*	*	*	*	*	*
Missing	*	*	*	*	*	*
No Economic Disadvantage	191	724.99	22.15	667	780	0.83
Economic Disadvantage	989	724.54	24.09	650	819	0.84
Female	604	729.55	23.42	656	819	0.83
Male	576	719.44	23.05	650	782	0.84
Language Proficiency NA	*	*	*	*	*	*
Language Proficiency NEP	604	719.00	23.65	650	819	0.84
Language Proficiency LEP	576	730.50	22.47	656	791	0.83
Language Proficiency FEP	*	*	*	*	*	*
Not Migrant	1,163	724.64	23.81	650	819	0.84
Migrant	17	722.82	21.94	675	766	0.83

*n-count less than 16

Table H.15. Performance by Subgroup—Science Grade 5

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	49,442	737.36	31.44	650	850	0.89
IEP	6,986	702.63	35.36	650	828	0.90
No Accommodation	52,265	734.95	33.08	650	850	0.89
Accommodation	4,163	709.33	35.54	650	808	0.90
Am. Indian/Alaska Native	359	718.09	34.55	650	809	0.89
Asian	1,967	743.48	31.65	650	828	0.89
Black	2,481	716.45	33.46	650	799	0.85
Hispanic	19,509	717.78	33.01	650	850	0.87
White	29,002	743.83	30.04	650	850	0.88
Hawaiian/Pacific Islander	201	713.37	33.88	650	787	0.87
Two or More Races	2,906	738.46	32.61	650	850	0.89
Missing	*	*	*	*	*	*
No Economic Disadvantage	34,695	743.48	30.24	650	850	0.88
Economic Disadvantage	21,733	716.43	32.88	650	830	0.86
Female	27,472	732.43	32.52	650	850	0.88
Male	28,956	733.66	35.22	650	850	0.90
Language Proficiency NA	45,968	737.92	32.25	650	850	0.89
Language Proficiency NEP	1,433	683.42	29.54	650	824	0.77
Language Proficiency LEP	5,151	703.01	28.49	650	802	0.78
Language Proficiency FEP	3,876	733.66	25.30	650	823	0.84
Not Migrant	56,232	733.17	33.90	650	850	0.89
Migrant	196	702.77	32.18	650	779	0.82

*n-count less than 16

Table H.16. Performance by Subgroup—Science Grade 8

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	45,490	734.22	31.32	650	831	0.92
IEP	5,457	700.59	31.12	650	806	0.90
No Accommodation	47,278	732.45	32.23	650	831	0.92
Accommodation	3,669	707.03	33.43	650	811	0.92
Am. Indian/Alaska Native	352	722.27	30.33	650	797	0.90
Asian	1,671	744.97	31.80	650	826	0.92
Black	2,199	716.73	32.35	650	795	0.90
Hispanic	19,164	716.28	31.85	650	821	0.90
White	25,117	741.48	29.19	650	831	0.91
Hawaiian/Pacific Islander	153	719.93	33.50	650	783	0.92
Two or More Races	2,286	736.36	31.63	650	818	0.92
Missing	*	*	*	*	*	*
No Economic Disadvantage	31,412	739.88	30.23	650	831	0.92
Economic Disadvantage	19,535	715.72	31.72	650	818	0.90
Female	24,168	731.51	31.69	650	827	0.92
Male	26,779	729.81	34.08	650	831	0.92

Subgroup	N	Mean	SD	Min.	Max.	Alpha
Language Proficiency NA	41,893	735.73	31.24	650	831	0.92
Language Proficiency NEP	1,397	682.70	26.90	650	776	0.77
Language Proficiency LEP	3,568	700.53	26.29	650	775	0.79
Language Proficiency FEP	4,089	720.82	27.65	650	802	0.87
Not Migrant	50,777	730.71	32.95	650	831	0.92
Migrant	170	702.89	30.93	650	787	0.90

*n-count less than 16

Table H.17. Performance by Subgroup—Science Grade 11

Subgroup	N	Mean	SD	Min.	Max.	Alpha
No IEP	28,972	731.31	28.38	650	842	0.89
IEP	2,795	704.99	29.74	650	796	0.85
No Accommodation	29,510	730.20	28.88	650	842	0.90
Accommodation	2,257	713.27	32.38	650	800	0.90
Am. Indian/Alaska Native	223	721.12	26.21	650	774	0.85
Asian	969	736.31	30.08	650	800	0.90
Black	1,399	715.67	29.72	650	795	0.87
Hispanic	13,088	718.12	28.11	650	807	0.86
White	14,679	739.17	26.47	650	842	0.89
Hawaiian/Pacific Islander	93	717.69	32.73	650	781	0.90
Two or More Races	1,309	734.57	29.59	650	836	0.90
Missing	*	*	*	*	*	*
No Economic Disadvantage	20,014	735.32	28.25	650	842	0.90
Economic Disadvantage	11,753	718.22	28.33	650	808	0.86
Female	14,935	728.68	27.15	650	824	0.88
Male	16,832	729.27	31.36	650	842	0.91
Language Proficiency NA	27,032	733.32	27.61	650	842	0.89
Language Proficiency NEP	814	683.63	26.65	650	752	0.47
Language Proficiency LEP	1,985	699.42	25.31	650	775	0.67
Language Proficiency FEP	1,936	718.05	22.37	650	783	0.78
Not Migrant	31,674	729.04	29.45	650	842	0.90
Migrant	93	712.71	29.12	650	765	0.85

*n-count less than 16

Appendix I: Summary Statistics for Points Earned by Subclaim

Table I.1. Points Earned Summary by Subclaim—Mathematics

Subclaim	Grade	Mean	SD	Min.	Max.	Average % Correct
Subclaim A	3	11.4	5.3	0	22	51.91
	4	10.9	6.2	0	24	45.45
	5	11.5	6.0	0	23	49.85
	6	6.8	4.8	0	20	33.79
	7	6.9	4.7	0	23	29.86
	8	9.0	5.2	0	24	37.62
Subclaim B	3	5.0	2.5	0	9	55.70
	4	2.7	1.8	0	7	38.32
	5	3.5	2.3	0	8	43.73
	6	4.7	2.4	0	11	42.26
	7	2.8	1.9	0	8	35.70
	8	2.8	1.9	0	7	40.17
Subclaim C	3	3.6	2.7	0	11	32.65
	4	3.5	3.3	0	11	31.44
	5	2.5	2.7	0	11	22.75
	6	3.4	3.1	0	11	30.99
	7	2.5	2.8	0	11	22.97
	8	1.9	2.4	0	11	16.84
Subclaim D	3	3.1	2.3	0	9	33.99
	4	2.3	2.6	0	9	25.56
	5	2.3	2.5	0	9	25.43
	6	2.0	2.4	0	9	22.33
	7	2.1	2.2	0	9	23.21
	8	1.8	2.4	0	9	19.69

Table I.2. Points Earned Summary by Subclaim—ELA

Subclaim	Grade	Mean	SD	Min.	Max.	Average % Correct
RL	3	6.6	4.2	0	17	38.85
	4	8.5	4.0	0	18	47.04
	5	8.4	4.5	0	18	46.86
	6	8.5	4.7	0	18	47.45
	7	7.0	4.0	0	18	39.08
	8	8.2	4.4	0	18	45.46
RI	3	7.7	3.7	0	14	54.98
	4	6.8	4.3	0	18	37.63
	5	6.3	3.3	0	18	35.14
	6	8.5	4.5	0	22	38.68
	7	7.5	4.6	0	22	33.94
	8	9.1	5.0	0	22	41.26
RV	3	6.2	2.8	0	10	61.63
	4	5.7	2.6	0	10	57.13
	5	4.5	2.0	0	8	56.46
	6	4.5	2.3	0	8	56.44
	7	4.8	2.8	0	10	48.31
	8	5.4	2.7	0	10	54.21
WE (unweighted)	3	1.1	1.1	0	6	18.39
	4	1.9	1.3	0	7	27.00
	5	1.9	1.4	0	7	27.09
	6	1.8	1.6	0	8	22.26
	7	2.4	1.9	0	8	30.02
	8	2.5	2.0	0	8	31.28
WKL	3	1.1	1.1	0	6	18.53
	4	1.5	1.4	0	6	24.34
	5	1.5	1.4	0	6	25.54
	6	1.7	1.6	0	6	28.13
	7	2.1	1.8	0	6	34.45
	8	2.4	2.0	0	6	39.53

Note. RL = Reading: Literary Text, RI = Reading: Informational Text, RV = Reading: Vocabulary, WE = Writing: Written Expression, WKL = Writing: Knowledge and Use of Language Conventions. Results for WE are unweighted.

Table I.3. Points Earned Summary by Subclaim—CSLA

Subclaim	Grade	Mean	SD	Min.	Max.	Average % Correct
RL	3	6.1	4.0	0	17	35.71
	4	8.3	4.4	0	20	41.25
RI	3	3.4	2.8	0	14	24.44
	4	4.3	3.1	0	17	23.76
RV	3	4.3	2.9	0	10	42.56
	4	4.0	2.3	0	8	50.64
WE (unweighted)	3	1.4	1.6	0	6	23.92
	4	1.9	1.8	0	7	26.86
WKL	3	1.6	1.6	0	6	26.97
	4	2.2	1.8	0	6	36.34

Note. RL = Reading: Literary Text, RI = Reading: Informational Text, RV = Reading: Vocabulary, WE = Writing: Written Expression, WKL = Writing: Knowledge and Use of Language Conventions. Results for WE are unweighted.

Table I.4. Points Earned Summary by Subclaim—Science

Subclaim	Grade	Mean	SD	Min.	Max.	Average % Correct
Physical Science	5	6.5	4.3	0	18	36.36
	8	7.3	4.4	0	21	35.04
	11	7.2	4.0	0	18	39.91
Life Science	5	5.2	2.7	0	12	42.92
	8	7.9	4.8	0	22	36.00
	11	5.7	3.3	0	16	35.73
Earth and Space Science	5	8.7	4.4	0	21	41.52
	8	5.7	3.7	0	18	31.71
	11	5.0	3.2	0	16	31.55

Appendix J: Classical Item-Level Statistics

Table J.1. SR Item Classical Statistics—Mathematics Grade 3

Item	Omit %	*P*-value	Item–Total Correlation
1	0.85	0.73	0.50
2	0.28	0.69	0.59
3	0.28	0.45	0.60
4	3.63	0.37	0.65
5	1.02	0.42	0.56
6	2.50	0.47	0.42
7	0.41	0.58	0.56
8	0.46	0.28	0.45
9	0.46	0.79	0.35
10	3.51	0.39	0.61
11	1.03	0.22	0.40
12	0.32	0.62	0.51
13	0.41	0.52	0.67
14	0.39	0.45	0.67
15	1.04	0.58	0.64
16	0.18	0.81	0.45
17	0.08	0.77	0.50
18	3.22	0.54	0.43
19	0.70	0.77	0.51
20	1.30	0.52	0.48
21	0.05	0.83	0.46
22	0.14	0.49	0.44
23	0.67	0.60	0.59
24	0.25	0.80	0.39
25	0.27	0.58	0.59
26	0.10	0.35	0.59
27	2.53	0.34	0.70
28	2.14	0.46	0.46
29	0.77	0.38	0.74
30	0.43	0.35	0.73
31	0.68	0.32	0.78

Table J.2. CR Item Classical Statistics—Mathematics Grade 3

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
1	3	2.70	35.79	17.25	32.43	11.82	–	0.39	0.61
2	4	1.13	32.79	39.66	17.39	7.52	1.51	0.26	0.65

Table J.3. SR Item Classical Statistics—Mathematics Grade 4

Item	Omit %	*P*-value	Item–Total Correlation
1	1.25	0.21	0.53
2	0.06	0.53	0.53
3	0.98	0.40	0.60
4	0.38	0.18	0.56
5	1.05	0.18	0.59
6	0.52	0.62	0.40
7	0.95	0.40	0.49
8	3.17	0.28	0.65
9	1.77	0.41	0.42
10	0.21	0.62	0.53
11	3.81	0.30	0.58
12	0.10	0.42	0.46
13	0.09	0.64	0.52
14	0.43	0.62	0.45
15	0.05	0.72	0.56
16	0.51	0.45	0.53
17	0.10	0.43	0.39
18	3.57	0.45	0.54
19	1.23	0.27	0.40
20	1.56	0.31	0.71
21	1.40	0.53	0.50
22	0.27	0.54	0.67
23	1.03	0.37	0.68
24	0.61	0.36	0.54
25	0.13	0.67	0.61
26	0.35	0.41	0.75
27	6.77	0.26	0.74

Table J.4. CR Item Classical Statistics—Mathematics Grade 4

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	5%	6%	P-value	Item-Total Correlation
1	3	2.33	60.04	11.86	10.47	15.31	–	–	–	0.26	0.69
2	3	3.57	74.06	7.36	5.18	9.83	–	–	–	0.16	0.63
3	6	1.02	41.51	10.15	15.19	6.21	9.34	7.56	9.02	0.31	0.78

Table J.5. SR Item Classical Statistics—Mathematics Grade 5

Item	Omit %	*P*-value	Item–Total Correlation
1	0.08	0.61	0.54
2	1.27	0.35	0.47
3	0.65	0.64	0.51
4	2.23	0.37	0.60
5	0.24	0.35	0.63
6	0.08	0.84	0.30
7	0.96	0.48	0.63
8	0.68	0.37	0.57
9	0.73	0.43	0.57
10	0.25	0.33	0.50
11	0.25	0.67	0.58
12	0.41	0.73	0.33
13	1.79	0.63	0.54
14	0.11	0.63	0.42
15	0.13	0.28	0.41
16	0.07	0.73	0.54
17	0.57	0.53	0.49
18	3.10	0.45	0.37
19	0.14	0.43	0.51
20	0.07	0.58	0.70
21	2.32	0.39	0.61
22	0.49	0.53	0.61
23	1.29	0.37	0.76
24	0.07	0.44	0.72
25	0.62	0.29	0.61
26	1.72	0.28	0.76
27	0.16	0.23	0.77
28	1.85	0.25	0.79

Table J.6. CR Item Classical Statistics—Mathematics Grade 5

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
1	3	5.03	54.18	25.89	10.79	4.10	.	0.20	0.68
2	4	2.18	53.17	14.45	15.12	6.05	9.04	0.25	0.72

Table J.7. SR Item Classical Statistics—Mathematics Grade 6

Item	Omit %	*P*-value	Item–Total Correlation
1	0.19	0.58	0.53
2	0.24	0.22	0.47
3	0.38	0.59	0.43
4	0.67	0.10	0.39
5	1.67	0.32	0.63
6	0.28	0.24	0.54
7	0.33	0.27	0.38
8	0.79	0.24	0.56
9	3.24	0.28	0.56
10	3.12	0.30	0.66
11	2.22	0.24	0.61
12	0.87	0.51	0.49
13	0.64	0.40	0.16
14	0.68	0.31	0.53
15	0.18	0.50	0.40
16	0.36	0.28	0.23
17	0.88	0.27	0.29
18	0.44	0.61	0.49
19	0.17	0.32	0.55
20	0.80	0.40	0.72
21	1.32	0.21	0.67
22	0.03	0.58	0.56
23	0.02	0.44	0.38
24	0.34	0.38	0.68
25	2.92	0.18	0.79

Table J.8. CR Item Classical Statistics—Mathematics Grade 6

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
1	3	0.76	37.04	11.92	29.19	21.09	–	0.45	0.73
2	3	0.82	47.98	18.76	23.83	8.61	–	0.31	0.70
3	4	2.48	67.61	12.60	6.55	7.34	3.42	0.15	0.71
4	4	1.62	40.78	11.53	15.90	20.23	9.95	0.36	0.75

Table J.9. SR Item Classical Statistics—Mathematics Grade 7

Item	Omit %	*P*-value	Item–Total Correlation
1	0.74	0.44	0.58
2	0.21	0.21	0.50
3	0.08	0.49	0.55
4	0.34	0.19	0.56
5	1.14	0.14	0.57
6	0.80	0.31	0.39
7	0.39	0.26	0.67
8	0.40	0.49	0.31
9	0.15	0.67	0.50
10	0.46	0.45	0.44
11	0.12	0.21	0.38
12	2.79	0.57	0.37
13	0.45	0.31	0.27
14	0.35	0.22	0.45
15	0.43	0.23	0.47
16	0.17	0.41	0.21
17	0.21	0.54	0.44
18	0.15	0.21	0.25
19	0.66	0.55	0.46
20	1.09	0.13	0.41
21	0.88	0.16	0.68
22	0.21	0.38	0.47
23	2.14	0.13	0.60
24	0.61	0.22	0.69
25	0.27	0.30	0.71
26	2.00	0.26	0.82

Table J.10. CR Item Classical Statistics—Mathematics Grade 7

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
1	3	4.13	56.03	14.53	15.65	9.66	–	0.25	0.77
2	3	2.80	56.76	29.01	9.56	1.88	–	0.18	0.65
3	4	2.65	56.62	18.62	6.02	7.33	8.76	0.22	0.77

Table J.11. SR Item Classical Statistics—Mathematics Grade 8

Item	Omit %	*P*-value	Item–Total Correlation
1	1.50%	0.15	0.65
2	0.72%	0.39	0.61
3	1.28%	0.26	0.60
4	2.15%	0.13	0.61
5	1.23%	0.33	0.52
6	0.17%	0.31	0.51
7	0.63%	0.26	0.59
8	0.09%	0.43	0.34
9	0.10%	0.41	0.33
10	0.17%	0.34	0.44
11	0.14%	0.29	0.44
12	0.13%	0.84	0.32
13	0.14%	0.73	0.38
14	0.47%	0.27	0.45
15	0.37%	0.66	0.45
16	0.07%	0.54	0.36
17	0.23%	0.48	0.29
18	0.24%	0.51	0.50
19	1.45%	0.26	0.71
20	0.11%	0.24	0.50
21	1.07%	0.38	0.66
22	0.08%	0.46	0.53
23	0.48%	0.33	0.73
24	0.68%	0.23	0.76
25	3.74%	0.18	0.83

Table J.12. CR Item Classical Statistics—Mathematics Grade 8

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
1	3	3.44	67.66	7.93	12.81	8.16	–	0.19	0.76
2	3	1.87	62.94	14.93	7.28	12.99	–	0.23	0.70
3	4	5.14	76.59	8.74	4.14	3.82	1.58	0.09	0.69

Table J.13. SR Item Classical Statistics—ELA Grade 3

Item	Omit %	*P*-value	Item–Total Correlation
1	0.59	0.52	0.50
2	0.89	0.37	0.57
3	0.76	0.47	0.59
4	0.69	0.45	0.68
5	1.59	0.77	0.61
6	2.42	0.78	0.52
7	2.22	0.54	0.60
8	2.70	0.57	0.59
9	1.97	0.57	0.52
10	0.03	0.59	0.54
11	0.14	0.45	0.56
12	0.10	0.48	0.54
13	0.24	0.50	0.60
14	1.46	0.29	0.52
15	0.13	0.57	0.58
16	0.18	0.61	0.53
17	0.23	0.55	0.67
18	0.45	0.54	0.60
19	0.51	0.28	0.43

Table J.14. CR Item Classical Statistics—ELA Grade 3

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
PCR_1_WE	3	1.58	52.05	40.50	5.67	0.20	–	0.17	0.70
PCR_1_WKL	3	1.58	48.96	43.56	5.66	0.24	–	0.19	0.61
PCR_2_WE	3	2.38	49.89	37.28	10.37	0.08	–	0.19	0.70
PCR_2_WKL	3	2.38	48.94	43.81	4.75	0.13	–	0.18	0.59

Table J.15. SR Item Classical Statistics—ELA Grade 4

Item	Omit %	*P*-value	Item–Total Correlation
1	0.09	0.37	0.36
2	0.06	0.50	0.56
3	0.05	0.41	0.41
4	0.07	0.64	0.64
5	0.46	0.40	0.52
6	1.44	0.39	0.43
7	0.62	0.27	0.48
8	0.72	0.30	0.52
9	1.10	0.39	0.35
10	0.85	0.44	0.36
11	1.30	0.39	0.47
12	0.34	0.46	0.47
13	0.23	0.75	0.54
14	0.35	0.44	0.61
15	0.43	0.30	0.39
16	0.40	0.47	0.61
17	0.16	0.61	0.59
18	0.15	0.63	0.44
19	0.16	0.74	0.47
20	0.19	0.50	0.69
21	0.17	0.68	0.65

Table J.16. CR Item Classical Statistics—ELA Grade 4

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
PCR_1_WE	3	0.82	24.91	44.32	24.74	5.21	–	0.36	0.75
PCR_1_WKL	3	0.82	42.36	38.80	14.45	3.57	–	0.26	0.70
PCR_2_WE	4	0.78	35.02	53.11	9.59	1.27	0.23	0.19	0.72
PCR_2_WKL	3	0.78	46.09	43.21	8.52	1.41	–	0.21	0.68

Table J.17. SR Item Classical Statistics—ELA Grade 5

Item	Omit %	*P*-value	Item–Total Correlation
1	0.25	0.78	0.52
2	0.24	0.59	0.57
3	0.41	0.48	0.54
4	0.48	0.49	0.52
5	0.36	0.27	0.44
6	0.33	0.63	0.66
7	0.41	0.58	0.64
8	0.19	0.53	0.50
9	0.18	0.44	0.50
10	0.17	0.32	0.36
11	0.19	0.49	0.49
12	0.18	0.32	0.44
13	0.05	0.35	0.31
14	0.18	0.53	0.32
15	0.08	0.50	0.41
16	0.09	0.22	0.34
17	0.39	0.50	0.57
18	0.05	0.48	0.44
19	0.34	0.37	0.45
20	0.61	0.44	0.49

Table J.18. CR Item Classical Statistics—ELA Grade 5

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
PCR_1_WE	3	0.65	23.45	34.76	34.23	6.92	–	0.41	0.79
PCR_1_WKL	3	0.65	38.76	36.66	19.62	4.32	–	0.30	0.75
PCR_2_WE	4	1.41	48.27	38.28	9.96	1.72	0.37	0.16	0.75
PCR_2_WKL	3	1.41	48.11	39.63	9.04	1.81	-	0.21	0.70

Table J.19. SR Item Classical Statistics—ELA Grade 6

Item	Omit %	*P*-value	Item–Total Correlation
1	0.14	0.60	0.44
2	0.19	0.41	0.63
3	0.17	0.36	0.42
4	0.17	0.39	0.51
5	0.21	0.67	0.53
6	0.16	0.57	0.45
7	0.15	0.41	0.55
8	0.15	0.48	0.62
9	0.12	0.47	0.55
10	0.14	0.48	0.63
11	0.44	0.39	0.46
12	0.54	0.34	0.38
13	0.33	0.42	0.38
14	0.40	0.42	0.36
15	0.05	0.55	0.55
16	0.13	0.41	0.54
17	0.05	0.51	0.57
18	0.04	0.57	0.55
19	0.15	0.36	0.48
20	0.10	0.32	0.37
21	0.03	0.52	0.49
22	0.06	0.64	0.53

Table J.20. CR Item Classical Statistics—ELA Grade 6

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
PCR_1_WE	4	0.95	42.46	35.13	16.61	3.95	0.92	0.21	0.78
PCR_1_WKL	3	0.95	46.89	31.08	15.95	5.14	–	0.26	0.77
PCR_2_WE	4	1.10	35.58	41.64	17.70	3.78	0.20	0.22	0.79
PCR_2_WKL	3	1.10	38.22	39.47	16.98	4.24	–	0.29	0.77

Table J.21. SR Item Classical Statistics—ELA Grade 7

Item	Omit %	*P*-value	Item–Total Correlation
1	0.46	0.39	0.43
2	0.34	0.36	0.56
3	0.48	0.22	0.46
4	0.39	0.49	0.57
5	0.45	0.29	0.35
6	0.50	0.16	0.37
7	1.04	0.38	0.47
8	1.81	0.34	0.51
9	1.80	0.25	0.53
10	1.30	0.42	0.55
11	0.16	0.25	0.42
12	0.10	0.54	0.41
13	0.22	0.42	0.45
14	0.18	0.58	0.58
15	0.25	0.34	0.48
16	0.24	0.29	0.33
17	0.30	0.35	0.52
18	0.24	0.50	0.52
19	0.26	0.50	0.59
20	0.25	0.44	0.39
21	0.25	0.55	0.61
22	0.26	0.58	0.60
23	0.28	0.41	0.51

Table J.22. CR Item Classical Statistics—ELA Grade 7

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
PCR_1_WE	4	1.57	31.01	28.69	24.40	11.46	2.87	0.31	0.82
PCR_1_WKL	3	1.57	36.91	25.80	22.59	13.12	–	0.37	0.81
PCR_2_WE	4	1.55	26.68	38.63	25.62	6.30	1.21	0.28	0.83
PCR_2_WKL	3	1.55	37.66	34.45	20.12	6.22	–	0.31	0.80

Table J.23. SR Item Classical Statistics—ELA Grade 8

Item	Omit %	*P*-value	Item–Total Correlation
1	0.23	0.64	0.59
2	0.31	0.50	0.54
3	0.17	0.25	0.26
4	0.14	0.60	0.60
5	0.03	0.72	0.43
6	0.13	0.56	0.51
7	0.28	0.27	0.37
8	0.66	0.63	0.43
9	1.00	0.46	0.58
10	0.86	0.31	0.50
11	1.25	0.34	0.45
12	0.36	0.54	0.66
13	0.28	0.50	0.56
14	0.34	0.44	0.50
15	0.37	0.43	0.47
16	0.30	0.35	0.49
17	0.32	0.43	0.49
18	0.19	0.56	0.58
19	0.32	0.48	0.57
20	0.32	0.39	0.48
21	0.37	0.58	0.54
22	0.45	0.46	0.45
23	0.46	0.25	0.44

Table J.24. CR Item Classical Statistics—ELA Grade 8

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
PCR_1_WE	4	1.85	33.09	26.86	25.42	10.81	1.97	0.30	0.85
PCR_1_WKL	3	1.85	34.41	26.31	24.63	12.80	–	0.38	0.83
PCR_2_WE	4	2.04	29.53	24.59	29.05	12.26	2.54	0.32	0.85
PCR_2_WKL	3	2.04	31.85	24.41	28.51	13.19	–	0.40	0.82

Table J.25. SR Item Classical Statistics—CSLA Grade 3

Item	Omit %	*P*-value	Item–Total Correlation
1	1.39	0.39	0.57
2	1.94	0.38	0.48
3	1.39	0.50	0.47
4	1.32	0.47	0.56
5	2.29	0.37	0.53
6	11.67	0.38	0.58
7	14.72	0.29	0.44
8	13.68	0.32	0.48
9	13.61	0.27	0.52
10	15.35	0.16	0.18
11	2.78	0.48	0.61
12	3.54	0.33	0.57
13	3.89	0.23	0.24
14	3.26	0.38	0.60
15	2.64	0.48	0.57
16	2.15	0.28	0.40
17	3.33	0.31	0.34
18	2.78	0.39	0.54
19	6.74	0.08	0.39

Table J.26. CR Item Classical Statistics—CSLA Grade 3

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
PCR_1_WE	3	6.74	40.49	33.33	14.38	5.07	–	0.26	0.81
PCR_1_WKL	3	6.74	37.29	36.81	13.19	5.97	–	0.27	0.61
PCR_2_WE	3	8.47	50.21	22.08	13.82	5.42	–	0.22	0.70
PCR_2_WKL	3	8.47	37.15	36.04	10.28	8.06	–	0.27	0.60

Table J.27. SR Item Classical Statistics—CSLA Grade 4

Item	Omit %	*P*-value	Item–Total Correlation
1	1.02	0.61	0.52
2	1.19	0.45	0.34
3	1.36	0.31	0.42
4	1.27	0.61	0.42
5	1.10	0.66	0.53
6	1.19	0.20	0.47
7	1.19	0.55	0.55
8	1.02	0.30	0.52
9	1.78	0.62	0.57
10	1.19	0.53	0.59
11	2.63	0.46	0.45
12	2.88	0.22	0.32
13	3.05	0.19	0.23
14	2.63	0.24	0.22
15	2.80	0.37	0.44
16	3.98	0.21	0.33
17	6.10	0.34	0.52
18	3.14	0.25	0.42
19	3.81	0.36	0.50
20	3.31	0.18	0.16
21	4.07	0.14	0.12

Table J.28. CR Item Classical Statistics—CSLA Grade 4

Item	Max. Points	Omit %	0%	1%	2%	3%	4%	P-value	Item-Total Correlation
PCR_1_WE	3	1.95	45.93	19.58	23.90	8.64	–	0.31	0.68
PCR_1_WKL	3	1.95	28.81	36.69	15.76	16.78	–	0.40	0.55
PCR_2_WE	4	2.71	43.73	26.53	15.51	8.98	2.54	0.24	0.79
PCR_2_WKL	3	2.71	35.17	36.61	13.64	11.86	–	0.33	0.57

Table J.29. SR Item Classical Statistics—Science Grade 5

Item	Omit %	*P*-value	Item–Total Correlation
1	0.64	0.60	0.59
2	0.04	0.41	0.38
3	0.26	0.45	0.59
4	0.38	0.22	0.33
5	1.07	0.23	0.32
6	1.06	0.51	0.32
7	0.10	0.54	0.46
8	1.45	0.15	0.35
9	1.00	0.41	0.45
10	1.30	0.26	0.53
11	0.08	0.41	0.28
12	0.05	0.34	0.59
13	0.08	0.38	0.53
14	0.07	0.31	0.45
15	0.01	0.78	0.46
16	0.29	0.36	0.20
17	0.45	0.49	0.24
18	0.28	0.32	0.22
19	0.21	0.61	0.34
20	0.23	0.32	0.28
21	0.25	0.28	0.25
22	0.67	0.55	0.33
23	0.12	0.42	0.27
24	0.01	0.56	0.32
25	0.41	0.58	0.46
26	0.01	0.70	0.41
27	0.08	0.46	0.45

Table J.30. CR Item Classical Statistics—Science Grade 5

Item	Max. Points	Omit %	0%	1%	2%	P-value	Item-Total Correlation
1	2	4.78	80.48	13.75	1.00	0.08	0.38
2	2	2.19	74.50	18.53	4.78	0.14	0.35
3	2	3.98	91.04	4.18	0.80	0.03	0.16
4	2	3.59	63.55	27.89	4.98	0.19	0.42
5	2	2.79	65.94	12.55	18.73	0.25	0.58
6	2	2.19	65.94	28.29	3.59	0.18	0.51
7	2	6.37	78.69	12.95	1.99	0.08	0.50
8	2	5.58	68.53	16.14	9.76	0.18	0.54
9	2	2.79	70.72	17.13	9.36	0.18	0.57
10	2	2.39	84.66	8.96	3.98	0.08	0.46
11	2	2.19	77.09	15.94	4.78	0.13	0.47
12	2	2.19	42.23	48.21	7.37	0.31	0.52
13	2	2.97	54.42	23.91	18.70	0.31	0.64
14	2	2.12	56.17	22.19	19.52	0.31	0.50
15	2	4.34	74.22	16.80	4.64	0.13	0.39

Item	Max. Points	Omit %	0%	1%	2%	P-value	Item-Total Correlation
16	2	1.79	52.86	31.09	14.26	0.30	0.50
17	2	1.07	33.52	14.78	50.62	0.58	0.58
18	2	1.51	44.54	41.07	12.87	0.33	0.56
19	2	5.07	65.56	19.80	9.58	0.19	0.52
20	2	3.02	37.82	31.84	27.32	0.43	0.68
21	2	1.57	36.72	20.40	41.31	0.52	0.65
22	2	1.72	59.03	19.07	20.18	0.30	0.61
23	2	1.49	47.18	18.62	32.72	0.42	0.62
24	2	0.97	19.20	40.02	39.81	0.60	0.64

Table J.31. SR Item Classical Statistics—Science Grade 8

Item	Omit %	*P*-value	Item–Total Correlation
1	0.01	0.52	0.38
2	0.03	0.68	0.46
3	0.11	0.31	0.61
4	0.14	0.45	0.41
5	0.42	0.16	0.48
6	0.06	0.13	0.29
7	0.46	0.10	0.32
8	1.03	0.29	0.38
9	0.19	0.15	0.15
10	0.06	0.61	0.40
11	0.20	0.31	0.49
12	0.31	0.60	0.46
13	0.33	0.26	0.57
14	0.38	0.33	0.40
15	0.40	0.13	0.10
16	0.53	0.40	0.44
17	0.34	0.62	0.47
18	0.06	0.07	0.26
19	0.14	0.16	0.29
20	0.08	0.40	0.45
21	1.09	0.60	0.55
22	0.13	0.38	0.49
23	0.22	0.28	0.37
24	0.09	0.54	0.40
25	0.19	0.21	0.36
26	0.27	0.54	0.20
27	0.05	0.58	0.44
28	0.35	0.49	0.56
29	0.09	0.46	0.54
30	0.07	0.62	0.41
31	0.06	0.71	0.31
32	0.08	0.46	0.38
33	0.16	0.75	0.48

Table J.32. CR Item Classical Statistics—Science Grade 8

Item	Max. Points	Omit %	0%	1%	2%	P-value	Item-Total Correlation
1	2	12.18	81.73	5.45	0.64	0.03	0.48
2	2	7.37	77.24	12.18	3.21	0.09	0.51
3	2	11.22	83.97	4.17	0.64	0.03	0.36
4	2	11.86	73.08	13.46	1.60	0.08	0.53
5	2	10.58	58.01	20.83	10.58	0.21	0.52
6	2	10.58	82.05	5.45	1.92	0.05	0.56
7	2	12.50	81.41	5.13	0.96	0.04	0.48
8	2	8.33	77.56	8.33	5.77	0.10	0.48
9	2	8.01	65.71	19.23	7.05	0.17	0.59
10	2	8.65	86.86	4.49	0.00	0.02	0.21
11	2	6.73	85.90	6.73	0.64	0.04	0.42
12	2	9.94	88.14	1.60	0.32	0.01	0.24
13	2	6.09	60.26	28.85	4.81	0.19	0.53
14	2	10.26	82.05	7.37	0.32	0.04	0.42
15	2	3.77	59.25	27.96	9.02	0.23	0.65
16	2	2.31	56.21	21.79	19.69	0.31	0.61
17	2	2.87	79.37	14.23	3.54	0.11	0.56
18	2	3.81	54.34	27.33	14.52	0.28	0.68
19	2	2.85	45.70	25.36	26.09	0.39	0.57
20	2	2.60	66.47	20.18	10.74	0.21	0.60
21	2	3.63	63.08	23.76	9.53	0.21	0.66
22	2	0.89	57.14	18.89	23.08	0.33	0.45
23	2	2.53	36.19	20.40	40.87	0.51	0.68
24	2	2.18	63.59	26.51	7.72	0.21	0.55
25	2	2.19	74.25	18.92	4.65	0.14	0.55
26	2	3.63	72.64	10.37	13.35	0.19	0.53
27	2	1.79	34.16	36.16	27.89	0.46	0.63
28	2	3.88	52.38	33.31	10.43	0.27	0.59

Table J.33. SR Item Classical Statistics—Science Grade 11

Item	Omit %	*P*-value	Item–Total Correlation
1	0.79	0.18	0.46
2	4.38	0.27	0.46
3	0.15	0.46	0.47
4	1.41	0.46	0.54
5	0.37	0.41	0.27
6	0.29	0.59	0.55
7	0.20	0.07	0.25
8	0.04	0.66	0.58
9	0.07	0.43	0.42
10	0.13	0.19	0.41
11	0.06	0.21	0.37
12	0.07	0.45	0.50
13	0.19	0.50	0.53

Item	Omit %	*P*-value	Item–Total Correlation
14	0.02	0.40	0.48
15	0.40	0.34	0.17
16	0.72	0.51	0.46
17	0.26	0.36	0.37
18	0.29	0.81	0.36
19	0.21	0.43	0.26
20	0.22	0.51	0.38
21	0.06	0.27	0.23
22	0.03	0.57	0.34
23	0.06	0.46	0.42
24	0.06	0.40	0.34
25	2.68	0.39	0.12
26	0.01	0.57	0.34

Table J.34. CR Item Classical Statistics—Science Grade 11

Item	Max. Points	Omit %	0%	1%	2%	P-value	Item-Total Correlation
1	2	18.66	77.03	3.83	0.48	0.02	0.14
2	2	20.10	66.99	12.44	0.48	0.07	0.26
3	2	15.31	65.55	12.44	6.70	0.13	0.53
4	2	10.53	50.72	37.80	0.96	0.20	0.49
5	2	18.18	48.80	27.75	5.26	0.19	0.50
6	2	10.05	78.95	10.05	0.96	0.06	0.33
7	2	16.75	53.11	25.36	4.78	0.17	0.44
8	2	15.31	55.50	25.84	3.35	0.16	0.46
9	2	11.48	52.15	33.49	2.87	0.20	0.46
10	2	13.40	63.16	22.49	0.96	0.12	0.47
11	2	13.88	76.56	9.09	0.48	0.05	0.23
12	2	29.19	68.42	2.39	0.00	0.01	0.04
13	2	8.29	81.60	8.68	1.44	0.06	0.31
14	2	8.61	46.25	20.74	24.39	0.35	0.67
15	2	4.39	37.96	21.69	35.97	0.47	0.69
16	2	3.85	39.19	43.41	13.56	0.35	0.61
17	2	6.14	37.89	33.92	22.05	0.39	0.61
18	2	6.40	52.94	26.70	13.96	0.27	0.67
19	2	6.90	39.79	34.64	18.67	0.36	0.61
20	2	6.12	34.94	27.46	31.48	0.45	0.68
21	2	3.61	44.42	46.88	5.09	0.29	0.54
22	2	5.81	43.42	36.20	14.56	0.33	0.60
23	2	3.78	81.32	12.05	2.85	0.09	0.41
24	2	12.92	74.33	10.85	1.90	0.07	0.42

Appendix K: Scree Plots

Figure K.1. Scree Plot—Mathematics Grade 3

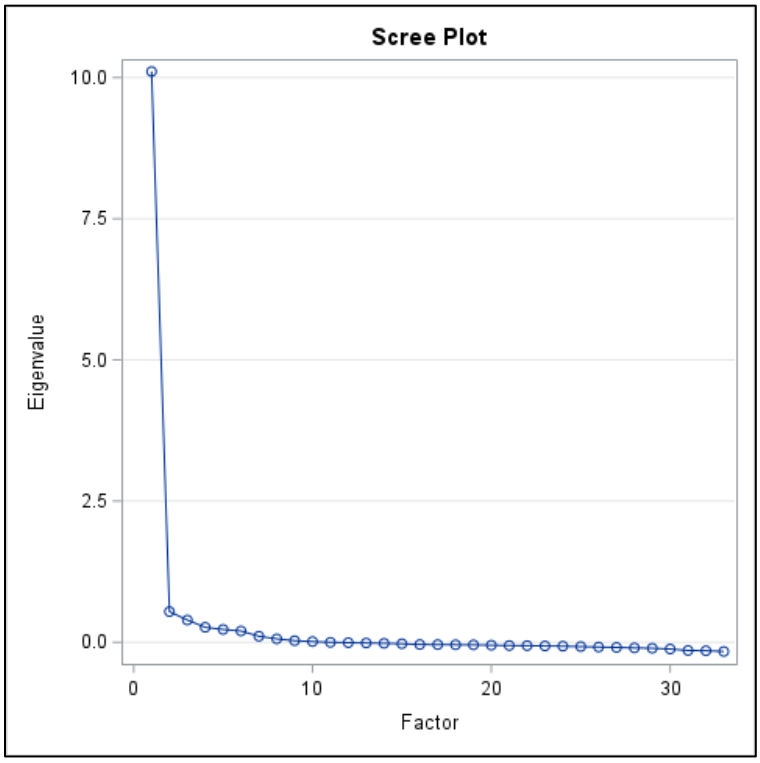


Figure K.2. Scree Plot—Mathematics Grade 4

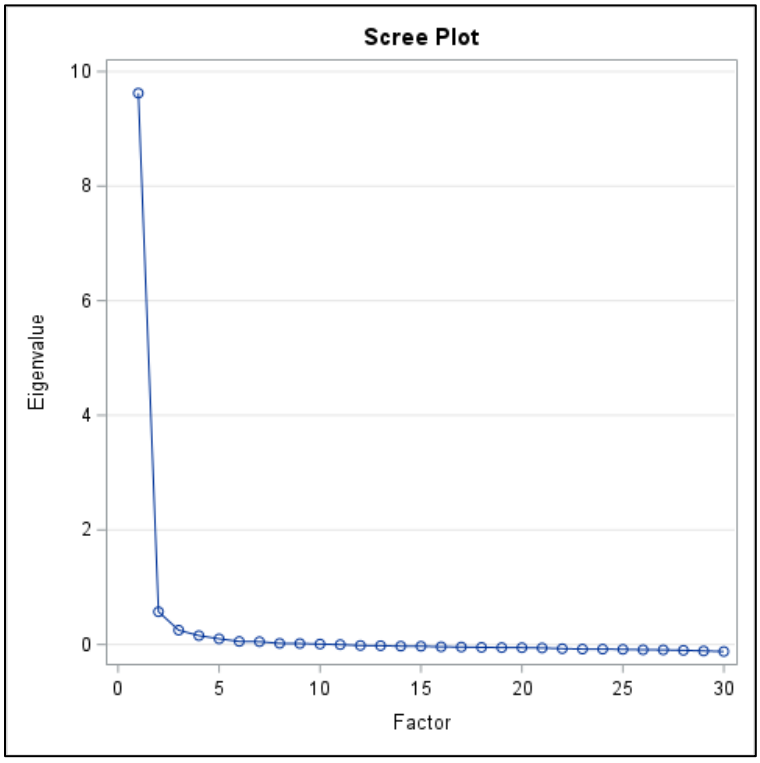


Figure K.3. Scree Plot—Mathematics Grade 5

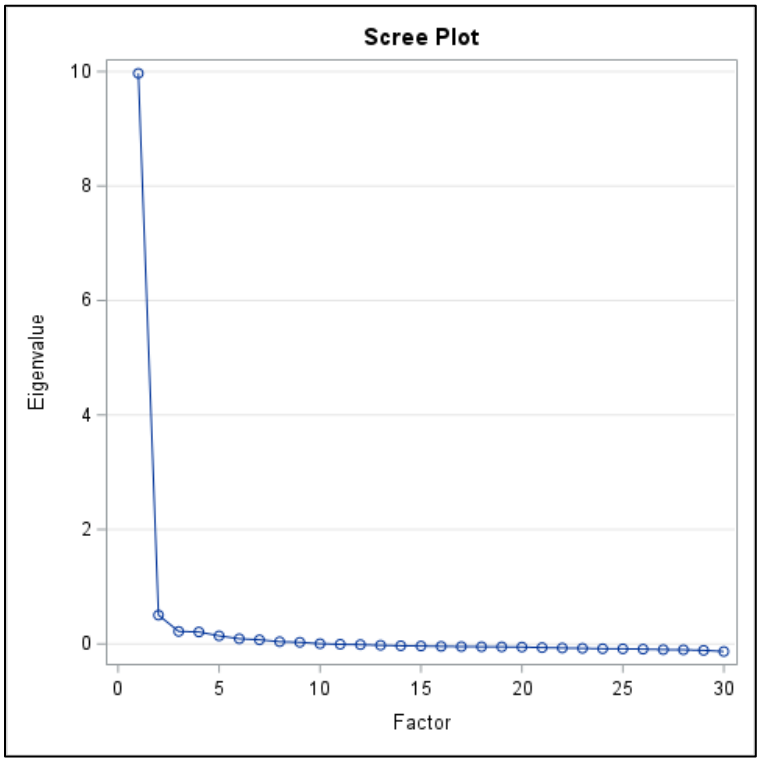


Figure K.4. Scree Plot—Mathematics Grade 6

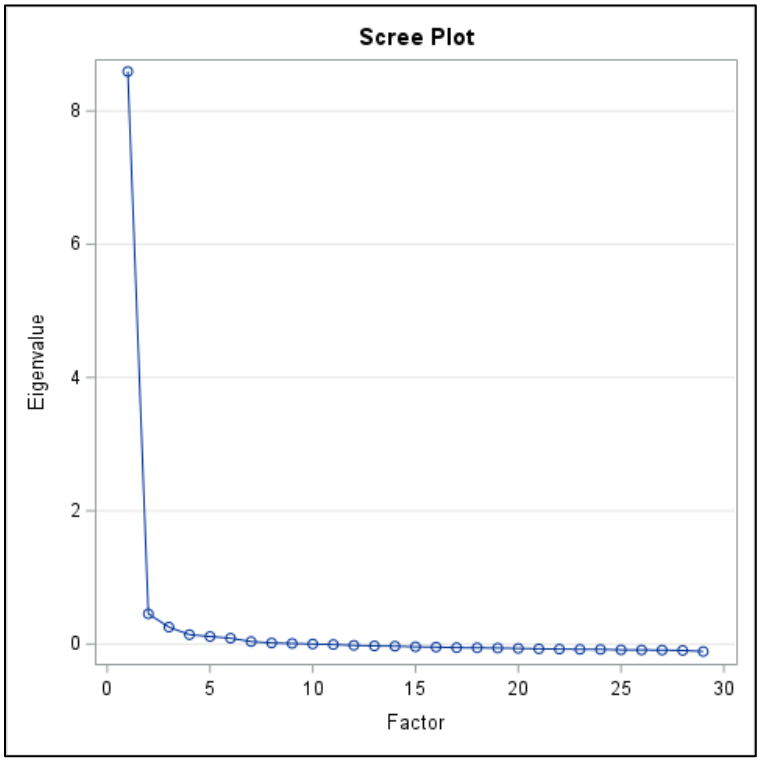


Figure K.5. Scree Plot—Mathematics Grade 7

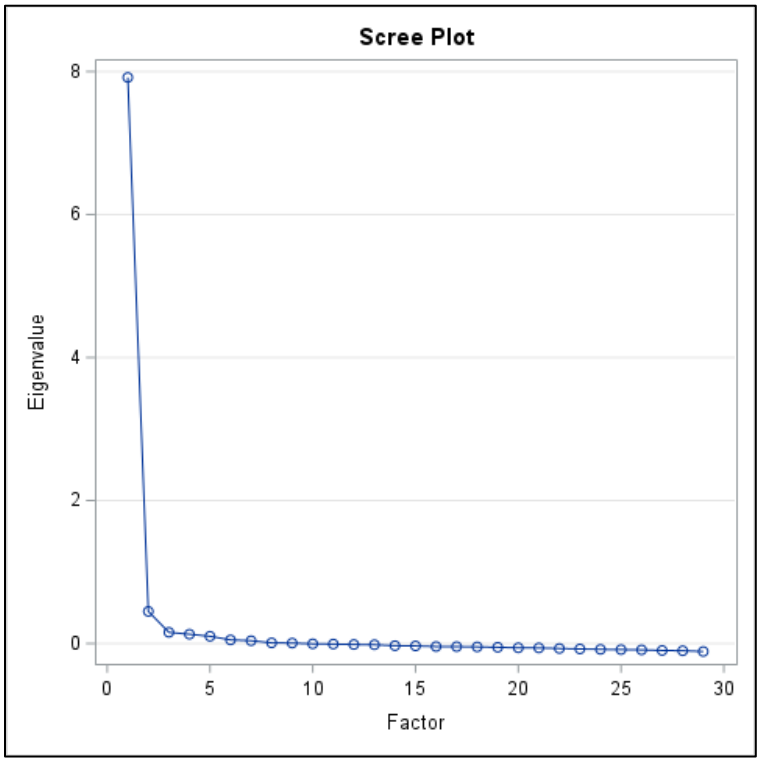


Figure K.6. Scree Plot—Mathematics Grade 8

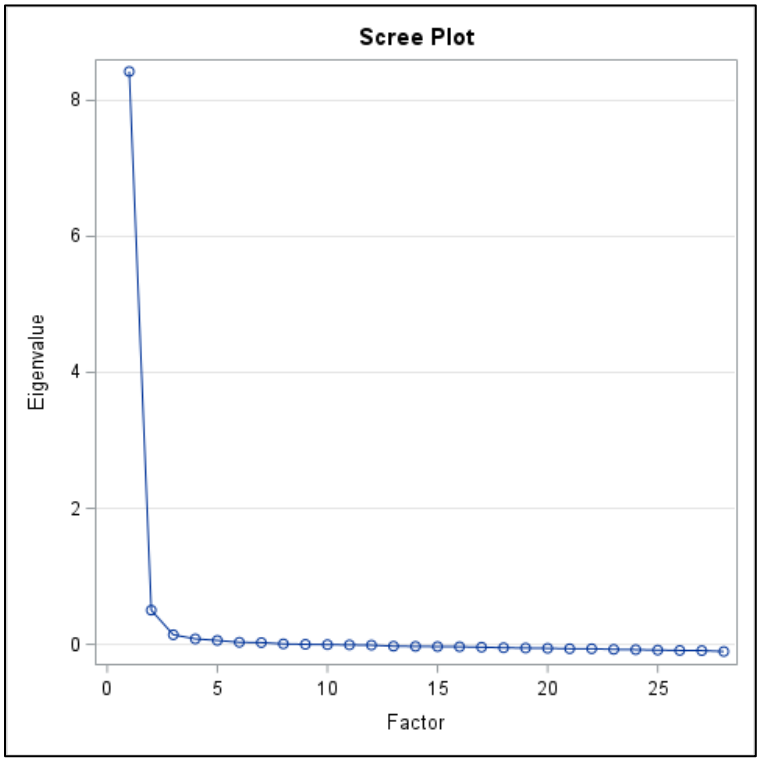


Figure K.7. Scree Plot—ELA Grade 3

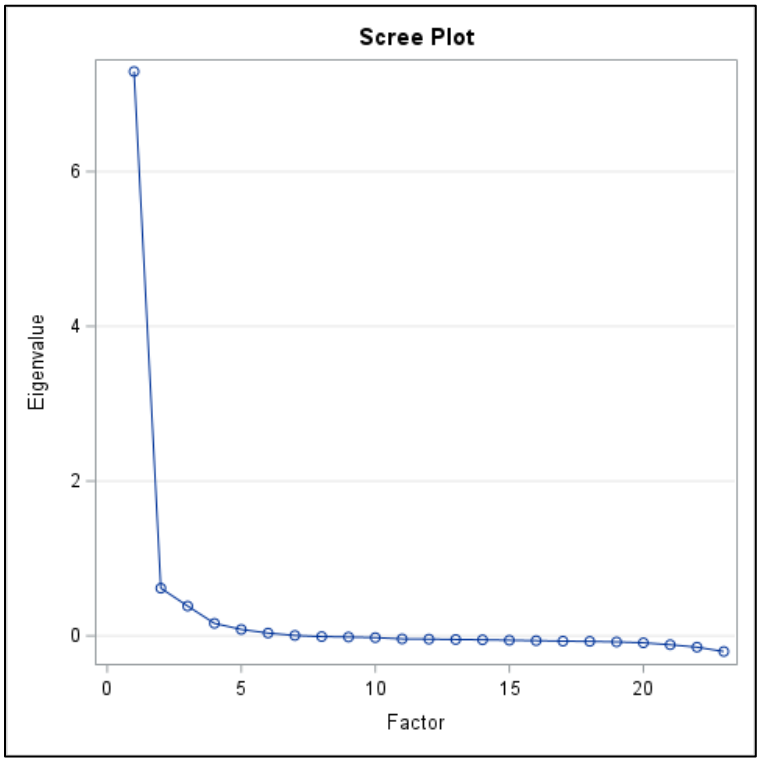


Figure K.8. Scree Plot—ELA Grade 4

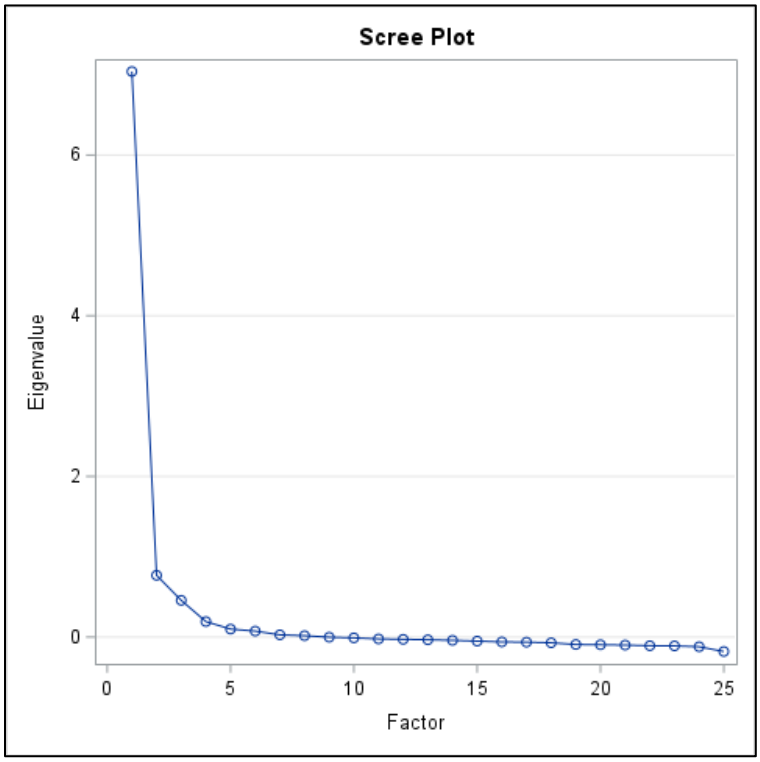


Figure K.9. Scree Plot—ELA Grade 5

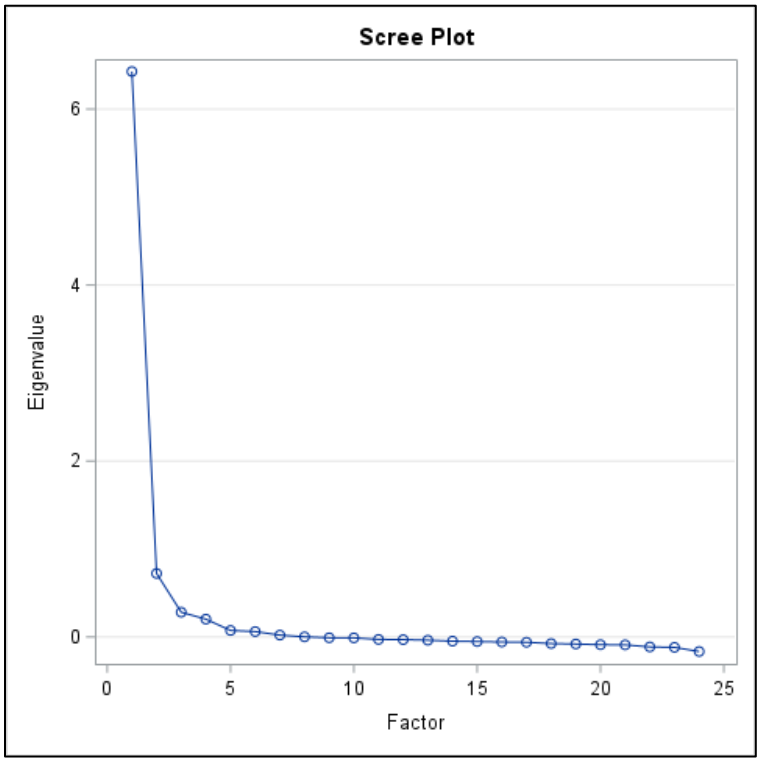


Figure K.10. Scree Plot—ELA Grade 6

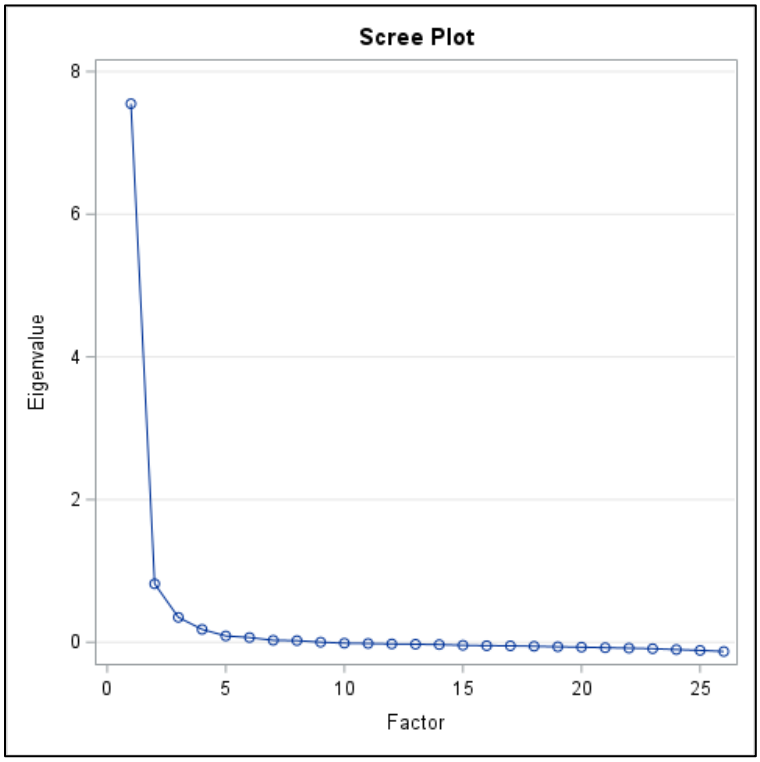


Figure K.11. Scree Plot—ELA Grade 7

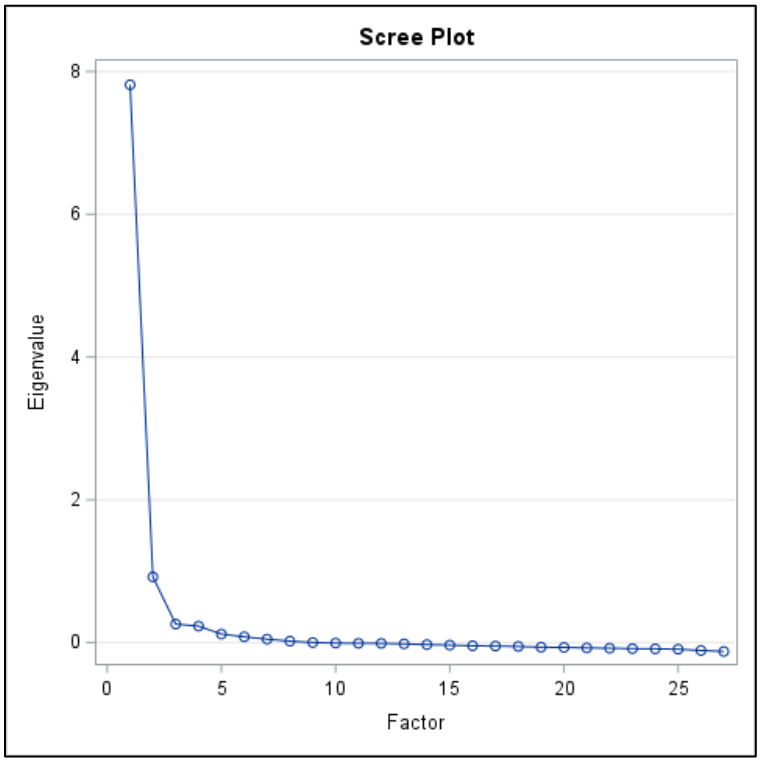


Figure K.12. Scree Plot—ELA Grade 8

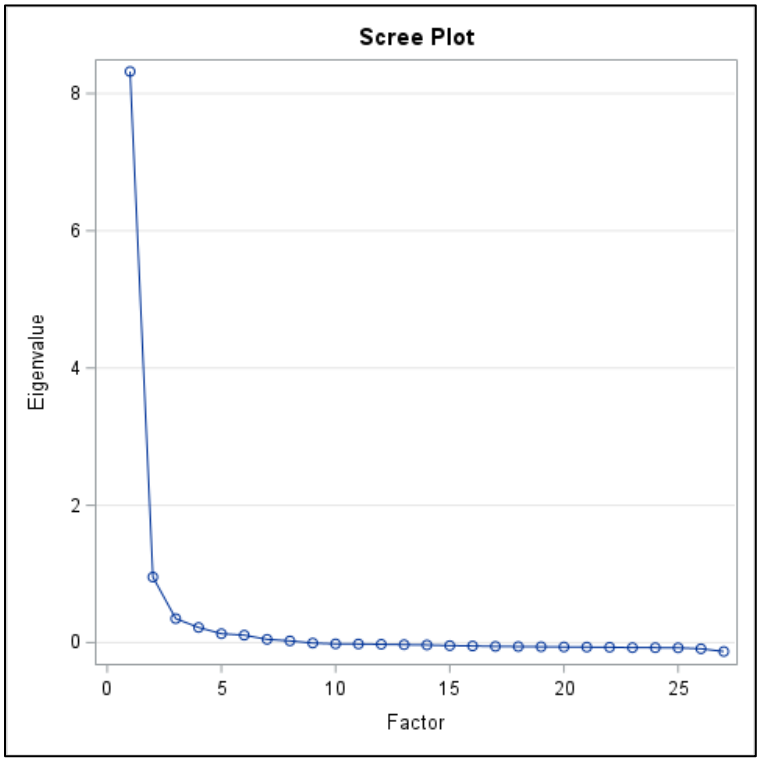


Figure K.13. Scree Plot—CSLA Grade 3

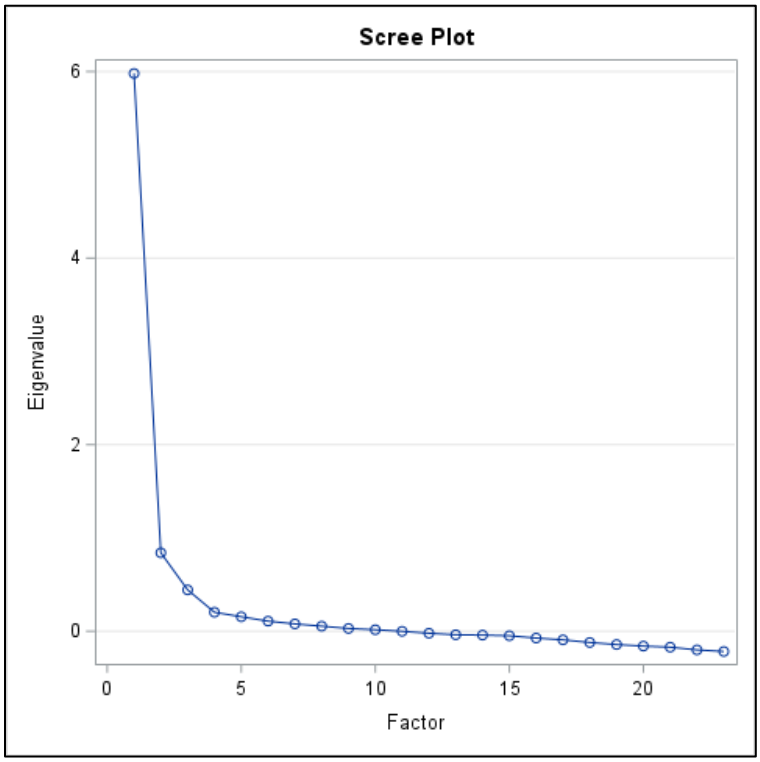


Figure K.14. Scree Plot—CSLA Grade 4

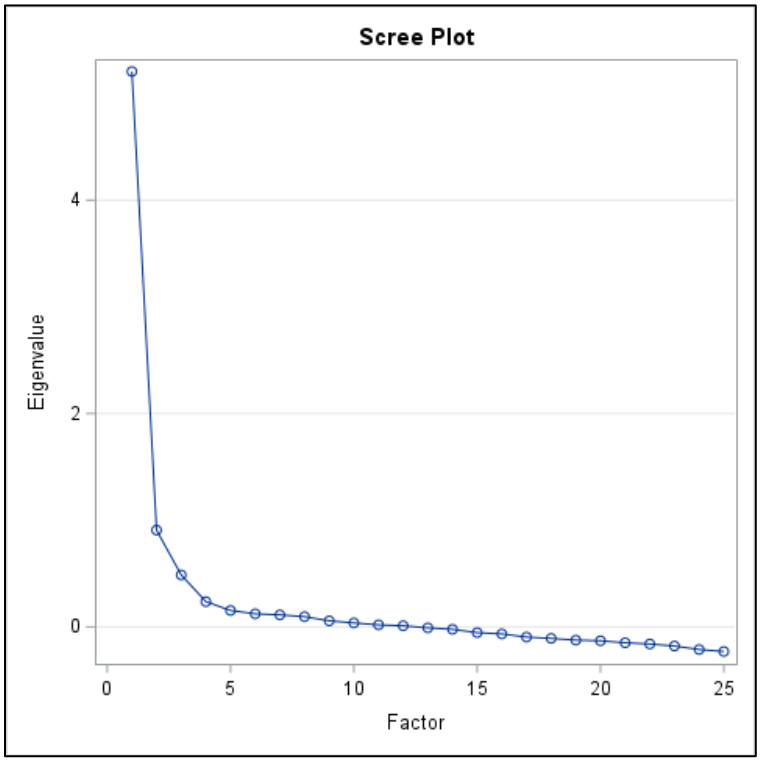


Figure K.15. Scree Plot—Science Grade 5

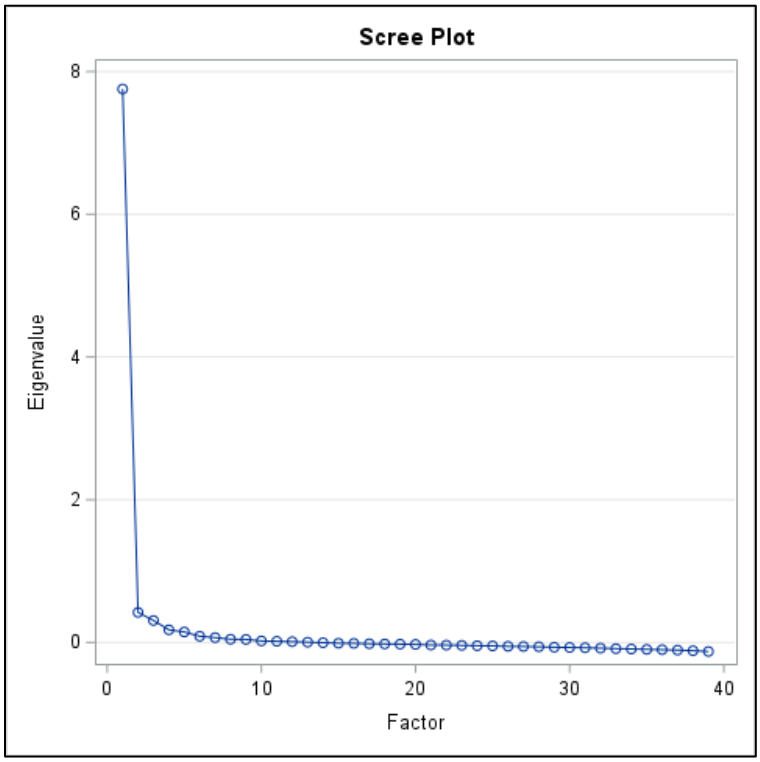


Figure K.16. Scree Plot—Science Grade 8

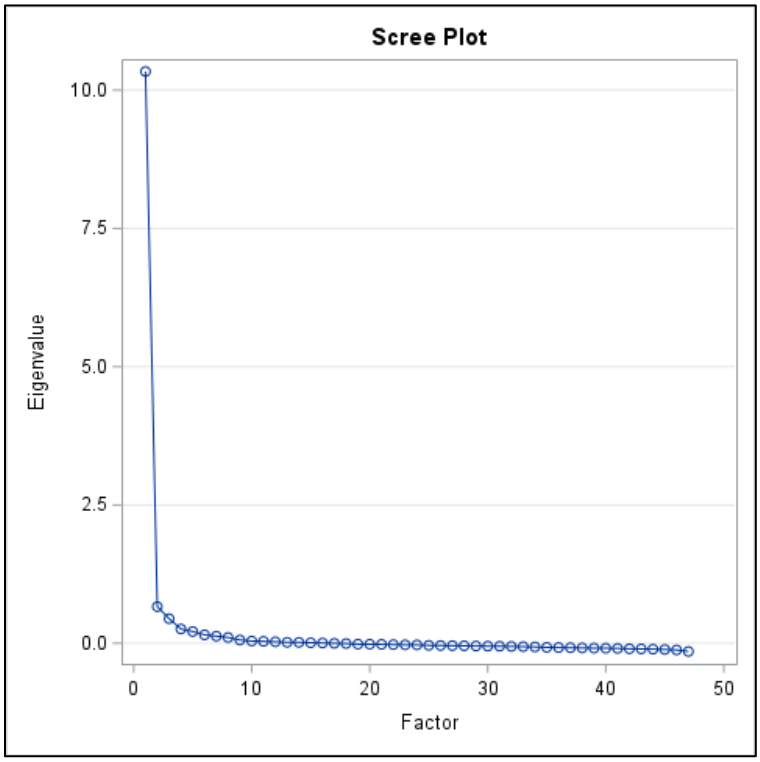
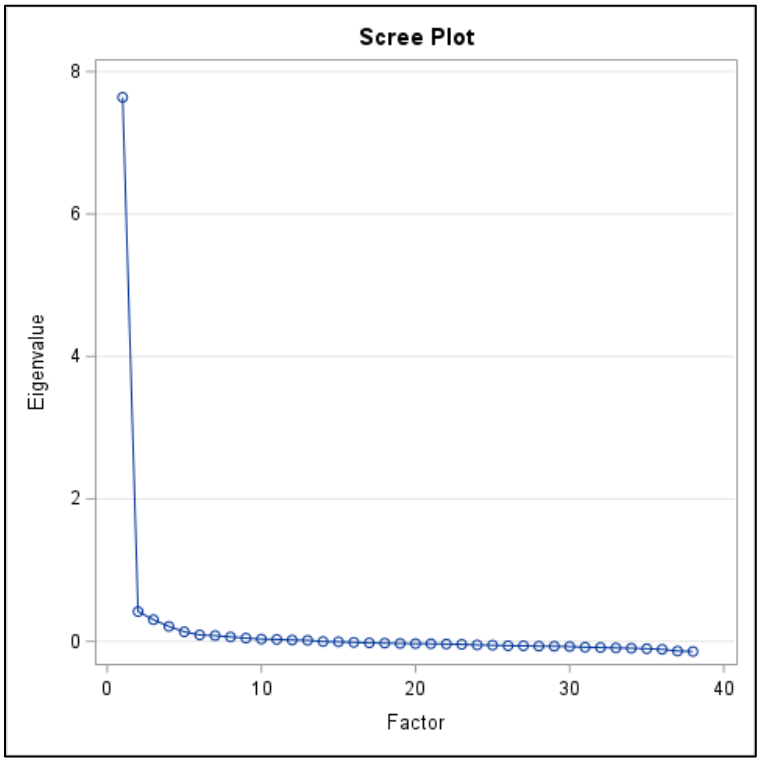


Figure K.17. Scree Plot—Science Grade 11



Appendix L: Mathematics 2022 Post-Equating Check TCCs

Figure L.1. 2022 Post-Equating Check TCCs—Mathematics Grade 3

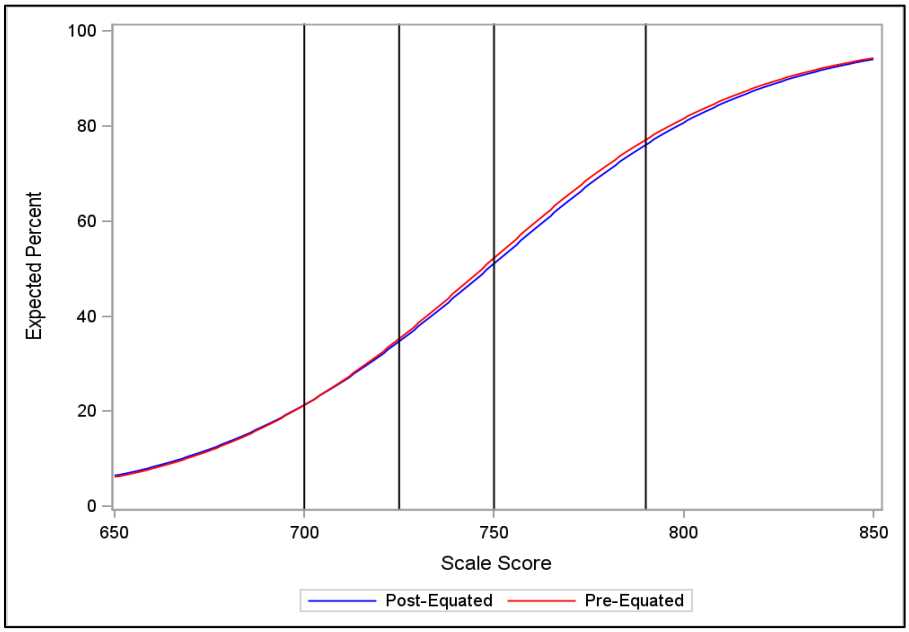


Figure L.2. 2022 Post-Equating Check TCCs—Mathematics Grade 4

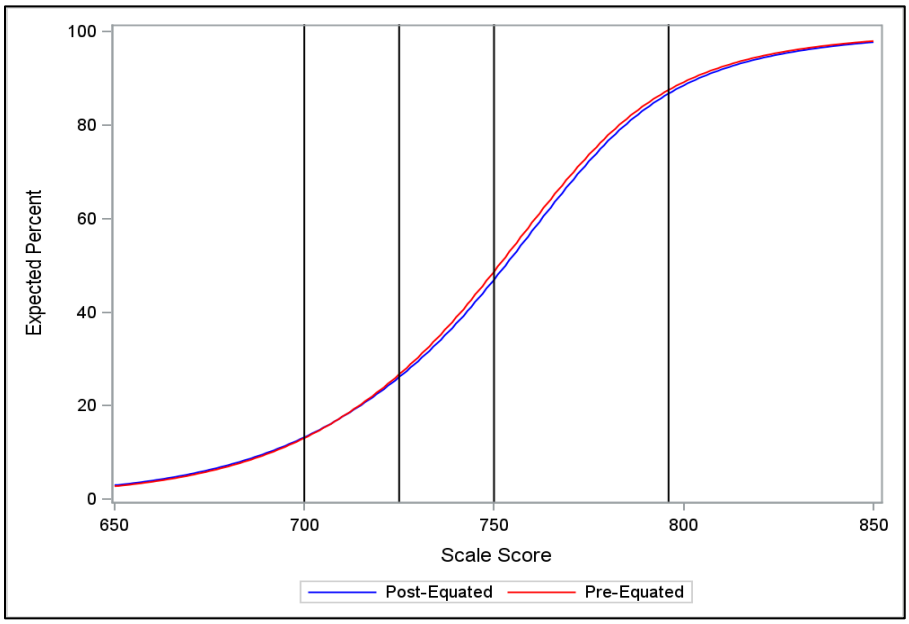


Figure L.3. 2022 Post-Equating Check TCCs—Mathematics Grade 5

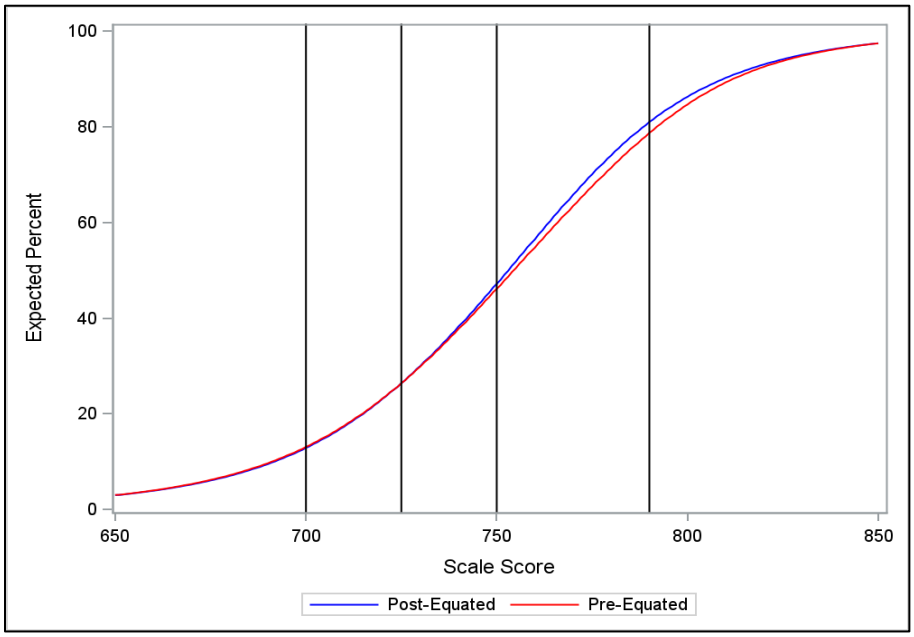


Figure L.4. 2022 Post-Equating Check TCCs—Mathematics Grade 6

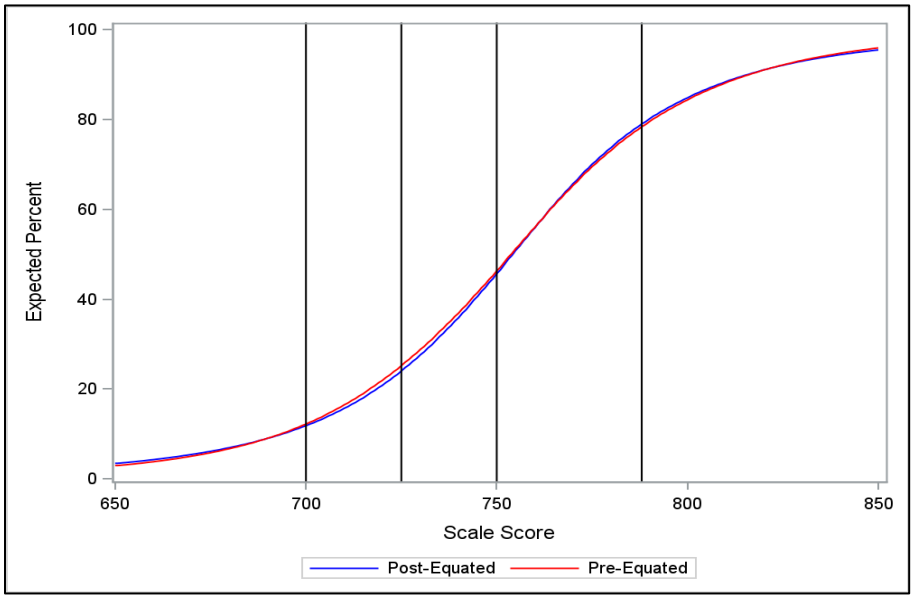


Figure L.5. 2022 Post-Equating Check TCCs—Mathematics Grade 7

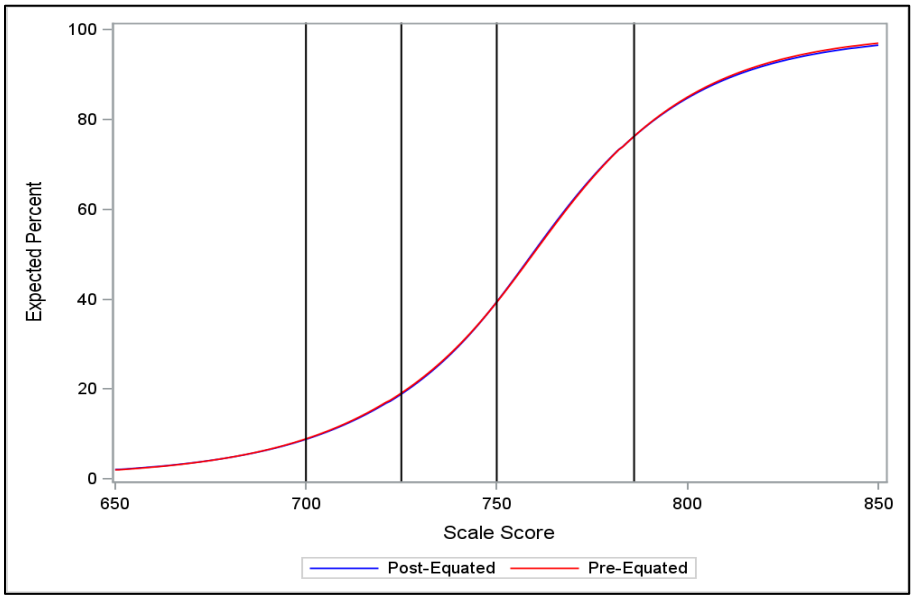
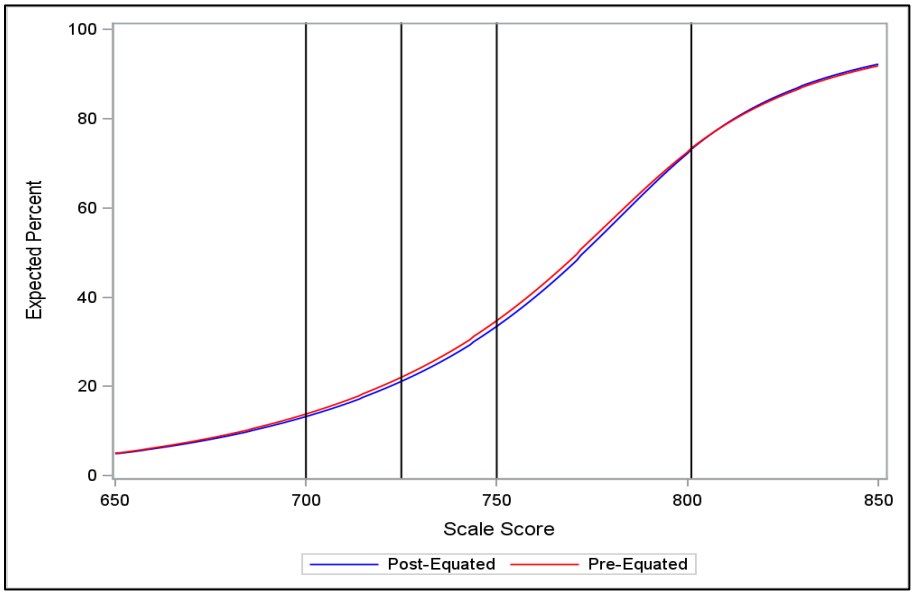


Figure L.6. 2022 Post-Equating Check TCCs—Mathematics Grade 8



Appendix M: IRT Item-Level Statistics

Table M.1. Operational Item Parameter Estimates—ELA Grade 3

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.695	2.497	0	2.378	-0.248	-2.130	–	No
2	CR	GPC	0.783	2.430	0	2.190	-0.173	-2.017	–	No
3	CR	GPC	0.674	2.748	0	2.618	-0.259	-2.360	–	No
4	CR	GPC	0.739	2.601	0	2.360	0.620	-2.980	–	No
5	XI	GPC	0.379	-0.062	0	-0.136	0.136	–	–	Yes
6	XI	GPC	0.439	0.601	0	-2.121	2.121	–	–	Yes
7	XI	GPC	0.479	0.184	0	-0.978	0.978	–	–	Yes
8	XI	GPC	0.722	0.258	0	-0.080	0.080	–	–	Yes
9	XI	GPC	0.893	-1.172	0	0.317	-0.317	–	–	Yes
10	XI	GPC	0.639	-1.478	0	0.413	-0.413	–	–	Yes
11	XI	GPC	0.497	-0.173	0	-0.558	0.558	–	–	No
12	XI	GPC	0.469	-0.307	0	-0.887	0.887	–	–	No
13	XI	GPC	0.355	-0.369	0	-1.668	1.668	–	–	Yes
14	XI	GPC	0.393	-0.389	0	-1.634	1.634	–	–	Yes
15	XI	GPC	0.403	0.294	0	-1.554	1.554	–	–	Yes
16	XI	GPC	0.513	0.201	0	0.952	-0.952	–	–	Yes
17	XI	GPC	0.715	0.061	0	1.264	-1.264	–	–	No
18	XI	GPC	0.526	1.345	0	0.998	-0.998	–	–	No
19	XI	GPC	0.453	-0.281	0	-0.913	0.913	–	–	Yes
20	XI	GPC	0.384	-0.468	0	-1.203	1.203	–	–	Yes
21	XI	GPC	0.712	-0.135	0	0.051	-0.051	–	–	No
22	XI	GPC	0.573	-0.144	0	0.407	-0.407	–	–	No
23	XI	GPC	0.368	1.714	0	1.034	-1.034	–	–	Yes

Table M.2. Operational Item Parameter Estimates—ELA Grade 4

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.748	1.272	0	1.393	-0.187	-1.207	–	No
2	CR	GPC	0.752	0.767	0	1.704	-0.095	-1.609	–	Yes
3	CR	GPC	0.757	1.669	0	1.692	-0.419	-1.273	–	No
4	CR	GPC	0.806	2.004	0	2.531	0.013	-1.058	-1.486	Yes
5	XI	GPC	0.216	0.989	0	-3.277	3.277	–	–	Yes
6	XI	GPC	0.431	0.030	0	-1.595	1.595	–	–	Yes
7	XI	GPC	0.313	0.631	0	0.456	-0.456	–	–	No
8	XI	GPC	0.768	-0.561	0	0.295	-0.295	–	–	No
9	XI	GPC	0.396	0.488	0	-1.779	1.779	–	–	Yes
10	XI	GPC	0.290	0.657	0	-1.815	1.815	–	–	Yes
11	XI	GPC	0.451	1.298	0	0.064	-0.064	–	–	No
12	XI	GPC	0.513	1.007	0	0.103	-0.103	–	–	Yes
13	XI	GPC	0.209	0.820	0	-2.357	2.357	–	–	Yes
14	XI	GPC	0.232	0.485	0	-0.262	0.262	–	–	Yes
15	XI	GPC	0.341	0.623	0	-1.253	1.253	–	–	Yes
16	XI	GPC	0.381	0.248	0	0.441	-0.441	–	–	No
17	XI	GPC	0.605	-1.183	0	0.355	-0.355	–	–	Yes
18	XI	GPC	0.534	0.270	0	-2.752	2.752	–	–	Yes
19	XI	GPC	0.297	1.390	0	-0.315	0.315	–	–	Yes
20	XI	GPC	0.534	0.146	0	-1.522	1.522	–	–	Yes
21	XI	GPC	0.580	-0.491	0	0.200	-0.200	–	–	No
22	XI	GPC	0.332	-0.752	0	-0.310	0.310	–	–	Yes
23	XI	GPC	0.355	-1.194	0	-3.161	3.161	–	–	Yes
24	XI	GPC	0.802	0.024	0	0.043	-0.043	–	–	Yes
25	XI	GPC	0.905	-0.752	0	0.711	-0.711	–	–	No

Table M.3. Operational Item Parameter Estimates—ELA Grade 5

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.814	1.297	0	1.316	0.033	-1.348	–	Yes
2	CR	GPC	0.792	0.785	0	1.429	0.289	-1.718	–	Yes
3	CR	GPC	0.745	1.871	0	1.553	-0.340	-1.213	–	Yes
4	CR	GPC	0.775	2.276	0	1.942	0.198	-0.821	-1.319	Yes
5	XI	GPC	0.576	-0.873	0	-1.434	1.434	–	–	Yes
6	XI	GPC	0.531	-0.130	0	-0.454	0.454	–	–	Yes
7	XI	GPC	0.438	0.352	0	-0.854	0.854	–	–	Yes
8	XI	GPC	0.436	0.335	0	-0.053	0.053	–	–	No
9	XI	GPC	0.350	1.545	0	-1.421	1.421	–	–	Yes
10	XI	GPC	0.890	-0.227	0	0.013	-0.013	–	–	Yes
11	XI	GPC	0.707	-0.018	0	-0.528	0.528	–	–	Yes
12	XI	GPC	0.388	0.111	0	-0.738	0.738	–	–	Yes
13	XI	GPC	0.519	0.607	0	1.130	-1.130	–	–	No
14	XI	GPC	0.274	1.653	0	0.287	-0.287	–	–	Yes
15	XI	GPC	0.384	0.333	0	-0.197	0.197	–	–	Yes
16	XI	GPC	0.330	1.324	0	-0.947	0.947	–	–	Yes
17	XI	GPC	0.237	1.774	0	1.358	-1.358	–	–	Yes
18	XI	GPC	0.181	0.040	0	-2.715	2.715	–	–	No
19	XI	GPC	0.429	0.239	0	2.077	-2.077	–	–	Yes
20	XI	GPC	0.256	2.306	0	-1.771	1.771	–	–	No
21	XI	GPC	0.630	0.253	0	1.003	-1.003	–	–	Yes
22	XI	GPC	0.378	0.404	0	0.844	-0.844	–	–	No
23	XI	GPC	0.349	1.037	0	-0.168	0.168	–	–	Yes
24	XI	GPC	0.471	0.628	0	1.094	-1.094	–	–	No

Table M.4. Operational Item Parameter Estimates—ELA Grade 6

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.833	1.274	0	1.010	0.015	-1.025	–	Yes
2	CR	GPC	0.829	1.709	0	1.689	0.409	-0.788	-1.310	Yes
3	CR	GPC	0.825	1.230	0	1.406	-0.119	-1.288	–	Yes
4	CR	GPC	0.817	1.967	0	2.274	0.627	-0.602	-2.299	Yes
5	XI	GPC	0.303	-0.408	0	-2.483	2.483	–	–	Yes
6	XI	GPC	0.642	0.516	0	0.117	-0.117	–	–	No
7	XI	GPC	0.283	1.000	0	-1.840	1.840	–	–	Yes
8	XI	GPC	0.391	0.697	0	-0.907	0.907	–	–	Yes
9	XI	GPC	0.553	-0.716	0	0.495	-0.495	–	–	Yes
10	XI	GPC	0.390	-0.331	0	1.127	-1.127	–	–	Yes
11	XI	GPC	0.449	0.565	0	-1.128	1.128	–	–	No
12	XI	GPC	0.631	0.236	0	0.196	-0.196	–	–	Yes
13	XI	GPC	0.433	0.307	0	-1.274	1.274	–	–	No
14	XI	GPC	0.594	0.226	0	-0.754	0.754	–	–	Yes
15	XI	GPC	0.404	0.825	0	0.660	-0.660	–	–	No
16	XI	GPC	0.257	1.241	0	-1.306	1.306	–	–	Yes
17	XI	GPC	0.262	0.681	0	-0.728	0.728	–	–	Yes
18	XI	GPC	0.283	0.883	0	1.309	-1.309	–	–	Yes
19	XI	GPC	0.439	-0.052	0	-1.130	1.130	–	–	Yes
20	XI	GPC	0.493	0.637	0	0.323	-0.323	–	–	No
21	XI	GPC	0.479	0.120	0	-0.956	0.956	–	–	No
22	XI	GPC	0.446	-0.136	0	-1.137	1.137	–	–	Yes
23	XI	GPC	0.375	0.969	0	-0.277	0.277	–	–	Yes
24	XI	GPC	0.344	1.717	0	1.549	-1.549	–	–	No
25	XI	GPC	0.480	0.031	0	1.226	-1.226	–	–	No
26	XI	GPC	0.450	-0.491	0	-1.161	1.161	–	–	No

Table M.5. Operational Item Parameter Estimates—ELA Grade 7

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.752	0.615	0	0.674	0.146	-0.820	–	Yes
2	CR	GPC	0.704	1.055	0	1.406	0.657	-0.492	-1.571	Yes
3	CR	GPC	0.878	0.924	0	1.139	0.012	-1.151	–	Yes
4	CR	GPC	0.885	1.236	0	1.934	0.582	-0.808	-1.708	Yes
5	XI	GPC	0.303	0.678	0	-1.923	1.923	–	–	Yes
6	XI	GPC	0.612	0.765	0	0.621	-0.621	–	–	Yes
7	XI	GPC	0.413	1.393	0	-1.971	1.971	–	–	Yes
8	XI	GPC	0.502	0.091	0	-0.875	0.875	–	–	Yes
9	XI	GPC	0.255	1.548	0	-1.398	1.398	–	–	Yes
10	XI	GPC	0.468	2.620	0	1.173	-1.173	–	–	No
11	XI	GPC	0.425	0.698	0	0.249	-0.249	–	–	Yes
12	XI	GPC	0.431	0.764	0	-1.296	1.296	–	–	Yes
13	XI	GPC	0.626	1.284	0	0.590	-0.590	–	–	No
14	XI	GPC	0.560	0.438	0	0.351	-0.351	–	–	No
15	XI	GPC	0.370	1.535	0	-0.334	0.334	–	–	No
16	XI	GPC	0.291	-0.151	0	-1.953	1.953	–	–	No
17	XI	GPC	0.383	0.594	0	0.614	-0.614	–	–	No
18	XI	GPC	0.573	-0.274	0	-0.464	0.464	–	–	Yes
19	XI	GPC	0.351	0.805	0	-3.614	3.614	–	–	Yes
20	XI	GPC	0.220	1.608	0	-2.742	2.742	–	–	Yes
21	XI	GPC	0.519	0.843	0	0.625	-0.625	–	–	No
22	XI	GPC	0.438	0.061	0	-0.744	0.744	–	–	Yes
23	XI	GPC	0.561	0.066	0	-0.441	0.441	–	–	Yes
24	XI	GPC	0.451	0.611	0	2.100	-2.100	–	–	Yes
25	XI	GPC	0.708	-0.147	0	0.402	-0.402	–	–	Yes
26	XI	GPC	0.758	-0.301	0	0.683	-0.683	–	–	Yes
27	XI	GPC	0.446	0.508	0	-0.086	0.086	–	–	Yes

Table M.6. Operational Item Parameter Estimates—ELA Grade 8

Item	Item Type	Model	A	B	D1	D2	D3	D4	D5	Misfit Flag
1	CR	GPC	0.756	0.467	0	0.811	0.201	-1.012	–	No
2	CR	GPC	0.741	1.101	0	1.495	0.822	-0.498	-1.819	Yes
3	CR	GPC	0.765	0.383	0	0.784	0.337	-1.121	–	No
4	CR	GPC	0.802	0.901	0	1.434	0.916	-0.597	-1.753	Yes
5	XI	GPC	0.542	-0.619	0	-1.311	1.311	–	–	Yes
6	XI	GPC	0.442	-0.057	0	-0.333	0.333	–	–	No
7	XI	GPC	0.208	2.859	0	1.320	-1.320	–	–	Yes
8	XI	GPC	0.583	-0.488	0	-0.561	0.561	–	–	No
9	XI	GPC	0.368	-1.329	0	-0.593	0.593	–	–	No
10	XI	GPC	0.499	-0.383	0	0.941	-0.941	–	–	Yes
11	XI	GPC	0.284	1.612	0	-0.605	0.605	–	–	Yes
12	XI	GPC	0.395	-0.954	0	0.929	-0.929	–	–	Yes
13	XI	GPC	0.558	0.091	0	0.077	-0.077	–	–	Yes
14	XI	GPC	0.395	0.857	0	-1.831	1.831	–	–	Yes
15	XI	GPC	0.391	0.885	0	0.467	-0.467	–	–	No
16	XI	GPC	0.897	-0.239	0	0.818	-0.818	–	–	No
17	XI	GPC	0.487	-0.056	0	-0.288	0.288	–	–	No
18	XI	GPC	0.415	0.263	0	0.446	-0.446	–	–	Yes
19	XI	GPC	0.444	0.461	0	1.424	-1.424	–	–	No
20	XI	GPC	0.364	0.683	0	-2.005	2.005	–	–	Yes
21	XI	GPC	0.325	0.288	0	-3.241	3.241	–	–	No
22	XI	GPC	0.509	-0.322	0	-1.068	1.068	–	–	Yes
23	XI	GPC	0.469	0.034	0	-1.023	1.023	–	–	No
24	XI	GPC	0.443	0.643	0	0.852	-0.852	–	–	No
25	XI	GPC	0.427	-0.414	0	-1.826	1.826	–	–	Yes
26	XI	GPC	0.370	0.182	0	0.838	-0.838	–	–	Yes
27	XI	GPC	0.369	1.420	0	-0.545	0.545	–	–	No

Table M.7. Operational Item Parameter Estimates—CSLA Grade 3

Item	Item Type	Model	B	D1	D2	D3	D4	Infit	Outfit
1	CR	Rasch	0.054	0	-0.904	0.694	0.210	0.88	0.85
2	CR	Rasch	0.377	0	-0.937	0.115	0.821	0.76	0.71
3	CR	Rasch	0.182	0	-0.880	0.688	0.192	1.02	0.97
4	CR	Rasch	0.480	0	-0.345	-0.300	0.645	0.93	0.87
5	XI	Rasch	-0.192	0	-1.290	1.290	–	0.80	0.80
6	XI	Rasch	-0.290	0	0.577	-0.577	–	1.07	1.23
7	XI	Rasch	-0.755	0	-0.773	0.773	–	1.04	1.06
8	XI	Rasch	-0.653	0	0.156	-0.156	–	0.96	0.95
9	XI	Rasch	-0.287	0	0.339	-0.339	–	1.03	1.07
10	XI	Rasch	-0.295	0	0.767	-0.767	–	0.92	0.92
11	XI	Rasch	0.119	0	0.204	-0.204	–	1.14	1.22
12	XI	Rasch	-0.081	0	1.273	-1.273	–	1.09	1.17
13	XI	Rasch	0.272	0	0.121	-0.121	–	0.99	1.01
14	XI	Rasch	0.891	0	-0.171	0.171	–	1.26	1.64
15	XI	Rasch	-0.706	0	0.436	-0.436	–	0.85	0.81
16	XI	Rasch	-0.105	0	0.276	-0.276	–	0.92	0.89
17	XI	Rasch	0.450	0	-0.180	0.180	–	1.30	1.53
18	XI	Rasch	-0.329	0	0.788	-0.788	–	0.90	0.93
19	XI	Rasch	-0.712	0	-0.090	0.090	–	0.90	0.88
20	XI	Rasch	0.157	0	-0.022	0.022	–	1.13	1.18
21	XI	Rasch	-0.062	0	0.539	-0.539	–	1.30	1.48

Table M.8. Operational Item Parameter Estimates—CSLA Grade 4

Item	Item Type	Model	B	D1	D2	D3	D4	D5	Infit	Outfit
1	CR	Rasch	-0.605	0	-0.812	0.608	0.204	–	1.03	1.01
2	CR	Rasch	0.069	0	0.133	-0.881	0.748	–	0.94	0.94
3	CR	Rasch	-0.227	0	-0.754	0.613	0.142	–	1.01	1.00
4	CR	Rasch	0.302	0	-0.722	-0.349	0.012	1.059	0.83	0.77
5	XI	Rasch	-1.251	0	-0.197	0.197	–	–	0.89	0.86
6	XI	Rasch	-0.496	0	-0.650	0.650	–	–	1.12	1.13
7	XI	Rasch	-0.162	0	0.908	-0.908	–	–	1.05	1.07
8	XI	Rasch	-1.167	0	0.668	-0.668	–	–	1.07	1.10
9	XI	Rasch	-1.463	0	-0.205	0.205	–	–	0.86	0.85
10	XI	Rasch	0.792	0	-0.786	0.786	–	–	0.89	0.87
11	XI	Rasch	-0.991	0	0.699	-0.699	–	–	0.88	0.89
12	XI	Rasch	-0.149	0	0.836	-0.836	–	–	0.91	0.84
13	XI	Rasch	-1.225	0	0.582	-0.582	–	–	0.83	0.78
14	XI	Rasch	-0.906	0	0.911	-0.911	–	–	0.84	0.82
15	XI	Rasch	-0.626	0	0.399	-0.399	–	–	1.01	1.05
16	XI	Rasch	0.090	0	0.704	-0.704	–	–	1.02	1.15
17	XI	Rasch	1.102	0	-0.997	0.997	–	–	1.08	1.19
18	XI	Rasch	0.261	0	-0.260	0.260	–	–	1.16	1.30
19	XI	Rasch	-0.250	0	-0.523	0.523	–	–	0.97	0.96
20	XI	Rasch	0.248	0	0.811	-0.811	–	–	1.08	1.51
21	XI	Rasch	-0.006	0	1.434	-1.434	–	–	0.99	1.07
22	XI	Rasch	-0.320	0	0.227	-0.227	–	–	0.95	0.95
23	XI	Rasch	0.446	0	0.435	-0.435	–	–	1.24	1.83
24	XI	Rasch	0.807	0	0.011	-0.011	–	–	1.19	1.80
25	SR	Rasch	-0.197	0	-0.162	0.162	–	–	0.90	0.86

Table M.9. Operational Item Parameter Estimates—Science Grade 5

Item	Item Type	Model	A	B	C	D1	D2	D3	Misfit Flag
1	CR	GPC	0.789	0.752	–	0	0.127	-0.127	Yes
2	CR	GPC	0.443	0.964	–	0	-0.279	0.279	Yes
3	CR	GPC	0.494	2.175	–	0	0.153	-0.153	No
4	CR	GPC	0.504	1.055	–	0	0.432	-0.432	Yes
5	CR	GPC	0.533	-0.291	–	0	-0.837	0.837	Yes
6	CR	GPC	0.674	0.827	–	0	0.826	-0.826	No
7	CR	GPC	0.649	1.378	–	0	0.094	-0.094	No
8	CR	GPC	0.919	0.273	–	0	0.365	-0.365	Yes
9	CR	GPC	0.715	-0.025	–	0	-0.197	0.197	Yes
10	CR	GPC	0.678	0.800	–	0	-0.191	0.191	Yes
11	CR	GPC	0.606	0.328	–	0	-0.417	0.417	Yes
12	CR	GPC	0.850	-0.391	–	0	0.672	-0.672	Yes
13	XI	3PL	1.306	-0.169	0.088	–	–	–	Yes
14	XI	3PL	0.694	0.845	0.144	–	–	–	No
15	XI	3PL	1.137	0.267	0.028	–	–	–	No
16	XI	3PL	0.505	1.733	0.009	–	–	–	No
17	XI	3PL	1.301	1.489	0.128	–	–	–	No
18	XI	3PL	0.508	0.577	0.182	–	–	–	No
19	XI	3PL	0.768	0.104	0.118	–	–	–	Yes
20	XI	3PL	1.969	1.511	0.076	–	–	–	Yes
21	XI	3PL	0.662	0.443	0.013	–	–	–	Yes
22	XI	3PL	1.049	0.914	0.006	–	–	–	Yes
23	XI	3PL	0.420	1.134	0.119	–	–	–	Yes
24	XI	3PL	1.239	0.615	0.023	–	–	–	No
25	XI	3PL	0.942	0.536	0.022	–	–	–	No
26	XI	3PL	1.092	1.003	0.098	–	–	–	Yes
27	SR	3PL	1.101	-0.708	0.264	–	–	–	Yes
28	SR	3PL	0.746	2.025	0.273	–	–	–	No
29	SR	3PL	0.702	1.479	0.360	–	–	–	No
30	SR	3PL	0.991	1.840	0.248	–	–	–	No
31	SR	3PL	1.137	0.694	0.437	–	–	–	No
32	SR	3PL	0.730	1.648	0.179	–	–	–	Yes
33	SR	3PL	0.553	1.964	0.132	–	–	–	No
34	SR	3PL	1.210	0.860	0.394	–	–	–	No
35	SR	3PL	1.156	1.389	0.324	–	–	–	Yes
36	SR	3PL	0.420	-0.124	0.082	–	–	–	No
37	SR	3PL	1.029	0.233	0.242	–	–	–	Yes
38	SR	3PL	0.720	-0.564	0.170	–	–	–	Yes
39	SR	3PL	0.753	0.393	0.081	–	–	–	No

Table M.10. Operational Item Parameter Estimates—Science Grade 8

Item	Item Type	Model	A	B	C	D1	D2	D3	Misfit Flag
1	CR	GPC	1.056	1.077	–	0	0.495	-0.495	Yes
2	CR	GPC	0.655	0.801	–	0	-0.063	0.063	Yes
3	CR	GPC	1.176	1.619	–	0	0.314	-0.314	Yes
4	CR	GPC	1.007	0.831	–	0	0.359	-0.359	Yes
5	CR	GPC	0.575	0.496	–	0	-0.025	0.025	Yes
6	CR	GPC	0.851	1.166	–	0	0.161	-0.161	Yes
7	CR	GPC	1.123	1.087	–	0	0.369	-0.369	Yes
8	CR	GPC	0.344	0.971	–	0	-0.896	0.896	Yes
9	CR	GPC	0.843	0.001	–	0	-0.089	0.089	Yes
10	CR	GPC	0.730	1.348	–	0	0.493	-0.493	Yes
11	CR	GPC	0.928	1.567	–	0	0.390	-0.390	Yes
12	CR	GPC	0.610	1.309	–	0	-0.769	0.769	Yes
13	CR	GPC	0.737	0.190	–	0	0.489	-0.489	Yes
14	CR	GPC	0.761	1.046	–	0	0.622	-0.622	Yes
15	XI	2PL	0.497	-0.098	–	–	–	–	Yes
16	XI	2PL	0.733	-0.776	–	–	–	–	Yes
17	XI	2PL	1.221	0.697	–	–	–	–	Yes
18	XI	2PL	0.548	0.257	–	–	–	–	Yes
19	XI	2PL	1.086	1.390	–	–	–	–	Yes
20	XI	2PL	0.485	2.531	–	–	–	–	Yes
21	XI	2PL	0.732	2.218	–	–	–	–	Yes
22	XI	3PL	0.608	1.203	0.040	–	–	–	No
23	XI	3PL	0.314	4.146	0.042	–	–	–	No
24	XI	3PL	0.605	-0.318	0.098	–	–	–	Yes
25	XI	3PL	0.937	0.904	0.044	–	–	–	Yes
26	XI	3PL	0.743	-0.316	0.058	–	–	–	Yes
27	XI	3PL	1.162	0.889	0.001	–	–	–	Yes
28	XI	3PL	0.775	1.069	0.094	–	–	–	Yes
29	XI	3PL	0.199	7.158	0.045	–	–	–	No
30	XI	3PL	0.830	0.731	0.105	–	–	–	Yes
31	XI	3PL	0.869	-0.146	0.164	–	–	–	No
32	XI	3PL	1.420	2.085	0.030	–	–	–	Yes
33	XI	3PL	1.710	1.667	0.100	–	–	–	Yes
34	XI	3PL	1.253	0.860	0.194	–	–	–	Yes
35	XI	3PL	1.014	-0.273	0.031	–	–	–	Yes
36	XI	3PL	0.774	0.550	0.006	–	–	–	Yes
37	XI	3PL	0.555	1.261	0.016	–	–	–	Yes
38	SR	3PL	1.277	0.669	0.329	–	–	–	Yes
39	SR	3PL	0.943	1.530	0.077	–	–	–	Yes
40	SR	3PL	0.672	1.577	0.428	–	–	–	Yes
41	SR	3PL	1.061	0.309	0.278	–	–	–	Yes
42	SR	3PL	1.246	0.289	0.115	–	–	–	Yes
43	SR	3PL	1.074	0.373	0.099	–	–	–	Yes
44	SR	3PL	1.347	0.462	0.400	–	–	–	Yes
45	SR	3PL	0.463	-1.247	0.019	–	–	–	Yes
46	SR	3PL	1.178	0.898	0.281	–	–	–	Yes
47	SR	3PL	0.981	-0.945	0.019	–	–	–	Yes

Table M.11. Operational Item Parameter Estimates—Science Grade 11

Item	Item Type	Model	A	B	C	D1	D2	D3	Misfit Flag
1	CR	GPC	0.559	2.854	–	0	0.066	-0.066	No
2	CR	GPC	0.765	0.571	–	0	-0.076	0.076	Yes
3	CR	GPC	0.835	0.129	–	0	-0.042	0.042	Yes
4	CR	GPC	0.792	0.691	–	0	0.847	-0.847	Yes
5	CR	GPC	0.680	0.481	–	0	0.438	-0.438	Yes
6	CR	GPC	0.963	0.870	–	0	0.360	-0.360	Yes
7	CR	GPC	0.711	0.613	–	0	0.506	-0.506	Yes
8	CR	GPC	0.824	0.191	–	0	0.189	-0.189	Yes
9	CR	GPC	0.715	1.290	–	0	1.313	-1.313	No
10	CR	GPC	0.714	0.790	–	0	0.628	-0.628	Yes
11	CR	GPC	0.645	2.226	–	0	0.135	-0.135	No
12	CR	GPC	0.785	2.234	–	0	0.281	-0.281	Yes
13	XI	2PL	0.857	1.395	–	–	–	–	No
14	XI	2PL	0.699	1.054	–	–	–	–	Yes
15	XI	3PL	0.835	0.476	0.119	–	–	–	Yes
16	XI	3PL	0.880	0.223	0.019	–	–	–	Yes
17	XI	3PL	0.315	0.846	0.020	–	–	–	No
18	XI	3PL	1.202	-0.062	0.140	–	–	–	Yes
19	XI	3PL	1.389	2.187	0.033	–	–	–	No
20	XI	3PL	1.436	-0.294	0.145	–	–	–	No
21	XI	3PL	0.743	0.666	0.127	–	–	–	No
22	XI	3PL	1.006	1.484	0.044	–	–	–	No
23	XI	3PL	1.027	1.534	0.081	–	–	–	Yes
24	XI	3PL	0.828	0.327	0.058	–	–	–	No
25	SR	3PL	1.220	0.336	0.165	–	–	–	Yes
26	SR	3PL	0.728	0.515	0.023	–	–	–	Yes
27	SR	3PL	1.806	1.741	0.293	–	–	–	No
28	SR	3PL	0.911	0.383	0.184	–	–	–	Yes
29	SR	3PL	0.535	0.940	0.055	–	–	–	Yes
30	SR	3PL	0.845	-0.719	0.390	–	–	–	Yes
31	SR	3PL	1.065	1.449	0.329	–	–	–	No
32	SR	3PL	0.660	0.449	0.183	–	–	–	Yes
33	SR	3PL	0.756	2.055	0.180	–	–	–	Yes
34	SR	3PL	0.496	0.069	0.164	–	–	–	Yes
35	SR	3PL	0.984	0.723	0.221	–	–	–	Yes
36	SR	3PL	0.901	1.188	0.236	–	–	–	No
37	SR	3PL	1.234	2.206	0.363	–	–	–	Yes
38	SR	3PL	0.483	0.00	0.132	–	–	–	Yes

Appendix N: TCC, TIC, and CSEM Curves

Figure N.1. Mathematics Grade 3 TCC

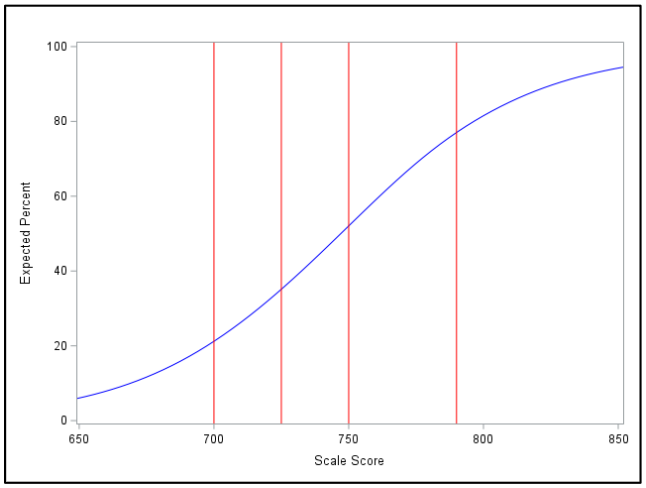


Figure N.2. Mathematics Grade 3 TIC

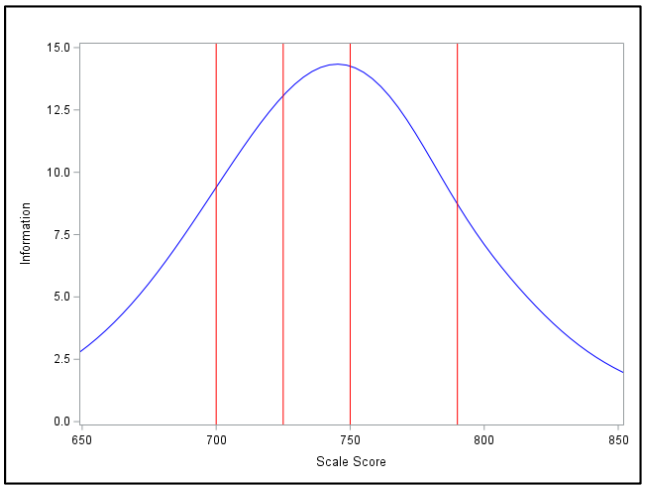


Figure N.3. Mathematics Grade 3 CSEM Curve

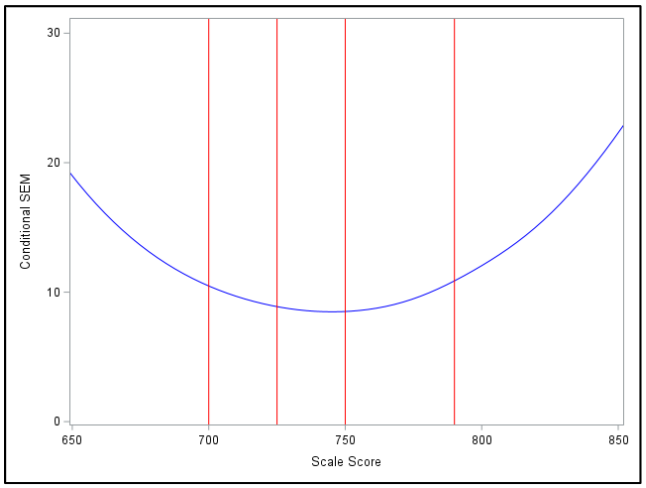


Figure N.4. Mathematics Grade 4 TCC

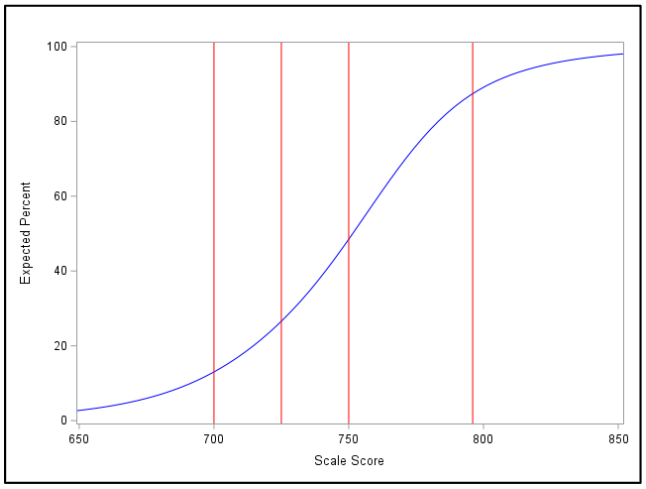


Figure N.5. Mathematics Grade 4 TIC

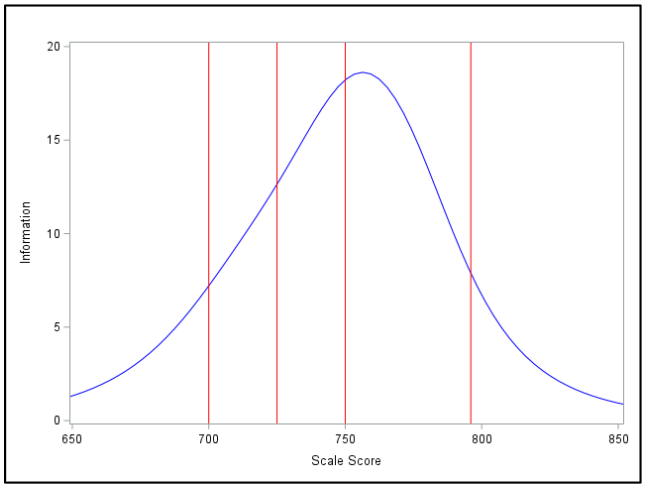


Figure N.6. Mathematics Grade 4 CSEM Curve

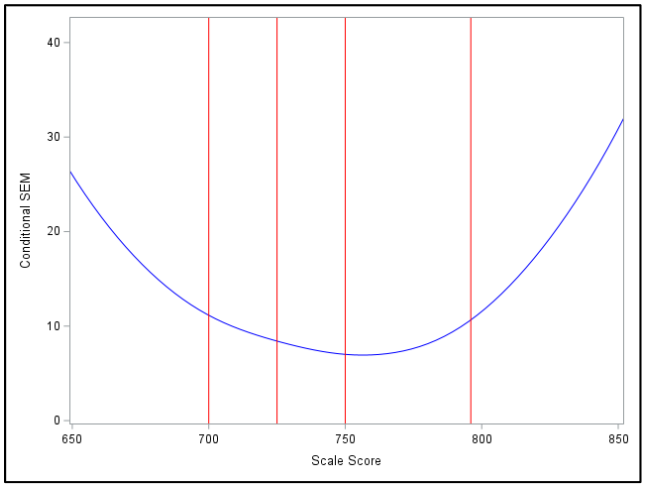


Figure N.7. Mathematics Grade 5 TCC

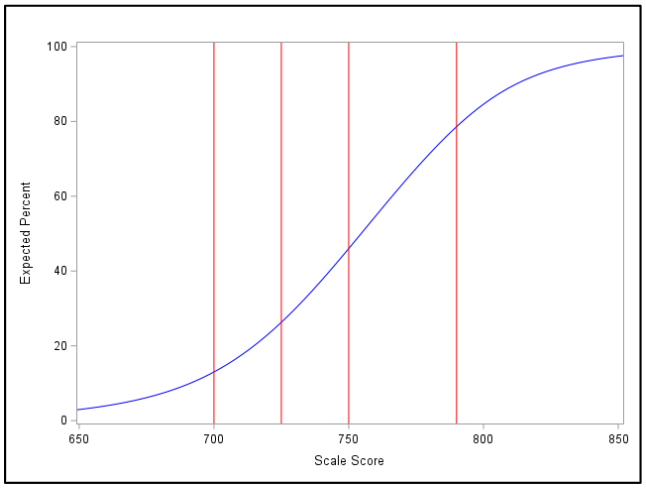


Figure N.8. Mathematics Grade 5 TIC

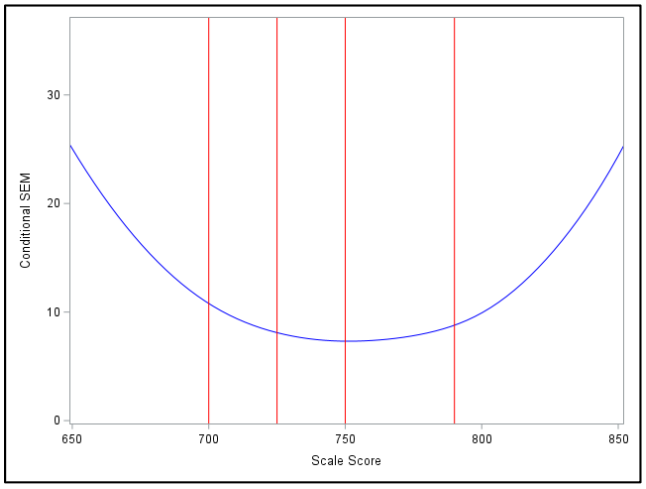


Figure N.9. Mathematics Grade 5 CSEM Curve

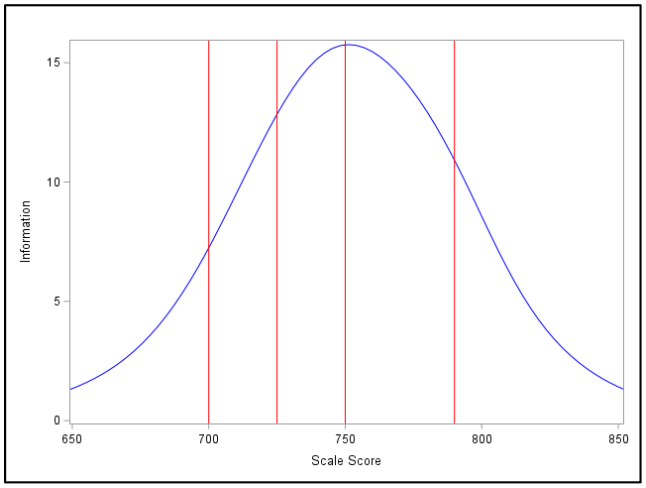


Figure N.10. Mathematics Grade 6 TCC

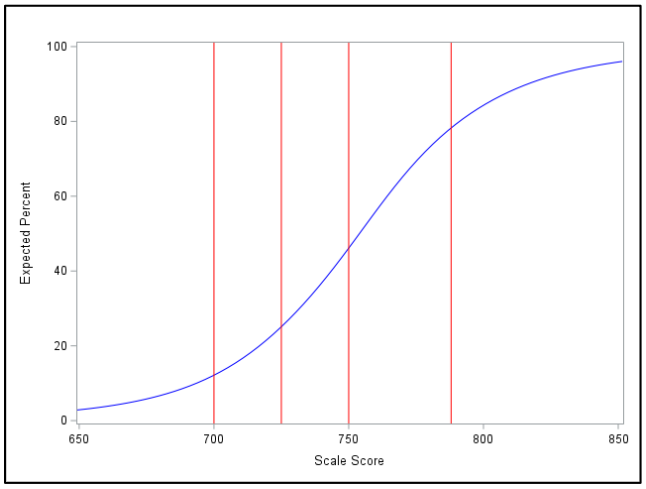


Figure N.11. Mathematics Grade 6 TIC

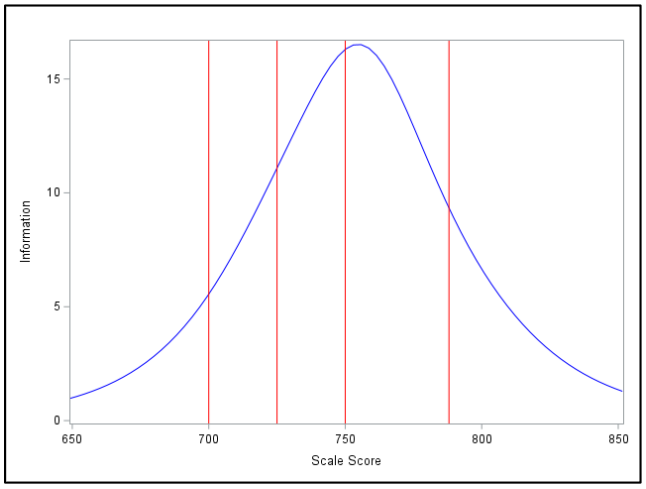


Figure N.12. Mathematics Grade 6 CSEM Curve

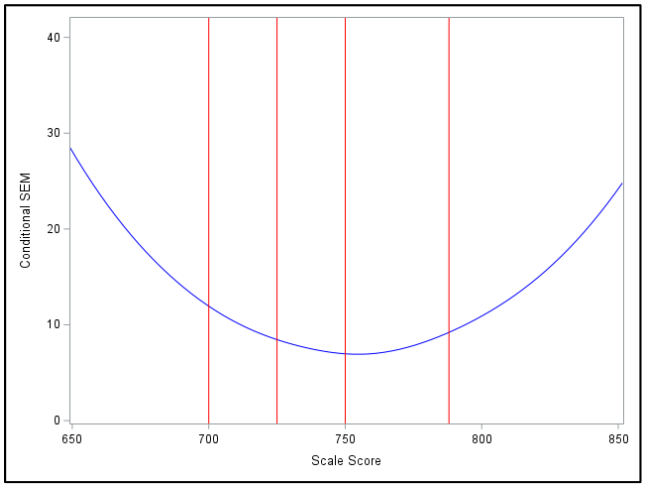


Figure N.13. Mathematics Grade 7 TCC

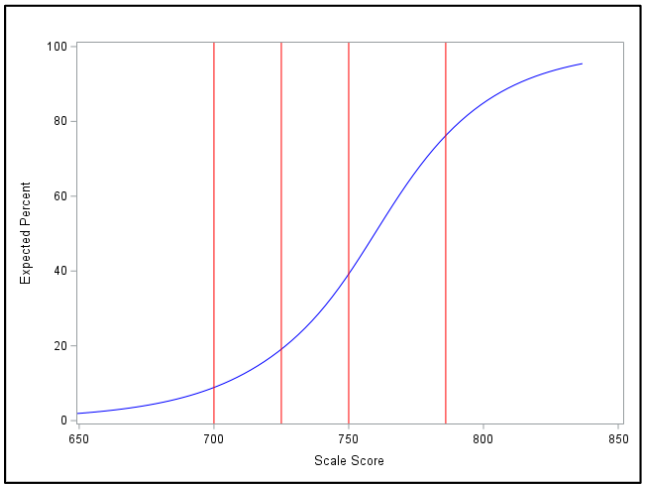


Figure N.14. Mathematics Grade 7 TIC

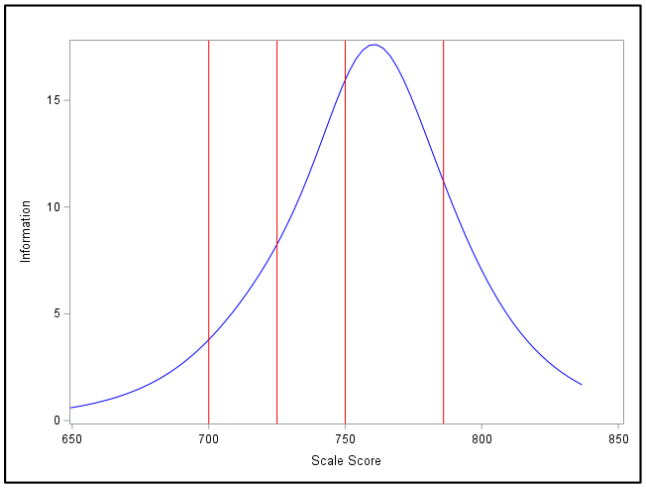


Figure N.15. Mathematics Grade 7 CSEM Curve

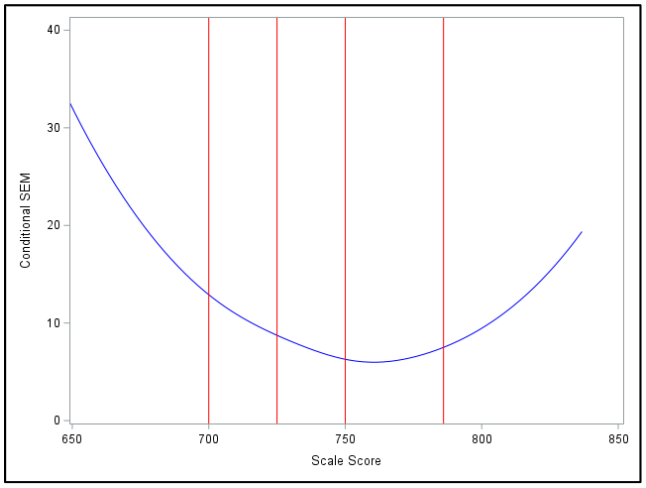


Figure N.16. Mathematics Grade 8 TCC

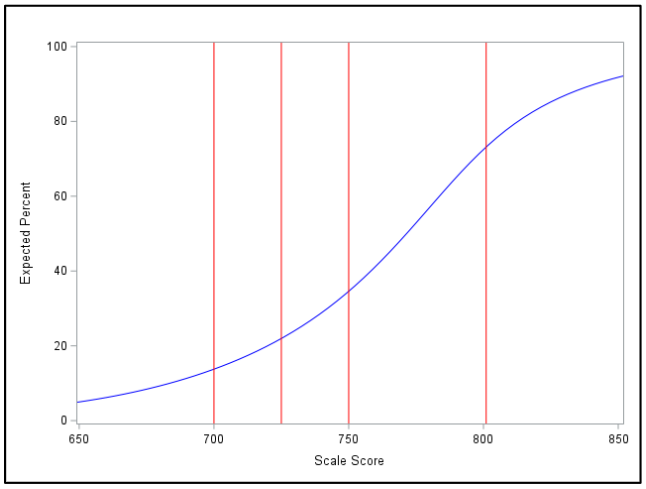


Figure N.17. Mathematics Grade 8 TIC

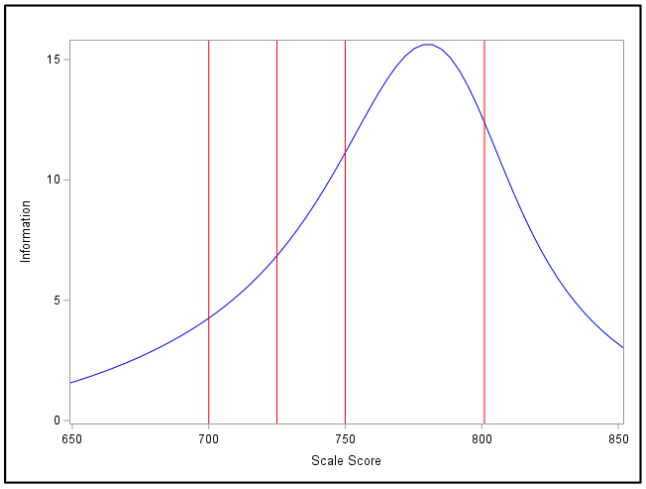


Figure N.18. Mathematics Grade 8 CSEM Curve

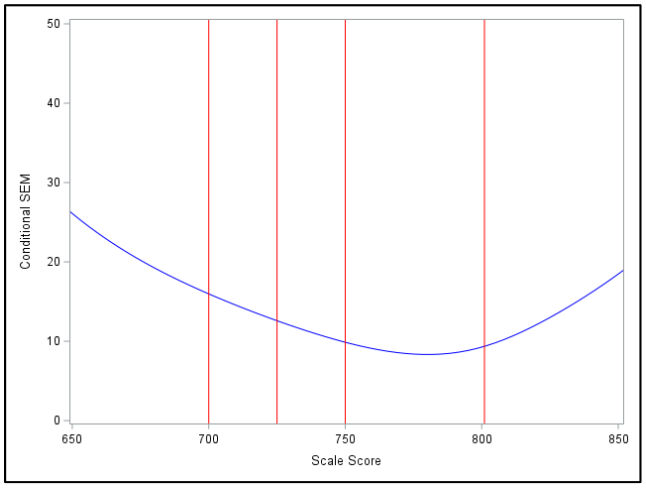


Figure N.19. ELA Grade 3 TCC

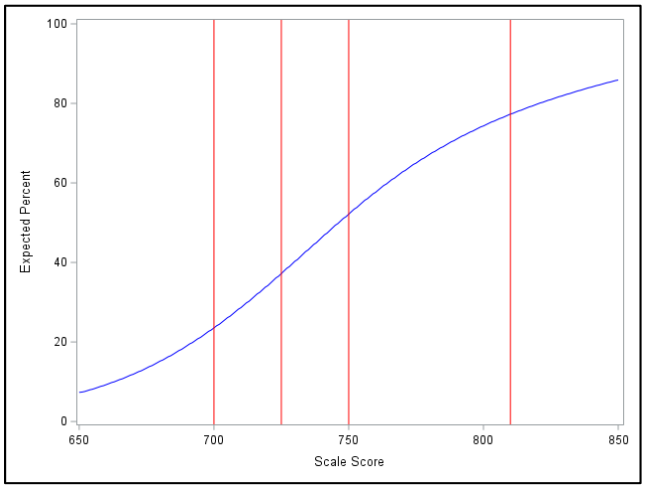


Figure N.20. ELA Grade 3 TIC

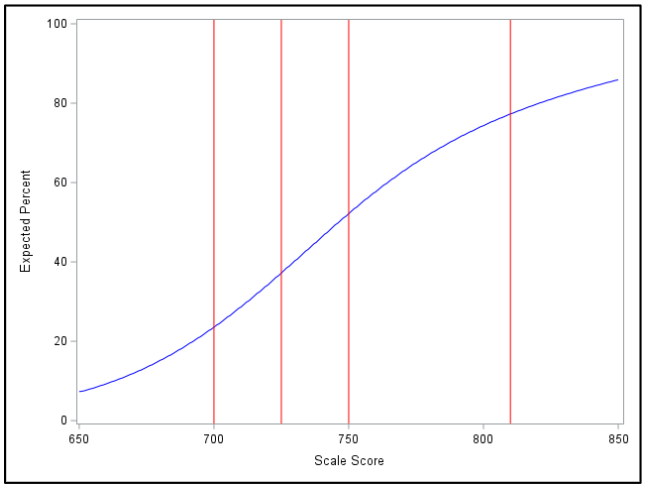


Figure N.21. ELA Grade 3 CSEM Curve

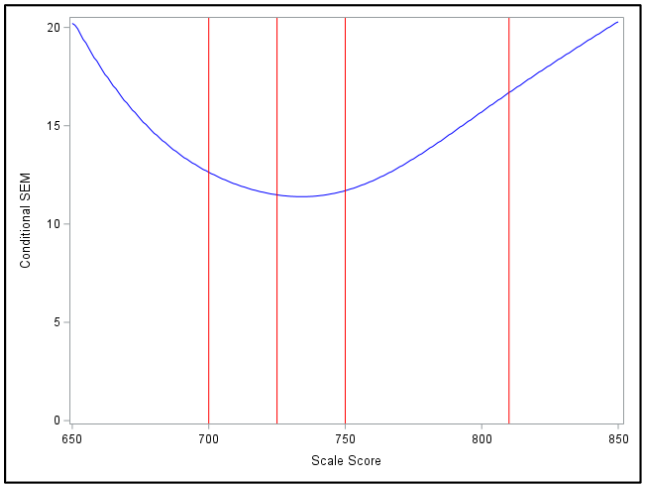


Figure N.22. ELA Grade 4 TCC

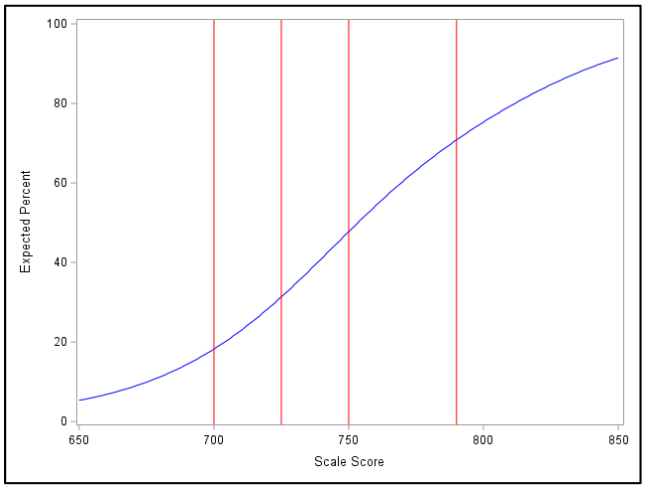


Figure N.23. ELA Grade 4 TIC

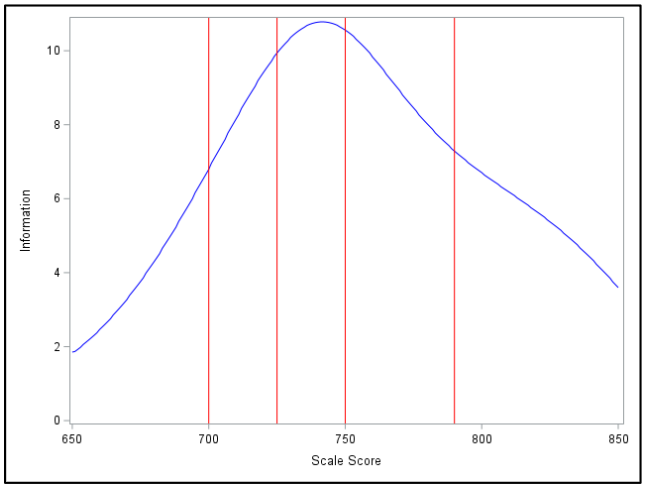


Figure N.24. ELA Grade 4 CSEM Curve

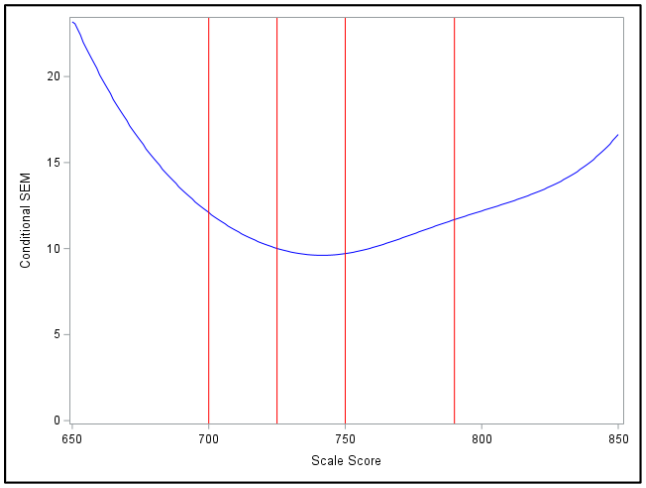


Figure N.25. ELA Grade 5 TCC

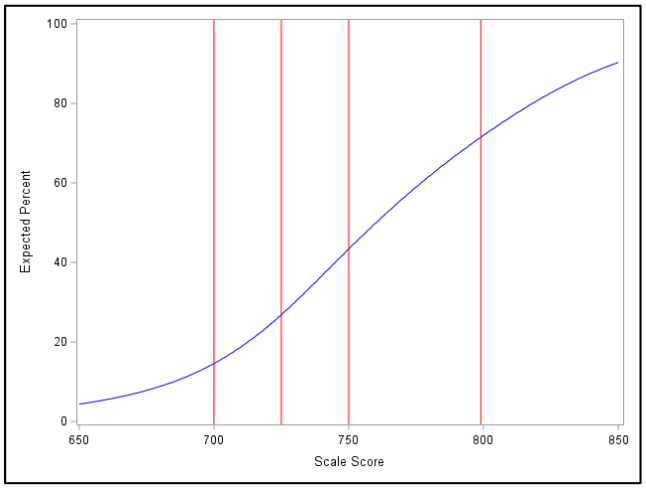


Figure N.26. ELA Grade 5 TIC

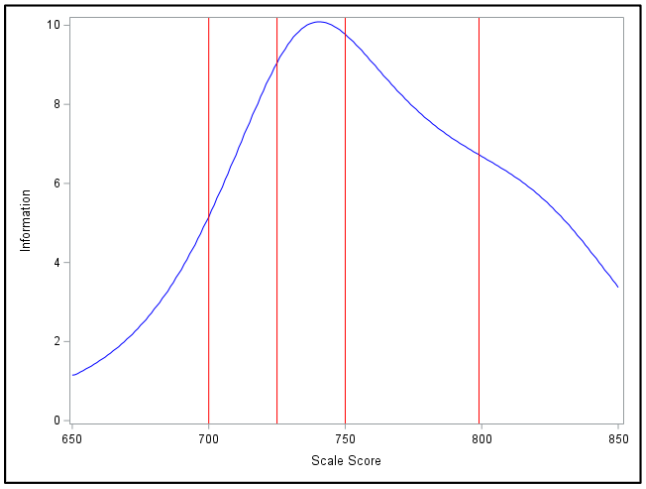


Figure N.27. ELA Grade 5 CSEM Curve

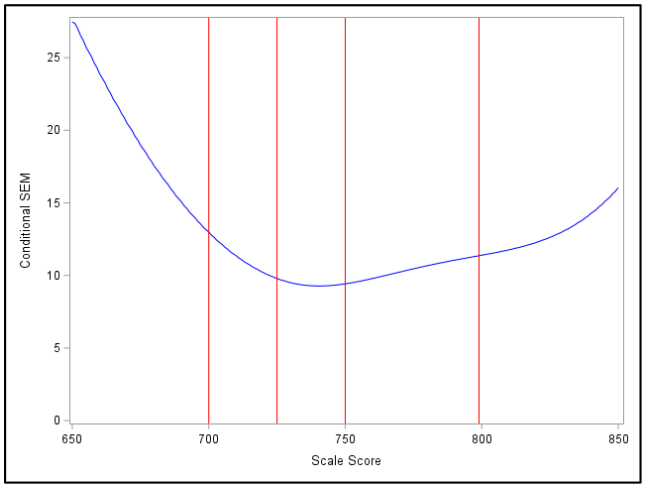


Figure N.28. ELA Grade 6 TCC

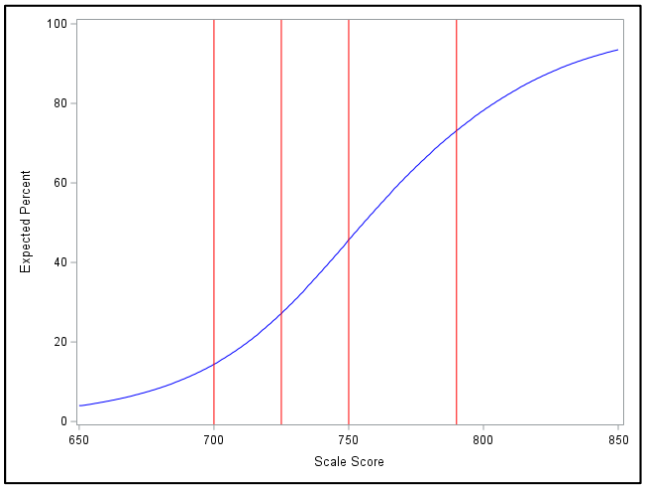


Figure N.29. ELA Grade 6 TIC

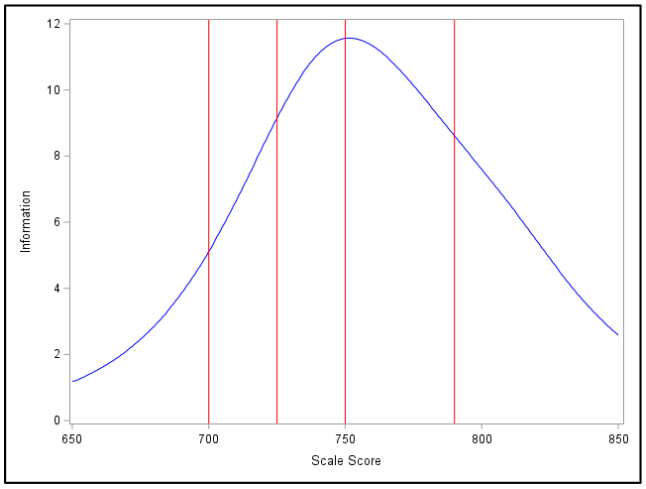


Figure N.30. ELA Grade 6 CSEM Curve

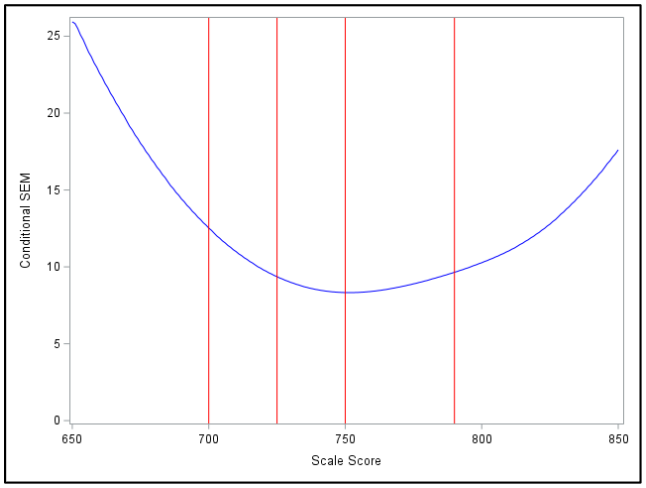


Figure N.31. ELA Grade 7 TCC

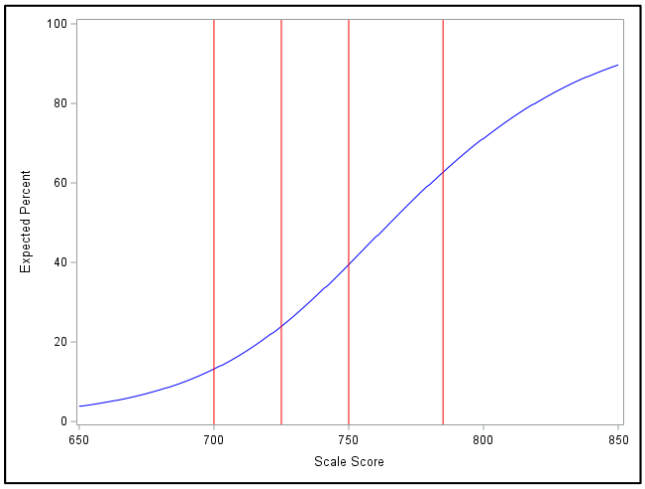


Figure N.32. ELA Grade 7 TIC

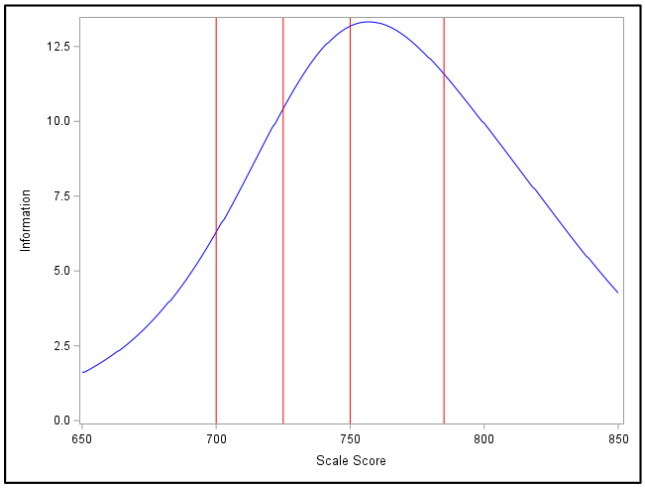


Figure N.33. ELA Grade 7 CSEM Curve

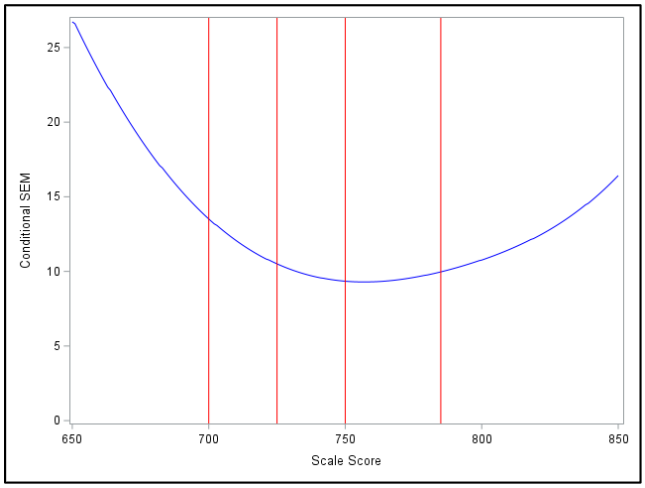


Figure N.34. ELA Grade 8 TCC

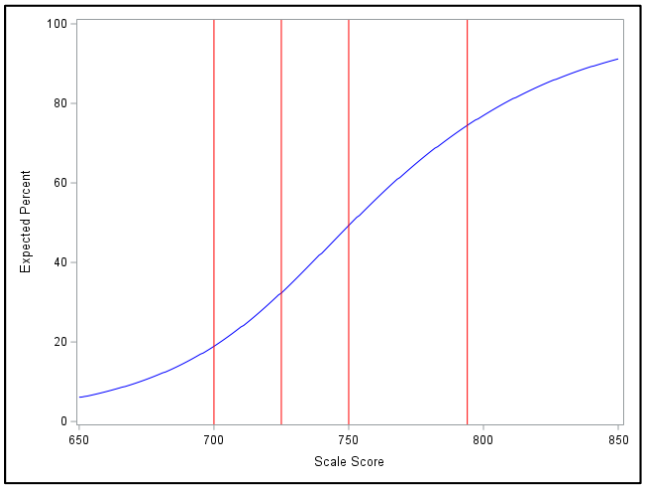


Figure N.35. ELA Grade 8 TIC

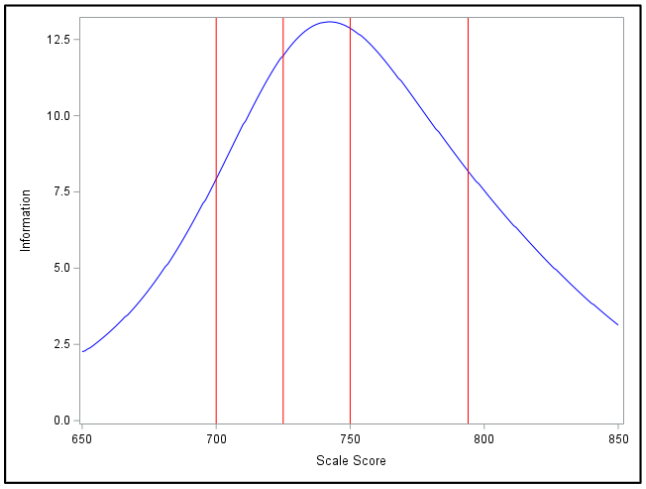


Figure N.36. ELA Grade 8 CSEM Curve

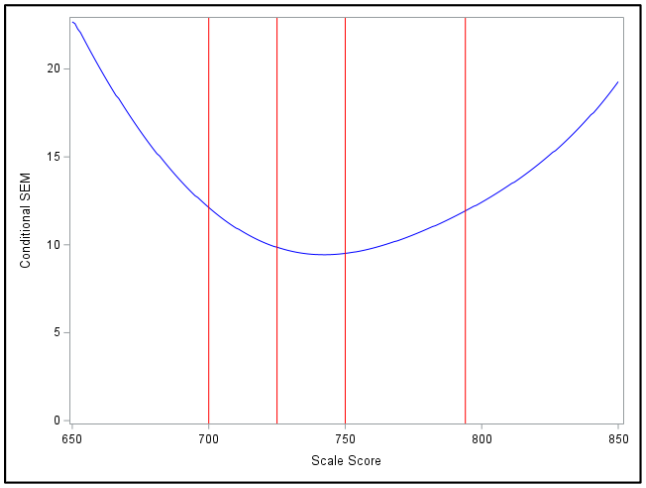


Figure N.37. CSLA Grade 3 TCC

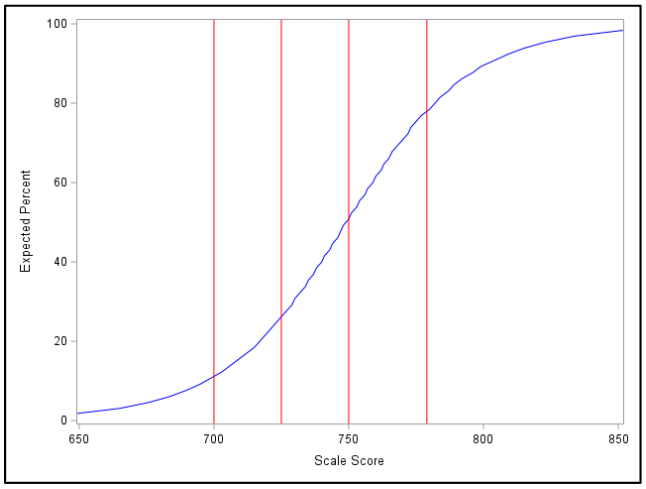


Figure N.38. CSLA Grade 3 TIC

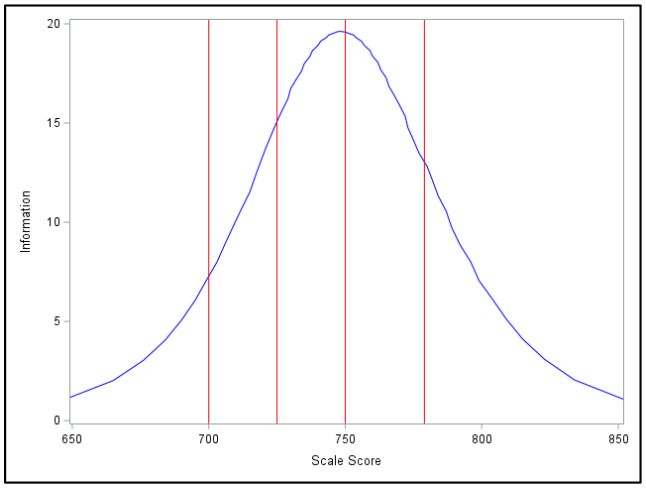


Figure N.39. CSLA Grade 3 CSEM Curve

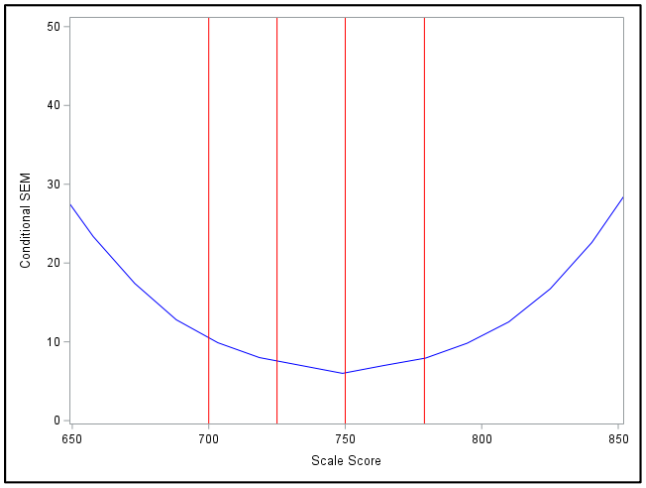


Figure N.40. CSLA Grade 4 TCC

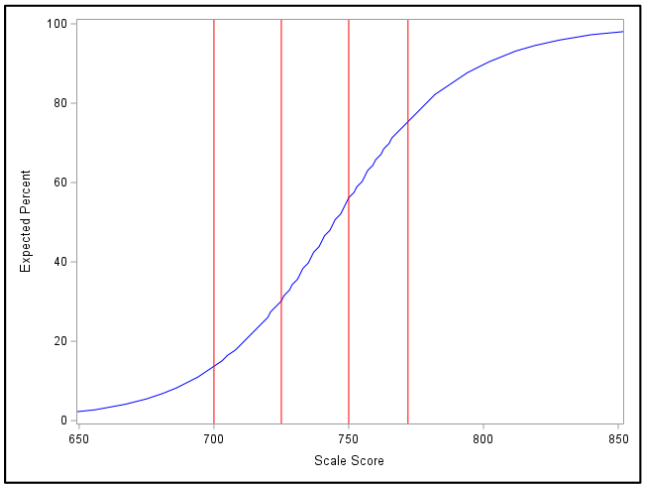


Figure N.41. CSLA Grade 4 TIC

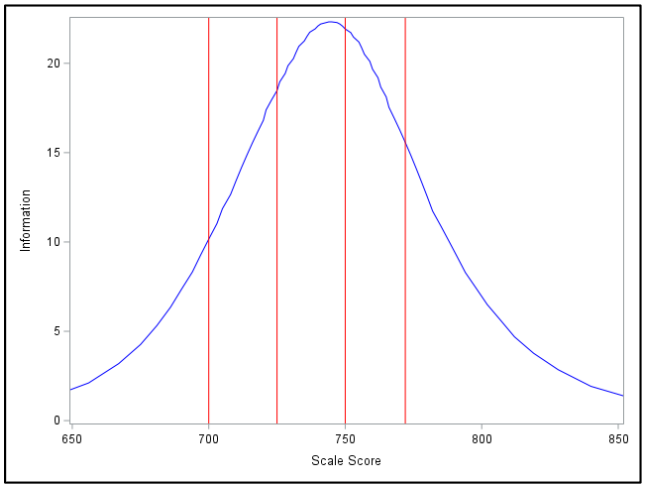


Figure N.42. CSLA Grade 4 CSEM Curve

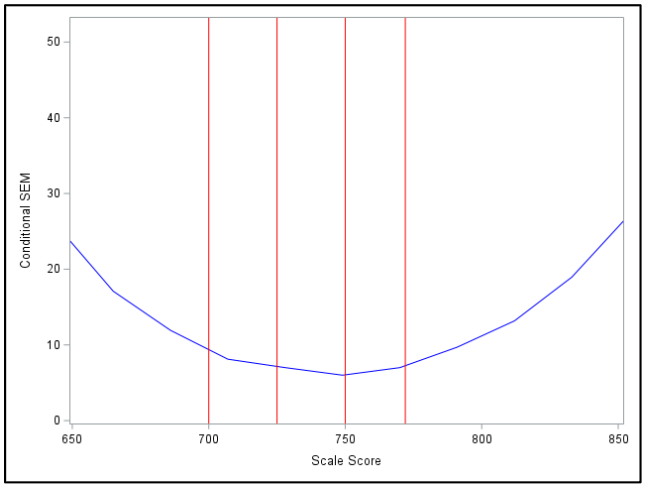


Figure N.43. ELA Reading Grade 3 TCC

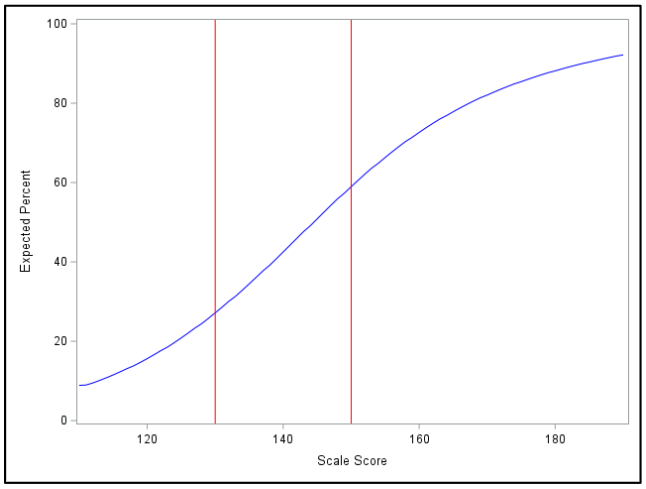


Figure N.44. ELA Reading Grade 3 TIC

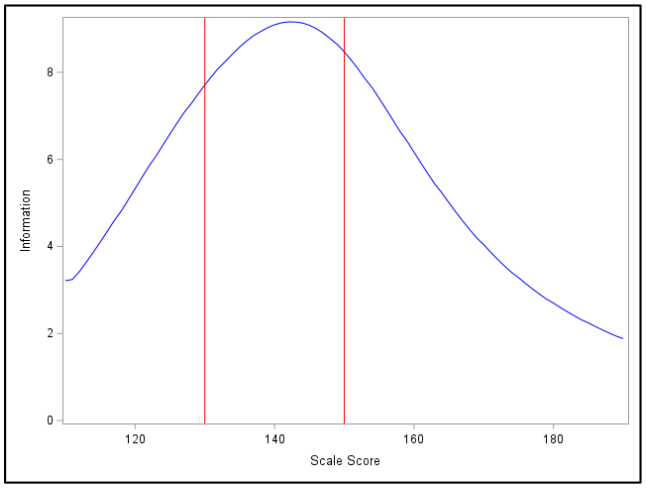


Figure N.45. ELA Reading Grade 3 CSEM Curve

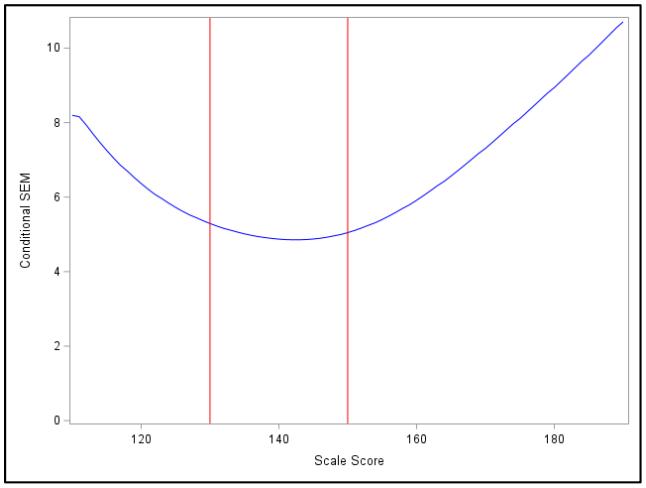


Figure N.46. ELA Reading Grade 4 TCC

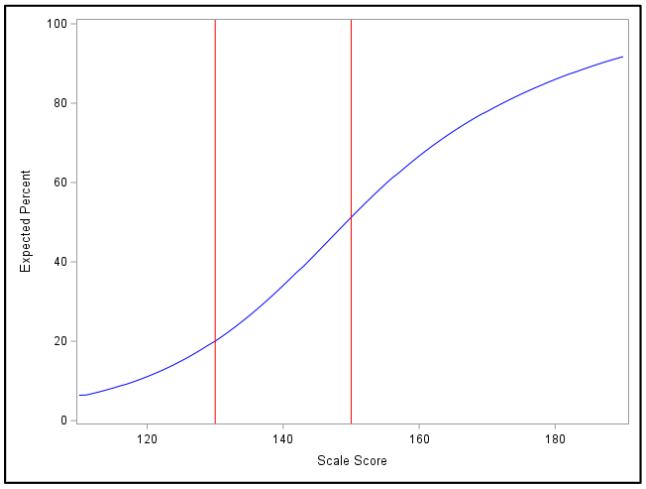


Figure N.47. ELA Reading Grade 4 TIC

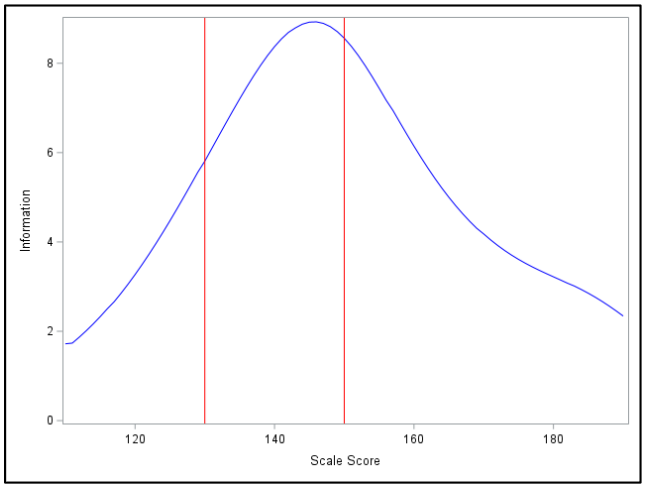


Figure N.48. ELA Reading Grade 4 CSEM Curve

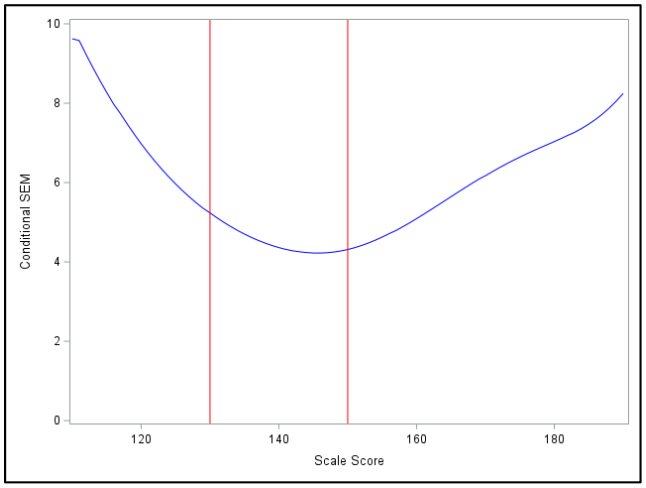


Figure N.49. ELA Reading Grade 5 TCC

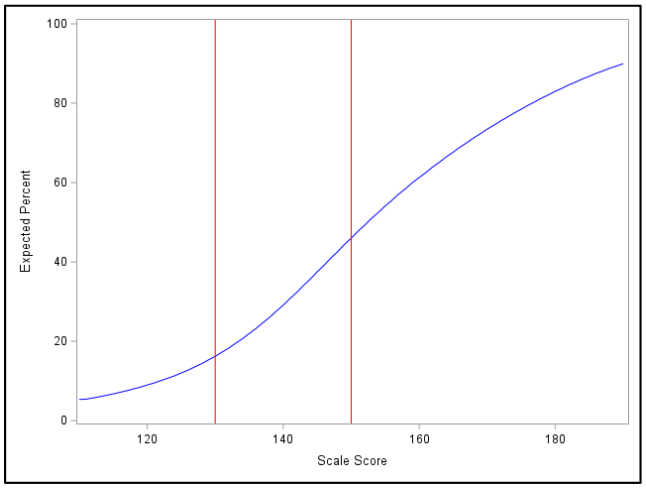


Figure N.50. ELA Reading Grade 5 TIC

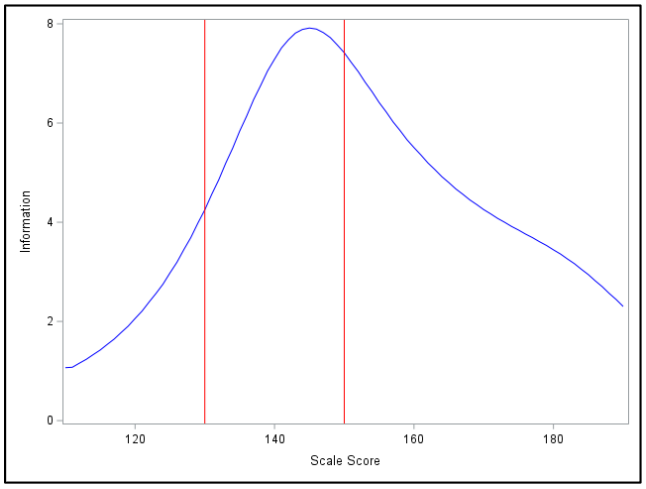


Figure N.51. ELA Reading Grade 5 CSEM Curve

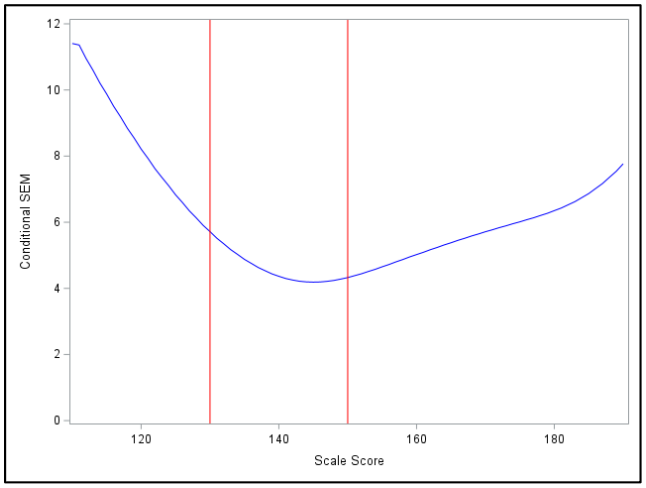


Figure N.52. ELA Reading Grade 6 TCC

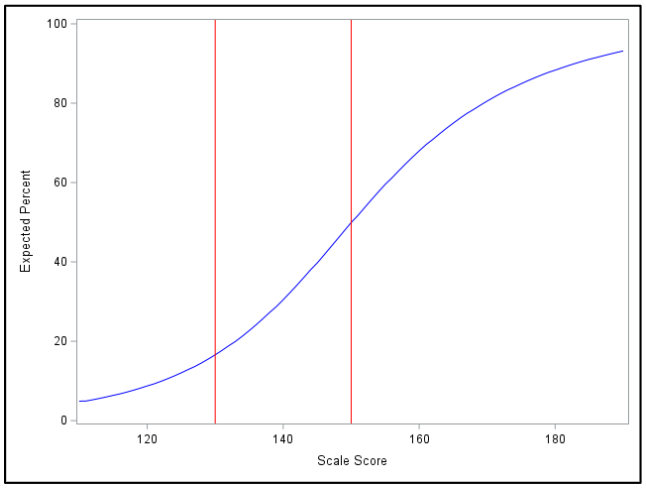


Figure N.53. ELA Reading Grade 6 TIC

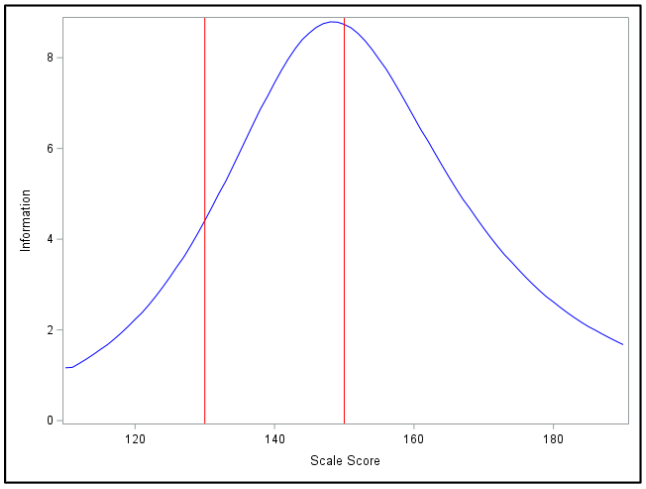


Figure N.54. ELA Reading Grade 6 CSEM Curve

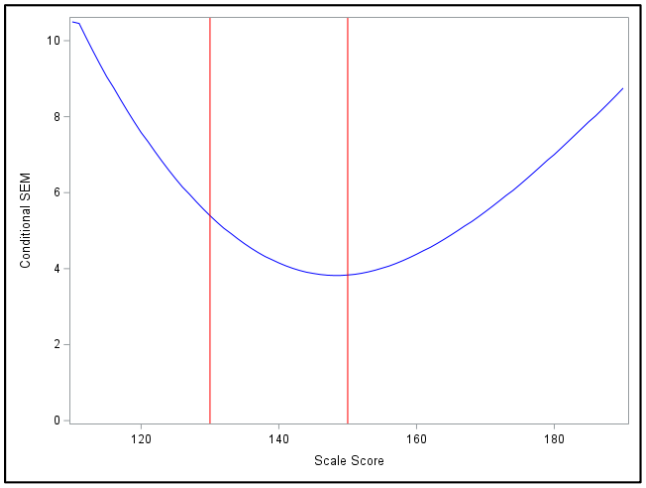


Figure N.55. ELA Reading Grade 7 TCC

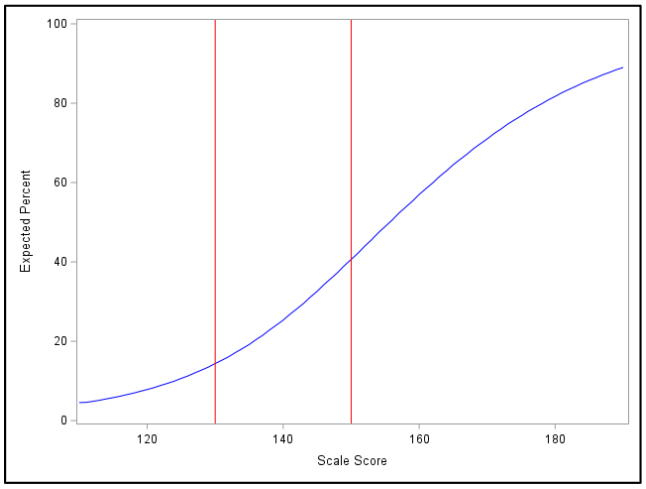


Figure N.56. ELA Reading Grade 7 TIC

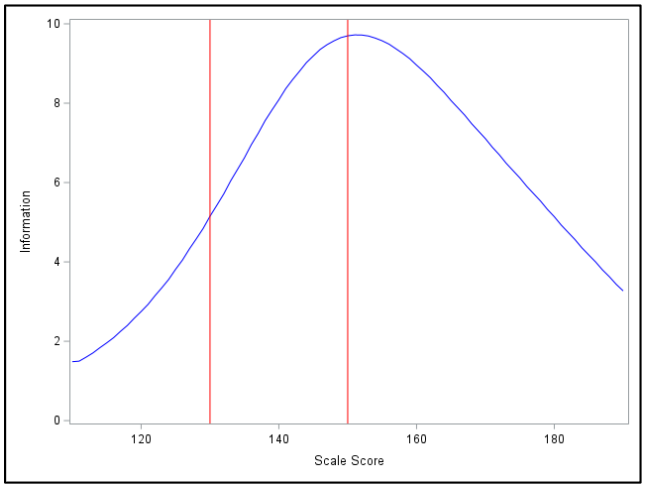


Figure N.57. ELA Reading Grade 7 CSEM Curve

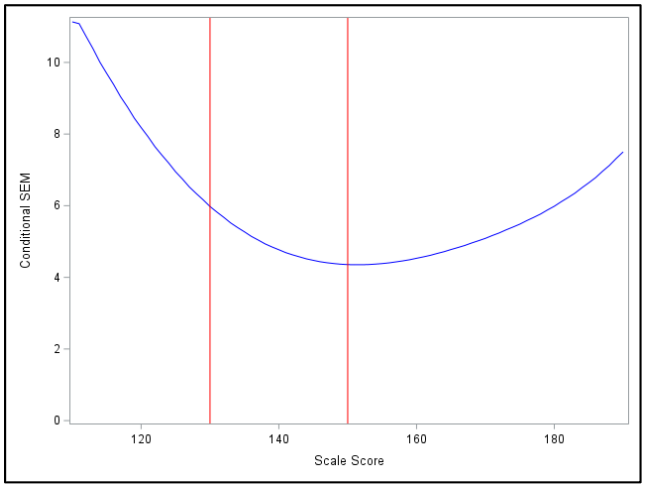


Figure N.58. ELA Reading Grade 8 TCC

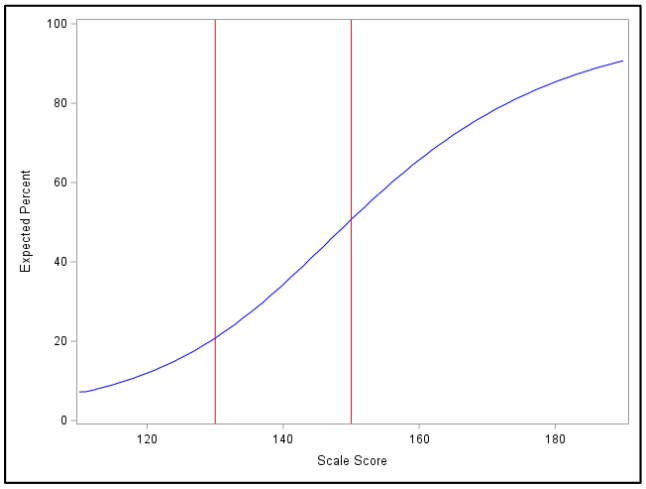


Figure N.59. ELA Reading Grade 8 TIC

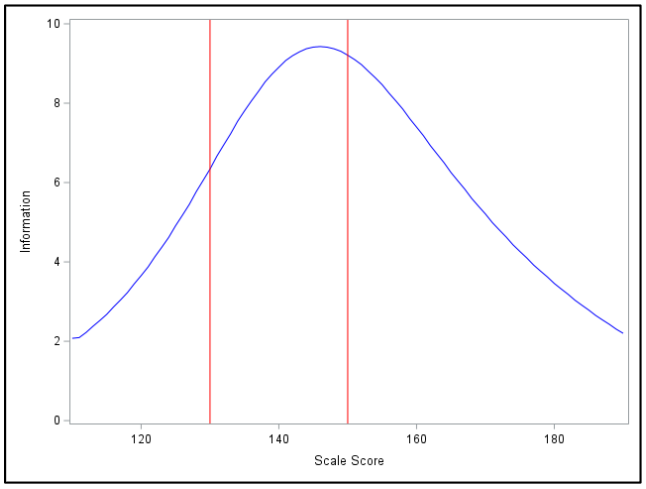


Figure N.60. ELA Reading Grade 8 CSEM Curve

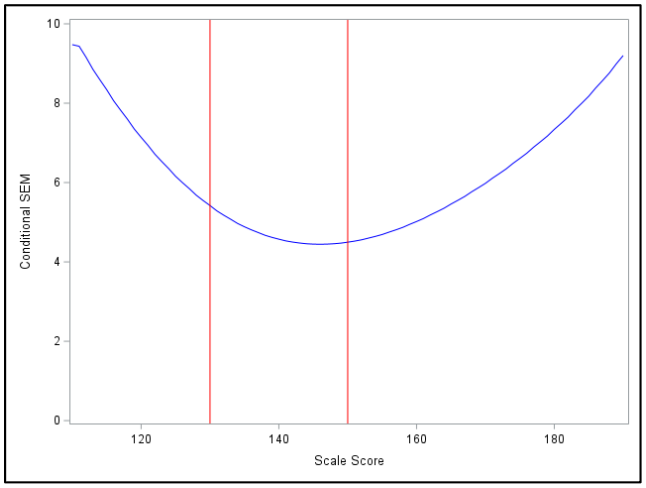


Figure N.61. CSLA Reading Grade 3 TCC

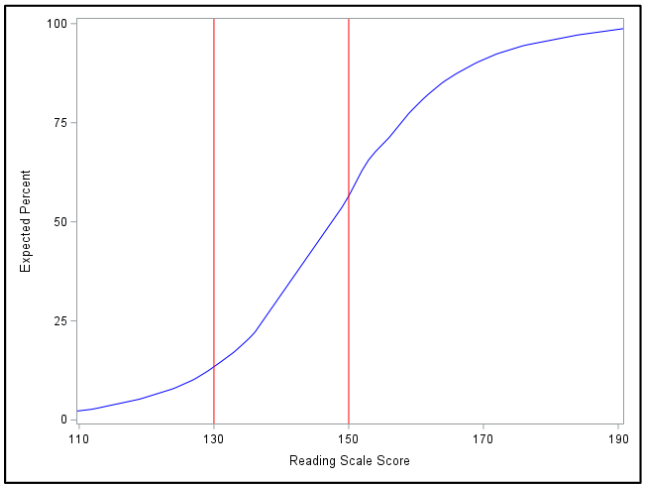


Figure N.62. CSLA Reading Grade 3 TIC

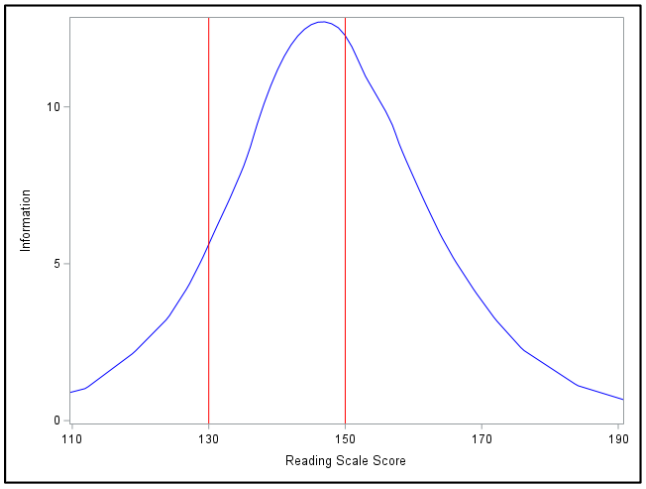


Figure N.63. CSLA Reading Grade 3 CSEM Curve

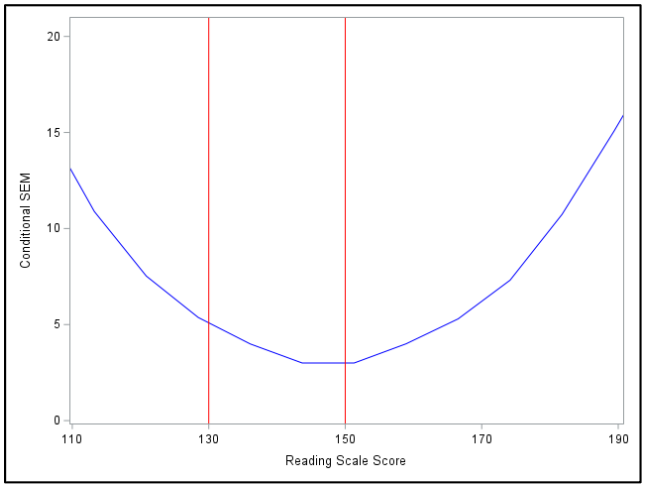


Figure N.64. CSLA Reading Grade 4 TCC

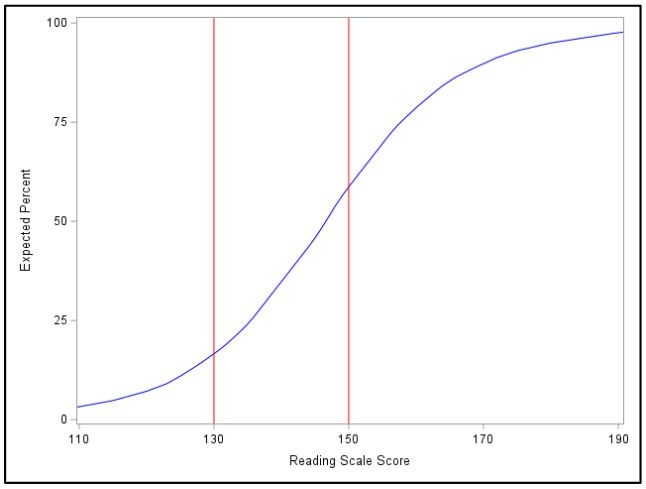


Figure N.65. CSLA Reading Grade 4 TIC

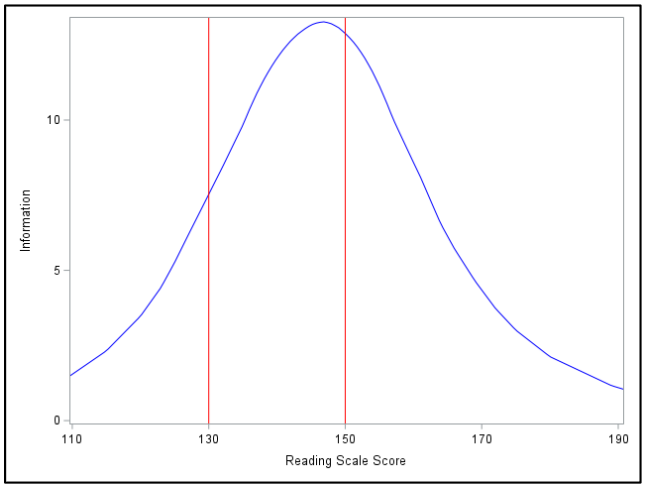


Figure N.66. CSLA Reading Grade 4 CSEM Curve

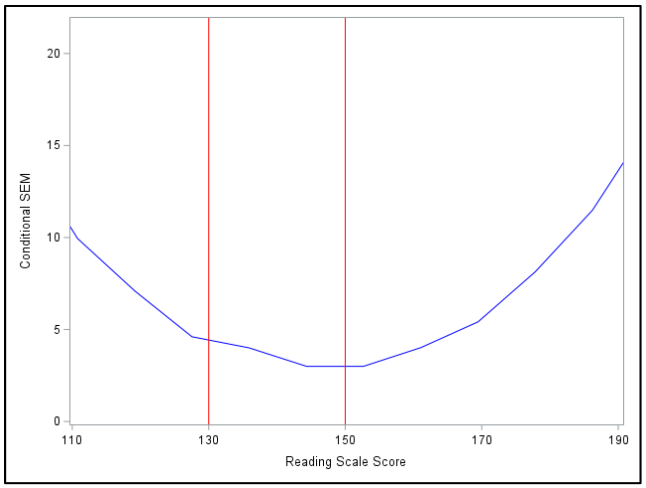


Figure N.67. Science Grade 5 TCC

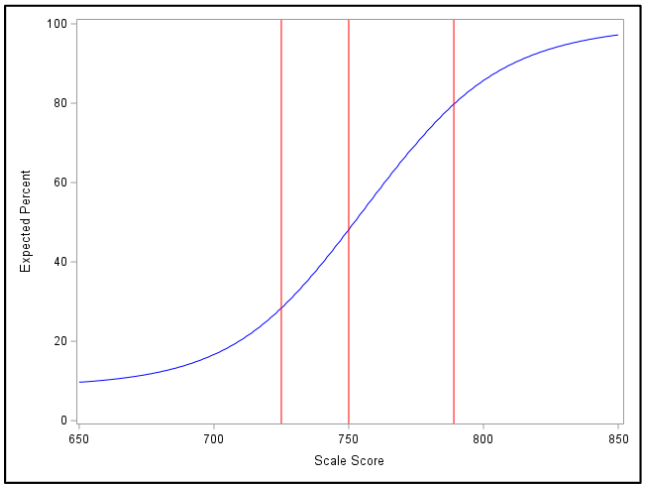


Figure N.68. Science Grade 5 TIC

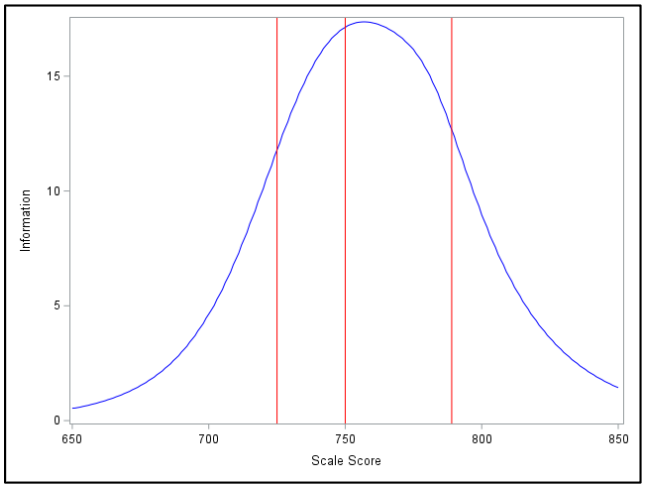


Figure N.69. Science Grade 5 CSEM

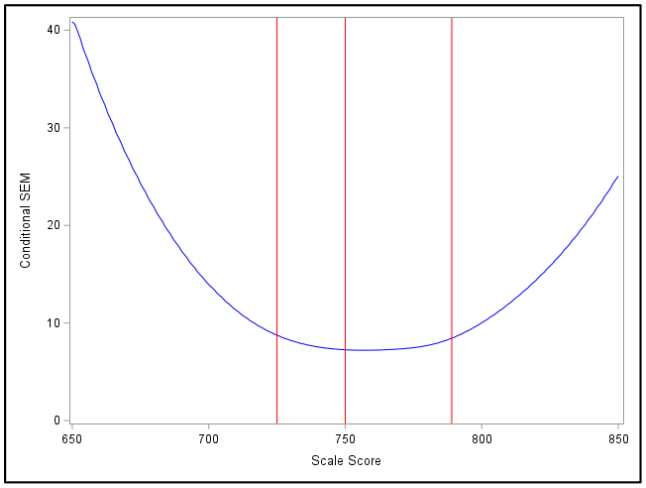


Figure N.70. Science Grade 8 TCC

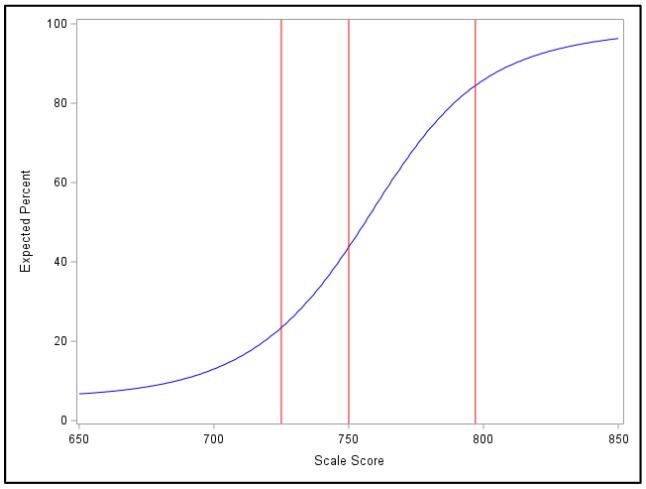


Figure N.71. Science Grade 8 TIC

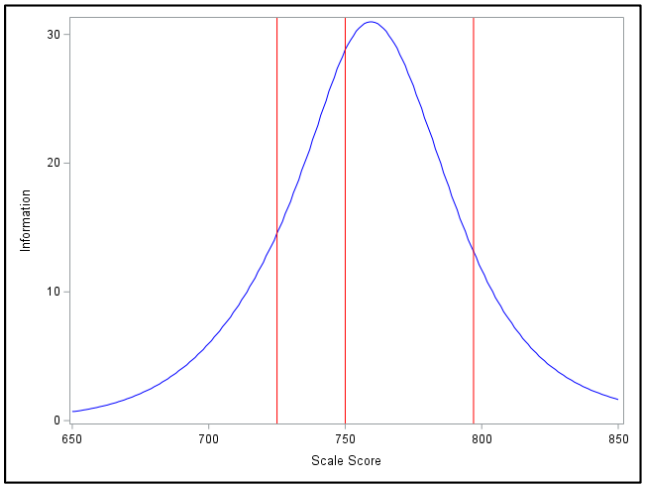


Figure N.72. Science Grade 8 CSEM

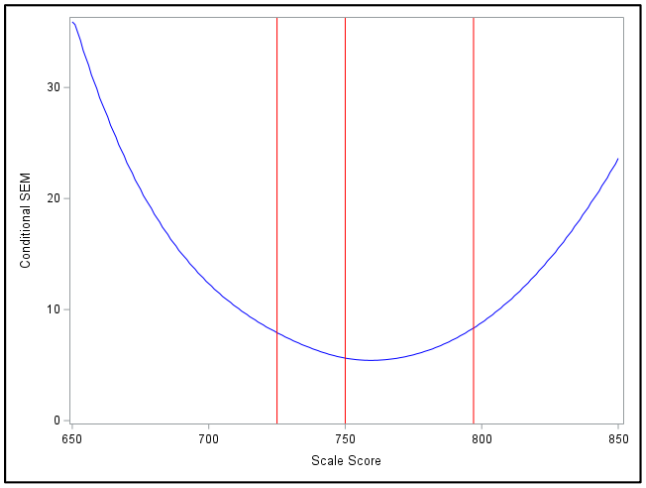


Figure N.73. Science Grade 11 TCC

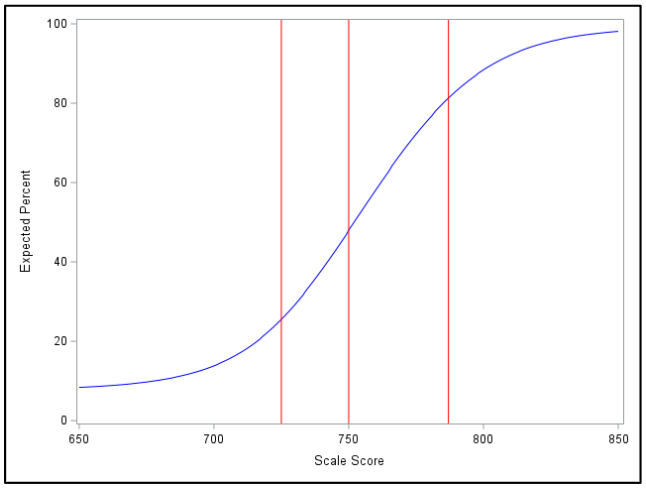


Figure N.74. Science Grade 11 TIC

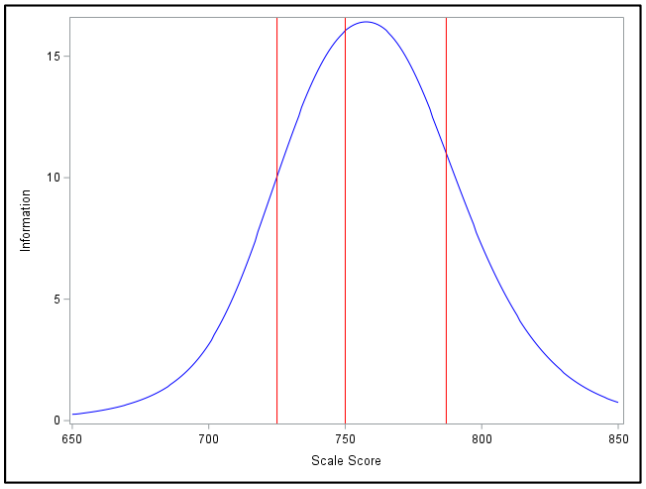
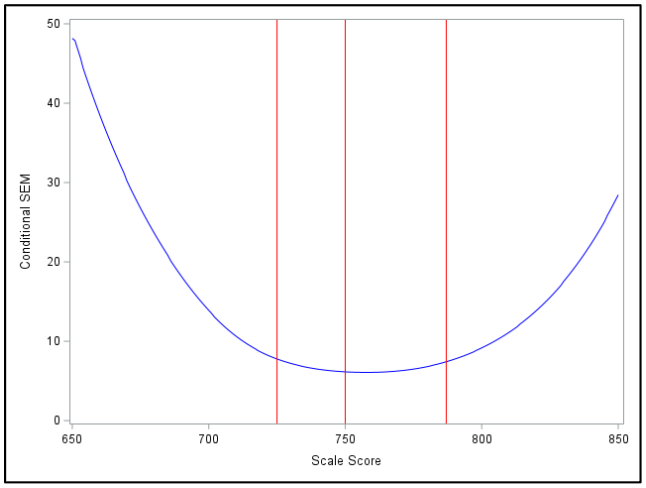


Figure N.75. Science Grade 11 CSEM



Appendix O: Inter-Rater Agreement

Table O.1. Operational Rater Agreement Statistics—Mathematics Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	5,428	88.0	11.6	0.4	0.94	0.00	0.94
Item 2	4	5,510	83.4	15.7	1.0	0.90	0.01	0.90
Item 3_PartB	2	5,059	91.4	8.3	0.2	0.92	0.02	0.92
Item 4_PartB	3	5,056	85.8	13.6	0.6	0.93	0.00	0.93
Item 4_PartC	4	4,997	92.0	6.9	1.1	0.95	0.00	0.95

Table O.2. Operational Rater Agreement Statistics—Mathematics Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	5,430	89.3	9.8	0.9	0.95	0.00	0.95
Item 2_PartB	3	5,044	93.4	5.7	0.9	0.97	0.00	0.97
Item 3	3	5,326	95.2	3.2	1.5	0.95	0.00	0.95
Item 4_PartA	2	5,118	89.5	10.2	0.2	0.90	0.01	0.90
Item 4_PartB	2	4,946	94.0	5.8	0.2	0.94	0.01	0.94
Item 5	6	5,493	89.6	9.1	1.3	0.98	0.00	0.98

Table O.3. Operational Rater Agreement Statistics—Mathematics Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1_PartA	3	5,494	92.0	7.4	0.7	0.96	0.00	0.96
Item 1_PartB	3	5,263	93.9	5.0	1.1	0.95	0.00	0.95
Item 2	3	5,335	90.9	8.6	0.5	0.93	0.01	0.93
Item 3_PartB	3	5,169	87.5	11.7	0.8	0.89	0.01	0.89
Item 4	4	5,449	80.9	17.1	2.1	0.92	0.00	0.92
Item 5_PartB	2	5,392	94.1	5.7	0.1	0.94	0.00	0.94

Table O.4. Operational Rater Agreement Statistics—Mathematics Grade 6

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1_PartA	4	5,371	91.0	7.0	2.0	0.95	0.00	0.95
Item 1_PartB	2	5,244	97.3	2.6	0.1	0.96	0.01	0.96
Item 2	3	5,507	88.5	11.1	0.4	0.95	0.00	0.95
Item 3	4	5,389	85.1	12.7	2.2	0.91	0.00	0.91
Item 4	4	5,439	75.0	22.1	2.9	0.92	0.01	0.92
Item 5	3	5,478	80.7	17.4	1.9	0.88	0.01	0.88

Table O.5. Operational Rater Agreement Statistics—Mathematics Grade 7

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1_PartA	3	5,215	87.0	10.0	3.0	0.85	0.00	0.85
Item 1_PartB	3	5,063	91.7	6.6	1.7	0.94	0.00	0.94
Item 2	3	5,147	88.9	10.8	0.3	0.95	0.00	0.95
Item 3	4	5,190	82.8	15.5	1.7	0.93	0.01	0.93
Item 4_PartB	3	5,128	88.9	10.9	0.2	0.91	0.01	0.91
Item 5	3	5,184	84.3	14.8	0.9	0.83	0.02	0.83

Table O.6. Operational Rater Agreement Statistics—Mathematics Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	4,942	92.1	7.6	0.3	0.96	0.01	0.96
Item 2_PartB	3	4,677	88.0	11.2	0.8	0.88	0.01	0.88
Item 3	4	4,838	89.4	9.6	1.0	0.91	0.00	0.91
Item 4	3	5,033	87.9	10.7	1.4	0.93	0.00	0.93
Item 5_PartA	4	4,914	93.4	6.0	0.6	0.96	0.00	0.96
Item 5_PartB	2	4,809	90.2	9.5	0.3	0.92	0.00	0.92

Table O.7. Field Test Rater Agreement Statistics—Mathematics Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	2,945	94.8	3.8	1.4	0.97	0.01	0.97
Item 2	3	2,929	92.8	6.7	0.5	0.97	0.00	0.97
Item 3	3	2,899	87.4	11.8	0.8	0.94	0.01	0.94
Item 4_PartA	2	2,762	92.0	7.7	0.2	0.93	0.00	0.93
Item 5	3	2,865	92.3	6.8	0.9	0.92	0.00	0.92
Item 6_PartB	3	2,997	87.5	12.3	0.3	0.94	0.01	0.94
Item 7_PartA	2	2,957	92.7	7.2	0.1	0.95	0.00	0.95
Item 7_PartB	2	2,956	91.0	8.9	0.1	0.92	0.02	0.92
Item 8_PartC	2	2,997	90.8	9.1	0.0	0.91	0.00	0.91
Item 9_PartB	2	2,831	95.3	4.6	0.1	0.97	0.00	0.97
Item 10_PartC	2	2,899	97.2	2.7	0.1	0.97	0.00	0.97
Item 10_PartD	2	2,899	95.4	4.5	0.1	0.96	0.00	0.96
Item 11	3	2,993	90.4	9.1	0.5	0.96	0.00	0.96
Item 12	3	2,996	81.1	17.6	1.3	0.91	0.00	0.91
Item 13_PartB	2	2,881	92.4	7.4	0.2	0.94	0.00	0.94
Item 14	3	2,875	85.1	13.8	1.1	0.89	0.01	0.89
Item 15_PartA	2	2,911	90.3	9.4	0.3	0.93	0.00	0.93
Item 15_PartB	2	2,908	92.2	7.6	0.2	0.92	0.02	0.92
Item 16_PartB	3	2,963	85.2	13.8	1.0	0.91	0.02	0.91
Item 17_PartA	2	2,995	94.3	5.6	0.1	0.94	0.01	0.95
Item 17_PartB	2	2,995	90.9	9.1	0.0	0.92	0.02	0.92
Item 18	3	2,996	86.7	12.5	0.8	0.94	0.01	0.94
Item 19_PartA	2	2,908	90.3	9.6	0.1	0.90	0.01	0.90
Item 19_PartB	2	2,908	91.8	8.0	0.1	0.93	0.01	0.93
Item 20_PartA	2	2,873	97.0	2.9	0.2	0.98	0.00	0.98
Item 20_PartB	2	2,873	97.3	2.6	0.1	0.97	0.01	0.97
Item 21_PartA	2	2,901	91.2	8.7	0.1	0.93	0.00	0.93
Item 21_PartB	2	2,900	93.1	6.9	0.0	0.94	0.01	0.94
Item 23_PartB	2	2,852	97.3	2.7	0.0	0.98	0.00	0.98
Item 24_PartA	4	2,985	93.5	5.7	0.8	0.98	0.01	0.98
Item 24_PartB	2	2,983	94.8	5.0	0.2	0.97	0.00	0.97
Item 25_PartB	2	2,984	93.9	5.5	0.5	0.95	0.00	0.95
Item 25_PartC	3	2,983	90.3	9.3	0.4	0.95	0.00	0.95

Table O.8. Field Test Rater Agreement Statistics—Mathematics Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1_PartB	2	2,998	90.2	9.7	0.1	0.88	0.00	0.88
Item 2	3	2,997	80.7	18.1	1.1	0.89	0.01	0.89
Item 3	3	2,988	89.4	9.9	0.8	0.90	0.01	0.90
Item 4_PartA	2	2,838	94.9	4.7	0.4	0.95	0.01	0.95
Item 4_PartB	2	2,832	94.1	5.8	0.2	0.90	0.00	0.90
Item 5_PartA	2	2,992	92.1	7.7	0.2	0.95	0.00	0.95
Item 5_PartB	2	2,985	90.5	9.0	0.5	0.89	0.01	0.89
Item 6	3	2,990	92.5	6.1	1.3	0.89	0.01	0.89
Item 7	3	2,973	86.2	13.0	0.8	0.94	0.01	0.94
Item 8_PartB	2	2,969	94.1	5.6	0.4	0.95	0.01	0.95
Item 9	3	2,994	81.9	16.7	1.4	0.87	0.01	0.87
Item 10_PartB	3	2,992	82.6	16.2	1.2	0.91	0.01	0.91
Item 11_PartA	2	2,994	93.6	6.2	0.2	0.96	0.00	0.96
Item 11_PartB	2	2,990	90.0	9.7	0.3	0.93	0.00	0.93
Item 12_PartA	2	2,828	92.6	7.2	0.2	0.95	0.01	0.95
Item 12_PartB	2	2,823	93.2	6.7	0.1	0.95	0.00	0.95
Item 13_PartA	2	2,878	90.6	9.1	0.3	0.89	0.01	0.89
Item 13_PartB	2	2,870	92.0	7.8	0.2	0.84	0.00	0.84
Item 14	3	2,992	80.7	17.9	1.4	0.90	0.00	0.90
Item 15	3	2,944	79.8	19.4	0.8	0.87	0.00	0.87
Item 16_PartA	2	2,866	92.6	7.2	0.1	0.91	0.01	0.91
Item 17_PartA	2	2,888	89.6	10.0	0.3	0.93	0.00	0.93
Item 17_PartB	2	2,886	91.0	8.9	0.1	0.93	0.01	0.93
Item 18	4	2,896	84.2	13.0	2.8	0.91	0.01	0.91
Item 19_PartA	2	2,818	90.2	9.6	0.2	0.92	0.01	0.92
Item 19_PartB	2	2,814	94.7	5.2	0.1	0.96	0.01	0.96
Item 20_PartB	3	2,975	91.7	7.7	0.7	0.95	0.00	0.95
Item 20_PartC	2	2,972	90.6	6.8	2.6	0.89	0.00	0.89
Item 21_PartB	3	2,888	87.4	11.3	1.4	0.89	0.01	0.89
Item 21_PartC	2	2,884	96.7	2.6	0.7	0.97	0.01	0.97
Item 22	3	2,994	91.1	8.2	0.7	0.94	0.00	0.94

Table O.9. Field Test Rater Agreement Statistics—Mathematics Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1_PartC	4	2,986	85.0	14.0	1.0	0.95	0.01	0.95
Item 2	3	2,989	82.3	16.4	1.3	0.89	0.01	0.89
Item 3_PartA	2	2,995	90.2	9.7	0.1	0.89	0.03	0.89
Item 3_PartB	2	2,993	96.4	3.1	0.5	0.97	0.00	0.97
Item 4_PartB	3	2,983	94.7	5.0	0.3	0.94	0.00	0.94
Item 5_PartB	3	2,992	94.2	5.7	0.0	0.95	0.01	0.95
Item- 6	4	2,990	84.5	15.5	0.0	0.92	0.01	0.92
Item 7_PartB	3	2,994	96.5	3.1	0.4	0.99	0.01	0.99
Item 8_PartA	3	2,983	88.4	10.9	0.7	0.93	0.00	0.93
Item 8_PartB	3	2,986	91.6	7.4	1.0	0.96	0.01	0.96
Item 9	3	2,988	93.0	6.4	0.6	0.97	0.01	0.97
Item 10	3	2,992	90.9	8.9	0.2	0.97	0.00	0.97
Item 11	3	2,994	89.1	10.4	0.6	0.93	0.02	0.93
Item 12	3	2,991	83.7	14.4	1.9	0.90	0.01	0.90
Item 13_PartA	2	2,989	94.4	4.3	1.3	0.93	0.01	0.93
Item 13_PartB	2	2,990	92.5	7.4	0.1	0.93	0.01	0.93
Item 14	3	2,992	86.0	13.5	0.5	0.94	0.01	0.95
Item 15	3	2,987	90.9	8.1	0.9	0.94	0.01	0.94
Item 16	3	2,996	91.7	8.0	0.3	0.97	0.01	0.97
Item 17	3	2,990	86.9	13.1	0.0	0.95	0.00	0.95
Item 18	4	2,997	87.7	10.9	1.4	0.92	0.00	0.92
Item 19	4	2,993	77.7	19.3	3.0	0.92	0.00	0.92
Item 19	3	2,990	84.7	14.7	0.6	0.90	0.01	0.90
Item 20_PartA	2	2,997	94.3	5.4	0.3	0.96	0.00	0.96
Item 20_PartB	4	2,996	87.0	12.1	0.9	0.96	0.00	0.96
Item 21_PartB	3	2,992	88.3	10.6	1.1	0.93	0.00	0.93
Item 21_PartC	2	2,991	93.5	6.1	0.4	0.95	0.00	0.95

Table O.10. Field Test Rater Agreement Statistics—Mathematics Grade 6

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	2,982	81.4	16.1	2.5	0.84	0.02	0.84
Item 2	3	2,991	92.2	7.3	0.5	0.95	0.00	0.95
Item 3	3	2,985	86.2	13.2	0.6	0.92	0.01	0.92
Item 4_PartB	2	2,988	90.3	9.3	0.4	0.88	0.01	0.88
Item 5	3	2,974	92.7	7.1	0.2	0.91	0.01	0.91
Item 6	3	2,994	88.6	10.5	0.9	0.94	0.01	0.94
Item 7	4	2,983	83.5	15.1	1.4	0.93	0.01	0.93
Item 8	3	2,995	94.4	5.6	0.0	0.97	0.00	0.97
Item 9	3	2,989	92.3	7.3	0.4	0.96	0.00	0.96
Item 10	3	2,987	89.3	9.7	1.0	0.93	0.00	0.93
Item 11	3	2,987	85.4	13.6	1.0	0.85	0.02	0.85
Item 12_PartA	2	2,985	90.8	9.1	0.2	0.91	0.00	0.91
Item 12_PartB	2	2,987	92.6	7.3	0.1	0.89	0.00	0.89
Item 13_PartA	2	2,989	89.8	10.1	0.1	0.90	0.00	0.90
Item 13_PartB	2	2,985	93.0	6.9	0.1	0.93	0.01	0.93
Item 14_PartA	2	2,993	95.6	4.0	0.3	0.97	0.00	0.97
Item 14_PartB	2	2,989	90.2	9.6	0.2	0.91	0.00	0.91
Item 15	3	2,997	85.8	13.9	0.3	0.84	0.01	0.84
Item 16	3	2,993	89.0	9.8	1.2	0.89	0.00	0.89
Item 17	4	2,990	91.5	8.2	0.2	0.97	0.00	0.97
Item 18	4	2,989	88.8	10.4	0.8	0.97	0.00	0.97
Item 19_PartA	2	2,989	91.6	8.2	0.2	0.90	0.01	0.90
Item 19_PartB	2	2,982	96.8	3.0	0.3	0.89	0.01	0.89
Item 20	3	2,984	86.4	11.8	1.8	0.91	0.00	0.91
Item 21	3	2,992	90.7	9.1	0.2	0.96	0.01	0.96
Item 22_PartA	3	2,994	91.0	7.4	1.6	0.95	0.00	0.95
Item 22_PartB	3	2,991	87.5	12.0	0.5	0.92	0.01	0.92

Table O.11. Field Test Rater Agreement Statistics—Mathematics Grade 7

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	2,968	94.2	5.1	0.7	0.86	0.00	0.86
Item 2_PartA	2	2,988	94.1	5.4	0.5	0.95	0.00	0.95
Item 2_PartB	2	2,987	95.8	4.1	0.1	0.97	0.00	0.97
Item 3	3	2,978	91.9	7.2	0.9	0.95	0.00	0.95
Item 4	3	2,985	92.2	7.4	0.4	0.95	0.00	0.95
Item 5	3	2,985	91.6	7.2	1.2	0.92	0.00	0.92
Item 6	3	2,969	96.2	3.3	0.5	0.95	0.00	0.95
Item 7_PartB	3	2,986	92.7	7.0	0.2	0.97	0.01	0.97
Item 8	4	2,979	82.8	15.4	1.7	0.94	0.00	0.94
Item 9	4	2,986	83.1	15.8	1.1	0.93	0.01	0.93
Item 10	4	2,983	92.3	7.2	0.6	0.92	0.00	0.92
Item 11	4	2,980	87.6	11.7	0.7	0.96	0.00	0.96
Item 12_PartB	3	2,980	92.7	6.9	0.4	0.96	0.00	0.96
Item 13	3	2,983	88.9	10.5	0.6	0.94	0.01	0.94
Item 14	3	2,974	84.4	14.1	1.4	0.90	0.00	0.90
Item 15	3	2,989	94.3	5.2	0.5	0.97	0.01	0.97
Item 16	3	2,986	91.3	8.4	0.3	0.96	0.00	0.96
Item 17	4	2,991	93.2	5.8	1.0	0.98	0.00	0.98
Item 18	4	2,984	84.6	13.3	2.1	0.95	0.01	0.95
Item 19	4	2,979	92.6	6.5	0.9	0.98	0.00	0.98
Item 20	3	2,992	93.8	5.7	0.5	0.96	0.00	0.96
Item 21	6	2,984	84.9	13.7	1.4	0.97	0.00	0.97
Item 22_PartA	4	2,991	90.8	7.3	1.9	0.96	0.00	0.96
Item 22_PartB	2	2,980	97.1	2.7	0.2	0.93	0.00	0.93

Table O.12. Field Test Rater Agreement Statistics—Mathematics Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	3	2,977	93.1	6.6	0.3	0.96	0.00	0.96
Item 2_PartA	2	2,982	92.7	6.9	0.4	0.95	0.01	0.95
Item 2_PartB	2	2,967	97.0	3.0	0.1	0.96	0.00	0.96
Item 3	6	2,964	90.1	7.8	2.1	0.93	0.00	0.93
Item 4	3	2,972	88.9	10.1	1.0	0.93	0.00	0.93
Item 5	3	2,968	86.6	12.4	1.0	0.95	0.01	0.95
Item 6_PartA	4	2,980	92.9	6.6	0.4	0.97	0.01	0.97
Item 6_PartB	2	2,975	97.9	1.9	0.2	0.95	0.00	0.95
Item 7	3	2,988	92.2	7.7	0.1	0.96	0.01	0.96
Item 8_PartA	2	2,975	94.1	5.8	0.1	0.95	0.00	0.95
Item 8_PartB	2	2,967	97.8	2.2	0.0	0.99	0.00	0.99
Item 9	3	2,980	87.3	12.0	0.6	0.93	0.00	0.93
Item 10	4	2,988	93.2	6.2	0.6	0.98	0.00	0.98
Item 11	3	2,976	95.6	4.0	0.4	0.97	0.00	0.97
Item 12_PartB	2	2,964	97.9	2.0	0.1	0.96	0.01	0.96
Item 13	3	2,979	83.8	15.6	0.6	0.91	0.00	0.91
Item 14	4	2,990	98.2	1.5	0.3	0.96	0.00	0.96
Item 15	6	2,979	86.7	12.5	0.8	0.98	0.00	0.98

Table O.13. Operational Rater Agreement Statistics—ELA Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	5,578	85.7	14.3	0.0	0.81	0.01	0.81
PCR 1 WE	3	5,578	82.8	17.1	0.1	0.77	0.01	0.77
PCR 2 WKL	3	5,580	86.1	13.9	0.0	0.80	0.01	0.80
PCR 2 WE	3	5,580	81.9	17.9	0.1	0.79	0.01	0.79

Table O.14. Operational Rater Agreement Statistics—ELA Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	5,545	80.9	18.8	0.2	0.85	0.02	0.85
PCR 1 WE	3	5,545	80.9	19.0	0.2	0.86	0.01	0.86
PCR 2 WKL	3	5,557	80.3	19.6	0.1	0.77	0.03	0.77
PCR 2 WE	4	5,557	74.9	24.9	0.2	0.71	0.01	0.71

Table O.15. Operational Rater Agreement Statistics—ELA Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	5,678	80.8	19.1	0.1	0.87	0.01	0.87
PCR 1 WE	3	5,678	80.9	19.1	0.1	0.88	0.01	0.88
PCR 2 WKL	3	5,662	80.8	18.6	0.6	0.78	0.01	0.78
PCR 2 WE	4	5,662	74.8	24.1	1.1	0.74	0.00	0.74

Table O.16. Operational Rater Agreement Statistics—ELA Grade 6

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	5,569	80.7	19.3	0.0	0.88	0.00	0.88
PCR 1 WE	4	5,569	79.6	20.4	0.0	0.88	0.01	0.88
PCR 2 WKL	3	5,561	79.6	19.8	0.6	0.84	0.01	0.84
PCR 2 WE	4	5,561	77.6	22.0	0.4	0.83	0.01	0.83

Table O.17. Operational Rater Agreement Statistics—ELA Grade 7

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	5,393	79.9	19.6	0.5	0.90	0.01	0.90
PCR 1 WE	4	5,393	77.4	22.1	0.5	0.90	0.01	0.90
PCR 2 WKL	3	5,387	79.7	19.8	0.5	0.87	0.01	0.87
PCR 2 WE	4	5,387	74.7	24.8	0.5	0.84	0.01	0.84

Table O.18. Operational Rater Agreement Statistics—ELA Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	5,200	79.3	19.9	0.8	0.89	0.00	0.89
PCR 1 WE	4	5,200	77.5	21.7	0.8	0.89	0.00	0.89
PCR 2 WKL	3	5,205	79.5	20.4	0.1	0.91	0.04	0.91
PCR 2 WE	4	5,205	80.0	20.0	0.0	0.92	0.01	0.92

Table O.19. Field Test Rater Agreement Statistics—ELA Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	3,082	88.1	11.9	0.0	0.81	0.02	0.81
PCR 1 WE	3	3,082	83.6	16.4	0.0	0.69	0.03	0.69
PCR 2 WKL	3	3,066	91.3	8.7	0.0	0.87	0.00	0.87
PCR 2 WE	3	3,066	80.6	19.3	0.0	0.71	0.01	0.71
PCR 3 WKL	3	3,062	85.2	14.5	0.3	0.82	0.01	0.82
PCR 3 WE	3	3,062	79.9	19.9	0.2	0.73	0.02	0.73
PCR 4 WKL	3	3,059	84.3	15.7	0.0	0.78	0.00	0.78
PCR 4 WE	3	3,059	83.8	16.2	0.0	0.66	0.02	0.66
PCR 5 WKL	3	3,073	84.8	15.2	0.0	0.79	0.02	0.79
PCR 5 WE	3	3,073	82.5	17.5	0.0	0.65	0.02	0.65
PCR 6 WKL	3	3,061	85.2	14.8	0.0	0.79	0.02	0.79
PCR 6 WE	3	3,061	79.6	20.4	0.0	0.63	0.02	0.63
PCR 7 WKL	3	3,000	85.7	14.3	0.0	0.78	0.01	0.78
PCR 7 WE	3	3,000	80.7	19.2	0.0	0.66	0.02	0.66
PCR 8 WKL	3	3,099	85.8	14.2	0.0	0.80	0.02	0.80
PCR 8 WE	3	3,099	80.3	19.6	0.1	0.65	0.01	0.65
PCR 9 WKL	3	3,075	83.4	16.6	0.0	0.76	0.00	0.76
PCR 9 WE	3	3,075	80.8	19.1	0.0	0.59	0.06	0.59

Table O.20. Field Test Rater Agreement Statistics—ELA Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	3,101	79.7	20.3	0.0	0.80	0.01	0.80
PCR 1 WE	4	3,101	78.9	21.0	0.1	0.78	0.03	0.78
PCR 2 WKL	3	3,085	80.2	19.5	0.3	0.76	0.01	0.76
PCR 2 WE	4	3,085	76.9	22.6	0.4	0.75	0.02	0.75
PCR 3 WKL	3	3,072	86.5	13.5	0.0	0.91	0.00	0.91
PCR 3 WE	3	3,072	87.4	12.6	0.0	0.91	0.01	0.91
PCR 4 WKL	3	2,999	85.8	14.2	0.0	0.88	0.01	0.89
PCR 4 WE	3	2,999	84.8	15.2	0.0	0.87	0.00	0.87
PCR 5 WKL	3	3,047	81.1	18.7	0.1	0.78	0.02	0.78
PCR 5 WE	4	3,047	73.3	26.7	0.0	0.73	0.04	0.73
PCR 6 WKL	3	3,040	83.0	17.0	0.0	0.80	0.02	0.80
PCR 6 WE	4	3,040	75.3	24.7	0.0	0.71	0.01	0.71
PCR 7 WKL	3	3,080	80.6	19.2	0.2	0.77	0.00	0.78
PCR 7 WE	4	3,080	73.2	26.8	0.0	0.72	0.01	0.72
PCR 8 WKL	3	3,060	80.3	19.7	0.0	0.78	0.01	0.78
PCR 8 WE	4	3,060	74.0	26.0	0.0	0.72	0.02	0.72
PCR 9 WKL	3	3,081	81.3	18.2	0.4	0.75	0.01	0.75
PCR 9 WE	4	3,081	75.0	23.7	1.3	0.64	0.02	0.64
PCR 10 WKL	3	3,042	82.6	17.4	0.0	0.77	0.03	0.77
PCR 10 WE	4	3,042	76.2	23.8	0.0	0.67	0.02	0.67

Table O.21. Field Test Rater Agreement Statistics—ELA Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	3,064	81.1	18.3	0.6	0.83	0.01	0.83
PCR 1 WE	4	3,064	78.4	20.8	0.8	0.82	0.02	0.82
PCR 2 WKL	3	2,993	80.0	19.7	0.4	0.82	0.07	0.82
PCR 2 WE	4	2,993	72.9	26.6	0.5	0.79	0.01	0.79
PCR 3 WKL	3	2,997	80.2	19.7	0.1	0.85	0.10	0.85
PCR 3 WE	4	2,997	70.2	29.7	0.1	0.79	0.05	0.79
PCR 4 WKL	3	2,996	80.9	19.0	0.1	0.83	0.01	0.83
PCR 4 WE	4	2,996	79.7	20.3	0.0	0.82	0.00	0.82
PCR 5 WKL	3	2,996	80.4	19.6	0.0	0.81	0.01	0.81
PCR 5 WE	4	2,996	74.3	25.7	0.0	0.78	0.01	0.78
PCR 6 WKL	3	3,108	79.4	20.5	0.1	0.84	0.02	0.84
PCR 6 WE	4	3,108	78.1	21.7	0.2	0.82	0.00	0.82

Table O.22. Field Test Rater Agreement Statistics—ELA Grade 6

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	2,985	80.0	19.3	0.7	0.85	0.00	0.85
PCR 1 WE	4	2,985	79.6	19.7	0.8	0.85	0.00	0.85
PCR 2 WKL	3	3,033	79.9	19.9	0.2	0.85	0.02	0.85
PCR 2 WE	4	3,033	79.8	19.8	0.4	0.85	0.01	0.85
PCR 3 WKL	3	2,978	81.3	18.4	0.2	0.84	0.03	0.84
PCR 3 WE	4	2,978	80.4	19.5	0.2	0.83	0.02	0.84
PCR 4 WKL	3	2,982	83.2	16.6	0.3	0.85	0.04	0.85
PCR 4 WE	4	2,982	79.9	19.8	0.2	0.83	0.01	0.83
PCR 5 WKL	3	2,980	80.0	19.9	0.2	0.80	0.03	0.80
PCR 5 WE	4	2,980	78.1	21.9	0.0	0.79	0.01	0.79
PCR 6 WKL	3	3,054	79.9	19.9	0.2	0.84	0.02	0.84
PCR 6 WE	4	3,054	82.8	17.1	0.1	0.87	0.02	0.87
PCR 7 WKL	3	3,021	80.2	19.7	0.1	0.88	0.00	0.88
PCR 7 WE	4	3,021	74.7	25.3	0.1	0.88	0.03	0.88
PCR 8 WKL	3	2,983	80.2	19.8	0.0	0.89	0.01	0.89
PCR 8 WE	4	2,983	78.0	21.9	0.0	0.89	0.03	0.89
PCR 9 WKL	3	2,992	80.1	19.9	0.0	0.88	0.01	0.88
PCR 9 WE	4	2,992	71.4	28.6	0.0	0.89	0.01	0.89
PCR 10 WKL	3	2,982	80.9	19.0	0.0	0.89	0.01	0.89
PCR 10 WE	4	2,982	76.1	23.9	0.0	0.90	0.00	0.90

Table O.23. Field Test Rater Agreement Statistics—ELA Grade 7

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	3,000	80.1	19.9	0.0	0.88	0.00	0.88
PCR 1 WE	4	3,000	76.9	23.1	0.0	0.87	0.00	0.87
PCR 2 WKL	3	3,000	79.9	19.3	0.8	0.88	0.01	0.88
PCR 2 WE	4	3,000	74.7	24.6	0.7	0.88	0.01	0.88
PCR 3 WKL	3	2,981	80.1	19.8	0.2	0.87	0.02	0.88
PCR 3 WE	4	2,981	74.3	25.4	0.3	0.87	0.01	0.87
PCR 4 WKL	3	3,081	81.3	18.3	0.4	0.89	0.01	0.89
PCR 4 WE	4	3,081	79.8	20.0	0.2	0.90	0.01	0.90
PCR 5 WKL	3	2,998	79.9	19.4	0.8	0.89	0.00	0.89
PCR 5 WE	4	2,998	80.4	19.3	0.3	0.91	0.00	0.91
PCR 6 WKL	3	2,994	79.9	18.9	1.1	0.89	0.01	0.89
PCR 6 WE	4	2,994	77.5	21.7	0.9	0.89	0.02	0.89
PCR 7 WKL	3	2,997	81.3	18.6	0.1	0.88	0.01	0.88
PCR 7 WE	4	2,997	81.8	18.1	0.1	0.90	0.01	0.90
PCR 8 WKL	3	2,996	80.0	19.9	0.1	0.89	0.09	0.89
PCR 8 WE	4	2,996	73.7	26.1	0.1	0.87	0.04	0.87
PCR 9 WKL	3	3,076	79.2	19.7	1.1	0.86	0.02	0.86
PCR 9 WE	4	3,076	81.4	17.9	0.7	0.89	0.02	0.89
PCR 10 WKL	3	2,999	79.9	19.8	0.3	0.86	0.01	0.86
PCR 10 WE	4	2,999	77.9	21.7	0.4	0.87	0.01	0.87
PCR 11 WKL	3	2,998	80.9	18.1	1.0	0.90	0.02	0.90

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 11 WE	4	2,998	79.4	19.7	1.0	0.91	0.00	0.91
PCR 12 WKL	3	2,993	80.5	19.3	0.3	0.89	0.06	0.89
PCR 12 WE	4	2,993	73.1	26.5	0.4	0.87	0.01	0.87

Table O.24. Field Test Rater Agreement Statistics—ELA Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WKL	3	3,077	81.1	18.5	0.4	0.91	0.00	0.91
PCR 1 WE	4	3,077	84.0	15.9	0.1	0.93	0.01	0.93
PCR 2 WKL	3	2,995	80.2	19.8	0.1	0.91	0.01	0.91
PCR 2 WE	4	2,995	83.5	16.5	0.0	0.93	0.01	0.93
PCR 3 WKL	3	3,123	83.1	16.2	0.7	0.91	0.01	0.91
PCR 3 WE	4	3,123	79.9	19.4	0.8	0.91	0.00	0.91
PCR 4 WKL	3	2,986	79.8	19.2	1.0	0.90	0.01	0.90
PCR 4 WE	4	2,986	77.0	22.1	0.9	0.91	0.00	0.91
PCR 5 WKL	3	2,986	81.7	18.0	0.3	0.91	0.02	0.91
PCR 5 WE	4	2,986	73.7	25.7	0.6	0.89	0.02	0.89
PCR 6 WKL	3	2,989	83.8	15.4	0.7	0.92	0.00	0.92
PCR 6 WE	4	2,989	79.6	19.5	0.9	0.91	0.00	0.91
PCR 7 WKL	3	3,035	80.0	19.5	0.5	0.90	0.02	0.90
PCR 7 WE	4	3,035	78.6	21.0	0.3	0.91	0.01	0.91
PCR 8 WKL	3	2,994	82.0	17.6	0.4	0.91	0.04	0.91
PCR 8 WE	4	2,994	74.9	25.0	0.1	0.90	0.01	0.90
PCR 9 WKL	3	2,987	81.0	18.9	0.0	0.92	0.08	0.92
PCR 9 WE	4	2,987	73.6	26.4	0.0	0.90	0.01	0.90
PCR 1 WKL	3	2,990	81.0	19.0	0.0	0.92	0.01	0.92
PCR 1 WE	4	2,990	80.4	19.5	0.1	0.92	0.01	0.92
PCR 2 WKL	3	2,980	81.9	18.1	0.0	0.91	0.01	0.91
PCR 2 WE	4	2,980	83.5	16.4	0.1	0.93	0.01	0.93
PCR 3 WKL	3	2,982	80.5	19.5	0.0	0.91	0.01	0.91
PCR 3 WE	4	2,982	85.6	14.3	0.1	0.94	0.00	0.94

Table O.25. Operational Rater Agreement Statistics—CSLA Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	147	86.8	13.2	0.0	0.90	0.02	0.90
PCR 1 WKL	3	147	89.7	10.3	0.0	0.92	0.02	0.92
PCR 2 WE	3	147	98.5	1.5	0.0	0.99	0.00	0.99
PCR 2 WKL	3	147	93.9	6.1	0.0	0.96	0.00	0.96

Table O.26. Operational Rater Agreement Statistics—CSLA Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	118	88.0	12.0	0.0	0.95	0.00	0.95
PCR 1 WKL	3	118	89.7	10.3	0.0	0.95	0.00	0.95
PCR 2 WE	4	119	84.7	15.3	0.0	0.94	0.07	0.94
PCR 2 WKL	3	119	91.9	8.1	0.0	0.96	0.04	0.96

Table O.27. Field Test Rater Agreement Statistics—CSLA Grade 3

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	3	360	96.4	3.6	0.0	0.97	0.02	0.97
PCR 1 WKL	3	360	90.8	9.2	0.0	0.94	0.06	0.95
PCR 2 WE	3	369	98.2	1.8	0.0	0.99	0.00	0.99
PCR 2 WKL	3	369	92.4	7.6	0.0	0.95	0.04	0.95
PCR 3 WE	3	364	94.7	5.3	0.0	0.97	0.01	0.97
PCR 3 WKL	3	364	94.4	5.6	0.0	0.96	0.00	0.96
PCR 4 WE	3	365	91.5	8.5	0.0	0.94	0.01	0.94
PCR 4 WKL	3	365	91.2	8.8	0.0	0.95	0.04	0.95

Table O.28. Field Test Rater Agreement Statistics—CSLA Grade 4

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
PCR 1 WE	4	588	91.1	8.9	0.0	0.96	0.01	0.96
PCR 1 WKL	3	588	94.8	5.2	0.0	0.97	0.01	0.97
PCR 2 WE	4	595	89.3	10.7	0.0	0.89	0.01	0.90
PCR 2 WKL	3	595	91.4	8.6	0.0	0.95	0.00	0.95

Table O.29. Operational Rater Agreement Statistics—Science Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	5,622	90.0	8.8	1.2	0.89	0.00	0.89
Item 2	2	5,623	89.9	9.7	0.3	0.91	0.00	0.91
Item 3	2	5,623	91.0	8.7	0.3	0.84	0.01	0.84
Item 4	2	5,363	90.0	8.3	1.7	0.85	0.01	0.85
Item 5	2	5,364	90.1	8.6	1.3	0.92	0.00	0.92
Item 6	2	5,361	90.1	9.0	0.9	0.87	0.00	0.87
Item 7	2	5,363	89.6	9.5	0.9	0.85	0.01	0.85
Item 8	2	5,363	89.9	9.0	1.1	0.90	0.01	0.90
Item 9	2	5,364	89.7	9.3	1.0	0.92	0.01	0.92
Item 10	2	5,623	90.2	9.3	0.5	0.91	0.00	0.91
Item 11	2	5,623	90.1	9.6	0.4	0.93	0.02	0.93
Item 12	2	5,623	89.9	9.8	0.3	0.90	0.00	0.90

Table O.30. Operational Rater Agreement Statistics—Science Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	5,122	90.1	9.9	0.0	0.88	0.03	0.88
Item 2	2	5,121	90.0	9.4	0.7	0.91	0.01	0.91
Item 3	2	5,122	90.3	9.1	0.7	0.77	0.02	0.77
Item 4	2	4,918	90.7	9.2	0.2	0.91	0.00	0.91
Item 5	2	4,918	90.2	8.5	1.3	0.90	0.00	0.90
Item 6	2	4,916	89.7	8.7	1.6	0.84	0.00	0.84
Item 7	2	4,915	87.1	12.4	0.6	0.83	0.01	0.83
Item 8	2	5,122	90.3	9.6	0.0	0.93	0.01	0.93
Item 9	2	5,122	89.9	9.9	0.2	0.93	0.01	0.93
Item 10	2	5,122	90.5	9.5	0.0	0.88	0.01	0.88
Item 11	2	5,122	90.2	9.4	0.4	0.82	0.00	0.82

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 12	2	4,916	96.2	2.9	0.9	0.94	0.01	0.94
Item 13	2	4,916	89.9	10.0	0.0	0.92	0.00	0.92
Item 14	2	4,916	90.0	10.0	0.0	0.89	0.00	0.89

Table O.31. Operational Rater Agreement Statistics—Science Grade 11

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	3,215	93.7	6.3	0.0	0.77	0.01	0.77
Item 2	2	3,215	89.2	10.3	0.5	0.91	0.00	0.91
Item 3	2	3,134	90.7	9.3	0.0	0.94	0.00	0.94
Item 4	2	3,133	89.9	9.8	0.3	0.89	0.01	0.89
Item 5	2	3,133	89.1	9.7	1.2	0.88	0.01	0.88
Item 6	2	3,215	89.7	9.9	0.4	0.90	0.01	0.90
Item 7	2	3,134	90.3	9.7	0.0	0.91	0.00	0.91
Item 8	2	3,133	89.4	9.6	0.9	0.91	0.00	0.91
Item 9	2	3,215	90.5	9.3	0.2	0.85	0.01	0.85
Item 10	2	3,215	89.9	9.1	0.9	0.88	0.01	0.88
Item 11	2	3,215	93.2	6.8	0.0	0.84	0.01	0.84
Item 12	2	3,215	89.6	9.5	0.9	0.66	0.04	0.67

Table O.32. Field Test Rater Agreement Statistics—Science Grade 5

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	3,244	91.1	8.2	0.8	0.87	0.01	0.87
Item 2	2	3,226	91.2	8.6	0.2	0.93	0.00	0.93
Item 3	2	3,519	92.0	8.0	0.0	0.85	0.01	0.85
Item 4	2	3,122	90.5	8.5	1.0	0.81	0.01	0.81
Item 5	2	3,469	89.8	10.0	0.1	0.88	0.01	0.88
Item 6	2	3,159	89.7	10.0	0.3	0.87	0.02	0.87
Item 7	2	3,338	93.2	6.6	0.2	0.95	0.01	0.95
Item 8	2	3,698	94.3	5.5	0.2	0.81	0.02	0.81
Item 9	2	3,391	88.9	10.0	1.1	0.79	0.01	0.79
Item 10	2	3,311	85.1	12.8	2.2	0.82	0.01	0.82
Item 11	2	3,089	90.2	9.7	0.1	0.92	0.01	0.92
Item 12	2	3,230	89.8	10.0	0.2	0.90	0.01	0.90
Item 13	2	3,048	90.0	9.8	0.2	0.92	0.01	0.92
Item 14	2	3,099	91.7	8.1	0.1	0.90	0.00	0.90
Item 15	2	3,113	89.9	10.0	0.1	0.82	0.00	0.82
Item 16	2	3,124	89.6	9.9	0.5	0.85	0.01	0.85
Item 17	2	3,081	90.2	9.1	0.7	0.90	0.00	0.90
Item 18	2	3,108	96.0	3.4	0.6	0.97	0.00	0.97
Item 19	2	3,085	90.1	9.4	0.5	0.92	0.01	0.92
Item 20	2	3,092	89.7	9.4	0.9	0.87	0.01	0.87
Item 21	2	3,059	89.9	6.8	3.3	0.83	0.03	0.84
Item 22	2	3,045	90.0	9.9	0.1	0.89	0.02	0.89
Item 23	2	3,043	90.2	9.6	0.2	0.87	0.02	0.87

Table O.33. Field Test Rater Agreement Statistics—Science Grade 8

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	3,149	96.0	4.0	0.0	0.96	0.00	0.96
Item 2	2	3,222	89.6	10.1	0.3	0.91	0.01	0.91
Item 3	2	3,195	89.7	8.3	2.0	0.89	0.02	0.89
Item 4	2	3,493	88.7	11.3	0.0	0.88	0.00	0.88
Item 5	2	3,660	90.6	9.3	0.1	0.91	0.00	0.91
Item 6	2	3,212	90.2	9.7	0.1	0.87	0.01	0.87
Item 7	2	3,435	89.8	9.6	0.6	0.89	0.00	0.89
Item 8	2	3,306	89.8	9.7	0.5	0.86	0.01	0.86
Item 9	2	3,307	89.2	10.6	0.2	0.86	0.03	0.86
Item 10	2	3,239	89.6	9.6	0.8	0.90	0.00	0.90
Item 11	2	3,233	89.5	7.4	3.1	0.76	0.01	0.76
Item 12	2	3,339	89.3	10.3	0.5	0.92	0.01	0.92
Item 13	2	3,226	89.3	10.7	0.0	0.91	0.05	0.91
Item 14	2	3,191	89.4	10.6	0.0	0.92	0.00	0.92
Item 15	2	3,444	92.2	6.9	0.9	0.90	0.00	0.90
Item 16	2	3,283	93.5	6.1	0.4	0.81	0.01	0.81
Item 17	2	3,207	89.5	9.8	0.8	0.86	0.01	0.86
Item 18	2	3,211	91.3	6.9	1.8	0.90	0.00	0.90
Item 19	2	3,193	89.6	9.9	0.5	0.86	0.01	0.86
Item 20	2	3,225	94.6	4.9	0.5	0.95	0.00	0.95
Item 21	2	3,148	90.0	9.1	0.9	0.91	0.01	0.91
Item 22	2	3,206	89.2	10.8	0.0	0.86	0.00	0.86
Item 23	2	3,187	89.3	10.7	0.0	0.88	0.01	0.88
Item 24	2	3,182	90.7	9.3	0.0	0.84	0.00	0.84
Item 25	2	3,204	90.0	9.5	0.6	0.90	0.00	0.90
Item 26	2	3,145	89.5	10.5	0.0	0.91	0.00	0.91
Item 27	2	3,176	89.4	10.5	0.1	0.90	0.01	0.90

Table O.34. Field Test Rater Agreement Statistics—Science Grade 11

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 1	2	3,083	92.4	7.6	0.0	0.88	0.01	0.88
Item 2	2	2,937	88.9	11.1	0.0	0.91	0.02	0.91
Item 3	2	2,948	88.7	11.3	0.0	0.87	0.00	0.87
Item 4	2	2,996	94.4	5.6	0.0	0.94	0.00	0.94
Item 5	2	2,914	89.1	10.9	0.0	0.90	0.01	0.90
Item 6	2	2,993	98.6	1.4	0.0	0.99	0.00	0.99
Item 7	2	2,994	97.1	2.9	0.0	0.96	0.01	0.96
Item 8	2	2,940	91.5	8.5	0.0	0.90	0.01	0.90
Item 9	2	2,968	92.8	7.2	0.0	0.90	0.01	0.90
Item 10	2	3,083	90.8	9.2	0.1	0.88	0.01	0.88
Item 11	2	3,049	89.3	10.0	0.7	0.89	0.01	0.89
Item 12	2	3,065	88.8	11.2	0.0	0.90	0.00	0.90
Item 13	2	3,325	89.0	11.0	0.0	0.89	0.00	0.89
Item 14	2	4,026	89.8	8.8	1.4	0.78	0.01	0.78

Item	Max. Points	N	%Exact	%Adjacent	%Non-Adjacent	Kappa	MD	Corr.
Item 15	2	3,291	91.5	8.4	0.1	0.88	0.01	0.88
Item 16	2	3,098	88.8	11.2	0.0	0.84	0.00	0.84
Item 17	2	3,011	78.0	21.7	0.2	0.83	0.01	0.83
Item 18	2	2,983	90.2	9.8	0.0	0.63	0.02	0.63
Item 19	2	3,018	93.0	7.0	0.0	0.65	0.01	0.65
Item 20	2	3,048	90.7	9.3	0.0	0.78	0.01	0.78
Item 21	2	2,935	83.1	16.9	0.0	0.82	0.01	0.82
Item 22	2	2,936	94.3	5.6	0.1	0.91	0.02	0.91
Item 23	2	2,968	91.0	8.9	0.1	0.76	0.02	0.76
Item 24	2	3,048	91.1	8.8	0.1	0.75	0.02	0.75
Item 25	2	3,081	97.9	2.1	0.0	0.98	0.01	0.98
Item 26	2	3,003	89.2	10.8	0.0	0.87	0.01	0.87
Item 27	2	3,002	89.3	10.7	0.0	0.79	0.01	0.79
Item 28	2	3,060	91.9	7.5	0.6	0.77	0.00	0.77