

Final Technical Report for 2016 Administration

Pearson

January 10, 2017



Table of Contents

List of Tables	vii
List of Figures	xiv
Executive Summary	1
Section 1: Introduction	7
1.1 Background	7
1.2 Purpose of the Operational Tests	8
1.3 Composition of Operational Tests	8
1.4 Intended Population	9
1.5 Groups and Organizations Involved with PARCC	9
1.6 Overview of the Technical Report.....	10
1.7 Glossary of Abbreviations	13
Section 2: Test Development.....	15
2.1 Overview of the PARCC Assessment, Claims, and Design.....	15
2.1.1 English Language Arts/Literacy (ELA/L) Assessments – Claims and Subclaims.....	15
2.1.2 Mathematics Assessments – Claims and Subclaims	16
2.2 Test Development Activities	17
2.2.1 Item Development Process	17
2.2.2 Item and Text Review Committees.....	18
2.2.3 Operational Test Construction.....	20
2.2.4 Linking Design of the Operational Test.....	24
2.2.5 Graphical Representation of PARCC Operational Test Linking Design	25
2.2.6 Field Test Data Collection Overview	27
Section 3: Test Administration	29
3.1 Testing Windows.....	29
3.2 Test Security and Administration Policies.....	29
3.3 Accessibility Features and Accommodations.....	31
3.3.1 Participation Guidelines for PARCC Assessments	31
3.3.2 PARCC Accessibility System.....	32
3.3.3 What are Accessibility Features?	32
3.3.4 Accommodations for Students with Disabilities and English Learners.....	32

3.3.5 Unique Accommodations.....	34
3.3.6 Emergency Accommodations	34
3.3.7 Student Refusal Form.....	35
3.4 Testing Irregularities and Security Breaches.....	35
3.5 Data Forensics Analyses.....	37
3.5.1 Response Change Analysis	37
3.5.2 Plagiarism Analysis.....	37
3.5.3 Internet and Social Media Monitoring.....	38
3.5.4 Off-Hours Testing Monitoring.....	38
Section 4: Item Scoring	39
4.1 Machine Scored Items	39
4.1.1 Key Based Items	39
4.1.2 Rule Based Items.....	39
4.2 Human or Handscored Items	40
4.2.1 Scorer Training	41
4.2.2 Scorer Qualification.....	44
4.2.3 Managing Scoring.....	45
4.2.4 Monitoring Scoring	45
4.3 Automated Scoring for PARCC PCRs	49
4.3.1 Changes to the ELA/L Scoring Rubric	49
4.3.2 Continuous Flow	49
4.3.3 Calibration of IEA Using Operational Data.....	49
4.3.4 Smart Routing	50
4.3.5 Quality Criteria for Evaluating Automated Scoring.....	50
4.3.6 Hierarchy of Assigned Scores for Reporting	50
4.3.7 Sampling Responses Used for Training IEA.....	51
4.3.8 Primary Criteria for Evaluating IEA Performance.....	51
4.3.9 Contingent Primary Criteria for Evaluating IEA Performance.....	52
4.3.10 Applying Smart Routing	52
4.3.11 Evaluation of Secondary Criteria for Evaluating IEA Performance	53
Section 5: Test Taker Characteristics.....	56
5.1 Overview of Test Taking Population	56

5.2 Composition of Operational Forms.....	56
5.3 Rules for Inclusion of Students in Analyses	57
5.4 Test Takers by Grade, Mode, and Gender	57
5.5 Demographics	59
Section 6: Classical Item Analysis	60
6.1 Overview	60
6.2 Data Screening Criteria	60
6.3 Description of Classical Item Analysis Statistics.....	60
6.4 Summary of Classical Item Analysis Flagging Criteria	62
6.5 Classical Item Analysis Results	63
Section 7: Differential Item Functioning.....	68
7.1 Overview	68
7.2 DIF Procedures	68
7.3 Operational Analysis DIF Comparison Groups.....	70
7.4 Operational Differential Item Functioning Results	72
Section 8: Reliability	75
8.1 Overview	75
8.2 Reliability and SEM Estimation	76
8.3 Reliability Results for Total Group	77
8.4 Reliability Results for Subgroups of Interest.....	79
8.5 Reliability Results for English Language Arts/Literacy Subscores	103
8.6 Reliability Results for Mathematics Subscores	106
8.7 Reliability of Classification	108
8.8 Inter-rater Agreement	113
Section 9: Validity	114
9.1 Overview	114
9.2 Evidence Based on Test Content.....	114
9.3 Evidence Based on Internal Structure.....	116
9.3.1 Intercorrelations	117
9.3.2 Reliability.....	130
9.3.3 Local Item Dependence	130
9.4 Evidence Based on Relationships to Other Variables	135

9.5 Evidence from the Special Studies	141
9.5.1 Content Alignment Studies	141
9.5.2 Benchmarking Study	142
9.5.3 Mode Comparability Study	143
9.5.4 Device Comparability Study	144
9.6 Evidence Based on Response Processes	144
9.7 Interpretations of Test Scores.....	145
9.8 Evidence Based on the Consequences to Testing.....	145
9.9 Summary	146
Section 10: IRT Calibration and Scaling in Operational Year Two	148
10.1 Overview	148
10.2 IRT Data Preparation.....	148
10.2.1 Overview	148
10.2.2 Student Inclusion/Exclusion Rules	149
10.2.3 Items Excluded from IRT Sparse Matrices	150
10.2.4 Omitted, Not Reached and Not Presented Items	150
10.2.5 Quality Control of the IRT Sparse Matrix Data Files	150
10.3 Description of the Calibration Process.....	151
10.3.1 Two-Parameter Logistic/Generalized Partial Credit Model.....	151
10.3.2 Treatment of Prose Constructed Response (PCR) Tasks.....	152
10.3.3 IRT Item Exclusion Rules (Before Calibration).....	152
10.3.4 IRTPRO Calibration Procedures and Convergence Criteria.....	153
10.3.5 Calibration Quality Control	154
10.4 Model Fit Evaluation Criteria	155
10.5 Items Excluded from Score Reporting	158
10.5.1 Item Review Process	158
10.5.2 Count and Percentage of Items Excluded from Score Reporting	159
10.6 Scaling Parameter Estimates.....	162
10.6.1 Scaling Procedures (Year-to-Year)	162
10.6.2 Scaling Procedures (PBT to CBT)	162
10.6.3 Scaling Procedures (Traditional-to-Integrated Mathematics)	163
10.6.4 Comparability across Spanish and English Versions	163

10.6.5 Scaling Quality Control.....	164
10.7 Items Excluded from Linking Sets	164
10.8 Correlations and Plots of Scaling Item Parameter Estimates	167
10.9 Scaling Constants	174
10.10 Summary Statistics and Distributions from IRT Analyses	175
10.10.1 IRT Summary Statistics for English Language Arts/Literacy.....	176
10.10.2 IRT Summary Statistics for Mathematics.....	178
Section 11: Performance Level Setting.....	182
11.1 Performance Standards	182
11.2 Performance Levels and Policy Definitions.....	182
11.3 Performance Level Setting Process for the PARCC Assessment System	184
11.3.1 PARCC Research Studies	185
11.3.2 PARCC Pre-Policy Meeting	185
11.3.3 Performance Level Setting Meetings.....	185
11.3.4 PARCC Post-Policy Reasonableness Review.....	187
Section 12: Scale Scores.....	188
12.1 Operational Test Content (Claims and Subclaims)	188
12.1.1 English Language Arts/Literacy.....	188
12.1.2 Mathematics	190
12.2 Establishing the Reporting Scales	190
12.2.1 Full Summative Score Scale and Performance Levels.....	191
12.2.2 ELA/L Reading and Writing Claim Scale	192
12.2.3 Subclaims Scale	193
12.3 Creating Conversion Tables	193
12.4 Score Distributions.....	197
12.4.1 Score Distributions for ELA/L	197
12.4.2 Score Distributions for Mathematics.....	206
12.5 Interpreting Claim Scores and Subclaim Scores.....	212
12.5.1 Interpreting Claim Scores.....	212
12.5.2 Interpreting Subclaim Scores	212
Section 13: Quality Control Procedures	214
13.1 Quality Control of the Item Bank	214

13.2 Quality Control of Test Form Development.....	215
13.3 Quality Control of Test Materials.....	216
13.4 Quality Control of Scanning	217
13.5 Quality Control of Image Editing.....	217
13.6 Quality Control of Answer Document Processing and Scoring.....	218
13.7 Quality Control of Psychometric Processes	219
13.7.1 Pearson Psychometric Quality Control Process	219
13.7.2 HumRRO Psychometric Quality Control Process	221
13.7.3 Measured Progress Psychometric Quality Control Process.....	222
References	224
Appendices	227
Appendix 5: Test Takers by Grade and Mode, for Each State	227
Appendix 7: Summary of Differential Item Function (DIF) Results.....	260
Appendix 8: Reliability of Classification by Content and Grade Level	282
Appendix 10.1: IRT Results for Spring 2016 English Language Arts/Literacy (ELA/L)	303
Appendix 10.2: IRT Results for Spring 2016 Mathematics.....	314
Appendix 12.1: Form Composition	335
Appendix 12.2: Scaling Constants and Associated Information	341
Appendix 12.3: Raw-to-Scale Conversion Tables for Performance Level Setting (PLS) Forms.....	346
Appendix 12.4: IRT Test Characteristic Curves, Information Curves and CSEM Curves	409
Appendix 12.5: Subgroup Scale Score Performance.....	430
Addendum: Statistical Summary of the Fall/Winter Block 2015 Administration	460
Addendum 5: Test Taker Characteristics	460
Addendum 8: Reliability.....	470
Addendum 9: Validity	485
Addendum 12: Scale Scores.....	490
Addendum 13: Inter-rater Agreement for Prose Constructed Response.....	499

List of Tables

Table 1.1 Defined Summative Scale Scores and Cut Scores	5
Table 1.2 Defined Scaled Scores and Cut Scores for Reading and Writing Claim Scores	6
Table 1.3 Glossary of PARCC Abbreviations and Acronyms.....	13
Table 2.1 Number of Core Operational Forms per Grade/Subject and Mode for ELA/L and Mathematics	21
Table 3.1 PARCC Fall/Winter Block 2015 and Spring 2016 Testing Windows	29
Table 4.1 Training Materials Used During Scoring.....	43
Table 4.2 Mathematics Qualification Requirements	45
Table 4.3 Scoring Validity Agreement Requirements	47
Table 4.4 Inter-rater Agreement Expectations and Results.....	48
Table 5.1 ELA/L Test Takers by Grade and Mode: All States Combined	58
Table 5.2 Mathematics Test Takers by Grade and Mode: All States Combined	58
Table 5.3 Spanish-Language Mathematics Test Takers, by Grade and Mode: All States Combined.....	59
Table 6.1 Summary of p Values for ELA/L Operational Items by Grade and Mode.....	64
Table 6.2 Summary of p Values for Mathematics Operational Items by Grade and Mode	65
Table 6.3 Summary of Item-total Polyserial Correlations for ELA/L Operational Items by Grade and Mode	66
Table 7.1 DIF Categories for Dichotomous Selected Response and Constructed Response Items	70
Table 7.2 DIF Categories for Polytomous Constructed Response Items.....	70
Table 7.3 Traditional DIF Comparison Groups	71
Table 8.1 Summary of ELA/L Test Reliability Estimates for Total Group	78
Table 8.2 Summary of Mathematics Test Reliability Estimates for Total Group	79
Table 8.3 Summary of Test Reliability Estimates for Subgroups: Grade 3 ELA/L.....	82
Table 8.4 Summary of Test Reliability Estimates for Subgroups: Grade 4 ELA/L.....	83
Table 8.5 Summary of Test Reliability Estimates for Subgroups: Grade 5 ELA/L.....	84
Table 8.6 Summary of Test Reliability Estimates for Subgroups: Grade 6 ELA/L.....	85
Table 8.7 Summary of Test Reliability Estimates for Subgroups: Grade 7 ELA/L.....	86
Table 8.8 Summary of Test Reliability Estimates for Subgroups: Grade 8 ELA/L.....	87
Table 8.9 Summary of Test Reliability Estimates for Subgroups: Grade 9 ELA/L.....	88
Table 8.10 Summary of Test Reliability Estimates for Subgroups: Grade 10 ELA/L.....	89
Table 8.11 Summary of Test Reliability Estimates for Subgroups: Grade 11 ELA/L.....	90
Table 8.12 Summary of Test Reliability Estimates for Subgroups: Grade 3 Mathematics	91
Table 8.13 Summary of Test Reliability Estimates for Subgroups: Grade 4 Mathematics	92
Table 8.14 Summary of Test Reliability Estimates for Subgroups: Grade 5 Mathematics	93
Table 8.15 Summary of Test Reliability Estimates for Subgroups: Grade 6 Mathematics	94
Table 8.16 Summary of Test Reliability Estimates for Subgroups: Grade 7 Mathematics	95
Table 8.17 Summary of Test Reliability Estimates for Subgroups: Grade 8 Mathematics	96
Table 8.18 Summary of Test Reliability Estimates for Subgroups: Algebra I	97
Table 8.19 Summary of Test Reliability Estimates for Subgroups: Geometry	98
Table 8.20 Summary of Test Reliability Estimates for Subgroups: Algebra II	99

Table 8.21 Summary of Test Reliability Estimates for Subgroups: Integrated Mathematics I	100
Table 8.22 Summary of Test Reliability Estimates for Subgroups: Integrated Mathematics II	101
Table 8.23 Summary of Test Reliability Estimates for Subgroups: Integrated Mathematics III	102
Table 8.24 Descriptions of ELA/L Claims and Subclaims.....	103
Table 8.25 Average ELA/L Reliability Estimates for Total Test and Subscores	105
Table 8.26 Average Mathematics Reliability Estimates for Total Test and Subscores	107
Table 8.27 Reliability of Classification: Summary for ELA/L.....	109
Table 8.28 Reliability of Classification: Grade 3 ELA/L.....	110
Table 8.29 Reliability of Classification: Summary for Mathematics	112
Table 8.30 Inter-rater Agreement Expectations and Results.....	113
Table 9.1 Average Intercorrelations and Reliability between Grade 3 ELA/L Subclaims.....	119
Table 9.2 Average Intercorrelations and Reliability between Grade 4 ELA/L Subclaims.....	119
Table 9.3 Average Intercorrelations and Reliability between Grade 5 ELA/L Subclaims.....	120
Table 9.4 Average Intercorrelations and Reliability between Grade 6 ELA/L Subclaims.....	120
Table 9.5 Average Intercorrelations and Reliability between Grade 7 ELA/L Subclaims.....	121
Table 9.6 Average Intercorrelations and Reliability between Grade 8 ELA/L Subclaims.....	121
Table 9.7 Average Intercorrelations and Reliability between Grade 9 ELA/L Subclaims.....	122
Table 9.8 Average Intercorrelations and Reliability between Grade 10 ELA/L Subclaims.....	122
Table 9.9 Average Intercorrelations and Reliability between Grade 11 ELA/L Subclaims.....	123
Table 9.10 Average Intercorrelations and Reliability between Grade 3 Mathematics Subclaims	123
Table 9.11 Average Intercorrelations and Reliability between Grade 4 Mathematics Subclaims	124
Table 9.12 Average Intercorrelations and Reliability between Grade 5 Mathematics Subclaims	124
Table 9.13 Average Intercorrelations and Reliability between Grade 6 Mathematics Subclaims	125
Table 9.14 Average Intercorrelations and Reliability between Grade 7 Mathematics Subclaims	125
Table 9.15 Average Intercorrelations and Reliability between Grade 8 Mathematics Subclaims	126
Table 9.16 Average Intercorrelations and Reliability between Algebra I Subclaims	126
Table 9.17 Average Intercorrelations and Reliability between Geometry Subclaims	127
Table 9.18 Average Intercorrelations and Reliability between Algebra II Subclaims	127
Table 9.19 Average Intercorrelations and Reliability between Integrated Mathematics I Subclaims	128
Table 9.20 Average Intercorrelations and Reliability between Integrated Mathematics II Subclaims	128
Table 9.21 Average Intercorrelations and Reliability between Integrated Mathematics III Subclaims	129
Table 9.22 Conditions used in LID Reliability Investigation and Results.....	133
Table 9.23 Summary of Q_3 Values for ELA/L Grade 4 and Integrated Mathematics II (Spring 2015).....	133
Table 9.24 Correlations between ELA/L and Mathematics for Grade 3	136
Table 9.25 Correlations between ELA/L and Mathematics for Grade 4	136
Table 9.26 Correlations between ELA/L and Mathematics for Grade 5	137
Table 9.27 Correlations between ELA/L and Mathematics for Grade 6	137
Table 9.28 Correlations between ELA/L and Mathematics for Grade 7	138
Table 9.29 Correlations between ELA/L and Mathematics for Grade 8	138
Table 9.30 Correlations between ELA/L and Mathematics for High School	139
Table 9.31 Correlations between ELA/L Reading and Mathematics for High School.....	139
Table 10.1 N-Counts, Percent of Students, and Number of Items in the ELA/L IRT Calibration Files	151

Table 10.2 N-Counts, Percent of Students, and Number of Items in the Mathematics IRT Calibration Files	151
Table 10.3 Number and Percentage of ELA/L Items Excluded from IRT Calibration	160
Table 10.4 Number and Percentage of Mathematics Items Excluded from IRT Calibration	161
Table 10.5 WRMSD Flagging Criteria for Inspection and Possible Removal of Linking Items	165
Table 10.6 Number of ELA/L Items Excluded from the Year-to-Year Linking Sets.....	166
Table 10.7 Number of Mathematics Items Excluded from the Year-to-Year Linking Sets	166
Table 10.8 Number of ELA/L Items Excluded from the CBT/PBT Linking Sets	167
Table 10.10 Number of Items, Number of Points and Correlations for ELA/L Year-to-Year Linking Items	168
Table 10.11 Number of Items, Number of Points and Correlations for Mathematics Year-to-Year Linking Items	168
Table 10.12 Number of Items, Number of Points and Correlations for ELA/L CBT/PBT Linking Items	169
Table 10.13 Number of Items, Number of Points and Correlations for Mathematics CBT/PBT Linking Items	169
Table 10.14 Scaling Constants Spring 2015 to Spring 2016 for ELA/L	174
Table 10.15 Scaling Constants Spring 2015 to Spring 2016 for Mathematics	174
Table 10.16 Scaling Constants Spring 2016 PBT to CBT for ELA/L	175
Table 10.18 CBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade.....	176
Table 10.19 PBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade.....	176
Table 10.20 CBT IRT Standard Errors of Parameter Estimates for All Items for ELA/L by Grade	177
Table 10.21 PBT IRT Standard Errors of Parameter Estimates for All Items for ELA/L by Grade.....	177
Table 10.22 CBT IRT Model Fit for All Items for ELA/L by Grade	178
Table 10.23 PBT IRT Model Fit for All Items for ELA/L by Grade	178
Table 10.24 CBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject ..	179
Table 10.25 PBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject ..	179
Table 10.26 CBT IRT Standard Errors of Parameter Estimates for All Items for Mathematics by Grade/Subject	180
Table 10.27 PBT IRT Standard Errors of Parameter Estimates for All Items for Mathematics by Grade/Subject	180
Table 10.28 CBT IRT Model Fit for All Items for Mathematics by Grade/Subject.....	181
Table 10.29 PBT IRT Model Fit for All Items for Mathematics by Grade/Subject.....	181
Table 11.1 PARCC PLS Committee Meetings and Dates	187
Table 12.1 Form Composition for ELA/L Grade 3	189
Table 12.2 Contribution of Prose Constructed Response Items to ELA/L	190
Table 12.3 Mathematics Form Composition for Grade 3	190
Table 12.4 Defined Summative Scale Scores	191
Table 12.5 Defined Scaled Scores for Reading and Writing Claim Scores	192
Table 12.6 Calculating Scaling Constants for Reading and Writing Claim Scores.....	193
Table 12.7 Subgroup Performance for ELA/L: Grade 3.....	203
Table 12.8 Subgroup Performance for ELA/L: Grade 9.....	205
Table 12.9 Subgroup Performance for Mathematics Scale Scores: Grade 3.....	209

Table 12.10 Subgroup Performance for Mathematics Scale Scores: Algebra I	210
Table 12.11 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics I.....	212
Table A.5.1 ELA/L Test Takers, by State, and Grade	227
Table A.5.2 Mathematics Test Takers, by State, and Grade	230
Table A.5.3 Spanish-Language Mathematics Test Takers, by State, and Grade	233
Table A.5.4 All States Combined: ELA/L Test Takers by Grade, Mode, and Gender.....	236
Table A.5.5 All States Combined: All Mathematics Test Takers by Grade, Mode, and Gender	237
Table A.5.6 All States Combined: Spanish-Language Mathematics Test Takers, by Grade, Mode, and Gender	238
Table A.5.7 Demographic Information for Grade 3 ELA/L, Overall and by State.....	239
Table A.5.8 Demographic Information for Grade 4 ELA/L, Overall and by State.....	240
Table A.5.9 Demographic Information for Grade 5 ELA/L, Overall and by State.....	241
Table A.5.10 Demographic Information for Grade 6 ELA/L, Overall and by State.....	242
Table A.5.11 Demographic Information for Grade 7 ELA/L, Overall and by State.....	243
Table A.5.12 Demographic Information for Grade 8 ELA/L, Overall and by State.....	244
Table A.5.13 Demographic Information for Grade 9 ELA/L, Overall and by State.....	245
Table A.5.14 Demographic Information for Grade 10 ELA/L, Overall and by State.....	246
Table A.5.15 Demographic Information for Grade 11 ELA/L, Overall and by State.....	247
Table A.5.16 Demographic Information for Grade 3 Mathematics, Overall and by State	248
Table A.5.17 Demographic Information for Grade 4 Mathematics, Overall and by State	249
Table A.5.18 Demographic Information for Grade 5 Mathematics, Overall and by State	250
Table A.5.19 Demographic Information for Grade 6 Mathematics, Overall and by State	251
Table A.5.20 Demographic Information for Grade 7 Mathematics, Overall and by State	252
Table A.5.21 Demographic Information for Grade 8 Mathematics, Overall and by State	253
Table A.5.22 Demographic Information for Algebra I, Overall and by State	254
Table A.5.23 Demographic Information for Geometry, Overall and by State	255
Table A.5.24 Demographic Information for Algebra II, Overall and by State	256
Table A.5.25 Demographic Information for Integrated Mathematics I, Overall and by State	257
Table A.5.26 Demographic Information for Integrated Mathematics II, Overall and by State	258
Table A.5.27 Demographic Information for Integrated Mathematics III, Overall and by State	259
Table A.7.1 Differential Item Functioning for ELA/L Grade 3	260
Table A.7.2 Differential Item Functioning for ELA/L Grade 4	261
Table A.7.3 Differential Item Functioning for ELA/L Grade 5	262
Table A.7.4 Differential Item Functioning for ELA/L Grade 6	263
Table A.7.5 Differential Item Functioning for ELA/L Grade 7	264
Table A.7.6 Differential Item Functioning for ELA/L Grade 8	265
Table A.7.7 Differential Item Functioning for ELA/L Grade 9	266
Table A.7.8 Differential Item Functioning for ELA/L Grade 10	267
Table A.7.9 Differential Item Functioning for ELA/L Grade 11	268
Table A.7.10 Differential Item Functioning for Mathematics Grade 3	269
Table A.7.11 Differential Item Functioning for Mathematics Grade 4	270
Table A.7.12 Differential Item Functioning for Mathematics Grade 5	271

Table A.7.13 Differential Item Functioning for Mathematics Grade 6	272
Table A.7.14 Differential Item Functioning for Mathematics Grade 7	273
Table A.7.15 Differential Item Functioning for Mathematics Grade 8	274
Table A.7.16 Differential Item Functioning for Algebra I	275
Table A.7.17 Differential Item Functioning for Geometry	276
Table A.7.18 Differential Item Functioning for Algebra II	277
Table A.7.19 Differential Item Functioning for Integrated Mathematics I	278
Table A.7.20 Differential Item Functioning for Integrated Mathematics II	279
Table A.7.21 Differential Item Functioning for Integrated Mathematics III	280
Table A.7.22 Differential Item Functioning for Mathematics: Spanish-Language vs. English-Language Forms	281
Table A.8.1 Reliability of Classification: Grade 3 ELA/L	282
Table A.8.2 Reliability of Classification: Grade 4 ELA/L	283
Table A.8.3 Reliability of Classification: Grade 5 ELA/L	284
Table A.8.4 Reliability of Classification: Grade 6 ELA/L	285
Table A.8.5 Reliability of Classification: Grade 7 ELA/L	286
Table A.8.6 Reliability of Classification: Grade 8 ELA/L	287
Table A.8.7 Reliability of Classification: Grade 9 ELA/L	288
Table A.8.8 Reliability of Classification: Grade 10 ELA/L	289
Table A.8.9 Reliability of Classification: Grade 11 ELA/L	290
Table A.8.10 Reliability of Classification: Grade 3 Mathematics	291
Table A.8.11 Reliability of Classification: Grade 4 Mathematics	292
Table A.8.12 Reliability of Classification: Grade 5 Mathematics	293
Table A.8.13 Reliability of Classification: Grade 6 Mathematics	294
Table A.8.14 Reliability of Classification: Grade 7 Mathematics	295
Table A.8.15 Reliability of Classification: Grade 8 Mathematics	296
Table A.8.16 Reliability of Classification: Algebra I	297
Table A.8.17 Reliability of Classification: Geometry	298
Table A.8.18 Reliability of Classification: Algebra II	299
Table A.8.19 Reliability of Classification: Integrated Mathematics I	300
Table A.8.20 Reliability of Classification: Integrated Mathematics II	301
Table A.8.21 Reliability of Classification: Integrated Mathematics III	302
Table A.10.1 CBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade	303
Table A.10.2 PBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade	305
Table A.10.3 CBT IRT Standard Errors of Parameter Estimates for All Items for ELA/L by Grade	307
Table A.10.4 PBT IRT Standard Errors of Parameter Estimates for All Items for ELA/L by Grade	309
Table A.10.5 CBT IRT Model Fit for All Items for ELA/L by Grade	311
Table A.10.6 PBT IRT Model Fit for All Items for ELA/L by Grade	312
Table A.10.7 CBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject.	314
Table A.10.8 PBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject.	318
Table A.10.9 CBT IRT Standard Errors of Parameter Estimates for All Items for Mathematics by Grade/Subject	321

Table A.10.10 PBT IRT Standard Errors of Parameter Estimates for All Items for Mathematics by Grade/Subject	325
Table A.10.11 CBT IRT Model Fit for All Items for Mathematics by Grade/Subject	328
Table A.10.12 PBT IRT Model Fit for All Items for Mathematics by Grade/Subject	332
Table A.12.1 Form Composition for ELA/L Grade 3	335
Table A.12.2 Form Composition for ELA/L Grade 4	335
Table A.12.3 Form Composition for ELA/L Grade 5	335
Table A.12.4 Form Composition for ELA/L Grade 6	336
Table A.12.5 Form Composition for ELA/L Grade 7	336
Table A.12.6 Form Composition for ELA/L Grade 8	336
Table A.12.7 Form Composition for ELA/L Grade 9	337
Table A.12.8 Form Composition for ELA/L Grade 10	337
Table A.12.9 Form Composition for ELA/L Grade 11	337
Table A.12.10 Form Composition for Mathematics Grade 3	338
Table A.12.11 Form Composition for Mathematics Grade 4	338
Table A.12.12 Form Composition for Mathematics Grade 5	338
Table A.12.13 Form Composition for Mathematics Grade 6	338
Table A.12.14 Form Composition for Mathematics Grade 7	339
Table A.12.15 Form Composition for Mathematics Grade 8	339
Table A.12.16 Form Composition for Algebra I	339
Table A.12.17 Form Composition for Geometry	339
Table A.12.18 Form Composition for Algebra II	340
Table A.12.19 Form Composition for Integrated Mathematics I	340
Table A.12.20 Form Composition for Integrated Mathematics II	340
Table A.12.21 Form Composition for Integrated Mathematics III	340
Table A.12.22 Threshold Scores and Scaling Constants for ELA/L Grades 3 to 8	341
Table A.12.23 Threshold Scores and Scaling Constants for Mathematics Grades 3 to 8	342
Table A.12.25 Threshold Scores and Scaling Constants for High School Mathematics	344
Table A.12.26 Scaling Constants for Reading and Writing Grades 3 to 8	345
Table A.12.27 Conversion Table for Performance Level Setting Form: ELA/L Grade 3	346
Table A.12.28 Conversion Table for Performance Level Setting Form: ELA/L Grade 4	349
Table A.12.29 Conversion Table for Performance Level Setting Form: ELA/L Grade 5	352
Table A.12.30 Conversion Table for Performance Level Setting Form: ELA/L Grade 6	355
Table A.12.31 Conversion Table for Performance Level Setting Form: ELA/L Grade 7	359
Table A.12.32 Conversion Table for Performance Level Setting Form: ELA/L Grade 8	363
Table A.12.33 Conversion Table for Performance Level Setting Form: ELA/L Grade 9	367
Table A.12.34 Conversion Table for Performance Level Setting Form: ELA/L Grade 10	371
Table A.12.35 Conversion Table for Performance Level Setting Form: ELA/L Grade 11	375
Table A.12.36 Conversion Table for Performance Level Setting Form: Mathematics Grade 3	379
Table A.12.37 Conversion Table for Performance Level Setting Form: Mathematics Grade 4	381
Table A.12.38 Conversion Table for Performance Level Setting Form: Mathematics Grade 5	383
Table A.12.39 Conversion Table for Performance Level Setting Form: Mathematics Grade 6	385

Table A.12.40 Conversion Table for Performance Level Setting Form: Mathematics Grade 7	387
Table A.12.41 Conversion Table for Performance Level Setting Form: Mathematics Grade 8	389
Table A.12.42 Conversion Table for Performance Level Setting Form: Algebra I	391
Table A.12.43 Conversion Table for Performance Level Setting Form: Algebra II	394
Table A.12.44 Conversion Table for Performance Level Setting Form: Geometry	397
Table A.12.45 Conversion Table for Performance Level Setting Form: Integrated Mathematics I	400
Table A.12.46 Conversion Table for Performance Level Setting Form: Integrated Mathematics II	403
Table A.12.47 Conversion Table for Performance Level Setting Form: Integrated Mathematics III	406
Table A.12.48 Subgroup Performance for ELA/L Scale Scores: Grade 3	430
Table A.12.49 Subgroup Performance for ELA/L Scale Scores: Grade 4	432
Table A.12.50 Subgroup Performance for ELA/L Scale Scores: Grade 5	434
Table A.12.51 Subgroup Performance for ELA/L Scale Scores: Grade 6	436
Table A.12.52 Subgroup Performance for ELA/L Scale Scores: Grade 7	438
Table A.12.53 Subgroup Performance for ELA/L Scale Scores: Grade 8	440
Table A.12.54 Subgroup Performance for ELA/L Scale Scores: Grade 9	442
Table A.12.55 Subgroup Performance for ELA/L Scale Scores: Grade 10	444
Table A.12.56 Subgroup Performance for ELA/L Scale Scores: Grade 11	446
Table A.12.57 Subgroup Performance for Mathematics Scale Scores: Grade 3	448
Table A.12.58 Subgroup Performance for Mathematics Scale Scores: Grade 4	449
Table A.12.59 Subgroup Performance for Mathematics Scale Scores: Grade 5	450
Table A.12.60 Subgroup Performance for Mathematics Scale Scores: Grade 6	451
Table A.12.61 Subgroup Performance for Mathematics Scale Scores: Grade 7	452
Table A.12.62 Subgroup Performance for Mathematics Scale Scores: Grade 8	453
Table A.12.63 Subgroup Performance for Mathematics Scale Scores: Algebra I	454
Table A.12.64 Subgroup Performance for Mathematics Scale Scores: Geometry	455
Table A.12.65 Subgroup Performance for Mathematics Scale Scores: Algebra II	456
Table A.12.66 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics I	457
Table A.12.67 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics II	458
Table A.12.68 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics III	459

List of Figures

Figure 2.1 PARCC ELA/L CBT and PBT Linking Design (Grades 6-11)	26
Figure 2.2 PARCC Mathematics CBT and PBT Linking Design (Grades 3-6)	27
Figure 3.2 Claims from the Administration Phase of the PARCC Theory of Action.	Error! Bookmark not defined.
Figure 9.1 Comparison of Score Reliability by Item Reliability and Cluster (Testlet) Reliability Approaches	132
Figure 9.2 Distribution of Q_3 Values for Grade 4 ELA/L (Spring 2015).....	134
Figure 9.3 Distribution of Q_3 Values for Integrated Mathematics II (Spring 2015).....	134
Figure 10.1 An example ELA/L 5-Category Item, 2 PL/GPC Model, N-count 44,658, $Q_1=1266.64$, $ZQ_1=147.21$ and a criterion $ZQ_{1,crit}= 237.02$	158
Figure 10.2 An example ELA/L 3-Category Item, 2 PL/GPC Model, N-count 175,839	159
Figure 10.3 ELA/L Grade 8 Transformed New a - vs. Reference a -Parameter Estimates for Year-to-Year Linking	170
Figure 10.4 ELA/L Grade 8 Transformed New b - vs. Reference b -Parameter Estimates for Year-to-Year Linking	170
Figure 10.5 Mathematics Grade 5 Transformed New a - vs. Reference a -Parameter Estimates for Year-to-Year Linking	171
Figure 10.6 Mathematics Grade 5 Transformed New b - vs. Reference b -Parameter Estimates for Year-to-Year Linking	171
Figure 10.7 ELA/L Grade 4 Transformed New a - vs. Reference a -Parameter Estimates for PBT to CBT Linking	172
Figure 10.8 ELA/L Grade 4 Transformed New b - vs. Reference b -Parameter Estimates for PBT to CBT Linking	172
Figure 10.9 Mathematics Grade 8 Transformed New a - vs. Reference a -Parameter Estimates for PBT to CBT Linking	173
Figure 10.10 Mathematics Grade 8 Transformed New b - vs. Reference b -Parameter Estimates for PBT to CBT Linking	173
Figure 12.1 Test Characteristic Curves, Information Curves, and Conditional Standard Error of Measurement Curve for ELA/L Grade 3	196
Figure 12.4 Distributions of ELA/L Scale Scores: Grades 3 to 11	197
Figure 12.4 (continued) Distributions of ELA/L Scale Scores: Grades 3 to 11	198
Figure 12.5 Distributions of Reading Scale Scores: Grades 3 to 11	199
Figure 12.5 (continued) Distributions of Reading Scale Scores: Grades 3 to 11.....	200
Figure 12.6 Distributions of Writing Scale Scores: Grades 3 to 11	Error! Bookmark not defined.
Figure 12.6 (continued) Distributions of Writing Scale Scores: Grades 3 to 11.....	202
Figure 12.7 Distributions of Mathematics Scale Scores: Grades 3 to 8	207
Figure 12.8 Distributions of Mathematics Scale Scores: High School.....	209
Figure A.12.1 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 3	409
Figure A.12.2 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 4	410
Figure A.12.3 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 5	411

Figure A.12.4 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 6	412
Figure A.12.5 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 7	413
Figure A.12.6 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 8	414
Figure A.12.7 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 9	415
Figure A.12.8 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 10 ...	416
Figure A.12.9 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 11 ...	417
Figure A.12.10 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 3	418
Figure A.12.11 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 4	419
Figure A.12.12 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 5	420
Figure A.12.13 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 6	421
Figure A.12.14 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 7	422
Figure A.12.15 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 8	423
Figure A.12.16 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Algebra I	424
Figure A.12.17 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Geometry	425
Figure A.12.18 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Algebra II	426
Figure A.12.19 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Integrated Mathematics I	427
Figure A.12.20 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Integrated Mathematics II	428
Figure A.12.21 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Integrated Mathematics III	429

Executive Summary

The Partnership for Assessment of Readiness for College and Careers (PARCC) is a state-led consortium creating next-generation assessments that, compared to traditional K-12 assessments, more accurately measure student progress toward college and career readiness. The PARCC assessments are aligned to the Common Core State Standards (CCSS) and were administered operationally for the first time in the 2014-2015 academic year. PARCC comprises assessments in both English language arts/literacy (ELA/L) and Mathematics in grades 3 to 8 and high school.

The information provided in this technical report is intended for use by those who evaluate tests, interpret scores, or use test results in making educational decisions. It is assumed that the reader has technical knowledge of test construction and measurement procedures, as stated in *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). The purpose of this technical report is to describe the second operational administration of the PARCC assessments in the 2015-2016 academic year and includes the following topics:

- Background and purpose of the assessments,
- Test development of items and forms,
- Test administration, security, and scoring,
- Test taker characteristics,
- Classical item analyses and differential item functioning,
- Reliability and validity of scores,
- Item response theory (IRT) calibration and scaling,
- Performance level setting,
- Development of the score reporting scales and student performance, and
- Quality control procedures.

Background and Purpose

Assessments for the first operational administration were constructed in 2014. Eleven states and the District of Columbia participated in the first administration of the PARCC assessments during the 2014-2015 school year. A small subset of students were tested in fall 2014. ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered in the fall; these assessments were administered on paper only. The majority of students tested during the spring 2015 window when all grades and content areas were administered online and on paper. Seven states, the Bureau of Indian Education, and District of Columbia participated in the second administration in school year 2015-2016. Not all participating states had students testing in all grades. In fall 2015 ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered online and on paper. The majority of students tested during the spring 2016 window when all grades and content areas were administered online and on paper.

The PARCC assessments are designed to achieve several purposes. First, the tests are intended to provide evidence to determine whether students are on track for college- and career-readiness. Second, the tests are structured to access the full range of CCSS and measure the total breadth of student performance. Finally, the tests are designed to provide data to help inform classroom instruction, student interventions and professional development.

The fall 2015 operational administration of the PARCC assessment included two separate components: the Performance-Based Assessment (PBA) and the End-of-Year (EOY) assessment. Both components were administered as computer-based tests (CBT) and as paper-based tests (PBT). A valid score in both the PBA and EOY assessments was required for a student to receive a summative score. The spring 2016 operational administration of the PARCC assessment combined the Performance-Based Assessment (PBA) and End-of-Year (EOY) into one testing window.

Item Types

The tests contain selected response, brief and extended constructed response, technology-enabled, and technology-enhanced items (TEI), as well as performance tasks. Technology-enabled items are single-response or constructed-response items that involve some type of digital stimulus or open-ended response box with which the students engage in answering questions. Technology-enhanced items involve specialized student interactions for collecting performance data. Therefore, the act of performing the task is the way in which data are collected. Students may be asked, among other tasks, to categorize information, organize or classify data, order a series of events, plot data, generate equations, highlight text, or fill in a blank. One example of a TEI is an interaction in which students are asked to drag response options onto a Venn diagram to show the relationship among ideas.

Classical and IRT Item Analysis

Classical item analyses and differential item functioning analyses were performed on the data to evaluate the psychometric characteristics of the operational test items after items were administered and before scores were reported. The two-parameter logistic/generalized partial credit (2PL/GPC) IRT models were used for calibrations and scaling. Multiple operational core forms were administered for each grade in ELA/L and mathematics. The forms included sets of embedded common items to provide data to support horizontal linking across test forms within a grade and content area and across years.

The purpose of the IRT calibration and scaling was to place all operational items for a single grade/subject onto a common scale. The results of the 2014 field test dimensionality study indicated that multidimensional models, based on predetermined test structures (e.g., PBA versus EOY, and ELA/reading versus ELA/L writing), did not provide significantly better model fit compared to a unidimensional model, for both ELA/L and mathematics. A mode comparability study based on the 2014 field test data did not provide evidence to assume that scores resulting from PBT and CBT forms were strictly comparable between modes, particularly for PBA. Based on the findings from these two studies, the operational data were calibrated concurrently across forms, and calibrations were conducted separately for PBT and CBT response data using IRT models consistent with mixed format data.

After calibration and model fit evaluation was completed, a master list of all items flagged as problematic was compiled and brought to the PARCC Priority Alert Task Force.¹ The Task Force reviewed each item, its content and the statistical properties, and made decisions about whether to include the item in the operational scores. Sometimes, an item was rejected because it appeared to have content issues, and sometimes an item was excluded because it could not be calibrated or showed extremely poor IRT model fit. Ultimately the decision about whether to keep or exclude each flagged item was made by the Task Force. The goals of the Task Force were to: a) minimize the number of items excluded from the operational test forms, and b) avoid advantaging or disadvantaging any test takers.

Once the item response data from the computer-based tests (CBT) and the paper-based tests (PBT) were calibrated for all grades and content areas, all available item parameter estimates of common items across modes, were used to transform the PBT item parameter estimates onto the CBT scales. The software program STUIRT (Kim & Kolen, 2004) was used to obtain Stocking and Lord (1983) transformation values to link the PBT scales to the CBT scales.

The PBT forms for all grades and content areas were generated using items from the CBT forms. In response to several practical constraints based on the number of forms constructed for each mode and to meet the blueprints (e.g., inclusion of TEI on CBT forms), there was no single CBT form that was administered intact in the paper delivery mode at any grade level. For example, TEI from online forms were replaced in the paper forms with items having similar content, but appropriate for paper-based testing. However, for both ELA/L and mathematics, the content on PBT forms significantly overlapped content on the CBT forms. A mode comparability study was conducted in 2015 and the results are presented as a separate special report. The study evaluated the extent to which scores from CBT and PBT forms could be considered as comparable with regard to psychometric characteristics. A major finding was that score comparability was inconsistent across the content domains and grade levels investigated.

Overall Scale Scores, Claim Scores, and Subclaim Scores

The PARCC ELA/L and mathematics scores are expressed as various types of scale scores (both total scores and claim scores, related to the claims structures described below), as well as by performance levels used to describe how well students meet the academic standards for their grade level. On the basis of a student's total score, an inference is drawn about how much knowledge and skill in the content area the student has acquired. The total score is also used to classify students in terms of the level of knowledge and skill in the content area as students progress in their K-12 education. These levels are called performance levels and are reported as:

- Level 5: Exceeded expectations
- Level 4: Met expectations
- Level 3: Approached expectations
- Level 2: Partially met expectations
- Level 1: Did not yet meet expectations

¹ The Priority Alert Task Force comprised Parcc Inc. staff, state leads, and state staff.

Students classified as either Level 4 or Level 5 are meeting or exceeding the grade level expectations. Additionally, information on more specific skills is provided and is reported as *Below Expectations*, *Nearly Meets Expectations*, and *Meets or Exceeds Expectations*.

PARCC has developed performance level descriptors (PLDs) to assist with the understanding and interpretations of the ELA/L and mathematics scores (<http://www.parcconline.org/assessments/test-design/ela-literacy/ela-performance-level-descriptors> and <http://www.parcconline.org/assessments/test-design/mathematics/math-performance-level-descriptors>). Additionally, resource information is available online to educators, parents, and students (<http://avocet.pearson.com/PARCC/Home#15727>), which includes information on understanding and interpreting the ELA/L and mathematics score reports.

The claim structures for ELA/L and for mathematics, grounded in the Common Core State Standards, informs the design and development of the summative assessments.

Claim Structure for ELA/L

Master Claim. The master claim is the overall performance goal for the PARCC ELA/L Assessment System—students must demonstrate that they are college- and career-ready or on track to readiness as demonstrated through reading and comprehending of grade-level texts of appropriate complexity and writing effectively when using and/or analyzing sources.

Major Claims: 1) reading and comprehending a range of sufficiently complex texts independently, and 2) writing effectively when using and/or analyzing sources.

Subclaims: The subclaims further explicate what is measured on the PARCC assessments and include claims about student performance on the standards and evidences outlined in the PARCC evidence tables for reading and writing (<http://www.parcconline.org/assessments/test-design/ela-literacy/test-specifications-documents>). The claims and evidences are grouped into the following categories.

1. Vocabulary, Interpretation, and Use
2. Reading Literature
3. Reading Informational Text
4. Written Expression
5. Knowledge of Language and Conventions

Claim Structure for Mathematics

Master Claim. The degree to which a student is college- or career-ready or on track to being ready in mathematics. The student solves grade-level/course-level problems aligned to the Standards for Mathematical Content with connections to the Standards for Mathematical Practice.

Subclaims: The subclaims further explicate what is measured on the PARCC assessments and include claims about student performance on the standards and evidences outlined in the PARCC evidence statement tables for mathematics

(<http://www.parcconline.org/assessments/test-design/mathematics/math-test-specifications-documents>). The claims and evidence are grouped into the following categories.

Subclaim A: Major Content with Connections to Practices

Subclaim B: Additional and Supporting Content with Connections to Practices

Subclaim C: Highlighted Practices with Connections to Content: Expressing Mathematical Reasoning by constructing viable arguments, critiquing the reasoning of others, and/or attending to precision when making mathematical statements

Subclaim D: Highlighted Practice with Connections to Content: Modeling/Application by solving real-world problems by applying knowledge and skills articulated in the standards.

Score Scales

Scale scores were defined for each test as a linear transformation of the IRT theta (θ) scale. The test characteristic curves associated with the performance level setting forms were used to identify the theta values associated with the Level 2 and Level 4 point scores. By defining Level 2 and 4 scale scores to be 700 and 750, respectively, the linear relationship between theta and scale scores was established.

The result is 201 defined full summative scale score points for each ELA/L and mathematics assessment, ranging from 650 to 850. A scale score of 700 is always the minimum for Level 2 performance, a scale score of 750 is always the minimum for Level 4 performance.

The thresholds for summative performance levels on the scale score metric recommended by the scale score task force are described in Table 1.1.

Table 1.1 Defined Summative Scale Scores and Cut Scores

	Lowest Obtainable Scale Score	Cut Score Level 2	Cut Score Level 4	Highest Obtainable Scale Score
Full Summative	650	700	750	850

As with the full summative scores, scale scores for Reading and Writing were defined for each test as a linear transformation of the IRT theta (θ) scale. The same IRT theta scale was used for Reading and Writing as was used for the ELA/L full summative scores. The theta values associated with the Level 2 and Level 4 performance levels were identified using the test characteristic curves associated with the performance level setting forms. Parallel to the full summative scores, the relationship between theta and scale scores was established with Level 2 and 4 theta scores and the corresponding predefined scale scores.

The result was 81 defined scale score points for Reading, ranging from 10 to 90. A scale score of 30 is the cut score for minimum Level 2 performance, a scale score of 50 is the cut score for minimum Level 4 performance. There are 51 defined scale score points for Writing, ranging from 10 to 60. A scale score of 25 is the cut score for minimum Level 2 performance, a scale score of 35 is the cut score for minimum Level 4 performance. The threshold Reading and Writing performance levels on the scale score metric recommended by the scale score task force are described in Table 1.2.

Table 1.2 Defined Scaled Scores and Cut Scores for Reading and Writing Claim Scores

	Lowest Obtainable Scale Score	Cut Score Level 2	Cut Score Level 4	Highest Obtainable Scale Score
Reading	10	30	50	90
Writing	10	25	35	60

Regarding the subclaim scores, the Level 4 cut is defined as *Meets or Exceeds Expectations* because grade 3-8, and high school students at Level 4 or above are likely to have the skills and knowledge to meet the definition of career and college readiness. Subclaim outcomes center on that performance level and are reported at three levels:

- *Below Expectations;*
- *Nearly Meets Expectations; or*
- *Meets or Exceeds Expectations.*

Quality Control

To ensure IRT calibrations, scaling and conversion tables were produced accurately, HumRRO replicated the data processing, IRT calibrations and scale score transformations carried out by Pearson, and the generation of the score conversion tables. Pearson and HumRRO independently generated incomplete data matrices and conducted the calibrations using IRTPRO (Cai, Thissen & du Toit, 2011) calibration software. Pearson and HumRRO both used STUIRT software to transform 2015 item parameter estimates to the 2016 IRT scale and to transform PBT item parameter estimates onto the CBT scales for each grade/subject. Pearson's scaling constants were compared to those generated by HumRRO and found to be consistent. Measured Progress (MP) performed independent quality control comparisons between the Pearson and HumRRO item parameter estimates to identify any differences. In addition, MP independently made certain that the same items were excluded from the linking sets, and compared transformed parameter estimates computed by Pearson and HumRRO. If items had large differences across years or modes, the items were discussed and any remaining issues resolved. Measured Progress prepared reports documenting their findings. Exact matches were found between all Pearson and HumRRO conversion tables before scores were reported.

Section 1: Introduction

1.1 Background

States associated with the Partnership for Assessment of Readiness for College and Careers (PARCC) came together in early 2010 with a shared vision of ensuring that all students - regardless of income, family background or geography - have equal access to a world-class education that will prepare them for success after high school in college and/or careers. The PARCC goal was to develop new assessments that tie into more rigorous academic expectations and help prepare students for success in college and the workforce, as well as to provide information back to teachers and parents about where students are on their path to success. Calling on the expertise of thousands of teachers, higher education faculty and other educators in multiple states, the PARCC assessment system is a high quality set of summative assessments, diagnostic assessments, formative tasks, and other support materials for teachers including professional development and communications tools.

The PARCC consortium develops and administers next-generation assessments that, compared to traditional K-12 assessments, more accurately measure student progress toward college and career readiness. The assessments are aligned to the Common Core State Standards (CCSS) and include both English language arts/literacy (ELA/L) assessments (grades three through eleven) and mathematics assessments (grades three through eight, and high school). Compared to traditional standardized tests, these assessments are intended to measure more complex skills like critical-thinking, persuasive writing, and problem-solving.

In 2013, the PARCC Governing Board launched Parcc Inc., a non-profit organization designed to support the successful delivery of the tests in 2014-15, and the long-term success of the multi-state partnership. States continue to govern decisions about the assessment system; the non-profit organization is their “agent” for overseeing the many vendors involved in the PARCC assessment system, coordinating the multiple work groups and committees (including Governing Board meetings), managing the PARCC intellectual property, overseeing the research agenda and the Technical Advisory Committee, and developing and launching the multiple non-summative tools.

Summative assessments for the first operational administration were constructed in 2014. Eleven states and the District of Columbia participated in the first administration of the PARCC assessments during the 2014-2015 school year. A small subset of students tested in Fall 2014. ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered in the fall; these assessments were administered on paper only. The majority of students tested during the Spring 2015 window when all grades and content areas were administered online and on paper. Seven states, the Bureau of Indian Education, and District of Columbia participated in the second administration in school year 2015-2016. Not all participating states had students testing in all grades. In fall 2015 ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II were administered online and paper. The majority of students tested during the spring 2016 window when all grades and content areas were administered online and on paper.

The purpose of this technical report is to describe the second operational administration of the PARCC summative assessments in the 2015-2016 academic year, including test form construction, test administration, item scoring, test taker characteristics, classical item analysis results, reliability results, evidence of validity, item response theory (IRT) calibrations and scaling, performance level setting procedure, and quality control procedures.

1.2 Purpose of the Operational Tests

The PARCC assessments are designed to achieve several purposes. First, the tests are intended to provide evidence to determine whether students are on track for college- and career-readiness. Second, the tests are structured to access the full range of CCSS and measure the total breadth of student performance. Finally, the tests are designed to provide data to help inform classroom instruction, student interventions, and professional development.

1.3 Composition of Operational Tests

Each operational test form was constructed to reflect the full test blueprint in terms of content, standards measured, and item types. Sets of common items, included to provide data to support horizontal linking across test forms within a grade and content area, were proportionally representative of the operational test blueprint.

The fall 2015 operational administration of the PARCC assessment included two separate components: the Performance-Based Assessment (PBA) and the End-of-Year (EOY) assessment. Both components were administered as computer-based tests (CBT) and as paper-based tests (PBT). A valid score in both the PBA and EOY assessments was required for a student to receive a summative score. The spring 2016 operational administration of the PARCC assessment combined the Performance-Based Assessment (PBA) and End-of-Year (EOY) into one testing window.

The Fall PBA and EOY components utilized somewhat different item types. The PBA was administered after approximately 75 percent of instructional time was complete. The purpose of the PBA component was to measure critical thinking, reasoning, and the ability to apply skills and knowledge in reading, writing, and mathematics. The ELA/L PBA component comprised three types of tasks: literary analysis, narrative writing, and research simulation. For each task, students were instructed to read one or more texts, answer several brief questions, and then write an essay based on the material they read. The mathematics PBA consisted of tasks designed to assess a student's ability to use mathematics to solve real-life problems. Some of the tasks required that students describe how they solved a problem, while other tasks measured conceptual understanding and ability to apply concepts by means of selected-response or technology-enhanced items.

The Fall EOY administration occurred after approximately 90 percent of instruction was complete. Students were required to demonstrate their skills and knowledge by answering innovative selected-response and short-answer questions that measured concepts and skills. The ELA/L EOY assessment had between two and four literary and informational texts; each text had five or six brief comprehension and vocabulary

questions. The mathematics EOY assessment contained tasks that measured a combination of conceptual understanding, applications, skills, and procedures.

The spring 2016 PARCC assessments were administered in either computer-based or paper-based format. PBA and EOY components were combined into one testing window in an effort to shorten the test, and make it easier for schools to administer and for students to take. English language arts/literacy (ELA/L) assessments focused on writing effectively when analyzing text. Mathematics assessments focused on applying skills and concepts, and understanding multi-step problems that require abstract reasoning and modeling real-world problems, precision, perseverance, and strategic use of tools. In both content areas, students also demonstrated their acquired skills and knowledge by answering selected response items and fill-in-the-blank questions.

Each spring assessment was comprised of multiple units, and additionally, one of the mathematics units was split into two sections: a non-calculator section and a calculator section.

1.4 Intended Population

The PARCC tests are intended for students taking ELA/L and/or mathematics in grades 3 through 11, as well as students taking high school mathematics (i.e., Algebra I, Geometry, Algebra II, Integrated Mathematics I – III). For these students, the PARCC tests measured whether students were meeting state academic standards and mastering the knowledge and skills needed to progress in their K-12 education and beyond.

1.5 Groups and Organizations Involved with PARCC

- Parcc Inc. is a nonprofit organization that assumes the responsibility for management of the PARCC consortium, as well as the development and implementation of PARCC assessments.

A number of committees of educators, state education agency staff, and national experts lead the work of the PARCC consortium. These committees include:

- the PARCC consortium Governing Board that makes major policy and operational decisions,
- the Technical Advisory Committee that helps ensure all assessments will provide reliable results to inform valid instructional and accountability decisions,
- the K-12 State Leads that coordinates all aspects of development of the PARCC assessment system and serves as the conduit to the Technical Advisory Committee and the Governing Board,
- the Advisory Committee on College Readiness which includes higher education executive officers from PARCC states and other state- and nationally-recognized leaders in the postsecondary community, and

- the Higher Education Leadership Team which is responsible for coordinating higher education engagement in the PARCC assessment system and works closely with the Advisory Committee on College Readiness.
- Test and item development activities were conducted by Pearson and WestEd under the guidance and oversight of PARCC leadership.²
- Pearson served as the primary contractor for the PARCC operational administration and was responsible for developing test forms, production of all testing materials, packaging and distribution, receiving and scanning of materials, and scoring, as well as program management and customer service.
- Pearson Psychometrics was responsible for all psychometric analyses of the PARCC operational test data. This included classical item analyses, differential item functioning (DIF) analyses, item calibrations based on item response theory (IRT), scaling, and development of all conversion tables.
- HumRRO served as a subcontractor and was responsible for replicating item calibrations based on item response theory (IRT), scaling, and development of all conversion tables.
- Measured Progress (MP) served as a subcontractor to conduct external evaluations; they were responsible for reviewing and comparing the psychometric IRT calibrations performed by Pearson, which were replicated by HumRRO. MP also provided comparisons of results obtained independently from Pearson and from HumRRO for raw-to-theta (RST) conversion tables, summative and claim scale scores, performance level classifications, and subclaim performance level classifications.

1.6 Overview of the Technical Report

This report begins by providing explanations of the test form construction process, test administration, and scoring of the test items. Subsequent sections of the report present descriptions of test taker characteristics, results of classical item analyses, results of reliability analyses, evidence of validity, item response theory (IRT) calibrations and scaling, performance level setting procedure, and quality control procedures.

The technical report contains the following sections:

- Section 2 – Test Development

² PARCC leadership includes the following groups: PARCC Governing Board, K-12 State Leads, Higher Education Leadership Team, Technical Advisory Committee, Operational Working Group members from each of the member states, and staff members from Parcc, Inc., the project management partner for the PARCC Consortium.

This section describes the PARCC test design and the procedures followed during the development of operational test forms.

- Section 3 – Test Administration

This section presents the operational administration schedule, information regarding test security and confidentiality, accessibility features and accommodations, and testing irregularities and security breaches.

- Section 4 – Scoring of the Items

The key-based and rule-based processes for machine scored items, as well as the training and monitoring processes for human scored items are provided in this section.

- Section 5 – Test Taker Characteristics

This section describes the composition of test forms, rules for inclusion of students in analyses, distributions of test takers by grade, mode, and gender, and distributions of demographic variables of interest.

- Section 6 – Classical Item Analyses

The classical item-level statistics calculated for the operational test data, the flagging criteria used to identify items that performed differently than expected, and the results of these analyses are presented in this section.

- Section 7 – Differential Item Functioning

In this section, the methods for conducting differential item functioning analyses as well as corresponding flagging criteria are described. This is followed by definitions of the comparison groups and subsequent results for the comparison groups.

- Section 8 - Reliability

The results of internal consistency reliability analyses and corresponding standard errors of measurement, for each grade, content area, and mode (CBT or PBT) for all test takers, and for subgroups of interest, is provided in this section. This is followed by reliability results for subscores and reliability of classification (i.e., decision accuracy and decision consistency). Finally, expectations and results for interrater agreement for hand scored items are summarized.

- Section 9 – Validity

Validity evidence based on analyses of the internal structure of the tests is provided in this section. Correlations between subscores are reported by grade, content area, and mode (CBT or PBT) for all test takers.

- Section 10 - IRT Calibration and Scaling

This section presents the information related to the calibration and scaling of item response data including: data preparation, the calibration process, model fit evaluation, and items excluded from score reporting. In addition, the scaling process (paper to online) is described and evaluated.

- Section 11 – Performance Level Setting (PLS) Procedure and Results

Performance levels and policy definitions, as well as the processes followed to establish performance level thresholds are described in this section.

- Section 12 - Scale Scores

This section provides an overview of the claims and subclaims, describes the development of the reporting scales and conversion tables, and presents scale score distributions. Finally, information regarding the interpretation of claim scores and subclaim scores is presented.

- Section 13 – Quality Control Procedures

All aspects of quality control are presented in this section. These activities range from quality assurance of item banking, test form construction, and all testing materials to quality control of scanning, image editing, and scoring. This is followed by a detailed description of the steps taken to ensure that all psychometric analyses were of the highest quality.

- References

- Appendices

To facilitate utility, tables in the appendices are numbered sequentially according to the section represented by the tables. For example, the first appendix table for Section 5 is numbered A.5.1, the second appendix table for Section 5 is numbered A.5.2, and so on.

- Addendum

The addendum presents the results of analyses for the Fall 2015 operational administration. These results are reported separately from the Spring 2016 results because fall testing involved a nonrepresentative subset of students testing only ELA/L grades 9, 10, and 11, as well as Algebra I, Geometry, and Algebra II.

To organize the addendum, tables are numbered sequentially according to the section represented by the tables. For example, the first addendum table for Section 5 is numbered ADD.5.1, the second addendum table for section 5 is numbered ADD.5.2, and so on.

1.7 Glossary of Abbreviations

Table 1.3 Glossary of PARCC Abbreviations and Acronyms

Abbreviation/Acronym	Definition
1PL/PC	One-parameter/Partial Credit Model
2PL/GPC	Two-parameter Logistic/Generalized Partial Credit Model
3PL/GPC	Three-parameter Logistic/Generalized Partial Credit Model
AAF	Accessibility, Accommodations, and Fairness
ABBI	Assessment Banking for Building and Interoperability
AERA	American Educational Research Association
AIS	Average Item Score
AIQ	Assessment and Information Quality
APA	American Psychological Association
ASC	Additional and Supporting Content (Mathematics)
ASL	American Sign Language
ATA	Automatic Test Assembler
CBT	Computer-Based Test
CCSS	Common Core State Standards
CDQ	Customer Data Quality
CSEM	Conditional Standard Error of Measurement
DIF	Differential Item Functioning
DPL	Digital Production Line
DPP	Digital Pre-press
EBSS	Evidence-based Standard Setting
ELA/L	English Language Arts/Literacy
EL	English Learners
EOC	End-of-Course
EOY	End-of-Year
ePEN2	Electronic Performance Evaluation Network
ESEA	Elementary and Secondary Education Act
FRL	Free or Reduced-price Lunch
FS	Full Summative
FT	Field Test
IA	Item Analysis
ICC	Item Characteristic Curve
IDEA	Individuals with Disabilities Education Act
IEP	Individualized Education Program
INF	Information Curve
IRA	Inter-rater Agreement
IRF	Item Response File
IRT	Item Response Theory
IRS	Individual Student Report
K-12	Kindergarten to Grade 12
LEA	Local Education Agency
LID	Local Item Dependence
MAD	Mean Absolute Difference
MC	Major Content (Mathematics)
MH	Mantel-Haenszel
MP	Measured Progress

MP	Modeling Practice (Mathematics)
MR	Mathematical Reasoning
NAEP	National Assessment of Educational Progress
NCLB	No Child Left Behind
NCME	National Council on Measurement in Education
NSLP	National School Lunch Program
OE responses	Open-ended responses
OMR	Optical Mark Reading
OWG	Operational Working Group
PARCC	Partnership for Assessment of Readiness for College and Careers
PBA	Performance-Based Assessment
PBT	Paper-Based Test
PCR	Prose Constructed Response (ELA/L)
PEJ	Postsecondary Educators' Judgment
PLD	Performance Level Descriptor
PLS	Performance Level Setting
PV	Product Validation
QA	Quality Assurance
RD	Reading (ELA/L)
RI	Reading Information (ELA/L)
RL	Reading Literature (ELA/L)
RMSD	Root Mean Square Difference
RV	Reading Vocabulary (ELA/L)
RST	Raw-score-to-theta
SD	Standard Deviation
SDF	Student Data File
SE	Standard Error
SEJ	Standard Error of Judgment
SEM	Standard Error of Measurement
SIRB	Scored Item Response Block
SMD	Standardized Mean Difference
SSMC	Single Select Multiple Choice
SWD	Students with Disabilities
TCC	Test Characteristic Curve
TTS	Text to Speech
UIN	Unique Item Number
WE	Writing Written Expression (ELA/L)
WKL	Writing Knowledge Language and Conventions (ELA/L)
WLS	Weighted Least Squares
WR	Writing (ELA/L)
WRMSD	Weighted Root Mean Square Difference

Section 2: Test Development

2.1 Overview of the PARCC Assessment, Claims, and Design

Aligned to the Common Core State Standards (CCSS) as articulated in the PARCC Model Content Frameworks, the PARCC assessments are designed to determine whether students are college- and career-ready or on track, assess the full range of the CCSS, measure the full range of student performance, and provide data to help inform instruction, interventions, and professional development. Test development is an ongoing process involving educators, researchers, psychometricians, subject matter professionals, and assessment experts who participate in the development of the PARCC test design and its underlying foundational documents; develop and review passages and items used to build the PARCC assessments; monitor the program for quality, accessibility, and fairness for all students; and construct, review, and score the assessments

The PARCC summative assessments include both English language arts/literacy (ELA/L) and mathematics assessments in grades 3 to 8 and high school. The high school mathematics tests include traditional mathematics and integrated mathematics course pathways. Tests contain selected response, brief and extended constructed response, technology-enabled and technology-enhanced items (TEI), as well as performance tasks. Technology-enabled items are single-response or constructed-response items that involve some type of digital stimulus or open-ended response box with which the students engage in answering questions. Technology-enhanced items involve specialized student interactions for collecting performance data. In other words, the act of performing the task is the way in which data is collected. Students may be asked, among other interactions, to categorize information, organize or classify data, order a series of events, plot data, generate equations, highlight text, or fill in a blank. One example of a TEI is an interaction in which students are asked to drag response options onto a Venn diagram to show the relationship among ideas.

The PARCC assessments offer a wide range of accessibility features for all students and accommodations for students with disabilities (e.g., screen reader, assistive technology, braille, large print, text-to-speech, and ASL video versions of the test, as well as response accommodations that allow students to respond to test items using different formats). For English learners who are native Spanish speakers, PARCC offers a paper-based edition of the mathematics assessment in Spanish, and both large print and Text-to-Speech versions of the test in Spanish (refer to the PARCC Accessibility Features and Accommodations Manual for in-depth information).

2.1.1 English Language Arts/Literacy (ELA/L) Assessments – Claims and Subclaims

The ELA/L summative assessment at each grade level consists of three task types: Literary Analysis, Research Simulation, and Narrative Writing. For each performance-based task, students are asked to read or view one or more texts, answer comprehension and vocabulary questions, and write an extended response that requires them to draw evidence from text(s). The summative assessment also contains literary and informational reading passages with comprehension and vocabulary questions.

The claim structure, grounded in the CCSS, undergirds the design and development of the ELA/L summative assessments.

Master Claim. The master claim is the overall performance goal for the PARCC ELA/Literacy Assessment System—students must demonstrate that they are college- and career-ready or on track to readiness as demonstrated through reading and comprehending of grade-level texts of appropriate complexity and writing effectively when using and/or analyzing sources.

Major Claims: 1) reading and comprehending a range of sufficiently complex texts independently, and 2) writing effectively when using and/or analyzing sources.

Sub Claims: The sub claims further explicate what is measured on the PARCC assessments and include claims about student performance on the standards and evidences outlined in the PARCC evidence tables for reading and writing (refer to *PARCC Test Specifications Documents*). The claims and evidences are grouped into the following categories.

1. Vocabulary Interpretation and Use
2. Reading Literature
3. Reading Informational Text
4. Written Expression
5. Knowledge of Language and Conventions

2.1.2 Mathematics Assessments – Claims and Subclaims

The summative mathematics assessment at each grade level includes both short- and extended-response questions focused on applying skills and concepts to solve problems that require demonstration of the mathematical practices from the Common Core State Standards with a focus on modeling and reasoning with precision. The assessments also include performance-based short-answer questions focused on conceptual understanding, procedural skills, and application.

The claim structure, grounded in the CCSS, undergirds the design and development of the summative assessments.

Master Claim. The degree to which a student is college- or career-ready or on track to being ready in mathematics. The student solves grade-level/course-level problems aligned to the Standards for Mathematical Content with connections to the Standards for Mathematical Practice.

Sub Claims: The sub claims further explicate what is measured on the PARCC assessments and include claims about student performance on the standards and evidences outlined in the PARCC evidence statement tables for mathematics (refer to *PARCC Test Specifications Documents*). The claims and evidence are grouped into the following categories.

Subclaim A: Major Content with Connections to Practices.

Subclaim B: Additional and Supporting Content with Connections to Practices.

Subclaim C: Highlighted Practices with Connections to Content: Expressing Mathematical Reasoning by constructing viable arguments, critiquing the reasoning of others, and/or attending to precision when making mathematical statements.

Subclaim D: Highlighted Practice with Connections to Content: Modeling/Application by solving real-world problems by applying knowledge and skills articulated in the standards.

2.2 Test Development Activities

Test development activities began with the standards and model content frameworks. From these, PARCC, in collaboration with more than 2,000 educators, researchers, and psychometricians, has developed the PARCC test specifications documents that guide the development of test items and the composition of the tests. These documents include the College- and Career-Ready determinations and Performance-Level Descriptions, Claim Structure, Evidence Statement Tables, Blueprints, Informational Guides, Passage Selection Guidelines, Mathematics Sequencing Guidelines, Task Generation Models, Fairness and Sensitivity Guidelines, Text Selection Guidelines, and the Style Guide. Refer to the PARCC website for further information about these documents.

2.2.1 Item Development Process

PARCC test and item development activities were conducted by Pearson and WestEd under the guidance and oversight of PARCC leadership, including the PARCC Governing Board, the K-12 State Leads, the Higher Education Leadership Team, the Technical Advisory Committee, the Operational Working Group members from each of the member states, the PARCC State Text and Content Review Committees, and staff members from Parcc, Inc., the project management partner for the PARCC Consortium.

Developing high quality assessment content with authentic stimuli for computer-based tests (CBT) and paper-based tests (PBT) measuring rigorous standards is a complex process involving the services of many experts including assessment designers, psychometricians, managers, trainers, content providers, content experts, editors, artists, programmers, technicians, human scorers, advisors, and members of the PARCC Operational Working Groups.

Bank Analysis and Item Development Plan

The PARCC summative item bank houses passages and items at each assessed grade level and subject. The bank supports the administration of the assessments, along with item release and practice tests. Items are developed and field tested annually. Prior to the annual item development cycle, the item development teams, in conjunction with members of the Operational Working Groups (OWGs) for ELA/L and mathematics, evaluated the strengths of the bank and considered the needs for future tests to establish an item development plan.

Text Selection for ELA/L

Using the PARCC Passage Selection Guidelines, English language arts subject matter experts were trained to search for appropriate passages to support an annual pool of passages for consideration. Guided by the PARCC test specifications documents, Pearson and WestEd recruited, trained, and managed the contracted subject matter experts to deliver the number of texts specified in the annual item development plan. The Passage Selection Guidelines provided a text complexity framework, and guidance on selecting of a variety of text types and passages that allow for a range of standards/evidences to be demonstrated to meet the PARCC claims. PARCC ELA/L tests are based on authentic texts, including multi-media stimulus. Authentic texts are grade-appropriate texts that are not developed for the purposes of the assessment or to achieve a particular readability metric, but reflect the original language of the authors. Pearson and WestEd content experts reviewed the passages for adherence to the PARCC passage selection guidelines (guidelines available here: <https://prc.parcconline.org/library/parcc-passage-selection-guidelines>) to meet to the annual item development plan described above in the number and distribution of genres and topics prior to review and consideration by the State Text Review Committee. ELA/L item development was not conducted until after texts were approved by the State Text Review committee.

Item Development

Guided by the PARCC foundational documents, Pearson and WestEd recruited and trained the item writers and managed the item writing to develop the number of items specified in the annual item development plan. Prior to further committee reviews, the assessment teams at Pearson and WestEd reviewed the items - for content accuracy, alignment to the standards, range of difficulty, adherence to universal design principles (which maximize the participation of the widest possible range of students), bias and sensitivity, and copy edit to enable the accurate measurement of the PARCC standards.

2.2.2 Item and Text Review Committees

Members of the PARCC OWGs for ELA/L and mathematics, state-level experts, local educators, postsecondary faculty, and community members from the PARCC states conducted rigorous reviews of every item and passage being developed for the PARCC assessment system to ensure all test items are of the highest quality, aligned to the standards, and fair for all student populations. All PARCC reviewers were nominated by their state education agency. The purpose of the educator reviews was to provide feedback to Pearson and WestEd, and PARCC on the quality, accuracy, alignment, and appropriateness of the test passages and items developed annually for the summative PARCC assessments. The meetings were conducted either in person or virtually and included large group training on the expectations and processes of each meeting, followed by break outs into grade/subject working committees where additional training was provided.

State Text Review

The State Text Review is a review and approval by the State Text Review Committee of the texts eligible for item development. Participants reviewed and provided feedback to Pearson, WestEd, and PARCC about the grade-level appropriateness, content, and potential bias concerns, and reached consensus about which texts would move forward for development. The State Text Review Committee was made up of both State Content and Bias and Sensitivity committee members.

State Content Item Review

During State Content Item Review, committees reviewed and edited test items for adherence to the PARCC foundational documents, basic universal design principles, PARCC Accessibility Guidelines, associated item metadata, and PARCC Style Guide. Committees accessed the item content within the Pearson Assessment Banking for Building and Interoperability (ABBI) system that previews how the passages and items will be displayed in an operational online environment. Committees also verified that the appropriate scoring rule had been applied to each item. The Content Review committees were made up of Operational Working Group members and educators nominated by PARCC member states.

State Bias and Sensitivity Review

Educators and community members make up the committee that reviews items and tasks to confirm that there are no bias or sensitivity issues that would interfere with a student's ability to achieve his or her best performance. The committee reviewed items and tasks to evaluate adherence to the Fairness and Sensitivity Guidelines, and to ensure that items and tasks do not unfairly advantage or disadvantage one student or group of students over another. Bias and Sensitivity Committee members made edits and modifications to items and passages to eliminate sources of bias and improve accessibility for all students.

Editorial Review

The PARCC editorial review committee is comprised of state-level editors who reviewed up to 10 percent of the items and tasks. The committee reviewed the items for copy edit, clarity, and adherence to the PARCC Style Guide.

Data Review Committee

Following the field test, educator and bias committee members met to evaluate test items and associated performance data with regard to appropriateness, level of difficulty, and potential gender, ethnic, or other bias, then recommended acceptance or rejection of each field-test item for inclusion on an operational assessment. The committee also made recommendations that items be revised and re-field tested. Items that were approved by the committee are eligible for use on operational summative assessments.

2.2.3 Operational Test Construction

Under the guidance in the Operational Test Form Creation Specifications, Pearson constructed the operational forms to adhere to the test blueprints and the assessment goals outlined in the form creation specifications. These goals were:

- Test forms designed to measure well across the full range of student ability;
- Scores that are comparable among forms and across test administrations;
- Scales that support classification of students into performance levels;
- The number of parallel forms are maximized;
- Overexposure of items is minimized; and
- Adherence to standards for validity, reliability, and fairness (*Standards for Educational and Psychological Testing, 2014*).

Each content-area and grade-level assessment was based on a specific test blueprint that guided how each test is built. Test blueprints determined the range and distribution of content, and the distribution of points across the PARCC subclaims and task types.

Multiple operational forms were constructed for each grade/subject. These forms were designed to facilitate psychometric equating through a common item linking strategy (described in Section 2.2.4) and to be constructed as “parallel” as possible from a content and test-taking experience. Evaluation criteria for parallelism included adherence to blueprint; sequencing of content across the forms; statistical averages and distributions for difficulty (e.g., p value) and discrimination (e.g., polyserial correlation); item type and cognitive complexity; and passage characteristics for ELA/L including genre, topics, word count, and text complexity.

Core forms are the operational test forms consisting of only those items that will count towards a student’s score. Core forms are constructed to meet the blueprint and psychometric properties outlined in the test construction specifications. PARCC creates multiple core forms for a given assessment to enhance test security and to support opportunity for item release. The number of core operational forms per grade/subject is provided in Table 2.1. Additionally, appropriate forms were identified as accessibility and accommodated forms; and the core forms for all mathematics assessments included embedded field test items. A sample of students were administered ELA/L core forms with embedded field test items. Accessibility and accommodated forms and embedded field testing are described later in this section.

Table 2.1 Number of Core Operational Forms per Grade/Subject and Mode for ELA/L and Mathematics

Grade/ Subject	ELA/L		Mathematics	
	CBT	PBT	CBT	PBT
Grade 3	3	2	3	2
Grade 4	3	2	3	2
Grade 5	3	2	3	2
Grade 6	3	2	3	2
Grade 7	3	2	3	2
Grade 8	3	2	3	2
Grade 9	3	2		
Grade 10	3	2		
Grade 11	3	2		
Algebra I			3	2
Geometry			3	2
Algebra II			3	2
Integrated Mathematics I			1	1
Integrated Mathematics II			1	1
Integrated Mathematics III			1	1

Test Construction Activities

After the Data Review Meetings and prior to the Test Construction Meetings, Pearson assessment specialists constructed initial versions of all of the core forms, as depicted in Table 2.1. The construction model varied slightly between the two subject areas.

For ELA/L, content specialists constructed the initial core forms shown in Table 2.1 based on the support documents and specific processes to achieve fair parallel forms. The following steps were used to construct the operational core ELA/L form inputs taken to the Test Construction Committee for review.

1. Constructed the online forms to match blueprint and test construction specifications
2. Constructed the paper forms to match the blueprint and test construction specifications
3. Identified Accommodated and Accessibility Forms by evaluating the constructed forms for eligibility

The ELA/L construction process included iterative steps between content specialists and psychometricians. Custom PARCC test construction reports (i.e., SAS Reports) generated by the Pearson psychometric team provided information on adherence to blueprint and statistical averages/distributions of item difficulty and discrimination describing the forms and allowing comparison of the forms. These reports facilitated content changes to better achieve the test construction goals.

For mathematics, Pearson employed the use of an automatic test assembler (ATA) to select the items for the initial forms. Based on the blueprints and other test construction goals and specifications, the ATA was able to create sets of items best satisfying the statistical parameters outlined in the test construction specifications; however the ATA was unable to sequence the items as required by the PARCC Mathematics Sequencing Guidelines. Sequencing was conducted by assessment specialists who ordered the items according to the sequencing guidelines. To achieve the appropriate linking design, assessment specialists created linking item sets from an ATA-generated linking blueprint; these sets are shared across forms using the strategy described later in Section 2.2.4. The following steps were used to construct the linking sets and operational core form inputs taken to the Test Construction Committee for review.

1. ATA pulls linking online blueprint
2. Blueprint sequenced and linking item sets created
3. ATA uses linking sets and pulls online forms
4. Construct the online forms
5. Construct the paper forms
6. Identify Accommodated and Accessibility Forms

Similar to the ELA/L construction process, mathematics included iterative steps between assessment specialists and psychometricians. Custom PARCC test construction reports (i.e., SAS Reports) generated by the Pearson psychometric team provided information on adherence to blueprint and statistical averages/distributions of item difficulty and discrimination allowing a comparison of the forms and facilitating content changes to better achieve the test construction goals. Since the mathematics forms were generated by the ATA, psychometricians could also generate the SAS reports prior to content experts reviewing the forms.

Pearson assessment specialists identified forms for each grade/subject suitable for use as the accommodated forms. The content of these forms was also reviewed by Pearson accessibility specialists allowing for content changes prior to the Test Construction Meetings.

These test construction activities provided significant inputs to commence the Test Construction Meetings including:

- The proposed items for the initial operational core forms and the accommodated forms described above
- SAS reports describing each form and comparing parallel forms
- Recommended accommodated forms

Test Construction Meeting to Review Test Construction Inputs

Members of the State Item Content Committees and the Accessibility, Accommodations, and Fairness (AAF) Operational Working Group (OWG) participated in the building of operational core forms that met PARCC summative assessments requirements. In that process, they met in an in-person meeting to

review and made recommendations for changes so that test forms conformed to both the content and psychometric requirements of the assessment.

Accommodated Form Review Process

In addition to participating in many of the development activities including the State Text Review and the State Bias and Sensitivity Review meetings, the Accessibility, Accommodations, and Fairness (AAF) Operational Working Group (OWG) reviewed the proposed accommodated forms at the Test Construction Meeting for accessibility to make sure that the content can be accommodated for students with disabilities and English learners without changing the underlying measured construct.

Forms were identified to support the following accommodations:

Group 1

- Braille (Tactile Graphics available)
- Large Print
- Refreshable braille (ELA/L only) (Tactile Graphics available)
 - Also supports Screen Reader Assistive Technology
- Screen Reader Assistive Technology (Mathematics)
- Spanish Paper (Mathematics only)

Group 2

- Closed Captioning (ELA/L only)
- Signing: ASL (ELA/L only)
- Online TTS (Text and Graphics only)
 - ELA/L
 - Mathematics
 - Need to support text only, and text and graphics
- Online Spanish/TTS (Mathematics only)

At the conclusion of the meetings, all test forms were constructed to meet test blueprints and PARCC requirements, and to the extent possible, reflect the operational linking design. Each test form reflected the test blueprint in terms of content, item types, and test length, as well as *expected* difficulty and performance along the ability continuum. Linking sets were proportionally representative of the operational test blueprint. The operational core forms, linking set forms, and field test forms were reviewed by PARCC Forms Review committees and approved prior to the test administration.

Spanish-Language Assessments for Mathematics

For English learners, PARCC offers a paper-based edition of the mathematics assessment in Spanish, as well as large print and Text-to-Speech versions of the test in Spanish. Once the operational form was approved, the form was sent to Pearson's subcontractor, Teneo, for transadaption of the items. Transadaption differs from translation in that it takes into consideration the grade-level appropriateness of the words, as well as the linguistic and cultural differences that exist between speakers of two different languages. Accounting for these differences allows the item to measure the achievement of Spanish language speakers in the same way that the original version of the item does for native speakers.

of English. The PARCC Spanish Glossary provided guidance to the translator conducting the transadaptation in grade-level and culturally-appropriate ways of transadapting the items. For the Spanish language text-to-speech form, the alternate text (used for description and/or text in art and graphics) was transadapted from the alternate text for the English language version of the Text-to-Speech form. Phonetic mark-up, which guides how the text-to-speech reader pronounces content-specific words and phrases, was also applied in this process.

In addition to the expert review of potential content for all accommodated forms conducted by the AAF OWG with assistance from content experts at the Test Construction Meetings, the transadapted forms underwent three additional quality checks: a Pearson Spanish copy edit services review and approval, an AAF OWG review and approval, and a Spanish DIF analysis after the administration.

2.2.4 Linking Design of the Operational Test

This section begins with a discussion of special considerations for selection of linking items, followed by two examples of the graphical representations of the linking designs for ELA/L and for mathematics. To support the goal of score comparability within and across administrations and years, PARCC implemented a hybrid approach that incorporated the strengths of common item linking and randomly equivalent groups. The use of repeated operational core items was leveraged for common item linking. In addition, all forms were available throughout the operational administration, with spiraling at the student level, leveraged to support linking through randomly equivalent groups.

The PARCC operational test forms involved various types of linking: horizontal linking, testing mode linking, and across administration linking. Horizontal linking consisted of linking items, or common items, included in multiple forms in a single administration. The horizontal linking was achieved through a daisy-chain strategy. This strategy links multiple operational forms together in a ring; where each operational form shares some items with a preceding form and some items with a following form, and the last form also shares some items with the first form. Together, all the shared items make up the horizontal linking set. All forms for the grade and subject are connected, but not identical (e.g., A is connected to B, B is connected to C, and C is connected to A). Testing mode linking consisted of common items placed in computer-based forms and paper-based forms within an administration to support the development of scores on the same reporting scale. Across administration linking, or year-to-year linking, consisted of common items included in two different administrations. The placement of linking items across forms or administrations supports the development of comparable scores.

Linking item sets can be internal or external linking sets. Internal linking sets consist of common items in operational positions such that the items contribute to the students' scores. External linking sets consist of common items in positions resulting in the items not contributing to students' scores. The 2015-2016 linking designs included both external and internal linking sets. The horizontal linking across forms within an administration and across administration linking included internal linking sets. The testing mode linking included both internal and external linking sets.

For ELA/L, the horizontal linking designs for the Spring 2016 online test forms were based on the number of unique test forms constructed for a grade. After constructing the unique test forms, the test

forms were divided into sections and sections were dispersed across additional forms such that each section appeared on two forms. As a result, the operational linking sets represented full test blueprints. This means that linking items were selected to reflect the content balance, task models, types of items, and cognitive complexity of the full PARCC assessment.

For mathematics, the ATA pulled an initial blueprint linking set that was divided into item sets and distributed across the Spring 2016 online forms following a daisy-chain strategy, as depicted below in Figure 2.2.

The paper forms for both subjects were generated from the online forms. In response to several practical constraints based on the number of forms constructed for each mode and to meet the blueprints (e.g., inclusion of technology enhanced items in CBT forms), there was no one online form that was administered intact in the paper delivery mode at any grade level. For example, technology enhanced items from online forms were replaced in the paper forms with items from similar content, but appropriate for paper-based testing. However, for both subjects, the content on paper forms significantly overlaps with that on the online forms.

2.2.5 Graphical Representation of PARCC Operational Test Linking Design

This section includes two examples of graphical representations to illustrate the horizontal linking designs described above. Designs for across administrations linking are not included in these graphs. Note that to the extent possible, item set (a) for within year equating and item set (b) for across year equating will be the same set of items. Limitations in achieving identical linking sets include a change in blueprint from last year, the release of some of last year's content, and exposure concerns.

Figure 2.1 illustrates the linking design for ELA/L for grades 6-11 CBT forms (O1, O2, O3) and PBT forms (P1, P2). Each form was constructed to include two linking sets, one text and one task, accounting for approximately 33% of the score points. The set of forms for a grade/subject linked through a "daisy chain" model such that each form was linked to two adjacent forms. Additional new content had exposure on multiple forms, providing stronger within year linking. Grades 3-5 linking design was similar.

PARCC ELA/L Grades 6-11

Form		P1	P2	O1	O2	O3
Unit 1	LAT	Linking				Set 1
	PCR					
	Long			Linking Set 2		
Unit 2	RST					
	PCR					
	Short	Linking Set 3				
Unit 3	NT		Linking Set 4			
	PCR					
	Pair				Linking Set 5	

Figure 2.1 PARCC ELA/L CBT and PBT Linking Design (Grades 6-11)

Figure 2.2 illustrates the linking design for mathematics for grades 3-5 CBT forms (O1, O2, O3) and PBT forms (P1, P2). Each form was constructed to include three linking sets, accounting for approximately 38% of the score points. Linking sets were positioned in different units. The set of forms for a grade/subject linked through a “daisy chain” model such that each form was linked to two adjacent forms. Additional new content had exposure on multiple forms, providing stronger within year linking. Grades 6-8 and high school linking designs were similar.

PARCC Mathematics Grades 3-5						
Form		P1	P2	O1	O2	O3
Unit 1	Item set	Linking Set 1				
	Item set			Linking Set 2		
Unit 2	Item set		Linking Set 3			
	Item set	Linking				Set 4
Unit 3	Item set			Linking Set 5		
	Item set	Linking Set 6				
Unit 4	Item set				Linking Set 7	
	Item set					

Figure 2.2 PARCC Mathematics CBT and PBT Linking Design (Grades 3-6)

2.2.6 Field Test Data Collection Overview

Field test items were embedded in the 2016 spring operational forms to collect data for psychometric analysis necessary to support the assessment system for future administrations. Field test administration entailed paper and computer administration modes, with computer administration as the dominant mode. The ELA/L embedded field test items were administered to a sample of students.

The initial data collection design entailed two conditions. Condition one, which comprised the mathematics assessment, was an embedded census field test model in which all students taking the summative assessment participated in the field test. Field test sets were constructed to balance the expected cognitive load and difficulty across forms, reflected in the number of points, distribution of task types, and balance of passages for ELA/L. Forms for each content area were spiraled at the student level.

Under Condition 2, which comprised the ELA/L assessment, PARCC sampled approximately one third of the schools across the consortium states. Students in the sampled schools took forms containing ELA/L embedded field-test tasks. Schools were selected so that the sample for each ELA/L assessment was representative of the general PARCC testing populations in terms of achievement (i.e., average scale

score and percentage of student at Levels 4 and 5 in the previous year) and demographics (i.e., ethnicity composition, percentage of economically disadvantaged, English learners, and students with disabilities). A three-year sampling plan was created such that if a given school was part of the ELA/L field test one year (e.g., spring 2016), it would not be required to participate in the field test for the subsequent two years (e.g., spring 2017 and 2018).

Section 3: Test Administration

3.1 Testing Windows

Fall/Winter Block 2015 (ELA/L) and Spring 2016 (ELA/L) assessments focused on writing effectively when analyzing text. Fall/Winter Block 2015 Mathematics and Spring 2016 Mathematics assessments focused on applying skills and concepts, and understanding multi-step problems that require abstract reasoning and modeling real-world problems, precision, perseverance, and strategic use of tools. In both content areas, students demonstrated their acquired skills and knowledge by answering selected response items and fill-in-the-blank questions.

The 2015-2016 operational administration of PARCC assessments included a Fall/Winter Block administration beginning in the fall of 2015 as well as a Spring administration in the spring of 2016. The Fall/Winter Block 2015 operational administration of the PARCC assessment included two separate components: the Performance-Based Assessment (PBA) and the End-of-Year (EOY) assessment. Both components were administered as computer-based tests (CBT) and as paper-based tests (PBT). A student must have participated in both content-specific windows (PBA and EOY) in order to receive a score for the assessment. The Spring 2016 operational administration of the PARCC assessment combined the Performance-Based Assessment (PBA) and End-of-Year (EOY) into one testing window. Each PARCC assessment was comprised of multiple units, and additionally, one of the mathematics units for grade 7 and high school course assessments, was split into two sections: a non-calculator section and a calculator section.

Table 3.1 PARCC Fall/Winter Block 2015 and Spring 2016 Testing Windows

Fall/Winter Block 2015	Performance-Based Assessment (PBA) November 9, 2015 – December 18, 2015 End-of-Year (EOY) Assessment December 7, 2015 – January 29, 2016
Spring 2016	March 7, 2016 – June 10, 2016

3.2 Test Security and Administration Policies

The administration of any PARCC assessment is a secure testing event. Maintaining the security of test materials before, during, and after the test administration is crucial to obtaining valid and reliable results. School Test Coordinators are responsible for ensuring that all personnel with authorized access to secure materials are trained in and subsequently act in accordance with all security requirements.

School Test Coordinators must implement chain-of-custody requirements for specified materials. School Test Coordinators are responsible for distributing materials to Test Administrators, collecting materials from Test Administrators, returning secure test materials and securely destroying certain specified materials after testing.

The administration of PARCC assessment includes both secure and non-secure materials, and these materials are further delineated by whether they are “scorable” or “nonscorable,” depending on whether the assessments were administered via paper/pencil (e.g., paper-based assessments) or online (e.g., computer-based assessments). For the 2015-2016 paper-based administration, students used paper-based answer documents (except in grade 3 where students responded directly into test booklets). About 87% of the PARCC assessments administered during the 2015-2016 administration were online assessments, and about 13% were paper-based assessments.

Secure vs. Non-Secure Materials

PARCC defines secure materials as those that must be closely monitored and tracked to prevent unauthorized access to or prohibited use or distribution of secure content such as test items, reading passages, student work, etc. For paper-based tests, secure materials include both used and unused test booklets and used scratch paper while for computer-based tests, secure materials include student testing tickets, secure administration scripts (e.g., mathematics read-aloud), and used scratch paper. PARCC defines non-secure materials as any authorized testing materials that do not include secure content (e.g., test items or student work). These include test administration manuals, unused scratch paper, and mathematics reference sheets that have not been written upon, etc.

Scorable vs. Nonscorable Materials

Paper-based assessments have both scorable and nonscorable materials while computer-based assessments only have nonscorable materials. Scorable materials for paper-based assessments comprise of used (e.g., includes student work) test booklets (grade 3) and answer documents (grades 4 and above) only. Scorable materials must be returned to the vendor to be scored. All other materials for paper-based testing, such as blank (e.g., unused) test booklets, test administration manuals, scratch paper, mathematics reference sheets, etc., are deemed non-scorable. For computer-based tests, there are no scorable materials as student work is submitted electronically for scoring; thus there are limited physical materials to return (e.g., secure administration scripts for certain accommodations).

Students taking the computer-based test may not have access to secure test materials before testing, including printed Student Testing Tickets. Printed Mathematics Reference Sheets (if applicable) and scratch paper must be new and unmarked.

Students taking the paper-based test may not have access to scorable or nonscorable secure test content before or after testing. Scorable secure materials that are to be provided by Test Administrators to students include Test Booklets (Grade 3) or Answer Documents (Grades 4-high school). Nonscorable secure materials that are distributed by Test Administrators to paper-based testing students include

Large Print Test Booklets, braille Test Booklets, scratch paper (paper used by students to take notes and work through items), and printed Mathematics Reference Sheets (grades 5–8 and high school).

School Test Coordinators are required to maintain a tracking log to account for collection and destruction of test materials, including mathematics reference sheets written on by students and scratch paper written on by students. As part of the test administration policy, schools are required to maintain the Chain-of-Custody Form or tracking log of secure materials for at least three years unless otherwise directed by state policy. Copies of the Chain-of-Custody Form for paper-based testing are included in each Local Education Agency (LEA) or school's test materials shipment.

Test Administrators are not to have extended access to test materials before or after administration (except for certain accessibility or accommodations purposes). Test Administrators must document the receipt and return of all secure test materials (used and unused) to the School Test Coordinator immediately after testing.

All PARCC test security and administration policies are found in the *PARCC Test Coordinator Manual* and the *PARCC Test Administrator Manuals*. Archived versions of test administration manuals from past administration years can be found at: <http://parconline.org/assessments/administration/archived-testing-manuals>. State security and administration policies may exceed that of the PARCC policies. State-specific policies are included in **Appendix C** of the *Test Coordinator Manual*.

3.3 Accessibility Features and Accommodations

3.3.1 Participation Guidelines for PARCC Assessments

All students, including students with disabilities and English learners, are required to participate in statewide assessments and have their assessment results be part of the state's accountability systems, with narrow exceptions for English learners in their first year in a U.S. school, and certain students with disabilities who have been identified by the Individualized Education Program (IEP) team to take their state's alternate assessment. All eligible students will participate in the PARCC ELA/literacy and mathematics assessments. Federal laws governing student participation in statewide assessments include the No Child Left Behind Act of 2001 (NCLB), the Individuals with Disabilities Education Improvement Act of 2004 (IDEA), Section 504 of the Rehabilitation Act of 1973 (reauthorized in 2008), and the Elementary and Secondary Education Act (ESEA) of 1965, as amended. All students can receive accessibility features on PARCC assessments.

Four distinct groups of students may receive accommodations on PARCC assessments:

1. **Students with disabilities who have an Individualized Education Program (IEP);**
2. **Students with a Section 504 plan** who have a physical or mental impairment that substantially limits one or more major life activities, have a record of such an impairment, or are regarded as having such an impairment, but who do not qualify for special education services;

3. **Students who are English learners;** and
4. **Students who are English learners with disabilities who have an IEP or 504 plan.** These students are eligible for accommodations intended for both students with disabilities and English learners.

Testing accommodations for students with disabilities or students who are English learners (EL) must be documented according to the guidelines and requirements outlined in the PARCC Accessibility Features and Accommodations Manual, Archived versions of past editions of the Accessibility Features and Accommodations Manual can be found at:

<http://parcconline.org/assessments/administration/archived-testing-manuals>.

3.3.2 PARCC Accessibility System

Through a combination of universal design principles and accessibility features, PARCC has designed an inclusive assessment system by considering accessibility from initial design through item development, field testing, and implementation of the assessments for all students, including students with disabilities, English learners, and English learners with disabilities. Accommodations may still be needed for some students with disabilities and English learners to assist in demonstrating what they know and can do. However, the accessibility features available to students should minimize the need for accommodations during testing and ensure the inclusive, accessible, and fair testing of the diverse students being assessed.

3.3.3 What are Accessibility Features?

On the PARCC computer-based assessments, accessibility features are tools or preferences that are either built into the assessment system or provided externally by Test Administrators, and may be used by any student taking the PARCC assessments (i.e., students with and without disabilities, gifted students, English learners, and English learners with disabilities). Since accessibility features are intended for all students, they are not classified as accommodations. Students should have the opportunity to select and practice using them prior to testing to determine which are appropriate for use on the PARCC assessment. Consideration should be given to the supports a student finds helpful and consistently uses during instruction. Practice tests that include accessibility features are available for teacher and student use throughout the year. Practice tests are available at parcc.pearson.com.

3.3.4 Accommodations for Students with Disabilities and English Learners

It is important to ensure that performance in the classroom and on assessments is influenced minimally, if at all, by a student's disability or linguistic/cultural characteristics that may be unrelated to the content being assessed. For PARCC assessments, accommodations are considered to be adjustments to the testing conditions, test format, or test administration that provide equitable access during assessments for students with disabilities and students who are English learners. In general, the administration of the assessment should not be the first occasion on which an accommodation is introduced to the student. To the extent possible, accommodations should:

- provide equitable access during instruction and assessments;
- mitigate the effects of a student's disability;
- not reduce learning or performance expectations;
- not change the construct being assessed; and
- not compromise the integrity or validity of the assessment.

Accommodations are intended to reduce and/or eliminate the effects of a student's disability and/or English language proficiency level; however, **accommodations should never reduce learning expectations by reducing the scope, complexity, or rigor of an assessment.** Moreover, accommodations provided to a student on the PARCC assessments must be generally consistent with those provided for classroom instruction and classroom assessments. There are some accommodations that may be used for instruction and for formative assessments that are not allowed for the summative assessment because they impact the validity of the assessment results; for example, allowing a student to use a thesaurus or access the Internet during a PARCC assessment. There may be consequences (e.g., excluding a student's test score) for the use of non-allowable accommodations during PARCC assessments. It is important for educators to become familiar with PARCC policies regarding accommodations used for assessments.

To the extent possible, accommodations should adhere to the following principles:

- Accommodations enable students to participate more fully and fairly in instruction and assessments and to demonstrate their knowledge and skills.
- Accommodations should be based upon an individual student's needs rather than on the category of a student's disability, level of English language proficiency alone, level of or access to grade-level instruction, amount of time spent in a general classroom, current program setting, or availability of staff.
- Accommodations should be based on a documented need in the instruction/assessment setting and should not be provided for the purpose of giving the student an enhancement that could be viewed as an unfair advantage.
- Accommodations for students with disabilities must be described and documented in the student's appropriate plan (i.e., either a 504 plan or an approved IEP); and must be provided if they are listed.
- Accommodations for English learners should be described and documented.
- Students who are English learners with disabilities are eligible to receive accommodations for both students with disabilities and English learners.

- Accommodations should become part of the student's program of daily instruction as soon as possible after completion and approval of the appropriate plan.
- Accommodations should not be introduced for the first time during the testing of a student.
- Accommodations should be monitored for effectiveness.
- Accommodations used for instruction should also be used, if allowable, on local district assessments and state assessments.

In the following scenarios, the school must follow each state's policies and procedures for notifying the state assessment office:

- A student **was provided a test accommodation that was not listed** in his or her IEP/504 plan/documentation for an English learner, or
- A student **was not provided a test accommodation that was listed** in his or her IEP/504 plan/documentation for an English learner.

3.3.5 Unique Accommodations

PARCC provides a comprehensive list of accessibility features and accommodations in the PARCC Accessibility Features and Accommodations Manual that are designed to increase access to PARCC assessments and that will result in valid, comparable assessment scores. However, students with disabilities or English learners may require additional accommodations that are not already listed. PARCC states individually review requests for unique accommodations in their respective states and provide a determination as to whether the accommodation would result in a valid score for the student, and if so, would approve the request.

3.3.6 Emergency Accommodations

An emergency accommodation may be appropriate for a student who incurs a temporary disabling condition that interferes with test performance shortly before or during the PARCC assessment window. A student, whether or not they already have an IEP or 504 plan, may require an accommodation as a result of a recently-occurring accident or illness. Cases include students who have a recently-fractured limb (e.g., arm, wrist, or shoulder); whose only pair of eyeglasses has broken; or a student returning to school after a serious or prolonged illness or injury. An emergency accommodation should be given only if the accommodation will result in a valid score for the student (i.e., does not change the construct being measured by the test[s]). If the principal (or designee) determines that a student requires an emergency accommodation on the PARCC assessment, an Emergency Accommodation Form must be completed and maintained in the student's assessment file. If required by a PARCC state, the school may need to consult with the state or district assessment office for approval. **The parent must be notified that an emergency accommodation was provided.** If appropriate, the Emergency Accommodation Form may also be submitted to the district assessment coordinator to be retained in the student's central

office file. Requests for emergency accommodations will be approved after it is determined that use of the accommodation would result in a valid score for the student.

3.3.7 Student Refusal Form

If a student refuses an accommodation listed in his or her IEP, 504 plan, or if required by the PARCC member state, an English learner plan, the school should document in writing that the student refused the accommodation, and the accommodation must be offered and remain available to the student during testing. This form must be completed and placed in the student's file and a copy must be sent to the parent on the day of refusal. Principals (or designee) should work with Test Administrators to determine who, if any others, should be informed when a student refuses an accommodation documented in an IEP, 504 plan, or (if required by the PARCC member state) English learner plan.

3.4 Testing Irregularities and Security Breaches

Any action that compromises test security or score validity is prohibited. These may be classified as testing irregularities or security breaches. Below are examples of activities that compromise test security or score validity (note that these lists are not exhaustive). It is highly recommended that School Test Coordinators discuss other possible testing irregularities and security breaches with Test Administrators during training.

Examples of test security breaches and irregularities include but are not limited to:

- **Electronic Devices**
 - Using a cell phone or other prohibited handheld electronic device (e.g., smartphone, iPod, smart watch, personal scanner) while secure test materials are still distributed, while students are testing, after a student turns in his or her test materials, or during a break.
 - Exception: Test Coordinators, Technology Coordinators, Test Administrators, and Proctors are permitted to use cell phones in the testing environment only in cases of emergencies or when timely administration assistance is needed. LEAs may set additional restrictions on allowable devices as needed.
- **Test Supervision**
 - Coaching students during testing, including giving students verbal or nonverbal cues, hints, suggestions, or paraphrasing or defining any part of the test.
 - Engaging in activities (e.g., grading papers, reading a book, newspaper, or magazine) that prevent proper student supervision at all times while secure test materials are still distributed or while students are testing.
 - Leaving students unattended for any period of time while secure test materials are still distributed or while students are testing.
 - Deviating from testing time procedures.

- Allowing cheating of any kind.
 - Providing unauthorized persons with access to secure materials.
 - Unlocking a test in PearsonAccess^{next} during non-testing times.
 - Failing to provide a student with a documented accommodation or providing a student with an accommodation that is not documented and therefore is not appropriate.
 - Allowing students to test before or after the state's test administration window.
- **Test Materials**
 - Losing a student test booklet or answer document.
 - Losing a student testing ticket.
 - Leaving test materials unattended or failing to keep test materials secure at all times.
 - Reading or viewing the passages or test items before, during, or after testing.
 - Exception: Administration of a Human Reader/Signer accessibility feature for mathematics or accommodation for English language arts/literacy which requires a Test Administrator to access passages or test items.
 - Copying or reproducing (e.g., taking a picture of) any part of the passages or test items or any secure test materials or online test forms.
 - Revealing or discussing passages or test items with anyone, including students and school staff, through verbal exchange, email, social media, or any other form of communication.
 - Removing secure test materials from the school's campus or removing them from locked storage for any purpose other than administering the test.
 - **Testing Environment**
 - Allowing unauthorized visitors in the testing environment.
 - Failing to follow administration directions exactly as specified in the *Test Administrator Manual*.
 - Displaying testing aids in the testing environment (e.g., a bulletin board containing relevant instructional materials) during testing.

All instances of security breaches and testing irregularities must be reported to the School Test Coordinator immediately. The Form to Report a Testing Irregularity or Security Breach must be completed within two school days of the incident.

If any situation occurred that could cause any part of the test administration to be compromised, schools referred to the *PARCC Test Coordinator Manual* for each state's policy and immediately followed those steps. Instructions for the School Test Coordinator or LEA Test Coordinator to report a testing irregularity or security breach was available in the *PARCC Test Coordinator Manual*.

3.5 Data Forensics Analyses

Maintaining the validity of test scores is essential in any high-stakes assessment program, and misconduct represents a serious threat to test score validity. When used appropriately, data forensic analyses can serve as an integral component of a wider test security protocol. The results of these data forensic analyses may be instrumental in identifying potential cases of misconduct for further follow-up and investigation.

In 2015-2016, PARCC conducted the following four data forensics analyses on its operational assessments:

- Response Change Analysis
- Plagiarism Analysis
- Internet and Social Media Monitoring
- Off Hours Testing Monitoring

An overview of each data forensics analysis method is provided next.

3.5.1 Response Change Analysis

Response change analysis looks at how often student answers are changed, focusing specifically on an excessive number of wrong answers changed to right answers. In traditional paper-based, multiple-choice testing programs, this is sometimes referred to as “erasure analysis”³. The rationale for erasure analysis is that a teacher or administrator who is intent on improving classroom performance might be motivated to change student responses after the answer sheets are collected. A clustered number of student answer documents from the same school or classroom with unusually high numbers of answers changed from wrong to right might provide evidence to support follow-up investigation. PARCC’s response change analysis extended the traditional erasure method to account for issues specific to computer-based testing as well as the variety of items types on the PARCC assessments, such as partial-credit, multi-part, and multiple-select items.

3.5.2 Plagiarism Analysis

Plagiarism analysis compares the responses given for a group of written composition items, looking for high degrees of similarity. For the PARCC assessments, the primary item type of interest was the prose constructed response (PCR) tasks in the English Language Arts and Literacy (ELA/L) content area. This analysis was conducted for PCR tasks administered online using some of the same artificial intelligence (AI) techniques that are applied in automated essay scoring. Specifically, this method was based on

³ The term “erasure analysis” is sometimes objected to because it is inferential rather than descriptive. A more descriptive term is “mark discrimination analysis” which recognizes that the scanning approach makes discriminations among the darkness of selected answer choices when multiple responses to a multiple-choice item are detected during answer sheet processing.

Latent Semantic Analysis (LSA) technology to detect possible plagiarism. Using LSA, the content of each constructed response was compared against the content of every other constructed response and a measure that indicated the degrees of similarity was generated for each pair of response comparison. Because LSA provided a semantic representation of language, rather than a syntactic or word-based representation, it allowed the detection of potential copying behaviors, even when test takers or administrators substituted synonymous words or phrases.

3.5.3 Internet and Social Media Monitoring

Internet and social media monitoring was conducted by Caveon, LLC. Caveon's team monitored English language websites and searchable forums that were publicly available for suspected proxy testing solicitations and website postings that contain, or appear to contain, infringements of PARCC's protected operational test content. The Internet and social media outlets monitored included popular websites (such as Facebook and Twitter), blogs, discussion forums, video archives, document archives, brain dumps, auction sites, media outlets, peer-to-peer servers, etc. Caveon's process generated regular updates that categorize identified threats by level of actual or potential risk based upon the representations made on the web sites, or actual analysis of the proffered content. For example, categorizations typically ranged from "cleared" (lowest risk but bookmarked for continued monitoring) to "severe" (highest risk). Note that this process only considered potential breaches of secure item content, not violations of testing administration policies. Potential breaches were reported directly to the state(s) implicated for further action. Summary reports describing the threats were provided to PARCC through notification emails.

3.5.4 Off-Hours Testing Monitoring

Off-hours testing monitoring checks for suspicious testing activities at test administration locations occurring outside of the set windows for computer-based testing sessions. PARCC states established set start and end time for administering computer-based assessments. Based on these hours, authorized users (that is, users with the State Role) were allowed to override the start and end times for a test session. The off-hours testing monitoring process tracked such occurrences and logged them in an operational report, which listed the sessions within an organization that selected to test outside the set window. PARCC states could use this report to follow-up with the organizations identified in the report.

Section 4: Item Scoring

4.1 Machine Scored Items

4.1.1 Key Based Items

Pearson performed a key review prior to the test administration to verify that the scoring (answer) keys were correct for each item. Once the forms were constructed and approved by PARCC for publication, an independent key review was performed by an experienced third party vendor. The vendor reviewed each item and confirmed that the key was correct. If discrepancies were identified, a Pearson senior content specialist or content manager reviewed the flagged item(s) and worked with the item developers to resolve the issue.

4.1.2 Rule Based Items

Rule based scoring refers to item types that use various scoring models. PARCC uses QTI (Question and Test Interoperability) item type implementation based on scoring model rules. Examples of these item types include “choice interaction” which presents a set of choices where one or more choices can be selected; text entry, where the response is entered in a text box; hot spot or text interaction, where an area in a graph or text in a paragraph (for example) can be highlighted or match interaction, where an association can be made between pairs of choices in a set. These items include the scoring rules and correct responses as part of their item XML (markup language) coding.

During the initial stages of item development for PARCC, Pearson staff worked closely with PARCC to first delineate the rules for the scoring rubrics and then to adjust those rules based on student responses. During the Item Tryout⁴ planning phase, Pearson content staff received input from PARCC staff to develop a thorough rule based scoring process that met PARCC needs.

Pearson worked with the item developers to review initial scoring rules created during the item development. Once the rule based scoring process was approved by PARCC, and prior to test construction, Pearson content staff worked closely with the item developers to finalize scoring rubrics for items to be scored via the rule based scoring method. The proposed scoring rubrics were sent to PARCC for review, and if any additional changes were needed or new rules added, Pearson documented and applied the requested edits.

During test construction, Pearson monitored and evaluated the scoring and updated the scoring keys/ scoring rules in the item bank. After the tryout items were scored, Pearson prepared a frequency distribution of student responses for each item or task scored using a rule based approach and compared this to the expected response based on correct answers to ensure that scoring keys and rules were appropriately applied. The content team does this by analyzing the student response data to determine if scoring is acceptable using the item metadata and the student response file in conjunction with any potential item issues as flagged by psychometrics. These frequency distributions included an indication of right/wrong and other identifying information defined by PARCC and those items that

⁴ The item tryout was a set of item studies conducted in spring 2015.

showed a statistical anomaly, whereby the frequency distribution was outside of the expected range, were sent to content experts to verify that the items were coded with the correct key.

Following the Rule Based Scoring Educator Committee's review, which occurred prior to year 1 test construction, Pearson analyzed the feedback from the committees and made recommendations about adjustments to the scoring rubrics, based on the results of the reviews. Upon submission of the results, Pearson worked with PARCC staff to discuss these findings and determine next steps prior to the completion of scoring. In subsequent years as scoring inquiries arise throughout the process of test construction, forms creation, testing, scoring, and psychometric analysis, items with scoring discrepancies are brought before the PARCC Task Force for resolution. This committee consists of representatives from each state as well as PARCC and Pearson content specialists.

Following the initial development of the PARCC rule based scoring rubrics, Pearson has continued to monitor and evaluate new item development to ensure the scoring rules established are maintained within all item types as approved.

Pearson continues to use several avenues to monitor scoring each year. Prior to testing a third party key review by which reviewers check operational and field test items for correct keys. Any disputed items go to a 2nd review with Pearson content experts and anything still in question is taken before the PARCC task force for review and possible key change. During testing, Pearson creates early testing files for frequency distribution analysis whereby items where an incorrect key receives a high distribution of responses are further evaluated for accuracy. After testing during psychometric analysis all responses are again evaluated for distribution of responses and potential scoring abnormalities. Any change in scoring that may be requested as a result of the psychometric analysis is also taken before the PARCC task force for decisions. These processes are the same for both paper and online modes of testing.

4.2 Human or Handscored Items

PARCC 2016 constructed-response items were scored by human scorers in a process referred to as handscoring. Online training units were used to train all scorers. The online training units included prompts (items), passages, rubrics, training sets, and qualification sets. Scorers who successfully completed the training and qualified, demonstrating they could correctly score student responses based on the guidelines in the online training units, were permitted to score student responses using the ePEN2 (electronic Performance Evaluation Network, second generation) scoring platform. All online and paper responses were scored within the ePEN2 system. Scorer quality was monitored throughout scoring.

Pearson staff roles and responsibilities were as follows:

- Scorers were individuals who applied scores to student responses.
- Scoring Supervisors monitored the work of a team of scorers through review of scorer statistics and backreading, which is a review of responses scored by each scorer. When backreading, a

supervisor sees the scores applied by scorers, which helps the supervisor provide additional coaching or instruction to the scorer being backread.

- Scoring Directors managed the scoring quality of a subset of items and monitored the work of scoring supervisors and scorers for their assigned items. Scoring Directors backread responses scored by supervisors and scorers as part of their quality monitoring duties.
- ELA/Literacy and Mathematics Content Specialists managed the scoring quality and monitored the work of the Scoring Directors.
- Project Managers documented the procedures, identified risks, and managed day-to-day administrative matters.
- A Program Manager provided oversight for the entire scoring process.

All Pearson employees involved in the scoring or the supervision of scoring possessed at least a four-year college degree.

4.2.1 Scorer Training

Key steps in the development of scorer training materials were Rangefinding and Rangefinder Review meetings where educators and administrators from PARCC states met to interpret the scoring rubrics and determine consensus scores for student responses. Rangefinding meetings were held prior to scoring field test items, and Rangefinder Review meetings were held prior to scoring operational items.

At Rangefinding meetings, educators and administrators from PARCC states reviewed student responses and used scoring rubrics to determine consensus scores. Those responses scored in Rangefinding were used to create field test scorer training sets. After PARCC reviewed scoring statistics from field test scoring, items were selected for operational testing. For items selected to be on the operational assessment, Pearson filled out field test scorer training materials with additional student responses in order to create proposed operational scorer training sets. PARCC educators and administrators then attended Rangefinder Review meetings to review and approve proposed operational training sets.

When developing scorer training materials, Pearson Scoring Directors carefully reviewed detailed notes and records from PARCC Rangefinding and Rangefinder Review committee meetings. Training sets were developed using the responses scored by the committees and additional suitable student response samples (as needed). PARCC reviewers reviewed and approved all scorer training sets prior to scorer training.

During training, scorers reviewed training sets of scored student responses with annotations that explained the rationale for the score assigned. The anchor set was the primary reference for scorers as they internalized the rubric during training. Each anchor set consisted of responses that were clear examples of student performance at each score point. The responses selected were representative of typical approaches to the task and arranged to reflect a continuum of performance. All scorers had

access to the anchor set when they were training and scoring and were directed to refer to it regularly during scoring.

Practice set papers are used to further define the lines of each score point and may not be as clear cut as the anchor responses.

Qualification sets were used to confirm that scorers understood how to score student responses accurately. Qualification sets were composed of responses that were clear examples of score points. Scorers were required to meet specified agreement percentages on qualification sets in order to score student responses.

Pearson developed two types of training sets to train scorers: prototype and abbreviated sets. “Prototype” training sets were complete training sets consisting of anchor, practice, and qualification sets (refer to 4.2.2 for information on qualification process). In English language arts/literacy (ELA/L) there was one prototype training set per task type (Research Simulation Task, Literary Analysis Task, and Narrative Writing Task) at each of the nine grade levels (grades 3-11). In mathematics, a prototype training set was built for each group of items containing the same number of open-ended points in each grade or course, for a total of approximately three to five prototype sets per grade level or course.

The prototype training approach promoted consistency in scoring, as each subsequent abbreviated training set for the ELA/L task type or mathematics item grouping was based on the prototype. Once a prototype was chosen, full training materials were developed for that item, and at each grade level, scorers were trained to score a particular task type using the prototype training materials for that type.

Abbreviated training sets were prepared for all items not selected for prototype training sets. The abbreviated training sets included an anchor set and two practice sets so scorers could internalize the scoring standards for these new items, which were similar to prototype items they had previously scored.

Anchor and practice sets for both prototype and abbreviated items included annotations for each response. Annotations are formal written explanations of the score for each student response.

The table below details the composition of the anchor sets, practice sets, and qualification sets.

Table 4.1 Training Materials Used During Scoring

Training Set Development	
Description	Specification
Anchor Set	
The anchor set is the primary reference for scorers as they internalize the rubric during training. All scorers have access to the anchor set when they are training and scoring, and are directed to refer to it regularly.	<p>The anchor set for mathematics prototype items comprises 3 annotated responses per score point.</p> <p>The anchor set for subsequent abbreviated items for mathematics comprise 1-3 annotated responses per score point.</p>
The anchor set comprises clear examples of student performance at each score point. The responses selected may be representative of typical approaches to the task or arranged to reflect a continuum of performance.	The anchor sets for ELA/L prototype items comprise 3 annotated responses per score point. Anchor sets for prototype items include separate complete anchor sets for each applicable scoring trait (Reading Comprehension and Written Expression and Conventions).
Practice Sets	
Practice sets are used to help trainees develop experience in independently applying the scoring guide (the rubric) to student responses. Some of these responses clearly reinforce the scoring guidelines presented in the anchor set. Other responses are selected because they are more difficult to evaluate, fall near the boundary between two score categories, or represent unusual approaches to the task.	<p>The practice sets for mathematics prototype and abbreviated items include two to three sets of ten annotated responses.</p>
The practice sets provide guidance and practice for trainees in defining the line between score categories, as well as applying the scoring criteria to a wider range of types of responses	ELA/L practice sets for prototype items include two sets of five annotated responses and two sets of ten annotated responses.
	The subsequent ELA/L practice sets for abbreviated items include two sets of ten annotated responses.
Qualification Sets	
	<p>The qualification sets for mathematics prototype items include 3 sets of 10 responses each (not annotated).</p> <p>The subsequent mathematics abbreviated items for mathematics do not include qualification sets.</p>

Table 4.1 Training Materials Used During Scoring

Training Set Development	
Description	Specification
Qualification sets are used to confirm that scorer trainees understand the scoring criteria and are able to assign scores to student responses accurately. The responses in these sets are selected to reinforce the application of the scoring criteria illustrated in the anchor set.	The qualification sets for ELA/L prototype items include 3 sets of 10 responses each (not annotated).
Scorer trainees must demonstrate acceptable performance on these sets by meeting a pre-determined standard for accuracy in order to qualify to score. Pearson scoring staff define and document qualifying standards in conjunction with PARCC prior to scoring. The qualification sets for mathematics prototype items include 3 sets of 10 responses each (not annotated).	The subsequent ELA/L abbreviated items do not include qualification sets.

4.2.2 Scorer Qualification

In order to score items, scorers were required to show that they were able to apply PARCC scoring methodology accurately through a qualification process. Scorers were asked to apply scores to three qualification sets consisting of 10 responses each. ELA/L scorers applied a score for each part on each response in the qualification sets. Literary Analysis, Research Simulation, and the Narrative Writing Tasks each had two parts: the Reading Comprehension and Written Expression part and the Conventions part. Mathematics scorers applied a score for each part of an item that was a constructed response. The number of constructed-response parts for each mathematics item ranged from one to four. Scorers were required match the PARCC-approved score at a percentage agreed to by PARCC in order to qualify.

For ELA/L qualification, scorers were required to meet the following three conditions:

1. On at least one of the three qualifying sets, at least 70% of the ratings on each of the three scoring parts (considered separately), must agree exactly with the PARCC-approved scores.
2. On at least two of the three qualifying sets, at least 70% of the ratings (combined across the three scoring parts) must agree exactly with the PARCC-approved scores.
3. Combining over the three qualifying sets and across the three scoring parts, at least 95% of the ratings must be within one point of the PARCC-approved scores.

For mathematics qualification, the requirements were based on the item types and score point ranges. Because mathematics items can have one or more scoring parts, a scorer needed to achieve the following requirements separately for each scoring part (when applicable to the item):

Table 4.2 Mathematics Qualification Requirements

Category	Score Point Range	Perfect Agreement	Within One Point
2-category	0-1	90%	100%
3-category	0-2	80%	96%
4-category	0-3	70%	96%
5-category	0-4	70%	96%
7-category	0-6	70%	95%

On at least two of the three qualifying sets, a scorer was required to meet the “Perfect Agreement” percentage indicated in the table above for each category. “Perfect agreement” was achieved when the scores applied exactly matched the PARCC-approved scores. Over the three qualifying sets, a scorer was required to meet the “Within One Point” percentage indicated in the table above for each category. The average is exclusive to each part, so an item with multiple scoring parts would have multiple part rating averages within one point of the PARCC-approved score.

4.2.3 Managing Scoring

Pearson created a Handscoring Specifications document that detailed the handscoring schedule, customer requirements, rangefinding plans, Rangefinding Review, quality management plans, item information, and staffing plans for each scoring administration.

4.2.4 Monitoring Scoring

Second Scoring. Second scoring for ELA/L was performed by human scorers for online responses that received first scores from the Intelligent Essay Assessor and for all ELA/L responses from paper testing. Online ELA/L responses that received first scores from humans were second scored by the Intelligent Essay Assessor. If the first and second scores applied were non-adjacent, a third and occasionally a fourth score was assigned to resolve scorer disagreements. When a resolution score (i.e. 3rd score) was nonadjacent to one and/or both of the first and second scores, the Content Specialist or Scoring Director would apply an adjudication score (4th score).

If a response was scored more than once, the following rules were applied to determine the final score:		
Score Type	Rank	Final Score Calculation
Adjudication	1	If an Adjudication score is assigned, this is the final score.

Resolution	2	If no Adjudication score is assigned, this is the final score.
Backread	3	If no Adjudication or Resolution score is assigned, the latest backreading score is the final score.
Human First score	4	If no Adjudication, Resolution or Backreading score is assigned, this is the final score.
Human Second Score	5	If no Adjudication, Resolution, Backreading or Human First score is assigned, this is the final score.
Artificial Intelligence Score	6	If no Human Score is assigned, this is the final score.

Backreading. Backreading was one of the major responsibilities of Pearson Scoring Supervisors and a primary tool for proactively guarding against scorer drift where scorers score responses in comparison to one another instead of in comparison to the training responses. Scoring supervisory staff used the ePEN2 backreading tool to review scores assigned to individual student responses by any given scorer in order to confirm that the scores were correctly assigned and to give feedback and remediation to individual scorers. Pearson backread approximately five percent of the handscored responses. Backreading scores did not override the original score but were used to monitor scorer performance.

Validity. Validity responses are pre-scored responses strategically interspersed in the pool of live responses. These responses were not distinguishable from any other responses so that scorers were not aware they were scoring validity responses rather than live responses. The use of validity responses provided an objective measure that helped ensure that scorers were applying the same standards throughout the project. In addition, validity was at times shared with scorers in a process known as “validity as review.” Validity as review provided scorers automated, immediate feedback: a chance to review responses they mis-scored, with reference to the correct score and a brief explanation of that score. One validity response was sent to scorers for every 25 “live” responses scored.

PARCC validity agreement requirements for scorers are listed in Table 4.2. Scorers had to meet the required validity agreement percentages to continue working on the PARCC project. Scorers who did not maintain expected agreement statistics were given a series of interventions culminating in a targeted calibration set: a test of scorer knowledge. Scorers who did not pass targeted calibration were removed from scoring the item, and all the scores they assigned were deleted.

Table 4.3 Scoring Validity Agreement Requirements

Subject	Score Point Range	Perfect Agreement	Within One Point
Mathematics	0-1	90%	96%*
Mathematics	0-2	80%	96%
Mathematics	0-3	70%	96%
Mathematics	0-4	65%	95%
Mathematics	0-6	65%	95%
ELA/L	Multi-trait	65%	96%

Calibration Responses. Calibration responses are special sets created during scoring to help train scorers on particular areas of concern or focus. Scoring directors used calibration responses to reinforce rangefinding standards, introduce scoring decisions, or address scoring issues and trends. Calibration was used either to correct a scoring issue or trend, or to continue scorer training by introducing a scoring decision. Calibration was administered regularly throughout scoring.

Inter-rater Agreement. Inter-rater agreement is the agreement between the first and second scores assigned to student responses and is the measure of how often scorers agree with each other. Pearson scoring staff used inter-rater agreement statistics as one factor in determining the needs for continuing training and intervention on both individual and group levels. PARCC inter-rater agreement expectations were as follows:

Table 4.4 Inter-rater Agreement Expectations and Results

Subject	Score Point Range	Perfect Agreement Expectation	Perfect Agreement Result	Within One Point Expectation	Within One Point Result
Mathematics	0-1	90%	94%	96%	100%
Mathematics	0-2	80%	92%	96%	99%
Mathematics	0-3	70%	89%	96%	99%
Mathematics	0-4	65%	86%	95%	98%
Mathematics	0-6	65%	89%	95%	97%
ELA/L	Multi-trait	65%	72%	96%	99%

Pearson's ePEN2 scoring system included comprehensive inter-rater agreement reports that allowed supervisory personnel to monitor both individual and group performance. Based on reviews of these reports, scoring experts targeted individuals for increased backreading and feedback, and if necessary, retraining.

The perfect agreement rate for all mathematics responses scored by two scorers was 90% and the within one point rate was 99%. For all ELA/L responses scored by two scorers, the perfect agreement rate was 72% and the within one point rate was 99%.

The results by grade level for ELA/L are provided in Addendum 13: Inter-rater Agreement for Prose Constructed Response.

4.3 Automated Scoring for PARCC PCRs

Automated scoring performed by Pearson’s Intelligent Essay Assessor (IEA) was the default option for scoring the PARCC assessment’s online Prose Constructed Response (PCR) tasks in 2016. Under the default option, it was assumed that operational scores for approximately 67 percent of the online PCR responses would be assigned by IEA for the spring 2016 administration. The operational scores for the remaining online responses were assigned by human scorers. Human scoring was applied to responses that were scored while IEA was being trained as well as to additional responses routed to human scoring when there was uncertainty about the automated scores. Two states (Massachusetts for all grades and New Jersey in certain grades) adopted an option to have human scores assigned as the first score for 100 percent of responses.

For 10 percent of responses, a second “reliability” score was assigned. The purpose of the reliability score was to provide data for evaluating the consistency of scoring, which is done by evaluating scoring agreement. When IEA provided the first score of record, the second reliability score was a human score. For those states choosing the human scoring option, the second reliability score was assigned by IEA.

4.3.1 Changes to the ELA/L Scoring Rubric

For the 2016 administration, the scoring rubrics for the Literary Analysis tasks (LAT) and Research Simulation Tasks (RST) were updated to combine the Reading Comprehension and Written Expression traits. Therefore, these PCR task items were scored on two traits instead of three: (1) Reading Comprehension and Written Expression and (2) Knowledge of Language and Conventions. Narrative Writing tasks (NWT) continued to be scored on the same two traits as in 2015: (1) Written Expression and (2) Knowledge of Language and Conventions.

One implication of the ELA/L scoring rubric change was that IEA had to be trained on operational data before it could be applied operationally to score the LATs and RSTs. However, it was possible to train IEA on 2015 operational data for any NWTs that were administered online in 2015, and so IEA began scoring those NWT prompts where the training results met the established quality criteria from the outset of the 2016 operational administration.

4.3.2 Continuous Flow

Continuous flow scoring results in an integrated connection between human scoring and automated scoring. It refers to a system of scoring where either an automated score, a human score, or both can be assigned based on a predetermined asynchronous operational flow.

4.3.3 Calibration of IEA Using Operational Data

Continuous flow scoring facilitates the training of IEA using human scores assigned to operational online data collected early in the administration. Once IEA obtains sufficient data to train, it can be “turned on” and becomes the primary source of scoring (although human scoring continues for the 10% reliability sample and other responses that may be routed accordingly).

4.3.4 Smart Routing

Smart routing refers to the practice of using automated scoring results to detect responses that are likely to be challenging to score, and applying automated routing rules to obtain one or more additional human scores. Smart routing can be applied prompt by prompt to the extent needed to meet scoring quality criteria for automated scoring.

4.3.5 Quality Criteria for Evaluating Automated Scoring

The PARCC State Leads approved specific quality criteria for evaluating automated scoring at the time IEA was trained. The primary evaluation criteria for IEA was based on responses to validity papers with “known” scores assigned by experts. For each prompt scored, a set of validity papers is used to monitor the human scoring process over time. Validity papers are seeded into human scoring throughout the administration. The expectation is that IEA can score validity papers at least as accurately as humans can.

Additional measures of inter-rater agreement for evaluating automated scoring were proposed based on the research literature (Williamson, Xi, & Breyer 2012). These measures were previously utilized in Pearson’s automated scoring research and include Pearson correlation, Kappa, Quadratic-Weighted Kappa, exact agreement, and standardized mean difference. These measures are computed between pairs of human scores, as well as between IEA and the humans, to evaluate how performance was the same or different. Criteria for evaluating the training of IEA given these measures include the following:

- Pearson correlation between IEA-human should be within 0.1 of human-human
- Kappa between IEA-human should be within 0.1 of human-human
- Quadratic-Weighted Kappa between IEA-human should be within 0.1 of human-human
- Exact agreement between IEA-human should be within 5.25% of human-human
- Standardized mean difference between IEA-human should be less than 0.15

The specific criteria for evaluating IEA included both primary and secondary criteria and are noted below.

1. Primary Criteria – Based on responses to validity papers: *With smart routing applied as needed, IEA agreement is as good as or better than human agreement for each trait score.*
2. Contingent Primary Criteria – Based on the training responses if validity responses are not available: *With smart routing applied as needed, IEA-Human exact agreement is within 5.25% of Human-Human exact agreement for each trait score.*
3. Secondary Criteria - Based on the training responses: *With smart routing applied as needed, IEA-human differences on statistical measures for each trait score are within the Williamson et al. tolerances for subgroups with at least 50 responses.*

4.3.6 Hierarchy of Assigned Scores for Reporting

When multiple scores are assigned for a given response, the following hierarchy determines which score was 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 back read score and human and/or IEA scores are assigned, the back read score is reported;
- If a resolution score is assigned and an adjudicated score is not assigned, the resolution score is reported (note that if nonadjacent scores are encountered, responses are automatically routed to resolution);
- If an adjudicated score is assigned, it is reported (note that if a resolution score is nonadjacent to the other scores assigned, responses are automated routed to adjudication).

4.3.7 Sampling Responses Used for Training IEA

As previously mentioned, the responses used for training IEA differed for the Narrative prompts and the Literary and Research prompts. Specifically, for the 2016 online Narrative prompts that were also administered online in 2015, IEA was trained using 2015 data and evaluated prior to the start of operational scoring. For the Literary and Research prompts, IEA was trained using early 2016 operational responses.

For prompts trained using operational data, the early performance of human scoring was closely monitored to verify that an appropriate set of data would be available for training IEA. In particular, several characteristics of the human scoring data were monitored, including:

- Exact agreement between human scorers (the goal was for this to be at least 65% for each trait)
- Exact agreement between human scores conditioned on score point (the goal was for this to be at least 50% for each trait)
- The number of responses at each score point (the goal was to have at least 40 responses at the highest score points in the training samples used by IEA)
- Number of responses with two human scores assigned (note that IEA “ordered” additional scoring of responses during the sampling period as needed).

Although the desired characteristics of the training data were easily achieved for some prompts, they were more challenging to achieve for others. For some prompts, a subset of scores were reset and clarifying directions were provided to scorers to improve human-human agreement. For other prompts, special sampling approaches were used to increase the numbers of responses that received top scores. In addition, a healthy percentage of responses were back read during the sampling period and these scores as well as double human scores were all part of the data used to train IEA.

4.3.8 Primary Criteria for Evaluating IEA Performance

The primary criteria for evaluating IEA performance is based on evaluating validity papers and 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.*

To operationalize the primary criteria for a given prompt, the following general steps are undertaken:

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 validity agreement with the human agreement.
4. If the IEA validity agreement is greater than or equal to the human agreement for each trait, IEA can be deployed operationally.

In addition to looking at overall validity agreement, conditional agreement was also examined. In general, it was desirable for IEA to exceed 65 percent agreement at every score point as well as be close to or exceed the human validity agreement at each score point.

One limitation in applying the validity criteria is that validity papers were seeded into the human scoring and accumulated over the full 2016 scoring effort. As a result, sufficient human scored validity papers were not available for evaluating IEA. For this reason, the human agreement rates on validity papers scored in 2015 were used to evaluate several of the prompts. For these comparisons, the 2015 human agreement rates with validity papers for the old Reading and Expressions traits were combined to establish the criteria for the new Expressions trait.

4.3.9 Contingent Primary Criteria for Evaluating IEA Performance

For many of the prompts trained in 2016, it was not possible to utilize human scored validity responses in evaluating IEA performance. In these cases, IEA was evaluated based on IEA-Human exact agreement for each trait score and compared to agreement based on responses that were double-scored by humans. A portion of the data were held out for evaluating IEA-Human exact agreement 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. If the IEA-human agreement is within 5.25% of the human-human agreement, IEA can be deployed operationally.

In addition to the overall comparison, the following performance thresholds in the test data set were targeted: 1) at least 65 percent overall IEA human agreement; and 2) 50 percent IEA-human agreement by score point (i.e., conditioned on the human score). These targets went beyond the contingent primary criteria approved by the State Leads.

4.3.10 Applying Smart Routing

With smart routing, the quality of automated scoring can be increased by routing responses that are more likely to disagree with a human score to receive an additional human score.

When human scorers read a paper, they typically apply integer scores based on a scoring rubric. When there is strong agreement between two independent human readers, they might both assign a score of “3” such that the average score over both raters is also a 3 (i.e., $(3+3)/2 = 3$). IEA simulates this behavior, but because its scores come from an artificial intelligence algorithm, it generates continuous (i.e., decimalized) scores. In this case, the IEA score might be a 2.9 or 3.1. When human readers disagree on the score for a paper, say one reader gives the paper a score of 3 and another reader gives the paper a score of 4, the average of the two scores would be 3.5 (i.e., $3+4=7/2=3.5$). For this paper, IEA would likely provide a score between 3 and 4, say 3.4 or 3.6. Because this continuous score needs to be rounded to an integer score for reporting, it might be reported as a 3 or a 4, depending on the rounding rules. Smart routing involves routing those responses with “in between” IEA scores to additional human

scoring because the nature of the responses suggest there may be less confidence in the IEA score. Since these “in between” IEA scores are based on modeling human scores, it follows that human scores may be less certain as well, and thus such responses tend to be the ones that it makes sense to have double-scored and possibly to resolve if the IEA and human scores are non-adjacent.

Smart routing was utilized as needed to help IEA achieve targeted quality metrics (e.g., validity agreement or agreement with human scorers). Smart routing involved the application of the following four steps:

1. The continuous IEA score for each of the two trait scores was rounded to the nearest score interval of 0.2, starting from zero. For example, IEA scores between 0 and 0.1 were rounded to an interval score of 0, scores between 0.1 and 0.3 were rounded to an interval score of 0.2, scores between 0.3 and 0.5 were rounded to an interval score of 0.4, and so on.
2. Within each of these intervals, the percentage of exact agreement between IEA integer scores and the human scores was calculated for each trait.
3. For each prompt, agreement rates were evaluated by rounding interval. Those intervals for which the agreement rates were below a designated threshold for either trait were identified.
4. Once IEA scoring was implemented, responses within intervals for which IEA – Human agreement was below the designated threshold were routed for additional human scoring.

In training IEA, we first evaluated the scoring models without smart routing by applying either the primary validity criteria or the contingent criteria as described in sections 3.2 and 3.3 of these procedures. For those prompts that did not meet these criteria, we applied increasing smart routing thresholds in an iterative fashion to filter scores and evaluate the remaining scores against the criteria. That is, in any one iteration a particular smart routing threshold was applied such that only scores falling in intervals for which exact agreement exceeded the threshold were included in evaluating the criteria. If the primary or contingent criteria were not met with this level of smart routing, we repeated the analysis applying an increased smart routing threshold. If the primary or contingent criteria were still not met, we repeated the analysis applying a still higher threshold. If the criteria were still not met after a maximum threshold was applied, we investigated different models and/or utilized additional human scoring data until an IEA scoring model was found that met the criteria.

4.3.11 Evaluation of Secondary Criteria for Evaluating IEA Performance

The secondary criteria for evaluating IEA performance involved comparing agreement indices for IEA-human scoring for various demographic subgroups. Because of the importance of protecting personally identifiable information (PII), student demographic data is stored and managed separately from the performance scoring data. For this reason, it was not possible to evaluate subgroup performance in real time as IEA was being trained.

For those prompts trained on early operational data, attempts were made to prioritize the data being returned from the field to include data from states or districts where more diverse populations of students were anticipated. In addition, requests for additional human scores were made to increase the likelihood that there would be sufficient numbers of responses with two human scores for most of the demographic subgroups of interest.

Once IEA was trained and deployed, scoring sets used in training were matched to demographic information so that agreement between IEA and human scorers could be evaluated across subgroups. The analysis were conducted for the following 10 comparison groups:

Group Type	Comparison Groups
Sex	Female Male
Ethnicity	American Indian/Alaska Native Asian Black/African American Hispanic/Latino Native Hawaiian or Other Pacific Islander White
Special Instructional Needs	English Language Learners (ELL) Students with Disabilities (SWD)

IEA-human agreement indices were calculated for all cases with an IEA score and at least one human score. Human-human agreement was calculated for all cases with two human scores.

To evaluate the training of IEA for subgroups, we applied the following criteria approved by the state leads for subgroups with at least 50 IEA-human scores and at least 50 human-human scores:

- Pearson correlation between IEA-human should be within 0.1 of human-human
- Kappa between IEA-human should be within 0.1 of human-human
- Quadratic-Weighted Kappa between IEA-human should be within 0.1 of human-human
- Exact agreement between IEA-human should be within 5.25% of human-human
- Standardized mean difference between IEA-human should be less than ± 0.15 (this criterion was applied to subgroups with at least 50 IEA-human scores)

Although it was not expected that these criteria would be met for all subgroups for all prompts, If results of the evaluation between IEA and human scoring for subgroups for any prompt indicated that IEA performance persistently failed on the criteria listed above, consideration 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 criteria approved by the State Leads, we also compared the performance of IEA to 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

These targets were not intended to be directly applied in decisions about whether to deploy IEA operationally or not. Such targets may or may not be met by human scoring for any particular prompt

and/or subgroup, and if they are not met by human scoring they are unlikely to be met by IEA scoring. Nevertheless, comparisons to these targets provided additional information about IEA performance (and human scoring) in an absolute sense.

Section 5: Test Taker Characteristics

5.1 Overview of Test Taking Population

Approximately three million students participated in the second operational administration of the PARCC assessments during the 2015–2016 school year in Colorado, Bureau of Indian Education, District of Columbia, Illinois, Maryland, Massachusetts, New Jersey, New Mexico, and Rhode Island. Not all participating states had students testing in all grades. Assessments were administered for English language arts/literacy (ELA/L) in grades 3 through 11; mathematics assessments were administered in grades 3 through 8, as well as for traditional high school mathematics (Algebra I, Geometry, and Algebra II) and integrated high school mathematics (Integrated Mathematics I – III). A small subset of students tested in ELA/L grades 9, 10, and 11, and Algebra I, Geometry, and Algebra II during fall of 2015. Test takers characteristics for this group are presented in an addendum.⁵ The majority of students tested during the Spring 2016 window when all grades and content areas were administered online and on paper.

5.2 Composition of Operational Forms

The fall 2015 operational administration of the PARCC assessment included two separate components: the Performance-Based Assessment (PBA) and the End-of-Year (EOY) assessment. Both components were administered as computer-based tests (CBT) and as paper-based tests (PBT). A valid score in both the PBA and EOY assessments was required for a student to receive a summative score. The spring 2016 operational administration of the PARCC assessment combined the Performance-Based Assessment (PBA) and End-of-Year (EOY) into one testing window.

The PBA and EOY components measured different types of knowledge and skills. The PBA was administered after approximately 75 percent of instructional time was complete. The PBA component consisted of relatively long questions, many of which required multiple steps. The purpose of this component was to measure critical thinking, reasoning, and the ability to apply skills and knowledge in reading, writing, and mathematics. The ELA/L PBA focused on writing effectively when analyzing text. The mathematics PBA focused on applying skills and concepts, and on understanding multistep problems that require abstract reasoning, precision, and perseverance.

The EOY administration occurred after approximately 90 percent of instruction was complete. Students were required to demonstrate their skills and knowledge by answering innovative selected-response and short-answer questions that measured concepts and skills. In the ELA/L EOY component students demonstrated their understanding of literary and informational passages. The mathematics EOY component required students to show their understanding of concepts, procedures, and short applications.

The spring 2016 PARCC assessments were administered in either computer-based or paper-based format. English language arts/literacy (ELA/L) assessments focused on writing effectively when analyzing

⁵ Addendum 5 presents a summary of the test taker characteristics for the Fall 2015 administration.

text. Mathematics assessments focused on applying skills and concepts, and understanding multi-step problems that require abstract reasoning and modeling real-world problems, precision, perseverance, and strategic use of tools. In both content areas, students also demonstrated their acquired skills and knowledge by answering selected response items and fill-in-the-blank questions.

Each spring assessment was comprised of multiple units, and additionally, one of the mathematics units was split into two sections: a non-calculator section and a calculator section.

5.3 Rules for Inclusion of Students in Analyses

Criteria for inclusion of students were implemented prior to all operational analyses. These rules were established by Pearson psychometricians in consultation with PARCC to determine which, if any, student records should be removed from analyses. This data screening process resulted in higher quality, albeit slightly smaller, data sets.

Student response data were included in analyses if:

- 1) Valid form numbers were observed for each unit for online assessments or a form for paper assessments,
- 2) Student records were not flagged as “void” (i.e., do not score), and
- 3) The student attempted at least 25% of the items in each unit or form,

Additionally, in cases where students had more than one valid record, the record with the higher raw score was chosen. Records for students with administration issues or anomalies were excluded from analyses.

5.4 Test Takers by Grade, Mode, and Gender

Table 5.1 presents, for each grade of ELA/L, the number and percentage of students who took the test in each mode (CBT or PBT). This information is provided for all participating states combined. Table 5.2 presents the same type of information for all students who took the mathematics assessments, and Table 5.3 provides this information for students who took the mathematics assessments in Spanish.

Markedly more students tested online than on paper, across all grades for both content areas. For ELA/L the percentages of online test takers, for all states combined, ranged from 79.1% to 95.6% while the percentages of paper test takers ranged from 4.4% to 20.9%. For all mathematics test takers, the percentages of students testing online ranged from 79.1% to 99.5%, whereas the percentages of students testing on paper ranged from 0.5% to 20.9%. The percentages of mathematics online students taking Spanish-language forms ranged from 82% to 100% and the percentages of mathematics students taking paper Spanish-language forms ranged from 1.9% to 18%. Generally, the percentage of students who tested online increased steadily from the lower grades to the higher grades. For example, about 80% of the ELA/L grade 3 students tested online, while 96% of the grade 11 students tested online. Overall, fewer students tested at the higher grades for both content areas.

Table 5.1 ELA/L Test Takers by Grade and Mode: All States Combined

Grade	No. of Valid Cases	CBT		PBT	
		N	%	N	%
3	471,801	373,061	79.1	98,740	20.9
4	461,204	378,410	82.0	82,794	18.0
5	455,980	405,897	89.0	50,083	11.0
6	455,888	403,788	88.6	52,100	11.4
7	449,801	396,461	88.1	53,340	11.9
8	440,160	390,037	88.6	50,123	11.4
9	275,158	260,548	94.7	14,610	5.3
10	192,956	184,539	95.6	8,417	4.4
11	136,934	130,884	95.6	6,050	4.4
Grand Total	3,339,882	2,923,625		416,257	

Note: Includes students taking accommodated forms of ELA/L.

Table 5.2 Mathematics Test Takers by Grade and Mode: All States Combined

Grade	No. of Valid Cases	CBT		PBT	
		N	%	N	%
3	476,620	377,172	79.1	99,448	20.9
4	464,485	380,072	81.8	84,413	18.2
5	458,218	406,685	88.8	51,533	11.2
6	457,815	405,870	88.7	51,945	11.3
7	435,545	383,442	88.0	52,103	12.0
8	359,231	314,746	87.6	44,485	12.4
A1	323,701	304,078	93.9	19,623	6.1
GO	145,270	140,056	96.4	5,214	3.6
A2	139,956	132,009	94.3	7,947	5.7
M1	16,581	16,492	99.5	89	0.5
M2	4,655	4,387	94.2	268	5.8
M3	2,371	2,229	94.0	142	6.0
Grand Total	3,284,448	2,867,238		417,210	

Note: Includes students taking mathematics in English, students taking Spanish-language forms for mathematics, and students taking accommodated forms. A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table 5.3 Spanish-Language Mathematics Test Takers, by Grade and Mode: All States Combined

Grade	No. of Valid Cases	CBT		PBT	
		N	%	N	%
3	5,924	4,857	82.0	1,067	18
4	3,490	3,068	87.9	422	12.1
5	2,743	2,675	97.5	68	2.5
6	2,306	2,221	96.3	85	3.7
7	2,325	2,215	95.3	110	4.7
8	2,119	1,980	93.4	139	6.6
A1	3,556	3,333	93.7	223	6.3
GO	1,813	1,761	97.1	52	2.9
A2	911	894	98.1	17	1.9
M1	84	82	97.6	2	2.4
M2	68	68	100	n/a	n/a
M3	5	5	100	n/a	n/a
Grand Total	25,344	23,159		2,185	

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. n/a = not applicable.

The number and percentage of students with valid test scores in each content area, grade, and mode of assessment are presented, for all states combined and for each state separately, in **Appendix 5** as Tables A.5.1, A.5.2, and A.5.3, for ELA/L test takers, all mathematics test takers, and students taking the Spanish-language mathematics tests, respectively. Table A.5.4 presents the ELA/L distribution by grade, mode, and gender, for all states combined; Tables A.5.5, and A.5.6 present similar information for all mathematics test takers and for students taking the Spanish-language mathematics tests, respectively.

5.5 Demographics

Also presented in **Appendix 5** is student demographic information for the following characteristics: economically disadvantaged, students with disabilities, English learners (EL), gender, and race/ethnicity (American Indian/Alaska Native; Black/African American; Hispanic/Latino; White/Caucasian; Native Hawaiian or Other Pacific Islander; two or more races reported; race not reported). Student demographic information was provided by the states and district and captured in PearsonAccess^{next} by means of a student data upload. The demographic data was verified by the states and district prior to score reporting.

Tables A.5.7 through A.5.15 provide demographic information for students with valid ELA/L scores, and Tables A.5.16 through A.5.27 present demographics for students with valid mathematics scores. All tables of demographic information are organized by grade; the results are first aggregated across all PARCC states and then presented for each state. Percentages are not reported for any states in which fewer than 20 students tested in a grade/content area.

Section 6: Classical Item Analysis

6.1 Overview

This section describes the results of the classical item analysis conducted for data obtained from the operational test items. Item analysis serves two purposes: to inform item exclusion decisions for IRT analysis, and to provide item statistics for the item bank.

PARCC item analysis included data from the following types of items: key-based selected response items, rule-based machine-scored items, and hand-scored constructed response items. For each item, the analysis produced item difficulty, item discrimination, and item response frequencies.

6.2 Data Screening Criteria

Item analyses were conducted by test form based on administration mode. In preparation for item analysis, student response files were processed to verify that the data were free of errors. Pearson Customer Data Quality (CDQ) staff ran predefined checks on all data files and verified that all fields and data needed to perform the statistical analyses were present and within expected ranges.

Before beginning item analysis, Pearson performed the following data screening operations:

1. All records with an invalid form number were excluded.
2. All records that were flagged as “void” were excluded.
3. All records where the student attempted fewer than 25% of items were excluded.
4. For students with more than one valid record, the record with the higher raw score was chosen.
5. Records for students with administration issues or anomalies were excluded.

6.3 Description of Classical Item Analysis Statistics

A set of classical item statistics were computed for each operational item by form and by administration mode. Each statistic was designed to evaluate the performance of each item.

The following statistics and associated flagging rules were used to identify items that were not performing as expected:

- 1. Classical item difficulty indices (*p* value and average item score).** When constructing PARCC tests, a wide range of item difficulties is desired (i.e., from easy to hard items) so that students of all ability levels can be assessed with precision. At the operational stage, item difficulty statistics are used by test developers to build forms that meet desired test difficulty targets. Some of the items proved to be unexpectedly difficult. This may be due to students’ lack of familiarity with the item type or students’ limited opportunity to learn the content represented in the item.

For dichotomously scored items, item difficulty is indicated by its p value, which is the proportion of test takers who answered that item correctly. The range for p values is from .00 to 1.00. Items with high p values are easy items and those with low p values are difficult items. Dichotomously scored items were flagged for review if the p value was above .95 (i.e., too easy) or below .25 (i.e., too difficult).

For polytomously scored items, difficulty is indicated by the average item score (AIS). The AIS can range from .00 to the maximum total possible points for an item. To facilitate interpretation, the AIS values for polytomously scored items are often expressed as percentages of the maximum possible score, which are equivalent to the p values of dichotomously scored items. The desired p value range for polytomously scored items is .30 to .80; items with values outside this range were flagged for review.

2. **The percentage of students choosing each response option.** Selected response items on PARCC assessments refer primarily to single-select multiple choice items. These items require that the test taker select a response from a number of answer options. These statistics for single-select multiple choice items indicate the percentage of students who select each of the answer options and the percentage that omit the item. The percentages are also computed for the high-performing subgroup of students who scored at the top 20% on the assessment. Items were flagged for review if more high-performing test takers chose the incorrect option than the correct response. Such a result could indicate that the item has multiple correct answers or is miskeyed.
3. **Item-total correlation.** This statistic describes the relationship between test takers' performance on a specific item and their performance on the total test. The item-total correlation is usually referred to as the item discrimination index. For PARCC operational item analysis, the total score on the assessment was used as the total test score. The polyserial correlation was calculated for both selected response items and constructed response items as an estimate of the correlation between an observed continuous variable and an unobserved continuous variable hypothesized to underlie the variable with ordered categories (Olsson, Drasgow, and Dorans, 1982). Item-total correlations can range from -1.00 to 1.00. Desired values are positive and larger than .20. Negative item-total correlations indicate that low ability test takers perform better on an item than high ability test takers, an indication that the item may be potentially flawed. Item-total correlations below .20 were flagged for review. Items with extremely low or negative values were considered for exclusion from IRT calibrations or linking (refer to Section 10 for details on item inclusion and exclusion criteria for IRT analyses).
4. **Distractor-total correlation.** For selected response items, this estimate describes the relationship between selecting an incorrect response (i.e., a distractor) for a specific item and performance on the total test. The polyserial correlation is calculated (refer to #3 above) for the distractors. Items with distractor-total correlations above .00 were flagged for review as these items may have multiple correct answers, be miskeyed, or have other content issues.

- 5. Percentage of students omitting or not reaching each item.** For both selected response and constructed response items, this statistic is useful for identifying problems with test features such as testing time and item/test layout. Typically, if students have an adequate amount of testing time, approximately 95 percent of students should attempt to answer each question on the test. A distinction is made between “omit” and “not reached” for items without responses:
- An item is considered “omit” if the student responded to subsequent items.
 - An item is considered “not reached” if the student did not respond to any subsequent items.

Patterns of high omit or not reached rates for items located near the end of a test section may indicate that test takers did not have adequate time. Items with high omit rates were flagged. Omit rates for constructed response items tend to be higher than for selected response items. Therefore, the omit rate for flagging individual items was 5% for selected response items and 15% for constructed response items. If a test taker omitted an item, then the test taker received a score of ‘0’ for that item and was included in the N-count for that item. However, if an item was near the end of the test and classified as not reached, the test taker did not receive a score and was not included in the N-count for that item.

- 6. Distribution of item scores.** For constructed response items, examination of the distribution of scores is helpful to identify how well the item is functioning. If no students’ responses are assigned the highest possible score point, this may indicate that the item is not functioning as expected (e.g., the item could be confusing, poorly worded, or just unexpectedly difficult), the scoring rubric is flawed, and/or test takers did not have an opportunity to learn the content. In addition, if all or most test takers score at the extreme ends of the distribution (e.g., 0 and 2 for a 3-category item), this may indicate that there are problems with the item or the rubric so that test takers can receive either full credit or no credit at all, but not partial credit.

The raw score frequency distributions for constructed response items were computed to identify items with few or no observations at any score points. Items with no observations or a low percentage (i.e., <3%) of test takers obtaining any score point were flagged. In addition, constructed response items were flagged if they had U-shaped distributions, with high frequencies for extreme scores and very low frequencies for middle score categories. Items with such response patterns may pose problems during the IRT calibrations and therefore may need to be excluded (refer to Section 10 for more information).

6.4 Summary of Classical Item Analysis Flagging Criteria

In summary, items are flagged for review if the item analysis yielded any of the following results:

- p value above .95 for dichotomous items and above .80 for polytomous items
- p value below .25 for dichotomous items, and below .30 for polytomous items
- Item-total correlation below .20

4. Any distractor-total correlation above .00
5. Greater number of high-performing students (top 20%) choosing a distractor than the keyed response
6. High percentage of omits: above 5% for selected response items, and above 15% for constructed response items
7. High percentage that did not reach the item: above 5% for selected response items, and above 15% for constructed response items
8. Constructed response items with a score value obtained by less than 3% percent of responses

Pearson's psychometric staff carefully reviewed the flagged items and brought items to the PARCC Priority Alert Task Force to decide if the items were problematic and should be excluded from scoring.

6.5 Classical Item Analysis Results

This section presents tables summarizing the item analysis results for the Spring 2016 operational items.

- Tables 6.1 and 6.2 present p value information by grade and mode for the ELA/L and mathematics operational items from the Spring 2016 operational administration.
- Tables 6.3 and 6.4 present item-total correlations by grade and mode for the ELA/L and mathematics operational items from the Spring 2016 operational administration.

An operational item could appear on multiple test forms. The tables list only unique items in each test mode, and the reported item statistics are based on student responses across multiple occurrences of an item. The Integrate Mathematics I, II, and III paper tests were pre-equated due to small sample sizes and are not included in Tables 6.2 and 6.4.

Spoiled or 'do not score' items were excluded from the total test score in item analysis. These items were removed from scoring because of item performance, technical scoring issues, content concerns, or multiple/no correct answers. Additionally, some items were dropped during item calibrations due to:

- A low weighted polyserial,
- A low p value (e.g., extremely difficult item), or
- Extremely poor IRT model fit or item not able to calibrate.

Tables 10.3 and 10.4 in section 10 present the count and percentage of CBT and PBT items excluded from IRT calibration along with the reasons the items were excluded for ELA/L and mathematics, respectively. The tables in this section and in Addendum 6 include only those items that were used for operational scoring.

The Fall 2015 forms were based on the spring 2015 operational forms; therefore, the item analyses for these forms were reported in the 2014-2015 Technical Report.

Table 6.1 Summary of *p* Values for ELA/L Operational Items by Grade and Mode

Grade	Mode	N of Unique Items	Mean <i>p</i> Value	SD <i>p</i> Value	Min <i>p</i> Value	Max <i>p</i> Value	Median <i>p</i> Value
3	CBT	64	0.43	0.17	0.12	0.80	0.45
3	PBT	52	0.43	0.14	0.18	0.82	0.40
4	CBT	85	0.44	0.14	0.13	0.79	0.42
4	PBT	50	0.40	0.13	0.14	0.70	0.40
5	CBT	84	0.40	0.15	0.14	0.78	0.37
5	PBT	44	0.43	0.13	0.17	0.78	0.44
6	CBT	80	0.42	0.13	0.22	0.70	0.39
6	PBT	56	0.48	0.13	0.27	0.77	0.47
7	CBT	88	0.45	0.14	0.16	0.84	0.44
7	PBT	72	0.48	0.13	0.24	0.80	0.45
8	CBT	84	0.46	0.14	0.18	0.86	0.42
8	PBT	56	0.50	0.14	0.17	0.83	0.49
9	CBT	89	0.40	0.12	0.14	0.69	0.41
9	PBT	66	0.48	0.14	0.18	0.82	0.47
10	CBT	102	0.40	0.12	0.16	0.79	0.37
10	PBT	58	0.43	0.12	0.21	0.69	0.42
11	CBT	78	0.37	0.13	0.13	0.73	0.37
11	PBT	64	0.40	0.12	0.19	0.76	0.40

Note: CBT = computer-based testing (online); PBT = paper-based testing (paper).

Table 6.2 Summary of p Values for Mathematics Operational Items by Grade and Mode

Grade	Mode	N of Unique Items	Mean p Value	SD p Value	Min p Value	Max p Value	Median p Value
3	CBT	97	0.54	0.25	0.12	0.94	0.55
3	PBT	75	0.54	0.22	0.14	0.93	0.55
4	CBT	101	0.54	0.25	0.09	0.95	0.53
4	PBT	65	0.50	0.22	0.11	0.93	0.53
5	CBT	95	0.46	0.21	0.06	0.94	0.46
5	PBT	69	0.49	0.22	0.11	0.91	0.51
6	CBT	101	0.45	0.24	0.06	0.93	0.44
6	PBT	63	0.46	0.21	0.12	0.93	0.43
7	CBT	103	0.34	0.24	0.04	0.88	0.30
7	PBT	65	0.37	0.22	0.07	0.83	0.33
8	CBT	89	0.35	0.22	0.04	0.81	0.33
8	PBT	63	0.39	0.20	0.08	0.82	0.35
A1	CBT	100	0.30	0.18	0.02	0.73	0.29
A1	PBT	69	0.34	0.19	0.02	0.75	0.34
GO	CBT	109	0.33	0.21	0.01	0.90	0.30
GO	PBT	79	0.31	0.19	0.05	0.77	0.30
A2	CBT	101	0.25	0.15	0.05	0.83	0.22
A2	PBT	72	0.29	0.18	0.03	0.83	0.25
M1	CBT	42	0.30	0.19	0.03	0.74	0.34
M2	CBT	41	0.25	0.17	0.01	0.62	0.22
M3	CBT	40	0.28	0.17	0.04	0.82	0.26

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based testing (online); PBT = paper-based testing (paper).

Table 6.3 Summary of Item-total Polyserial Correlations for ELA/L Operational Items by Grade and Mode

Grade	Mode	N of Unique Items	Mean Polyserial	SD Polyserial	Min Polyserial	Max Polyserial	Median Polyserial
3	CBT	64	0.53	0.12	0.21	0.73	0.54
3	PBT	52	0.54	0.12	0.19	0.73	0.53
4	CBT	85	0.48	0.15	0.19	0.79	0.46
4	PBT	50	0.46	0.15	0.18	0.78	0.45
5	CBT	84	0.48	0.16	0.11	0.80	0.45
5	PBT	44	0.46	0.16	0.20	0.76	0.43
6	CBT	80	0.47	0.16	0.16	0.82	0.43
6	PBT	56	0.49	0.14	0.24	0.80	0.49
7	CBT	88	0.48	0.15	0.22	0.84	0.44
7	PBT	72	0.47	0.14	0.18	0.80	0.44
8	CBT	84	0.48	0.17	0.15	0.85	0.46
8	PBT	56	0.48	0.15	0.16	0.81	0.47
9	CBT	89	0.48	0.17	0.12	0.84	0.47
9	PBT	66	0.49	0.13	0.26	0.80	0.49
10	CBT	102	0.49	0.16	0.20	0.84	0.47
10	PBT	58	0.52	0.15	0.24	0.83	0.52
11	CBT	78	0.45	0.18	0.11	0.82	0.42
11	PBT	64	0.45	0.17	0.17	0.82	0.43

Note: CBT = computer-based testing (online); PBT = paper-based testing (paper).

Table 6.4 Summary of Item-total Correlations for Mathematics Operational Items by Grade and Mode

Grade	Mode	N of Unique Items	Mean Polyserial	SD Polyserial	Min Polyserial	Max Polyserial	Median Polyserial
3	CBT	97	0.48	0.13	0.20	0.78	0.50
3	PBT	75	0.50	0.13	0.22	0.82	0.51
4	CBT	101	0.48	0.13	0.22	0.77	0.47
4	PBT	65	0.50	0.13	0.28	0.74	0.52
5	CBT	95	0.48	0.14	0.18	0.80	0.47
5	PBT	69	0.50	0.12	0.26	0.74	0.49
6	CBT	101	0.50	0.13	0.14	0.79	0.49
6	PBT	63	0.52	0.13	0.17	0.74	0.51
7	CBT	103	0.48	0.13	0.21	0.80	0.46
7	PBT	65	0.48	0.14	0.27	0.77	0.45
8	CBT	89	0.46	0.12	0.24	0.75	0.45
8	PBT	63	0.49	0.13	0.22	0.81	0.49
A1	CBT	100	0.46	0.14	0.13	0.79	0.45
A1	PBT	69	0.45	0.13	0.15	0.72	0.44
GO	CBT	109	0.49	0.14	0.20	0.81	0.49
GO	PBT	79	0.47	0.15	0.16	0.81	0.47
A2	CBT	101	0.50	0.14	0.15	0.81	0.51
A2	PBT	72	0.45	0.15	0.10	0.77	0.44
M1	CBT	42	0.44	0.15	0.17	0.79	0.42
M2	CBT	41	0.40	0.13	0.16	0.67	0.38
M3	CBT	40	0.46	0.16	0.20	0.75	0.47

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based testing (online); PBT = paper-based testing (paper)

Section 7: Differential Item Functioning

7.1 Overview

Differential item functioning (DIF) analyses were conducted using the data obtained from the operational items. If an item performs differentially across identifiable subgroups (e.g., gender or ethnicity) when students are matched on ability, the item may be measuring something other than the intended construct (i.e., possible evidence of DIF). It is important, however, to recognize that item performance differences flagged for DIF might be related to actual differences in relevant knowledge or skills (item impact) or statistical Type I error. As a result, DIF statistics are used to identify *potential* item bias. Subsequent reviews by content experts and bias/sensitivity committees are required to determine the source and meaning of performance differences.

7.2 DIF Procedures

Dichotomous Items. The Mantel-Haenszel (MH) DIF statistic was calculated for selected-response items and for dichotomously-scored constructed-response items. In this method, test takers are classified to relevant subgroups of interest (e.g., gender or ethnicity). Using the PARCC raw score total as the criteria, test takers in a certain total score category in the focal group (e.g., females) are compared with examinees in the same total score category in the reference group (e.g., males). For each item, test takers in the focal group are also compared to test takers in the reference group who performed equally well on the test as a whole. The common odds ratio is estimated across all categories of matched test taker ability using the following formula (Dorans & Holland, 1993), and the resulting estimate is interpreted as the relative likelihood of success on a particular item for members of two groups when matched on ability.

$$\hat{\alpha}_{MH} = \frac{\sum_{s=1}^S R_{rs} W_{fs} / N_{ts}}{\sum_{s=1}^S R_{fs} W_{rs} / N_{ts}}, \quad (7-1)$$

in which:

S = the number of score categories,

R_{rs} = the number of test takers in the reference group who answer the item correctly,

W_{fs} = the number of test takers in the focal group who answer the item incorrectly,

R_{fs} = the number of test takers in the focal group who answer the item correctly,

W_{rs} = the number of test takers in the reference group who answer the item incorrectly, and

N_{ts} = the total number of test takers.

To facilitate the interpretation of MH results, the common odds ratio is frequently transformed to the delta scale using the following formula (Holland & Thayer, 1988):

$$MH\ D-DIF = -2.35 \ln (\hat{\alpha}_{MH}) \quad (7-2)$$

Positive values indicate DIF in favor of the focal group (i.e., positive DIF items are differentially easier for the focal group), whereas negative values indicate DIF in favor of the reference group (i.e., negative DIF item are differentially easier for the reference group).

Polytomous Items. For polytomously scored constructed-response items, the MH D-DIF statistic is not calculated; instead the standardization DIF (Dorans & Schmitt, 1991; Zwick, Thayer & Mazzeo, 1997; Dorans, 2013), in conjunction with the Mantel chi-square statistic (Mantel, 1963; Mantel & Haenszel, 1959), is used to identify items with DIF.

The standardization DIF compares the item means of the two groups after adjusting for differences in the distribution of test takers across the values of the matching variable (i.e., total test score) and is calculated using the following formula:

$$STD - EISDIF = \frac{\sum_{s=1}^S N_{fs} * E_f(Y|X = s)}{\sum_{s=1}^S N_{fs}} - \frac{\sum_{s=1}^S N_{rs} * E_r(Y|X = s)}{\sum_{s=1}^S N_{rs}} = \frac{\sum_{s=1}^S D_s}{\sum_{s=1}^S N_{fs}}, \quad (7-3)$$

in which:

X = the total score

Y = the item score

S = the number of score categories on X ,

N_{rs} = the number of test takers in the reference group in score category s ,

N_{fs} = the number of test takers in the focal group in score category s ,

E_r = the expected item score for reference group, and

E_f = the expected item score for reference group.

A positive *STD-EISDIF* value means that, conditional on the total test score, the focal group has a higher mean item score than the focal group. In contrast, a negative *STD-EISDIF* value means that, conditional on the total test score, the focal group has a lower mean item score than the reference group.

Classification. Based on the DIF statistics and significance tests, items are classified into three categories and assigned values of A, B, or C (Zieky, 1993). Category A items contain negligible DIF, Category B items exhibit slight to moderate DIF, and Category C items possess moderate to large DIF values. Positive values indicate that, conditional on the total score, the focal group has a higher mean item score than the reference group. In contrast, negative DIF values indicate that, conditional on the total test score, the focal group has a lower mean item score than the reference group. The flagging criteria for dichotomously scored items are presented in Table 7.1; the flagging criteria for polytomously scored constructed response items are provided in Table 7.2.

Table 7.1 DIF Categories for Dichotomous Selected Response and Constructed Response Items

DIF Category	Criteria
A (negligible)	Absolute value of the MH D-DIF is not significantly different from zero, or is less than one.
B (slight to moderate)	1. Absolute value of the MH D-DIF is significantly different from zero but not from one, and is at least one; OR 2. Absolute value of the MH D-DIF is significantly different from one, but is less than 1.5. Positive values are classified as “B+” and negative values as “B-”.
C (moderate to large)	Absolute value of the MH D-DIF is significantly different from one, and is at least 1.5. Positive values are classified as “C+” and negative values as “C-”.

Table 7.2 DIF Categories for Polytomous Constructed Response Items

DIF Category	Criteria
A (negligible)	Mantel Chi-square <i>p value</i> > 0.05 or $ STD-EISDIF/SD \leq 0.17$
B (slight to moderate)	Mantel Chi-square <i>p value</i> < 0.05 and $ STD-EISDIF/SD > 0.17$
C (moderate to large)	Mantel Chi-square <i>p value</i> < 0.05 and $ STD-EISDIF/SD > 0.25$

Note: *STD-EISDIF* = standardized DIF; *SD* = total group standard deviation of item score.

7.3 Operational Analysis DIF Comparison Groups

Traditional Comparisons. DIF analyses were conducted on each test form for designated comparison groups defined on the basis of demographic variables including: gender, race/ethnicity, economic disadvantage, and special instructional needs such as students with disabilities (SWD) or English learners (EL). Student demographic information was provided by the states and district and captured in PearsonAccess by means of a student data upload. The demographic data was verified by the states and district prior to score reporting. These comparison groups are specified in Table 7.3.

Table 7.3 Traditional DIF Comparison Groups

Grouping Variable	Focal Group	Reference Group
Gender	Female	Male
Ethnicity	American Indian/Alaska Native (AmerIndian)	White
	Asian	White
	Black or African American	White
	Hispanic/Latino	White
	Native Hawaiian or Pacific Islander	White
	Multiple Race Selected	White
Economic Status*	Economically Disadvantaged (EcnDis)	Not Economically Disadvantaged (NoEcnDis)
Special Instructional Needs	English Learner (ELY)	Non English Learner (ELN)
	Students with Disabilities (SWDY)	Students without Disabilities (SWDN)

Note: * Economic status was based on participation in National School Lunch Program: receipt of free or reduced-price lunch.

Comparison across Languages. DIF analyses were also conducted for Spanish-language items vs. English-language items in mathematics for items that previously were not evaluated for Spanish-language DIF. The purpose of the Spanish vs. English DIF analysis was to evaluate how similarly the items functioned between the two languages because the data from the Spanish-language forms were not separately calibrated using IRT. The item parameter estimates based on the English speaking test takers were used to generate conversion tables for the Spanish-language forms. Spanish-language mathematics items flagged for C-DIF were reviewed by content specialists and the PARCC Priority Alert Task Force to decide if the items were problematic and should be excluded from scoring. An item could be dropped from a Spanish-language form but remain in the English-language form if no other issues were detected; in those cases separate conversion tables were generated for the two versions of the form which had different numbers of items.

The Spanish-language forms did not have a non-accommodated English-language form counterpart with the same set of items (refer to Section 2 for more information on the development of Spanish-language forms). Most of the Spanish-language items were previously evaluated for Spanish-language DIF. For items that had not been evaluated for Spanish-language DIF, the analyses were conducted for items that had an English-language item on a non-accommodated test form.

Sample Size Requirement. DIF analyses were conducted when the following sample size requirements were met:

- The smaller group, reference or focal, had at least 100 students, and
- The combined group, reference and focal, had at least 400 students.

Spanish-language vs. English-language DIF analyses were not conducted for Integrated Mathematics I, II and III because of insufficient sample sizes.

7.4 Operational Differential Item Functioning Results

Appendix 7 presents tables summarizing the DIF results for the Spring 2016 operational items, with one table prepared for each content and grade level (e.g., ELA/L Grade 3). The Fall 2015 forms were based on spring 2015 operational forms. The DIF analyses for these forms are reported in the 2014-2015 Technical Report.

Spoiled or ‘do not score’ items were excluded from the total test score for each form in DIF analysis. These items were removed from scoring because of item performance, technical scoring issues, content concerns, multiple correct answers, or no correct answers. However, the tables in this section may include items for certain grade levels that were excluded from scoring based on later analyses (refer to Section 10.5 Items Excluded from Score Reporting for more information).

In the DIF results tables, the column “DIF Comparisons” identifies the focal and reference groups for the analysis performed; the column “Mode” identifies the test delivery mode. “Total N of Unique Items” reports the number of unique items included in the analysis, whereas “Total N of Item Occurrences” reports the number of times items were used on test forms. An item could be used in multiple test forms; therefore, items were counted according to the occurrences. For example, if the same item appeared in five test forms, it was counted as five occurrences; if this item was classified as B+ on one form and C+ on another form, both occurrences were reported in the corresponding columns. For the Spanish-language DIF, “Total N of Item Occurrences” reports the number of items previously not analyzed for Spanish-language DIF and a non-accommodated English-language item was available in spring 2016. “Total N of Item Occurrences Included in DIF Analysis” reports the number of occurrences with sufficient sample sizes to be included in DIF analyses. In addition, “0” indicates that the DIF analysis did not classify any items in the particular DIF category, while “n/a” indicates that the DIF analysis was not performed due to insufficient sample sizes.

Table 7.4 Differential Item Functioning for ELA/L Grade 3

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	71	105	105			2	2	101	96	2	2
	PBT	57	70	70			1	1	65	93	4	6
White vs AmerIndian	CBT	71	105	105			1	1	104	99		
	PBT	57	70	70	1	1	8	11	61	87		
White vs Asian	CBT	71	105	105			1	1	102	97	2	2
	PBT	57	70	70			1	1	67	96	2	3
White vs Black	CBT	71	105	105			2	2	103	98		
	PBT	57	70	70			4	6	66	94		
White vs Hispanic	CBT	71	105	105	1	1	2	2	102	97		
	PBT	57	70	70	1	1	7	10	62	89		
White vs Pacific Islander	CBT	71	105	105					105	100		
	PBT	57	70	52	1	2	2	4	49	94		
White vs Multiracial	CBT	71	105	105					105	100		
	PBT	57	70	70					70	100		
NoEcnDis vs EcnDis	CBT	71	105	105					105	100		
	PBT	57	70	70			3	4	67	96		
ELN vs ELY	CBT	71	105	105	1	1	2	2	102	97		
	PBT	57	70	70	3	4	3	4	64	91		
SWDN vs SWDY	CBT	71	105	105					105	100		
	PBT	57	70	70					70	100		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability.

Table 7.5 Differential Item Functioning for Mathematics Grade 3

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	97	129	129	1	1	4	3	124	96				
	PBT	75	86	86			1	1	85	99				
White vs AmerIndian	CBT	97	129	129			2	2	127	98				
	PBT	75	86	86	3	3	7	8	74	86	2	2		
White vs Asian	CBT	97	129	129					111	86	16	12	2	2
	PBT	75	86	86			1	1	82	95	3	3		
White vs Black	CBT	97	129	129			2	2	124	96	3	2		
	PBT	75	86	86			7	8	78	91	1	1		
White vs Hispanic	CBT	97	129	129			1	1	128	99				
	PBT	75	86	86	1	1	7	8	76	88	2	2		
White vs Pacific Islander	CBT	97	129	129					127	98	2	2		
	PBT	75	86	40	1	3	1	3	33	83	5	13		
White vs Multiracial	CBT	97	129	129					129	100				
	PBT	75	86	86					86	100				
NoEcnDis vs EcnDis	CBT	97	129	129			1	1	128	99				
	PBT	75	86	86	1	1	3	3	80	93	1	1	1	1
ELN vs ELY	CBT	97	129	129			3	2	126	98				
	PBT	75	86	86			4	5	82	95				
SWDN vs SWDY	CBT	97	129	129			4	3	125	97				
	PBT	75	86	86			1	1	84	98	1	1		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Section 8: Reliability

8.1 Overview

Reliability focuses on the extent to which differences in test scores reflect true differences in the knowledge, ability, or skill being tested rather than fluctuations due to chance. Thus, reliability measures the consistency of the scores across conditions that can be assumed to differ at random, especially which form of the test the test taker is administered and which persons are assigned to score responses to constructed-response questions. In statistical terms, the variance in the distributions of test scores, essentially the differences among individuals, is partly due to real differences in the knowledge, skill, or ability being tested (true variance) and partly due to random errors in the measurement process (error variance). Reliability is an estimate of the proportion of the total variance that is true variance.

There are several different ways of estimating reliability. The type of reliability estimate reported here is an internal-consistency measure, which is derived from analysis of the consistency of the performance of individuals across items within a test. It is used because it serves as a good estimate of alternate forms reliability, but it does not take into account form-to-form variation due to lack of test form parallelism, nor is it responsive to day-to-day variation due to, for example, the examinee's state of health or the testing environment.

Reliability coefficients range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely individuals would be to obtain very similar scores upon repeated testing occasions, if the students do not change in their level of the knowledge or skills measured by the test. The reliability estimates in the tables to follow attempt to answer the question, "How consistent would the scores of these test takers be over replications of the entire testing process?"

Reliability of classification estimates the proportion of students who are accurately classified into proficiency levels. There are two kinds of classification reliability statistics: decision accuracy and decision consistency. Decision accuracy is the agreement between the classifications actually made and the classifications that would be made if the test scores were perfectly reliable. Decision consistency is the agreement between the classifications that would be made on two independent different forms of the test.

Another index is inter-rater reliability for the human scored constructed-response items, which measures the agreement between individual raters (scorers). The inter-rater reliability coefficient answers the question, "How consistent would the scores of these test takers be over replication of scoring of the same responses by different scorers?"

Standard error of measurement (SEM) quantifies the amount of error in the test scores. SEM is the extent by which test takers' scores tend to differ from the scores they would receive if the test were perfectly reliable. As the SEM increases, the variability of student's observed scores is likely to increase across repeated testing. Observed scores with large SEMs pose a challenge to the valid interpretation of a single test score.

Reliability and SEM estimates were calculated at the full assessment level (both PBT and CBT), and at the claim and subclaim levels. In addition, conditional SEMs were calculated and reported in Section 12 and Appendix 12.4.

8.2 Reliability and SEM Estimation

Coefficient alpha (Cronbach, 1951), which measures internal consistency reliability, is the most commonly used measure of reliability. Coefficient alpha is estimated by substituting sample estimates for the parameters in the formula below:

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_x^2} \right], \quad (8-1)$$

where n is the number of items, σ_i^2 is the variance of scores on the i -th item, and σ_x^2 is the variance of the total score (sum of scores on the individual items). Other things being equal, the more items a test includes, the higher the internal consistency reliability.

Since PARCC test forms have mixed item types (dichotomous and polytomous items), it is more appropriate to report stratified alpha (Feldt & Brennan, 1989). Stratified alpha is a weighted average of coefficient alphas for item sets with different maximum score points or “strata.” Stratified alpha is a reliability estimate computed by dividing the test into parts (“strata”), computing alpha separately for each part, and using the results to estimate a reliability coefficient for the total score. Stratified alpha is used here because different parts of the test consist of different item types and may measure different skills. The formula for the stratified alpha is:

$$\rho_{strata} = 1 - \frac{\sum_{j=1}^J \sigma_{x_j}^2 (1 - \alpha_j)}{\sigma_x^2} \quad (8-2)$$

where $\sigma_{x_j}^2$ is the variance for part j of the test, σ_x^2 is the variance of the total scores, and α_j is coefficient alpha for part j of the test. Estimates of stratified alpha are computed by substituting sample estimates for the parameters in the formula. The average stratified alpha is a weighted average of the stratified alphas across the test forms.

The formula for the standard error of measurement is:

$$\sigma_E = \sigma_x \sqrt{1 - \rho_{XX'}} \quad (8-3)$$

where σ_x is the standard deviation of the test score, either total raw score or scale scores, and $\rho_{XX'}$ is the reliability estimated by substitution of appropriate statistics for the parameters in equation 8-1 or 8-2.

8.3 Reliability Results for Total Group

Tables 8.1 and Table 8.2 summarize test reliability estimates for the total testing group for English language arts/literacy (ELA/L) and mathematics, respectively. The section includes only Spring 2016 results. The Fall 2015 results are located in the Addendum.⁶ The tables provide the average reliability, which is estimated by averaging the internal consistency estimates computed for all the individual forms of the test, and both the raw score and scale score SEMs, separately for the computer-based and paper-based tests within each grade level. In addition, the number of forms, the total sample size across all forms, and the average maximum possible score for each set of tests are provided. Estimates were calculated only for groups of 100 or more students administered a specific test form.

English Language Arts/Literacy

The average reliability estimates for the CBT tests for grades 3-11 English language arts/literacy (ELA/L) range from a low of .91 to a high of .93. The average reliability estimates for the PBT tests for ELA/L grades 3-11 range from a low of .89 to a high of .94. The tests for grades 3-5 have fewer maximum possible points than for the grades 6-11 tests. The average reliability estimates are at least .90 except for grades 4 and 5 PBT tests which are .89.

The average raw score SEM is consistently between a very reasonable 4% and 6% of the maximum possible score. The scale score SEMs are the lowest for grade 6 and the highest for grade 3. Across the nine grade levels, the raw score SEMs and scale score SEMs for the PBT assessments are higher than those for the CBT assessments except for grade 10.

⁶ Addendum 8 provides a summary of reliability information for the Fall 2015 administration.

Table 8.1 Summary of ELA/L Test Reliability Estimates for Total Group

Grade Level	Testing Mode	Number of Forms	Total Sample Size	Average Maximum Possible Score	Average Reliability	Average Raw Score SEM	Average Scale Score SEM
3	CBT	5	371,885	93	0.91	5.21	12.02
	PBT	3	98,738	94	0.91	5.46	12.24
4	CBT	5	377,002	106	0.91	5.78	10.59
	PBT	3	82,792	106	0.89	6.25	11.47
5	CBT	5	404,383	106	0.91	5.56	9.83
	PBT	3	50,081	106	0.89	6.07	10.37
6	CBT	5	402,155	121	0.92	6.28	8.79
	PBT	3	52,096	121	0.92	6.72	8.75
7	CBT	5	395,258	121	0.93	6.37	9.57
	PBT	3	53,335	121	0.92	6.92	10.97
8	CBT	5	388,964	121	0.93	6.43	10.05
	PBT	3	50,121	121	0.92	6.76	10.72
9	CBT	6	259,459	121	0.93	5.97	9.33
	PBT	3	14,606	121	0.92	6.66	10.70
10	CBT	6	183,504	121	0.93	6.24	11.96
	PBT	3	8,407	121	0.94	6.54	11.95
11	CBT	6	129,937	121	0.92	6.17	10.89
	PBT	3	6,045	121	0.91	6.55	12.12

Note: ELA grade 3 CBT tests have a lower average maximum possible score due to a spoiled item.

Mathematics

The average reliability estimates for the grades 3-8 mathematics and end-of-course (EOC) assessments range from .86 to .93 for the CBT tests and from .75 to .93 for the PBT tests. Most of the average reliability estimates are above .90 except for some of the Integrated Mathematics tests. Integrated Mathematics I for PBT did not have sufficient sample sizes per form to estimate reliability.

The SEM as percentage of total score consistently range from 4% to 5% of the maximum score. The SEMs for the scale scores are the highest for Integrated Mathematics I and III and grade 8 and the lowest for geometry and grades 6 and 7. The PBT scale score SEMs are within one scale score point of the CBT scale score SEMs.

Table 8.2 Summary of Mathematics Test Reliability Estimates for Total Group

Grade Level	Testing Mode	Number of Forms	Total Sample Size	Average Maximum Possible Score	Average Reliability	Average Raw Score SEM	Average Scale Score SEM
3	CBT	7	375,519	66	0.93	3.46	9.03
	PBT	4	99,447	66	0.93	3.63	9.34
4	CBT	7	378,225	66	0.93	3.35	8.34
	PBT	4	84,410	66	0.93	3.56	8.88
5	CBT	7	405,033	66	0.92	3.56	8.64
	PBT	3	51,463	66	0.93	3.55	8.50
6	CBT	7	404,238	66	0.93	3.51	8.15
	PBT	3	51,856	66	0.93	3.61	8.16
7	CBT	7	382,190	66	0.92	3.34	8.19
	PBT	4	52,101	66	0.92	3.55	8.03
8	CBT	7	314,017	66	0.91	3.25	11.07
	PBT	4	44,484	66	0.91	3.58	12.11
A1	CBT	7	301,139	81	0.91	3.91	10.19
	PBT	4	19,605	81	0.92	3.77	10.28
GO	CBT	6	138,781	81	0.93	3.47	7.12
	PBT	3	5,156	81	0.93	3.65	7.35
A2	CBT	6	130,338	81	0.93	3.62	10.53
	PBT	2	7,839	81	0.91	3.86	11.52
M1	CBT	2	16,275	81	0.90	3.45	10.65
	PBT						
M2	CBT	2	4,313	80	0.86	3.27	10.84
	PBT	1	266	80	0.84	3.56	10.33
M3	CBT	1	2,142	81	0.92	3.79	11.36
	PBT	1	114	80	0.75	3.25	13.32

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. M1 sample size was insufficient to report the results. A2 PBT and M3 PBT tests have lower average maximum possible scores due to a spoiled item.

8.4 Reliability Results for Subgroups of Interest

When sample size was sufficient, score reliability and scale score SEM were estimated for the groups identified for DIF analysis. Estimates were calculated only for groups of 100 or more students administered a specific test form.

Tables 8.3 through 8.11 summarize test reliability for groups of interest for English Language Arts/Literacy grades 3-11, and Tables 8.12 through 8.23 summarize test reliability for groups of interest

for mathematics grades/subjects. Note that reliability estimates are dependent on score variance, and subgroups with smaller variance are likely to have lower reliability estimates than the total group.

Gender

English Language Arts/Literacy

Both the average reliability estimates and average SEMs for males and females are similar to the corresponding reliabilities and SEMs for the total group. Eleven of the eighteen reliabilities are .01 higher for males than for females. The SEMs for females are all higher than for males.

Mathematics

As with the English Language Arts/Literacy test components, the reliability estimates and SEMs for males and females reflect the corresponding reliabilities and SEMs for the total group. Typically, the reliabilities are .01 higher for males than for females. The SEMs are generally very similar for females and males.

Ethnicity

English Language Arts/Literacy

The majority of the reliabilities for the ethnicity groups are .01-.02 lower than for the total group. There is not a consistent difference among the test reliabilities for White, African-American, Asian/Pacific Islander, Hispanic, and multiple ethnicity students, with the majority of the reliabilities between .89 and .94. However, the majority of the reliabilities for American Indian/Alaskan Native students range from .88 to .91. In general, the SEMs are similar to the total group SEMs. Nevertheless, for most grade levels, the SEMs are highest for Asian/Pacific Islander students.

Mathematics

As with the English Language Arts/Literacy reliabilities, the reliabilities for ethnicity groups are marginally lower than for the total group of students. Once again the average SEMs reflect the total group SEMs. While there is variation across tests, the average reliabilities are generally highest for Asian/Pacific Islander students. The American Indian/Alaskan Native groups and the African-American group has the lowest reliabilities.

Special Education Needs

English Arts/Literacy

The reliabilities for five groups of students (Economically Disadvantaged, Not Economically Disadvantaged, Non English Learner, Students with Disabilities, and Students without Disabilities) are generally .01 to .02 less than those for the total group of students. The majority of the reliabilities range from .88 to .93. The average reliabilities for English Learner students are lower, ranging from .84 to .88.

The SEMs are generally similar to the total group SEMs, however, for 16 of the 18 sets of SEMs, the lowest SEM is for Students with Disabilities.

Mathematics

The average reliabilities for the larger student groups (Not Economically Disadvantaged, Non English Learner, and Students without Disabilities) are quite similar to the total group of students. For Economically Disadvantaged, English Learners, and Students with Disabilities, the average reliabilities average .04-.05 lower than those for the total group. The Economically Disadvantaged group has the highest reliabilities and lowest SEMs on average. The English Learner group has the lowest reliabilities and highest SEMs on average.

Students Taking Accommodated Forms

English Arts/Literacy

Two of the four accommodation form types had sufficient sample sizes to allow for estimation of reliability and SEM. The other two groups did not have at least 100 students take any specific form. Within grades, the reliabilities of the Closed Caption forms, which range from .92 to .95, are higher than the average reliabilities for total group. For the Text-to-Speech forms, the reliabilities, which range from .84 to .88, are lower than for the total group reliabilities.

Mathematics

Only the Text-to-Speech forms had sufficient sample sizes for reliability and SEM estimation. With the exception of the Integrated Mathematics I, II and III courses, the Text-to-Speech reliabilities are very close to the total group reliabilities. The corresponding SEMs were somewhat higher than those for the total group SEMs.

Students Taking Translated Forms

Mathematics

With the exception of Integrated Mathematics I, II and III, there were sufficient numbers of students taking the Spanish Language form for reliability and SEM estimation. The reliabilities average .14 less than for the total group, with the largest differences being for grades 7 and 8. The corresponding SEMs are generally higher for the students administered the Spanish language forms. The moderate to high correlations suggest the translated forms are sufficient for individual student reporting.

Table 8.3 Summary of Test Reliability Estimates for Subgroups: Grade 3 ELA/L

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	371,885	93	0.91	12.02	98,738	94	0.91	12.24
Gender								
Male	189,541	93	0.91	11.76	50,305	94	0.91	11.98
Female	182,210	93	0.91	12.27	48,433	94	0.91	12.50
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	174,764	93	0.89	12.09	36,973	94	0.90	12.22
African American	58,106	93	0.90	11.77	20,440	94	0.89	12.23
Asian/Pacific Islander	22,624	93	0.89	12.90	5,128	94	0.89	12.76
American Indian/Alaska Native	3,546	93	0.89	11.39	1,765	94	0.87	12.17
Hispanic	99,656	93	0.90	11.79	30,995	94	0.89	12.20
Multiple	12,624	93	0.91	11.98	3,126	94	0.91	12.42
Special Instructional Needs								
Economically Disadvantaged	171,175	93	0.89	11.72	56,791	94	0.89	12.19
Not-economically Disadvantaged	196,369	93	0.89	12.26	41,104	94	0.89	12.37
English Learner (EL)	47,433	93	0.88	11.60	20,930	94	0.87	12.25
Non English Learner	318,219	93	0.90	12.09	77,523	94	0.91	12.27
Students with Disabilities (SWD)	49,831	93	0.91	11.31	13,872	93	0.90	11.84
Students without Disabilities	206,901	93	0.90	12.21	37,583	94	0.90	12.24
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	204	94	0.92	12.13	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	8,147	94	0.85	12.03	-	-	-	-

Table 8.4 Summary of Test Reliability Estimates for Subgroups: Grade 4 ELA/L

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	377,002	106	0.91	10.59	82,792	106	0.89	11.47
Gender								
Male	192,184	106	0.91	10.45	42,126	106	0.89	11.26
Female	184,715	106	0.90	10.70	40,666	106	0.88	11.63
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	179,996	106	0.89	10.51	31,237	106	0.88	11.31
African American	57,342	106	0.89	10.63	16,946	106	0.86	11.67
Asian/Pacific Islander	23,244	106	0.89	11.03	4,214	106	0.88	11.63
American Indian/Alaska Native	3,688	106	0.89	10.36	1,594	106	0.83	12.30
Hispanic	100,111	106	0.89	10.63	26,187	106	0.86	11.53
Multiple	12,263	106	0.91	10.51	2,351	106	0.90	11.50
Special Instructional Needs								
Economically Disadvantaged	170,854	106	0.89	10.61	47,039	106	0.86	11.61
Not-economically Disadvantaged	201,717	106	0.89	10.60	34,928	106	0.88	11.37
English Learner (EL)	31,107	106	0.86	10.64	10,791	106	0.83	11.92
Non English Learner	339,651	106	0.90	10.58	71,743	106	0.88	11.42
Students with Disabilities (SWD)	55,923	106	0.91	10.31	13,933	106	0.88	11.13
Students without Disabilities	204,686	106	0.90	10.66	29,891	106	0.88	11.33
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	256	106	0.92	10.66	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	9,492	106	0.84	10.43	-	-	-	-

Table 8.5 Summary of Test Reliability Estimates for Subgroups: Grade 5 ELA/L

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	404,383	106	0.91	9.83	50,081	106	0.89	10.37
Gender								
Male	206,355	106	0.91	9.66	25,720	106	0.89	10.19
Female	197,962	106	0.91	9.98	24,361	106	0.88	10.55
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	186,556	106	0.90	9.73	25,127	106	0.88	10.27
African American	65,802	106	0.89	9.87	6,433	106	0.87	10.46
Asian/Pacific Islander	24,634	106	0.90	10.09	2,954	106	0.88	10.54
American Indian/Alaska Native	3,587	106	0.88	9.72	1,480	106	0.84	10.80
Hispanic	111,190	106	0.89	9.95	12,142	106	0.86	10.51
Multiple	12,156	106	0.91	9.75	1,731	106	0.89	10.30
Special Instructional Needs								
Economically Disadvantaged	188,495	106	0.88	9.94	21,888	106	0.87	10.49
Not-economically Disadvantaged	211,619	106	0.90	9.78	27,252	106	0.88	10.32
English Learner (EL)	26,430	106	0.84	10.02	5,441	106	0.83	10.90
Non English Learner	371,582	106	0.90	9.80	44,506	106	0.88	10.33
Students with Disabilities (SWD)	61,747	106	0.90	9.44	10,735	106	0.88	10.23
Students without Disabilities	199,926	106	0.90	9.91	28,683	106	0.87	10.33
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	331	106	0.94	8.65	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	9,972	106	0.87	8.81	-	-	-	-

Table 8.6 Summary of Test Reliability Estimates for Subgroups: Grade 6 ELA/L

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	402,155	121	0.92	8.79	52,096	121	0.92	8.75
Gender								
Male	205,544	121	0.92	8.68	26,754	121	0.92	8.55
Female	196,559	121	0.92	8.87	25,342	121	0.91	8.93
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	189,660	121	0.91	8.69	25,491	121	0.91	8.63
African American	64,711	121	0.91	8.92	7,696	121	0.91	8.84
Asian/Pacific Islander	24,305	121	0.92	8.89	3,329	121	0.92	8.97
American Indian/Alaska Native	3,604	121	0.89	8.84	1,352	121	0.87	9.54
Hispanic	108,122	121	0.91	8.92	12,153	121	0.91	8.77
Multiple	11,339	121	0.92	8.76	1,853	121	0.92	8.69
Special Instructional Needs								
Economically Disadvantaged	181,767	121	0.90	8.93	22,590	121	0.91	8.84
Not-economically Disadvantaged	215,759	121	0.91	8.72	28,728	121	0.91	8.71
English Learner (EL)	21,540	121	0.86	9.20	4,582	121	0.87	9.32
Non English Learner	374,639	121	0.92	8.76	47,327	121	0.92	8.71
Students with Disabilities (SWD)	62,422	121	0.91	8.70	10,828	121	0.91	8.68
Students without Disabilities	196,486	121	0.92	8.80	29,585	121	0.91	8.72
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	362	121	0.94	8.55	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	9,491	121	0.88	8.62	-	-	-	-

Table 8.7 Summary of Test Reliability Estimates for Subgroups: Grade 7 ELA/L

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	395,258	121	0.93	9.57	53,335	121	0.92	10.97
Gender								
Male	202,017	121	0.93	9.44	27,311	121	0.92	10.80
Female	193,089	121	0.93	9.69	26,024	121	0.91	11.14
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	187,804	121	0.92	9.48	26,445	121	0.91	10.82
African American	63,924	121	0.92	9.69	7,761	121	0.90	11.16
Asian/Pacific Islander	24,000	121	0.93	9.64	3,451	121	0.91	10.96
American Indian/Alaska Native	3,680	121	0.91	9.62	1,065	121	0.86	11.86
Hispanic	104,904	121	0.92	9.71	12,602	121	0.90	11.11
Multiple	10,550	121	0.93	9.49	1,807	121	0.92	10.98
Special Instructional Needs								
Economically Disadvantaged	174,771	121	0.91	9.72	22,126	121	0.89	11.16
Not-economically Disadvantaged	215,769	121	0.93	9.51	30,409	121	0.91	10.89
English Learner (EL)	20,935	121	0.87	10.00	4,422	121	0.85	11.61
Non English Learner	368,716	121	0.93	9.54	48,719	121	0.91	10.93
Students with Disabilities (SWD)	61,370	121	0.92	9.54	10,782	121	0.91	10.90
Students without Disabilities	194,254	121	0.93	9.56	30,723	121	0.91	10.92
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	330	121	0.95	9.73	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	8,719	121	0.87	10.25	-	-	-	-

Table 8.8 Summary of Test Reliability Estimates for Subgroups: Grade 8 ELA/L

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	388,964	121	0.93	10.05	50,121	121	0.92	10.72
Gender								
Male	199,370	121	0.93	9.93	25,582	121	0.92	10.47
Female	189,539	121	0.92	10.13	24,539	121	0.91	10.96
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	186,013	121	0.92	9.98	25,616	121	0.91	10.68
African American	64,138	121	0.91	10.15	7,056	121	0.91	10.62
Asian/Pacific Islander	23,023	121	0.92	10.21	3,108	121	0.91	11.18
American Indian/Alaska Native	3,715	121	0.90	10.11	1,051	121	0.87	10.90
Hispanic	101,788	121	0.91	10.15	11,565	121	0.91	10.62
Multiple	9,867	121	0.93	9.97	1,541	121	0.93	10.78
Special Instructional Needs								
Economically Disadvantaged	170,454	121	0.91	10.13	20,110	121	0.91	10.64
Not-economically Disadvantaged	213,685	121	0.92	10.03	29,155	121	0.91	10.78
English Learner (EL)	21,111	121	0.87	10.41	4,248	121	0.87	10.96
Non English Learner	362,665	121	0.92	10.03	45,715	121	0.91	10.71
Students with Disabilities (SWD)	59,625	121	0.91	10.00	10,084	121	0.91	10.35
Students without Disabilities	190,453	121	0.92	10.11	31,359	121	0.91	10.86
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	370	121	0.95	9.77	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	7,750	121	0.88	10.29	-	-	-	-

Table 8.9 Summary of Test Reliability Estimates for Subgroups: Grade 9 ELA/L

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	259,459	121	0.93	9.33	14,606	121	0.92	10.70
Gender								
Male	132,978	121	0.93	9.22	7,390	121	0.92	10.51
Female	126,369	121	0.93	9.40	7,216	121	0.91	10.85
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	115,923	121	0.93	9.25	8,362	121	0.91	10.68
African American	36,248	121	0.92	9.44	1,035	121	0.90	10.60
Asian/Pacific Islander	16,357	121	0.94	9.43	1,045	121	0.91	10.93
American Indian/Alaska Native	3,302	121	0.90	9.35	203	121	0.85	11.18
Hispanic	81,648	121	0.92	9.44	3,281	121	0.91	10.52
Multiple	5,292	121	0.94	9.27	420	121	0.92	10.64
Special Instructional Needs								
Economically Disadvantaged	112,710	121	0.92	9.45	4,392	121	0.91	10.61
Not-economically Disadvantaged	141,435	121	0.93	9.30	9,643	121	0.91	10.62
English Learner (EL)	18,073	121	0.86	9.85	714	121	0.87	10.39
Non English Learner	236,609	121	0.93	9.30	13,677	121	0.91	10.70
Students with Disabilities (SWD)	37,254	121	0.91	9.23	2,139	121	0.92	10.37
Students without Disabilities	124,682	121	0.93	9.39	2,334	121	0.92	10.96
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	196	121	0.94	9.43	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	2,129	121	0.84	9.99	-	-	-	-

Table 8.10 Summary of Test Reliability Estimates for Subgroups: Grade 10 ELA/L

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	183,504	121	0.93	11.96	8,407	121	0.94	11.95
Gender								
Male	93,553	121	0.93	11.67	4,374	121	0.94	11.72
Female	89,841	121	0.93	12.19	3,888	121	0.93	12.22
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	77,904	121	0.93	12.03	4,643	121	0.93	11.84
African American	37,978	121	0.92	11.89	1,602	121	0.92	12.22
Asian/Pacific Islander	13,758	121	0.93	12.28	263	121	0.93	11.98
American Indian/Alaska Native	2,801	121	0.90	11.80	103	121	0.90	11.71
Hispanic	47,137	121	0.92	11.89	1,159	121	0.93	12.13
Multiple	3,458	121	0.94	11.86	291	121	0.93	12.10
Special Instructional Needs								
Economically Disadvantaged	71,859	121	0.92	11.85	3,223	121	0.92	12.12
Not-economically Disadvantaged	106,854	121	0.93	12.08	4,386	121	0.93	11.78
English Learner (EL)	9,297	121	0.87	11.35	256	121	0.83	12.99
Non English Learner	173,554	121	0.93	11.98	7,992	121	0.93	11.94
Students with Disabilities (SWD)	28,424	121	0.91	11.42	1,461	121	0.93	11.13
Students without Disabilities	145,456	121	0.93	12.03	5,885	121	0.93	12.15
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	138	121	0.94	11.11	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	1,272	121	0.84	11.36	-	-	-	-

Table 8.11 Summary of Test Reliability Estimates for Subgroups: Grade 11 ELA/L

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	129,937	121	0.92	10.89	6,045	121	0.91	12.12
Gender								
Male	67,136	121	0.92	10.72	3,226	121	0.91	12.01
Female	62,700	121	0.91	10.99	2,758	121	0.90	12.23
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	58,944	121	0.92	10.79	4,062	121	0.91	11.98
African American	24,584	121	0.91	11.08	755	121	0.89	12.47
Asian/Pacific Islander	6,833	121	0.92	11.18				
American Indian/Alaska Native	2,470	121	0.88	10.58	366	121	0.87	12.58
Hispanic	34,786	121	0.91	10.95	472	121	0.91	12.10
Multiple	1,990	121	0.92	10.81	213	121	0.89	12.34
Special Instructional Needs								
Economically Disadvantaged	54,356	121	0.91	10.93	2,036	121	0.89	12.30
Not-economically Disadvantaged	75,462	121	0.92	10.90	3,805	121	0.91	12.13
English Learner (EL)	4,730	121	0.88	10.82				
Non English Learner	124,753	121	0.92	10.89	5,748	121	0.91	12.09
Students with Disabilities (SWD)	19,930	121	0.90	10.72	952	121	0.91	11.38
Students without Disabilities	87,689	121	0.91	10.98	2,284	121	0.90	12.07
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	1,928	121	0.85	10.52	-	-	-	-

Table 8.12 Summary of Test Reliability Estimates for Subgroups: Grade 3 Mathematics

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	375,519	66	0.93	9.03	99,447	66	0.93	9.34
Gender								
Male	191,565	66	0.94	9.01	50,623	66	0.93	9.33
Female	183,954	66	0.93	9.04	48,824	66	0.93	9.34
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	174,999	66	0.93	8.94	36,988	66	0.92	9.21
African American	58,252	66	0.93	9.05	20,313	66	0.92	9.36
Asian/Pacific Islander	22,881	66	0.93	9.29	5,308	66	0.93	9.47
American Indian/Alaska Native	3,576	66	0.92	8.91	1,748	66	0.91	8.90
Hispanic	102,601	66	0.92	9.03	31,645	66	0.92	9.29
Multiple	12,556	66	0.94	8.99	3,188	66	0.93	9.33
Special Instructional Needs								
Economically Disadvantaged	174,078	66	0.92	9.04	57,373	66	0.92	9.32
Not-economically Disadvantaged	197,206	66	0.93	9.00	41,162	66	0.92	9.27
English Learner (EL)	51,285	66	0.92	9.13	21,657	66	0.92	9.33
Non English Learner	317,896	66	0.93	9.02	77,414	66	0.93	9.33
Students with Disabilities (SWD)	50,119	66	0.93	9.29	13,808	66	0.92	9.75
Students without Disabilities	209,299	66	0.93	8.98	38,060	66	0.93	9.25
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	93,286	66	0.94	9.36	-	-	-	-
Students Taking Translated Forms								
Spanish Language Form	4,462	66	0.90	9.54	1,067	66	0.90	9.95

Table 8.13 Summary of Test Reliability Estimates for Subgroups: Grade 4 Mathematics

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	378,225	66	0.93	8.34	84,410	66	0.93	8.88
Gender								
Male	192,826	66	0.94	8.28	42,975	66	0.93	8.80
Female	185,399	66	0.93	8.39	41,435	66	0.93	8.95
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	179,725	66	0.92	8.26	31,735	66	0.92	8.92
African American	57,364	66	0.92	8.53	16,958	66	0.91	8.80
Asian/Pacific Islander	23,504	66	0.93	8.39	4,405	66	0.93	9.09
American Indian/Alaska Native	3,668	66	0.91	8.38	1,569	66	0.90	8.81
Hispanic	101,307	66	0.91	8.41	27,076	66	0.91	8.77
Multiple	12,178	66	0.94	8.30	2,426	66	0.93	8.84
Special Instructional Needs								
Economically Disadvantaged	171,913	66	0.91	8.48	48,084	66	0.91	8.79
Not-economically Disadvantaged	201,945	66	0.93	8.28	35,454	66	0.92	8.97
English Learner (EL)	33,549	66	0.90	8.52	11,532	66	0.91	8.62
Non English Learner	338,473	66	0.93	8.33	72,502	66	0.93	8.91
Students with Disabilities (SWD)	55,871	66	0.92	8.54	13,900	66	0.91	8.69
Students without Disabilities	205,974	66	0.93	8.29	30,882	66	0.93	9.02
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	90,839	66	0.94	8.36	-	-	-	-
Students Taking Translated Forms								
Spanish Language Form	2,876	66	0.87	8.69	422	66	0.86	8.82

Table 8.14 Summary of Test Reliability Estimates for Subgroups: Grade 5 Mathematics

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	405,033	66	0.92	8.64	51,463	66	0.93	8.50
Gender								
Male	206,669	66	0.93	8.65	26,439	66	0.93	8.53
Female	198,364	66	0.92	8.60	25,024	66	0.92	8.45
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	186,430	66	0.92	8.46	25,566	66	0.92	8.42
African American	65,877	66	0.90	8.97	6,463	66	0.90	8.62
Asian/Pacific Islander	24,894	66	0.93	8.45	3,061	66	0.93	8.56
American Indian/Alaska Native	3,559	66	0.90	8.90	1,466	66	0.88	8.63
Hispanic	111,738	66	0.90	8.90	12,871	66	0.91	8.54
Multiple	12,047	66	0.93	8.55	1,828	66	0.94	8.51
Special Instructional Needs								
Economically Disadvantaged	188,907	66	0.90	8.95	22,709	66	0.91	8.58
Not-economically Disadvantaged	211,927	66	0.92	8.45	27,810	66	0.93	8.44
English Learner (EL)	28,266	66	0.88	9.34	5,840	66	0.89	8.64
Non English Learner	370,375	66	0.92	8.59	45,418	66	0.93	8.48
Students with Disabilities (SWD)	61,696	66	0.90	9.14	10,733	66	0.90	8.74
Students without Disabilities	200,834	66	0.92	8.50	29,684	66	0.93	8.41
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	110,215	66	0.93	8.76	-	-	-	-
Students Taking Translated Forms								
Spanish Language Form	2,537	66	0.84	10.16	-	-	-	-

Table 8.15 Summary of Test Reliability Estimates for Subgroups: Grade 6 Mathematics

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	404,238	66	0.93	8.15	51,856	66	0.93	8.16
Gender								
Male	206,623	66	0.93	8.12	26,624	66	0.93	8.15
Female	197,615	66	0.92	8.15	25,232	66	0.93	8.14
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	189,772	66	0.92	8.01	25,586	66	0.92	8.04
African American	65,552	66	0.90	8.43	6,840	66	0.90	8.44
Asian/Pacific Islander	24,575	66	0.93	8.19	3,326	66	0.94	8.19
American Indian/Alaska Native	3,605	66	0.90	8.42	1,337	66	0.90	8.32
Hispanic	109,052	66	0.91	8.31	12,725	66	0.90	8.26
Multiple	11,323	66	0.93	8.11	1,848	66	0.93	8.13
Special Instructional Needs								
Economically Disadvantaged	183,318	66	0.90	8.37	22,522	66	0.90	8.32
Not-economically Disadvantaged	216,433	66	0.92	8.04	28,556	66	0.93	8.09
English Learner (EL)	23,471	66	0.88	8.64	4,977	66	0.89	8.50
Non English Learner	374,708	66	0.93	8.10	46,655	66	0.93	8.12
Students with Disabilities (SWD)	62,599	66	0.91	8.42	10,652	66	0.89	8.57
Students without Disabilities	198,406	66	0.93	8.08	29,117	66	0.93	8.07
Students Taking Accommodated Forms								
A: ASL	102	66	0.85	8.92	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	94,673	66	0.94	8.16	-	-	-	-
Students Taking Translated Forms								
Spanish Language Form	2,163	66	0.83	9.70	-	-	-	-

Table 8.16 Summary of Test Reliability Estimates for Subgroups: Grade 7 Mathematics

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	382,190	66	0.92	8.19	52,101	66	0.92	8.03
Gender								
Male	195,431	66	0.92	8.20	26,584	66	0.92	8.01
Female	186,759	66	0.91	8.13	25,407	66	0.92	7.99
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	180,836	66	0.91	7.86	25,660	66	0.92	7.70
African American	62,201	66	0.88	9.02	6,849	66	0.88	8.83
Asian/Pacific Islander	20,885	66	0.93	7.67	3,221	66	0.94	7.61
American Indian/Alaska Native	3,614	66	0.88	8.61	1,050	66	0.88	8.42
Hispanic	104,244	66	0.88	8.81	13,293	66	0.89	8.49
Multiple	10,038	66	0.92	7.94	1,644	66	0.93	7.66
Special Instructional Needs								
Economically Disadvantaged	173,458	66	0.88	8.84	21,881	66	0.89	8.60
Not-economically Disadvantaged	204,153	66	0.92	7.83	29,314	66	0.93	7.71
English Learner (EL)	22,803	66	0.84	9.80	4,922	66	0.85	9.26
Non English Learner	353,929	66	0.92	8.07	46,935	66	0.92	7.90
Students with Disabilities (SWD)	60,884	66	0.88	9.35	10,561	66	0.89	8.74
Students without Disabilities	183,814	66	0.92	7.89	29,263	66	0.93	7.75
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	84,756	66	0.91	9.20	-	-	-	-
Students Taking Translated Forms								
Spanish Language Form	2,145	66	0.67	13.83	110	66	0.70	11.69

Table 8.17 Summary of Test Reliability Estimates for Subgroups: Grade 8 Mathematics

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	314,017	66	0.91	11.07	44,484	66	0.91	12.11
Gender								
Male	162,931	66	0.91	10.98	22,760	66	0.91	12.15
Female	151,086	66	0.90	11.10	21,585	66	0.91	11.98
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	142,124	66	0.91	10.87	22,123	66	0.91	11.82
African American	57,743	66	0.86	11.90	5,513	66	0.87	12.72
Asian/Pacific Islander	13,402	66	0.94	10.58	2,599	66	0.93	11.46
American Indian/Alaska Native	3,194	66	0.83	11.66	1,043	66	0.85	12.51
Hispanic	89,300	66	0.87	11.61	11,636	66	0.88	12.50
Multiple	7,876	66	0.92	10.79	1,254	66	0.93	12.07
Special Instructional Needs								
Economically Disadvantaged	153,286	66	0.87	11.66	18,929	66	0.88	12.52
Not-economically Disadvantaged	157,815	66	0.91	10.80	24,902	66	0.91	11.82
English Learner (EL)	21,725	66	0.82	12.04	4,641	66	0.84	12.88
Non English Learner	288,184	66	0.91	11.02	39,637	66	0.91	11.99
Students with Disabilities (SWD)	55,668	66	0.87	11.68	9,454	66	0.87	12.68
Students without Disabilities	129,959	66	0.90	11.06	26,404	66	0.92	11.89
Students Taking Accommodated Forms								
A: ASL	107	66	0.78	13.68	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	69,266	66	0.91	11.33	-	-	-	-
Students Taking Translated Forms								
Spanish Language Form	1,941	66	0.78	13.63	139	66	0.70	14.83

Table 8.18 Summary of Test Reliability Estimates for Subgroups: Algebra I

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	301,139	81	0.91	10.19	19,605	81	0.92	10.28
Gender								
Male	155,264	81	0.91	10.15	10,145	81	0.92	10.30
Female	145,875	81	0.90	10.19	9,460	81	0.91	10.22
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	127,928	81	0.91	9.82	10,639	81	0.92	10.11
African American	55,822	81	0.85	11.12	2,893	81	0.90	10.43
Asian/Pacific Islander	19,339	81	0.93	9.28	923	81	0.93	9.58
American Indian/Alaska Native	3,681	81	0.83	10.64	108	81	0.86	10.43
Hispanic	87,420	81	0.85	11.01	4,067	81	0.88	10.66
Multiple	6,556	81	0.92	9.91	648	81	0.92	10.17
Special Instructional Needs								
Economically Disadvantaged	132,298	81	0.85	11.04	6,734	81	0.89	10.66
Not-economically Disadvantaged	163,856	81	0.92	9.81	12,108	81	0.92	10.06
English Learner (EL)	22,732	81	0.80	11.22	1,111	81	0.81	11.30
Non English Learner	273,887	81	0.91	10.10	18,222	81	0.92	10.18
Students with Disabilities (SWD)	45,380	81	0.87	10.61	2,967	81	0.91	10.49
Students without Disabilities	182,451	81	0.91	9.87	9,263	81	0.92	10.16
Students Taking Accommodated Forms								
A: ASL	153	81	0.54	15.20	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	58,724	81	0.89	11.67	-	-	-	-
Students Taking Translated Forms								
Spanish Language Form	3,181	81	0.66	13.05	222	81	0.59	14.88

Table 8.19 Summary of Test Reliability Estimates for Subgroups: Geometry

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	138,781	81	0.93	7.12	5,156	81	0.93	7.35
Gender								
Male	70,817	81	0.93	7.08	2,658	81	0.94	7.34
Female	67,964	81	0.92	7.13	2,378	81	0.92	7.27
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	65,066	81	0.93	6.99	2,578	81	0.93	7.16
African American	19,318	81	0.88	7.87	660	81	0.89	8.61
Asian/Pacific Islander	10,547	81	0.95	6.59	115	81	0.93	6.50
American Indian/Alaska Native	2,596	81	0.83	8.03				
Hispanic	38,872	81	0.88	7.86	1,143	81	0.90	7.74
Multiple	2,123	81	0.94	7.11				
Special Instructional Needs								
Economically Disadvantaged	51,231	81	0.88	7.85	2,305	81	0.89	8.02
Not-economically Disadvantaged	82,787	81	0.93	6.94	2,121	81	0.93	7.06
English Learner (EL)	6,752	81	0.83	8.38	221	81	0.88	7.43
Non English Learner	130,857	81	0.93	7.05	4,777	81	0.93	7.26
Students with Disabilities (SWD)	20,227	81	0.89	7.70	771	81	0.91	7.62
Students without Disabilities	104,632	81	0.93	6.98	3,023	81	0.93	7.25
Students Taking Accommodated Forms								
A: ASL								
C: Closed-Caption								
R: Screen Reader								
T: Text-to-Speech	12,644	81	0.90	7.68				
Students Taking Translated Forms								
Spanish Language Form	1,734	81	0.70	10.33				

Table 8.20 Summary of Test Reliability Estimates for Subgroups: Algebra II

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	130,338	81	0.93	10.53	7,839	80	0.91	11.52
Gender								
Male	64,586	81	0.93	10.43	3,887	81	0.92	11.50
Female	65,752	81	0.92	10.61	3,952	80	0.91	11.53
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	61,040	81	0.92	10.59	4,659	81	0.91	11.26
African American	21,883	81	0.87	11.41	1,370	80	0.89	11.75
Asian/Pacific Islander	10,544	81	0.94	10.05	270	81	0.92	11.49
American Indian/Alaska Native	2,083	81	0.84	11.66				
Hispanic	32,573	81	0.88	11.24	1,049	81	0.87	12.00
Multiple	1,879	81	0.93	10.41	317	80	0.91	11.53
Special Instructional Needs								
Economically Disadvantaged	47,894	81	0.88	11.33	2,696	80	0.88	11.94
Not-economically Disadvantaged	82,352	81	0.93	10.45	5,107	80	0.91	11.35
English Learner (EL)	4,041	81	0.87	10.94				
Non English Learner	125,957	81	0.93	10.51	7,553	80	0.91	11.49
Students with Disabilities (SWD)	14,346	81	0.90	10.66	664	80	0.90	11.71
Students without Disabilities	97,168	81	0.93	10.42	2,957	80	0.90	11.07
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	12,156	81	0.91	11.34	-	-	-	-
Students Taking Translated Forms								
Spanish Language Form	879	81	0.79	12.42	-	-	-	-

Table 8.21 Summary of Test Reliability Estimates for Subgroups: Integrated Mathematics I

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	16,275	81	0.90	10.65	-	-	-	-
Gender								
Male	8,266	81	0.91	10.57	-	-	-	-
Female	8,009	81	0.89	10.66	-	-	-	-
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	6,865	81	0.91	10.28	-	-	-	-
African American	1,878	81	0.84	11.89	-	-	-	-
Asian/Pacific Islander	538	81	0.93	10.22	-	-	-	-
American Indian/Alaska Native	124	81	0.82	12.03	-	-	-	-
Hispanic	6,400	81	0.87	11.17	-	-	-	-
Multiple	391	81	0.92	10.32	-	-	-	-
Special Instructional Needs								
Economically Disadvantaged	8,066	81	0.86	11.31	-	-	-	-
Not-economically Disadvantaged	8,205	81	0.91	10.29	-	-	-	-
English Learner (EL)	1,997	81	0.74	12.86	-	-	-	-
Non English Learner	13,890	81	0.90	10.51	-	-	-	-
Students with Disabilities (SWD)	1,967	81	0.82	11.91	-	-	-	-
Students without Disabilities	4,031	81	0.88	11.00	-	-	-	-
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	1,480	81	0.84	12.43	-	-	-	-
Students Taking Translated Forms								
Spanish Language Form	-	-	-	-	-	-	-	-

Table 8.22 Summary of Test Reliability Estimates for Subgroups: Integrated Mathematics II

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	4,313	80	0.86	10.84	266	80	0.84	10.33
Gender								
Male	2,119	80	0.87	10.74	117	80	0.86	9.97
Female	2,194	80	0.85	10.92	149	80	0.82	10.61
Unknown/Missing	-	-	-	-	-	-	-	-
Ethnicity								
White	1,325	80	0.87	10.66				
African American	484	80	0.73	12.33				
Asian/Pacific Islander								
American Indian/Alaska Native	182	80	0.69	12.08	202	80	0.83	10.54
Hispanic	2,121	80	0.80	11.19				
Multiple								
Special Instructional Needs								
Economically Disadvantaged	2,786	80	0.79	11.41	215	80	0.84	10.48
Not-economically Disadvantaged	1,487	80	0.88	10.76				
English Learner (EL)	210	80	0.51	13.56				
Non English Learner	3,982	80	0.86	10.85	189	80	0.83	10.14
Students with Disabilities (SWD)	388	81	0.60	12.67				
Students without Disabilities	1,441	80	0.88	10.64	190	80	0.83	10.34
Students Taking Accommodated Forms								
A: ASL	-	-	-	-	-	-	-	-
C: Closed-Caption	-	-	-	-	-	-	-	-
R: Screen Reader	-	-	-	-	-	-	-	-
T: Text-to-Speech	499	81	0.65	13.53	-	-	-	-
Students Taking Translated Forms								
Spanish Language Form	-	-	-	-	-	-	-	-

Table 8.23 Summary of Test Reliability Estimates for Subgroups: Integrated Mathematics III

	CBT				PBT			
	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM	Total Sample Size	Average Maximum Possible Raw Scores	Average Reliability	Average Scale Score SEM
Total Group	2,142	81	0.92	11.36	114	80	0.75	13.32
Gender								
Male	987	81	0.92	11.23				
Female	1,155	81	0.91	11.45				
Unknown/Missing								
Ethnicity								
White	1,345	81	0.91	11.28				
African American								
Asian/Pacific Islander								
American Indian/Alaska Native	118	81	0.85	12.16				
Hispanic	529	81	0.93	11.28				
Multiple								
Special Instructional Needs								
Economically Disadvantaged	1,059	81	0.91	11.57				
Not-economically Disadvantaged	1,082	81	0.92	11.08				
English Learner (EL)								
Non English Learner	2,031	81	0.92	11.34	103	80	0.77	13.26
Students with Disabilities (SWD)	124	81	0.84	10.99				
Students without Disabilities	680	81	0.89	11.59				
Students Taking Accommodated Forms								
A: ASL								
C: Closed-Caption								
R: Screen Reader								
T: Text-to-Speech								
Students Taking Translated Forms								
Spanish Language Form	-	-	-	-				

8.5 Reliability Results for English Language Arts/Literacy Subscores

PARCC developed subclaims in addition to major claims based on the Common Core State Standards. English Language Arts/Literacy has two Major Claims relating to Reading Complex Text and Writing. The Major Claim for Reading Complex Text is that students read and comprehend a range of sufficiently complex texts independently. The Major Claim for Writing is that students write effectively when using and/or analyzing sources. Refer to Table 8.24 for a summary of the English language arts/literacy claims and subclaims.

Table 8.24 Descriptions of ELA/L Claims and Subclaims

English Language Arts/Literacy		
Major Claim	Subclaim	Description
Reading	Reading Literature	Students demonstrate comprehension and draw evidence from readings of grade-level, complex literary text
Reading	Reading Information	Students demonstrate comprehension and draw evidence from readings of grade-level, complex informational text
Reading	Reading Vocabulary	Students use context to determine the meaning of words and phrases
Writing	Writing Written Expression	Students produce clear and coherent writing in which the development, organization, and style are appropriate to the task, purpose, and audience
Writing	Writing Knowledge of Language and Conventions	Students demonstrate knowledge of conventions and other important elements of language

Reliability indices were calculated for each major claim and subclaim. Table 8.25 presents the average reliability estimates for all forms of the test at the specified grade and testing mode for the English Language Arts/Literacy tests. In order to assist in understanding the reliability estimates, the average maximum number of points for each major claim and subclaim are also provided.

The reliabilities for the Reading Complex Text claim for grades 3-11 ranges from .87 to .91 for CBT and from .86 to .92 for PBT. The reliability for grades 3-5 ranges from .86 to .90 and the average reliability for grades 6-11 ranges from .88 to .92.

The Writing claim reliabilities are lower than those for the Reading claim. The reliabilities for the Writing claim for grade 3 are based on 36 points and the average reliabilities for the grades 4-11 Writing claims are based on 45 points. The reliability for grades 3-5 ranges from .79 to .84 with a median of .82, and the average reliability for grades 6-11 ranges from .84 to .88, with a median of .86. Taking the number of points into consideration, the per-point information of the two claims are quite similar, as are the per-point information when comparing grades 3-5 with grades 6-11.

Reliability of the Reading Literature subclaim scores over testing mode and grade has a median average of .78, and the reliabilities vary from .71 to .83. For grades 3-5, the Reading Information subclaim reliabilities are based on 25 points and have a median of .73. For grades 6-11, the Reading Information subclaim is based on an average of 37 points, and the median subclaim reliability is .80. Once again, when taking the number of points into consideration, the per-point information of the claim is quite similar when comparing grades 3-5 with grades 6-11. The Reading Vocabulary subclaim is based on the fewest points, ranging from 12 to 22 points. The average subclaim reliability has a median of .67 and ranges from .58 to .74. The lower reliabilities for the Reading Vocabulary subclaim are reflected in the smaller range of raw score points.

The Writing: Written Expression subclaim is based on 27 points for grades 3-5 and 36 points for grades 6-11. The median average reliability for grades 3-5 is .73, and the median average reliability for grade 6-11 is .82. The Writing: Knowledge of Language and Conventions subclaims are all based on nine points. The average reliabilities are consistent, varying from .76 to .83, with a median of .80.

Table 8.25 Average ELA/L Reliability Estimates for Total Test and Subscores

Grade Level	Testing Mode	Reading: Total		Reading: Literature		Reading: Information		Reading: Vocabulary		Writing: Total		Writing: Written Expression		Writing: Knowledge Language and Conventions	
		Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability
3	CBT	57	0.90	26	0.83	18	0.67	13	0.72	36	0.80	27	0.71	9	0.83
	PBT	58	0.90	27	0.80	19	0.73	12	0.73	36	0.79	27	0.70	9	0.80
4	CBT	64	0.88	25	0.73	26	0.74	13	0.65	42	0.84	33	0.80	9	0.84
	PBT	64	0.87	26	0.74	25	0.72	13	0.59	42	0.81	33	0.75	9	0.79
5	CBT	64	0.87	25	0.71	24	0.77	15	0.57	42	0.84	33	0.79	9	0.85
	PBT	64	0.86	23	0.65	26	0.74	15	0.62	42	0.82	33	0.75	9	0.81
6	CBT	76	0.89	24	0.73	36	0.80	16	0.64	45	0.86	36	0.85	9	0.86
	PBT	76	0.91	25	0.76	33	0.81	18	0.71	45	0.85	36	0.83	9	0.83
7	CBT	76	0.90	25	0.76	33	0.81	17	0.64	45	0.87	36	0.87	9	0.88
	PBT	76	0.89	21	0.72	35	0.79	20	0.70	45	0.84	36	0.82	9	0.83
8	CBT	76	0.89	24	0.68	34	0.78	18	0.71	45	0.87	36	0.87	9	0.88
	PBT	76	0.90	22	0.75	35	0.78	19	0.72	45	0.85	36	0.83	9	0.83
9	CBT	76	0.90	23	0.76	37	0.83	17	0.64	45	0.88	36	0.87	9	0.88
	PBT	76	0.91	31	0.80	29	0.80	16	0.67	45	0.85	36	0.81	9	0.82
10	CBT	76	0.91	27	0.75	35	0.84	15	0.64	45	0.87	36	0.86	9	0.87
	PBT	76	0.92	23	0.77	35	0.86	18	0.72	45	0.86	36	0.85	9	0.86
11	CBT	76	0.88	28	0.74	32	0.75	16	0.60	45	0.87	36	0.85	9	0.86
	PBT	76	0.88	29	0.75	30	0.75	17	0.62	45	0.86	36	0.84	9	0.84

8.6 Reliability Results for Mathematics Subscores

For mathematics, there are four subclaims related to the major claim that students are on track or ready for college and careers:

- Subclaim A: Students solve problems involving the *major content* for their grade level with connections to the Standards for Mathematical Practice.
- Subclaim B: Students solve problems involving the *additional and supporting content* for their grade level with connections to the Standards for Mathematical Practice.
- Subclaim C: Students express grade/course-level appropriate *mathematical reasoning* by constructing viable mathematical arguments and critiquing the reasoning of others, and/or attending to precision when making mathematical statements
- Subclaim D: Students solve real-world problems with a degree of difficulty appropriate to the grade/course by applying knowledge and skills articulated in the standards and by engaging particularly in the *modeling practice*.

Reliability estimates were calculated for each subclaim for mathematics. Table 8.26 presents the average reliability estimates for mathematics subclaims by mode (CBT and PBT) and grade/subject. The sample size for Integrated Mathematics I PBT was not sufficient for reliability analyses.

Subclaims with greater numbers of points tend to have greater internal consistency reliability estimates. The Major Content subclaim has the largest number of points for each assessment and accordingly has higher average reliabilities than the other three subclaims. For grades 3 through 8, the average reliability for the subclaim is .86. The Major Content reliabilities were lower for the six EOC assessments than for grade level assessments. The average reliability for the Major Content subclaim for the traditional EOC tests is .81 and the average reliability for the integrated EOC tests is 0.66.

The average reliability for the Additional and Supporting Content subclaim range for grades 3 through 8 is .66 and the average reliability for the traditional EOC tests is .73. Due to the number of subclaim items being more similar across grades and courses, the subclaim reliabilities for Mathematics Reasoning are less variable than those for the Additional and Supporting Content subclaim. The Mathematics Reasoning subclaim reliability ranges from .43 for Integrated Mathematics II PBT test forms to .77 for the grade 7 PBT test forms.

For the Modeling Practice subclaim, the average reliability is .62 for grades 3 through 8 and .64 for the traditional and integrated mathematics EOC test forms.

Table 8.26 Average Mathematics Reliability Estimates for Total Test and Subscores

Grade Level	Testing Mode	Major Content		Additional & Supporting Content		Mathematics Reasoning		Modeling Practice	
		Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability	Average Max Possible Raw Score	Average Reliability
3	CBT	30	0.89	10	0.68	14	0.71	12	0.70
	PBT	30	0.88	10	0.70	14	0.69	12	0.66
4	CBT	31	0.87	9	0.63	14	0.76	12	0.61
	PBT	31	0.88	9	0.64	14	0.75	12	0.60
5	CBT	30	0.87	10	0.56	14	0.73	12	0.58
	PBT	30	0.86	10	0.64	14	0.72	12	0.61
6	CBT	26	0.86	14	0.71	14	0.70	12	0.63
	PBT	26	0.84	14	0.75	14	0.74	12	0.65
7	CBT	29	0.85	11	0.61	14	0.70	12	0.61
	PBT	29	0.83	11	0.68	14	0.77	12	0.55
8	CBT	27	0.80	13	0.67	14	0.66	12	0.59
	PBT	27	0.83	13	0.64	14	0.63	12	0.59
A1	CBT	27	0.79	22	0.72	14	0.65	18	0.64
	PBT	27	0.78	22	0.75	14	0.66	18	0.72
GO	CBT	30	0.85	19	0.71	14	0.69	18	0.69
	PBT	30	0.84	19	0.73	14	0.70	18	0.67
A2	CBT	28	0.80	21	0.75	14	0.72	18	0.67
	PBT	28	0.78	21	0.71	14	0.65	18	0.61
M1	CBT	31	0.77	18	0.65	14	0.66	18	0.69
	PBT								
M2	CBT	30	0.68	18	0.58	14	0.52	18	0.68
	PBT	29	0.58	19	0.48	14	0.43	18	0.62
M3	CBT	26	0.78	23	0.67	14	0.73	18	0.64
	PBT	26	0.48	22	0.27	14	0.51	18	0.41

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. Integrated Mathematics I PBT had insufficient sample sizes.

8.7 Reliability of Classification

The reliability of the classifications for the test takers was calculated using the computer program BB-CLASS (Brennan, 2004), which operationalizes a statistical method developed by Livingston and Lewis (1993, 1995). As Livingston and Lewis (1993, 1995) explain, this method uses information from the administration of one test form (i.e., distribution of scores, the minimum and maximum possible scores, the cut points used for classification, and the reliability coefficient) to estimate two kinds of statistics, “decision accuracy” and “decision consistency.” Decision accuracy refers to the extent to which the classifications of test takers based on their scores on the test form agree with the classifications made on the basis of the classifications that would be made if the test scores were perfectly reliable. Decision consistency refers to the agreement between these classifications based on two non-overlapping, equally difficult forms of the test.

Decision consistency values are always lower than the corresponding decision accuracy values, because in decision consistency, both of the classifications of the student are based on scores that depend on which form of the test the student took. In decision accuracy, only one of the classifications is based on a score that can vary in this way. It is not possible to know which students were accurately classified, but it is possible to estimate the proportion of the students who were accurately classified. Similarly, it is not possible to know which students would be consistently classified if they were retested with another form, but it is possible to estimate the proportion of the students who would be consistently classified.

English Language Arts/Literacy

Table 8.27 provides information about the accuracy and the consistency of two classifications made on the basis of the scores on the grades 3-11 English Language Arts/Literacy assessments. The columns labeled as “Exact level” provide the classification of the student into one of five achievement levels. The columns labeled as “Level 4 or higher vs. 3 or lower” provide the classification of the student as being either in one of the upper two levels (Levels 4 and 5) or in one of the lower three levels (Levels 1, 2, and 3).

The table shows that for classifying each student into one of the five achievement levels, the proportion accurately classified ranges from .71 to .77; the proportion who would be consistently classified on two different test forms ranges from .60 to .68. For classifying each student simply as being at Level 4 or higher vs. being at Level 3 or lower, the proportion accurately classified ranges from .90 to .93; the proportion who would be consistently classified on two different test forms ranges from .86 to .90.

Table 8.27 Reliability of Classification: Summary for ELA/L

Grade Level	Testing Mode	Decision Accuracy: Proportion Accurately Classified		Decision Consistency: Proportion Consistently Classified	
		Level 4 or higher vs. 3 or lower		Level 4 or higher vs. 3 or lower	
		Exact level		Exact level	
3	CBT	0.73	0.91	0.64	0.88
	PBT	0.74	0.91	0.65	0.87
4	CBT	0.73	0.91	0.63	0.87
	PBT	0.71	0.90	0.60	0.86
5	CBT	0.76	0.91	0.67	0.87
	PBT	0.75	0.90	0.66	0.86
6	CBT	0.76	0.91	0.67	0.88
	PBT	0.77	0.91	0.68	0.88
7	CBT	0.75	0.92	0.65	0.89
	PBT	0.73	0.91	0.63	0.88
8	CBT	0.76	0.92	0.67	0.89
	PBT	0.74	0.91	0.65	0.88
9	CBT	0.76	0.92	0.66	0.89
	PBT	0.74	0.91	0.64	0.88
10	CBT	0.73	0.92	0.64	0.89
	PBT	0.75	0.93	0.66	0.90
11	CBT	0.74	0.92	0.64	0.88
	PBT	0.71	0.91	0.61	0.87

Table 8.28 provides more detailed information about the accuracy and the consistency of the classification of students into proficiency levels for ELA/L grade 3. Each cell in the 5-by-5 table shows the estimated proportion of students who would be classified into a particular combination of proficiency levels. The sum of the five bold italicized values on the diagonal should equal the exact level of decision accuracy or consistency presented in Table 8.27. For “Level 4 and higher vs. 3 and lower” found in Table 8.27, the sum of the shaded values in Table 8.28 should equal the level of decision accuracy or consistency presented in Table 8.27. Note that the sums based on values in Table 8.28 may not match exactly to the values in Table 8.27 due to truncation and rounding.

Detailed information for all ELA/L Spring results are provided in Appendix 8 Tables A.8.1 to A.8.9. Fall block results for ELA/L grades 9-11 are provided in the addendum to Section 8. The structure of these tables is the same as that of Table 8.28 and the values in the tables should be interpreted in the same manner as Table 8.28.

Table 8.28 Reliability of Classification: Grade 3 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.15	0.03	0.00	0.00	0.00	0.18
		700 – 724	0.04	0.12	0.04	0.00	0.00	0.20
		725 – 749	0.00	0.04	0.15	0.04	0.00	0.23
		750 – 809	0.00	0.00	0.05	0.29	0.02	0.36
		810 – 850	0.00	0.00	0.00	0.01	0.02	0.03
	Decision Consistency	650 – 699	0.14	0.04	0.01	0.00	0.00	0.19
		700 – 724	0.04	0.09	0.06	0.01	0.00	0.20
		725 – 749	0.00	0.05	0.11	0.05	0.00	0.22
		750 – 809	0.00	0.01	0.06	0.27	0.02	0.35
		810 – 850	0.00	0.00	0.00	0.02	0.02	0.04
PBT	Decision Accuracy	650 – 699	0.14	0.02	0.00	0.00	0.00	0.16
		700 – 724	0.03	0.10	0.04	0.00	0.00	0.18
		725 – 749	0.00	0.04	0.14	0.05	0.00	0.22
		750 – 809	0.00	0.00	0.05	0.33	0.02	0.40
		810 – 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 – 699	0.13	0.03	0.01	0.00	0.00	0.17
		700 – 724	0.04	0.08	0.05	0.01	0.00	0.18
		725 – 749	0.00	0.04	0.11	0.06	0.00	0.21
		750 – 809	0.00	0.01	0.06	0.30	0.02	0.39
		810 – 850	0.00	0.00	0.00	0.02	0.03	0.05

Note: This table includes the same information as Table A.8.1. The sum of the five bold italicized values on the diagonal should equal the exact level of decision accuracy or consistency presented in Table 8.27. For “Level 4 and higher vs. 3 and lower” presented in Table 8.27, the sum of the shaded values in Table 8.28 should equal the level of decision accuracy or consistency presented in Table 8.27. Any differences between the sums based on values in Table 8.28 and the values in Table 8.27 are due to truncation and rounding.

Mathematics

Table 8.29 provides information about the accuracy and the consistency of two classifications made on the basis of the scores on the mathematics assessments. For the grades 3-8 mathematics tests, the table shows that for classifying each student into one of the five achievement levels, the proportion accurately classified ranges from .71 to .78; the proportion who would be consistently classified on two different test forms ranges from .62 to .69. For the six high school mathematics courses, the table shows that for classifying each student into one of the five achievement levels, the proportion accurately classified ranges from .68 to .79; the proportion who would be consistently classified on two different test forms ranges from .57 to .70.

For classifying each student simply as being at Level 4 or higher vs. being at Level 3 or lower, for the grades 3-8 mathematics tests, the proportion accurately classified ranges from .91 to .92; the proportion who would be consistently classified on two different test forms ranges from .87 to .89. For the six high school mathematics courses, the proportion accurately classified as being at Level 4 or higher vs. being at Level 3 or lower ranges from .90 to .97; the proportion who would be consistently classified on two different test forms ranges from .86 to .95.

Appendix 8 tables A.8.10 to A.8.21 provide more detailed information about the accuracy and the consistency of the classification of students into proficiency levels on the basis of the mathematics. Each cell in the 5-by-5 table shows the estimated proportion of students who would be classified into a particular combination of proficiency levels. Fall block results for Algebra I, Geometry, and Algebra II are provided in the addendum to Section 8.

Table 8.29 Reliability of Classification: Summary for Mathematics

Grade Level	Testing Mode	Decision Accuracy: Proportion Accurately Classified		Decision Consistency: Proportion Consistently Classified	
		Level 4 or higher vs. 3 or lower		Level 4 or higher vs. 3 or lower	
		Exact Level		Exact Level	
3	CBT	0.76	0.92	0.67	0.89
	PBT	0.76	0.92	0.66	0.89
4	CBT	0.78	0.92	0.69	0.89
	PBT	0.77	0.92	0.68	0.89
5	CBT	0.76	0.92	0.67	0.88
	PBT	0.77	0.92	0.68	0.89
6	CBT	0.78	0.92	0.69	0.89
	PBT	0.77	0.92	0.68	0.89
7	CBT	0.77	0.92	0.68	0.88
	PBT	0.76	0.92	0.67	0.88
8	CBT	0.73	0.92	0.64	0.89
	PBT	0.71	0.91	0.62	0.87
A1	CBT	0.75	0.92	0.65	0.88
	PBT	0.76	0.92	0.67	0.88
GO	CBT	0.79	0.93	0.70	0.90
	PBT	0.78	0.92	0.70	0.89
A2	CBT	0.77	0.94	0.69	0.91
	PBT	0.74	0.92	0.66	0.88
M1	CBT	0.73	0.91	0.63	0.88
	PBT	--	--	--	--
M2	CBT	0.68	0.92	0.57	0.89
	PBT	0.69	0.90	0.58	0.86
M3	CBT	0.76	0.93	0.67	0.90
	PBT	0.70	0.97	0.57	0.95

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. "--" means insufficient sample size (< 100 students).

8.8 Inter-rater Agreement

Inter-rater Agreement is the agreement between the first and second scores assigned to student responses. Inter-rater agreement measurements include exact, adjacent, and nonadjacent agreement. Pearson scoring staff used these statistics as one factor in determining the needs for continuing training and intervention on both individual and group levels. Table 8.30 displays both PARCC's expectations and the actual Spring 2015 agreement percentages for perfect agreement and perfect plus adjacent agreement.

Table 8.30 Inter-rater Agreement Expectations and Results

Subject	Score Point Range	Perfect Agreement Expectation	Perfect Agreement Result	Within One Point Expectation	Within One Point Result
Mathematics	0-1	90%	94%	96%*	100%
Mathematics	0-2	80%	92%	96%	99%
Mathematics	0-3	70%	89%	96%	99%
Mathematics	0-4	65%	86%	95%	98%
Mathematics	0-6	65%	89%	95%	97%
ELA/L	Multi-trait	65%	72%	96%	99%

Note: *A 0 or 1 score compared to a blank score will have a disagreement greater than 1 point.

Pearson's ePEN2 scoring system included comprehensive inter-rater agreement reports that allowed supervisory personnel to monitor both individual and group performance. Based on reviews of these reports, scoring experts targeted individuals for increased backreading and feedback and, if necessary, retraining. Table 8.30 shows that the actual percentages for both exact reader agreement and the percentages of agreement within one-point were higher than the inter-rater agreement expectations. Refer to Section 4 for more information on handscoring.

Section 9: Validity

9.1 Overview

The Standards for Educational and Psychological Testing, issued jointly by the American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME] (2014) reports:

Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests. The process of validation involves accumulating relevant evidence to provide a sound scientific basis for the proposed score interpretations (p. 11).

The purpose of test validation is not to validate the test itself but to validate interpretations of the test scores for particular uses. Test validation is not a quantifiable property but an ongoing process, beginning at initial conceptualization and continuing throughout the lifetime of an assessment. Every aspect of an assessment provides evidence in support of its validity (or evidence of lack of validity), including design, content specifications, item development, and psychometric characteristics. The 2016 operational assessment provided an opportunity to gather evidence of validity based on both test content and on the internal structure of the tests.

Pearson applies the principles of Universal Design, as articulated in materials developed by the National Center for Educational Outcomes (NCEO) at the University of Minnesota.

9.2 Evidence Based on Test Content

Evidence based on content of achievement tests 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 PARCC Assessments adhere to the principles of evidence-centered design, in which the standards to be measured (the Common Core State Standards) are identified, and the performance a student needs to achieve to meet those standards is delineated in the PARCC evidence statements. Test items were reviewed for adherence to universal design principles, which maximize the participation of the widest possible range of students.

Pearson and PARCC have built spreadsheets at the evidence statement level that incorporates the probability statements from the test blueprints and attrition rates at committee review and data review. The basis of our entire item development will be driven by the use of these item development target spreadsheets provided by PARCC. Before beginning item development, Pearson will use these target spreadsheets to develop an internal item development plan to correlate with the expectations of the test design. These will be reviewed and approved by PARCC as discussed in V.A.1.A. We acknowledge that each assessment has multiple parts and each part specifies the types of tasks and standards eligible for assessment.

In addition to the PARCC evidence statements, content is aligned through the articulation of performance in the performance level descriptors. At the policy level, the performance level descriptors 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 students performing at a particular achievement level are able to demonstrate. Those policy-level descriptors are the foundation for the subject- and grade-specific performance level descriptors which, along with the PARCC Evidence frameworks, guide the development of the items and tasks.

The PARCC College- and Career-Ready determinations (CCRD) in English Language Arts/literacy and mathematics describe the academic knowledge, skills and practices students must demonstrate to show readiness for success in entry-level, credit-bearing college courses and relevant technical courses. The PARCC states determined that this level means graduating from high school and having at least a 75% likelihood of earning a grade of “C” or better in credit-bearing courses without the need for remedial coursework. After reviewing the standards and assessment design, the PARCC Governing Board (made up of the K-12 education chiefs in PARCC states) in conjunction with the PARCC Advisory Committee on College Readiness (composed of higher education chiefs in the PARCC states), determined that students who achieve at levels 4 and 5 on the final PARCC high school assessments are likely to have acquired the skills and knowledge to meet the definition of college- and career-readiness. To validate the determinations, PARCC conducted a Postsecondary Educator Judgment Study and a Benchmark study of the SAT, ACT, National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), Programme of International Student Assessment (PISA), and Progress in International Reading Literacy Study (PIRLS) tests (McClarty, Korbin, Moyer, Griffin, Huth, Carey, and Medberry, 2015).

Gathering construct validity evidence for PARCC is embedded in the process by which the PARCC assessment content is developed and validated. At each step in the assessment development process, PARCC states involved hundreds of educators, assessment experts, and bias and sensitivity experts in review of text, items and tasks for accuracy, appropriateness, alignment to the instructional standards, and freedom from bias. See Chapter 2 for an overview of the content development process. In the early stages of development, Pearson conducted research studies to validate the PARCC item and task development approach. One such study was a Student Task Interaction Study designed to collect data on the student’s experience with the assessment tasks and technological functionalities, as well as the amount of time needed for answering each task. Pearson also conducted a Rubric Choice Study that compared the functioning of two rubrics developed to score the Prose Constructed Response (PCR) tasks in ELA. Quantitative and qualitative evidence was collected to support the use of a condensed or expanded trait scoring rubric in scoring student responses.

PARCC items and tasks were field tested prior to their use on an assessment. During the initial field test administration in 2014, PARCC states collected feedback from students, test administrators, test coordinators, and classroom teachers on their experience with the PARCC assessments, including the quality of test items and student experience. A summary of the feedback can be found at: <http://parcconline.org/files/79/College%20and%20Career%20Ready/91/PARCCCCRDandPLDPublicFeedbackSummaryReport-FINAL.pdf>. The feedback from that survey was used to inform test directions, test

timing, and the function of online task interactions. Performance data from the field test also informed the future development of additional items and tasks.

All item developers and item writers are provided an electronic version of PARCC Accessibility Guidelines and PARCC's Linguistic Complexity Rubric. Items and passages are reviewed internally by accessibility and fairness experts trained in the principles of Universal Design and who become well versed in PARCC's Accessibility Guidelines. Items received internal review for alignment to PARCC evidence tables, Task Generation Model, item selection guidelines, and accessibility and fairness reviews.

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 subgroups of test takers. PARCC convened bias and sensitivity committees to review all items. Additionally, content experts facilitated reviews of all items. All reviewers were trained using PARCC Bias and Sensitivity Guidelines, and the Guidelines were used to review items and ELA/L passages. Accommodations were made available based on individual need documented in the student's approved IEP, 504 Plan, or if required by the PARCC member state, an English learner (EL) Plan (refer to Section 3.4). An accessibility specialist worked in consultation with the PARCC accessibility specialist to review forms and determine which forms should be used for students with accommodations.

The ELA/L and mathematics operational test forms, as described in Section 2, were carefully constructed to align with the test blueprints and specifications that are based on the Common Core State Standards (CCSS). During the fall of 2014, content experts representing Parcc, Inc. and various PARCC states, along with content experts, held a series of meetings to review the operational forms for ELA/L and mathematics. These meetings provided opportunity to evaluate tests forms in their entirety and recommend changes. Requested item replacements were accommodated to the extent possible while striving to maintain the integrity of the various linking designs required for the operational test analyses. Psychometricians were available throughout this process to provide guidance with regard to implications of item replacements for the linking and statistical requirements.

Further information regarding the PARCC assessment college- and career-ready content standards, performance level descriptors, and accessibility features and accommodations is provided at the following URL: <http://www.parcconline.org/policies-and-guidance>.

9.3 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, APA, & NCME, 2014, p. 16). The term construct is used here to refer to the characteristics that a test is intended to measure; in the case of the PARCC operational tests, the characteristics of interest are the knowledge and skills defined by the test blueprint for ELA/L and for mathematics.

The PARCC assessments provide a full summative test score, Reading claim score, and Writing claim score as well as ELA/L subclaims and mathematics subclaim scores. 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. This information can then be used by teachers to plan for further instruction, to plan for curriculum development, and to report progress to parents. The results can also be used as one factor in making administrative decisions about program effectiveness, teacher effectiveness, class grouping, and needs assessment.

9.3.1 Intercorrelations

The ELA/L full summative tests comprise two claim scores: Reading (RD) and Writing (WR) and five subclaim scores: Reading Literature (RL), Reading Information (RI), Reading Vocabulary (RV), Writing Written Expression (WE), and Writing Knowledge Language and Conventions (WKL). The RD claim score is a composite of RL, RI, and RV. The writing claim score, a composite of WE and WKL, comprises only PCR items and the same PCR items are in each subclaim. The ELA/L operational test analyses were performed by evaluating the separate trait scores of WE and WKL, and for some PCR items also RL or RI, therefore the trait scores were used for the intercorrelations.

The mathematics full summative tests have four subclaim scores: Major Content (MC), Mathematical Reasoning (MR), Modeling Practice (MP), and Additional and Supporting Content (ASC).

High total group internal consistencies as well as similar reliabilities across subgroups provide additional evidence of validity. High reliability of test scores implies that the test items within a domain are measuring a single construct, which is a necessary condition for validity when the intention is to measure a single construct. Refer to Section 8 for reliability estimates for the overall population, subgroups of interest, as well as for subscores for ELA/L and subclaims for mathematics.

Another way to assess the internal structure of a test is through the evaluation of correlations among subscores. These analyses were conducted between the ELA/L reading and writing claim scores and the ELA/L subclaims (RL, RI, RV, WE, and WKL) and between the mathematics subclaims. If these components within a content area are strongly related to each other, this is evidence of unidimensionality.

A series of tables are provided to summarize the results for the Spring 2016 administration.⁷ Tables 9.1 through 9.9 present the Pearson correlations observed between the ELA/L reading and writing claim scores and subclaim scores for each grade; correlations are reported separately for online (CBT) and paper (PBT) versions of the tests. The tables provide the weighted average intercorrelations by averaging the intercorrelations computed for all the core operational forms of the test, separately for the CBT and PBT tests within each grade level. The total sample size across all forms is provided in the upper triangle portion of the tables. The subclaim reliabilities (from Section 8) are reported along the diagonal. The WR, WE, and WKL scores tended to be highly correlated; this is expected given that these three intercorrelations are based on the same Writing items. RL, RI, and RV, all subclaims of Reading, are moderately to highly correlated. Additionally, the WR claim and the WE and WKL subclaims, are

⁷ Addendum 9 provides a summary of results for the Fall 2015 administration.

moderately correlated with RD subclaims (of RL, RI, and RV). These moderate to high ELA/L intercorrelations amongst the subclaims are sufficiently high enough to provide evidence that the ELA/L tests are unidimensional. The moderate intercorrelations among the subclaims and claims suggest the claims may be sufficient for individual student reporting.

The intercorrelations and reliability estimates for mathematics are provided in Tables 9.10 to 9.21. The mathematics intercorrelations are moderate. The only observable pattern in the mathematics intercorrelations is that the MC subclaim has consistently slightly higher correlations with the ASC, MR, and MP subclaims; the intercorrelations amongst the ASC, MR, and MP subclaims are slightly lower. The mathematics intercorrelations are sufficiently high enough to suggest that the mathematics tests are likely to be unidimensional with some minor secondary dimensions.

Additionally, the ELA/L and mathematics correlations for the two modes, PBT and CBT, displayed similar patterns of intercorrelations suggesting that the structure of the PBT assessments and CBT assessments are similar. The one exception is the Integrated Mathematics II and III test for which the intercorrelations are noticeably lower for the PBT assessments. This is due to the substantially smaller number of students that took the PBT version of these tests.

Table 9.1 Average Intercorrelations and Reliability between Grade 3 ELA/L Subclaims

CBT								PBT							
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.90	371,885	371,885	371,885	371,885	371,885	371,885	RD	0.90	98,738	98,738	98,738	98,738	98,738	98,738
RL	0.94	0.83	371,885	371,885	371,885	371,885	371,885	RL	0.93	0.80	98,738	98,738	98,738	98,738	98,738
RI	0.87	0.71	0.67	371,885	371,885	371,885	371,885	RI	0.90	0.74	0.73	98,738	98,738	98,738	98,738
RV	0.89	0.76	0.67	0.72	371,885	371,885	371,885	RV	0.87	0.73	0.70	0.73	98,738	98,738	98,738
WR	0.69	0.64	0.65	0.57	0.80	371,885	371,885	WR	0.75	0.70	0.70	0.61	0.79	98,738	98,738
WE	0.66	0.62	0.63	0.54	0.99	0.71	371,885	WE	0.73	0.69	0.68	0.59	0.99	0.70	98,738
WKL	0.66	0.62	0.61	0.56	0.90	0.81	0.83	WKL	0.70	0.65	0.65	0.59	0.90	0.82	0.80

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.2 Average Intercorrelations and Reliability between Grade 4 ELA/L Subclaims

CBT								PBT							
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.88	377,002	377,002	377,002	377,002	377,002	377,002	RD	0.87	82,792	82,792	82,792	82,792	82,792	82,792
RL	0.91	0.73	377,002	377,002	377,002	377,002	377,002	RL	0.91	0.74	82,792	82,792	82,792	82,792	82,792
RI	0.91	0.72	0.74	377,002	377,002	377,002	377,002	RI	0.90	0.70	0.72	82,792	82,792	82,792	82,792
RV	0.84	0.68	0.67	0.65	377,002	377,002	377,002	RV	0.82	0.64	0.64	0.59	82,792	82,792	82,792
WR	0.75	0.70	0.70	0.57	0.84	377,002	377,002	WR	0.74	0.69	0.68	0.54	0.81	82,792	82,792
WE	0.74	0.70	0.69	0.57	0.99	0.80	377,002	WE	0.73	0.68	0.68	0.53	0.99	0.75	82,792
WKL	0.71	0.67	0.66	0.55	0.94	0.90	0.84	WKL	0.70	0.65	0.65	0.52	0.94	0.90	0.79

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.3 Average Intercorrelations and Reliability between Grade 5 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.87	404,383	404,383	404,383	404,383	404,383	404,383	RD	0.86	50,081	50,081	50,081	50,081	50,081	50,081
RL	0.90	0.71	404,383	404,383	404,383	404,383	404,383	RL	0.87	0.65	50,081	50,081	50,081	50,081	50,081
RI	0.91	0.72	0.77	404,383	404,383	404,383	404,383	RI	0.91	0.66	0.74	50,081	50,081	50,081	50,081
RV	0.82	0.61	0.63	0.57	404,383	404,383	404,383	RV	0.84	0.62	0.66	0.62	50,081	50,081	50,081
WR	0.75	0.69	0.71	0.54	0.84	404,383	404,383	WR	0.74	0.67	0.68	0.57	0.82	50,081	50,081
WE	0.74	0.69	0.71	0.54	0.99	0.79	404,383	WE	0.73	0.67	0.68	0.56	0.99	0.75	50,081
WKL	0.72	0.66	0.68	0.53	0.94	0.90	0.85	WKL	0.70	0.63	0.65	0.54	0.94	0.89	0.81

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.4 Average Intercorrelations and Reliability between Grade 6 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.89	402,155	402,155	402,155	402,155	402,155	402,155	RD	0.91	52,096	52,096	52,096	52,096	52,096	52,096
RL	0.90	0.73	402,155	402,155	402,155	402,155	402,155	RL	0.90	0.76	52,096	52,096	52,096	52,096	52,096
RI	0.94	0.76	0.80	402,155	402,155	402,155	402,155	RI	0.94	0.76	0.81	52,096	52,096	52,096	52,096
RV	0.83	0.66	0.69	0.64	402,155	402,155	402,155	RV	0.87	0.70	0.74	0.71	52,096	52,096	52,096
WR	0.74	0.72	0.70	0.53	0.86	402,155	402,155	WR	0.75	0.73	0.71	0.57	0.85	52,096	52,096
WE	0.73	0.71	0.69	0.52	1.00	0.85	402,155	WE	0.74	0.73	0.70	0.57	1.00	0.83	52,096
WKL	0.73	0.71	0.70	0.53	0.97	0.95	0.86	WKL	0.73	0.71	0.70	0.56	0.97	0.95	0.83

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.5 Average Intercorrelations and Reliability between Grade 7 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.90	395,258	395,258	395,258	395,258	395,258	395,258	RD	0.89	53,335	53,335	53,335	53,335	53,335	53,335
RL	0.91	0.76	395,258	395,258	395,258	395,258	395,258	RL	0.87	0.72	53,335	53,335	53,335	53,335	53,335
RI	0.94	0.77	0.81	395,258	395,258	395,258	395,258	RI	0.94	0.72	0.79	53,335	53,335	53,335	53,335
RV	0.85	0.67	0.71	0.64	395,258	395,258	395,258	RV	0.87	0.68	0.72	0.70	53,335	53,335	53,335
WR	0.75	0.72	0.72	0.57	0.87	395,258	395,258	WR	0.74	0.71	0.70	0.58	0.84	53,335	53,335
WE	0.75	0.72	0.71	0.56	1.00	0.87	395,258	WE	0.73	0.70	0.69	0.57	1.00	0.82	53,335
WKL	0.75	0.72	0.72	0.57	0.98	0.96	0.88	WKL	0.73	0.70	0.69	0.57	0.98	0.96	0.83

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.6 Average Intercorrelations and Reliability between Grade 8 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.89	388,964	388,964	388,964	388,964	388,964	388,964	RD	0.90	50,121	50,121	50,121	50,121	50,121	50,121
RL	0.88	0.68	388,964	388,964	388,964	388,964	388,964	RL	0.89	0.75	50,121	50,121	50,121	50,121	50,121
RI	0.93	0.72	0.78	388,964	388,964	388,964	388,964	RI	0.93	0.74	0.78	50,121	50,121	50,121	50,121
RV	0.87	0.69	0.72	0.71	388,964	388,964	388,964	RV	0.88	0.73	0.73	0.72	50,121	50,121	50,121
WR	0.76	0.70	0.73	0.61	0.87	388,964	388,964	WR	0.76	0.72	0.71	0.61	0.85	50,121	50,121
WE	0.76	0.70	0.72	0.60	1.00	0.87	388,964	WE	0.75	0.72	0.71	0.61	1.00	0.83	50,121
WKL	0.76	0.69	0.73	0.61	0.97	0.95	0.88	WKL	0.75	0.71	0.70	0.61	0.97	0.95	0.83

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.7 Average Intercorrelations and Reliability between Grade 9 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.90	259,459	259,459	259,459	259,459	259,459	259,459	RD	0.91	14,606	14,606	14,606	14,606	14,606	14,606
RL	0.90	0.76	259,459	259,459	259,459	259,459	259,459	RL	0.93	0.80	14,606	14,606	14,606	14,606	14,606
RI	0.95	0.78	0.83	259,459	259,459	259,459	259,459	RI	0.93	0.78	0.80	14,606	14,606	14,606	14,606
RV	0.85	0.68	0.71	0.64	259,459	259,459	259,459	RV	0.86	0.73	0.72	0.67	14,606	14,606	14,606
WR	0.76	0.74	0.73	0.57	0.88	259,459	259,459	WR	0.74	0.71	0.70	0.57	0.85	14,606	14,606
WE	0.76	0.74	0.73	0.57	1.00	0.87	259,459	WE	0.73	0.71	0.70	0.56	1.00	0.81	14,606
WKL	0.76	0.73	0.73	0.58	0.98	0.97	0.88	WKL	0.74	0.71	0.70	0.57	0.99	0.98	0.82

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.8 Average Intercorrelations and Reliability between Grade 10 ELA/L Subclaims

	CBT								PBT						
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.91	183,504	183,504	183,504	183,504	183,504	183,504	RD	0.92	8,407	8,407	8,407	8,407	8,407	8,407
RL	0.91	0.75	183,504	183,504	183,504	183,504	183,504	RL	0.92	0.77	8,407	8,407	8,407	8,407	8,407
RI	0.95	0.78	0.84	183,504	183,504	183,504	183,504	RI	0.96	0.82	0.86	8,407	8,407	8,407	8,407
RV	0.83	0.67	0.71	0.64	183,504	183,504	183,504	RV	0.89	0.75	0.79	0.72	8,407	8,407	8,407
WR	0.77	0.74	0.74	0.58	0.87	183,504	183,504	WR	0.79	0.77	0.76	0.64	0.86	8,407	8,407
WE	0.76	0.74	0.73	0.57	1.00	0.86	183,504	WE	0.78	0.76	0.76	0.64	1.00	0.85	8,407
WKL	0.77	0.74	0.74	0.59	0.98	0.96	0.87	WKL	0.78	0.76	0.76	0.64	0.98	0.96	0.86

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.9 Average Intercorrelations and Reliability between Grade 11 ELA/L Subclaims

CBT								PBT							
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.88	129,937	129,937	129,937	129,937	129,937	129,937	RD	0.88	6,045	6,045	6,045	6,045	6,045	6,045
RL	0.91	0.74	129,937	129,937	129,937	129,937	129,937	RL	0.91	0.75	6,045	6,045	6,045	6,045	6,045
RI	0.92	0.74	0.75	129,937	129,937	129,937	129,937	RI	0.91	0.73	0.75	6,045	6,045	6,045	6,045
RV	0.81	0.64	0.65	0.60	129,937	129,937	129,937	RV	0.84	0.67	0.67	0.62	6,045	6,045	6,045
WR	0.74	0.69	0.72	0.50	0.87	129,937	129,937	WR	0.75	0.72	0.70	0.55	0.86	6,045	6,045
WE	0.73	0.69	0.72	0.50	1.00	0.85	129,937	WE	0.75	0.72	0.70	0.55	1.00	0.84	6,045
WKL	0.73	0.69	0.72	0.50	0.98	0.96	0.86	WKL	0.73	0.70	0.68	0.54	0.98	0.96	0.84

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary, WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.10 Average Intercorrelations and Reliability between Grade 3 Mathematics Subclaims

CBT					PBT				
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.89	375,552	375,552	375,552	MC	0.88	99,447	99,447	99,447
ASC	0.78	0.68	375,552	375,552	ASC	0.80	0.70	99,447	99,447
MR	0.74	0.64	0.71	375,552	MR	0.75	0.67	0.69	99,447
MP	0.78	0.66	0.71	0.70	MP	0.77	0.67	0.71	0.66

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.11 Average Intercorrelations and Reliability between Grade 4 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.87	378,251	378,251	378,251	MC	0.88	84,410	84,410	84,410
ASC	0.70	0.63	378,251	378,251	ASC	0.72	0.64	84,410	84,410
MR	0.79	0.65	0.76	378,251	MR	0.77	0.67	0.75	84,410
MP	0.73	0.61	0.72	0.61	MP	0.71	0.61	0.68	0.60

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.12 Average Intercorrelations and Reliability between Grade 5 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.87	405,077	405,077	405,077	MC	0.86	51,531	51,531	51,531
ASC	0.68	0.56	405,077	405,077	ASC	0.72	0.64	51,531	51,531
MR	0.79	0.62	0.73	405,077	MR	0.79	0.66	0.72	51,531
MP	0.75	0.60	0.70	0.58	MP	0.72	0.63	0.70	0.61

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.13 Average Intercorrelations and Reliability between Grade 6 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.86	404,256	404,256	404,256	MC	0.84	51,941	51,941	51,941
ASC	0.78	0.71	404,256	404,256	ASC	0.79	0.75	51,941	51,941
MR	0.77	0.69	0.70	404,256	MR	0.79	0.72	0.74	51,941
MP	0.75	0.68	0.73	0.63	MP	0.72	0.68	0.70	0.65

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.14 Average Intercorrelations and Reliability between Grade 7 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.85	382,198	382,198	382,198	MC	0.83	52,101	52,101	52,101
ASC	0.71	0.61	382,198	382,198	ASC	0.76	0.68	52,101	52,101
MR	0.76	0.64	0.70	382,198	MR	0.79	0.72	0.77	52,101
MP	0.75	0.64	0.68	0.61	MP	0.77	0.69	0.72	0.55

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.15 Average Intercorrelations and Reliability between Grade 8 Mathematics Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.80	314,025	314,025	314,025	MC	0.83	44,484	44,484	44,484
ASC	0.75	0.67	314,025	314,025	ASC	0.77	0.64	44,484	44,484
MR	0.72	0.68	0.66	314,025	MR	0.75	0.69	0.63	44,484
MP	0.67	0.64	0.67	0.59	MP	0.71	0.66	0.66	0.59

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.16 Average Intercorrelations and Reliability between Algebra I Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.79	301,139	301,139	301,139	MC	0.78	19,605	19,605	19,605
ASC	0.78	0.72	301,139	301,139	ASC	0.79	0.75	19,605	19,605
MR	0.72	0.69	0.65	301,139	MR	0.67	0.66	0.66	19,605
MP	0.77	0.72	0.67	0.64	MP	0.72	0.70	0.66	0.72

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.17 Average Intercorrelations and Reliability between Geometry Subclaims

CBT					PBT				
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.85	138,856	138,856	138,856	MC	0.84	5,208	5,208	5,208
ASC	0.77	0.71	138,856	138,856	ASC	0.77	0.73	5,208	5,208
MR	0.74	0.64	0.69	138,856	MR	0.66	0.62	0.70	5,208
MP	0.79	0.70	0.75	0.69	MP	0.80	0.71	0.66	0.67

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.18 Average Intercorrelations and Reliability between Algebra II Subclaims

CBT					PBT				
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.80	130,408	130,408	130,408	MC	0.78	7,929	7,929	7,929
ASC	0.79	0.75	130,408	130,408	ASC	0.77	0.71	7,929	7,929
MR	0.76	0.75	0.72	130,408	MR	0.71	0.70	0.65	7,929
MP	0.75	0.73	0.72	0.67	MP	0.71	0.70	0.69	0.61

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.19 Average Intercorrelations and Reliability between Integrated Mathematics I Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.77	16,364	16,364	16,364	MC				
ASC	0.72	0.65	16,364	16,364	ASC				
MR	0.63	0.57	0.66	16,364	MR				
MP	0.73	0.65	0.65	0.69	MP				

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Integrated Mathematics I PBT had insufficient sample sizes. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.20 Average Intercorrelations and Reliability between Integrated Mathematics II Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.68	4,381	4,381	4,381	MC	0.58	268	268	268
ASC	0.61	0.58	4,381	4,381	ASC	0.58	0.48	268	268
MR	0.63	0.55	0.52	4,381	MR	0.64	0.58	0.43	268
MP	0.67	0.59	0.60	0.68	MP	0.62	0.49	0.69	0.62

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

Table 9.21 Average Intercorrelations and Reliability between Integrated Mathematics III Subclaims

	CBT					PBT			
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.78	2,226	2,226	2,226	MC	0.48	142	142	142
ASC	0.74	0.67	2,226	2,226	ASC	0.38	0.27	142	142
MR	0.75	0.69	0.73	2,226	MR	0.46	0.39	0.51	142
MP	0.70	0.67	0.69	0.64	MP	0.56	0.44	0.53	0.41

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table. Please refer to Appendix A.12.1 (Form Composition) for information about the number of items and number of score points in each claim and subclaim.

9.3.2 Reliability

Additionally, the reliability analyses presented in Section 8 of this technical report provide information about the internal consistency of the PARCC assessments. Internal consistency is typically measured via correlations amongst the items on an assessment and provides an indication of how much the items measure the same general construct. The reliability estimates, computed using coefficient alpha (Cronbach, 1951) are presented in Tables 8.1 and 8.2 and are along the diagonals of Tables 9.1 through 9.21.⁸ The average reliabilities for ELA/L PARCC assessments range from .89 to .94 and for the mathematics assessments range from .75 up to .93. Tables 8.3 through 8.11 summarize test reliability for groups of interest for English Language Arts/Literacy grades 3-11, and Tables 8.12 through 8.23 summarize test reliability for groups of interest for mathematics grades/subjects. Along with the subclaim intercorrelations, the reliability estimates indicate that the items within each PARCC assessment are measuring the same construct and provides further evidence of unidimensionality.

9.3.3 Local Item Dependence

In addition to the intercorrelations for ELA/L and mathematics, local item independence was evaluated. Local independence is one of the primary assumptions of IRT that states the probability of success on one item is not influenced by performance on other items, when controlling for ability level. This implies that ability or theta accounts for the associations among the observed items. *Local item dependence* (LID) when present essentially overstates the amount of information predicted by the IRT model. It can exert other undesirable psychometric effects and represents a threat to validity since other factors besides the construct of interest are present. Classical statistics are also affected when LID is present since estimates of test reliability like IRT information can be inflated (Zenisky, Hambleton, & Sireci, 2003).

The LID issue affects the choice of item scoring in IRT calibrations. Specifically, if evidence suggests these items indeed have local dependence, then it might be preferable to sum the item scores into clusters or testlets as a method of minimizing LID. However, if these items do not appear to have strong local item dependence, then retaining the scores as individual item scores in an IRT calibration is preferred since more information concerning item properties is retained. During the initial operational administration of the PARCC assessments in spring 2015, a study that included two methods of investigating the presence of LID was conducted. A description of the methods along with study findings are summarized below.

First, analyses of the internal consistency in items and testlets were conducted under classical test theory (Wainer & Thissen, 2001) as a way to evaluate the degree of LID. Two estimates of Cronbach's alpha (Cronbach, 1951) were compared based on individual items in a test and those clustered into testlets. Cronbach's alpha is formulated as:

⁸ Section 8 provides information on the computations of the reliability estimates.

$$\alpha = \frac{k}{k-1} \frac{\sum_{i \neq i'} \sigma_{ii'}}{\sigma_T^2}, \quad (9-1)$$

where k is the total number of items, $\sigma_{ii'}$ is the covariance of items i and i' ($i \neq i'$), and σ_T^2 is the variance of total scores. To compute an alpha coefficient, sample standard deviations and variances are substituted for the $\sigma_{ii'}$ and σ_T^2 . The alpha for the total test based on individual items is compared with those that form testlets based on larger subparts. If the item-level configuration has appreciably higher levels of internal consistency compared with the testlets, LID may be present.

For IRT based methods, local dependence can be evaluated using statistics such as Q_3 (Yen, 1984). The item residual is the difference between observed and expected performance. The Q_3 index is the correlation between residuals of each item pair defined as

$$\begin{aligned} d_i &= (O - \hat{E}), \\ Q_3 &= r(d_i, d_{i'}) \end{aligned} \quad (9-2)$$

where O is the observed score and \hat{E} is the expected value of O under a proposed IRT model and the index is defined as the correlation between the two item residuals.

LID manifests itself as a residual correlation that is nonzero and large. For Q_3 , LID can be either positive or negative. Positive (negative) LID indicates that performance is higher (lower) than expectation. The residual Q_3 correlation matrix can be inspected to determine if there are any blocks of locally dependent items (e.g., perhaps blocks of items belonging to the same reading passage). For Q_3 , the null hypothesis is that local independence holds. The expected value of Q_3 is $-1/n-1$ where n is the number of items such that the statistic shows a small negative bias. As a rule-of-thumb, item pairs with moderate levels of LID for Q_3 are $|0.20|$ or greater. Significant levels of LID are present when the statistic is greater than $|0.40|$. An alternative is to use the Fisher r to z transformation and evaluate the resulting p values.

For the LID comparisons, the follows eight test levels administered in spring 2015 were selected:

1. Grade 4 for span 3-5 in ELA/L,
2. Grade 4 for span 3-5 in mathematics,
3. Grade 7 for span 6-8 in ELA/L,
4. Grade 7 for span 6-8 in mathematics,
5. Grade 10 for span 9-11 in ELA/I,
6. Integrated Mathematics II for Integrated Mathematics I-III,
7. Algebra I, and
8. Algebra II.

One Spring 2015 CBT form for each of the eight tests was selected that was roughly at the median in terms of test difficulty. For ELA/L, reading items were summed according to passage assignment. For mathematics, items were summed according to subclaims. Cronbach's alpha was computed for the entire forms using the two different approaches as described above, one involving calculations at the

item level and the second utilizing scores on summed items (i.e., testlets). Further description of the data is given in Table 9.22.

To cross-validate the internal consistency analysis, the Q_3 statistic was computed from spring CBT data based on Grade 4 ELA/L and Integrated Mathematics II items. All items in the pool at that test level were included. The CBT item pool for grade 4 ELA/L contained 125 items while Integrated Mathematics Two had 77 items.

The results for the internal consistency analysis are shown in Figure 9.1. In every instance, the item-level Cronbach's alpha is higher than in the testlet configuration. The greatest difference was for Algebra II which showed a difference of 0.07. Although this was not unexpected, the magnitude of the differences in the respective alpha coefficients in general do not suggest a concerning level of LID. Table 9.23 shows the summary for the Q_3 values. Figures 9.2 and 9.3 show graphs of the distribution of Q_3 values. Most of the Q_3 values were small and negative, again suggesting that LID is not at a level of concern.

For these two test levels, the difference in the alpha coefficients was 0.03 and was consistent with the low values of Q_3 .

In summary, this investigation did not find evidence for the existence of pervasive LID. The results of both the internal consistency analyses and Q_3 methods support a claim of minimal LID. For a multiple-choice only test containing four reading passages with 5 to 12 items associated with a reading passage, Sireci, Thissen, and Wainer (1991) reported that testlet alpha was approximately 10 percent lower than the item-level coefficient. In comparison, PARCC tests have complex test structures and exhibited smaller differences in alpha coefficients. In addition, the median Q_3 values presented in Table 9.23 centered around the expectation of $-1/n-1$.

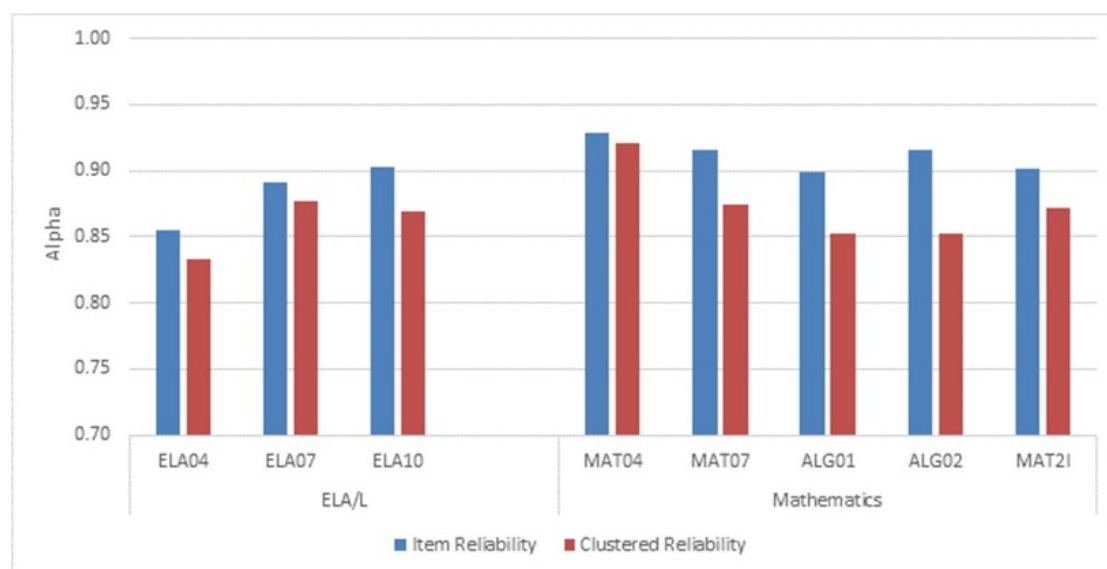


Figure 9.1 Comparison of Internal Consistency by Item and Cluster (Testlet)

Table 9.22 Conditions used in LID Investigation and Results

Content	Grade	<i>N</i> Valid	<i>N</i> Complete	Percent Incomplete	No. Items	No. Tasks	Item Rel.	Task Rel.
ELA/L								
ELA/L	4	13,660	13,518	1.04	31	5	0.86	0.83
ELA/L	7	12,757	12,685	0.56	41	7	0.89	0.88
ELA/L	10	3,097	3,033	2.07	41	7	0.90	0.87
Mathematics								
Math	4	10,332	10,255	0.75	53	4	0.93	0.92
Math	7	10,295	10,188	1.04	50	6	0.92	0.87
Math	A1	5,072	4,885	3.69	52	6	0.90	0.85
Math	A2	4,982	4,769	4.28	54	6	0.92	0.85
Math	M2	2,708	2,645	2.33	51	6	0.90	0.87

Note: A1 = Algebra I, A2 = Algebra II, M2 = Integrated Mathematics II.

Table 9.23 Summary of Q_3 Values for ELA/L Grade 4 and Integrated Mathematics II (Spring 2015)

Min.	Q_1	Median	Mean	Q_3	Max.	<i>SD</i>
ELA/L, Grade 4						
-0.138	-0.047	-0.031	-0.031	-0.017	0.279	0.030
Integrated Mathematics II						
-0.160	-0.038	-0.017	-0.019	0.001	0.280	0.032

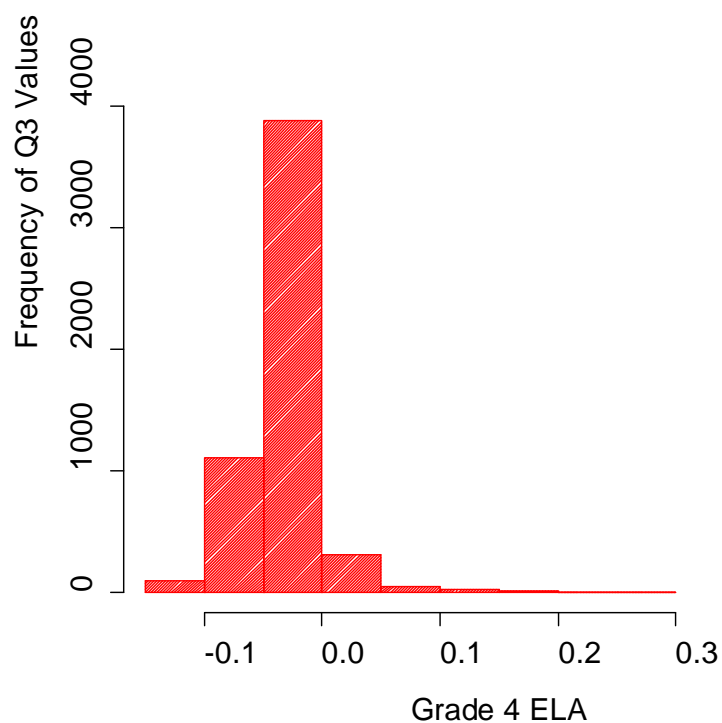


Figure 9.2 Distribution of Q_3 Values for Grade 4 ELA/L (Spring 2015)

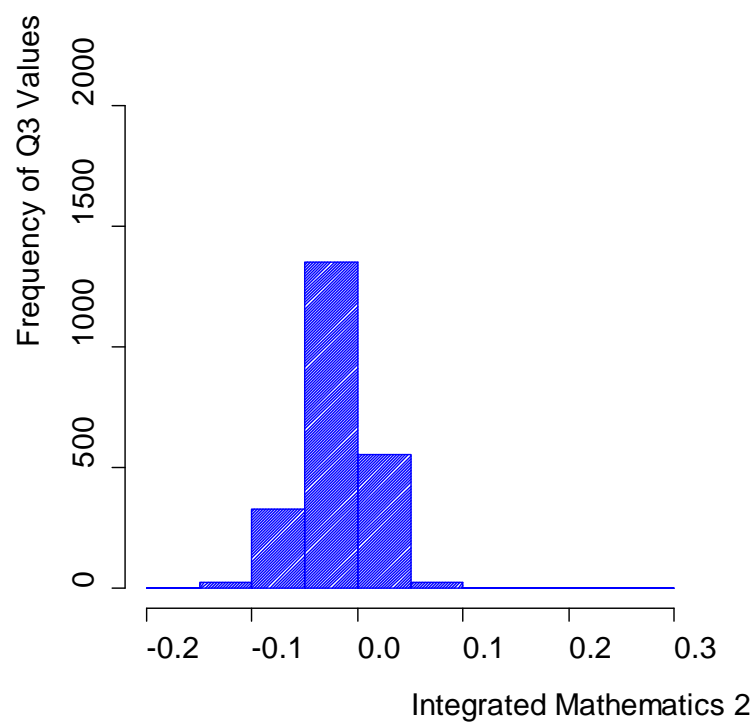


Figure 9.3 Distribution of Q_3 Values for Integrated Mathematics II (Spring 2015)

9.4 Evidence Based on Relationships to Other Variables

Empirical results concerning the relationships between score on a test and measures of other variables external to the test can also provide evidence of validity when these relationships are found to be consistent with the definition of the construct that the test is intended to measure. As indicated in the AERA, APA and NCME Standards (2014), the variables investigated can include other tests that measure the same construct and different constructs, criterion measures that scores on the test are expected to predict, as well as demographic characteristics of test takers that are expected to be related and unrelated to test performance.

The relationship of the scores across the ELA/L and mathematics assessments was evaluated using correlational analyses. Tables 9.24 through 9.29 present the Pearson correlations observed between the ELA/L scale scores and the mathematics scale scores for each grade; the correlations are reported separately for online (CBT) and paper (PBT) versions of the tests. For grades three through 8, students must have a valid test score for both ELA/L and mathematics at the same grade level to be included in the tables. These tables provide the correlation in the lower triangle and the sample size is provided in the upper triangle. In computing the correlations between a particular pair of ELA/L and mathematics tests, test takers must have taken both tests via the same mode in spring 2016. ELA/L, Reading (RD), Writing (WR), are moderately to highly correlated with mathematics; the correlations range from .64 up to .80 for grades 3 through 8, and range from .38 to .71 for the high school tests. These correlations suggest that the ELA/L and mathematics tests are assessing different content. The higher intercorrelations between the ELA/L, Reading (RD), and Writing (WR) suggest stronger internal relationships when compared to the correlations with the mathematics content area.

The ELA/L and mathematics correlations for the high school tests are presented in Tables 9.30 through 9.32. Because students in high school can take the mathematics courses in different years (e.g., one student may take Algebra I in grade 9 while another student may take Algebra I in grade 10), the high school mathematics scores were correlated with several of the ELA/L grades (e.g., Algebra I correlated with both grades 9 and 10). Only correlations for pairings with total sample sizes of at least 100 are shown in the tables. Correlations between high school mathematics scores and corresponding ELA/L scores demonstrate low to moderate correlations.

Table 9.24 Correlations between ELA/L and Mathematics for Grade 3

	CBT					PBT			
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		371,112	371,112	371,112	ELA/L		97,697	97,697	97,697
RD	0.96		371,112	371,112	RD	0.96		97,697	97,697
WR	0.86	0.70		371,112	WR	0.88	0.74		97,697
MA	0.80	0.78	0.68		MA	0.79	0.77	0.68	

Note: ELA/L = English language arts/Literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 3 ELA/L score *and* a valid grade 3 mathematics score to be included in this table.

Table 9.25 Correlations between ELA/L and Mathematics for Grade 4

	CBT					PBT			
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		375,844	375,844	375,844	ELA/L		82,186	82,186	82,186
RD	0.95		375,844	375,844	RD	0.95		82,186	82,186
WR	0.90	0.75		375,844	WR	0.90	0.73		82,186
MA	0.78	0.77	0.68		MA	0.76	0.75	0.66	

Note: ELA/L = English language arts/Literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 4 ELA/L score *and* a valid grade 4 mathematics score to be included in this table.

Table 9.26 Correlations between ELA/L and Mathematics for Grade 5

	CBT					PBT			
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		402,929	402,929	402,929	ELA/L		49,599	49,599	49,599
RD	0.96		402,929	402,929	RD	0.95		49,599	49,599
WR	0.87	0.72		402,929	WR	0.89	0.72		49,599
MA	0.77	0.75	0.67		MA	0.76	0.75	0.65	

Note: ELA/L = English language arts/Literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 5 ELA/L score *and* a valid grade 5 mathematics score to be included in this table.

Table 9.27 Correlations between ELA/L and Mathematics for Grade 6

	CBT					PBT			
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		400,243	400,243	400,243	ELA/L		50,107	50,107	50,107
RD	0.95		400,243	400,243	RD	0.95		50,107	50,107
WR	0.87	0.72		400,243	WR	0.88	0.73		50,107
MA	0.78	0.78	0.64		MA	0.79	0.78	0.66	

Note: ELA/L = English language arts/Literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 6 ELA/L score *and* a valid grade 6 mathematics score to be included in this table.

Table 9.28 Correlations between ELA/L and Mathematics for Grade 7

CBT					PBT				
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		377,972	377,972	377,972	ELA/L		50,271	50,271	50,271
RD	0.95		377,972	377,972	RD	0.95		50,271	50,271
WR	0.88	0.74		377,972	WR	0.90	0.73		50,271
MA	0.78	0.78	0.66		MA	0.79	0.79	0.67	

Note: ELA/L = English language arts/Literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 7 ELA/L score *and* a valid grade 7 mathematics score to be included in this table.

Table 9.29 Correlations between ELA/L and Mathematics for Grade 8

CBT					PBT				
	ELA/L	RD	WR	MA		ELA/L	RD	WR	MA
ELA/L		306,787	306,787	306,787	ELA/L		42,825	42,825	42,825
RD	0.95		306,787	306,787	RD	0.95		42,825	42,825
WR	0.88	0.72		306,787	WR	0.91	0.75		42,825
MA	0.74	0.74	0.62		MA	0.79	0.78	0.68	

Note: ELA/L = English language arts/Literacy, RD = Reading, WR = Writing, MA = Mathematics. The correlations are provided in the lower portion of the table and the sample sizes are provided in the upper portion of the table. Students must have a valid grade 8 ELA/L score *and* a valid grade 8 mathematics score to be included in this table.

Table 9.30 Correlations between ELA/L and Mathematics for High School

ELA/L	CBT						ELA/L	PBT					
	A1	GO	A2	M1	M2	M3		A1	GO	A2	M1	M2	M3
8	0.69 (69,594)	0.58 (7,976)	0.61 (379)	0.59 (1,313)			8	0.71 (4,457)	0.44 (797)				
9	0.67 (159,598)	0.67 (35,269)	0.65 (7,944)	0.70 (14,000)	0.60 (1,290)	0.62 (166)	9	0.71 (7,462)	0.64 (701)	0.54 (154)			
10	0.53 (11,827)	0.64 (76,161)	0.67 (32,522)		0.51 (2,357)		10	0.57 (557)	0.65 (2,280)	0.68 (920)			
11	0.44 (2,590)	0.50 (9,540)	0.59 (67,631)		0.59 (108)	0.66 (1,526)	11		0.52 (297)	0.68 (2,150)		0.68 (190)	

Note: ELA/L = English language arts/Literacy, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. The correlations are provided with the sample sizes, below in parentheses. Shaded cells indicate pairings with sample sizes less than 100.

Table 9.31 Correlations between ELA/L Reading and Mathematics for High School

RD	CBT						RD	PBT					
	A1	GO	A2	M1	M2	M3		A1	GO	A2	M1	M2	M3
8	0.69 (69,594)	0.57 (7,976)	0.63 (379)	0.58 (1,313)			8	0.71 (4,457)	0.48 (797)				
9	0.66 (159,598)	0.67 (35,269)	0.66 (7,944)	0.70 (14,000)	0.61 (1,290)	0.62 (166)	9	0.70 (7,462)	0.66 (701)	0.60 (154)			
10	0.53 (11,827)	0.65 (76,161)	0.69 (32,522)		0.53 (2,357)		10	0.57 (557)	0.65 (2,280)	0.67 (920)			
11	0.43 (2,590)	0.52 (9,540)	0.60 (67,631)		0.60 (108)	0.68 (1,526)	11		0.50 (297)	0.67 (2,150)		0.67 (190)	

Note: RD = Reading, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. The correlations are provided with the sample sizes, below in parentheses. Shaded cells indicate pairings with sample sizes less than 100.

Table 9.32 Correlations between ELA/L Writing and Mathematics for High School

WR	CBT						WR	PBT					
	A1	GO	A2	M1	M2	M3		A1	GO	A2	M1	M2	M3
8	0.58 (69,549)	0.49 (7,976)	0.50 (379)	0.49 (1,313)			8	0.57 (4,457)	0.30 (797)				
9	0.55 (159,598)	0.55 (35,269)	0.53 (7,944)	0.56 (14,000)	0.44 (1,290)	0.47 (166)	9	0.59 (7,462)	0.50 (701)	0.41 (154)			
10	0.43 (11,827)	0.52 (76,161)	0.53 (32,522)		0.37 (2,357)		10	0.46 (557)	0.53 (2,280)	0.58 (920)			
11	0.38 (2,590)	0.39 (9,540)	0.47 (67,631)		0.47 (108)	0.52 (1,526)	11		0.48 (297)	0.57 (2,150)		0.59 (190)	

Note: WR = Writing, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. The average correlations are provided with the sample sizes, below in parentheses. Shaded cells indicate pairings with sample sizes less than 100.

9.5 Evidence from the Special Studies

Several research studies were conducted to provide additional validity evidence for the PARCC's goals of assessing more rigorous academic expectations, helping to prepare students for college and careers, and providing information back to teachers and parents about their students' progress towards college and career readiness. Some of the special studies conducted include:

- content evaluation studies,
- benchmarking study,
- mode comparability study, and
- device comparability study.

The following paragraphs briefly describe each of these studies.

9.5.1 Content Alignment Studies

The content of the ELA/L assessments at grades 5, 8, and 11 and the Algebra II and Integrated Mathematics II assessment were evaluated to determine how well the PARCC assessments were aligned to the Common Core State Standards (CCSS; Doorey, & Polikoff, 2016, Schultz, Michaels, Dvorak, & Wiley, 2016). These content alignment studies were conducted by the Fordham Institute for grades 5 and 8 and by Human Resources Research Organization (HumRRO) for the high school assessments. Both of these studies used the same methodology by having content experts review the assessment items and answers (for the constructed response items the rubrics were reviewed). The content experts then judged how well the items aligned to the CCSS, the depth of knowledge of the items, and the accessibility of the items to all students, including English learners and students with disabilities. The authors of both studies noted that the content experts reviewing the assessments were required to be familiar with the CCSS but could not be employed by participating organizations or be the writers of the CCSS. Therefore, an effort was made to eliminate any potential conflicts of interest.

The content studies had the individual content experts review and rate each item then as a group the content experts came to a consensus on the final ratings for the content alignment, depth of knowledge, and accessibility to all students. In addition to the ratings, the content experts were asked to make comments that provided an explanation of their ratings; these comments were then used by the full group of content experts to provide narrative comments regarding the overall ratings and to provide feedback and recommendation about the assessment programs.

The PARCC assessment program was rated as *Excellent Match* for ELA/L content and depth and *Good Match* for mathematics content and depth for grades 5 and 8. However, for grade 11 ELA/L content was rated as *Excellent Match* but depth was rated as *Limited/Uneven Match*. The high school mathematics assessments were rated at *Excellent Match* for content and *Good Match* for depth.

The content studies noted some weaknesses and strengths of the PARCC assessments. For ELA/L it was noted that the assessments include complex texts, a range of cognitive demands, and have a variety of item types. Furthermore, the ELA/L "assessments require close reading, assess writing to sources,

research, and inquiry, and emphasize vocabulary and language skills” (Doorey & Polikoff, 2016). The grade 11 ELA/L assessment had a smaller range of depth and included items assessing the higher-demand cognitive level. A weakness of the ELA/L assessments is the lack of a listening and speaking component. It was also suggested that the ELA/L assessments could be enhanced by the inclusion of a research task that requires the use of two or more sources of information.

The strengths of the mathematics assessments include assessments that are aligned to the major work for each grade level. While the grade 5 assessment includes a range of cognitive demand, the grade 8 assessment includes a number of higher-demand items and may not fully assess the standards at the lowest level of cognitive demand. It was suggested that the grade 5 assessment could include more focus on the major work and the grade 8 assessment could include items at the lowest cognitive demand level. Additionally, the reviewers noted that some of the mathematics items should be carefully reviewed for editorial and mathematical accuracy.

The high school report noted that the PARCC assessment program incorporates a number of accessibility features and test accommodations for students with disabilities and for English language learners. Furthermore, the PARCC assessments included items designed to accommodate the needs of students with disabilities.

9.5.2 Benchmarking Study

The purpose of the PARCC Benchmarking Study (McClarty, Korbin, Moyer, Griffin, Huth, Carey, and Medberry, 2015) was to provide information that would inform the PARCC performance level setting (PLS) process. PARCC used an Evidence-Based Standard Setting approach (EBSS; McClarty, Way, Porter, Beimers, & Miles, 2013) to establish the performance levels for its assessments. In EBSS, the threshold scores for performance levels are set based on a combination of empirical research evidence and expert judgment. This benchmarking study provided one source of empirical evidence to inform the PARCC college and career readiness performance level (i.e., Level 4). The study findings were provided to PARCC’s pre-policy standard-setting committee. The charge of this committee was to suggest a reasonable range for the percentage of students meeting or exceeding the PARCC Level 4 threshold score and therefore considered college- and career-ready. Section 11.3.2 of this report provides more information about the PARCC pre-policy meeting. For the PARCC Benchmarking Study, external information was analyzed to provide information about the Level 4 threshold scores for the grade 11 ELA/literacy, Algebra II, and Integrated Mathematics III assessments, the grade 8 ELA/literacy and mathematics assessments, and the grade 4 ELA/literacy and mathematics assessments. The PARCC assessments and Level 4 expectations were compared with comparable assessments and expectations for the Programme of International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), National Assessment of Educational Progress (NAEP), ACT, SAT, the Michigan Merit Exam, and the Virginia End-of-Course exams. For each external assessment, the best-matched performance level was determined and the percentage of students reaching that level across the nation and in the PARCC states was determined. Across all grades and subjects, the data indicated approximately 25 to 50 percent of students were college- and career-ready or on track to readiness based on PARCC’s Level 4 expectations.

For details on how the benchmarking study was used during the standard setting process, refer to Section 11 of this technical report.

9.5.3 Mode Comparability Study

The PARCC (Operational) Mode Comparability Study was conducted using the 2015 operational data to support both computer-based testing (CBT) and paper-based testing (PBT) modes of administration of the PARCC assessments (Liu, Brown, Chen, Ali, Hou, & Costanzo, 2016). PARCC has plans to conduct a follow-up Mode Comparability Study after the spring 2017 administration. The following provides a summary of the study conducted in 2015.

For the spring 2015 operational administration, schools and districts within each state selected the mode of test administration. The resulting CBT and PBT test-taking groups were therefore not randomly equivalent. To improve the overall comparability of the CBT and PBT groups, propensity score matching, based on test-taker demographic information, was used. Then item-level analyses (e.g., p values, and differential item functioning) and test-level analyses (e.g., test characteristic curves) were conducted.

Item-level analyses showed that there were negligible to small differences in terms of p values and average item scores across modes for the majority of items in mathematics and ELA/L. Prose Constructed Response (PCR) task traits in ELA/L had larger p value effect sizes than other items, all favoring PBT. A very small percentage of items was identified as functioning differently (with C-level DIF) in the two modes. Many items ELA/L PCR task traits were also found to have B-level (DIF), favoring PBT.

Additionally, the item response theory (IRT) difficulty and discrimination parameters estimated separately within mode were highly correlated. For grade levels with lower correlations between modes, removing items with outlier parameter estimates provided substantial improvement in the correlation. As well, the overall the differences between common test characteristic curves (TCCs) of different modes were small and within 0.5 raw score points, except for ELA/L grade 9 and Geometry where TCC differences exceeded the differences that matter criterion in regions of the theta scale where large percentages of students were located. When comparing the performance on the common items, the effect sizes ranged from negligible to small for most of the tests evaluated. The directions of effect sizes were not consistent across subject/grade levels.

Additional analyses were conducted on students from one of the states that provided prior state assessment scores. Summary statistics of these students' prior state assessment scores suggested CBT and PBT samples from propensity score matching (PSM) were not comparable in their prior achievement. Therefore, poststratification weights based on prior state assessment score were used to calculate PBT students' PARCC scale score to minimize the impact of noncomparability of prior achievement across modes. The scale score differences were largely reduced for mathematics grade 5, 7 and Algebra I after weighting. Small effect sizes, in favor of PBT, were found for Geometry and ELA/L grade 9 and a negligible effect size was found for ELA/L grade 7 after poststratification weighting.

The PARCC (Operational) Mode Comparability Study found evidence that the score comparability was not consistent across all content domains and grade levels. As noted in the study, only one state

provided previous year's achievement data, therefore, the CBT and PBT groups were matched based on only demographic data. Furthermore, the additional analyses based on the one state that provided prior achievement data indicated that the CBT and PBT matched groups were not comparable in terms of their prior achievement. Thus, caution should be taken when interpreting the results of the Mode Comparability Study.

9.5.4 Device Comparability Study

In addition to the PARCC (Operational) Mode Comparability Study, the comparability across digital devices (e.g., tablet versus non-tablet) was evaluated using the 2015 operational data (Steedle, McBride, Johnson, & Keng, 2015). PARCC has plans to conduct a follow-up digital devices study after the spring 2017 administration. The following provides a summary of the study conducted in 2015.

PARCC allows students to take its assessments on a variety of digital devices, such as desktops, laptops, and tablets. It is therefore important to evaluate comparability across digital devices by investigating whether test items were of similar difficulty, whether psychometric properties of test scores were similar, and whether overall test score interpretation was similar across traditional (i.e., desktops and laptops) and non-traditional (i.e., tablet) computing devices. For the 2015 Device Comparability Study, any student who took one of the study forms on a tablet or non-tablet device were eligible for inclusion in the study. Students were matched on demographic information to create tablet and non-tablet samples that were considered randomly equivalent.

The 2015 Device Comparability Study found evidence of comparability between test scores from tablets and non-tablet devices. The item p values and IRT difficulty estimates were similar across tablets and non-tablet devices. A small number of items were flagged for device effects, and nearly all of them were part of high school mathematics assessments. The raw score and scale score distributions indicated similar overall performance on both components (PBA and EOY) of the 2015 PARCC assessments. Additionally, IRT true-score equating indicated that students who tested on non-tablet devices would be expected to score similarly had they taken the same PARCC assessment on tablets.

9.6 Evidence Based on Response Processes

As noted in the AERA, APA, and NCME Standards (2014), additional support for a particular score interpretation or use can be provided by theoretical and empirical evidence indicating that test takers are using the intended response processes when responding to the items in a test. This type of evidence may be gathered from interacting with test takers in order 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 constructed response items. Evidence may also be gathered by evaluating the correct and incorrect responses to short constructed response items (e.g., items requiring a few words to respond) or by evaluating the response patterns to multi-part items.

PARCC has undertaken research investigating the quality of the items, tasks, and stimuli, focusing on whether students interact with items/tasks as intended, whether they were given enough time to complete the assessments, and the degree to which PARCC scoring rubrics allow accurate and reliable scoring. In addition, PARCC has examined the accessibility of the test for students with disabilities and English learners. This research has included examining students' understanding of the format of the assessments and the use of technology. Although out of the purview of this technical report, several other PARCC research efforts have investigated questions relevant to response processes evidence.⁹

9.7 Interpretations of Test Scores

The PARCC ELA/L and mathematics scores are expressed as scale scores (both total scores and claim scores), along with performance levels to describe how well students met the academic standards for their grade level. Additionally, information on specific skills (the subclaims) is also provided and is reported as “Below Expectations”, “Nearly Meets Expectations” and, “Meets or Exceeds”. On the basis of a student’s total score, an inference is drawn about how much knowledge and skill in the content area the students has acquired. The total score is also used to classify students in terms of the level of knowledge and skill in the content area as students progress in the K-12 education. These levels are called performance levels and are reported as:

- Level 5: Exceeded expectations
- Level 4: Met expectations
- Level 3: Approached expectations
- Level 2: Partially met expectations
- Level 1: Did not yet meet expectations

Students classified as either Level 4 or Level 5 are meeting or exceeding the grade level expectations. PARCC has developed performance level descriptors (PLDs) to assist with the understanding and interpretations of the ELA/L and mathematics scores (<http://www.parcconline.org/news-and-video/230-performance-level-descriptors>). Additionally, resource information is available online to educators, parents, and students (<http://www.parcconline.org/resources>), which includes information on understanding and interpreting the ELA/L and mathematics score reports (<http://www.state.nj.us/education/assessment/parcc/scores/Fall14Spring15SRIG.pdf> and <http://www.parcconline.org/assessments/score-results>).

Section 12 of this technical report provides more information on the scale scores and the subclaim scores.

9.8 Evidence Based on the Consequences to Testing

The consequence of testing should also be investigated to support the validity evidence for the use of the PARCC assessments as the Standards note that tests are usually administered “with the expectation

⁹ Various PARCC research is described at: <http://www.parcconline.org/assessments/test-design/research>

that some benefit will be realized from the intended use of the scores” (AERA, APA, & NCME, 2014). When this is the case, evidence that the expected benefits accrue will provide support for the intended use of the scores. Evidence of the consequence of testing will also accrue with the continued implementation of the CCSS and the continued administration of the PARCC assessments.

Consequences of the PARCC tests may vary by state or by school district. For example, some states may require “passing” the PARCC assessments as one of several criteria for high school graduation, while other states/districts may not require students to “pass” the PARCC assessments for high school graduation. Additionally, some school districts may use the PARCC scores along with other information such as school grades and teacher recommendations for placing students into special programs (e.g., remedial support, gifted and talented program) or for course placement (e.g., Algebra I in grade 8). Because the consequences for the PARCC assessments can vary by each state, it is suggested that each PARCC member state provide school districts, teachers, parents, and students with information on how to interpret and use the PARCC scores. Additionally, the states should monitor how PARCC scores are used to ensure that the scores are being used as intended by PARCC.

9.9 Summary

In this section of the technical report, several aspects of validity were included, such as validity evidence based on content, the internal structure of the assessments, relationships across the content assessments, and from special studies.

The PARCC item development process involved educators, assessment experts, and bias and sensitivity experts in review of text, items and tasks for accuracy, appropriateness, alignment to the instructional standards, and freedom from bias. PARCC conducted several studies during the item development process to evaluate the item development process (e.g., technological functionalities, answer time required, and student experiences). Additionally, items were field tested prior to the initial operational administration and data and feedback from students, test administrators, classroom teachers was used to improve the operational administration of the items and to inform future item development. The multiple item and form reviews conducted by educators and studies to evaluate item administration help to ensure the integrity of the PARCC assessments.

The intercorrelations of the subclaims, the reliability analyses, and the local item dependence analyses indicated that the ELA/L and the mathematics assessments are both essentially unidimensional. Furthermore, the correlations between ELA/L and mathematics indicated that the two assessments are measuring different content. Also, the patterns of correlations for the CBT and PBT assessments were similar indicating that the structure of the assessments were similar across the two modes.

Several studies were conducted as part of the PARCC assessment program (e.g., benchmarking study, content evaluation/alignment studies, mode and device comparability studies). The benchmarking study was conducted in support of the standard setting meeting. This study indicated students performing at or above Level 4 could be considered to be college- and career-ready or on track to readiness.

The content evaluation/alignment studies performed by the Fordham Institute and HumRRO indicate that the PARCC assessments are good to excellent matches to the CCSS in terms of content and depth of knowledge. Thus, the PARCC assessments are assessing the college- and career-readiness standards. However, the reports noted that the PARCC program could improve by adding a wider range of depth of knowledge to some of the assessments. The reports also suggested enhancing the ELA/L assessments by including a research task that requires the use of two or more sources of information.

The mode comparability study indicated that the comparability across modes was inconsistent across content domains and grade levels. The mode comparability study indicated that outliers should be removed from the mode anchor set and that the PCR items should be considered for exclusion from the anchor set. Furthermore, the scoring of the PCR items should be carefully reviewed.

The device comparability study indicated that there were some, but small, effects of testing device when comparing tablet to non-tablet devices. While a small number of mathematics tasks were flagged for device effects, the raw and scale score distributions were similar across the testing devices. The equating analyses indicated that students could expect to receive a similar score regardless of the testing device.

In addition to the validity information presented in this Section of the technical report, other information in support of the uses and interpretations of the PARCC scores appear in the following sections:

Section 5 presents information regarding student characteristics for the spring administration of the ELA/L and mathematics administration.

Section 6 provides information concerning the test characteristics based on classical test theory.

Section 7 provides information regarding the differential item functioning analyses (DIF).

Section 8 provides information on the test reliability (total test score and for subclaims) and includes information on the interrater reliability/agreement.

Section 12 provides detailed information concerning the scores that were reported and the cut scores for ELA/L and mathematics.

The technical report addendum provides the test taker characteristics and test reliability (total test score and for subclaims) for the 2015 Fall block administration.

Section 10: IRT Calibration and Scaling in Operational Year Two

10.1 Overview

Multiple operational core forms were administered for each grade in English language arts/literacy (ELA/L) and mathematics. The purpose of the item response theory (IRT) calibration and scaling was to place all operational items for a single grade/subject onto a common scale. This section addresses procedures used to calibrate and scale the PARCC operational item response data using IRT. The operational data were calibrated concurrently across forms, and separately by mode (computer-based tests, or CBT, and, paper-based tests, or PBT) using IRT models consistent with mixed format data. The PBT IRT parameter estimates were then transformed onto the CBT scale using the Stocking and Lord (1983) procedure.

In this section of the technical report, the following topics related to IRT calibration and scaling are discussed:

Calibration:

- 10.2 IRT data preparation
- 10.3 Description of the calibration process
- 10.4 Model fit evaluation criteria
- 10.5 Items excluded from score reporting

Scaling:

- 10.6 Scaling Parameter Estimates
- 10.7 Items Excluded from Linking Sets
- 10.8 Correlations and Plots of Parameter Estimates
- 10.9 Scaling constants
- 10.10 Summary Statistics and Distributions from IRT Analyses

10.2 IRT Data Preparation

10.2.1 Overview

Post-equating was performed on an early sample of the student data. The Bureau of Indian Education (BIE), the District of Columbia (DC), and seven states were participating in the spring 2016 administration: Colorado (CO), Illinois (IL), Maryland (MD), Massachusetts (MA), New Jersey (NJ), New Mexico (NM), and Rhode Island (RI). Each state tested over multiple weeks. Student data were monitored to determine when the early equating sampling criteria were met. Student data were evaluated for the following:

- 1) Overall N-count, form count, and item count
- 2) Demographic Representation
- 3) State Representation

4) Summative Scale Score Distribution

Based on the Early Equating Report, using data from the spring 2015 PARCC administration, approximately 25% (75,000–90,000) of the online student data were sufficient for post-equating selected tests in grades 3–8 ELA/L and mathematics. Approximately 40% (70,000–90,000) of the online student data were sufficient for post-equating selected tests in grades 9–11 ELA/L and traditional mathematics. The larger percent of the student data for high school assessments was due to the high percent of students removed for not meeting attemptedness criteria and the need to obtain student responses for each score category for the more difficult items. Due to the small number of students taking the integrated mathematics assessments, approximately 90% of the student data were needed for post-equating. The results from the research study were used to determine criteria for sample size and acceptable differences between the baseline demographic distributions and the sample demographic distributions for the spring 2016 post-equating.

The resulting early equating samples for the spring 2016 administration exceeded state representation, exceeded the sample size criteria, met criteria for most of the demographic groups, and met criteria for the prior grade's PARCC performance level distributions for most of the grades/subjects. Tables 10.1 and 10.2 list the equating sample sizes by administration mode for each grade in ELA/L and mathematics, respectively.

All student response data in the early equating samples for operational items were used to create the IRT sparse data matrices for the concurrent calibration. IRT sparse data matrices combine student data across forms within administration mode. When duplicate records for a single student existed, the record with the largest raw score was included in the data file (and the other record was excluded). No student was included more than one time in the CBT and PBT IRT sparse data matrices file.

10.2.2 Student Inclusion/Exclusion Rules

The following are the IRT valid case criteria. These criteria are the same as the student inclusion/exclusion rules used to evaluate and filter data prior to conducting the operational item analysis (IA) and differential item functioning (DIF) analyses (steps 1-5). The rules were agreed upon with PARCC and applied to the scored data used in the IRT calibration.

1. All records with an invalid form number were excluded.
2. All records that were flagged as “void” were excluded.
3. Records in which the student attempted fewer than 25% of the items were excluded. An item was deemed “not attempted” if, it had a value of “M” (item omitted) or “Z” (item ‘spoiled’, do not score) in the scored item response block. For example, if there were 25 items on a form and two were flawed (“Z”), those two items were not included in the numerator or denominator of the percentage attempted calculation.
4. For students with more than one valid record, the record with the higher raw score was chosen.
5. Records for students with administration issues or anomalies were excluded.

An assumption was made that mathematics items translated into Spanish were equivalent to the same items in English. The results of Spanish versus English differential item functioning (DIF) analyses supported this assumption (see Appendix 7 Table A.7.22).

10.2.3 Items Excluded from IRT Sparse Matrices

Pearson conducted an initial scoring and key check. Items identified by Pearson as “spoiled” (also referred to as “do not use (DNU)”) were listed and excluded from the test maps. When the IRT sparse data matrices were created, all items were included in the files unless they were marked as “spoiled” by Pearson.

10.2.4 Omitted, Not Reached and Not Presented Items

In the student data files, ‘Z’ was used to represent “spoiled” or “not presented” items and ‘M’ was used to represent omitted items. For IA and IRT, *omitted* and *not reached* items were treated differently. Item response scores for omits were recoded as ‘O’ in the IRT sparse matrix files (i.e., *unless* the omitted item was a “not reached” item). Not reached items are omitted items at the end of the test or unit – items that the student probably did not reach or try to answer. Not Reached items were recoded from ‘M’ in the SIRB to ‘N’ (i.e., not presented) in the IRT sparse matrix files, if all items from that point, to the end of the form, are ‘M’ or ‘Z’. Not reached items were counted as *missing* or *no response*, and therefore did not contribute to the item statistics for IA and IRT calibration.

10.2.5 Quality Control of the IRT Sparse Matrix Data Files

The IRT sparse data matrices were created by the primary analysts and replicators from Pearson and HumRRO. The matrices were checked for quality and accuracy by comparing the number of students (*N*-counts), item category frequencies, and item statistics (e.g., AIS values) between Pearson and HumRRO. Since the same inclusion rules for students were used, all *N*-counts, category frequencies, and statistics for all items matched. All discrepancies in *N*-counts were resolved. The programs used to create the IRT statistics were independent, so the QC procedure involved parallel computing. Tables 10.1 and 10.2 show the *N*-Counts (*N*), percentage of students (Percent), and number of items (*n*-Items) in the CBT and PBT IRT sparse data matrices for each grade in ELA/L and mathematics, respectively.

Table 10.1 N-Counts, Percent of Students, and Number of Items in the ELA/L IRT Calibration Files

Grade	N			Percent		n-Items	
	ALL	CBT	PBT	CBT	PBT	CBT	PBT
3	321,130	258,460	62,670	80.5	19.5	65	52
4	290,320	230,933	59,387	79.5	20.5	85	50
5	309,881	277,599	32,282	89.6	10.4	84	44
6	282,219	255,682	26,537	90.6	9.4	80	56
7	369,023	337,798	31,225	91.5	8.5	88	72
8	382,313	351,012	31,301	91.8	8.2	84	56
9	266,750	252,364	14,386	94.6	5.4	89	66
10	136,305	128,773	7,532	94.5	5.5	102	58
11	129,964	124,097	5,867	95.5	4.5	78	64

Table 10.2 N-Counts, Percent of Students, and Number of Items in the Mathematics IRT Calibration Files

Grade/ Subject	N			Percent		n-Items	
	ALL	CBT	PBT	CBT	PBT	CBT	PBT
3	165,838	105,798	60,040	63.8	36.2	97	75
4	283,218	237,951	45,267	84.0	16.0	101	65
5	217,178	176,389	40,789	81.2	18.8	95	69
6	288,187	250,305	37,882	86.9	13.1	101	63
7	199,278	173,163	26,115	86.9	13.1	103	65
8	205,031	173,980	31,051	84.9	15.1	89	63
A1	211,197	193,784	17,413	91.8	8.2	100	69
GO	119,875	114,802	5,073	95.8	4.2	109	79
A2	117,223	109,395	7,828	93.3	6.7	101	73
M1	14,606	14,606	n/a	100	n/a	42	n/a
M2	3,766	3,766	n/a	100	n/a	42	n/a
M3	2,081	2,081	n/a	100	n/a	40	n/a

Note: A1 = Algebra I, A2 = Algebra II, GO = Geometry, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, and M3 = Integrated Mathematics III.

10.3 Description of the Calibration Process

The Item Response Theory (IRT) calibrations were performed separately by mode on the IRT sparse data matrix. The form-to-form linking is established through internal and external common-items selected during test construction to represent the blueprint.

10.3.1 Two-Parameter Logistic/Generalized Partial Credit Model

The operational IRT analyses were conducted by both Pearson and HumRRO. The operational items in the IRT sparse data matrix were concurrently calibrated with the two-parameter logistic/generalized partial credit model (2PL/GPC: Muraki, 1992). The 2PL/GPC is denoted

$$p_{im}(\theta_j) = \frac{\exp \left[\sum_{k=0}^m Da_i(\theta_j - b_i + d_{ik}) \right]}{\sum_{v=0}^{M_i-1} \exp \left[\sum_{k=0}^v Da_i(\theta_j - b_i + d_{iv}) \right]},$$

where $a_i(\theta_j - b_i + d_{io}) \equiv 0$; $p_{im}(\theta_j)$ is the probability of a test taker with θ_j getting score m on item i ; M_i is the number of score categories of item i with possible item scores as consecutive integers from 0 to $M_i - 1$; D is the IRT scale constant (1.7). IRT calibrations might also use a guessing parameter in special cases, if needed.

10.3.2 Treatment of Prose Constructed Response (PCR) Tasks

For prose constructed response (PCR) tasks, the student received a prompt and wrote a response which was then scored using a multi-trait rubric. An *aggregated* PCR item score was determined by adding together the multiple scores the student received on the two traits. The PCR scoring rubric for the Research Simulation Task and the Literary Analysis Task were modified for the 2016 administration. The original (2014-2015) Written Expression (WE) and Reading Comprehension (RC) traits were combined into a single trait called Reading Comprehension and Writing Expression (RCWE). Therefore, two trait scores were assigned for these PCRs: the Reading Comprehension and Writing Expression (RCWE) or Writing Knowledge and Conventions (WKL). One of the writing traits (Written Expression) was weighted by 4 to give it more emphasis in the total score and weighted by 3 to give it more emphasis in the writing claim score. The aggregated PCR scores have total maximum points possible range from 12 to 19 depending on the item and the grade.

The PCRs were calibrated at the trait score level (and not as aggregated scores). To address the issue of local independence related to PCR items, a single-calibration “model” approach was used. When sample sizes were large (i.e., greater than 10,000 test takers), the data were manipulated using random assignment, by selecting one of the two traits for each PCR item for each student. Then one calibration was run so that all trait parameters were independently estimated. When sample sizes were smaller (i.e., field-test samples), a multiple-calibration “model” approach was used. In this alternative approach, the same data set was calibrated two times, each trait represented in one of the two data sets for all students. Then the PCR traits were scaled onto the base scale using non-PCR items as anchor items. These two trait calibration approaches addressed the issue of local dependence, and also allowed for the accurate calculation of claim scores, and the proper weighting of traits in the summative scale scores.

10.3.3 IRT Item Exclusion Rules (Before Calibration)

In addition to checking IRT data for accuracy, Pearson conducted item analyses (IA) to identify items that were not performing as expected and should be considered for removal from calibration and score reporting. The following are the criteria Pearson used to flag extremely problematic items to be dropped from calibration. All “non-spoiled” items were included in the IRT data matrices, however, the IRTPRO

calibration software (Cai, Thissen, & du Toit, 2011) control files were used to exclude from calibration items flagged for the following reasons:

1. A weighted polyserial correlation less than 0.0
2. An average item score of 0.0
3. 100% of the students have the same item score, such as:
 - a. 100% omitted the item,
 - b. 100% received the same score,
 - c. 100% of the responses were at the same score after collapsing score categories due to low frequencies, or
 - d. 100% of the responses were not presented or not reached
4. Insufficient sample sizes for the selected IRT model combinations (i.e., 300 for the 2PL/GPC).
5. High omit rates (i.e., greater than 50%) on one or more forms (usually an indication that an item may not be functioning correctly on all forms).

A master list of all problematic items before and after calibration was maintained and all flagged and potentially flawed items were brought to the PARCC Priority Alert Task Force (consisting of Parcc Inc. and participating State Leads for PARCC member states) for content and statistical reviews. Ultimately, the decisions about whether to keep or exclude an item from score reporting was made by the PARCC Priority Task Force.

10.3.4 IRTPRO Calibration Procedures and Convergence Criteria

The data were calibrated concurrently across forms and separately by mode (CBT and PBT) using the 2PL/GPC model combination. The PBT parameter estimates were then transformed onto the CBT scale using the Stocking and Lord (1983) procedure. The primary goal was to place the operational item data within each content area and grade/subject on a common difficulty scale. The following are the steps used to calibrate the Spring 2016 operational item response data:

1. Using the IRT sparse data matrices, concurrent calibrations were conducted using commercially available IRTPRO for Windows (version 3.0) on CBT data, and separately on PBT data within each grade/subject.
2. The 2PL/GPC model combination was used for all grades and subjects for each content area. Thus, two calibrations were completed for each grade/subject.
3. IRTPRO Calibration Settings: The logistic partial credit model was specified using the scale constant of 1.0. The prior distributions for latent traits were set to a mean of 0.00 and a SD of 1.00. The number of quadrature points used in the estimation was set to 49. And, the slope starting value was set/updated before each run.
4. Each IRTPRO run was inspected for convergence and for any unexpected item-parameter estimates. The PRIORS command in IRTPRO provided a prior on IRT parameters to constrain the calibration so that convergence was more likely. Specifically, option “Guessing[0]” indicated that the prior is placed on the lower asymptote for the 3-PL model, and a normal

- distribution for the priors with mean of -1.4 and standard deviation 1. For these items, an inspection of item-level statistics and modal-data fit plots were sufficient to ensure that item parameters were acceptable if convergence was reached. Item information functions from the IRTPRO output may also be reviewed. Pearson verified that the maximum number of EM (expectation-maximization) cycles was not reached (which indicated the program did not converge).
5. To convert IRTPRO item parameters to the commonly used logistic parameter presentation (called new item parameters), the following formula was used since IRTPRO uses 1.0 for a scaling constant. There was no need to transfer *b*- and *c*-parameters from IRTPRO output. Please note that all un-scaled and scaled item parameters were kept on the theta scale. For 2PL or 3PL models:

$$\text{New } a\text{-parameter: } a_{\text{new}} = \frac{a_{\text{irtpro}}}{1.7}$$

6. Pearson reported any need for item-calibration decisions, including convergence issues and extreme parameter estimates, along with proposed resolutions, to the Priority Alert Task Force. Anticipated resolutions included fixing the slope parameters to a minimum .10 value, fixing the guessing parameter to a rational value (1 divided by number of options), and fixing the difficulty parameters at an upper or lower bound, depending on the nature of the problem. If extreme *b*-parameter values were observed (e.g., > 100) and the *a*-parameter values for these items were low (i.e., < 0.10), it was recommended that the prior for the *a*-parameter be set to 0.5.
7. Dropping an item from further processing or dropping an item and rerunning IRTPRO was performed only if it was needed after communication with HumRRO and the Priority Alert Task Force.
8. Inspection of model-data fit plots was helpful in deciding parameter constraints and acceptability of parameter fit. Documentation of each step, after resolution of any issues, was provided by Pearson to PARCC, HumRRO, and Measured Progress.

10.3.5 Calibration Quality Control

To ensure IRT calibrations and conversion tables were produced accurately, HumRRO replicated the IRT calibrations and the generation of the score conversion tables. Both Pearson and HumRRO used the same calibration software, IRTPRO. Daily meetings were held so that Pearson and HumRRO could provide status reports and discuss issues related to the IRT work. Measured Progress performed independent quality control comparisons between the Pearson and HumRRO item parameter estimates to identify any differences.

Specifically, Measured Progress completed the following quality control analyses/comparisons:

1. Made sure all items were treated the same way (e.g., if Pearson collapsed a category, made sure HumRRO collapsed the category in the same way for the item);
2. Compared IRT item parameter estimates by Pearson and HumRRO (i.e., IRT a -, b -, and d -parameter estimates);
3. Compared the scaling constants for the common item linking sets;
4. Compared transformed PBT parameter estimates generated by Pearson and HumRRO;
5. Compared all conversion tables produced Pearson and HumRRO to make sure they were accurate.

Measured Progress prepared reports documenting their findings. Exact matches were found between all Pearson and HumRRO conversion tables before scores were reported.

10.4 Model Fit Evaluation Criteria

The usefulness of IRT models is dependent on the extent to which they effectively reflect the data. As discussed by Hambleton, Swaminathan, and Rogers (1991), “The advantages of item response models can be obtained only when the fit between the model and the test data of interest is satisfactory. A poorly fitting IRT model will not yield invariant item and ability parameters” (p. 53).

After convergence was achieved for each IRT data set, the IRT model fit was evaluated by doing the following:

1. Calculating the Q_1 statistic and comparing it to a criterion score
2. Calculating the G^2 statistic and comparing it to a criterion score
3. Reviewing graphical output for all items

The Q_1 statistic (Yen, 1981) was used as an index of correspondence between observed and expected performance. To compute Q_1 , first the estimated item parameters and student response data (along with observed item scores) were used to estimate student ability ($\hat{\theta}$). Next, expected performance was computed on each item using students’ ability estimates in combination with estimated item parameters. Differences between expected item performance and observed item performance were then compared at 10 intervals across the range of student achievement (with approximately the same number of students per interval). Q_1 was computed as a ratio involving expected and observed item performance. Q_1 is interpretable as a chi-squared (χ^2) statistic, which can be compared to a critical chi-squared value to make a statistical inference about whether the data (observed item performance) were consistent with what might be observed if the IRT model was true (expected item performance). Q_1 is not directly comparable across different item types because items with different numbers of IRT parameters have different degrees of freedom (df). For that reason, a linear transformation (to a Z-score, Z_{Q_1}) was applied to Q_1 . This transformation also made item fit results easier to interpret and addresses the sensitivity of Q_1 to sample size.

To evaluate item fit, Yen's Q_1 statistic was calculated for all operational and field-test items. Q_1 is a fit statistics that compares observed and expected item performance. MAP (maximum *a posteriori*) estimates from IRTPRO was used as student ability estimates. For dichotomous items, Q_1 was computed as

$$Q_{1i} = \sum_{j=1}^j \frac{N_{ij}(O_{ij} - E_{ij})^2}{E_{ij}(1 - E_{ij})}$$

where N_{ij} was the number of examinees in interval (or group) j for item i , O_{ij} was the observed proportion of the examinees for the same cell, and E_{ij} was the expected proportions of the examinees for the same interval. The expected proportion was computed as

$$E_{ij} = \frac{1}{N_{ij}} \sum_{a \in j}^{N_{ij}} P_i(\hat{\theta}_a)$$

where $P_i(\hat{\theta}_a)$ was the item characteristic function for item i and examinee a . The summation is taken over examinees in interval j .

The generalization of Q_1 for items with multiple response categories is

$$Gen Q_{1i} = \sum_{j=1}^{10} \sum_{k=1}^{m_i} \frac{N_{ijk}(O_{ijk} - E_{ijk})^2}{E_{ijk}}$$

where

$$E_{ijk} = \frac{1}{N_{ij}} \sum_{a \in j}^{N_{ij}} P_{ik}(\hat{\theta}_a).$$

Both Q_1 and generalized Q_1 results were transformed to ZQ_1 and were compared to a criterion $ZQ_{1,crit}$ to determine acceptable fit. The conversion formula was

$$ZQ_1 = \frac{Q_1 - df}{\sqrt{2df}}$$

and

$$ZQ_{1,crit} = \frac{N}{1500} * 4,$$

Where df is the degrees of freedom. The degrees of freedom is equal to the number independent cells minus the number of independent item parameters. For example; the degrees of freedom for polytomous items equals $[10 \times (\text{number of score categories} - 1) - \text{number of independent item parameters}]$. For the GPCM, the number of independent item parameters equals 1 (for the a parameter)

plus the number of step values (e.g., for an item scored 0, 1, 2, 3: there are 3 independent step values—the b parameter is simply the mean of the step values and is not, therefore, independent).

If Q_1 is found to be excessively sensitive (i.e., a large number of items are flagged for poor fit, even if their item fit plots look okay), a likelihood-ratio chi-squared statistic may be computed for each item (Muraki, 1997):

$$G_i^2 = 2 \sum_{w=1}^{W_i} \sum_{h=1}^{m_i} r_{wih} \ln \frac{r_{wih}}{N_{wi} P_{ih}(\bar{\theta}_w)}$$

where r_{wih} is the observed frequency of the h^{th} categorical response to item i in interval w , N_{wi} is the number of examinees in interval w for item i , $P_{ih}(\bar{\theta}_w)$ is the expected probability of observing the h^{th} categorical response to item i for the mean θ in interval w , and W_i is the number of intervals remaining after neighboring intervals are merged, if necessary, to avoid expected values, $N_{wi} P_{ih}(\bar{\theta}_w)$, less than 5. To conduct a standard hypothesis test, the number of degrees of freedom is equal to the number of intervals, W_i , multiplied by $m_i - 1$.

As an alternative to a traditional hypothesis test, the “contingency coefficient” (effect size; Barton & Huynh, 2003) was computed:

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

In this formula, G_i^2 was substituted for χ^2 , and N is the sample size on which the IRT parameters were estimated. According to Cohen (1988, pp. 224-225) values of C below .10 are considered insignificant, .10+ small, .287+ medium, and .447+ large. A threshold of .35 is recommended (i.e., flag items for which $C \geq .35$).

An item fit-plot was created for each item. Item-fit plots show observed and expected average scores for each interval (e.g., figure below).

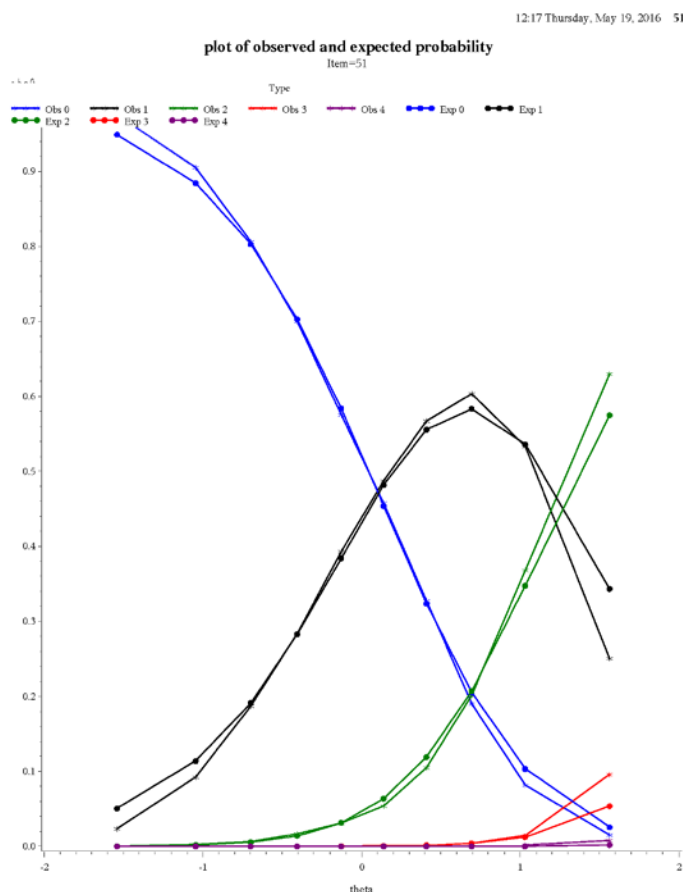


Figure 10.1 An example ELA/L 5-Category Item, 2 PL/GPC Model, N-count 44,658, $Q_1=1266.64$, $ZQ_1=147.21$ and a criterion $ZQ_{1,crit}=237.02$

10.5 Items Excluded from Score Reporting

As mentioned previously, after calibration and model fit evaluation was completed, a master list of all problematic items was compiled and potentially flawed items were brought to the PARCC Priority Alert Task Force. The Task Force reviewed each item, its content and the statistical properties, and made decisions about whether to include the item in the operational scores. Sometimes, an item was rejected because it appeared to have content issues, and sometimes an item was excluded because it could not be calibrated or showed extremely poor IRT model fit. Ultimately the decisions about whether to keep or exclude each flagged item was made by the Task Force.

10.5.1 Item Review Process

The following are the types of problematic items that were brought to the PARCC Priority Alert Task Force for evaluation and an “include or exclude” determination was made:

1. Extremely difficult items (e.g., an item with a p value less than 0.02),
2. Items with low a -parameter estimates (e.g., slope less than 0.10),

3. Items flagged for subgroup DIF or language DIF

Again, the primary goal was to minimize the number of items dropped from the operational test forms. An equally important goal was to not advantage or disadvantage any test takers.

10.5.2 Count and Percentage of Items Excluded from Score Reporting

All items that did not have IRT item parameter estimates were excluded from the student operational scores and the conversion tables used for score reporting. For ELA/L and mathematics, at most 2% of the items were excluded from score reporting for all grades/subjects. Figure 10.2 demonstrates an item that was flagged during the calibrations and item fit review for poor statistics (a-parameter=0.02 and b-parameter=33.48) and poor fit (e.g. $Q_1=4947.19$, $ZQ_1=845.52$ and a criterion $ZQ_{1,crit}=468.90$). This item was reviewed by the Priority Alert Task Force and removed from scoring.

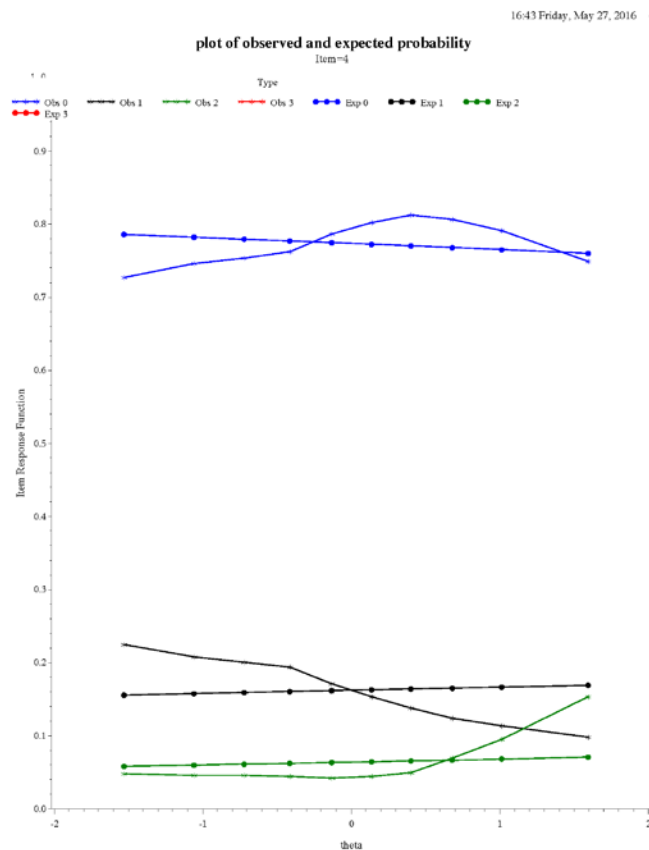


Figure 10.2 An example ELA/L 3-Category Item, 2 PL/GPC Model, N-count 175,839

Tables 10.3 and 10.4 present the count and percentage of CBT and PBT items excluded from IRT calibration along with the reasons the items were excluded for ELA/L and mathematics, respectively.

Table 10.3 Number and Percentage of ELA/L Items Excluded from IRT Calibration

Grade	Total <i>n</i> of CBT Items	<i>n</i> of CBT Items Excluded	Percent Excluded	Reason Excluded				Total <i>n</i> of PBT Items	<i>n</i> of PBT Items Excluded	Percent Excluded	Reason Excluded			
				Small Sample Size	Poor IA Stats	Did Not Calibrate	Other				Small Sample Size	Poor IA Stats.	Did Not Calibrate	Other
3	65	1	2%		1			52	0	0%				
4	85	0	0%					50	0	0%				
5	84	0	0%					44	0	0%				
6	80	0	0%					56	0	0%				
7	88	0	0%					72	0	0%				
8	84	0	0%					56	0	0%				
9	89	0	0%					66	0	0%				
10	102	0	0%					58	0	0%				
11	78	0	0%					64	0	0%				

Table 10.4 Number and Percentage of Mathematics Items Excluded from IRT Calibration

Grade/ Subject	Total n of CBT Items	n of CBT Items Excluded	Percent Excluded	Reason Excluded				Total n of PBT Items	n of PBT Items Excluded	Percent Excluded	Reason Excluded			
				Small Sample Size	Poor IA Stats	Did Not Calibrate	Other				Small Sample Size	Poor IA Stats.	Did Not Calibrate	Other
3	97	0	0%					75	0	0%				
4	101	0	0%					65	0	0%				
5	95	0	0%					69	0	0%				
6	101	0	0%					63	0	0%				
7	103	0	0%					65	0	0%				
8	89	0	0%					63	0	0%				
A1	100	0	0%					69	0	0%				
GO	101	0	0%					73	0	0%				
A2	109	0	0%					79	1	1%		1		
M1	42	0	0%					n/a						
M2	42	1	2%		1			n/a						
M3	40	0	0%					n/a						

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

10.6 Scaling Parameter Estimates

Three linking analyses to transform IRT parameters to a new IRT scale were a part of the spring 2016 psychometric process: (1) Year-to-year linking, (2) paper-to-online linking, and (3) traditional mathematics to integrated mathematics linking. The linking analyses included common-item sets. The linking methodology was based on the Stocking and Lord (1983) test characteristic curve scale transformation method.

10.6.1 Scaling Procedures (Year-to-Year)

Year-to-year linking transforms IRT parameters from different years (or administrations) onto the same underlying IRT scale. Due to the PARCC test redesign and updates to the trait scoring for the PCRs, the spring 2016 online IRT scale was established as the base IRT scale. The PARCC item bank consisting of spring 2015, fall 2015, and fall 2014 IRT parameters were on the 2015 IRT scale. The item bank was transformed to the spring 2016 IRT scale through a common item linking set consisting of spring 2016 operational online items that were operational or field-test items on the spring 2015 online administration.

Once the CBT item response data were calibrated for all grades/subjects for each content area, all available item parameter estimates of operational common items across years, were used to transform the spring 2015 item parameter estimates onto the spring 2016 scales. The software program STUIRT (Kim & Kolen, 2004) was used to obtain Stocking and Lord (1983) transformation values to link the scales.

The 2015 IRT parameters were then transformed to the 2016 IRT scale through the transformation values using the following formulas:

$$a_{2016} = a_{2015} / slope_{2015_to_2016},$$

where a_{2015} is the item discrimination parameter from 2015 for a given item, and a_{2016} is the 2016 item discrimination on 2016 scale for the same item, and $slope_{2015_to_2016}$ is the multiplicative coefficient from linking the IRT parameters in 2015 to the IRT parameters in 2016 resulting from the application of the Stocking and Lord methodology.

$$b_{2016} = b_{2015} * slope_{2015_to_2016} + intercept_{2015_to_2016},$$

where b_{2015} is the item difficulty parameter from 2015 for a given item, and b_{2016} is the item difficulty on 2016 scale, and $intercept_{2015_to_2016}$ is the additive coefficient from linking the IRT parameters in 2015 to the IRT parameters in 2016 based on the Stocking and Lord methodology.

10.6.2 Scaling Procedures (PBT to CBT)

PARCC tests are administered in two delivery modes: (1) online and (2) paper. Even though the tests are administered in different modes, the test forms have common items. The spring 2016 online IRT scale is the base scale for each PARCC assessments. The common items allow paper IRT parameters to be

transformed to the online IRT scale. The common item set consists of spring 2016 operational and field-test online items that are operational or field-test items in the spring 2016 paper administration.

Once the CBT and PBT item response data were calibrated for all grades/subjects for each content area, all available item parameter estimates of common items across modes (CBT and PBT), were used to transform the PBT item parameter estimates onto the CBT scales. The software program STUIRT (Kim & Kolen, 2004) was used to obtain Stocking and Lord (1983) transformation values to link the PBT scales.

10.6.3 Scaling Procedures (Traditional-to-Integrated Mathematics)

A special linking design between traditional mathematics and integrated mathematics was implemented using items that measure the same content standards in traditional and integrated mathematics. This linking design was necessary because the small sample sizes on the integrated mathematics assessments could not support extensive field testing in integrated mathematics. Establishing a link between traditional and integrated mathematics enriched the integrated mathematics item banks for test construction purposes and pre-equating form construction.

For linking purpose, some operational integrated mathematics items were also administered as field test items in the traditional mathematics forms. These items were used as common items to transform the traditional mathematics items onto the integrated mathematics IRT scales. This linking design was completed after the year-to-year linking and the paper-to-online linking are completed.

Due to the integrated mathematics assessments having content standards in common with multiple traditional mathematics assessments, there are multiple linking transforms planned. Where common items were available, each traditional mathematics assessment was scaled to each integrated mathematics scale. The software program STUIRT (Kim & Kolen, 2004) was used to obtain Stocking and Lord (1983) transformation values.

10.6.4 Comparability across Spanish and English Versions

All items on one CBT and one PBT form of the mathematics test at each grade/subject was translated from English into Spanish. However, data from the Spanish forms were not included in the calibration with the English data. The item parameter estimates based on data from the English forms was used to generate conversion tables for the Spanish forms. To check that the Spanish and English items were performing similarly across language versions, when sample size was large enough, Mantel-Haenszel (MH) and the Standardized Mean Difference (SMD) DIF procedures were run and items showing C-DIF were dropped from the Spanish forms. The DIF analyses¹⁰ required at least 100 students for the smaller group (either reference or focal group) and 400 student for the combined group (reference and focal groups). If either of these sample size requirements were not met, then the DIF analyses were not performed.

An item may be dropped from the Spanish form if: a) it appears the item was poorly translated, and b) if it provided either an advantage or a disadvantage to those students taking the Spanish forms. Spanish

¹⁰ Refer to Section 7 for more information on DIF analysis.

items flagged for C-DIF were reviewed by Pearson content specialists to decide if the translations were an issue. These items were also reviewed by the Priority Alert Task Force for determination of whether to exclude these items from score reporting. The number and content representation of the flagged items (to be dropped) was monitored closely to avoid dropping large number of items and points for a single form, and to avoid dropping too many items from a single subclaim. Items excluded from score reporting due to Spanish DIF were: one item each in mathematics grades 5, 6, 8, and Integrated Mathematics II; two items in Algebra II; three items in Integrated Mathematics I; and six items in Integrated Mathematics III.

10.6.5 Scaling Quality Control

HumRRO not only conducted independent calibrations of item response data using IRTPRO scaling software, they also used STUIRT (Kim & Kolen, 2004) software to transform their IRTPRO item parameter estimates onto the IRTPRO scales for each grade/subject. HumRRO's scaling constants were compared to those generated by Pearson and found exactly match. As described in Section 10.3.4, Measured Progress independently made certain that the same items were excluded from the linking sets, and compared transformed parameter estimates by Pearson and HumRRO.

10.7 Items Excluded from Linking Sets

Linking is an iterative process. Robust Z (Huynh & Meyer, 2010) and Weighted Root Mean Square Difference (WRMSD) were used to identify outlier items in the linking sets. Furthermore, items used in the paper-to-online linking were checked for differential item functioning (DIF). Mantel-Haenszel D-DIF procedures were used for dichotomous items and standardized mean difference (SMD) were calculated for polytomous items. The following rules were used to identify items for possible exclusion from the linking sets:

- 1) Exclude an item from the common item set if different amounts of collapsing resulted in a different number of response categories across modes or versions.
- 2) Flag and potentially exclude an item from the common item set if the online or paper weighted polyserial correlation, based on the item analysis, was less than 0.10.
- 3) Exclude items dropped by the PARCC Priority Alert Task Force (i.e., due to content or parameter estimation issues).
- 4) Flag and potentially exclude an item from the common item set if the mode DIF results indicated positive or negative C-DIF.

After removing items due to item performance issues as described above, the following steps were performed:

- 5) Implement the Robust Z approach to see if any common items are flagged
- 6) Run the initial Stocking and Lord procedure using the STUIRT software
- 7) Calculate WRMSD and check to see if any common items exceed the threshold
- 8) Re-run STUIRT after removing the items flagged by Robust Z
- 9) Compare the slopes and intercepts from steps 2 and 4

SAS code was developed to calculate WRMSD, Robust Z, compare the item characteristic curves (ICCs) across modes and to identify items for possible removal from the linking sets. Table 10.5 lists the flag criteria for the weighted root mean square difference (WRMSD). (*Note: these values were originally developed as part of the 2014 PARCC field test analysis.*)

Table 10.5 WRMSD Flagging Criteria for Inspection and Possible Removal of Linking Items

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

When inspecting items flagged for exclusion from the linking sets, content representation was also considered to avoid removing large numbers of items from the same subclaim. Tables 10.6 and 10.7 present the numbers of items excluded from the year-to-year linking sets for each grade/subject by content area. Tables 10.6 and 10.7 present: the total number of common items, items excluded from the year-to-year linking sets, and items kept in the linking sets for each grade/subject by content area. For ELA/L the final number of linking items ranged from 36 (in grade 3) to 80 (in grade 10). For mathematics, the final number of linking items ranged from 18 (in integrated mathematics 3) to 94 (in grade 7). For ELA/L, grade 6 had the largest number of items removed from the linking sets due to Robust Z for the a-parameter and b-parameter. For mathematics, grade 6 had the largest number of items removed from the linking sets due to Robust Z for the b-parameter.

Table 10.6 Number of ELA/L Items Excluded from the Year-to-Year Linking Sets

Grade	Total <i>n</i> of Common Items	Number Excluded	Final Number in Linking Set	Number of Excluded Items by Reason for Exclusion			
				Low Polyserial	Robust Z IRT_a	Robust Z IRT_b	High WRMSD
3	43	7	36		3	4	
4	58	11	47		4	7	
5	70	12	58		5	7	
6	73	15	58		7	8	
7	53	9	44		6	3	
8	64	4	60		3	1	
9	78	8	70		6	2	
10	88	8	80		6	2	
11	63	4	59		2	2	

Table 10.7 Number of Mathematics Items Excluded from the Year-to-Year Linking Sets

Grade /Subj.	Total <i>n</i> of Common Items	Number Excluded	Final Number in Linking Set	Number of Excluded Items by Reason for Exclusion			
				Low Polyserial	Robust Z IRT_a	Robust Z IRT_b	High WRMSD
3	87	9	78		3	6	
4	91	5	86		4	1	
5	80	12	68		5	7	
6	86	16	70		6	10	
7	97	3	94		2	1	
8	72	5	67		4	1	
A1	91	11	80		8	3	
GO	98	5	93		5		
A2	83	10	73		4	6	
M1	23	4	19		2	2	
M2	24	4	20	1	2	2	
M3	21	3	18		1	2	

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Tables 10.8 and 10.9 present the numbers of items excluded from the CBT/PBT linking sets for each grade/subject by content area. Tables 10.8 and 10.9 present: the total number of common items, items excluded from the CBT/PBT linking sets, and items kept in the linking sets for each grade/subject by content area. For ELA/L the final number of linking items ranged from 11 (in grade 10) to 26 (in grade 4). For mathematics, the final number of linking items ranged from 18 (in grade 4 and Algebra I) to 30 (in grades 6 and 8). For ELA/L, grade 7 had the largest number of items removed from the linking sets due

to Robust Z for the a -parameter. For mathematics, grade 4 had the largest number of items removed from the linking sets due to Robust Z for the a -parameter.

Table 10.8 Number of ELA/L Items Excluded from the CBT/PBT Linking Sets

Grade	Total n of Common Items	Number Excluded	Final Number in Linking Set	Number of Excluded Items by Reason for Exclusion				
				Low Polyserial	Mode C-DIF	Robust Z IRT_a	Robust Z IRT_b	High WRMSD
3	21	4	17		2	1	1	
4	30	4	26		2	2		
5	14	0	14					
6	25	2	23			2		
7	30	6	24		4	4	2	
8	29	4	25			3	1	
9	26	2	24		1	1	1	
10	12	1	11			1		
11	28	4	24		1	1	2	

Table 10.9 Number of Mathematics Items Excluded from the CBT/PBT Linking Sets

Grade /Subj.	Total n of Common Items	Number Excluded	Final Number in Linking Set	Number of Excluded Items by Reason for Exclusion				
				Low Polyserial	Mode C-DIF	Robust Z IRT_a	Robust Z IRT_b	High WRMSD
3	33	5	28			1	4	
4	29	11	18		3	8	2	
5	32	5	27			3	2	
6	32	2	30			1	1	
7	27	2	25		1		1	
8	34	4	30	1	1	1	2	
A1	22	4	18			3	1	
GO	25	5	20		1	3	1	
A2	22	2	20	1		2		

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II.

10.8 Correlations and Plots of Scaling Item Parameter Estimates

Once the final group of items for each linking set was determined, and the 2015 item parameter estimates were transformed onto the 2016 scales, the a - and b -parameter estimates across modes were plotted and the correlation between the a -parameter estimates and the b -parameter estimates were calculated. Tables 10.10 and 10.11 present the number of linking items, score points of the linking items, and the correlation of the a - and b -parameter estimates across years.

Table 10.10 Number of Items, Number of Points and Correlations for ELA/L Year-to-Year Linking Items

Grade	Number		Parameter Correlations	
	Items	Points	a -	b -
3	36	76	0.99	0.99
4	47	101	0.98	0.99
5	58	121	0.98	0.99
6	58	121	0.99	0.99
7	44	92	0.97	0.97
8	60	124	0.97	0.98
9	70	144	0.98	0.98
10	80	168	0.97	0.96
11	59	126	0.97	0.98

Table 10.11 Number of Items, Number of Points and Correlations for Mathematics Year-to-Year Linking Items

Grade/ Subject	Number		Parameter Correlations	
	Items	Points	a -	b -
3	78	111	0.97	0.99
4	86	129	0.95	0.99
5	68	121	0.97	0.99
6	70	108	0.98	0.99
7	94	158	0.97	0.99
8	67	120	0.96	0.99
A1	80	163	0.97	0.98
GE	93	178	0.96	0.99
A2	73	145	0.97	0.98
M1	19	36	0.98	0.98
M2	20	43	0.94	0.98
M3	18	36	0.93	0.98

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Tables 10.12 and 10.13 presents similar information for the PBT/CBT linking items; that is, the number of PBT/CBT linking items, score points of the linking items, and the correlation of the a - and b -parameter estimates across modes.

Table 10.12 Number of Items, Number of Points and Correlations for ELA/L CBT/PBT Linking Items

Number			Parameter Correlations	
Grade	Items	Points	a -	b -
3	17	37	0.98	0.99
4	26	55	0.98	0.99
5	14	32	0.98	0.97
6	23	49	0.97	0.99
7	24	48	0.91	0.98
8	25	53	0.98	0.98
9	24	51	0.97	0.99
10	11	25	0.99	0.99
11	24	50	0.97	0.99

Table 10.13 Number of Items, Number of Points and Correlations for Mathematics CBT/PBT Linking Items

Number			Parameter Correlations	
Grade/ Subject	Items	Points	a -	b -
3	28	38	0.92	0.99
4	18	24	0.97	0.98
5	27	50	0.94	1.00
6	30	44	0.96	0.99
7	25	32	0.9	0.99
8	30	58	0.96	0.99
A1	18	40	0.98	0.99
GE	20	56	0.93	0.99
A2	20	60	0.93	0.96

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Figures 10.3 to 10.10 are a selection of plots of the a - and b -parameter estimates for linking items for the year-to-year linking and CBT/PBT linking. ELA/L grade 8 (Figures 10.3 and 10.4) and mathematics grade 5 (Figures 10.5 and 10.6) are examples of the year-to-year linking. ELA/L grade 4 (Figures 10.7 and 10.8) and mathematics grade 8 (Figures 10.9 and 10.10) are examples of the CBT/PBT linking. For each plot, the x-axis is the original (reference) parameter and the y-axis is the new parameter after applying the scaling constants.

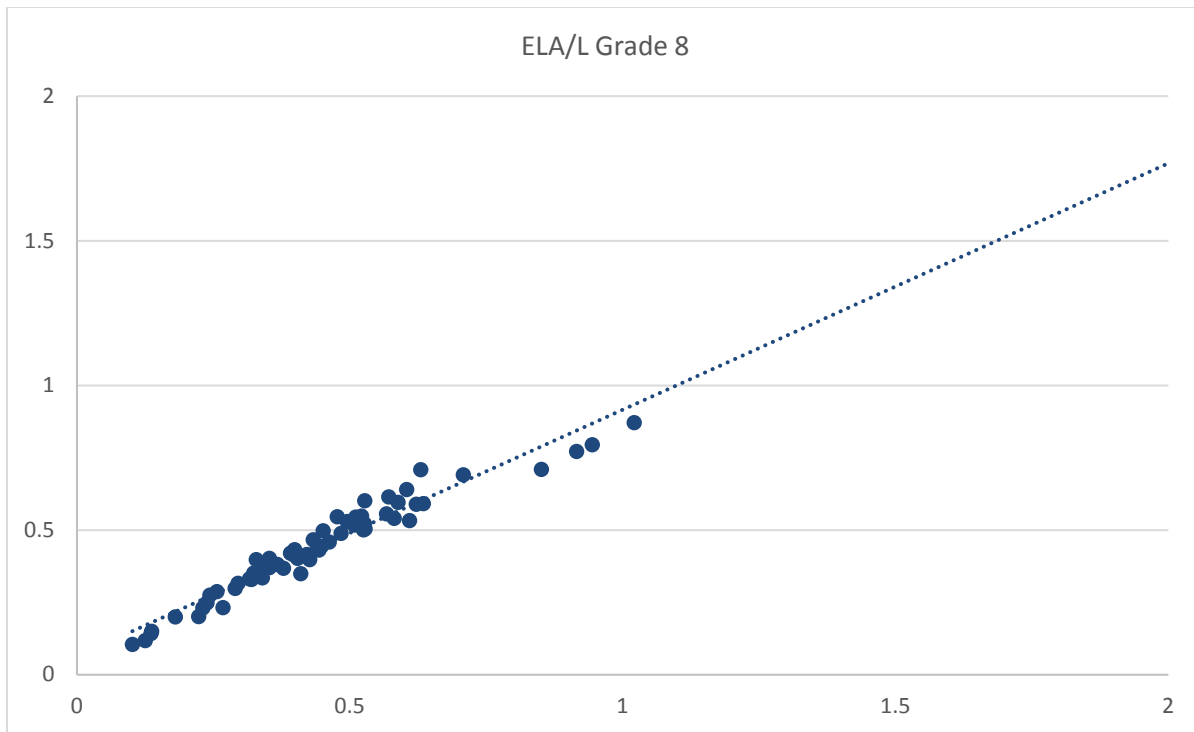


Figure 10.3 ELA/L Grade 8 Transformed New a - vs. Reference a -Parameter Estimates for Year-to-Year Linking

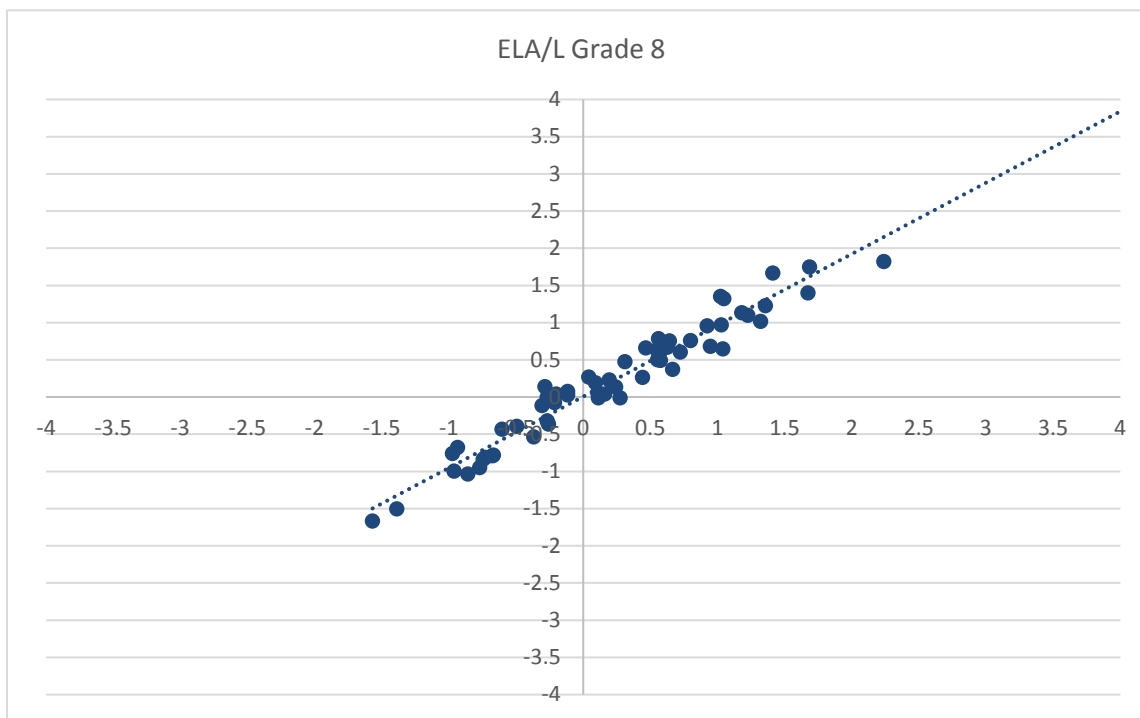


Figure 10.4 ELA/L Grade 8 Transformed New b - vs. Reference b -Parameter Estimates for Year-to-Year Linking

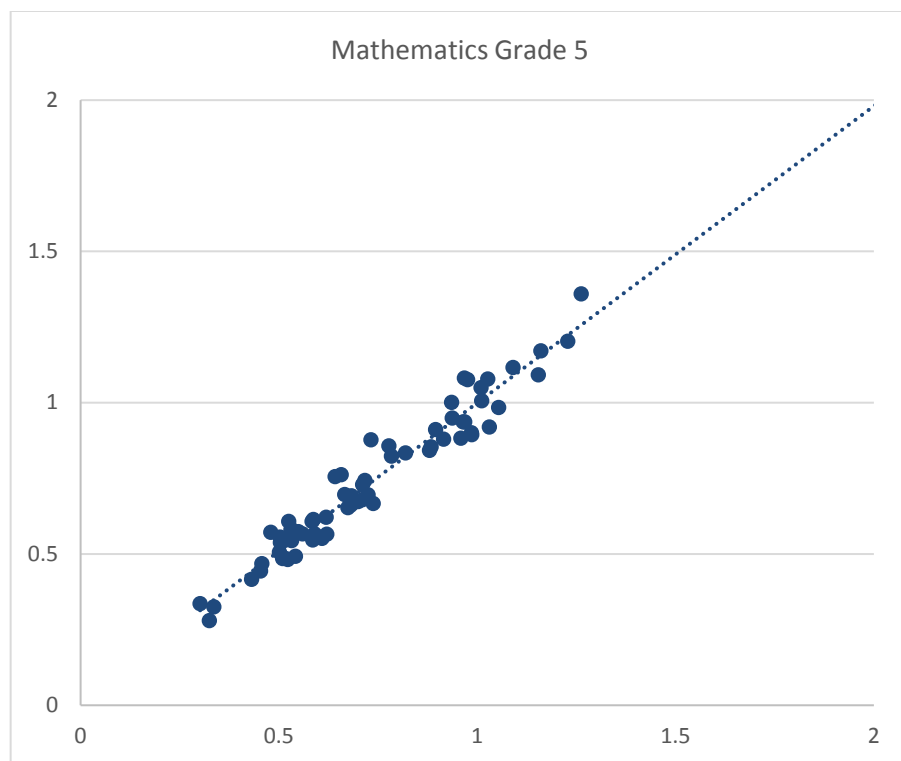


Figure 10.5 Mathematics Grade 5 Transformed New a - vs. Reference a -Parameter Estimates for Year-to-Year Linking

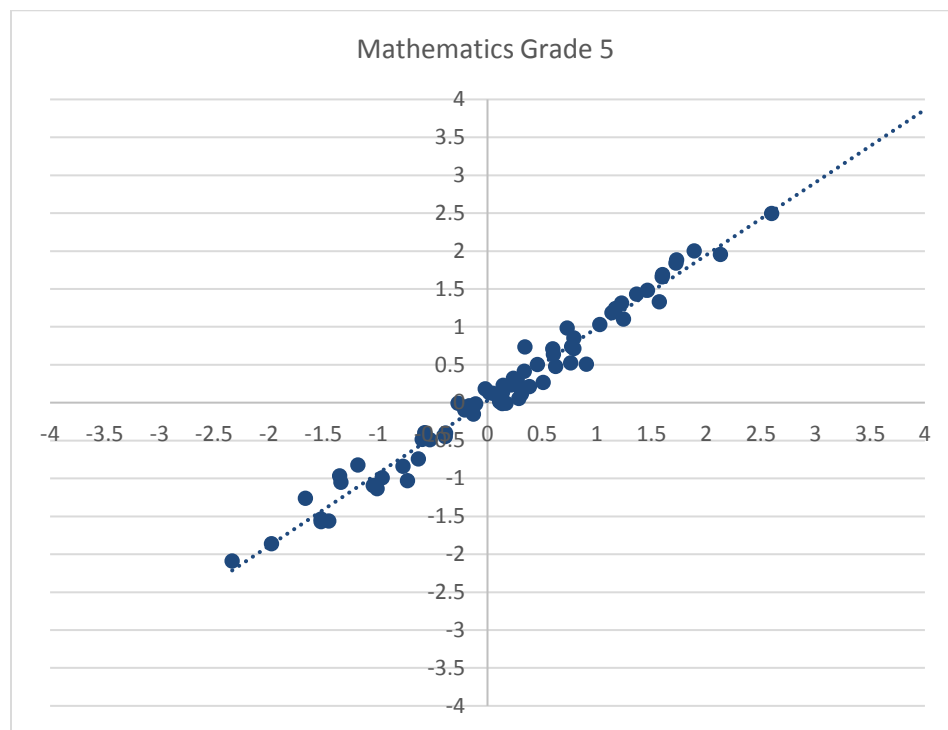


Figure 10.6 Mathematics Grade 5 Transformed New b - vs. Reference b -Parameter Estimates for Year-to-Year Linking

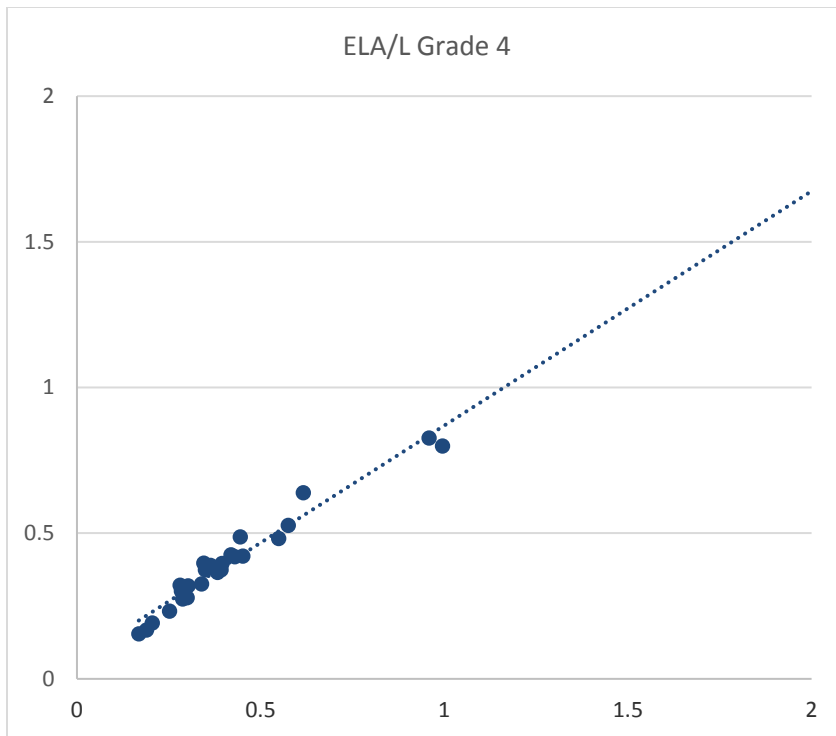


Figure 10.7 ELA/L Grade 4 Transformed New a - vs. Reference a -Parameter Estimates for PBT to CBT Linking

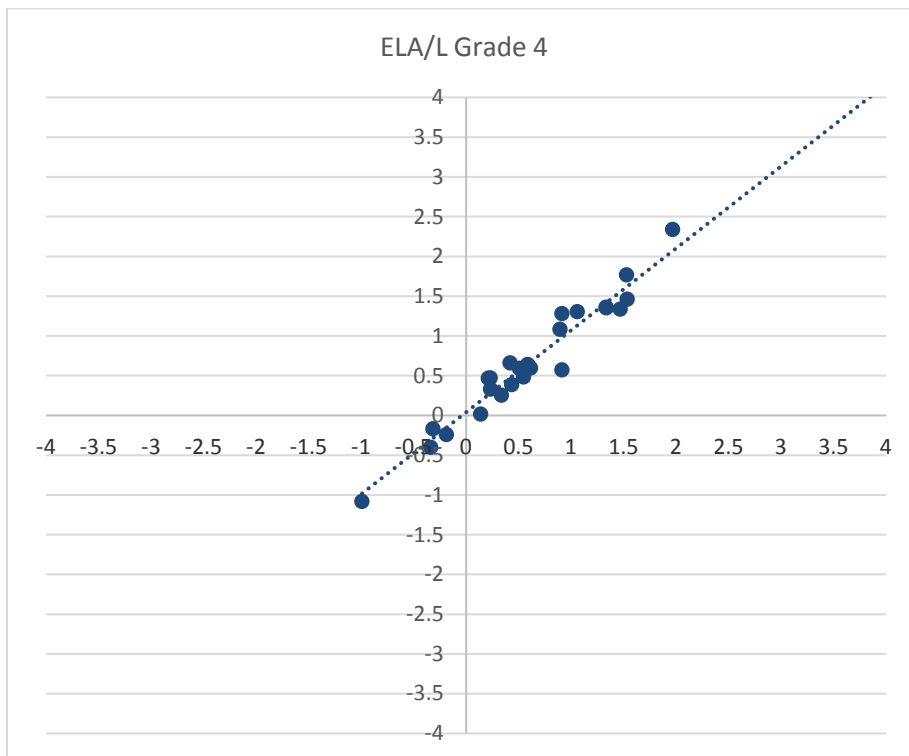


Figure 10.8 ELA/L Grade 4 Transformed New b - vs. Reference b -Parameter Estimates for PBT to CBT Linking

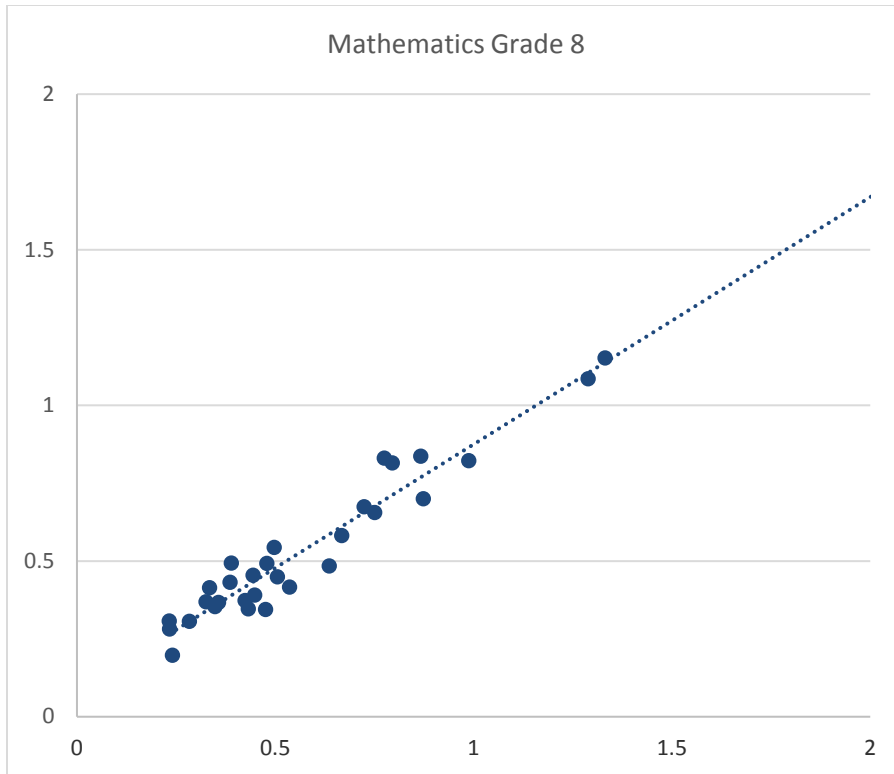


Figure 10.9 Mathematics Grade 8 Transformed New α - vs. Reference α -Parameter Estimates for PBT to CBT Linking

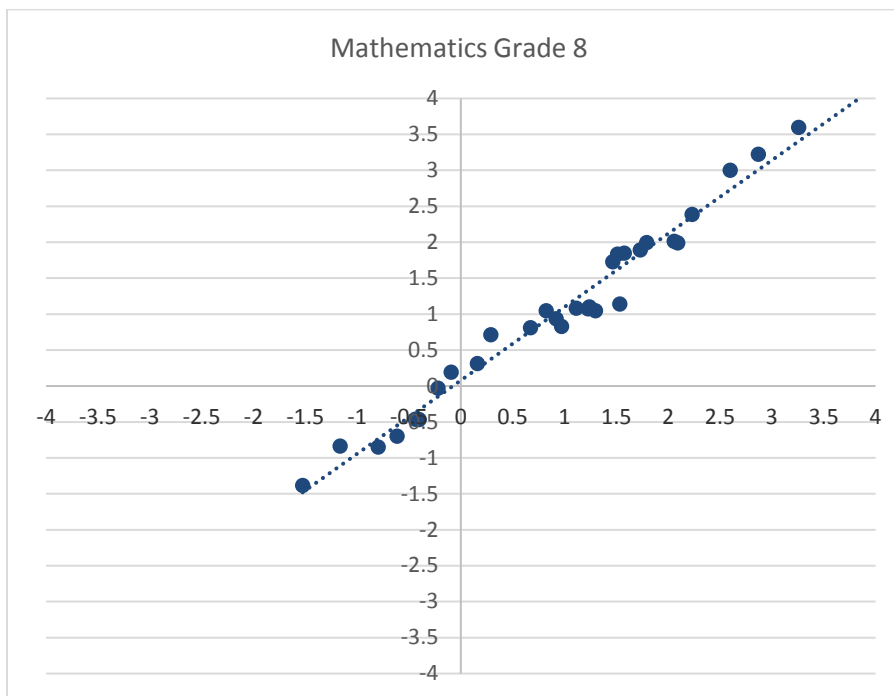


Figure 10.10 Mathematics Grade 8 Transformed New b - vs. Reference b -Parameter Estimates for PBT to CBT Linking

10.9 Scaling Constants

Tables 10.14 - 10.17 present the slope and intercept scaling constants for ELA/L and mathematics for the year-to-year linking and the PBT to CBT linking, respectively, derived from STUIRT (Kim & Kolen, 2004) using the Stocking and Lord (1983) test characteristic curve procedure.

Table 10.14 shows that, for ELA/L, the intercept is fairly consistent, ranging between -0.12 and 0.04 for grades 3 through 11. Table 10.15 shows that, for mathematics, the intercept is fairly consistent, ranging between -0.28 and 0.01 for grades 3 through 8 and high school.

Table 10.14 Scaling Constants Spring 2015 to Spring 2016 for ELA/L

Grade/Subject	Spring 2015 to Spring 2016	
	Slope	Intercept
3	0.9864	-0.0012
4	0.9759	-0.0047
5	1.0004	0.0363
6	1.0145	0.0191
7	0.9991	-0.0747
8	1.0062	-0.1229
9	1.0011	-0.0844
10	0.9844	-0.0685
11	1.0489	0.0404

Table 10.15 Scaling Constants Spring 2015 to Spring 2016 for Mathematics

Grade/ Subject	Spring 2015 to Spring 2016	
	Slope	Intercept
3	0.9918	-0.1840
4	0.9880	-0.1230
5	0.9983	-0.1863
6	1.0182	-0.0899
7	1.0229	-0.0673
8	1.0185	-0.0278
A1	0.9618	-0.1158
GO	0.9701	-0.0907
A2	0.9431	-0.0682
M1	1.0393	-0.1600
M2	1.0009	0.0131
M3	1.0582	-0.2764

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table 10.16 shows that, for ELA/L, the intercept is fairly consistent, ranging between 0.004 and 0.35 for grades 3 through 11. Table 10.17 shows that, for mathematics, the intercept ranged between -0.32 and 0.38 for grades 3 through 8 and high school.

Table 10.16 Scaling Constants Spring 2016 PBT to CBT for ELA/L

Grade/Subject	Spring 2016 PBT to CBT	
	Slope	Intercept
3	1.0329	0.1125
4	1.0062	0.0042
5	0.9547	0.3535
6	1.0476	0.2647
7	1.0498	0.0700
8	1.0186	0.2071
9	1.0615	0.2635
10	1.0255	0.2155
11	1.0965	0.2197

Table 10.17 Scaling Constants Spring 2016 PBT to CBT for Mathematics

Grade/ Subject	Spring 2015 PBT to CBT	
	Slope	Intercept
3	1.0604	-0.1093
4	1.0330	-0.3249
5	1.0554	0.0635
6	1.0655	0.0455
7	1.1436	0.1332
8	1.2392	0.3817
A1	1.0979	0.1167
GO	1.0260	0.1064
A2	0.9786	0.1854
M1	n/a	n/a
M2	n/a	n/a
M3	n/a	n/a

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

10.10 Summary Statistics and Distributions from IRT Analyses

Tables 10.18 through 10.29 present summary statistics for the IRT (b - and a -) parameter estimates and the standard errors (SEs) of the parameter estimates, as well as the IRT model fit values (chi-square and adjusted fit). The information is provided by content area (ELA/L and mathematics) and by mode (CBT and PBT) for all items at each grade level or course. The summary statistics shown include the total

number of items and score points, along with the mean, standard deviation (*SD*) and minimum and maximum.

10.10.1 IRT Summary Statistics for English Language Arts/Literacy

Tables 10.18 and 10.23 show the *b*- and *a*-parameter estimates for the ELA/L assessments. All item responses were estimated using the 2PL/GPC model combination. Tables 10.20 and 10.21 present the standard errors of estimate for CBT and PBT ELA/L assessments, and Tables 10.22 and 10.23 provide model fit information. IRT summary statistics are provided in the **Appendix 10** for ELA/L for all items, reading-only, and writing-only.

Table 10.18 CBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade

Grade	No. of Items	No. of Score Points	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
			Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
3	64	140	0.49	1.07	-1.42	3.25	0.55	0.19	0.16	1.07
4	85	189	0.46	1.09	-3.55	4.96	0.48	0.22	0.17	0.99
5	84	186	0.83	1.38	-1.31	7.91	0.48	0.25	0.05	1.12
6	80	178	0.55	0.87	-1.43	3.8	0.45	0.21	0.1	0.98
7	88	194	0.31	0.87	-1.81	3.18	0.45	0.21	0.17	1.02
8	84	186	0.38	0.95	-1.57	5.15	0.47	0.23	0.1	1.03
9	89	196	0.74	1.09	-0.78	6.49	0.48	0.25	0.08	1.16
10	102	225	0.70	0.83	-1.15	3.39	0.49	0.19	0.14	0.95
11	78	174	1.00	1.37	-1.15	8.68	0.47	0.22	0.09	0.96

Table 10.19 PBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade

Grade	No. of Items	No. of Score Points	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
			Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
3	52	114	0.66	0.88	-1.1	3.25	0.52	0.18	0.12	0.82
4	50	111	0.73	1.06	-1.13	5.09	0.41	0.17	0.15	0.82
5	44	98	0.91	0.83	-0.97	2.93	0.46	0.21	0.13	0.94
6	56	121	0.50	0.74	-0.82	2.14	0.43	0.16	0.16	0.78
7	72	159	0.28	0.83	-1.99	2.92	0.40	0.15	0.11	0.82
8	56	124	0.37	0.84	-1.17	2.97	0.43	0.18	0.09	0.88
9	66	144	0.47	0.83	-1.12	2.93	0.43	0.16	0.18	0.79
10	58	128	0.70	0.76	-0.53	2.59	0.50	0.19	0.18	0.9
11	64	140	1.06	0.99	-0.97	3.75	0.41	0.21	0.1	0.89

Table 10.20 CBT IRT Standard Errors of Parameter Estimates for All Items for ELA/L by Grade

Grade	No. of Items	No. of Score Points	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
			Mean	SD	Min	Max	Mean	SD	Min	Max
3	64	140	0.01	0.01	0.003	0.057	0.004	0.002	0.002	0.009
4	85	189	0.013	0.016	0.004	0.107	0.005	0.002	0.002	0.011
5	84	186	0.018	0.041	0.004	0.316	0.004	0.002	0.002	0.011
6	80	178	0.01	0.01	0.004	0.071	0.004	0.002	0.002	0.01
7	88	194	0.008	0.005	0.003	0.043	0.004	0.002	0.002	0.01
8	84	186	0.009	0.01	0.003	0.093	0.003	0.002	0.002	0.009
9	89	196	0.014	0.027	0.004	0.238	0.004	0.002	0.002	0.011
10	102	225	0.015	0.012	0.006	0.079	0.006	0.002	0.003	0.013
11	78	174	0.022	0.037	0.005	0.297	0.006	0.003	0.003	0.015

Table 10.21 PBT IRT Standard Errors of Parameter Estimates for All Items for ELA/L by Grade

Grade	No. of Items	No. of Score Points	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
			Mean	SD	Min	Max	Mean	SD	Min	Max
3	52	114	0.016	0.012	0.007	0.085	0.007	0.003	0.003	0.014
4	50	111	0.022	0.027	0.007	0.187	0.007	0.003	0.003	0.015
5	44	98	0.024	0.016	0.009	0.076	0.009	0.004	0.004	0.02
6	56	121	0.021	0.012	0.01	0.083	0.009	0.003	0.005	0.015
7	72	159	0.024	0.018	0.011	0.154	0.009	0.003	0.004	0.02
8	56	124	0.022	0.015	0.01	0.09	0.009	0.003	0.004	0.02
9	66	144	0.032	0.018	0.015	0.122	0.013	0.005	0.007	0.029
10	58	128	0.043	0.032	0.019	0.174	0.02	0.007	0.009	0.04
11	64	140	0.074	0.057	0.026	0.258	0.021	0.008	0.01	0.044

Table 10.22 CBT IRT Model Fit for All Items for ELA/L by Grade

Grade	No. of Items	No. of Score Points	G^2				Q_1			
			Mean	SD	Min	Max	Mean	SD	Min	Max
3	64	140	3136.9	2980	450.4	17317	3004.6	2881.8	418	17000.5
4	85	189	2181.4	1726.3	155.4	9059.9	2121.4	1741.4	148.9	9757.6
5	84	186	2521.2	2016.4	282.2	10665.5	2432.3	1982.7	278.3	10922.5
6	80	178	2898.3	2583.5	324.9	15422.4	2813.2	2623.8	320.2	16395.8
7	88	194	3447.8	3089.5	333.1	17360.3	3432.7	3351	300.3	21851.3
8	84	186	2919.8	3105.5	253.1	23884.3	2873.3	3175	248.4	24231.3
9	89	196	2629.2	2164.4	301.4	9393.9	2564	2245.2	289.5	10501.1
10	102	225	1382	987.7	172.2	5524	1309.8	971	168.4	5623.9
11	78	174	1367.9	1250.8	147.5	5899	1350.7	1291	143.6	6363.8

Table 10.23 PBT IRT Model Fit for All Items for ELA/L by Grade

Grade	No. of Items	No. of Score Points	G^2				Q_1			
			Mean	SD	Min	Max	Mean	SD	Min	Max
3	52	114	868.8	661.6	145.2	3861	813.8	604.8	133.3	3544.3
4	50	111	915.2	583.6	54.6	2234.4	872.1	564.8	56	2600.5
5	44	98	587.1	471.9	79.1	2011.4	582.3	486.6	73.8	2373.9
6	56	121	497.7	347.2	69.9	1732.8	482.7	346.1	66.8	1695
7	72	159	461.8	269	75.7	1484.6	457.6	278.1	72.9	1433.7
8	56	124	481.9	417.7	71.6	2489	505.8	555.4	68.6	3741.3
9	66	144	230.2	158.4	42.7	927.9	225.1	159.6	39.5	877.2
10	58	128	146.9	103.7	31.6	550.8	142.8	110.1	26.4	530
11	64	140	92.2	65	26.5	376.2	89.1	66.3	24.3	373.3

10.10.2 IRT Summary Statistics for Mathematics

Tables 10.24 and 10.29 show the b - and a -parameter estimates for the mathematics assessments. All item responses were estimated using the 2PL/GPC model combination. Due to small sample sizes, the Integrated Mathematics assessments were not post-equated; therefore, results are not presented in the PBT tables. Tables 10.26 and 10.27 present the standard errors of estimate for CBT and PBT mathematics assessments, and Tables 10.28 and 10.29 provide model fit information. IRT summary statistics are provided in the **Appendix 10** for mathematics for all items, single select multiple choice items, constructed response items, and subclaims.

Table 10.24 CBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject

Grade/ Subject	No. of Items	No. of Score Points	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
			Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
3	97	144	-0.23	1.34	-3.2	3.58	0.76	0.25	0.24	1.36
4	101	155	-0.23	1.29	-2.93	2.34	0.72	0.19	0.34	1.32
5	95	157	0.11	1.16	-3.41	2.60	0.67	0.24	0.24	1.26
6	101	169	0.25	1.27	-4.46	3.74	0.74	0.25	0.23	1.42
7	103	173	0.82	1.21	-2.23	3.24	0.76	0.32	0.23	1.84
8	89	162	0.87	1.24	-1.53	3.48	0.63	0.23	0.23	1.29
A1	100	192	1.19	1.08	-1.16	3.9	0.63	0.28	0.1	1.51
GO	109	203	0.87	1.13	-1.98	2.98	0.75	0.31	0.22	1.73
A2	101	198	1.35	0.81	-1.41	2.80	0.69	0.30	0.12	1.44
M1	42	81	1.20	1.06	-0.95	3.41	0.62	0.33	0.17	1.61
M2	41	80	1.90	1.45	-0.74	5.99	0.53	0.27	0.12	1.18
M3	40	81	1.27	1.08	-2.27	4.28	0.59	0.27	0.16	1.27

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table 10.25 PBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject

Grade/ Subject	No. of Items	No. of Score Points	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
			Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
3	75	111	-0.34	1.17	-3.22	2.25	0.72	0.24	0.28	1.32
4	65	109	-0.36	1.12	-2.69	1.53	0.69	0.22	0.3	1.24
5	69	115	0.06	1.15	-2.45	2.28	0.67	0.2	0.31	1.2
6	63	110	0.20	1.17	-3.54	2.47	0.67	0.22	0.23	1.4
7	65	102	0.86	1.31	-2.03	3.14	0.62	0.22	0.26	1.39
8	63	114	1.09	1.27	-1.37	3.54	0.54	0.22	0.18	1.07
A1	69	128	1.17	1.15	-0.99	3.36	0.55	0.23	0.13	1.3
GO	79	139	1.23	1.28	-1.51	5.98	0.67	0.3	0.18	1.41
A2	72	143	1.65	1.59	-1.14	9.06	0.60	0.27	0.12	1.41

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Table 10.26 CBT IRT Standard Errors of Parameter Estimates for All Items for Mathematics by Grade/Subject

Grade/ Subject	No. of Items	No. of Score Points	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
			Mean	SD	Min	Max	Mean	SD	Min	Max
3	97	144	0.018	0.015	0.005	0.086	0.011	0.004	0.004	0.023
4	101	155	0.011	0.008	0.003	0.038	0.007	0.002	0.002	0.013
5	95	157	0.012	0.01	0.003	0.061	0.007	0.002	0.002	0.013
6	101	169	0.011	0.012	0.003	0.104	0.007	0.002	0.003	0.014
7	103	173	0.015	0.01	0.004	0.051	0.009	0.005	0.003	0.041
8	89	162	0.015	0.01	0.005	0.045	0.007	0.003	0.003	0.02
A1	100	192	0.017	0.021	0.003	0.191	0.007	0.003	0.002	0.028
GO	109	203	0.016	0.011	0.005	0.066	0.01	0.004	0.004	0.022
A2	101	198	0.02	0.018	0.005	0.161	0.01	0.004	0.004	0.021
M1	42	81	0.051	0.038	0.021	0.199	0.017	0.007	0.008	0.04
M2	41	80	0.153	0.18	0.028	0.977	0.033	0.021	0.018	0.123
M3	40	81	0.108	0.107	0.038	0.631	0.041	0.018	0.015	0.097

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table 10.27 PBT IRT Standard Errors of Parameter Estimates for All Items for Mathematics by Grade/Subject

Grade/ Subject	No. of Items	No. of Score Points	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
			Mean	SD	Min	Max	Mean	SD	Min	Max
3	75	111	0.018	0.014	0.005	0.096	0.011	0.003	0.004	0.022
4	65	109	0.018	0.009	0.008	0.044	0.011	0.004	0.004	0.023
5	69	115	0.02	0.011	0.007	0.07	0.012	0.003	0.005	0.02
6	63	110	0.022	0.024	0.008	0.156	0.012	0.004	0.006	0.02
7	65	102	0.032	0.019	0.009	0.1	0.015	0.006	0.006	0.036
8	63	114	0.029	0.015	0.01	0.074	0.012	0.004	0.005	0.028
A1	69	128	0.047	0.044	0.015	0.293	0.017	0.01	0.007	0.077
GO	79	139	0.1	0.153	0.03	1.088	0.039	0.015	0.014	0.091
A2	72	143	0.112	0.215	0.021	1.482	0.029	0.013	0.011	0.096

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Table 10.28 CBT IRT Model Fit for All Items for Mathematics by Grade/Subject

Grade/ Subject	No. of Items	No. of Score Points	G^2				Q_1			
			Mean	SD	Min	Max	Mean	SD	Min	Max
3	97	144	292.4	396.5	23.3	2232.4	283.4	387.1	23.8	2184.7
4	101	155	762.6	1235.7	63.9	9556.2	741	1188.3	61.5	8903.2
5	95	157	743	742.7	67.4	5094.5	708.8	729.9	63.9	4918.4
6	101	169	925.5	1233.4	90.5	5676.1	906.3	1233.7	87	5694
7	103	173	727	934.8	29.1	5174.9	688.3	893	29.2	4958
8	89	162	739.4	886.9	60.1	6471.5	721.6	926.6	58.8	7013.4
A1	100	192	932.8	1111	75.4	8440.3	880.5	1008.2	73.5	7402.2
GO	109	203	537.5	509.7	15.4	2455.1	534.1	554.5	15.3	3491.6
A2	101	198	641.4	738.2	34	3247.4	592.6	693.4	32.8	3211.1
M1	42	81	165.2	145.5	18.6	780	288.4	852.7	18.1	5611.7
M2	41	80	57.3	55.9	9.2	257.6	54.5	56.2	9.1	234.1
M3	40	81	37.2	28.6	7.6	125.8	31.9	23.1	7	102.3

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table 10.29 PBT IRT Model Fit for All Items for Mathematics by Grade/Subject

Grade/ Subject	No. of Items	No. of Score Points	G^2				Q_1			
			Mean	SD	Min	Max	Mean	SD	Min	Max
3	75	111	257.8	265.3	33.1	1417.6	248.6	256.9	32.7	1368
4	65	109	272.4	302.9	28	1383.4	266.9	303.9	26.4	1356.7
5	69	115	211.3	173.7	19.5	742.4	199.3	160.7	19	680.1
6	63	110	268.8	279.8	19.2	1646.4	256	277.3	18.8	1648.5
7	65	102	188.6	250	15.6	1441.4	173.3	223.1	15.4	1282.4
8	63	114	242.9	281.6	15	1486.4	226.3	261	10.7	1309.9
A1	69	128	126.5	152.1	10.5	896.8	123.8	147.9	10.4	829.3
GO	79	139	45.9	44.2	4.6	221.6	41.7	39.7	4.3	197.7
A2	72	143	70.9	66.7	6.9	322.2	71.8	99.9	6.3	746.5

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Section 11: Performance Level Setting

11.1 Performance Standards

Performance standards relate levels of performance on an assessment directly to what students are expected to learn. This is done by establishing threshold scores that distinguish between performance levels. Performance level setting (PLS) is the process of establishing these threshold scores that define the performance levels for an assessment.

11.2 Performance Levels and Policy Definitions

For the PARCC assessments, the performance levels are

- Level 5: Exceeded expectations
- Level 4: Met expectations
- Level 3: Approached expectations
- Level 2: Partially met expectations
- Level 1: Did not yet meet expectations

More detailed descriptions of each performance level, known as policy definitions, are:

Level 5: Exceeded expectations

Students performing at this level **exceed academic expectations** for the knowledge, skills, and practices contained in the standards assessed at their grade level or course.

Grades 3-10: Students performing at this level **exceed academic expectations** for the knowledge, skills, and practices contained in the standards for English language arts/literacy or mathematics assessed at their grade level. They are **academically well prepared** to engage successfully in further studies in this content area.

Algebra II, Integrated Mathematics III, and ELA/L 11: Students performing at this level **exceed academic expectations** for the knowledge, skills, and practices contained in the Mathematics and ELA/L standards assessed at grade 11. They are very likely to engage successfully in entry-level, credit-bearing courses in Mathematics and ELA/L, as well as technical courses requiring an equivalent command of the content area. Students performing at this level are exempt from having to take and pass placement tests in two- and four-year public institutions of higher education designed to determine whether they are academically prepared for such courses without need for remediation.

Level 4: Met expectations

Students performing at this level **meet academic expectations** for the knowledge, skills, and practices contained in the standards assessed at their grade level or course.

Grades 3-10: Students performing at this level **meet academic expectations** for the knowledge, skills, and practices contained in the standards for English language arts/literacy or Mathematics assessed at

their grade level. They are **academically prepared** to engage successfully in further studies in this content area.

Algebra II, Integrated Mathematics III, and ELA/L 11: Students performing at this level **meet academic expectations** for the knowledge, skills and practices contained in Mathematics and ELA/L at grade 11. They are very likely to engage successfully in entry-level, credit bearing courses in mathematics and ELA/L, as well as technical courses requiring an equivalent command of the content area. Students performing at this level are exempt from having to take and pass placement tests in two- and four-year public institutions of higher education designed to determine whether they are academically prepared for such courses without need for remediation.

Level 3: Approached expectations

Students performing at this level **approach academic expectations** for the knowledge, skills, and practices contained in the standards assessed at their grade level or course

Grades 3-10: Students performing at this level **approach academic expectations** for the knowledge, skills, and practices contained in the standards for English language arts/literacy or Mathematics assessed at their grade level. They are likely prepared to engage successfully in further studies in this content area.

Algebra II, Integrated Mathematics III, and ELA/L 11: Students performing at this level **approach academic expectations** for the knowledge, skills, and practices contained in the ELA/L and mathematics standards assessed at grade 11. They are likely to engage successfully in entry-level, credit-bearing courses in mathematics and ELA/L, as well as technical courses requiring an equivalent command of the content area. **Students performing at Level 3 are strongly encouraged to continue to take challenging high school coursework in English and mathematics through graduation.** Postsecondary institutions are encouraged to use additional information about students performing at Level 3, such as course completion, course grades and scores on other assessments to determine whether to place them directly into entry level courses.

Level 2: Partially met expectations

Students performing at this level **partially meet academic expectations** for the knowledge, skills, and practices contained in the standards assessed at their grade level or course.

Grades 3-10: Students performing at this level **partially meet academic expectations** for the knowledge, skills, and practices contained in the standards for English language arts/literacy or Mathematics assessed at their grade level. They will likely need academic support to engage successfully in further studies in this content area.

Algebra II, Integrated Mathematics III, and ELA/L 11: Students performing at this level **partially meet academic expectations** for the knowledge, skills, and practices contained in the ELA/L and mathematics standards assessed at grade 11. They will likely need academic support to engage successfully in entry-level, credit-bearing courses, and technical courses requiring an equivalent command of the content area. Students performing at this level are not exempt from having to take and pass placement tests

designed to determine whether they are academically prepared for such courses without the need for remediation in two- and four-year public institutions of higher education.

Level 1: Did not meet expectations

Students performing at this level **do not yet meet academic expectations** for the knowledge, skills, and practices contained in the standards assessed at their grade level or course.

Grades 3-10: Students performing at this level **do not yet meet academic expectations** for the knowledge, skills, and practices contained in the standards for English language arts/literacy or Mathematics assessed at their grade level. They will need academic support to engage successfully in further studies in this content area.

Algebra II, Integrated Mathematics III, and ELA/L 11: Students performing at this level **do not yet meet academic expectations** for the knowledge, skills, and practices contained in the ELA/L and mathematics standards assessed at grade 11. They will need academic support to engage successfully in entry-level, credit-bearing courses in College Algebra, Introductory College Statistics, and technical courses requiring an equivalent level of mathematics. Students performing at this level are not exempt from having to take and pass placement tests in two- and four-year public institutions of higher education designed to determine whether they are academically prepared for such courses without need for remediation.

11.3 Performance Level Setting Process for the PARCC Assessment System

One of the main objectives of the PARCC assessment system is to provide information to students, parents, educators, and administrators as to whether students are on track in their learning for success after high school, defined as college- and career-readiness. To set performance levels associated with this objective, PARCC used the Evidence-Based Standard Setting (EBSS) method (Beimers, Way, McClarty, & Miles, 2012) for the PARCC Performance Level Setting (PLS) process. The EBSS method is a systematic method for combining various considerations into the process for setting performance levels, including policy considerations, content standards, educator judgment about what student should know and be able to demonstrate, and research to support PARCC's policy goals related to college- and career- readiness. A defined multistep process was used to allow a diverse set of stakeholders to consider the interaction of these elements in recommending performance level threshold scores for each PARCC assessment.

The seven steps of the EBSS process that were followed in order to establish performance standards for the PARCC assessments are:

- Step 1: Define outcomes of interest and policy goals
- Step 2: Develop research, data collection, and analysis plans
- Step 3: Synthesize the research results
- Step 4: Conduct pre-policy meeting
- Step 5: Conduct performance level setting (PLS) meetings with panels
- Step 6: Conduct reasonableness review with post-policy panel
- Step 7: Continue to gather evidence in support of standards

A summary of key components within these steps is provided below. Additional detail about each step in the PARCC Performance Level Setting (PLS) process is provided in the “PARCC Performance Level Setting Technical Report”.

11.3.1 PARCC Research Studies

PARCC conducted two research studies in support of their policy goals—the Benchmarking study and the Postsecondary Educators’ Judgment (PEJ) study. The Benchmarking study included a review of the literature relative to college- and career- readiness as well as consideration of the percentage of students obtaining a level equivalent to college- and career- readiness on a set of external assessments (e.g., ACT, SAT, NAEP). The PEJ study involved a group of nearly 200 college faculty reviewing items on the Algebra II and ELA/L 11 PARCC assessments and making judgments about the level of performance needed on each item to be academically ready for an entry-level college-credit bearing course in mathematics or ELA/L. Additional detail¹¹ about the Benchmarking study can be found in the “PARCC Performance Level Setting Technical Report” as well as in the “PARCC Benchmarking Study Report.” Additional detail about the PEJ study can be found in the in the “PARCC Performance Level Setting Technical Report” as well as in the “Postsecondary Educators’ Judgment Study Final Report.”

11.3.2 PARCC Pre-Policy Meeting

Prior to the PLS meetings, a pre-policy meeting was convened to determine reasonable ranges which would be shown to panelists during the high school PLS meetings. Pre-policy meeting participants included representatives from both K-12 and higher education who served in roles such as Commissioner/Superintendent, Deputy/Assistant Commissioner, State Board Member, Director of Assessment, Director of Academic Affairs, Senior Policy Associate, and so on. The reasonable ranges recommended by the pre-policy meeting defined the minimum and maximum percentage of students that would be expected to be classified as college- and career-ready. The pre-policy meeting participants reviewed the test purpose, how the performance standards will be used, and the results of the research studies to provide the recommendations for the reasonable ranges without viewing any student performance data.

11.3.3 Performance Level Setting Meetings

The task of the performance level setting committee was to recommend four threshold scores that would define the five performance levels for each PARCC assessment. PARCC solicited nominations from all states that had administered the PARCC assessments in 2014-2015 for panelists to serve on the performance level setting committees. Nominations were solicited both from state departments of public education (K-12) and higher education (primarily for participation on the high school panels). When selecting panelists, PARCC placed an emphasis on those educators who had content knowledge as

¹¹ More information can be requested online from <http://www.parcconline.org/assessments/test-design/research>.

well as experience with a variety of student groups and attempted to balance the panels in terms of state representation.

PARCC used an Extended Modified Angoff (Yes/No) method to collect educator judgments on the PARCC items. This method asked panelists to review each item on a reference form of the PARCC assessment and to make the following judgment:

How many points would a borderline student at each performance level likely earn if they answered the question?

This extension to the Yes/No standard setting method (Plake, Ferdous, Impara, & Buckendahl, 2005) allowed for incorporation of the multipoint PARCC items by asking educators to evaluate (Yes or No) whether a borderline student would earn the maximum number of points on an item, a lesser number of points on an item, or no points on the item. In the case of a single point or multiple-choice item, this task simplifies to the standard Yes/No method.

After receiving training on the PLS procedure, panelists participated in three rounds of judgments for each assessment. Within each round, panelists were asked to consider the items in the test form, starting with the PBA component and then the EOY component. Each panelist made a judgment for the Level 2 performance level, followed by judgments for the Level 3 performance level, the Level 4 performance level, and the Level 5 performance level, in this order. The panelists entered their item judgments for each round by completing an online item judgment survey. Educator judgments were summed across items to create an estimated total score on the reference form for each performance level threshold. Feedback data relative to panelist agreement, student performance on the items, and student performance on the test as a whole were provided in between each of the three rounds of judgment. Panelists were shown the pre-policy reasonable ranges prior to making their Round 1 judgments and again as feedback data following each round of judgment.

A dry-run of the PARCC PLS meeting process was held for Grade 11 English language arts/literacy (ELA/L) and Algebra II in order to evaluate the implementation of the performance level setting method with the innovative characteristics of the PARCC assessments. These content areas were selected because they combined all of the various aspects of the PARCC assessments, including the various types of items, scoring rules, and performance level decisions. The dry-run PLS meetings provided the opportunity to implement and evaluate multiple aspects of the operational plan for the actual PLS meeting, including pre-work, meeting materials, data analysis and feedback, and staff and panelist functions. The results of the dry-run PLS meeting were used to implement improvements in the process for the operational PLS meetings. Additional information about the methods and results of the dry-run PLS meeting is available in the full report for the “PARCC Performance Level Setting Dry-Run Meeting Report.”

The PLS meetings for the PARCC assessments were conducted during three one-week sessions. The dates of the twelve PLS committee meetings that were conducted are shown in Table 11.1.

Table 11.1 PARCC PLS Committee Meetings and Dates

Dates	Committees by Subjects and Grades
July 27 - 31, 2015	Algebra I/Integrated Mathematics I
	Geometry/Integrated Mathematics II
	Algebra II/Integrated Mathematics III
	Grade 9 English Language Arts/Literacy
	Grade 10 English Language Arts/Literacy
	Grade 11 English Language Arts/Literacy
August 17 - 21, 2015	Grades 7 & 8 Mathematics
	Grades 7 & 8 English Language Arts/Literacy
August 24 - 28, 2015	Grades 3 & 4 Mathematics
	Grades 5 & 6 Mathematics
	Grades 3 & 4 English Language Arts/Literacy
	Grades 5 & 6 English Language Arts/Literacy

Additional information about the methods and results of the PLS meetings is available in the “Performance Level Setting Technical Report.”

11.3.4 PARCC Post-Policy Reasonableness Review

Performance standards for all PARCC assessments were recommended by PLS committees and reviewed by the PARCC Governing Board and (for the Algebra II, Integrated Mathematics III, and ELA/L 11 assessments) the Advisory Committee on College Readiness as part of a post-policy reasonableness review. This group reviewed both the median threshold score recommendations from each committee and the variability in the threshold scores as represented by the Standard Error of Judgment (SEJ) of the committee. Adjustments to the median threshold scores that were within 2 SEJ were considered to be consistent with the PLS panels’ recommendation.

In addition to voting to adopt the performance standards based on the committees’ recommendations, this group also voted to conduct a shift in the PARCC performance levels to better meet the intended inferences about student performance. Holding the college- and career- ready (or on track) expectations (i.e., the current level 4) constant, performance levels above this expectation were combined and performance levels below this expectation were expanded to create the final system of performance levels with three below and two above the college- and career- ready (or on track) expectation. The shift in performance levels was accomplished using a scale anchoring process which involved two primary steps. In the first step, the top two performance levels, above college- and career-ready (or on-track), were combined into a single performance level and an additional performance level below college- and career- ready (or on track) was created by empirically determining the mid-point between the existing two levels. In the second step, the performance level descriptors (PLDs) were updated using items which discriminated student performance well at this level to create a PLD aligned with the new empirically determined performance level. At this same time, PLDs for all performance levels were reviewed for consistency and continuity. Members of the original PLS committees were recruited to participate in this process. Additional information about this process can be found in the Performance Level Setting Technical Report.

Section 12: Scale Scores

PARCC assessments are designed to measure and report results in categories called master claims and subclaims. Master claims (or simply “claims”) are at a higher level than subclaims with content representing multiple subclaims contributing to each claim outcome.

Four scale scores were reported for PARCC assessments.¹² A full summative (FS) claim score was reported for each mathematics assessment. A FS claim score, and separate claim scores for Reading and Writing were reported for each English language arts/literacy (ELA/L) assessment. PARCC reports results according to five performance levels that delineate the knowledge, skills, and practices students are able to demonstrate:

- Level 5: Exceeded expectations
- Level 4: Met expectations
- Level 3: Approached expectations
- Level 2: Partially met expectations
- Level 1: Did not yet meet expectations

Subclaim outcomes describe student performance for content-specific subsets of the item scores contributing to a particular claim. For example, Written Expression and Knowledge of Conventions subclaim outcomes are reported along with Writing claim scores. Subclaim outcomes are reported as *Below Expectations*, *Nearly Meets Expectations* or *Meets or Exceeds Expectations*.

12.1 Operational Test Content (Claims and Subclaims)

A claim is a statement about student performance based on how students respond to test questions. PARCC tests are designed to elicit evidence from students that support valid and reliable claims about the extent to which they are college and career ready or on track toward that goal and are making expected academic gains based on the Common Core State Standards (CCSS).

The number of items associated with each claim and subclaim outcome vary depending on test subject and grade. The item types vary in terms of the number of points associated with them, so that both the number of items and the number of points are important in evaluating the quality of a claim or subclaim score.

12.1.1 English Language Arts/Literacy

Table 12.1¹³ includes the number of items and the number of points by subclaim and claim for ELA/L Grade 3. Corresponding information is provided in **Appendix 12.1** for all ELA/L grades.

¹² Addendum 12 presents a summary of results on scale scores for the Fall 2015 administration.

¹³ Table A.12.1 in Appendix 12.1 is identical to Table 12.1.

Table 12.1 Form Composition for ELA/L Grade 3

Claims	Subclaims	Number of Items	Number of Points
Reading	Reading Literary Text	10 - 12	20 - 24
	Reading Informational Text	8 - 9	16 - 18
	Vocabulary	5 - 8	10 - 16
	Claim Total	26	52
Writing	Written Expression	2	30
	Knowledge of Conventions	1	12
	Claim Total	3	42
SUMMATIVE TOTAL		29	94

Note: Prose constructed responses (PCRs) consist of at least two writing traits (Written Expression and Writing Knowledge and Conventions) and, in some cases, a reading trait as well. An *aggregated* PCR item score is determined by summing the multiple scores the student received on two or three traits depending on the item. Therefore, each PCR trait is identified as a separate item in this table for the two writing subclaims and, in some cases, either the Reading Literary Text or the Reading Informational Text subclaim.

Each ELA/L form contains items of varying types. The prose constructed response (PCR) traits contribute to different claims and the aggregate of the traits contributes to the summative scale score. The following details the number of possible points and the associated subclaims for the three PCR tasks:

- Literary Analysis Task;
- Research Simulation Task;
- Narrative Writing Task.

The Literary Analysis Task and the Research Simulation Task are scored for two traits/subclaims: “Reading Comprehension & Written Expression” and “Knowledge of Conventions.” The Narrative Writing Task is scored for two traits/subclaims: Written Expression and Knowledge of Conventions. All traits/subclaims are initially scored as either 0-3 or 0-4; the Written Expression subclaims are multiplied by 3 (or weighted) to increase their contribution to the total score, making possible subclaim scores 0, 3, 6, and 9, or 0, 3, 6, 9, and 12. The maximum possible points for ELA/L PCR items are provided in Table 12.2.

Table 12.2 Contribution of Prose Constructed Response Items to ELA/L

Grade	Score	Possible Points		
		Literary Analysis Task	Research Simulation Task	Narrative Writing Task
3	Reading	3	3	0
	Written Expression	9	9	9
	Knowledge of Conventions	3	3	3
	Total	15	15	12
4-5	Reading	4	4	0
	Written Expression	12	12	9
	Knowledge of Conventions	3	3	3
	Total	19	19	12
6-11	Reading	4	4	0
	Written Expression	12	12	12
	Knowledge of Conventions	3	3	3
	Total	19	19	15

12.1.2 Mathematics

Table 12.3¹⁴ includes the numbers of items and points associated with subclaim scores for mathematics grade 3, as an example of the composition of the mathematics tests.

Table 12.3 Mathematics Form Composition for Grade 3

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	26	30
	Additional & Supporting Content	10	10
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		43	66

Because there is substantial variation in the composition of the tests, corresponding information is provided in **Appendix 12.1** tables for all mathematics grades/courses.

12.2 Establishing the Reporting Scales

PARCC reporting scales designate student performance into one of five Performance Levels¹⁵ with Level 1 indicating the lowest level of performance and Level 5 indicating the highest level of performance.

¹⁴ Table A.12.10 in Appendix 12.1 is identical to Table 12.3.

¹⁵ Section 11 provides an overview of the Performance Level Setting process, and detailed information can be found in the Performance Level Setting Technical Report.

Threshold or cut scores associated with performance levels were initially expressed as raw scores on the performance level setting (PLS) forms approved by the PARCC Governing Board.

A scale score task force was assembled by PARCC, which made recommendations about how threshold levels would be represented on the reporting scale.

12.2.1 Full Summative Score Scale and Performance Levels

There are 201 defined full summative scale score points for both ELA/L and mathematics, ranging from 650 to 850. A scale score of 700 is associated with minimum Level 2 performance, a scale score of 750 is associated with minimum Level 4 performance. The threshold for summative performance levels on the scale score metric recommended by the scale score task force are described in Table 12.4.

Table 12.4 Defined Summative Scale Scores

	Lowest Obtainable Scale Score	Level 2	Level 4	Highest Obtainable Scale Score
Full Summative	650	700	750	850

For spring 2015, scale scores were defined for each test as a linear transformation of the theta (θ_{2015}) scale. The theta values associated with the Level 2 and Level 4 performance levels were identified using the test characteristic curve associated with the performance level setting form. With Levels 2 and 4 scale scores fixed at 700 and 750, respectively, the relationship between theta (θ_{2015}) and scale scores ($Scale\ Score_{2015}$) was established as

$$Scale\ Score_{2015} = A_{2015} * \theta_{2015} + B_{2015}$$

where A_{2015} is the slope and B_{2015} is the intercept. The slope and intercept were established as

$$A_{2015} = \frac{750 - 700}{\theta_{2015_Level\ 4} - \theta_{2015_Level\ 2}}$$

and

$$B_{2015} = 750 - A_{2015} * \theta_{2015_Level\ 4}$$

As indicated by these formulas, the slope and intercept for the summative scale scores were based on the theta scale, and by default the IRT parameter scale, established in 2015. Since the spring 2016 IRT parameter scale is the base scale for the IRT parameters, the scaling constants A_{2015} and B_{2015} were updated in order to continue reporting performance levels, summative scale scores, claim scores, and sub-claim performance levels on the same scale as 2015. Maintaining the 2015 scale allows for prior year scores to be compared to current and future scores, and it maintains the performance levels cut scores.

New scaling constants for the summative scale score were needed for the linear transformation of the theta scale (θ_{2016}) to the 2015 reporting scale ($Scale\ Score_{2015}$):

$$Scale\ Score_{2015} = SA_{2016} * \theta_{2016} + SB_{2016}$$

The slope ($slope_{2015_to_2016}$) and intercept ($intercept_{2015_to_2016}$) generated during the year-to-year linking defined the linear relationship between the 2015 theta scale (θ_{2015}) and the 2016 theta scale (θ_{2016}). These values were included in the scale score formula, and the formulas were used to solve for the slope (SA_{2016}) and (SB_{2016}) intercept for 2016.

The slope (A_{2016}) was updated using the following formula:

$$SA_{2016} = A_{2015} / slope_{2015_to_2016},$$

where A_{2015} is the current scale score multiplicative constant, $slope_{2015_to_2016}$ is the multiplicative coefficient from the year-to-year linking, and SA_{2016} is the scale score slope constant for 2016 and beyond.

The intercept (B_{2016}) was updated using the following formula:

$$SB_{2016} = B_{2015} - A_{2016} * intercept_{2015_to_2016},$$

where B_{2015} is the current scale score additive constant, A_{2016} is the updated scale score slope, and (SB_{2016}) is the scale score intercept constant for 2016 and beyond.

In addition, new scaling constants for the reading and writing claim scales were needed. The same formulas were applied by replacing the slope (A_{2015}) and intercept (B_{2015}) with the reading claim slope and intercept and the writing claim slope and intercept.

A and B values resulting from these calculations as well as the theta values associated with the threshold performance levels are included in **Appendix 12.2**; **Appendix 12.3** includes raw to scale score conversion tables for the performance level setting forms.

12.2.2 ELA/L Reading and Writing Claim Scale

There are 81 defined scale score points for Reading, ranging from 10 to 90. A scale score of 30 is associated with minimum Level 2 performance, a scale score of 50 is associated with minimum Level 4 performance. There are 51 defined scale score points for Writing, ranging from 10 to 60. A scale score of 25 is associated with minimum Level 2 performance, a scale score of 35 is associated with minimum Level 4 performance. The threshold Reading and Writing performance levels on the scale score metric recommended by the scale score task force are described in Table 12.5.

Table 12.5 Defined Scaled Scores for Reading and Writing Claim Scores

	Lowest Obtainable Scale Score	Level 2	Level 4	Highest Obtainable Scale Score
Reading	10	30	50	90
Writing	10	25	35	60

As with the full summative scores, scale scores for Reading and Writing were defined for each test as a linear transformation of the IRT theta (Θ) scale. The same IRT theta scale was used for Reading and Writing as was used for the ELA/L full summative scores. The theta values associated with the Level 2 and Level 4 performance levels were identified using the test characteristic curve associated with the performance level setting form. As with the full summative scores, the relationship between theta and scale scores was established with Level 2 and 4 theta scores and the corresponding predefined scale scores. The formulas used for this are provided in Table 12.6.

Table 12.6 Calculating Scaling Constants for Reading and Writing Claim Scores

Reading	Writing
$Scale = A_R * \Theta + B_R$	$Scale = A_W * \Theta + B_W$
$A_R = \frac{50 - 30}{\Theta_{Level\ 4} - \Theta_{Level\ 2}}$	$A_W = \frac{35 - 25}{\Theta_{Level\ 4} - \Theta_{Level\ 2}}$
$B_R = 50 - A * \Theta_{Level\ 4}$	$B_W = 35 - A * \Theta_{Level\ 4}$

A and B values resulting from these calculations are included in **Appendix 12.2**.

12.2.3 Subclaims Scale

The Level 4 cut is defined as *Meets or Exceeds Expectations* because high school students at Level 4 or above are likely to have the skills and knowledge to meet the definition of career and college readiness. The Level 3 cut is defined as *Nearly Meets Expectations*. Subclaim outcomes center on the Level 3 and Level 4 performance levels and are reported at three levels:

- Below Expectations;
- Nearly Meets Expectations; or
- Meets or Exceeds Expectations.

The subclaim performance levels are designated through the IRT theta (Θ) scale for the items associated with a particular subclaim. The theta values and corresponding raw scores associated with the Level 3 and Level 4 performance levels were identified using the test characteristic curve. Students earning a raw subclaim score equal to or greater than the Level 4 threshold were designated as *Meets or Exceeds Expectations*. Students not earning a raw subclaim score equal to or greater than the Level 3 threshold were designated as *Below Expectations*. Other students, whose raw subclaim score fell between the Level 3 and 4 thresholds, were designated as *Nearly Meets Expectations*.

12.3 Creating Conversion Tables

A PARCC conversion table relates the number of points earned by a student on the ELA/L full summative score, the mathematics full summative score, the Reading claim score, or the Writing claim score to the corresponding scale score for the test form administered to that student. An IRT inverse test characteristic curve (TCC) approach is used to develop the relationship between point scores and IRT

ability estimates (θ_s). In carrying out the calculations, estimates of item parameters and thetas are substituted for parameters in the formulas in each step.

Step 1: Calculate the expected item score (i.e., estimated item true score) for every scale score in the selected range (determined by LOSS, HOSS, and scale score increment) based on the generalized partial credit model for both dichotomous and polytomous items:

$$s_i(\theta_j) = \sum_{m=0}^{M_i-1} m p_{im}(\theta_j), \quad (12-4)$$

$$p_{im}(\theta_j) = \frac{\exp \left[\sum_{k=0}^m D a_i(\theta_j - b_i + d_{ik}) \right]}{\sum_{v=0}^{M_i-1} \exp \left[\sum_{k=0}^v D a_i(\theta_j - b_i + d_{iv}) \right]}, \quad (12-5)$$

where $a_i(\theta_j - b_i + d_{i0}) \equiv 0$; $s_i(\theta_j)$ is the expected item score for item i on a scale score, θ_j ; $p_{im}(\theta_j)$ is the probability of a test taker with θ_j getting score m on item i ; M_i is the number of score categories of item i with possible item scores as consecutive integers from 0 to $M_i - 1$; D is the IRT scale constant (1.7); a_i is a slope parameter; b_i is a location parameter reflecting overall item difficulty; d_{ik} is a location parameter incrementing the overall item difficulty to reflect the difficulty of earning score category k ; v is the number of score categories.

Step 2: Calculate the expected (weighted) test score for every scale score in the selected range:

$$T_j = \sum_{i=1}^I w_i s_i(\theta_j), \quad (12-6)$$

where T_j is the expected (weighted) test score on a scale score, θ_j ; w_i is the item weight for item i (e.g., with $w_i = 2$, a dichotomous item is scored as 0 or 2, and a three-category item is scored as 0, 2, or 4); I is the total number of items in a test form.

Step 3: Calculate the estimated conditional standard error of measurement (CSEM) for each scale score in the selected range:

$$\text{CSEM}_j = \sqrt{\frac{1}{\sum_{i=1}^I L_i(\theta_j)}}, \quad (12-7)$$

$$L_i(\theta_j) = (D a_i)^2 [s_{i2}(\theta_j) - s_i^2(\theta_j)], \quad (12-8)$$

$$s_{i2}(\theta_j) = \sum_{m=0}^{M_i-1} m^2 p_{im}(\theta_j), \quad (12-9)$$

where $L_i(\theta_j)$ is the estimated item information function for item i on scale score θ_j .

Step 4: Match every raw score with a scale score. θ_j is the scale score for a raw score r_h if $T_j - r_h$ is minimum across all T_j s.

Figure 12.1 contains TCCs, estimated information (INF) curves, and estimated conditional standard error of measurement (CSEM) curves for ELA/L grade 3.¹⁶ The curves in each figure are for the three core online forms and two core paper forms. The average difficulty of each form is reported and the curves are reported on the theta scale. Vertical dotted lines indicate the performance level cuts on the theta scale. For ELA/L grade 3, the three CBT and two paper PBT forms had very similar TCCs. Information and CSEM curves were similar for CBT forms and PBT forms. **Appendix 12.4** contains TCC, INF, and CSEM curves for all ELA/L grades and all mathematics grades/courses.

¹⁶ Grade 3 TCC, INF, and CSEM curves are also included in Appendix 12.4 as Figures A.12.1.

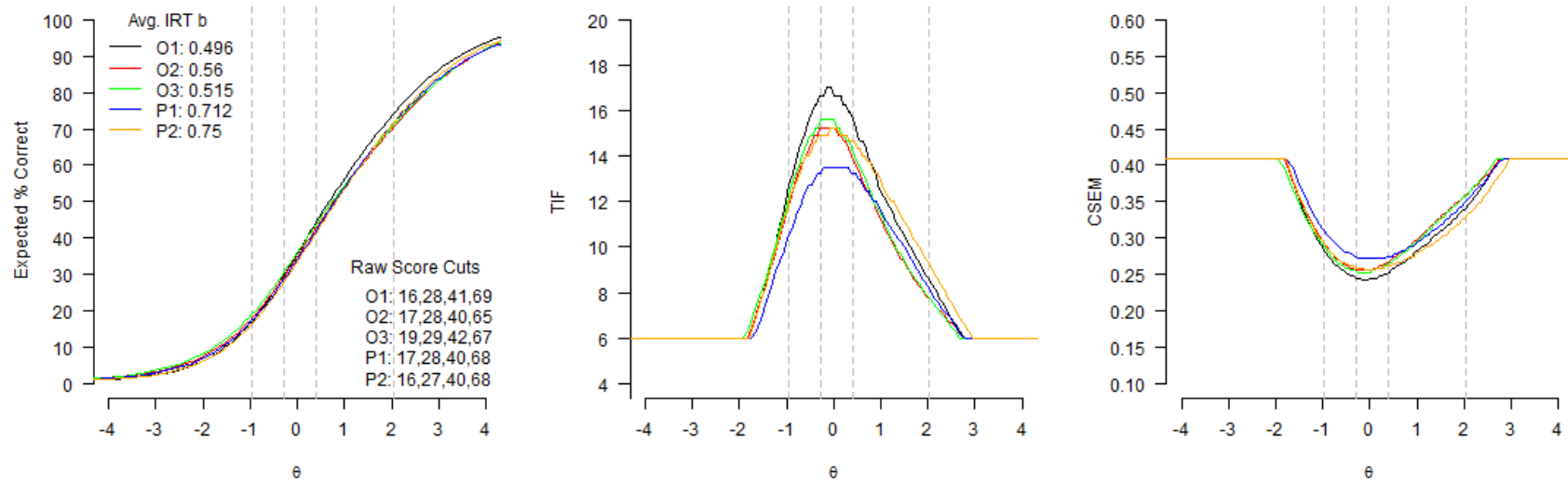


Figure 12.1 Test Characteristic Curves, Information Curves, and Conditional Standard Error of Measurement Curve for ELA/L Grade 3

12.4 Score Distributions

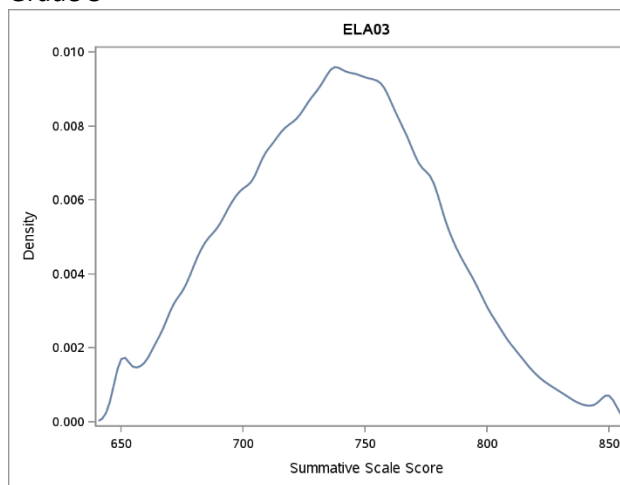
12.4.1 Score Distributions for ELA/L

All Students

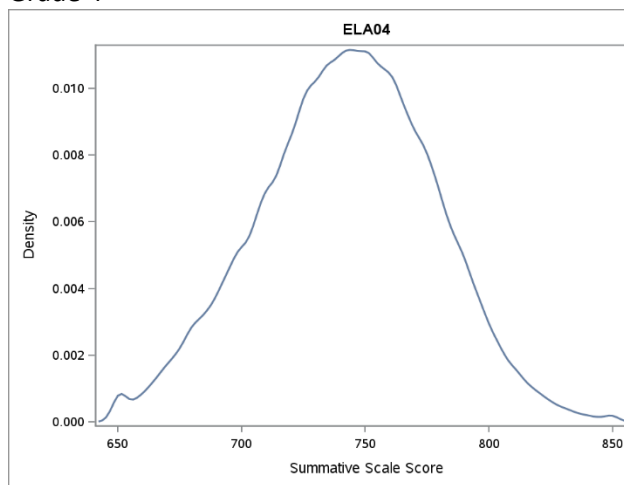
Figures 12.4 through 12.6 graphically represent the distributions of scale scores for grades 3 through 11 ELA/L FS, Reading, and Writing, respectively. The vertical axis of each graph, labeled “Density”, represents the proportion of students earning the scale score point indicated along the horizontal axis.

ELA/L scale scores that were a bit below the Level 4 cut score (i.e., 750) were most commonly observed for grades 3 to 11.

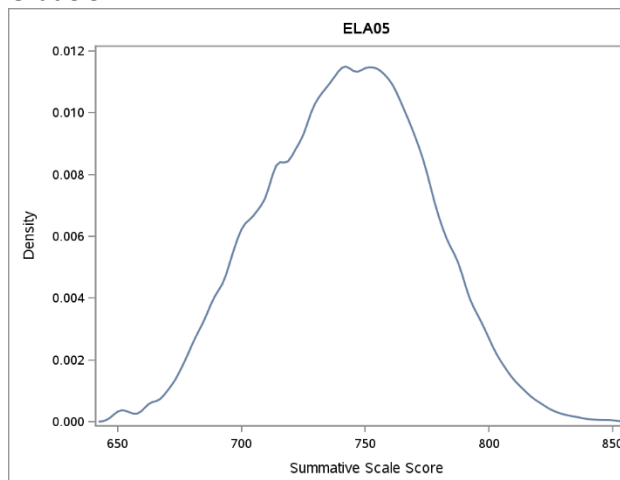
Grade 3



Grade 4



Grade 5



Grade 6

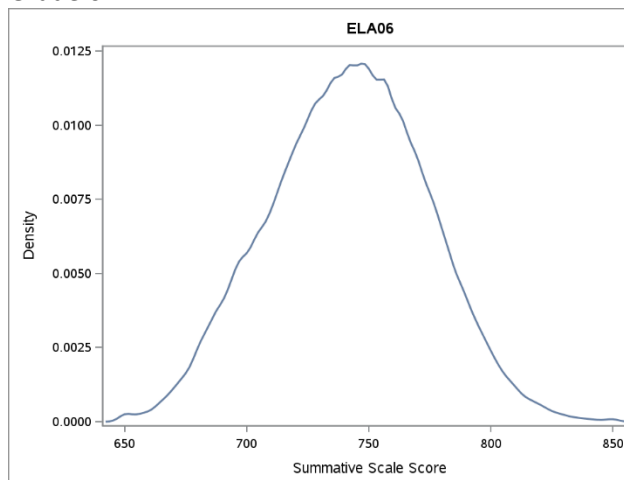
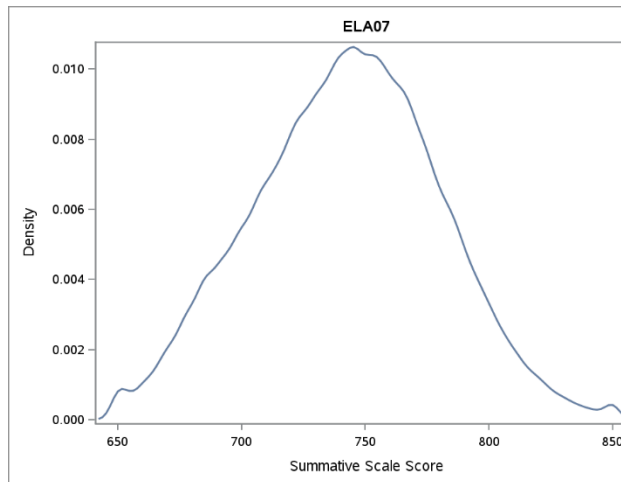
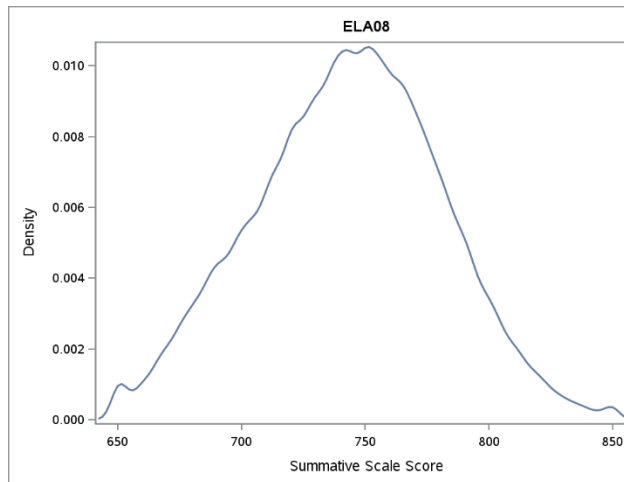


Figure 12.4 Distributions of ELA/L Scale Scores: Grades 3 to 11

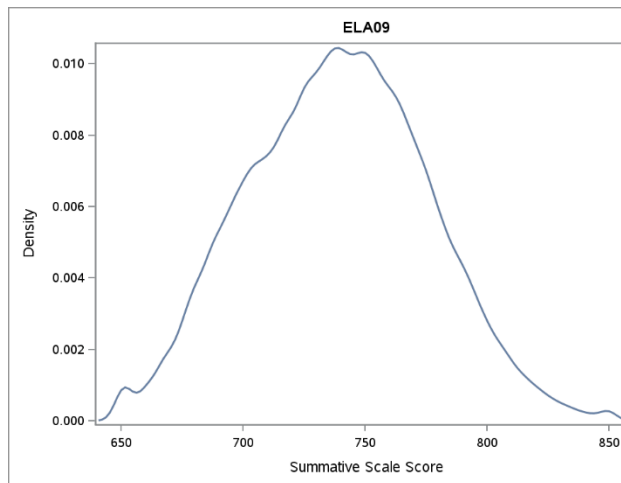
Grade 7



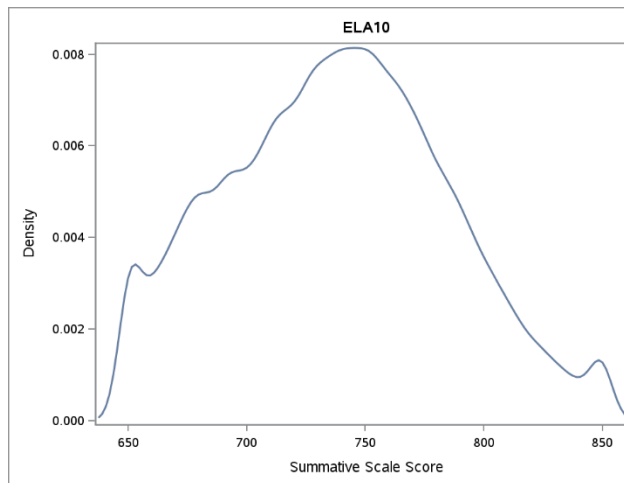
Grade 8



Grade 9



Grade 10



Grade 11

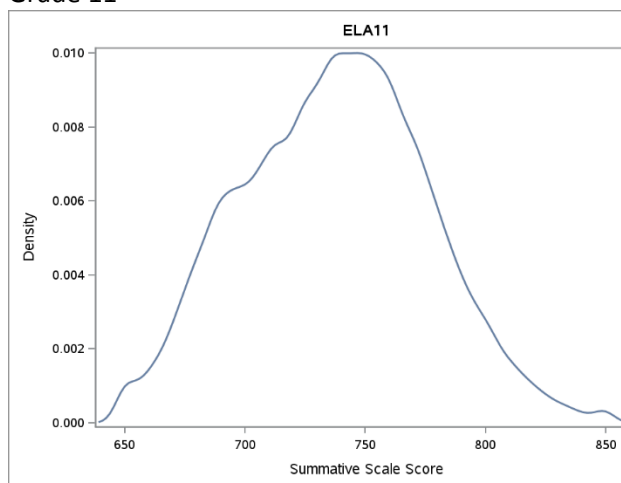
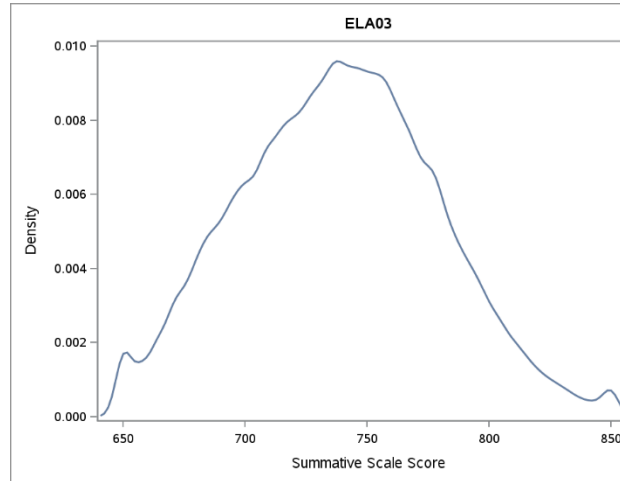


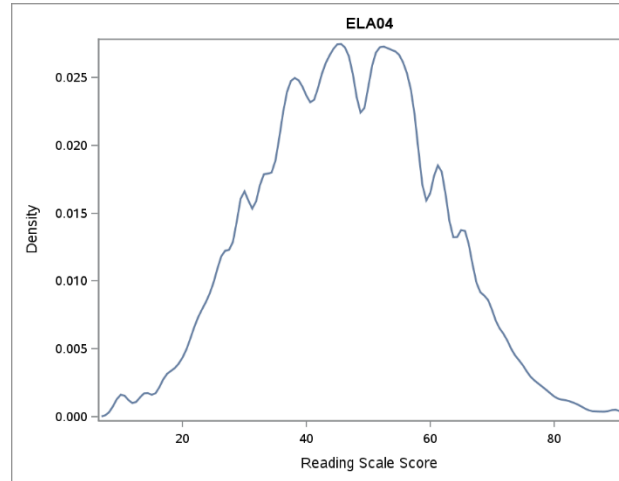
Figure 12.4 (continued) Distributions of ELA/L Scale Scores: Grades 3 to 11

Reading scale scores that were a bit below the Level 4 cut score of 50 were most often observed. Distributions were fairly symmetric, with scores below the Level 4 cut score being a bit more common than higher scores. A portion of this is due to larger numbers of students earning near zero raw scores than near perfect raw scores. Near zero raw scores can occur for a variety of reasons, such as student illness, and do not always indicate weak student skills.

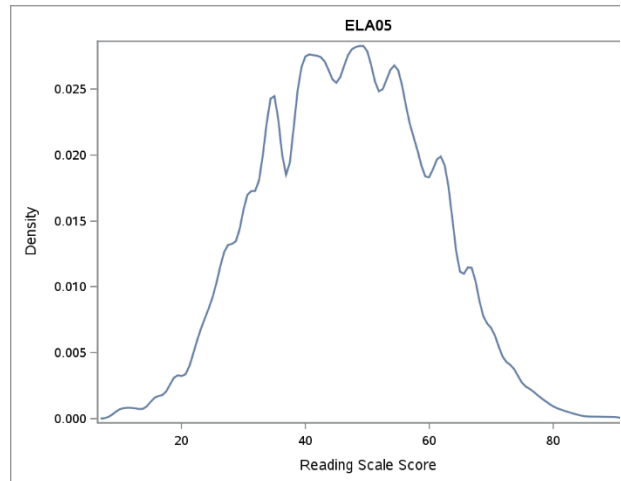
Grade 3



Grade 4



Grade 5



Grade 6

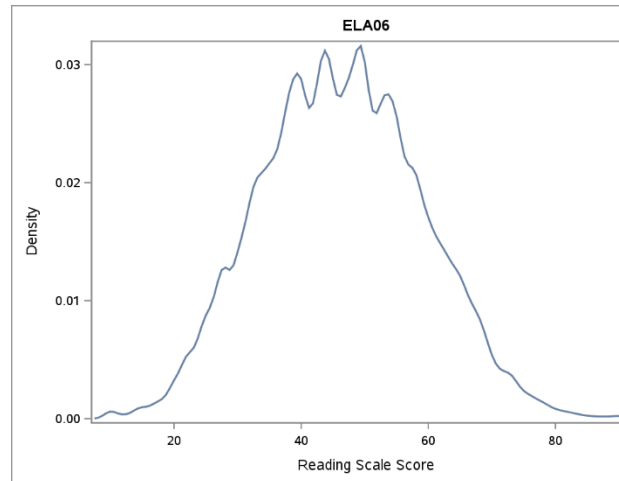
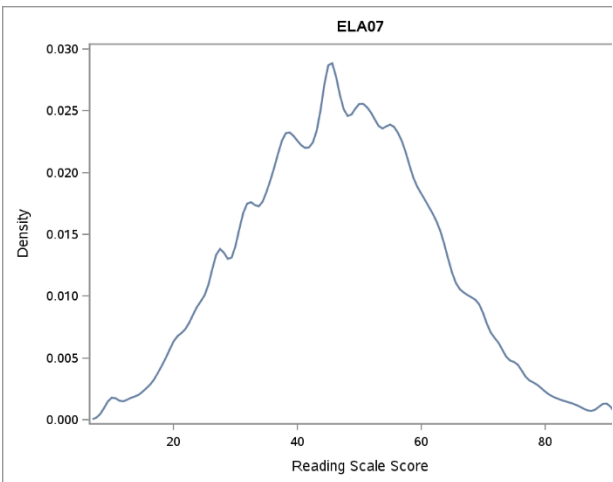
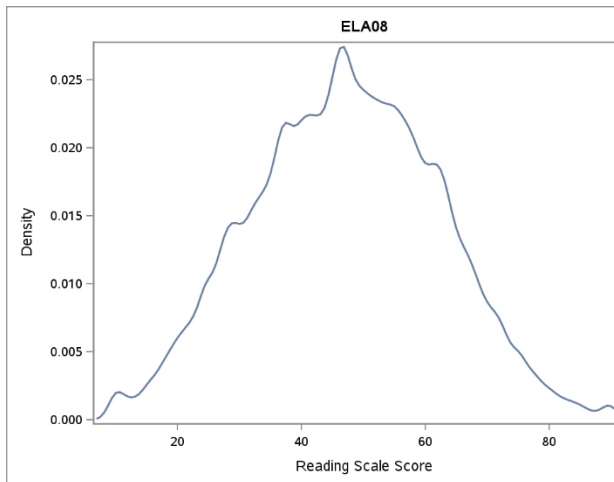


Figure 12.5 Distributions of Reading Scale Scores: Grades 3 to 11

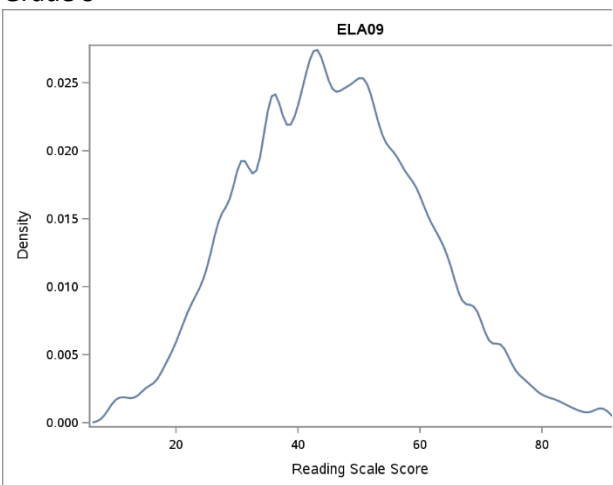
Grade 7



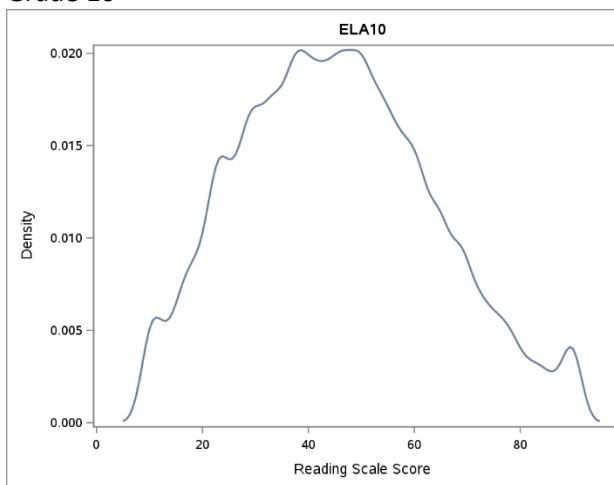
Grade 8



Grade 9



Grade 10



Grade 11

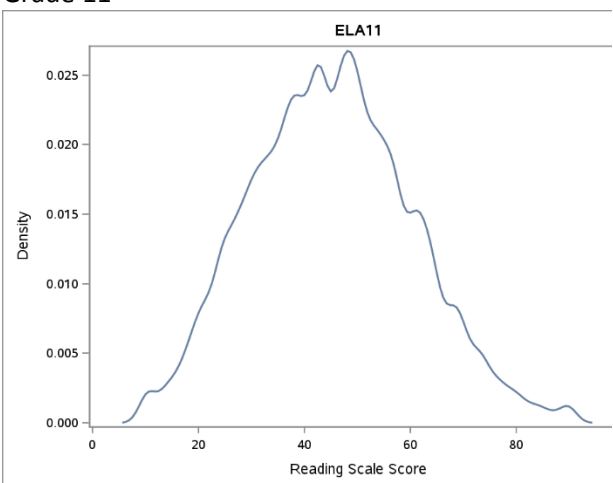
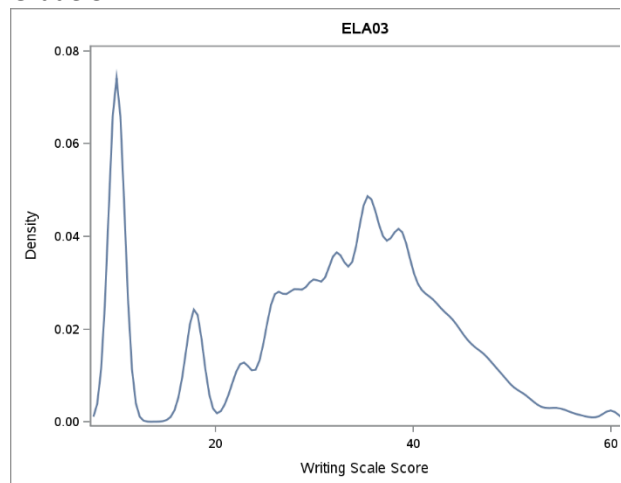


Figure 12.5 (continued) Distributions of Reading Scale Scores: Grades 3 to 11

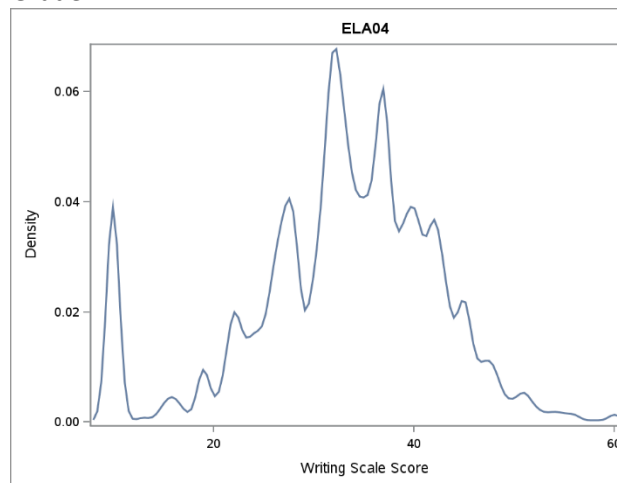
Writing scale score distributions are noticeably less smooth than Reading or ELA/L FS distributions due to peaks related to the weighting of the Written Expression portion of the PCR tasks.

The proportion of students earning the lowest obtainable scale score is fairly high for Writing. This occurred even though a score point of zero is typically necessary to obtain the minimum scale score. Writing items are embedded exclusively in PCR tasks, which tended to be difficult. Written Expression trait/subclaim also tended to be the most difficult of the PCR traits.

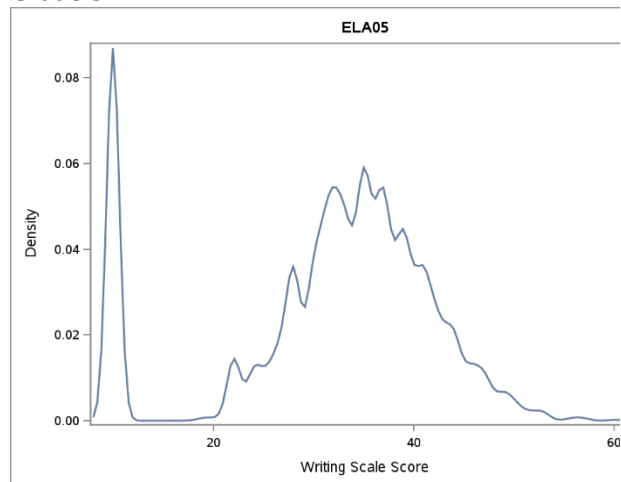
Grade 3



Grade 4



Grade 5



Grade 6

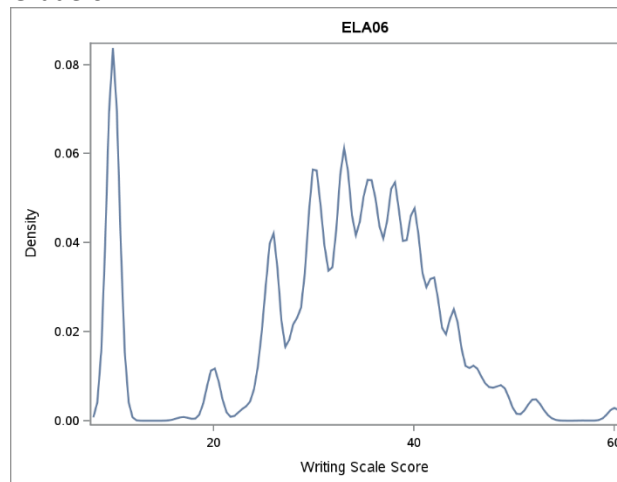
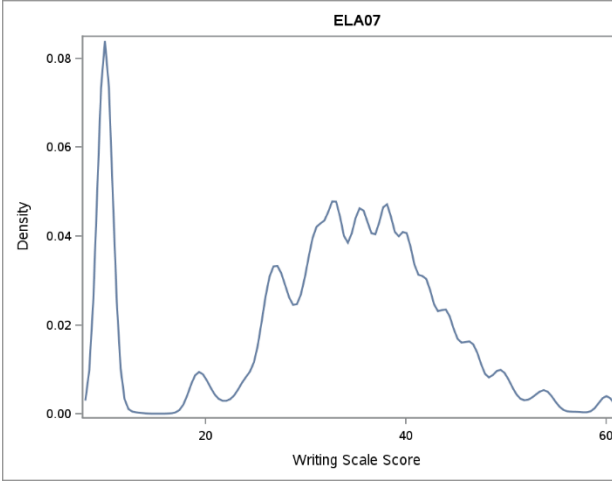
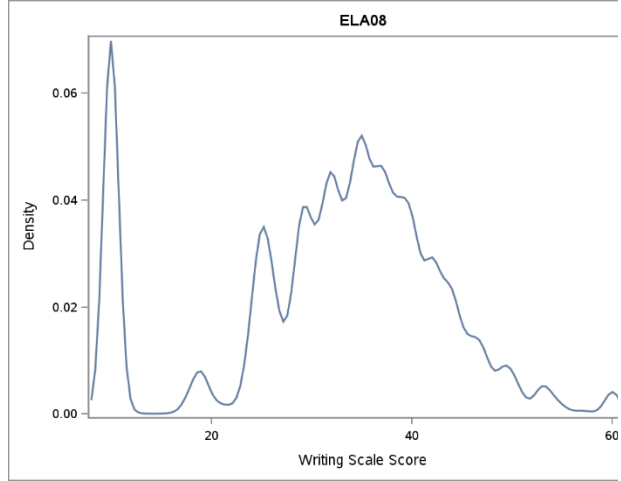


Figure 12.6 Distributions of Writing Scale Scores: Grades 3 to 11

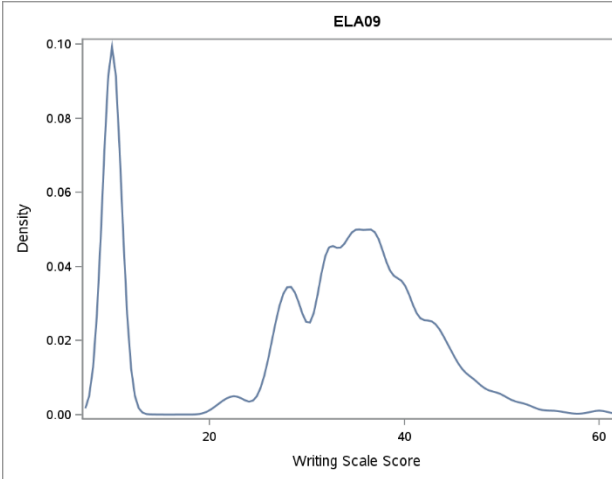
Grade 7



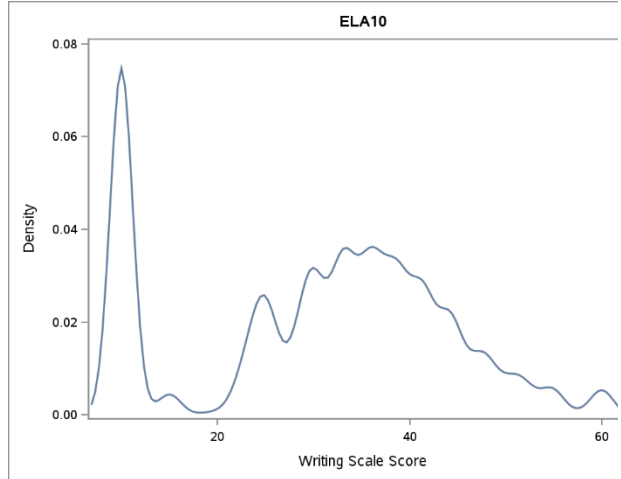
Grade 8



Grade 9



Grade 10



Grade 11

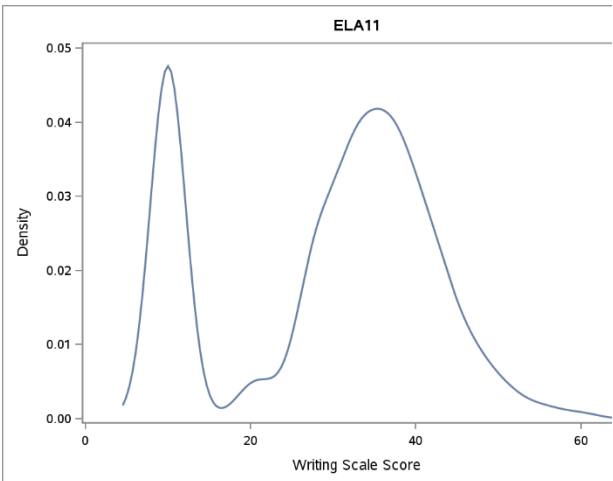


Figure 12.6 (continued) Distributions of Writing Scale Scores: Grades 3 to 11

Groups

Grade 3 group statistics for ELA/L FS, Reading, and Writing scale scores are presented in Table 12.7.¹⁷ Mean scores were higher for female students relative to male students. Mean scores were highest for Asian students followed by White students and were lowest for American Indian/Alaska Native students. Economically disadvantaged students performed less well than students who are not economically disadvantaged. English learners (EL) performed less well than Non English learner students. Students with disabilities (SWD) performed less well than students without disabilities.

Patterns of mean scale scores were extremely similar in other grades; corresponding tables for all grades are presented in **Appendix 12.5**.

Table 12.7 Subgroup Performance for ELA/L: Grade 3

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		471,801	738.48	40.58	650	850
Gender	Female	231,217	743.26	40.89	650	850
	Male	240,584	733.90	39.74	650	850
Ethnicity	American Indian/Alaska Native	5,431	716.71	34.76	650	850
	Asian	27,059	765.40	39.86	650	850
	Black or African American	78,661	722.26	37.94	650	850
	Hispanic/Latino	131,457	724.68	37.37	650	850
	Native Hawaiian or Pacific Islander	803	749.12	39.26	650	850
	Multiple Race Selected	15,836	744.02	40.43	650	850
	White	212,059	749.76	38.05	650	850
Economic Status*	Economically Disadvantaged	228,865	722.74	36.75	650	850
	Not Economically Disadvantaged	237,879	753.32	38.38	650	850
English Learner Status	English Learner (EL)	68,638	713.99	34.27	650	850
	Non English Learner	396,667	742.77	40.05	650	850
Disabilities	Students with Disabilities (SWD)	63,952	709.97	38.13	650	850
	Students without Disabilities	245,463	745.33	39.39	650	850
Reading Score		471,801	45.19	16.10	10	90
Gender	Female	231,217	46.47	16.18	10	90
	Male	240,584	43.96	15.93	10	90
Ethnicity	American Indian/Alaska Native	5,431	36.37	13.55	10	90
	Asian	27,059	55.21	16.06	10	90
	Black or African American	78,661	38.65	14.58	10	90
	Hispanic/Latino	131,457	39.37	14.46	10	90
	Native Hawaiian or Pacific Islander	803	48.52	15.20	10	90

¹⁷ Table A.12.48 in Appendix 12.5 is identical to Table 12.7.

Group Type	Group	N	Mean	SD	Min	Max
Economic Status*	Multiple Race Selected	15,836	47.78	16.18	10	90
	White	212,059	49.97	15.31	10	90
	Economically Disadvantaged	228,865	38.82	14.24	10	90
	Not Economically Disadvantaged	237,879	51.18	15.44	10	90
English Learner Status	English Learner (EL)	68,638	35.01	12.95	10	90
	Non English Learner	396,667	46.96	15.93	10	90
Disabilities	Students with Disabilities (SWD)	63,952	34.68	15.11	10	90
	Students without Disabilities	245,463	47.64	15.74	10	90
Writing Score		471,801	31.18	11.90	10	60
Gender	Female	231,217	33.13	11.50	10	60
	Male	240,584	29.32	11.97	10	60
Ethnicity	American Indian/Alaska Native	5,431	26.48	11.27	10	60
	Asian	27,059	38.17	10.46	10	60
	Black or African American	78,661	27.46	11.95	10	60
	Hispanic/Latino	131,457	28.43	11.73	10	60
	Native Hawaiian or Pacific Islander	803	34.50	11.54	10	60
	Multiple Race Selected	15,836	31.96	11.85	10	60
	White	212,059	33.44	11.24	10	60
Economic Status*	Economically Disadvantaged	228,865	27.62	11.72	10	60
	Not Economically Disadvantaged	237,879	34.55	11.05	10	60
English Learner Status	English Learner (EL)	68,638	26.17	11.51	10	60
	Non English Learner	396,667	32.07	11.74	10	60
Disabilities	Students with Disabilities (SWD)	63,952	23.14	11.94	10	60
	Students without Disabilities	245,463	33.17	11.31	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Grade 9 group statistics for ELA/L, Reading, and Writing scale scores are presented in Table 12.8.¹⁸ Mean scores were very similar to what was observed for grades 3 to 8. Mean scores were higher for female students than for male students. Mean scores were highest for Asian students followed by White students; scores were lowest for Black or African American students. Economically disadvantaged students performed less well than students who are not economically disadvantaged. English learners (EL) performed less well than Non English Learner students. Students with disabilities (SWD) performed less well than students without disabilities.

¹⁸ Table A.12.54 in Appendix 12.5 is identical to Table 12.8.

Very similar patterns are observed in other grades, and corresponding tables for all grades are presented in **Appendix 12.5**.

Table 12.8 Subgroup Performance for ELA/L: Grade 9

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		275,158	738.99	36.84	650	850
Gender	Female	134,144	746.59	35.93	650	850
	Male	141,014	731.76	36.23	650	850
Ethnicity	American Indian/Alaska Native	3,669	724.04	29.81	650	828
	Asian	17,145	767.48	37.14	650	850
	Black or African American	37,389	723.04	33.25	650	850
	Hispanic/Latino	85,831	727.43	33.73	650	850
	Native Hawaiian or Pacific Islander	417	748.40	36.39	650	850
	Multiple Race Selected	5,774	743.50	37.43	650	850
	White	124,609	748.09	34.93	650	850
Economic Status*	Economically Disadvantaged	117,958	724.96	33.12	650	850
	Not Economically Disadvantaged	151,423	749.76	35.89	650	850
English Learner Status	English Learner (EL)	19,031	702.15	26.81	650	846
	Non English Learner	251,285	741.91	35.92	650	850
Disabilities	Students with Disabilities (SWD)	39,611	712.05	32.17	650	850
	Students without Disabilities	127,890	746.64	36.61	650	850
Reading Score		275,158	45.80	14.81	10	90
Gender	Female	134,144	47.91	14.51	10	90
	Male	141,014	43.80	14.82	10	90
Ethnicity	American Indian/Alaska Native	3,669	38.99	11.87	10	79
	Asian	17,145	56.28	15.24	10	90
	Black or African American	37,389	39.61	13.11	10	90
	Hispanic/Latino	85,831	40.85	13.23	10	90
	Native Hawaiian or Pacific Islander	417	48.93	14.59	10	90
	Multiple Race Selected	5,774	48.26	15.36	10	90
	White	124,609	49.73	14.26	10	90
Economic Status*	Economically Disadvantaged	117,958	40.01	12.98	10	90
	Not Economically Disadvantaged	151,423	50.23	14.59	10	90
English Learner Status	English Learner (EL)	19,031	30.82	9.94	10	90
	Non English Learner	251,285	46.97	14.48	10	90
Disabilities	Students with Disabilities (SWD)	39,611	35.61	12.94	10	90
	Students without Disabilities	127,890	48.36	14.66	10	90
Writing Score		275,158	29.76	12.42	10	60
Gender	Female	134,144	33.10	11.31	10	60

Group Type	Group	N	Mean	SD	Min	Max
Ethnicity	Male	141,014	26.59	12.60	10	60
	American Indian/Alaska Native	3,669	27.23	11.25	10	56
	Asian	17,145	38.17	10.53	10	60
	Black or African American	37,389	24.88	12.40	10	60
	Hispanic/Latino	85,831	27.07	12.23	10	60
	Native Hawaiian or Pacific Islander	417	33.25	11.23	10	60
	Multiple Race Selected	5,774	30.18	12.51	10	60
Economic Status*	White	124,609	31.99	11.72	10	60
	Economically Disadvantaged	117,958	26.08	12.27	10	60
English Learner Status	Not Economically Disadvantaged	151,423	32.60	11.77	10	60
	English Learner (EL)	19,031	19.94	11.20	10	60
	Non English Learner	251,285	30.56	12.17	10	60
Disabilities	Students with Disabilities (SWD)	39,611	21.12	11.96	10	60
	Students without Disabilities	127,890	32.39	11.93	10	60

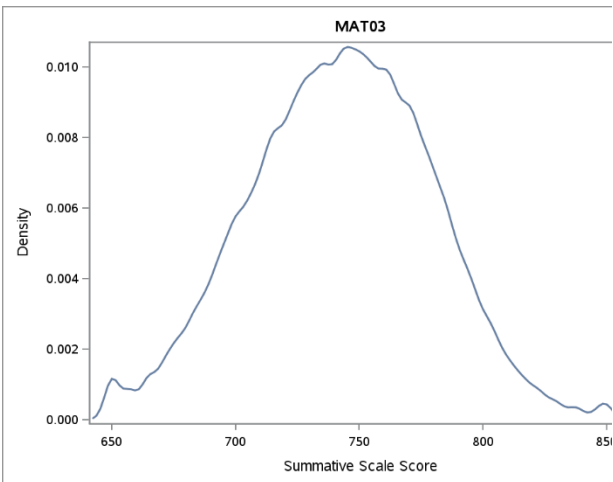
Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

12.4.2 Score Distributions for Mathematics

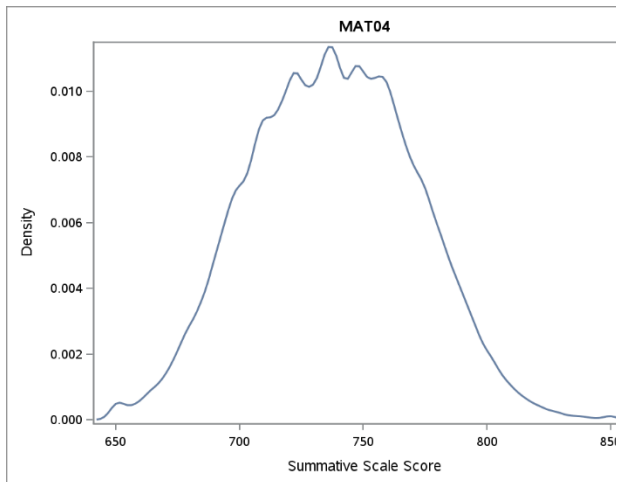
All Students

Figure 12.7 graphically represents the distributions of scale scores for grades 3 through 8 mathematics. Scale score distributions peaked midway between 700 and the Level 4 performance level cut of 750.

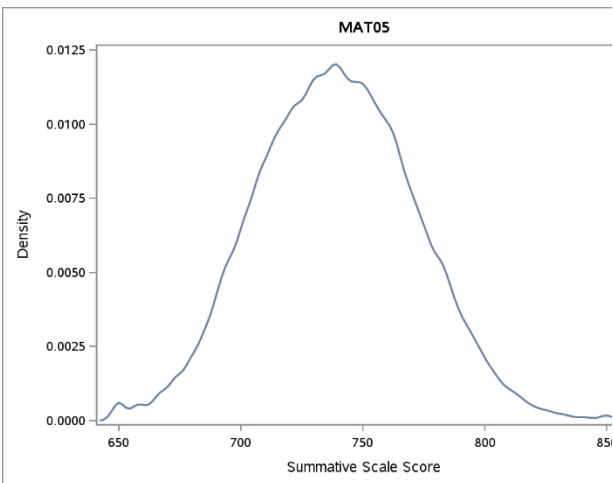
Grade 3



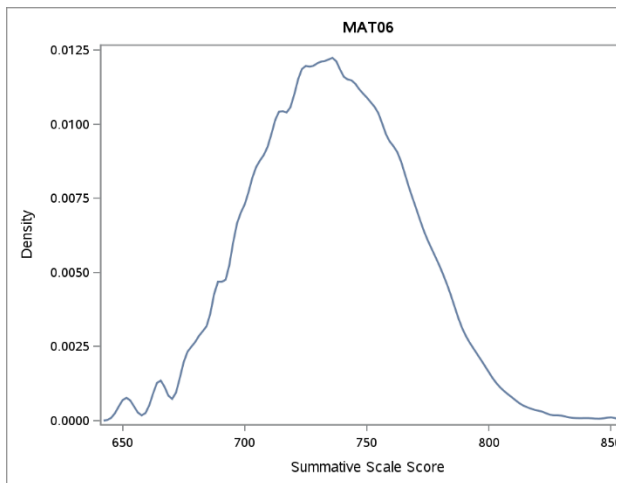
Grade 4



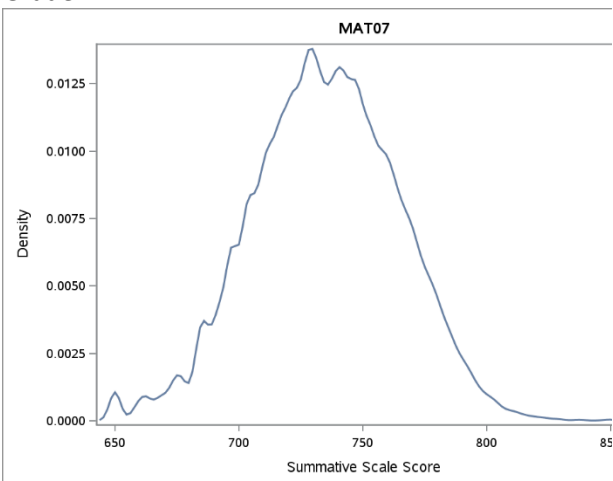
Grade 5



Grade 6



Grade 7



Grade 8

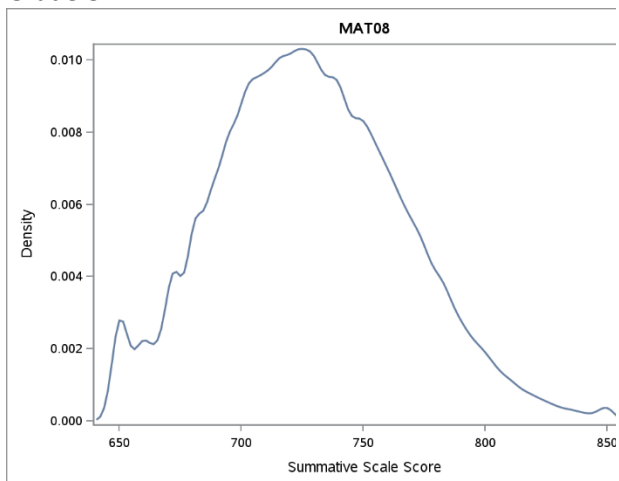
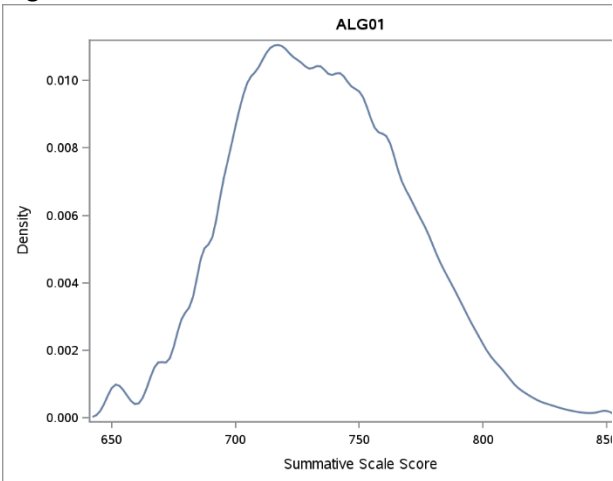


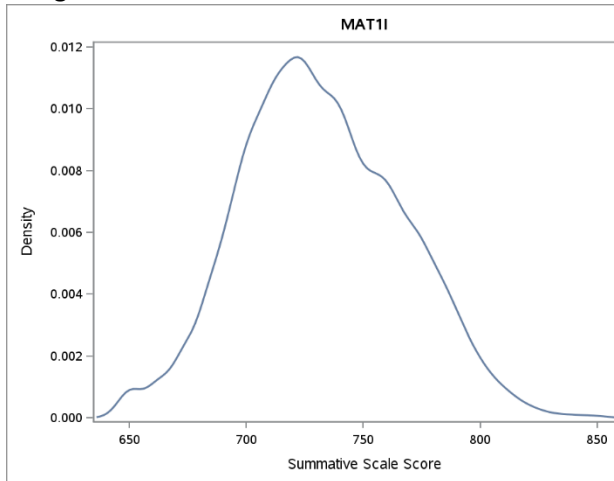
Figure 12.7 Distributions of Mathematics Scale Scores: Grades 3 to 8

Figure 12.8 graphically represents the distributions of scale scores for Algebra I, Geometry, Algebra II, and Integrated Mathematics I through III. Scale score distributions peaked between 700 and the 750 Level 4 performance level cut score. Distributions were similar for Algebra I and Integrated Mathematics I, for Geometry and Integrated Mathematics II, and for Algebra II and Integrated Mathematics III.

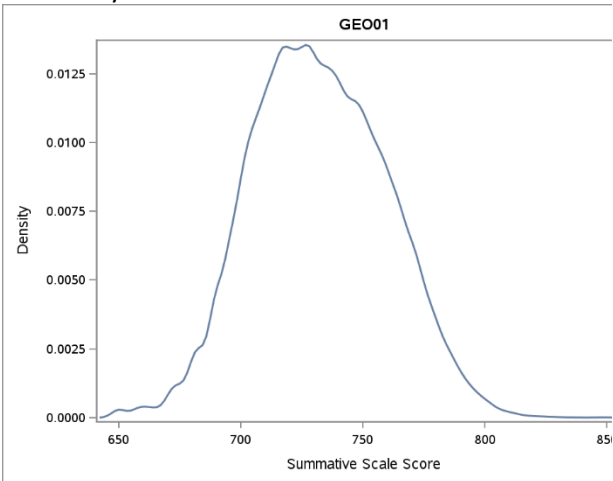
Algebra I



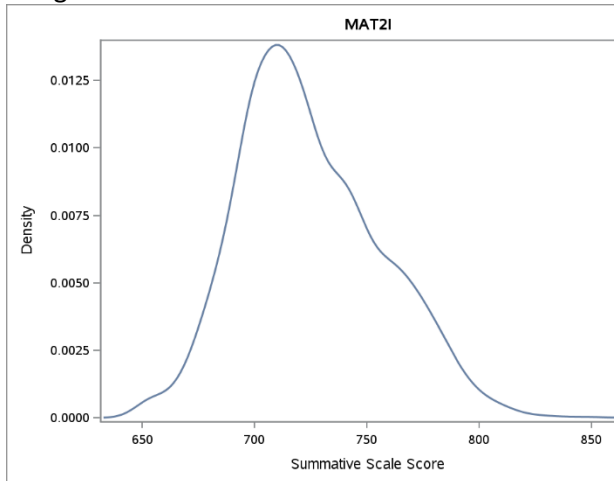
Integrated Mathematics I



Geometry



Integrated Mathematics II



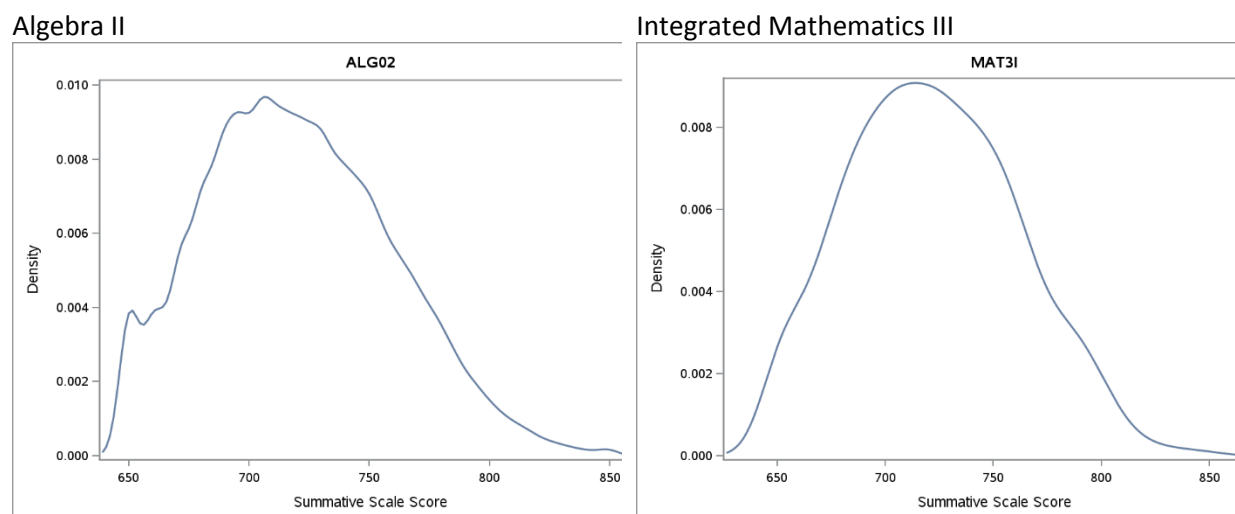


Figure 12.8 (continued) Distributions of Mathematics Scale Scores: High School

Groups

Grade 3 group statistics for mathematics scale scores are presented in Table 12.9.¹⁹ Mean scores were slightly higher for female students relative to male students. Mean scores were highest for Asian students followed by White students and were lowest for American Indian/Alaska Native students. Economically disadvantaged students performed less well than students who are not economically disadvantaged. English learners (EL) performed less well than Non English learner students. Students with disabilities (SWD) performed less well than students without disabilities.

Table 12.9 Subgroup Performance for Mathematics Scale Scores: Grade 3

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		476,620	742.64	36.55	650	850
Gender	Female	233,536	743.30	35.54	650	850
	Male	243,084	742.00	37.49	650	850
Ethnicity	American Indian/Alaska Native	5,436	723.43	31.20	650	845
	Asian	27,498	772.68	35.53	650	850
	Black or African American	78,668	726.06	34.28	650	850
	Hispanic/Latino	135,427	730.79	33.24	650	850
	Native Hawaiian or Pacific Islander	817	748.97	35.46	650	850
	Multiple Race Selected	15,843	746.02	36.86	650	850
	White	212,345	752.70	33.75	650	850
Economic Status*	Economically Disadvantaged	232,719	728.42	33.34	650	850
	Not Economically Disadvantaged	238,816	756.22	34.28	650	850

¹⁹ Table A.12.57 in Appendix 12.5 is identical to Table 12.9.

English Learner Status	English Learner (EL)	73,569	724.77	32.50	650	850
	Non English Learner	396,435	746.02	36.24	650	850
Disabilities	Students with Disabilities (SWD)	64,259	718.66	37.10	650	850
	Students without Disabilities	248,483	748.90	35.18	650	850

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Similar patterns were observed in other grades; corresponding tables for all grades are presented in **Appendix 12.5**.

Algebra I scale score statistics are presented in Table 12.10.²⁰ Mean scores were higher for female students relative to male students. Mean scores were highest for Asian students followed by White students and were lowest for American Indian/Alaska Native students. Economically disadvantaged students performed less well than students who are not economically disadvantaged. English learners (EL) performed less well than Non English learner students. Students with disabilities (SWD) performed less well than students without disabilities.

Table 12.10 Subgroup Performance for Mathematics Scale Scores: Algebra I

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		323,701	735.00	34.43	650	850
Gender	Female	156,671	737.07	33.14	650	850
	Male	167,030	733.05	35.49	650	850
Ethnicity	American Indian/Alaska Native	3,914	717.72	26.42	650	821
	Asian	19,893	765.60	36.37	650	850
	Black or African American	59,226	719.96	29.30	650	850
	Hispanic/Latino	93,396	723.59	29.61	650	850
	Native Hawaiian or Pacific Islander	504	743.90	36.53	650	850
	Multiple Race Selected	7,298	739.34	34.65	650	850
	White	139,104	744.96	33.14	650	850
Economic Status*	Economically Disadvantaged	140,960	721.81	29.38	650	850
	Not Economically Disadvantaged	177,020	745.33	34.58	650	850
English Learner Status	English Learner (EL)	24,388	709.23	25.78	650	850
	Non English Learner	294,669	737.12	34.14	650	850
Disabilities	Students with Disabilities (SWD)	48,956	713.19	29.77	650	850
	Students without Disabilities	194,227	741.88	34.77	650	850

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

²⁰ Table A.12.63 in Appendix 12.5 is identical to Table 12.10.

Very similar patterns were observed in Geometry and Algebra II; corresponding tables are presented in **Appendix 12.5**.

Integrated Mathematics I scale score statistics are presented in Table 12.11.²¹ Mean scores were higher for female students relative to male students. Mean scores were highest for Asian students followed by White students and were lowest for American Indian/Alaska Native students. Economically disadvantaged students performed less well than students who are not economically disadvantaged. English learners (EL) performed less well than non-English learner students. Students with disabilities (SWD) performed less well than students without disabilities.

²¹ Table A.12.66 in Appendix 12.5 is identical to Table 12.11.

Table 12.11 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics I

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		16,581	732.30	34.26	650	850
Gender	Female	8,128	734.36	33.01	650	841
	Male	8,453	730.33	35.30	650	850
Ethnicity	American Indian/Alaska Native	166	715.13	28.28	650	790
	Asian	551	749.85	37.96	650	850
	Black or African American	1,891	720.44	30.49	650	821
	Hispanic/Latino	6,552	723.95	30.77	650	835
	Native Hawaiian or Pacific Islander	30	735.00	36.69	673	810
	Multiple Race Selected	416	736.19	35.74	650	850
	White	6,968	742.18	34.65	650	850
Economic Status*	Economically Disadvantaged	8,230	722.92	30.91	650	838
	Not Economically Disadvantaged	8,347	741.57	34.87	650	850
English Learner Status	English Learner (EL)	2,106	710.19	25.55	650	815
	Non English Learner	14,074	736.33	33.95	650	850
Disabilities	Students with Disabilities (SWD)	2,019	709.23	29.76	650	830
	Students without Disabilities	4,216	725.38	31.59	650	848

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Very similar patterns were observed in Integrated Mathematics II and Integrated Mathematics III; corresponding tables are presented in **Appendix 12.5**.

12.5 Interpreting Claim Scores and Subclaim Scores

12.5.1 Interpreting Claim Scores

PARCC ELA/L assessments provide separate claim scale scores for both Reading and Writing. The claim scale scores and the summative scale score are on different scales; therefore, the sum of the scale scores for each claim will not equal the summative scale score. PARCC Reading scale scores range from 10 to 90 and PARCC Writing scale scores range from 10 to 60.

The claim scores can be interpreted by comparing a student's claim scale score to the average performance for the school, district, and state. The PARCC Individual Student Report (ISR) provides the student scale score results and the average scale score results for the school, district, and state.

12.5.2 Interpreting Subclaim Scores

Within each reporting category are specific skill sets (subclaims) students demonstrate on the PARCC assessments. Subclaim categories are not reported using scale scores or Performance Levels. Subclaim

performance for PARCC assessments is reported using graphical representations that indicate how the student performed relative to the Level 3 and Level 4 performance levels for the content area.

Subclaim indicators represent how well students performed in a subclaim category relative to Performance Level 3 and Level 4 thresholds for the items associated with the subclaim category. To determine a student's subclaim performance, the Performance Level 3 and Level 4 thresholds corresponding to the IRT based performance for the items for a given subclaim determined the reference points for "Approached Expectations" and "Did Not Yet Meet or Partially Met Expectations" respectively.

Student performance for each subclaim is marked with a subclaim performance indicator.

- An 'up' arrow for the specified subclaim for "Meets or Exceeds Expectations" indicates that the student's performance for the subclaim was equal to or better than the threshold for Performance Level 4 (i.e., students whose summative scale score was 750).
- A 'bidirectional' arrow for the specified subclaim for "Nearly Meets Expectations" indicates that the student's performance was below the Performance Level 4 threshold (i.e., summative scale score was 750) but better than or equal to the Performance Level 3 threshold (i.e., summative scale score was 725).
- A 'down' arrow for the specified subclaim for "Below Expectations" indicates that the student's performance for the subclaim was below the Performance Level 3 threshold (i.e., summative scale score was 725).

Section 13: Quality Control Procedures

Quality control in a testing program is a comprehensive and ongoing process. This section describes procedures put into place to monitor the quality of the item bank, test form, and ancillary material development. The quality checks for scanning, image editing, scoring, and data screening during psychometric analyses are also outlined. Additional quality information can be found in the PARCC Program Quality Plan document.

13.1 Quality Control of the Item Bank

The PARCC summative item bank consists of test passages and items, their associated metadata, and status (e.g., operational-ready, field-test ready, released, etc.). The items on the 2015-2016 assessments were developed by Pearson and West Ed and put in the item bank once created.

In 2015, the PARCC summative bank underwent a conversion from the existing repository to the Assessment Banking for Building and Interoperability (ABBI) system. The ABBI bank houses the passages and items, art, associated metadata, rubrics, alternate text for use on accommodated forms, and text complexity documentation. It provides an item previewer that allows items to be viewed and interacted with in the same way students see and interact with items and tools, and manages versioning of items with a date/time stamp. It allows PARCC reviewers to vote on item acceptance, and to record and retain their review notes for later reconciliation and reference. As part of the bank transition, quality processes were conducted to ensure that the content of the passages and items, and the underlying QTI structure associated with those items, remained consistent from the old bank to the new bank. A validation of scoring and metadata was conducted. Mathematics rubrics were loaded and the versions validated.

In 2015, the bank transition occurred after initial development, but prior to the PARCC editorial review. PARCC editorial review committee participants conducted their review in the item banking system. As with all reviews, the committee members viewed the items as the student would, and could vote to alter the item, accept the item or reject the item and record their comments in the system. After each meeting, reports were forwarded to Parcc Inc. The reports were generated by the item banking system and summarized feedback from the committee reviewers.

All new development for the PARCC assessments is being created within the ABBI system, which employs templates to control the consistency of the underlying scoring logic and QTI creation for each item type. The ABBI system incorporates a previewer that allows the PARCC reviewers to validate the content of the item and validate the expected scoring of tasks. It supports the full range of PARCC review activities, including content review, bias and sensitivity review, expert editorial review, data review, and test construction review. It provides insight into the item edit process through versioning. A series of metadata validations at key points in the development cycle provide support for metadata consistency. The bank can be queried on the full range of PARCC metadata values to support bank analysis.

13.2 Quality Control of Test Form Development

Test Forms were built based upon targets and the established blueprints set by PARCC. The construction process started with specification and requirement capture to create the Test Specification Document. From there items were pulled into forms based on the criteria approved in the Test Specifications document. Quality control steps were conducted on the items and forms evaluating several item characteristics (e.g., content accuracy, completeness, style guides conformity, tools function). After forms composition, the forms went through a review process that involved groups from Pearson, Parcc Inc., and the PARCC states. Revisions were incorporated into the forms before final review and approval. Section 2.2 provides more details on the form development process.

The forms quality assurance was performed by Pearson's Assessment and Information Quality (AIQ) organization. AIQ completed a comprehensive review of all *online* forms for the PARCC administration cycle. This group is part of Pearson's larger Organizational Quality group and operates exclusively to validate form operability. The group validates that the functionality of every online form is working to specifications. The overall functionality and maneuverability of each form is checked, and the behavior of each item within the form is verified. (Quality processes for paper forms are described in section 13.3).

The items within each form were tested to verify that they operated as expected for test takers. As a further aspect of the testing process, AIQ confirmed that forms were loaded correctly and that the audio was correct when compared to text. Sections, seal codes, and overviews were reviewed. Technology enhanced items also were tested as an additional measure. As enumerated in the Technology Guidelines for PARCC Assessments,²² user interfaces were compatible with a range of common computer devices, operating systems, and browsers.

Pearson also performed QC tests to verify that a standard set of responses was outputted to the XML as expected after PARCC had approved the final version of the form. These responses were based on the keys provided in the test map or a standard open-ended (OE) responses string that contained a valid range of characters. The test maps also were validated against the form layout and item types for correctness as part of these tests.

Pearson conducted a multifaceted validation of all item layout, rendering, and functionality. Reviewers conducted comparisons between the approved item and the item as it appeared in the field test form, validated that tools and functions in the test delivery system, TestNav, were accurately applied, and verified that the style and layout met requirements documented in the PARCC Publishing Style Guide as

²² This document is available online from:

http://www.parcconline.org/files/72/Technology%20Guidelines%20for%20PARCC%20Assessments/389/TechnologyGuidelinesPARCCAssessments-v5_0-August2015.pdf

part of the PARCC Item Development Technical Guide.²³ In addition, all answer keys were validated through a formal key review process. More details on the Test Development procedures are provided in Section 2.

13.3 Quality Control of Test Materials

Pearson provided high quality materials in a timely and efficient manner to meet PARCC's test administration needs. Since the majority of printing work was done in-house it was possible to fully control the production environment, press schedule, and quality process for print materials. Additionally, strict security requirements were employed to protect secure materials production; Section 3 provides details on the secure handling of test materials. Materials were produced according to the PARCC Style Guide and to the detailed specifications supplied in the Materials List.

Pearson Print Service operates within the sanctions of an ISO 9001:2008 Quality Management System, and practices process improvement through Lean principles and employee involvement.

Raw materials (paper and ink) used for scannable forms production were manufactured exclusively for Pearson Print Service using specifications created by Pearson Print Service. Samples of ink and paper were tested by Pearson prior to use in production. Project Specialists were the point of contact for incoming production.

Purchase orders and other order information were assessed against manufacturing capabilities and assigned to the optimal production methodology. PARCC expectations, quality requirements, and cost considerations were foremost in these decisions. Prior to release for manufacture, order information was checked against PARCC specifications, technical requirements, and other communication that includes expected outcomes. Records of these checks were maintained.

Files for image creation flow through one of two file preparation functions: Digital pre-press (DPP) for digital print methodology, or plateroom for offset print methodology. Both the DPP and plateroom functions verify content, file naming, imposition, pagination, numbering stream, registration of technical components, color mapping, workflow, and file integrity. Records of these checks are created and saved.

Offset production requires printing that uses a lithographic process. Offline finishing activities are required to create books and package offset output. Digital output may flow through an inkjet Digital Production Line (DPL) or a sheet-fed toner application process in the Xpress Center. A battery of quality checks was performed in these areas. The checks included color match, correct file selection, content match to proof, litho-code to serial number synchronization, registration of technical components, ink density controlled by densitometry, inspection for print flaws, perforations, punching, pagination, scanning requirements, and any unique features specified for the order. Records of these checks and

²³ The PARCC Item Development Technical Guide is available online from:
<http://parccinc.org/wp-content/uploads/2014/07/PARCCItemDevelopmentTechnicalGuidePUBLICDRAFTFORRELEASE-20130912.pdf>

samples pulled from planned production points were maintained. Offline finishing included cutting, shrink wrapping, folding, and collating. The collation process has three robust inline detection systems that inspected each book for:

- Caliper validation that detects too few or too many pages. This detector will stop the collator if an incorrect caliper reading is registered.
- An optical reader that will only accept one sheet. Two or zero sheets will result in a collator stoppage.
- The correct bar code for the signature being assembled. An incorrect or upside down signature will be rejected by the bar code scanner and will result in a collator stoppage.

Pearson's Quality Assurance (QA) department personnel inspected print output prior to collation and shipment. QA also supported process improvement, work area documentation, audited process adherence, and established training programs for employees.

13.4 Quality Control of Scanning

Establishing and maintaining the accuracy of scanning, editing, and imaging processes is a cornerstone of the Pearson scoring process. While the scanners are designed to perform with great precision, Pearson implements other quality assurance processes to confirm that the data captured from scan processing produce a complete and accurate map to the expected results.

Pearson pioneered optical mark reading (OMR) and image scanning, and continues to improve in-house scanners for this purpose. Software programs drive the capture of student demographic data and student responses from the test materials during scan processing. Routinely scheduled maintenance and adjustments to the scanner components (e.g., camera) maintain scanner calibration. Test sheets inserted into every batch test scanner accuracy and calibration.

Controlled processes for developing and testing software specifications included a series of validation and verification procedures to confirm the captured data can be mapped accurately and completely to the expected results and that editing application rules are properly applied.

13.5 Quality Control of Image Editing

The final step in producing accurate data for scoring is the editing process. Once information from the documents was captured in the scanning process, the scan program file was executed, comparing the data captured from the student documents to the project specifications. The result of the comparison was a report (or edit listing) of documents needing corrections or validation. Image Editing Services performed the tasks necessary to correct and verify the student data prior to scoring.

Using the report, editors verified that all unscanned documents were scanned, or the data were imported into the system through some other method such as flatbed scan or key entry.

Documents with missing or suspect data were pulled, verified, and corrections or additional data were entered. Standard edits included:

- Incorrect or double gridding
- Incorrect dates (including birth year)
- Mismatches between pre-ID label and gridded information
- Incomplete names

When all edits were resolved, corrections were incorporated into the document file containing student records.

Additional quality checks were also performed. These included student *N*-count checks to make certain:

- Students were placed under the correct header
- All sheets belonged to the appropriate document
- Documents were not scanned twice
- No blank documents existed

Finally, accuracy checks were performed by checking random documents against scanned data to verify the accuracy of the scanning process.

Once all corrections were made, the scan program was tested a second time to verify all data were valid. When the resulting output showed that no fields were flagged as suspect, the file was considered clean and scoring began. Once all scanning was completed, the right/wrong response data were securely handed off.

13.6 Quality Control of Answer Document Processing and Scoring

Quality control of answer document processing and scoring involves all aspects of the scoring procedures, including key-based and rule-based machine scoring and handscoring for constructed response items and performance tasks.

For the 2015 PARCC operational administration, Pearson's validation team prepared test plans used throughout the scoring process. Test plan preparation was organized around detailed specifications.

Based on lessons learned from previous administrations the following quality steps were implemented:

- Raw score validation (e.g., score key validation; evidence statement, field test non-score; double-grid combinations; possible correct combination, if applicable; out-of-range/negative test cases);
- Matching (e.g., validation of high-confidence criteria, low-confidence criteria, cross document, external or forced matching by customer; prior to and after data updates; extract file of matched and unmatched documents);
- Demographic update tests (e.g., verification of data extract against corresponding layout; valid values for updatable fields; invalid values for updatable/non-updatable fields; negative test for non-existing record or empty file).

The following components were added to the quality control process specifically for the PARCC program. These additional steps were introduced to address issues with item-level scoring that were identified in the 2014 PARCC field test administration:

- XML Validation: A combination of automated validation against 100% of item XMLs and human inspection of XML from selected difficult item types or composite items.
- Administration/End-to-End Data Validation: An automated generation of response data from approved test maps that have known conditions were executed against the operational scoring systems and data generation systems to verify scoring accuracy.
- Psychometric Validation: Verification of data integrity using criteria typically used in psychometric processes (e.g., statistical keychecks) and categorization of identified issues to help inform investigation by other groups
- Content Validation: An examination, by subject matter experts, of all items using a combination of automated tools to generate response and scoring data.

In addition to the steps described above, the following quality control process for answer keys and scoring that was implemented for the first PARCC operational administration were used:

1. Pearson's psychometrics team conducted empirical analyses based on preliminary data files and flagged items based on statistical criteria;
2. Pearson content team reviewed the flagged items and provided feedback on the accuracy of content, answer keys, and scoring;
3. Items potentially requiring changes were added to the product validation (PV) log for further investigation by other Pearson teams;
4. PARCC staff was notified of items for which keys or scoring changes were recommended;
5. PARCC approved/rejected scoring changes; and
6. All approved scoring changes were implemented and validated prior to the generation of the data files used for psychometric processing.

13.7 Quality Control of Psychometric Processes

High quality psychometric work for the 2015–2016 PARCC operational administrations was necessary to provide accurate and reliable results of student performance. Pearson, HumRRO, and Measured Progress implemented quality control procedures to ensure the quality of the work including:

- 1) Well-defined psychometric specifications
- 2) Consistently applied data cleaning rules
- 3) Clear and frequent communication
- 4) Test run analyses
- 5) Quality checks of the analyses
- 6) Checklists for statistical procedures

13.7.1 Pearson Psychometric Quality Control Process

Pearson was responsible for the psychometric analyses of the 2015-2016 PARCC operational administration and implemented measures to ensure the quality of work. The psychometric analyses were all conducted according to well-defined specifications. Data cleaning rules were clearly articulated and applied consistently throughout the process. Results from all analyses underwent comprehensive

quality checks by a team of psychometricians and data analysts. Detailed checklists were used by members of the team for each statistical procedure.

Described below is an overview of the quality control steps performed at different stages of the psychometric analyses. Greater detail is provided in Sections 6 (Classical Item Analysis), 7 (Differential Item Functioning), 10 (IRT Calibration and Scaling), and 12 (Scale Scores).

Data Screening

Data screening is an important first step to ensure quality data input for meaningful analysis. The Pearson Customer Data Quality (CDQ) team validated all student data files used in the operational psychometric analyses. The data validation for the student data files (SDF) and item response files (IRF) included the following steps:

1. Validated variables in the data file for values in acceptable ranges.
2. Validated that the test form ID, unique item numbers (UINs) and item sequence on the data file were consistent with the test form values on the corresponding Test Map
3. Computed the composite raw score, claim raw scores, and subclaim raw scores, given the item scores in the student data file.
4. Compared computed raw scores to the raw scores in the student data file.
5. Compared the student item response block (SIRB) to the item scores.
6. Flagged student records with inconsistencies for further investigation.

Pearson Psychometrics and HumRRO established predefined valid case criteria, which were implemented consistently throughout the process. Refer to Section 5.3 for rules for inclusion of students in analyses and Section 10.2 for IRT calibration data preparation criteria and procedures.

Classical Item Analysis

Classical item analysis (IA) produces item level statistics (e.g., item difficulty and item-total correlations). The IA results were reviewed by Pearson psychometricians. Items flagged for unusual statistical properties were reviewed by the content team. A subset of items identified as having key issues, scoring issues, or content issues was presented to the PARCC Priority Alert Task Force, which made decisions on whether to exclude them from the IRT calibration process and, consequently, the calculation of reported student scores. Refer to Section 6.4 for classical IA item flagging criteria.

Calibrations

Creation of item response theory (IRT) sparse data matrices is an important step before the calibrations can begin. Using the same scored item response data, Pearson and HumRRO teams filtered the data and generated their own sparse data matrices independently. Processing of all data was done in parallel by two psychometricians and compared for n-counts. This verification of the data preparation was important to ensure that student exclusion rules were applied consistently across the analyses.

During the calibration process, checks were made to ensure that the correct options for the analyses were selected. Checks were also made on the number of items, number of test takers with valid scores,

IRT item difficulties, standard errors for the item difficulties, and the consistency between selected IRT statistics to the corresponding statistics obtained during item analyses. Psychometricians also performed detailed reviews of statistics to investigate the extent to which the assumptions of the model fit the observed data. Refer to Section 10.4 for IRT model fit evaluation criteria.

Scaling

During the scaling process, checks were made on the number of linking items, the number of items that were excluded from linking during the stability check of the scaling process, and the scaling constants. Linking items that did not meet the anchor criteria were excluded as linking items. For example, C-DIF items flagged in the mode comparability study were dropped. Additionally, items with large weighted root mean square difference (WRMSD) values in Round 1 of scaling were excluded as linking items in Round 2. Finally, reviewers computed the linking constants and then checked that the linking constants were correctly applied. Refer to Section 10.6 for description of scaling process.

Conversion Tables

Conversion tables must be accurate because they are used to generate reported scores for test takers. Comprehensive records were meticulously maintained on item-level decisions, and thorough checks were made to ensure that the correct items were included in the final score. All conversion tables were processed in parallel by Pearson and HumRRO and completely matched. A reasonableness check was also conducted by psychometricians for each content and grade level to make sure the results were in alignment with observations during the analyses prior to conversion table creation. Refer to Section 12.3 for the procedure to create conversion tables.

Delivering Item Statistics

Item statistics based on classical item analyses and IRT analyses were obtained during the psychometric analysis process. The statistics were compiled by two data analysts independently to ensure that the correct statistics were delivered for the item bank.

13.7.2 HumRRO Psychometric Quality Control Process

HumRRO served as the psychometric replicator for the 2015-16 PARCC operational administration. HumRRO replicated the IRT analyses, scaling analyses, and the conversion file creations. The following steps outline the replication process:

1. Calibrated online and paper data separately.
2. Scaled the 2015 item parameter estimates to the 2016 online scale.
3. Scaled the paper item parameter estimates to the online scale.
4. Sent the item parameter estimates and scaling constants to Measured Progress for comparison.
5. Reconciled differences, if any, in results with Pearson and Measured Progress.
6. Generated the scale score scaling constants based on the 2016 online scale.
7. Sent data files to Measured Progress for comparison and reconciled differences, if any.
8. Generated the performance levels, summative, claim, and subclaim conversion tables.
9. Sent conversion tables to Measured Progress for comparison and reconciled differences, if any.

13.7.3 Measured Progress Psychometric Quality Control Process

Measured Progress (MP) served as the external evaluator for the 2015-16 PARCC operational administration. MP's main task was to evaluate the reasonableness of IRT calibration results, and to compare to the IRT calibration results, scaling constants, summative, claim and subclaim conversion tables created by HumRRO and Pearson.

IRT Calibrations Comparison

MP reviewed and compared the psychometric IRT calibrations performed primarily by Pearson and HumRRO for all grade levels in ELA/L and mathematics administered both online and on paper.

Pearson and HumRRO each provided comparison files containing IRT item parameter estimates, IRT model fit statistics, and classical item statistics (item-level mean score, item-total correlation). Pearson also provided the IRT model fit plots for the items. For each test, the reasonableness of IRT parameters and the comparability of IRT parameter estimates between Pearson and HumRRO were evaluated on the following aspects:

- Number of items and types of interventions in the IRT calibration process
- Descriptive statistics of the IRT a -, b -, and d -parameter estimates
- Scatterplot of IRT a -, b -, and d -parameter estimates
- Absolute differences in IRT a -, b -, and d -parameter estimates
- Mean absolute difference (MAD) and root mean square difference (RMSD) in IRT-model-predicted item mean scores if there were nontrivial absolute differences in IRT parameter estimates
- IRT model fit statistics and plots
- Item parameter linking status for paper forms

IRT Comparison Results

In general, MP observed highly comparable IRT item parameter estimates between Pearson and HumRRO across all grades and subjects and in both online and paper forms in the 2015-2016 PARCC operational analyses. The largest differences in item parameter estimates occurred at the fourth decimal place. In general, model fit ranged from good to reasonable, with a few items illustrating more variability when sample sizes were small. For a very few items across all the tests, MP observed extreme IRT parameter estimates and/or standard errors, and sent the findings to Pearson for further investigation. Those items were sent to the Priority Alert Task Force for review, and they were either spoiled in operational scoring or flagged for re-field-testing with larger sample sizes.

Conversion Files Comparison

Measured Progress provided comparison results for the scaling constants, the summative, claim, and subclaim score conversion tables and their performance levels of both regular and accommodated forms in each of the Grades 3-8 and High School tests.

The conversion tables were evaluated and compared in the following aspects:

- Form ID and the number of total score points

- Minimum and maximum score points
- Raw cut-scores
- Theta and scaled scores associated with each raw score point

Conversion Files Comparison Results

MP observed identical lower and upper limits of scale scores for summative and subclaim performance levels in each test. In the final comparison files and after any observed differences were reconciled, MP observed identical form IDs and number of raw score points for each test. The largest differences in the theta points only showed on the fourth decimals across all the conversion tables. Only very few scale score points differed by one point due to rounding errors out of the numerous forms across the grades and subjects. None of the differences occurred on the scale cut-scores. Overall, the final conversion files provided by HumRRO and Pearson were highly comparable.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Beimers, J. N., Way, W. D., McClarty, K. L., & Miles, J. A. (2012, January). Evidence based standard setting: Establishing cut scores by integrating research evidence with expert content judgments. Austin, TX: Pearson. Retrieved from http://researchnetwork.pearson.com/wp-content/uploads/Bulletin21_Evidence_Based_Standard_Setting.pdf
- Brennan, R. L. (2004). BB-CLASS [Computer software]. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment (CASMA).
- Cai, L., Thissen, D., & du Toit, S. (2011). IRTPRO 2.1 [Computer software]. Lincolnwood, IL: Scientific Software International.
- Council of Chief State School Officers. (2015). Common Core State Standards Initiative. Retrieved from <http://www.corestandards.org/>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Doorey, N. & Polikoff, M. (2016). *Evaluating the Content and Quality of Next Generation Assessments*. Washington, DC: Thomas B. Fordham Institute.
- Dorans, N. J., & Schmitt, A. P. (1991). *Constructed response and differential item functioning: A pragmatic approach*. (ETS Research Report No. 91-47). Princeton, NJ: Educational Testing Service.
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Erlbaum.
- Dorans, N. J. (2013). *ETS Contributions to the Quantitative Assessment of Item, Test and Score Fairness* (ETS R&D Science and Policy Contributions Series, ETS SPC-13-04). Princeton, NJ: Educational Testing Service.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performances and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.
- Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika*, 55, 577-601.

- Huynh, Huynh & Meyer, Patrick (2010). Use of Robust z in Detecting Unstable Items in Item Response Theory Models. *Practical Assessment, Research & Evaluation*, 15(2). Available online: <http://pareonline.net/getvn.asp?v=15&n=2>
- Kim, S., & Kolen M. J. (2004). *STUIRT: A Computer Program for Scale Transformation under Unidimensional Item Response Theory Models*. University of Iowa. Version 1.0.
- Kolen, M.J., & Brennan, R.L. (2004). *Test equating, scaling, and linking: Methods and practices*. New York: Springer.
- Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, 58, 690-700.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- McClary, K. L., Kobrin, J. L., Moyer, E, Griffin, S., Huth, K., Carey, S., & Medberry, S. (2015). *PARCC Benchmarking Study*. Pearson Educational Measurement, Iowa City, IA: Pearson.
- McClarty, K. L., Way, W. D., Porter, A. C., Beimers, J. N., & Miles, J. A. (2013). Evidence-based standard setting: establishing a validity framework for cut scores. *Educational Researcher*, 42(2), 78-88.
- Muraki, E., & Bock R. D. (2003). *PARSCALE: IRT Item Analysis and Test Scoring for Rating-Scale Data*. Chicago, IL: Scientific Software International. Version 4.1.
- Liu, J., Brown, T., Chen, J., Ali, U., Hou, L., & Costanzo, K. (2016). *Mode Comparability Study Based on Spring 2015 Operational Test Data*. Princeton, NJ: Educational Testing Service.
- Olsson, U., Drasgow, F., & Dorans, N. J. (1982), "The Polyserial Correlation Coefficient," *Biometrika*, 47, 337–347.
- Plake, B. S., Ferdous, A. A., Impara, J. C., & Buckendahl, C. W. (2005). Setting Multiple Performance Standards Using the Yes/No Method: An Alternative Item Mapping Method. *Meeting of the National Council on Measurement in Education*. Montreal. Canada.
- Schultz, S. R., Michaels, H. R., Dvorak, R. N., & Wiley, C. R. H. (2016). *Evaluating the Content and Quality of Next Generation High School Assessments*. (HumRRO Report 2016 No. 001). Alexandria, VA: Human Resources Research Organization.
- Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28, 237-247.
- Steedle, J., McBride, M. Johnson, M., & Keng, L. (2015). *Spring 2015 Digital Devices Comparability Research Study*. Pearson Educational Measurement, Iowa City, IA: Pearson.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.

- Thacker, A., Sinclair, A., Wise, L., & Becker, S. (April 10, 2014). *PARCC validity studies including predictive and longitudinal studies memorandum* (HumRRO Report 2014 No. 020). Alexandria, VA: Human Resources Research Organization.
- Wainer, H., & Thissen, D. (2001). *Test scoring*. Mahwah, NJ: Erlbaum.
- Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practices*, 31, 2–13.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8, 125-145.
- Zenisky, A. L., Hambleton, R. K., & Sireci, S.C. (2003). *Effects of local dependence on the validity of IRT item test, and ability statistics*. (Technical Report). American College Admissions Test.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–348). Hillsdale, NJ: Erlbaum
- Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). *Describing and Categorizing DIF in Polytomous Items (ETS Research Report RR-97-05)*. Princeton, NJ: Educational Testing Service.

Appendices

Appendix 5: Test Takers by Grade and Mode, for Each State

Table A.5.1 ELA/L Test Takers, by State, and Grade

State	Category	Total	English Language Arts/Literacy								
			Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
PARCC	N of Students	3,339,882	471,801	461,204	455,980	455,888	449,801	440,160	275,158	192,956	136,934
PARCC	N of CBT	2,923,625	373,061	378,410	405,897	403,788	396,461	390,037	260,548	184,539	130,884
PARCC	% of CBT	87.5	79.1	82	89	88.6	88.1	88.6	94.7	95.6	95.6
PARCC	N of PBT	416,257	98,740	82,794	50,083	52,100	53,340	50,123	14,610	8417	6,050
PARCC	% of PBT	12.5	20.9	18	11	11.4	11.9	11.4	5.3	4.4	4.4
BIE	N of Students	8,808	1,480	1,381	1,338	1,273	1,085	1,100	318	320	513
BIE	% of PARCC Data	0.3	0	0	0	0	0	0	0	0	0
BIE	N of CBT	1,559	158	139	141	120	253	250	162	162	174
BIE	% of CBT	17.7	10.7	10.1	10.5	9.4	23.3	22.7	50.9	50.6	33.9
BIE	N of PBT	7,249	1,322	1,242	1,197	1,153	832	850	156	158	339
BIE	% of PBT	82.3	89.3	89.9	89.5	90.6	76.7	77.3	49.1	49.4	66.1
CO	N of Students	407,867	63,303	62,964	61,847	59,887	57,989	53,765	48,112	n/a	n/a
CO	% of PARCC Data	12.2	1.9	1.9	1.9	1.8	1.7	1.6	1.4	n/a	n/a
CO	N of CBT	388,232	58,884	60,224	59,248	57,059	55,396	51,311	46,110	n/a	n/a
CO	% of CBT	95.2	93	95.6	95.8	95.3	95.5	95.4	95.8	n/a	n/a
CO	N of PBT	19,635	4,419	2,740	2,599	2,828	2,593	2,454	2,002	n/a	n/a
CO	% of PBT	4.8	7	4.4	4.2	4.7	4.5	4.6	4.2	n/a	n/a
DC	N of Students	37,874	6,211	5,512	5,029	4,450	4,343	4,159	3,410	4,701	59
DC	% of PARCC Data	1.1	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0
DC	N of CBT	37,319	6,139	5,454	4,999	4,437	4,338	4,134	3,326	4,433	59
DC	% of CBT	98.5	98.8	98.9	99.4	99.7	99.9	99.4	97.5	94.3	100
DC	N of PBT	555	72	58	30	n/a	n/a	25	84	268	n/a
DC	% of PBT	1.5	1.2	1.1	0.6	0.3	0.1	0.6	2.5	5.7	n/a
IL	N of Students	100,7955	149,169	145,294	144,640	147,351	144,775	142,058	94,466	11,451	28,751
IL	% of PARCC Data	30.2	4.5	4.4	4.3	4.4	4.3	4.3	2.8	0.3	0.9
IL	N of CBT	854,366	98,492	101,913	132,896	134,920	132,182	133,232	84,522	10,492	25,717

State	Category	English Language Arts/Literacy									
		Total	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
IL	% of CBT	84.8	66	70.1	91.9	91.6	91.3	93.8	89.5	91.6	89.4
IL	N of PBT	153,589	50,677	43,381	11,744	12,431	12,593	8,826	9,944	959	3,034
IL	% of PBT	15.2	34	29.9	8.1	8.4	8.7	6.2	10.5	8.4	10.6
MA	N of Students	295,830	50,160	48,971	48,585	48,954	49,647	49,513	n/a	n/a	n/a
MA	% of PARCC Data	8.9	1.5	1.5	1.5	1.5	1.5	1.5	n/a	n/a	n/a
MA	N of CBT	119,362	18,729	18,742	19,534	21,236	20,740	20,381	n/a	n/a	n/a
MA	% of CBT	40.3	37.3	38.3	40.2	43.4	41.8	41.2	n/a	n/a	n/a
MA	N of PBT	176,468	31,431	30,229	29,051	27,718	28,907	29,132	n/a	n/a	n/a
MA	% of PBT	59.7	62.7	61.7	59.8	56.6	58.2	58.8	n/a	n/a	n/a
MD	N of Students	459,985	67,660	65,842	64,227	62,919	62,378	61,753	643	58,220	16,343
MD	% of PARCC Data	13.8	2	2	1.9	1.9	1.9	1.8	0	1.7	0.5
MD	N of CBT	422,623	59,881	62,999	61,382	57,330	56,821	55,988	613	53,356	14,253
MD	% of CBT	91.9	88.5	95.7	95.6	91.1	91.1	90.7	95.3	91.6	87.2
MD	N of PBT	37,362	7,779	2,843	2,845	5,589	5,557	5,765	30	4,864	2,090
MD	% of PBT	8.1	11.5	4.3	4.4	8.9	8.9	9.3	4.7	8.4	12.8
NJ	N of Students	828,566	98,899	96,740	95,694	96,772	95,837	94,102	94,074	86,398	70,050
NJ	% of PARCC Data	24.8	3	2.9	2.9	2.9	2.9	2.8	2.8	2.6	2.1
NJ	N of CBT	825,528	98,695	96,529	95,511	96,467	95,514	93,748	93,557	85,910	69,597
NJ	% of CBT	99.6	99.8	99.8	99.8	99.7	99.7	99.6	99.5	99.4	99.4
NJ	N of PBT	3,038	204	211	183	305	323	354	517	488	453
NJ	% of PBT	0.4	0.2	0.2	0.2	0.3	0.3	0.4	0.5	0.6	0.6
NM	N of Students	211,318	24,252	23,990	24,242	23,853	23,596	23,402	24,022	22,828	21,133
NM	% of PARCC Data	6.3	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.6
NM	N of CBT	208,432	23,751	23,756	23,974	23,725	22,884	22,749	23,878	22,716	20,999
NM	% of CBT	98.6	97.9	99	98.9	99.5	97	97.2	99.4	99.5	99.4
NM	N of PBT	2,886	501	234	268	128	712	653	144	112	134
NM	% of PBT	1.4	2.1	1	1.1	0.5	3	2.8	0.6	0.5	0.6
RI	N of Students	81,554	10,660	10,500	10,373	10,423	10,143	10,301	10,113	9,009	32
RI	% of PARCC Data	2.4	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0
RI	N of CBT	66,079	8,325	8,644	8,207	8,488	8,325	8,237	8,380	7,441	32

State	Category	English Language Arts/Literacy									
		Total	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11
RI	% of CBT	81	78.1	82.3	79.1	81.4	82.1	80	82.9	82.6	100
RI	N of PBT	15,475	2,335	1,856	2,166	1,935	1,818	2,064	1,733	1,568	n/a
RI	% of PBT	19	21.9	17.7	20.9	18.6	17.9	20	17.1	17.4	n/a

Note: CBT = computer-based test; PBT = paper-based test; n/a = not applicable.

Table A.5.2 Mathematics Test Takers, by State, and Grade

State	Category	Mathematics												
		Total	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	A1	GO	A2	M1	M2	M3
PARCC	N of Students	3,284,448	476,620	464,485	458,218	457,815	435,545	359,231	323,701	145,270	139,956	16,581	4,655	2,371
PARCC	N of CBT	286,7238	377,172	380,072	406,685	405,870	383,442	314,746	304,078	140,056	132,009	16,492	4,387	2,229
PARCC	% of CBT	87.3	79.1	81.8	88.8	88.7	88	87.6	93.9	96.4	94.3	99.5	94.2	94
PARCC	N of PBT	417,210	99,448	84,413	51,533	51,945	52,103	44,485	19,623	5,214	7,947	89	268	142
PARCC	% of PBT	12.7	20.9	18.2	11.2	11.3	12	12.4	6.1	3.6	5.7	0.5	5.8	6
BIE	N of Students	8,734	1,479	1,368	1,336	1,272	1,072	1,063	281	334	281	29	206	n/a
BIE	% of PARCC Data	0.3	0	0	0	0	0	0	0	0	0	0	0	0
BIE	N of CBT	1,528	158	139	141	120	245	215	156	181	171	n/a	n/a	n/a
BIE	% of CBT	17.5	10.7	10.2	10.6	9.4	22.9	20.2	55.5	54.2	60.9	6.9	n/a	n/a
BIE	N of PBT	7,206	1,321	1,229	1,195	1,152	827	848	125	153	110	27	206	n/a
BIE	% of PBT	82.5	89.3	89.8	89.4	90.6	77.1	79.8	44.5	45.8	39.1	93.1	100	100
CO	N of Students	410,385	64,932	63,510	62,044	60,232	55,489	41,195	39,100	10,722	2,373	9,240	1,379	169
CO	% of PARCC Data	12.5	2	1.9	1.9	1.8	1.7	1.3	1.2	0.3	0.1	0.3	0	0
CO	N of CBT	382,434	59,967	59,519	58,136	56,119	51,726	38,106	36,082	9,856	2,163	9,212	1,379	169
CO	% of CBT	93.2	92.4	93.7	93.7	93.2	93.2	92.5	92.3	91.9	91.2	99.7	100	100
CO	N of PBT	27,951	4,965	3,991	3,908	4,113	3,763	3,089	3,018	866	210	28	n/a	n/a
CO	% of PBT	6.8	7.6	6.3	6.3	6.8	6.8	7.5	7.7	8.1	8.8	0.3	n/a	n/a
DC	N of Students	38,146	6,236	5,514	5,057	4,484	4,149	3,273	3,607	5,546	211	n/a	69	n/a
DC	% of PARCC Data	1.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.2	0	n/a	0	n/a
DC	N of CBT	37,594	6,163	5,459	5,027	4,471	4,144	3,248	3,523	5,279	211	n/a	69	n/a
DC	% of CBT	98.6	98.8	99	99.4	99.7	99.9	99.2	97.7	95.2	100	n/a	100	n/a
DC	N of PBT	552	73	55	30	n/a	n/a	25	84	267	n/a	n/a	n/a	n/a
DC	% of PBT	1.4	1.2	1	0.6	0.3	0.1	0.8	2.3	4.8	n/a	n/a	n/a	n/a
IL	N of Students	983,476	149,406	145,580	144,878	147,622	144,303	138,380	70,301	8,810	23,563	6,699	2,334	1,600
IL	% of PARCC Data	29.9	4.5	4.4	4.4	4.5	4.4	4.2	2.1	0.3	0.7	0.2	0.1	0
IL	N of CBT	831,095	98,593	102,045	133,105	135,155	131,657	129,610	63,385	7,991	19,141	6,668	2,274	1,471
IL	% of CBT	84.5	66	70.1	91.9	91.6	91.2	93.7	90.2	90.7	81.2	99.5	97.4	91.9
IL	N of PBT	152,381	50,813	43,535	11,773	12,467	12,646	8,770	6,916	819	4,422	31	60	129
IL	% of PBT	15.5	34	29.9	8.1	8.4	8.8	6.3	9.8	9.3	18.8	0.5	2.6	8.1

State	Category	Total	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Mathematics					
									A1	GO	A2	M1	M2	M3
MA	N of Students	296,763	50,408	49,230	48,759	49,079	49,752	44,367	5,168	n/a	n/a	n/a	n/a	n/a
MA	% of PARCC Data	9	1.5	1.5	1.5	1.5	1.5	1.4	0.2	n/a	n/a	n/a	n/a	n/a
MA	N of CBT	119,627	18,819	18,813	19,556	21,292	20,787	17,417	2,943	n/a	n/a	n/a	n/a	n/a
MA	% of CBT	40.3	37.3	38.2	40.1	43.4	41.8	39.3	56.9	n/a	n/a	n/a	n/a	n/a
MA	N of PBT	177,136	31,589	30,417	29,203	27,787	28,965	26,950	2,225	n/a	n/a	n/a	n/a	n/a
MA	% of PBT	59.7	62.7	61.8	59.9	56.6	58.2	60.7	43.1	n/a	n/a	n/a	n/a	n/a
MD	N of Students	448,656	68,084	66,260	64,677	63,169	53,779	43,368	63,591	5,588	20,140	n/a	n/a	n/a
MD	% of PARCC Data	13.7	2.1	2	2	1.9	1.6	1.3	1.9	0.2	0.6	n/a	n/a	n/a
MD	N of CBT	418,268	60,510	63,413	61,916	59,192	50,787	41,363	59,073	4,612	17,402	n/a	n/a	n/a
MD	% of CBT	93.2	88.9	95.7	95.7	93.7	94.4	95.4	92.9	82.5	86.4	n/a	n/a	n/a
MD	N of PBT	30,388	7,574	2,847	2,761	3,977	2,992	2,005	4,518	976	2,738	n/a	n/a	n/a
MD	% of PBT	6.8	11.1	4.3	4.3	6.3	5.6	4.6	7.1	17.5	13.6	n/a	n/a	n/a
NJ	N of Students	806,752	99,738	97,549	96,390	97,441	93,076	60,650	105,056	83,434	73,332	43	30	n/a
NJ	% of PARCC Data	24.6	3	3	2.9	3	2.8	1.8	3.2	2.5	2.2	0	0	0
NJ	N of CBT	803,623	99,520	97,308	96,186	97,121	92,739	60,329	104,423	82,943	72,968	43	30	13
NJ	% of CBT	99.6	99.8	99.8	99.8	99.7	99.6	99.5	99.4	99.4	99.5	100	100	100
NJ	N of PBT	3,129	218	241	204	320	337	321	633	491	364	n/a	n/a	n/a
NJ	% of PBT	0.4	0.2	0.2	0.2	0.3	0.4	0.5	0.6	0.6	0.5	n/a	n/a	n/a
NM	N of Students	211,379	25,555	24,841	24,590	23,982	23,703	19,227	25,771	21,879	20,049	569	637	576
NM	% of PARCC Data	6.4	0.8	0.8	0.7	0.7	0.7	0.6	0.8	0.7	0.6	0	0	0
NM	N of CBT	208,496	25,030	24,604	24,315	23,854	22,992	18,589	25,613	21,775	19,947	566	635	576
NM	% of CBT	98.6	97.9	99	98.9	99.5	97	96.7	99.4	99.5	99.5	99.5	99.7	100
NM	N of PBT	2,883	525	237	275	128	711	638	158	104	102	3	2	n/a
NM	% of PBT	1.4	2.1	1	1.1	0.5	3	3.3	0.6	0.5	0.5	0.5	0.3	n/a
RI	N of Students	80,030	10,775	10,623	10,482	10,528	10,214	7,701	10,742	8,957	n/a	n/a	n/a	n/a
RI	% of PARCC Data	2.4	0.3	0.3	0.3	0.3	0.3	0.2	0.3	0.3	0	0	n/a	n/a
RI	N of CBT	64,446	8,405	8,762	8,298	8,540	8,357	5,862	8,796	7,419	n/a	n/a	n/a	n/a
RI	% of CBT	80.5	78	82.5	79.2	81.1	81.8	76.1	81.9	82.8	85.7	100	n/a	n/a
RI	N of PBT	15,584	2,370	1,861	2,184	1,988	1,857	1,839	1,946	1,538	n/a	n/a	n/a	n/a
RI	% of PBT	19.5	22	17.5	20.8	18.9	18.2	23.9	18.1	17.2	14.3	n/a	n/a	n/a

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based test; PBT = paper-based test; n/a = not applicable.

Table A.5.3 Spanish-Language Mathematics Test Takers, by State, and Grade

State*	Category	Mathematics (Spanish-Language forms)												
		Total	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	A1	GO	A2	M1	M2	M3
PARCC	N of Students	25,344	5,924	3,490	2,743	2,306	2,325	2,119	3,556	1,813	911	84	68	n/a
PARCC	N of CBT	23,159	4,857	3,068	2,675	2,221	2,215	1,980	3,333	1,761	894	82	68	n/a
PARCC	% of CBT	91.4	82	87.9	97.5	96.3	95.3	93.4	93.7	97.1	98.1	97.6	100	100
PARCC	N of PBT	2,185	1,067	422	68	85	110	139	223	52	n/a	n/a	n/a	n/a
PARCC	% of PBT	8.6	18	12.1	2.5	3.7	4.7	6.6	6.3	2.9	1.9	2.4	n/a	n/a
BIE	N of Students	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
BIE	N of CBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
BIE	% of CBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
BIE	N of PBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
BIE	% of PBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
CO	N of Students	2,684	1,173	449	252	174	219	201	152	n/a	n/a	60	n/a	n/a
CO	% of PARCC Data	10.6	4.6	1.8	1	0.7	0.9	0.8	0.6	0	n/a	0.2	0	n/a
CO	N of CBT	2,538	1,101	438	244	158	201	185	147	n/a	n/a	60	n/a	n/a
CO	% of CBT	94.6	93.9	97.6	96.8	90.8	91.8	92	96.7	100	n/a	100	100	n/a
CO	N of PBT	146	72	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
CO	% of PBT	5.4	6.1	2.4	3.2	9.2	8.2	8	3.3	n/a	n/a	n/a	n/a	n/a
DC	N of Students	329	45	45	55	40	39	n/a	38	50	n/a	n/a	n/a	n/a
DC	% of PARCC Data	1.3	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.2	n/a	n/a	n/a	n/a
DC	N of CBT	329	45	45	55	40	39	n/a	38	50	n/a	n/a	n/a	n/a
DC	% of CBT	100	100	100	100	100	100	100	100	100	n/a	n/a	n/a	n/a
DC	N of PBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
DC	% of PBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
IL	N of Students	6,013	2,190	991	1,033	722	445	323	209	n/a	n/a	n/a	66	n/a
IL	% of PARCC Data	23.7	8.6	3.9	4.1	2.8	1.8	1.3	0.8	0	0.1	0.1	0.3	n/a
IL	N of CBT	4,633	1,264	619	1,008	707	427	311	208	n/a	n/a	n/a	66	n/a
IL	% of CBT	77	57.7	62.5	97.6	97.9	96	96.3	99.5	100	26.7	100	100	n/a
IL	N of PBT	1,380	926	372	25	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
IL	% of PBT	23	42.3	37.5	2.4	2.1	4	3.7	0.5	n/a	73.3	n/a	n/a	n/a

State *	Category	Mathematics (Spanish-Language forms)												
		Total	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	A1	GO	A2	M1	M2	M3
MA	N of Students	59	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
MA	% of PARCC Data	0.2	0	0	0	0.1	0	0.1	n/a	n/a	n/a	n/a	n/a	n/a
MA	N of CBT	57	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
MA	% of CBT	96.6	100	100	90.9	100	100	92.9	n/a	n/a	n/a	n/a	n/a	n/a
MA	N of PBT	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
MA	% of PBT	3.4	n/a	n/a	9.1	n/a	n/a	7.1	n/a	n/a	n/a	n/a	n/a	n/a
MD	N of Students	345	n/a	n/a	n/a	n/a	26	25	234	n/a	n/a	n/a	n/a	n/a
MD	% of PARCC Data	1.4	0	0.1	0	0.1	0.1	0.1	0.9	0	n/a	n/a	n/a	n/a
MD	N of CBT	218	n/a	n/a	n/a	n/a	26	25	107	n/a	n/a	n/a	n/a	n/a
MD	% of CBT	63.2	100	100	100	100	100	100	45.7	100	n/a	n/a	n/a	n/a
MD	N of PBT	127	n/a	n/a	n/a	n/a	n/a	n/a	127	n/a	n/a	n/a	n/a	n/a
MD	% of PBT	36.8	n/a	n/a	n/a	n/a	n/a	n/a	54.3	n/a	n/a	n/a	n/a	n/a
NJ	N of Students	11,129	1,121	1,047	947	1,045	1,219	1,185	2,369	1,401	795	n/a	n/a	n/a
NJ	% of PARCC Data	43.9	4.4	4.1	3.7	4.1	4.8	4.7	9.3	5.5	3.1	n/a	n/a	n/a
NJ	N of CBT	10,972	1,106	1,018	932	1,032	1,207	1,165	2,330	1,393	789	n/a	n/a	n/a
NJ	% of CBT	98.6	98.7	97.2	98.4	98.8	99	98.3	98.4	99.4	99.2	n/a	n/a	n/a
NJ	N of PBT	157	n/a	29	n/a	n/a	n/a	n/a	39	n/a	n/a	n/a	n/a	n/a
NJ	% of PBT	1.4	1.3	2.8	1.6	1.2	1	1.7	1.6	0.6	0.8	n/a	n/a	n/a
NM	N of Students	3,699	1,253	847	347	209	269	229	263	169	101	n/a	n/a	n/a
NM	% of PARCC Data	14.6	4.9	3.3	1.4	0.8	1.1	0.9	1	0.7	0.4	0	0	0
NM	N of CBT	3,670	1,234	846	343	208	269	227	263	169	101	n/a	n/a	n/a
NM	% of CBT	99.2	98.5	99.9	98.8	99.5	100	99.1	100	100	100	66.7	100	100
NM	N of PBT	29	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
NM	% of PBT	0.8	1.5	0.1	1.2	0.5	n/a	0.9	n/a	n/a	n/a	33.3	n/a	n/a
RI	N of Students	1,086	127	92	86	80	99	125	291	186	n/a	n/a	n/a	n/a
RI	% of PARCC Data	4.3	0.5	0.4	0.3	0.3	0.4	0.5	1.1	0.7	n/a	n/a	n/a	n/a
RI	N of CBT	742	92	83	71	40	37	37	240	142	n/a	n/a	n/a	n/a
RI	% of CBT	68.3	72.4	90.2	82.6	50	37.4	29.6	82.5	76.3	n/a	n/a	n/a	n/a
RI	N of PBT	344	35	n/a	n/a	40	62	88	51	44	n/a	n/a	n/a	n/a

State *	Category	Mathematics (Spanish-Language forms)												
		Total	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	A1	GO	A2	M1	M2	M3
RI	% of PBT	31.7	27.6	9.8	17.4	50	62.6	70.4	17.5	23.7	n/a	n/a	n/a	n/a

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based test; PBT = paper-based test; n/a = not applicable.

* No students in BIE tested in mathematics using Spanish-language forms.

Table A.5.4 All States Combined: ELA/L Test Takers by Grade, Mode, and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
3	All	471,801	231,217	49	240,584	51
3	CBT	373,061	182,783	49	190,278	51
3	PBT	98,740	48,434	49.1	50,306	50.9
4	All	461,204	226,107	49	235,097	51
4	CBT	378,410	185,441	49	192,969	51
4	PBT	82,794	40,666	49.1	42,128	50.9
5	All	455,980	223,111	48.9	232,869	51.1
5	CBT	405,897	198,749	49	207,148	51
5	PBT	50,083	24,362	48.6	25,721	51.4
6	All	455,888	222,704	48.9	233,184	51.1
6	CBT	403,788	197,362	48.9	206,426	51.1
6	PBT	52,100	25,342	48.6	26,758	51.4
7	All	449,801	219,732	48.9	230,069	51.1
7	CBT	396,461	193,704	48.9	202,757	51.1
7	PBT	53,340	26,028	48.8	27,312	51.2
8	All	440,160	214,660	48.8	225,500	51.2
8	CBT	390,037	190,121	48.7	199,916	51.3
8	PBT	50,123	24,539	49	25,584	51
9	All	275,158	134,144	48.8	141,014	51.2
9	CBT	260,548	126,927	48.7	133,621	51.3
9	PBT	14,610	7,217	49.4	7,393	50.6
10	All	192,956	94,310	48.9	98,646	51.1
10	CBT	184,539	90,364	49	94,175	51
10	PBT	8,417	3,946	46.9	4,471	53.1
11	All	136,934	66,006	48.2	70,928	51.8
11	CBT	130,884	63,185	48.3	67,699	51.7
11	PBT	6,050	2,821	46.6	3,229	53.4

Note: CBT = computer-based tests; PBT = paper-based tests.

Table A.5.5 All States Combined: All Mathematics Test Takers by Grade, Mode, and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
3	All	476,620	233,536	49	243,084	51
3	CBT	377,172	184,711	49	192,461	51
3	PBT	99,448	48,825	49.1	50,623	50.9
4	All	464,485	227,746	49	236,739	51
4	CBT	380,072	186,308	49	193,764	51
4	PBT	84,413	41,438	49.1	42,975	50.9
5	All	458,218	224,219	48.9	233,999	51.1
5	CBT	406,685	199,163	49	207,522	51
5	PBT	51,533	25,056	48.6	26,477	51.4
6	All	457,815	223,647	48.9	234,168	51.1
6	CBT	405,870	198,380	48.9	207,490	51.1
6	PBT	51,945	25,267	48.6	26,678	51.4
7	All	435,545	212,807	48.9	222,738	51.1
7	CBT	383,442	187,347	48.9	196,095	51.1
7	PBT	52,103	25,460	48.9	26,643	51.1
8	All	359,231	173,063	48.2	186,168	51.8
8	CBT	314,746	151,421	48.1	163,325	51.9
8	PBT	44,485	21,642	48.7	22,843	51.3
A1	All	323,701	156,671	48.4	167,030	51.6
A1	CBT	304,078	147,204	48.4	156,874	51.6
A1	PBT	19,623	9,467	48.2	10,156	51.8
GO	All	145,270	70,985	48.9	74,285	51.1
GO	CBT	140,056	68,538	48.9	71,518	51.1
GO	PBT	5,214	2,447	46.9	2,767	53.1
A2	All	139,956	70,612	50.5	69,344	49.5
A2	CBT	132,009	66,616	50.5	65,393	49.5
A2	PBT	7,947	3,996	50.3	3,951	49.7
M1	All	16,581	8,128	49	8,453	51
M1	CBT	16,492	8,095	49.1	8,397	50.9
M1	PBT	89	33	37.1	56	62.9
M2	All	4,655	2,384	51.2	2,271	48.8
M2	CBT	4,387	2,235	50.9	2,152	49.1
M2	PBT	268	149	55.6	119	44.4
M3	All	2,371	1,263	53.3	1,108	46.7
M3	CBT	2,229	1,191	53.4	1,038	46.6
M3	PBT	142	72	50.7	70	49.3

Note: Includes students taking English-language mathematics tests, students taking Spanish-language mathematics tests, and students taking accommodated forms. A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based test; PBT = paper-based test; and n/a = not applicable.

Table A.5.6 All States Combined: Spanish-Language Mathematics Test Takers, by Grade, Mode, and Gender

Grade	Mode	Valid Cases N	Gender			
			Female		Male	
			N	%	N	%
3	All	5,924	2,931	49.5	2,993	50.5
3	CBT	4,857	2,404	49.5	2,453	50.5
3	PBT	1,067	527	49.4	540	50.6
4	All	3,490	1,692	48.5	1,798	51.5
4	CBT	3,068	1,503	49	1,565	51
4	PBT	422	189	44.8	233	55.2
5	All	2,743	1,316	48	1,427	52
5	CBT	2,675	1,285	48	1,390	52
5	PBT	68	31	45.6	37	54.4
6	All	2,306	1,093	47.4	1,213	52.6
6	CBT	2,221	1,059	47.7	1,162	52.3
6	PBT	85	34	40	51	60
7	All	2,325	1,115	48	1,210	52
7	CBT	2,215	1,064	48	1,151	52
7	PBT	110	51	46.4	59	53.6
8	All	2,119	984	46.4	1,135	53.6
8	CBT	1,980	927	46.8	1,053	53.2
8	PBT	139	57	41	82	59
A1	All	3,556	1,598	44.9	1,958	55.1
A1	CBT	3,333	1,496	44.9	1,837	55.1
A1	PBT	223	102	45.7	121	54.3
GO	All	1,813	848	46.8	965	53.2
GO	CBT	1,761	826	46.9	935	53.1
GO	PBT	52	n/a	42.3	30	57.7
A2	All	911	482	52.9	429	47.1
A2	CBT	894	472	52.8	422	47.2
A2	PBT	n/a	n/a	58.8	n/a	41.2
M1	All	84	27	32.1	57	67.9
M1	CBT	82	26	31.7	56	68.3
M1	PBT	n/a	n/a	50	n/a	50
M2	All	68	38	55.9	30	44.1
M2	CBT	68	38	55.9	30	44.1
M3	All	n/a	n/a	40	n/a	60
M3	CBT	n/a	n/a	40	n/a	60

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III. CBT = computer-based test; PBT = paper-based test; and n/a = not applicable.

Table A.5.7 Demographic Information for Grade 3 ELA/L, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	48.5	99.7	45.4	56.6	53.8	37.4	47.0	41.5	72.5	53.5
SWD (%)	13.6	12.4	10.9	14.7	10.6	19.0	12.7	17.4	13.5	15.0
EL (%)	14.5	39.2	15.5	11.3	20.3	14.6	10.8	7.0	18.7	11.4
Male (%)	51.0	49.5	51.3	50.7	50.9	51.1	51.2	50.8	51.2	51.0
Female (%)	49.0	50.5	48.7	49.3	49.1	48.9	48.8	49.2	48.8	49.0
AmInd/ANat (%)	1.2	99.9	0.7	n/r	0.3	0.2	0.3	0.1	10.7	0.6
Asian (%)	5.7	n/r	3.0	1.3	4.7	6.2	6.4	10.1	1.0	3.0
Black/AA (%)	16.7	n/r	4.8	68.2	17.8	10.4	33.6	15.6	1.9	8.5
Hisp/Lat (%)	27.9	n/r	33.5	16.1	27.0	23.7	17.0	28.6	60.3	25.8
Wh/Caus (%)	44.9	n/r	53.2	11.7	46.6	55.4	38.3	43.1	23.8	56.9
NtvHawaii/Pacific (%)	0.2	n/r	0.3	n/r	0.1	0.1	0.2	0.3	0.2	n/r
Two or More (%)	3.4	n/r	4.5	1.9	3.4	3.7	4.3	2.1	1.9	4.9
Unknown (%)	0.1	n/r	0.0	0.6	0.1	0.3	n/r	0.1	0.2	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported.

Table A.5.8 Demographic Information for Grade 4 ELA/L, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	47.5	99.9	45.2	56.9	52.6	36.0	46.0	40.3	71.8	51.4
SWD (%)	15.2	15.2	11.8	16.1	12.8	20.7	14.5	18.9	14.8	14.6
EL (%)	9.1	51.3	14.0	8.3	10.2	11.2	5.5	4.1	13.8	9.0
Male (%)	51.0	47.0	51.1	50.5	50.8	50.9	51.3	51.0	50.6	51.9
Female (%)	49.0	53.0	48.9	49.5	49.2	49.1	48.7	49.0	49.4	48.1
AmInd/ANat (%)	1.2	99.9	0.7	n/r	0.3	0.2	0.3	0.1	10.7	0.6
Asian (%)	5.8	n/r	3.0	1.5	4.8	6.1	6.5	10.3	1.1	3.1
Black/AA (%)	16.1	n/r	4.6	68.9	16.9	10.2	33.6	15.3	1.9	7.9
Hisp/Lat (%)	27.6	n/r	33.8	16.8	26.8	23.3	16.2	27.6	60.4	25.6
Wh/Caus (%)	45.9	n/r	53.5	10.5	47.7	56.4	38.9	44.4	23.9	58.3
NtvHawaii/Pacific (%)	0.2	n/r	0.2	n/r	0.1	0.1	0.2	0.2	0.2	n/r
Two or More (%)	3.2	n/r	4.1	2.1	3.3	3.4	4.2	1.9	1.8	4.2
Unknown (%)	0.1	n/r	0.0	n/r	0.1	0.2	n/r	0.1	0.2	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table A.5.9 Demographic Information for Grade 5 ELA/L, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	46.4	99.8	45.2	56.8	51.3	34.3	44.4	39.0	72.2	50.0
SWD (%)	16.0	15.5	12.5	16.3	13.4	21.7	15.3	19.5	16.0	16.2
EL (%)	7.0	51.2	11.5	5.6	7.2	9.1	4.1	3.0	12.0	7.8
Male (%)	51.1	50.4	51.2	50.0	51.2	50.8	51.0	51.1	50.6	51.1
Female (%)	48.9	49.6	48.8	50.0	48.8	49.2	49.0	48.9	49.4	48.9
AmInd/ANat (%)	1.1	100.0	0.7	n/r	0.3	0.2	0.3	0.1	10.0	0.7
Asian (%)	5.9	n/r	3.3	1.3	4.9	6.0	6.8	10.3	1.1	3.0
Black/AA (%)	15.9	n/r	4.6	70.7	16.3	10.2	33.8	15.1	1.9	8.4
Hisp/Lat (%)	27.3	n/r	35.0	15.0	26.5	22.0	15.2	26.8	61.4	24.6
Wh/Caus (%)	46.5	n/r	52.2	10.7	48.5	57.8	39.7	45.7	23.6	59.0
NtvHawaii/Pacific (%)	0.2	n/r	0.2	n/r	0.1	0.1	0.2	0.2	0.1	n/r
Two or More (%)	3.1	n/r	3.9	2.1	3.2	3.5	4.0	1.7	1.8	4.1
Unknown (%)	0.1	n/r	0.0	n/r	0.1	0.2	n/r	0.0	0.2	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported.

Table A.5.10 Demographic Information for Grade 6 ELA/L, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	45.1	99.8	44.4	62.0	50.2	33.4	42.9	37.4	69.0	48.7
SWD (%)	16.1	13.9	12.5	19.5	13.8	22.3	15.5	19.3	15.4	16.9
EL (%)	5.8	48.2	10.2	5.0	5.3	7.5	3.3	2.6	11.8	5.7
Male (%)	51.1	49.6	51.3	50.2	51.1	51.1	51.0	51.4	50.8	51.4
Female (%)	48.9	50.4	48.7	49.8	48.9	48.9	49.0	48.6	49.2	48.6
AmInd/ANat (%)	1.1	99.9	0.8	n/r	0.2	0.3	0.3	0.1	10.2	0.8
Asian (%)	5.9	n/r	3.3	1.4	4.9	6.1	6.7	10.4	1.1	3.2
Black/AA (%)	15.9	n/r	4.5	72.1	16.6	10.2	34.0	15.0	1.9	8.2
Hisp/Lat (%)	26.6	n/r	34.8	14.6	26.0	21.4	14.3	25.6	61.7	23.7
Wh/Caus (%)	47.3	n/r	52.7	9.8	48.9	58.4	40.5	47.1	23.3	60.1
NtvHawaii/Pacific (%)	0.1	n/r	0.2	n/r	0.1	0.1	0.1	0.2	0.1	0.2
Two or More (%)	2.9	n/r	3.6	1.6	3.2	3.3	4.0	1.5	1.5	3.8
Unknown (%)	0.1	n/r	n/r	n/r	0.1	0.2	n/r	0.1	0.1	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported.

Table A.5.11 Demographic Information for Grade 7 ELA/L, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	44.0	99.8	43.5	60.2	49.2	31.9	42.0	36.3	68.3	47.4
SWD (%)	16.1	15.8	12.6	19.3	14.0	22.1	15.2	18.9	14.7	17.1
EL (%)	5.7	44.9	10.6	5.3	4.8	7.4	3.7	2.8	11.0	5.6
Male (%)	51.1	48.7	51.3	50.4	51.4	51.0	51.1	51.2	50.2	51.6
Female (%)	48.9	51.3	48.7	49.6	48.6	49.0	48.9	48.8	49.8	48.4
AmInd/ANat (%)	1.1	99.6	0.8	n/r	0.3	0.3	0.3	0.1	10.3	0.6
Asian (%)	6.0	n/r	3.4	1.8	4.8	6.1	6.9	10.4	1.2	2.8
Black/AA (%)	16.0	n/r	4.8	72.9	16.4	10.1	33.7	15.6	1.9	8.2
Hisp/Lat (%)	26.3	n/r	35.1	14.3	25.7	20.7	14.4	25.0	61.5	23.7
Wh/Caus (%)	47.7	n/r	52.1	9.1	49.6	59.5	40.7	47.4	23.4	61.3
NtvHawaii/Pacific (%)	0.2	n/r	0.3	n/r	0.1	0.1	0.1	0.2	0.1	n/r
Two or More (%)	2.8	n/r	3.6	1.5	3.0	3.0	4.0	1.2	1.5	3.2
Unknown (%)	0.1	n/r	n/r	n/r	0.1	0.2	n/r	0.1	0.1	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported.

Table A.5.12 Demographic Information for Grade 8 ELA/L, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	43.5	99.9	44.0	60.0	48.4	31.8	40.8	35.8	67.8	46.7
SWD (%)	15.9	15.9	12.1	20.5	13.8	21.7	15.2	18.7	14.0	17.0
EL (%)	5.8	45.6	11.4	5.6	4.8	7.1	3.8	2.9	11.2	5.7
Male (%)	51.2	46.0	52.0	50.4	51.1	50.5	51.1	51.4	51.2	52.8
Female (%)	48.8	54.0	48.0	49.6	48.9	49.5	48.9	48.6	48.8	47.2
AmInd/ANat (%)	1.1	100.0	0.8	n/r	0.3	0.2	0.3	0.1	10.4	0.7
Asian (%)	5.8	n/r	3.5	1.6	4.6	5.8	6.6	10.0	1.1	3.2
Black/AA (%)	16.2	n/r	5.0	72.3	16.5	10.5	34.0	15.6	2.1	7.9
Hisp/Lat (%)	25.9	n/r	35.5	15.5	25.2	20.4	13.9	24.5	61.7	23.7
Wh/Caus (%)	48.1	n/r	51.4	8.8	50.3	60.1	41.1	48.4	23.2	61.2
NtvHawaii/Pacific (%)	0.1	n/r	0.2	n/r	0.1	0.1	0.1	0.2	0.1	n/r
Two or More (%)	2.6	n/r	3.5	1.5	2.9	2.7	3.8	1.1	1.4	3.1
Unknown (%)	0.1	n/r	n/r	n/r	0.1	0.2	n/r	0.1	0.1	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table A.5.13 Demographic Information for Grade 9 ELA/L, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	42.9	99.7	42.9	48.3	44.1	n/r	49.5	35.9	62.3	46.4
SWD (%)	14.4	11.3	11.6	19.1	12.3	n/r	22.7	17.8	12.7	17.6
EL (%)	6.9	19.2	11.8	7.5	5.1	n/r	n/r	5.1	12.0	5.6
Male (%)	51.2	47.5	51.8	49.3	50.7	n/r	54.0	51.6	51.1	51.2
Female (%)	48.8	52.5	48.2	50.7	49.3	n/r	46.0	48.4	48.9	48.8
AmInd/ANat (%)	1.3	99.1	0.8	n/r	0.3	n/r	n/r	0.1	10.5	0.6
Asian (%)	6.2	n/r	3.4	1.9	5.8	n/r	n/r	9.9	1.3	3.0
Black/AA (%)	13.6	n/r	5.2	73.5	17.3	n/r	36.7	15.4	1.9	8.4
Hisp/Lat (%)	31.2	n/r	36.6	14.0	27.9	n/r	4.4	26.1	60.4	23.3
Wh/Caus (%)	45.3	n/r	50.3	7.3	45.9	n/r	51.3	47.2	24.4	61.3
NtvHawaii/Pacific (%)	0.2	n/r	0.2	n/r	0.1	n/r	n/r	0.2	0.1	n/r
Two or More (%)	2.1	n/r	3.4	1.0	2.6	n/r	6.2	1.0	1.4	3.2
Unknown (%)	0.1	n/r	n/r	2.1	0.1	n/r	n/r	0.1	n/r	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported due to n<20; n/a = not applicable.

Table A.5.14 Demographic Information for Grade 10 ELA/L, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	39.3	99.4	n/r	45.7	40.1	n/r	36.7	34.5	59.5	45.3
SWD (%)	15.6	15.3	n/r	19.5	10.1	n/r	14.2	17.9	12.0	16.2
EL (%)	5.1	10.9	n/r	6.9	2.3	n/r	5.0	4.5	8.6	5.4
Male (%)	51.1	53.4	n/r	49.3	51.4	n/r	50.8	51.5	50.6	51.8
Female (%)	48.9	46.6	n/r	50.7	48.6	n/r	49.2	48.5	49.4	48.2
AmInd/ANat (%)	1.6	99.1	n/r	n/r	0.2	n/r	0.3	0.1	10.8	0.6
Asian (%)	7.1	n/r	n/r	1.3	2.9	n/r	7.3	9.9	1.3	3.0
Black/AA (%)	20.6	n/r	n/r	74.5	10.6	n/r	35.2	15.4	2.2	8.1
Hisp/Lat (%)	25.4	n/r	n/r	15.1	21.1	n/r	13.9	25.7	59.7	22.8
Wh/Caus (%)	42.9	n/r	n/r	4.8	62.2	n/r	39.9	47.5	24.5	62.4
NtvHawaii/Pacific (%)	0.2	n/r	n/r	n/r	n/r	n/r	0.1	0.2	0.1	n/r
Two or More (%)	2.0	n/r	n/r	1.2	3.0	n/r	3.4	1.0	1.4	3.0
Unknown (%)	0.1	n/r	n/r	2.9	n/r	n/r	n/r	0.1	n/r	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table A.5.15 Demographic Information for Grade 11 ELA/L, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	41.8	99.8	n/r	67.8	41.5	n/r	42.4	36.5	57.5	71.9
SWD (%)	15.4	12.7	n/r	n/r	8.4	n/r	15.1	19.7	10.5	n/r
EL (%)	3.7	22.4	n/r	n/r	1.1	n/r	1.2	4.3	6.8	n/r
Male (%)	51.8	45.6	n/r	49.2	50.6	n/r	51.9	52.8	50.1	71.9
Female (%)	48.2	54.4	n/r	50.8	49.4	n/r	48.1	47.2	49.9	n/r
AmInd/ANat (%)	2.1	99.6	n/r	n/r	0.3	n/r	0.2	0.1	10.5	n/r
Asian (%)	4.9	n/r	n/r	n/r	1.3	n/r	2.2	8.2	1.5	n/r
Black/AA (%)	18.6	n/r	n/r	42.4	15.9	n/r	50.9	17.3	1.9	n/r
Hisp/Lat (%)	26.3	n/r	n/r	50.8	10.1	n/r	9.7	26.9	59.8	n/r
Wh/Caus (%)	46.1	n/r	n/r	n/r	69.3	n/r	34.0	46.4	24.7	n/r
NtvHawaii/Pacific (%)	0.2	n/r	n/r	n/r	0.1	n/r	0.1	0.2	0.1	n/r
Two or More (%)	1.6	n/r	n/r	n/r	3.0	n/r	2.9	0.8	1.5	n/r
Unknown (%)	0.1	n/r	n/r	n/r	n/r	n/r	n/r	0.1	n/r	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported due to n<20; n/a = not applicable.

Table A.5.16 Demographic Information for Grade 3 Mathematics, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	48.8	99.7	46.6	56.4	53.8	37.4	47.1	41.7	73.7	53.7
SWD (%)	13.5	12.4	10.9	14.5	10.5	18.9	12.6	17.3	13.3	14.9
EL (%)	15.4	39.3	17.5	11.9	20.6	14.9	11.4	7.7	22.7	12.4
Male (%)	51.0	49.6	51.3	50.6	50.9	51.1	51.2	50.8	51.2	51.1
Female (%)	49.0	50.4	48.7	49.4	49.1	48.9	48.8	49.2	48.8	48.9
AmInd/ANat (%)	1.1	99.9	0.6	n/r	0.3	0.2	0.3	0.1	10.1	0.6
Asian (%)	5.8	n/r	3.0	1.3	4.8	6.3	6.5	10.2	1.0	3.2
Black/AA (%)	16.5	n/r	4.7	67.8	17.7	10.4	33.5	15.5	1.8	8.6
Hisp/Lat (%)	28.4	n/r	35.1	16.3	27.0	23.8	17.2	28.9	62.3	26.2
Wh/Caus (%)	44.6	n/r	52.0	11.8	46.5	55.1	38.1	42.9	22.6	56.3
NtvHawaii/Pacific (%)	0.2	n/r	0.3	n/r	0.1	0.1	0.2	0.3	0.2	n/r
Two or More (%)	3.3	n/r	4.3	1.9	3.4	3.7	4.3	2.1	1.8	4.9
Unknown (%)	0.1	n/r	0.0	0.6	0.1	0.4	n/r	0.1	0.2	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported.

Table A.5.17 Demographic Information for Grade 4 Mathematics, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	47.7	99.9	45.6	56.5	52.6	36.0	46.1	40.5	72.7	51.7
SWD (%)	15.1	15.3	11.8	15.8	12.7	20.6	14.4	18.7	14.6	14.4
EL (%)	9.8	51.5	14.7	9.0	10.4	11.5	6.2	4.9	16.7	10.0
Male (%)	51.0	47.1	51.0	50.5	50.8	50.9	51.3	51.0	50.6	51.9
Female (%)	49.0	52.9	49.0	49.5	49.2	49.1	48.7	49.0	49.4	48.1
AmInd/ANat (%)	1.1	99.9	0.7	n/r	0.3	0.2	0.3	0.1	10.4	0.6
Asian (%)	5.9	n/r	3.0	1.5	4.9	6.2	6.6	10.4	1.0	3.3
Black/AA (%)	16.0	n/r	4.6	68.5	16.8	10.2	33.5	15.2	1.8	8.0
Hisp/Lat (%)	27.9	n/r	34.3	17.3	26.8	23.3	16.5	27.9	61.7	26.0
Wh/Caus (%)	45.6	n/r	53.0	10.4	47.7	56.2	38.7	44.2	23.1	57.8
NtvHawaii/Pacific (%)	0.2	n/r	0.2	n/r	0.1	0.1	0.2	0.2	0.2	n/r
Two or More (%)	3.2	n/r	4.1	2.1	3.3	3.4	4.2	1.9	1.7	4.2
Unknown (%)	0.1	n/r	0.0	n/r	0.1	0.4	n/r	0.1	0.1	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table A.5.18 Demographic Information for Grade 5 Mathematics, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	46.5	99.8	45.2	56.4	51.3	34.3	44.5	39.2	72.5	50.3
SWD (%)	15.9	15.6	12.6	16.3	13.4	21.5	15.2	19.4	15.9	16.0
EL (%)	7.5	51.3	11.6	6.3	7.4	9.4	4.8	3.8	13.1	8.8
Male (%)	51.1	50.4	51.2	50.0	51.2	50.8	51.0	51.1	50.7	51.1
Female (%)	48.9	49.6	48.8	50.0	48.8	49.2	49.0	48.9	49.3	48.9
AmInd/ANat (%)	1.1	100.0	0.7	n/r	0.3	0.2	0.3	0.1	9.8	0.7
Asian (%)	6.0	n/r	3.3	1.4	5.0	6.0	6.8	10.3	1.1	3.1
Black/AA (%)	15.8	n/r	4.6	70.0	16.3	10.2	33.6	15.0	1.9	8.5
Hisp/Lat (%)	27.4	n/r	35.0	15.6	26.5	22.1	15.5	27.1	61.8	25.1
Wh/Caus (%)	46.3	n/r	52.1	10.7	48.5	57.5	39.5	45.4	23.3	58.5
NtvHawaii/Pacific (%)	0.2	n/r	0.2	n/r	0.1	0.1	0.2	0.2	0.1	n/r
Two or More (%)	3.0	n/r	3.9	2.1	3.2	3.5	4.0	1.7	1.7	4.1
Unknown (%)	0.1	n/r	0.0	n/r	0.1	0.4	n/r	0.0	0.2	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported.

Table A.5.19 Demographic Information for Grade 6 Mathematics, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	45.2	99.8	44.4	61.3	50.2	33.4	43.2	37.7	69.3	49.0
SWD (%)	16.1	13.9	12.6	19.2	13.7	22.2	15.4	19.2	15.3	16.6
EL (%)	6.3	48.3	10.2	6.1	5.5	7.8	4.1	3.5	12.5	6.8
Male (%)	51.1	49.5	51.3	50.0	51.1	51.1	51.0	51.4	50.7	51.4
Female (%)	48.9	50.5	48.7	50.0	48.9	48.9	49.0	48.6	49.3	48.6
AmInd/ANat (%)	1.1	100.0	0.8	n/r	0.2	0.3	0.3	0.1	10.2	0.7
Asian (%)	6.0	n/r	3.3	1.5	5.0	6.1	6.7	10.4	1.1	3.3
Black/AA (%)	15.9	n/r	4.6	71.4	16.5	10.2	33.8	14.9	1.9	8.3
Hisp/Lat (%)	26.8	n/r	34.8	15.3	26.1	21.5	14.8	26.1	62.0	24.2
Wh/Caus (%)	47.1	n/r	52.7	9.8	48.9	58.2	40.3	46.8	23.0	59.5
NtvHawaii/Pacific (%)	0.1	n/r	0.2	n/r	0.1	0.1	0.1	0.2	0.1	0.2
Two or More (%)	2.9	n/r	3.6	1.7	3.2	3.3	4.0	1.5	1.5	3.7
Unknown (%)	0.1	n/r	n/r	n/r	0.1	0.4	n/r	0.1	0.1	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported.

Table A.5.20 Demographic Information for Grade 7 Mathematics, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	45.1	99.8	45.2	62.5	49.4	31.9	45.1	37.8	68.6	47.7
SWD (%)	16.5	15.9	13.0	19.8	14.0	22.0	17.0	19.4	14.5	16.9
EL (%)	6.4	45.1	11.1	6.6	5.1	7.7	5.3	3.8	12.0	6.6
Male (%)	51.1	48.6	51.0	50.2	51.4	51.0	51.3	51.1	50.3	51.6
Female (%)	48.9	51.4	49.0	49.8	48.6	49.0	48.7	48.9	49.7	48.4
AmInd/ANat (%)	1.1	99.7	0.8	n/r	0.3	0.3	0.3	0.1	10.2	0.6
Asian (%)	5.4	n/r	3.1	1.7	4.8	6.1	5.0	9.2	1.2	2.7
Black/AA (%)	15.9	n/r	4.9	75.1	16.4	10.1	34.5	16.0	1.9	8.3
Hisp/Lat (%)	27.2	n/r	36.5	15.4	25.8	20.9	15.6	26.1	62.1	24.2
Wh/Caus (%)	47.5	n/r	50.9	6.1	49.5	59.2	40.6	47.1	22.9	60.7
NtvHawaii/Pacific (%)	0.2	n/r	0.3	n/r	0.1	0.1	0.1	0.2	0.1	n/r
Two or More (%)	2.7	n/r	3.5	1.2	3.0	3.0	4.0	1.2	1.5	3.2
Unknown (%)	0.1	n/r	n/r	n/r	0.1	0.4	n/r	0.1	0.1	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported.

Table A.5.21 Demographic Information for Grade 8 Mathematics, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	48.1	99.9	51.7	67.0	48.5	33.6	48.9	44.1	72.9	54.1
SWD (%)	18.2	16.4	14.7	23.9	14.1	23.0	18.6	25.6	17.6	21.3
EL (%)	7.4	46.4	14.7	6.2	5.1	8.1	5.8	5.1	14.6	8.9
Male (%)	51.8	46.4	52.4	51.3	51.2	51.1	52.0	53.0	52.4	54.1
Female (%)	48.2	53.6	47.6	48.7	48.8	48.9	48.0	47.0	47.6	45.9
AmInd/ANat (%)	1.2	100.0	0.9	n/r	0.3	0.2	0.3	0.1	11.2	0.8
Asian (%)	4.4	n/r	2.7	1.2	4.6	5.6	4.9	5.3	0.7	2.8
Black/AA (%)	17.6	n/r	5.8	79.8	16.6	10.7	41.0	19.5	2.2	8.9
Hisp/Lat (%)	28.2	n/r	41.3	13.4	25.1	21.8	16.2	29.6	64.7	28.1
Wh/Caus (%)	45.7	n/r	45.7	4.2	50.4	58.5	33.9	44.2	19.9	55.8
NtvHawaii/Pacific (%)	0.1	n/r	0.2	n/r	0.1	0.1	0.1	0.2	n/r	n/r
Two or More (%)	2.6	n/r	3.3	1.1	2.9	2.7	3.6	1.0	1.2	3.3
Unknown (%)	0.1	n/r	n/r	n/r	0.1	0.4	n/r	0.1	n/r	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table A.5.22 Demographic Information for Algebra I, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	43.5	99.6	40.3	48.0	53.1	16.6	40.1	36.3	62.9	48.7
SWD (%)	15.1	9.3	11.8	19.4	14.2	8.6	15.0	17.5	13.4	15.9
EL (%)	7.5	15.7	10.0	8.9	8.5	1.3	7.0	5.2	12.6	8.3
Male (%)	51.6	46.3	52.3	50.8	51.7	46.4	51.9	51.5	51.4	51.7
Female (%)	48.4	53.7	47.7	49.2	48.3	53.6	48.1	48.5	48.6	48.3
AmInd/ANat (%)	1.2	99.3	0.7	n/r	0.3	n/r	0.3	0.1	10.6	0.7
Asian (%)	6.1	n/r	3.3	1.6	4.4	7.8	6.9	9.5	1.3	3.0
Black/AA (%)	18.3	n/r	4.5	70.5	21.0	8.4	34.8	15.3	2.0	9.1
Hisp/Lat (%)	28.9	n/r	33.6	16.5	32.8	9.0	15.5	26.6	60.6	25.1
Wh/Caus (%)	43.0	n/r	54.1	8.1	38.9	72.1	38.7	47.1	23.9	58.9
NtvHawaii/Pacific (%)	0.2	n/r	0.2	n/r	0.1	n/r	0.1	0.2	0.1	n/r
Two or More (%)	2.3	n/r	3.6	1.2	2.3	2.5	3.6	1.1	1.4	3.0
Unknown (%)	0.1	n/r	n/r	1.9	0.2	n/r	n/r	0.1	0.1	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table A.5.23 Demographic Information for Geometry, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	37.6	100.0	20.2	45.1	30.8	n/r	30.9	33.7	59.1	45.1
SWD (%)	14.7	14.4	3.6	17.7	8.9	n/r	15.5	17.4	11.0	14.7
EL (%)	5.0	8.4	2.3	7.4	1.4	n/r	4.4	4.3	9.0	6.8
Male (%)	51.1	51.5	50.5	48.4	50.8	n/r	53.8	51.4	50.7	51.2
Female (%)	48.9	48.5	49.5	51.6	49.2	n/r	46.2	48.6	49.3	48.8
AmInd/ANat (%)	1.9	99.4	0.5	n/r	0.3	n/r	0.4	0.1	10.2	0.4
Asian (%)	7.2	n/r	6.3	1.8	3.6	n/r	10.8	9.9	1.2	3.4
Black/AA (%)	13.8	n/r	2.2	69.0	8.7	n/r	23.8	15.2	2.2	8.5
Hisp/Lat (%)	28.2	n/r	18.4	17.0	12.4	n/r	10.7	25.3	60.1	23.7
Wh/Caus (%)	46.8	n/r	68.3	8.2	71.3	n/r	51.1	48.2	24.6	60.7
NtvHawaii/Pacific (%)	0.2	n/r	0.2	n/r	n/r	n/r	n/r	0.2	0.1	n/r
Two or More (%)	1.6	n/r	4.1	1.3	3.6	n/r	3.1	0.9	1.5	3.0
Unknown (%)	0.2	n/r	n/r	2.4	n/r	n/r	n/r	0.1	n/r	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported due to n<20; n/a = not applicable.

Table A.5.24 Demographic Information for Algebra II, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	36.9	100.0	9.8	32.7	38.8	n/r	33.2	32.9	55.2	n/r
SWD (%)	10.9	12.5	2.9	n/r	5.3	n/r	9.0	14.3	8.1	n/r
EL (%)	3.1	11.7	1.0	n/r	1.3	n/r	1.2	3.4	6.0	n/r
Male (%)	49.5	42.3	53.4	48.3	48.3	n/r	47.7	50.5	49.1	n/r
Female (%)	50.5	57.7	46.6	51.7	51.7	n/r	52.3	49.5	50.9	n/r
AmInd/ANat (%)	1.7	100.0	n/r	n/r	0.3	n/r	0.4	0.1	9.6	n/r
Asian (%)	7.6	n/r	8.1	n/r	2.0	n/r	6.8	11.2	1.7	n/r
Black/AA (%)	16.7	n/r	n/r	21.3	16.5	n/r	39.6	15.1	1.8	n/r
Hisp/Lat (%)	24.8	n/r	8.9	24.6	11.3	n/r	8.4	24.9	59.0	n/r
Wh/Caus (%)	47.3	n/r	77.1	45.5	66.6	n/r	42.2	47.4	26.3	n/r
NtvHawaii/Pacific (%)	0.2	n/r	n/r	n/r	0.1	n/r	0.1	0.3	n/r	n/r
Two or More (%)	1.7	n/r	4.6	n/r	3.1	n/r	2.5	0.9	1.5	n/r
Unknown (%)	0.1	n/r	n/r	n/r	n/r	n/r	n/r	0.1	n/r	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported due to n<20; n/a = not applicable.

Table A.5.25 Demographic Information for Integrated Mathematics I, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	49.6	96.6	49.7	n/r	46.8	n/r	n/r	62.8	78.6	n/r
SWD (%)	12.2	n/r	11.2	n/r	12.3	n/r	n/r	100.0	20.6	n/r
EL (%)	12.7	69.0	18.1	n/r	5.1	n/r	n/r	n/r	12.1	n/r
Male (%)	51.0	n/r	51.7	n/r	49.7	n/r	n/r	83.7	51.8	n/r
Female (%)	49.0	n/r	48.3	n/r	50.3	n/r	n/r	n/r	48.2	n/r
AmInd/ANat (%)	1.0	100.0	0.9	n/r	0.3	n/r	n/r	n/r	5.8	n/r
Asian (%)	3.3	n/r	4.5	n/r	1.9	n/r	n/r	n/r	n/r	n/r
Black/AA (%)	11.4	n/r	7.3	n/r	17.8	n/r	n/r	n/r	n/r	n/r
Hisp/Lat (%)	39.5	n/r	45.6	n/r	28.5	n/r	n/r	n/r	72.8	n/r
Wh/Caus (%)	42.0	n/r	38.7	n/r	48.7	n/r	n/r	55.8	18.1	n/r
NtvHawaii/Pacific (%)	0.2	n/r	0.3	n/r	n/r	n/r	n/r	n/r	n/r	n/r
Two or More (%)	2.5	n/r	2.6	n/r	2.6	n/r	n/r	n/r	n/r	n/r
Unknown (%)	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported due to n<20; n/a = not applicable.

Table A.5.26 Demographic Information for Integrated Mathematics II, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	65.9	99.0	34.4	36.2	75.6	n/r	n/r	80.0	90.4	n/r
SWD (%)	8.8	n/r	2.3	n/r	9.7	n/r	n/r	100.0	15.2	n/r
EL (%)	9.3	37.4	1.5	n/r	10.4	n/r	n/r	n/r	14.3	n/r
Male (%)	48.8	46.6	47.4	34.8	48.9	n/r	n/r	66.7	52.7	n/r
Female (%)	51.2	53.4	52.6	65.2	51.1	n/r	n/r	n/r	47.3	n/r
AmInd/ANat (%)	8.7	98.1	n/r	n/r	n/r	n/r	n/r	n/r	31.2	n/r
Asian (%)	1.9	n/r	4.5	n/r	1.0	n/r	n/r	n/r	n/r	n/r
Black/AA (%)	10.4	n/r	5.7	65.2	15.0	n/r	n/r	n/r	n/r	n/r
Hisp/Lat (%)	47.2	n/r	27.6	n/r	62.6	n/r	n/r	n/r	52.7	n/r
Wh/Caus (%)	29.7	n/r	56.6	n/r	20.7	n/r	n/r	n/r	14.3	n/r
NtvHawaii/Pacific (%)	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r
Two or More (%)	1.8	n/r	4.7	n/r	n/r	n/r	n/r	n/r	n/r	n/r
Unknown (%)	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported due to n<20; n/a = not applicable.

Table A.5.27 Demographic Information for Integrated Mathematics III, Overall and by State

Demographic	PARCC	BIE	CO	DC	IL	MA	MD	NJ	NM	RI
Econ Dis (%)	50.7	n/r	15.4	n/r	38.5	n/r	n/r	n/r	94.3	
SWD (%)	10.3	n/r	n/r	n/r	9.8	n/r	n/r	n/r	12.5	
EL (%)	2.6	n/r	n/r	n/r	n/r	n/r	n/r	n/r	8.2	
Male (%)	46.7	n/r	55.0	n/r	47.0	n/r	n/r	n/r	43.2	
Female (%)	53.3	n/r	45.0	n/r	53.0	n/r	n/r	n/r	56.8	
AmInd/ANat (%)	6.2	n/r	n/r	n/r	n/r	n/r	n/r	n/r	22.7	
Asian (%)	1.3	n/r	n/r	n/r	1.3	n/r	n/r	n/r	n/r	
Black/AA (%)	3.7	n/r	n/r	n/r	5.1	n/r	n/r	n/r	n/r	
Hisp/Lat (%)	24.3	n/r	14.8	n/r	11.4	n/r	n/r	n/r	62.8	
Wh/Caus (%)	63.0	n/r	76.3	n/r	80.6	n/r	n/r	n/r	12.3	
NtvHawaii/Pacific (%)	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	
Two or More (%)	1.3	n/r	n/r	n/r	1.3	n/r	n/r	n/r	n/r	
Unknown (%)	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	

Note: PARCC = data from all participating states combined. Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported due to n<20; n/a = not applicable.

Appendix 7: Summary of Differential Item Function (DIF) Results

Table A.7.1 Differential Item Functioning for ELA/L Grade 3

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	71	105	105			2	2	101	96	2	2
	PBT	57	70	70			1	1	65	93	4	6
White vs AmerIndian	CBT	71	105	105			1	1	104	99		
	PBT	57	70	70	1	1	8	11	61	87		
White vs Asian	CBT	71	105	105			1	1	102	97	2	2
	PBT	57	70	70			1	1	67	96	2	3
White vs Black	CBT	71	105	105			2	2	103	98		
	PBT	57	70	70			4	6	66	94		
White vs Hispanic	CBT	71	105	105	1	1	2	2	102	97		
	PBT	57	70	70	1	1	7	10	62	89		
White vs Pacific Islander	CBT	71	105	105					105	100		
	PBT	57	70	52	1	2	2	4	49	94		
White vs Multiracial	CBT	71	105	105					105	100		
	PBT	57	70	70					70	100		
NoEcnDis vs EcnDis	CBT	71	105	105					105	100		
	PBT	57	70	70			3	4	67	96		
ELN vs ELY	CBT	71	105	105	1	1	2	2	102	97		
	PBT	57	70	70	3	4	3	4	64	91		
SWDN vs SWDY	CBT	71	105	105					105	100		
	PBT	57	70	70					70	100		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability.

Table A.7.2 Differential Item Functioning for ELA/L Grade 4

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	92	111	111			3	3	99	89	9	8
	PBT	54	74	74			2	3	63	85	9	12
White vs AmerIndian	CBT	92	111	111			4	4	107	96		
	PBT	54	74	74	20	27	14	19	37	50	3	4
White vs Asian	CBT	92	111	111					104	94	7	6
	PBT	54	74	74			1	1	72	97	1	1
White vs Black	CBT	92	111	111					111	100		
	PBT	54	74	74	4	5	12	16	58	78		
White vs Hispanic	CBT	92	111	111			3	3	108	97		
	PBT	54	74	74	3	4	13	18	57	77	1	1
White vs Pacific Islander	CBT	92	111	111			1	1	110	99		
	PBT	54	74	0								
White vs Multiracial	CBT	92	111	111					111	100		
	PBT	54	74	74					74	100		
NoEcnDis vs EcnDis	CBT	92	111	111					111	100		
	PBT	54	74	74	2	3	4	5	68	92		
ELN vs ELY	CBT	92	111	111	2	2	11	10	98	88		
	PBT	54	74	74	2	3	9	12	63	85		
SWDN vs SWDY	CBT	92	111	111					111	100		
	PBT	54	74	74	9	12	4	5	61	82		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.3 Differential Item Functioning for ELA/L Grade 5

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	91	111	111			4	4	100	90	7	6
	PBT	48	74	74					64	86	10	14
White vs AmerIndian	CBT	91	111	111	1	1	2	2	108	97		
	PBT	48	74	74	15	20	19	26	40	54		
White vs Asian	CBT	91	111	111					110	99	1	1
	PBT	48	74	74					74	100		
White vs Black	CBT	91	111	111					111	100		
	PBT	48	74	74					74	100		
White vs Hispanic	CBT	91	111	111	2	2	1	1	108	97		
	PBT	48	74	74			3	4	71	96		
White vs Pacific Islander	CBT	91	111	111			2	2	109	98		
	PBT	48	74	0								
White vs Multiracial	CBT	91	111	111					111	100		
	PBT	48	74	74					74	100		
NoEcnDis vs EcnDis	CBT	91	111	111					111	100		
	PBT	48	74	74					74	100		
ELN vs ELY	CBT	91	111	111	3	3			108	97		
	PBT	48	74	74	2	3	3	4	69	93		
SWDN vs SWDY	CBT	91	111	111					111	100		
	PBT	48	74	74	1	1			73	99		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.4 Differential Item Functioning for ELA/L Grade 6

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	86	129	129	1	1	9	7	101	78	18	14		
	PBT	59	86	86	2	2	4	5	65	76	3	3	12	14
White vs AmerIndian	CBT	86	129	129			6	5	123	95				
	PBT	59	86	86	37	43	20	23	29	34				
White vs Asian	CBT	86	129	129					120	93	9	7		
	PBT	59	86	86					77	90	9	10		
White vs Black	CBT	86	129	129			1	1	128	99				
	PBT	59	86	86	1	1	1	1	84	98				
White vs Hispanic	CBT	86	129	129			5	4	124	96				
	PBT	59	86	86	1	1	1	1	84	98				
White vs Pacific Islander	CBT	86	129	129			2	2	127	98				
	PBT	59	86	0										
White vs Multiracial	CBT	86	129	129					129	100				
	PBT	59	86	86			2	2	84	98				
NoEcnDis vs EcnDis	CBT	86	129	129					129	100				
	PBT	59	86	86			1	1	85	99				
ELN vs ELY	CBT	86	129	129	3	2	6	5	120	93				
	PBT	59	86	86	1	1	6	7	79	92				
SWDN vs SWDY	CBT	86	129	129					129	100				
	PBT	59	86	86	3	3	10	12	73	85				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.5 Differential Item Functioning for ELA/L Grade 7

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	94	129	129	1	1	10	8	115	89	3	2		
	PBT	77	86	86			2	2	70	81	8	9	6	7
White vs AmerIndian	CBT	94	129	129	1	1	3	2	123	95	2	2		
	PBT	77	86	86	27	31	20	23	33	38	6	7		
White vs Asian	CBT	94	129	129			1	1	128	99				
	PBT	77	86	86					80	93	6	7		
White vs Black	CBT	94	129	129					129	100				
	PBT	77	86	86			3	3	83	97				
White vs Hispanic	CBT	94	129	129	1	1	2	2	126	98				
	PBT	77	86	86	3	3	3	3	77	90	3	3		
White vs Pacific Islander	CBT	94	129	129			1	1	125	97	3	2		
	PBT	77	86	0										
White vs Multiracial	CBT	94	129	129					129	100				
	PBT	77	86	86			1	1	85	99				
NoEcnDis vs EcnDis	CBT	94	129	129					129	100				
	PBT	77	86	86			1	1	85	99				
ELN vs ELY	CBT	94	129	129	2	2	6	5	118	91	3	2		
	PBT	77	86	86	3	3	7	8	72	84	3	3	1	1
SWDN vs SWDY	CBT	94	129	129					129	100				
	PBT	77	86	86	4	5	4	5	78	91				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.6 Differential Item Functioning for ELA/L Grade 8

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	90	129	129	5	4	10	8	102	79	12	9		
	PBT	60	86	86	2	2	6	7	67	78	10	12	1	1
White vs AmerIndian	CBT	90	129	129					129	100				
	PBT	60	86	86	36	42	15	17	30	35	5	6		
White vs Asian	CBT	90	129	129			2	2	127	98				
	PBT	60	86	86					83	97	3	3		
White vs Black	CBT	90	129	129					129	100				
	PBT	60	86	86			3	3	83	97				
White vs Hispanic	CBT	90	129	129			6	5	123	95				
	PBT	60	86	86	1	1	9	10	76	88				
White vs Pacific Islander	CBT	90	129	129			1	1	128	99				
	PBT	60	86	0										
White vs Multiracial	CBT	90	129	129					129	100				
	PBT	60	86	86			3	3	83	97				
NoEcnDis vs EcnDis	CBT	90	129	129					129	100				
	PBT	60	86	86			1	1	85	99				
ELN vs ELY	CBT	90	129	129	1	1	4	3	124	96				
	PBT	60	86	86	11	13	13	15	55	64	7	8		
SWDN vs SWDY	CBT	90	129	129					129	100				
	PBT	60	86	86			5	6	81	94				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.7 Differential Item Functioning for ELA/L Grade 9

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	95	129	129			11	9	112	87	3	2	3	2
	PBT	70	86	86			8	9	70	81	8	9		
White vs AmerIndian	CBT	95	129	129			2	2	127	98				
	PBT	70	86	86	2	2	3	3	80	93	1	1		
White vs Asian	CBT	95	129	129					129	100				
	PBT	70	86	86					86	100				
White vs Black	CBT	95	129	129			1	1	128	99				
	PBT	70	86	86			1	1	85	99				
White vs Hispanic	CBT	95	129	129			2	2	127	98				
	PBT	70	86	86	1	1	1	1	84	98				
White vs Pacific Islander	CBT	95	129	129					129	100				
	PBT	70	86	0										
White vs Multiracial	CBT	95	129	129					129	100				
	PBT	70	86	86	1	1			85	99				
NoEcnDis vs EcnDis	CBT	95	129	129					129	100				
	PBT	70	86	86			1	1	85	99				
ELN vs ELY	CBT	95	129	129	2	2	1	1	126	98				
	PBT	70	86	86	3	3	2	2	76	88	5	6		
SWDN vs SWDY	CBT	95	129	129					129	100				
	PBT	70	86	86			1	1	84	98	1	1		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.8 Differential Item Functioning for ELA/L Grade 10

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	109	129	129	4	3	7	5	111	86	7	5
	PBT	62	86	86			10	12	69	80	7	8
White vs AmerIndian	CBT	109	129	129	3	2	3	2	121	94	2	2
	PBT	62	86	86	2	2	8	9	76	88		
White vs Asian	CBT	109	129	129					129	100		
	PBT	62	86	86			1	1	85	99		
White vs Black	CBT	109	129	129					129	100		
	PBT	62	86	86					86	100		
White vs Hispanic	CBT	109	129	129	1	1	3	2	125	97		
	PBT	62	86	86			5	6	81	94		
White vs Pacific Islander	CBT	109	129	129			1	1	128	99		
	PBT	62	86	0								
White vs Multiracial	CBT	109	129	129					129	100		
	PBT	62	86	86			4	5	82	95		
NoEcnDis vs EcnDis	CBT	109	129	129			3	2	126	98		
	PBT	62	86	86					86	100		
ELN vs ELY	CBT	109	129	129	3	2	3	2	123	95		
	PBT	62	86	86	1	1	9	10	76	88		
SWDN vs SWDY	CBT	109	129	129					129	100		
	PBT	62	86	86					86	100		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.9 Differential Item Functioning for ELA/L Grade 11

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	84	129	129	2	2	2	2	125	97				
	PBT	68	86	86	3	3	8	9	69	80	4	5	2	2
White vs AmerIndian	CBT	84	129	129	2	2	7	5	118	91	2	2		
	PBT	68	86	86	6	7	18	21	48	56	5	6	9	10
White vs Asian	CBT	84	129	129					128	99			1	1
	PBT	68	86	0										
White vs Black	CBT	84	129	129			1	1	128	99				
	PBT	68	86	86			1	1	85	99				
White vs Hispanic	CBT	84	129	129			4	3	125	97				
	PBT	68	86	86			3	3	83	97				
White vs Pacific Islander	CBT	84	129	129			6	5	121	94	2	2		
	PBT	68	86	0										
White vs Multiracial	CBT	84	129	129					129	100				
	PBT	68	86	86					85	99	1	1		
NoEcnDis vs EcnDis	CBT	84	129	129					129	100				
	PBT	68	86	86					86	100				
ELN vs ELY	CBT	84	129	129	3	2	6	5	120	93				
	PBT	68	86	0										
SWDN vs SWDY	CBT	84	129	129					129	100				
	PBT	68	86	86			1	1	83	97	2	2		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.10 Differential Item Functioning for Mathematics Grade 3

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	97	129	129	1	1	4	3	124	96				
	PBT	75	86	86			1	1	85	99				
White vs AmerIndian	CBT	97	129	129			2	2	127	98				
	PBT	75	86	86	3	3	7	8	74	86	2	2		
White vs Asian	CBT	97	129	129					111	86	16	12	2	2
	PBT	75	86	86			1	1	82	95	3	3		
White vs Black	CBT	97	129	129			2	2	124	96	3	2		
	PBT	75	86	86			7	8	78	91	1	1		
White vs Hispanic	CBT	97	129	129			1	1	128	99				
	PBT	75	86	86	1	1	7	8	76	88	2	2		
White vs Pacific Islander	CBT	97	129	129					127	98	2	2		
	PBT	75	86	40	1	3	1	3	33	83	5	13		
White vs Multiracial	CBT	97	129	129					129	100				
	PBT	75	86	86					86	100				
NoEcnDis vs EcnDis	CBT	97	129	129			1	1	128	99				
	PBT	75	86	86	1	1	3	3	80	93	1	1	1	1
ELN vs ELY	CBT	97	129	129			3	2	126	98				
	PBT	75	86	86			4	5	82	95				
SWDN vs SWDY	CBT	97	129	129			4	3	125	97				
	PBT	75	86	86			1	1	84	98	1	1		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.11 Differential Item Functioning for Mathematics Grade 4

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	101	120	120			3	3	113	94	4	3
	PBT	65	80	80			2	3	75	94	3	4
White vs AmerIndian	CBT	101	120	120			1	1	119	99		
	PBT	65	80	80	4	5	8	10	67	84	1	1
White vs Asian	CBT	101	120	120					119	99	1	1
	PBT	65	80	80			1	1	75	94	4	5
White vs Black	CBT	101	120	120					120	100		
	PBT	65	80	80	1	1	5	6	73	91	1	1
White vs Hispanic	CBT	101	120	120					120	100		
	PBT	65	80	80			2	3	78	98		
White vs Pacific Islander	CBT	101	120	120			1	1	119	99		
	PBT	65	80	0								
White vs Multiracial	CBT	101	120	120					120	100		
	PBT	65	80	80					80	100		
NoEcnDis vs EcnDis	CBT	101	120	120					120	100		
	PBT	65	80	80			5	6	74	93	1	1
ELN vs ELY	CBT	101	120	120			1	1	119	99		
	PBT	65	80	80			1	1	79	99		
SWDN vs SWDY	CBT	101	120	120					120	100		
	PBT	65	80	80			1	1	79	99		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English Learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.12 Differential Item Functioning for Mathematics Grade 5

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	80	100	100	1	1	1	1	95	95	3	3		
	PBT	69	80	80			6	8	74	93				
White vs AmerIndian	CBT	80	100	100					100	100				
	PBT	69	80	80	1	1	2	3	77	96				
White vs Asian	CBT	80	100	100					97	97	3	3		
	PBT	69	80	80					79	99	1	1		
White vs Black	CBT	80	100	100					99	99	1	1		
	PBT	69	80	80	1	1	3	4	73	91	3	4		
White vs Hispanic	CBT	80	100	100					100	100				
	PBT	69	80	80			2	3	78	98				
White vs Pacific Islander	CBT	80	100	98			1	1	96	98	1	1		
	PBT	69	80	0										
White vs Multiracial	CBT	80	100	100					100	100				
	PBT	69	80	80					80	100				
NoEcnDis vs EcnDis	CBT	80	100	100					100	100				
	PBT	69	80	80					76	95	4	5		
ELN vs ELY	CBT	80	100	100			2	2	98	98				
	PBT	69	80	80			1	1	79	99				
SWDN vs SWDY	CBT	80	100	100			3	3	96	96			1	1
	PBT	69	80	80					80	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.13 Differential Item Functioning for Mathematics Grade 6

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	101	114	114			4	4	110	96				
	PBT	63	76	76			2	3	74	97				
White vs AmerIndian	CBT	101	114	114			1	1	113	99				
	PBT	63	76	76	5	7	3	4	68	89				
White vs Asian	CBT	101	114	114					111	97	3	3		
	PBT	63	76	76					69	91	7	9		
White vs Black	CBT	101	114	114					113	99	1	1		
	PBT	63	76	76					75	99	1	1		
White vs Hispanic	CBT	101	114	114			2	2	112	98				
	PBT	63	76	76			1	1	75	99				
White vs Pacific Islander	CBT	101	114	114					114	100				
	PBT	63	76	0										
White vs Multiracial	CBT	101	114	114					114	100				
	PBT	63	76	76			1	1	75	99				
NoEcnDis vs EcnDis	CBT	101	114	114					114	100				
	PBT	63	76	76					76	100				
ELN vs ELY	CBT	101	114	114			2	2	112	98				
	PBT	63	76	76			3	4	73	96				
SWDN vs SWDY	CBT	101	114	114					112	98			2	2
	PBT	63	76	76					76	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.14 Differential Item Functioning for Mathematics Grade 7

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	103	114	114	1	1	11	10	101	89	1	1		
	PBT	65	76	76			2	3	72	95	2	3		
White vs AmerIndian	CBT	103	114	114	4	4	1	1	109	96				
	PBT	65	76	76			6	8	70	92				
White vs Asian	CBT	103	114	114			1	1	93	82	15	13	5	4
	PBT	65	76	76			2	3	60	79	9	12	5	7
White vs Black	CBT	103	114	114			3	3	107	94	3	3	1	1
	PBT	65	76	76					76	100				
White vs Hispanic	CBT	103	114	114	3	3	3	3	108	95				
	PBT	65	76	76					76	100				
White vs Pacific Islander	CBT	103	114	114			1	1	113	99				
	PBT	65	76	0										
White vs Multiracial	CBT	103	114	114					114	100				
	PBT	65	76	76					76	100				
NoEcnDis vs EcnDis	CBT	103	114	114	1	1	5	4	108	95				
	PBT	65	76	76					76	100				
ELN vs ELY	CBT	103	114	114	5	4	8	7	101	89				
	PBT	65	76	76			1	1	74	97			1	1
SWDN vs SWDY	CBT	103	114	114					114	100				
	PBT	65	76	76					76	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.15 Differential Item Functioning for Mathematics Grade 8

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	74	92	92			3	3	89	97				
	PBT	63	72	72			3	4	69	96				
White vs AmerIndian	CBT	74	92	92	1	1	4	4	87	95				
	PBT	63	72	72	2	3	5	7	65	90				
White vs Asian	CBT	74	92	92					85	92	6	7	1	1
	PBT	63	72	72					63	88	7	10	2	3
White vs Black	CBT	74	92	92					92	100				
	PBT	63	72	72					72	100				
White vs Hispanic	CBT	74	92	92					92	100				
	PBT	63	72	72					72	100				
White vs Pacific Islander	CBT	74	92	89			2	2	87	98				
	PBT	63	72	0										
White vs Multiracial	CBT	74	92	92					92	100				
	PBT	63	72	72					72	100				
NoEcnDis vs EcnDis	CBT	74	92	92					92	100				
	PBT	63	72	72					72	100				
ELN vs ELY	CBT	74	92	92	1	1	3	3	88	96				
	PBT	63	72	72			2	3	69	96	1	1		
SWDN vs SWDY	CBT	74	92	92					92	100				
	PBT	63	72	72					72	100				

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.16 Differential Item Functioning for Algebra I

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	100	126	126			3	2	123	98				
	PBT	69	84	84	1	1	3	4	79	94	1	1		
White vs AmerIndian	CBT	100	126	126					126	100				
	PBT	69	84	84	1	1	3	4	79	94	1	1		
White vs Asian	CBT	100	126	126			2	2	118	94	5	4	1	1
	PBT	69	84	84			2	2	77	92	5	6		
White vs Black	CBT	100	126	126					126	100				
	PBT	69	84	84			1	1	82	98	1	1		
White vs Hispanic	CBT	100	126	126					126	100				
	PBT	69	84	84					84	100				
White vs Pacific Islander	CBT	100	126	126	1	1			122	97	3	2		
	PBT	69	84	0										
White vs Multiracial	CBT	100	126	126					126	100				
	PBT	69	84	84			1	1	83	99				
NoEcnDis vs EcnDis	CBT	100	126	126					126	100				
	PBT	69	84	84			1	1	83	99				
ELN vs ELY	CBT	100	126	126			3	2	121	96	2	2		
	PBT	69	84	84	1	1			83	99				
SWDN vs SWDY	CBT	100	126	126					126	100				
	PBT	69	84	84					83	99	1	1		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.17 Differential Item Functioning for Geometry

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	109	129	129			3	2	126	98				
	PBT	79	86	86	1	1	5	6	80	93				
White vs AmerIndian	CBT	109	129	129			4	3	125	97				
	PBT	79	86	86	2	2	3	3	78	91	3	3		
White vs Asian	CBT	109	129	129					121	94	8	6		
	PBT	79	86	86			3	3	77	90	5	6	1	1
White vs Black	CBT	109	129	129			2	2	126	98	1	1		
	PBT	79	86	86	1	1	2	2	79	92	3	3	1	1
White vs Hispanic	CBT	109	129	129			1	1	128	99				
	PBT	79	86	86			1	1	84	98	1	1		
White vs Pacific Islander	CBT	109	129	129	3	2	4	3	118	91	4	3		
	PBT	79	86	0										
White vs Multiracial	CBT	109	129	129					129	100				
	PBT	79	86	86			3	3	81	94	2	2		
NoEcnDis vs EcnDis	CBT	109	129	129			1	1	128	99				
	PBT	79	86	86					85	99	1	1		
ELN vs ELY	CBT	109	129	129	5	4	4	3	118	91	2	2		
	PBT	79	86	0										
SWDN vs SWDY	CBT	109	129	129			1	1	128	99				
	PBT	79	86	86	1	1	2	2	81	94	2	2		

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.18 Differential Item Functioning for Algebra II

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	101	123	123	1	1	1	1	121	98				
	PBT	73	82	82			2	2	80	98				
White vs AmerIndian	CBT	101	123	123			3	2	115	93	4	3	1	1
	PBT	73	82	0										
White vs Asian	CBT	101	123	123			1	1	105	85	17	14		
	PBT	73	82	82	1	1	1	1	78	95	2	2		
White vs Black	CBT	101	123	123					122	99	1	1		
	PBT	73	82	82	1	1			80	98	1	1		
White vs Hispanic	CBT	101	123	123			1	1	122	99				
	PBT	73	82	82			2	2	80	98				
White vs Pacific Islander	CBT	101	123	123			3	2	115	93	5	4		
	PBT	73	82	0										
White vs Multiracial	CBT	101	123	123					123	100				
	PBT	73	82	82			3	4	79	96				
NoEcnDis vs EcnDis	CBT	101	123	123					123	100				
	PBT	73	82	82					82	100				
ELN vs ELY	CBT	101	123	123	1	1	3	2	112	91	4	3	3	2
	PBT	73	82	0										
SWDN vs SWDY	CBT	101	123	123					123	100				
	PBT	73	82	82	1	1	3	4	71	87	6	7	1	1

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.19 Differential Item Functioning for Integrated Mathematics I

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	42	42	42			1	2	41	98				
White vs AmerIndian	CBT	42	42	42			2	5	40	95				
White vs Asian	CBT	42	42	42			1	2	40	95	1	2		
White vs Black	CBT	42	42	42			1	2	41	98				
White vs Hispanic	CBT	42	42	42			2	5	40	95				
White vs Pacific Islander	CBT	42	42	0										
White vs Multiracial	CBT	42	42	42					42	100				
NoEcnDis vs EcnDis	CBT	42	42	42					42	100				
ELN vs ELY	CBT	42	42	42	2	5	2	5	38	90				
SWDN vs SWDY	CBT	42	42	42			1	2	39	93	1	2	1	2

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.20 Differential Item Functioning for Integrated Mathematics II

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	42	42	42					40	95	2	5		
White vs AmerIndian	CBT	42	42	42	1	2	1	2	39	93	1	2		
White vs Asian	CBT	42	42	41			1	2	39	95	1	2		
White vs Black	CBT	42	42	42	1	2	2	5	37	88	2	5		
White vs Hispanic	CBT	42	42	42			2	5	40	95				
White vs Pacific Islander	CBT	42	42	0										
White vs Multiracial	CBT	42	42	0										
NoEcnDis vs EcnDis	CBT	42	42	42	1	2	1	2	40	95				
ELN vs ELY	CBT	42	42	42	2	5			37	88	3	7		
SWDN vs SWDY	CBT	42	42	42	2	5	1	2	32	76	6	14	1	2

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.21 Differential Item Functioning for Integrated Mathematics III

DIF Comparisons	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
Male vs Female	CBT	40	40	40					38	95	2	5
White vs AmerIndian	CBT	40	40	40			2	5	35	88	3	8
White vs Asian	CBT	40	40	0								
White vs Black	CBT	40	40	0								
White vs Hispanic	CBT	40	40	40			2	5	37	93	1	3
White vs Pacific Islander	CBT	40	40	0								
White vs Multiracial	CBT	40	40	0								
NoEcnDis vs EcnDis	CBT	40	40	40			1	3	39	98		
ELN vs ELY	CBT	40	40	0								
SWDN vs SWDY	CBT	40	40	40	1	3	1	3	33	83	5	13

Note: AmerIndian = American Indian/Alaska Native, Black = Black/African American, Hispanic = Hispanic/Latino, Pacific Islander = Native Hawaiian or Pacific Islander, Multiracial = Multiple Race Selected, NoEcnDis = not economically disadvantaged, EcnDis = economically disadvantaged, ELN = not an English learner, ELY = English learner, SWDN = not student with disability(ies), SWDY = student with disability. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Table A.7.22 Differential Item Functioning for Mathematics: Spanish-Language vs. English-Language Forms

Grade	Mode	Total N of Unique Items	Total N of Item Occurrences	Total N of Item Occurrences Included in DIF Analysis	C- DIF		B- DIF		A DIF		B+ DIF		C+ DIF	
					N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF	N of Occurrences	% of Total Occurrences in DIF
3	CBT	43	1	1	1	100								
	PBT	43	0	n/a										
4	CBT	40	4	4					4	100				
	PBT	40	1	1					1	100				
5	CBT	40	3	3					3	100				
	PBT	40	3	0										
6	CBT	38	4	4					2	50			2	50
	PBT	38	1	0										
7	CBT	38	3	3					3	100				
	PBT	38	1	1					1	100				
8	CBT	36	4	4					4	100				
	PBT	36	4	3					3	100				
A1	CBT	42	0	n/a										
	PBT	42	3	3					2	67			1	33
GO	CBT	43	7	7			1	14	6	86				
	PBT	43	10	0										
A2	CBT	41	3	3					3	100				
	PBT	41	8	0										

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II. Small sample sizes may result in fewer items in the column *Total N of Item Occurrences Included in DIF Analysis*.

Appendix 8: Reliability of Classification by Content and Grade Level

Table A.8.1 Reliability of Classification: Grade 3 ELA/L

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
		Summative Scale Score						
CBT	Decision Accuracy	650 - 699	0.15	0.03	0.00	0.00	0.00	0.18
		700 - 724	0.04	0.12	0.04	0.00	0.00	0.20
		725 - 749	0.00	0.04	0.15	0.04	0.00	0.23
		750 - 809	0.00	0.00	0.05	0.29	0.02	0.36
		810 - 850	0.00	0.00	0.00	0.01	0.02	0.03
	Decision Consistency	650 - 699	0.14	0.04	0.01	0.00	0.00	0.19
		700 - 724	0.04	0.09	0.06	0.01	0.00	0.20
		725 - 749	0.00	0.05	0.11	0.05	0.00	0.22
		750 - 809	0.00	0.01	0.06	0.27	0.02	0.35
		810 - 850	0.00	0.00	0.00	0.02	0.02	0.04
PBT	Decision Accuracy	650 - 699	0.14	0.02	0.00	0.00	0.00	0.16
		700 - 724	0.03	0.10	0.04	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.14	0.05	0.00	0.22
		750 - 809	0.00	0.00	0.05	0.33	0.02	0.40
		810 - 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 - 699	0.13	0.03	0.01	0.00	0.00	0.17
		700 - 724	0.04	0.08	0.05	0.01	0.00	0.18
		725 - 749	0.00	0.04	0.11	0.06	0.00	0.21
		750 - 809	0.00	0.01	0.06	0.30	0.02	0.39
		810 - 850	0.00	0.00	0.00	0.02	0.03	0.05

Table A.8.2 Reliability of Classification: Grade 4 ELA/L

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.09	0.02	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.12	0.04	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.18	0.05	0.00	0.27
		750 - 789	0.00	0.00	0.05	0.28	0.03	0.36
		790 - 850	0.00	0.00	0.00	0.02	0.06	0.08
	Decision Consistency	650 - 699	0.09	0.03	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.10	0.06	0.00	0.00	0.19
		725 - 749	0.00	0.05	0.15	0.06	0.00	0.26
		750 - 789	0.00	0.00	0.07	0.25	0.03	0.35
		790 - 850	0.00	0.00	0.00	0.03	0.06	0.09
PBT	Decision Accuracy	650 - 699	0.09	0.02	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.12	0.04	0.00	0.00	0.20
		725 - 749	0.00	0.05	0.18	0.05	0.00	0.27
		750 - 789	0.00	0.00	0.05	0.27	0.03	0.35
		790 - 850	0.00	0.00	0.00	0.02	0.05	0.07
	Decision Consistency	650 - 699	0.09	0.03	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.10	0.06	0.01	0.00	0.19
		725 - 749	0.00	0.05	0.14	0.06	0.00	0.26
		750 - 789	0.00	0.01	0.07	0.23	0.03	0.34
		790 - 850	0.00	0.00	0.00	0.04	0.05	0.08

Table A.8.3 Reliability of Classification: Grade 5 ELA/L

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
		Summative Scale Score						
CBT	Decision Accuracy	650 - 699	0.08	0.02	0.00	0.00	0.00	0.10
		700 - 724	0.03	0.14	0.04	0.00	0.00	0.21
		725 - 749	0.00	0.04	0.19	0.05	0.00	0.28
		750 - 798	0.00	0.00	0.05	0.32	0.02	0.38
		799 - 850	0.00	0.00	0.00	0.01	0.02	0.03
	Decision Consistency	650 - 699	0.08	0.03	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.12	0.05	0.00	0.00	0.20
		725 - 749	0.00	0.05	0.16	0.06	0.00	0.27
		750 - 798	0.00	0.00	0.06	0.29	0.02	0.38
		799 - 850	0.00	0.00	0.00	0.02	0.02	0.04
PBT	Decision Accuracy	650 - 699	0.04	0.01	0.00	0.00	0.00	0.05
		700 - 724	0.02	0.10	0.03	0.00	0.00	0.15
		725 - 749	0.00	0.04	0.18	0.05	0.00	0.27
		750 - 798	0.00	0.00	0.05	0.40	0.03	0.48
		799 - 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 - 699	0.04	0.02	0.00	0.00	0.00	0.07
		700 - 724	0.02	0.08	0.05	0.00	0.00	0.16
		725 - 749	0.00	0.04	0.15	0.07	0.00	0.26
		750 - 798	0.00	0.00	0.07	0.36	0.03	0.46
		799 - 850	0.00	0.00	0.00	0.03	0.03	0.06

Table A.8.4 Reliability of Classification: Grade 6 ELA/L

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.09	0.01	0.00	0.00	0.00	0.10
		700 - 724	0.03	0.14	0.04	0.00	0.00	0.20
		725 - 749	0.00	0.04	0.21	0.04	0.00	0.30
		750 - 789	0.00	0.00	0.04	0.28	0.02	0.35
		790 - 850	0.00	0.00	0.00	0.01	0.04	0.05
	Decision Consistency	650 - 699	0.08	0.03	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.12	0.06	0.00	0.00	0.20
		725 - 749	0.00	0.05	0.18	0.06	0.00	0.28
		750 - 789	0.00	0.00	0.06	0.25	0.02	0.34
		790 - 850	0.00	0.00	0.00	0.02	0.04	0.06
PBT	Decision Accuracy	650 - 699	0.05	0.01	0.00	0.00	0.00	0.06
		700 - 724	0.02	0.10	0.03	0.00	0.00	0.15
		725 - 749	0.00	0.03	0.19	0.05	0.00	0.27
		750 - 789	0.00	0.00	0.05	0.35	0.03	0.42
		790 - 850	0.00	0.00	0.00	0.02	0.07	0.09
	Decision Consistency	650 - 699	0.05	0.02	0.00	0.00	0.00	0.07
		700 - 724	0.02	0.09	0.05	0.00	0.00	0.16
		725 - 749	0.00	0.04	0.16	0.06	0.00	0.26
		750 - 789	0.00	0.00	0.06	0.31	0.03	0.41
		790 - 850	0.00	0.00	0.00	0.04	0.07	0.10

Table A.8.5 Reliability of Classification: Grade 7 ELA/L

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT		650 - 699	0.11	0.02	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.13	0.04	0.00	0.00	0.19
		725 - 749	0.00	0.04	0.17	0.04	0.00	0.25
		750 - 784	0.00	0.00	0.04	0.24	0.03	0.31
		785 - 850	0.00	0.00	0.00	0.02	0.10	0.12
	Decision	650 - 699	0.11	0.03	0.00	0.00	0.00	0.14
		700 - 724	0.03	0.10	0.05	0.00	0.00	0.19
		725 - 749	0.00	0.04	0.14	0.06	0.00	0.24
		750 - 784	0.00	0.00	0.06	0.21	0.03	0.30
		785 - 850	0.00	0.00	0.00	0.04	0.09	0.13
PBT		650 - 699	0.09	0.02	0.00	0.00	0.00	0.10
		700 - 724	0.02	0.10	0.03	0.00	0.00	0.16
		725 - 749	0.00	0.04	0.15	0.05	0.00	0.23
		750 - 784	0.00	0.00	0.05	0.25	0.03	0.34
		785 - 850	0.00	0.00	0.00	0.03	0.14	0.17
	Decision	650 - 699	0.08	0.03	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.08	0.05	0.00	0.00	0.16
		725 - 749	0.00	0.04	0.12	0.06	0.00	0.22
		750 - 784	0.00	0.00	0.06	0.22	0.04	0.32
		785 - 850	0.00	0.00	0.00	0.05	0.13	0.18

Table A.8.6 Reliability of Classification: Grade 8 ELA/L

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.12	0.02	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.12	0.04	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.17	0.04	0.00	0.25
		750 - 793	0.00	0.00	0.04	0.30	0.02	0.36
		794 - 850	0.00	0.00	0.00	0.02	0.06	0.07
	Decision Consistency	650 - 699	0.11	0.03	0.00	0.00	0.00	0.14
		700 - 724	0.03	0.10	0.05	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.14	0.06	0.00	0.24
		750 - 793	0.00	0.00	0.06	0.27	0.03	0.35
		794 - 850	0.00	0.00	0.00	0.03	0.06	0.09
PBT	Decision Accuracy	650 - 699	0.09	0.02	0.00	0.00	0.00	0.10
		700 - 724	0.02	0.10	0.04	0.00	0.00	0.16
		725 - 749	0.00	0.04	0.15	0.04	0.00	0.23
		750 - 793	0.00	0.00	0.05	0.31	0.03	0.39
		794 - 850	0.00	0.00	0.00	0.02	0.09	0.11
	Decision Consistency	650 - 699	0.08	0.03	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.08	0.05	0.00	0.00	0.16
		725 - 749	0.00	0.04	0.12	0.06	0.00	0.22
		750 - 793	0.00	0.00	0.06	0.28	0.04	0.38
		794 - 850	0.00	0.00	0.00	0.04	0.09	0.13

Table A.8.7 Reliability of Classification: Grade 9 ELA/L

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Summative Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.13	0.02	0.00	0.00	0.00	0.15
		700 - 724	0.03	0.14	0.04	0.00	0.00	0.21
		725 - 749	0.00	0.04	0.18	0.04	0.00	0.26
		750 - 790	0.00	0.00	0.04	0.25	0.02	0.31
		791 - 850	0.00	0.00	0.00	0.01	0.06	0.07
	Decision Consistency	650 - 699	0.12	0.03	0.00	0.00	0.00	0.16
		700 - 724	0.03	0.11	0.05	0.00	0.00	0.20
		725 - 749	0.00	0.05	0.14	0.05	0.00	0.25
		750 - 790	0.00	0.00	0.06	0.23	0.02	0.31
		791 - 850	0.00	0.00	0.00	0.03	0.06	0.08
PBT	Decision Accuracy	650 - 699	0.10	0.02	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.11	0.04	0.00	0.00	0.17
		725 - 749	0.00	0.04	0.16	0.04	0.00	0.24
		750 - 790	0.00	0.00	0.05	0.29	0.03	0.37
		791 - 850	0.00	0.00	0.00	0.02	0.08	0.11
	Decision Consistency	650 - 699	0.10	0.03	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.09	0.05	0.00	0.00	0.17
		725 - 749	0.00	0.04	0.12	0.06	0.00	0.23
		750 - 790	0.00	0.00	0.06	0.25	0.03	0.35
		791 - 850	0.00	0.00	0.00	0.04	0.08	0.12

Table A.8.8 Reliability of Classification: Grade 10 ELA/L

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Summative Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.20	0.02	0.00	0.00	0.00	0.23
		700 - 724	0.03	0.09	0.04	0.00	0.00	0.17
		725 - 749	0.00	0.03	0.12	0.04	0.00	0.20
		750 - 793	0.00	0.00	0.04	0.22	0.03	0.29
		794 - 850	0.00	0.00	0.00	0.02	0.09	0.12
	Decision Consistency	650 - 699	0.19	0.04	0.01	0.00	0.00	0.24
		700 - 724	0.04	0.07	0.05	0.01	0.00	0.17
		725 - 749	0.00	0.04	0.09	0.05	0.00	0.19
		750 - 793	0.00	0.01	0.05	0.19	0.03	0.28
		794 - 850	0.00	0.00	0.00	0.04	0.09	0.13
PBT	Decision Accuracy	650 - 699	0.17	0.02	0.00	0.00	0.00	0.19
		700 - 724	0.03	0.09	0.04	0.00	0.00	0.15
		725 - 749	0.00	0.03	0.12	0.04	0.00	0.19
		750 - 793	0.00	0.00	0.04	0.25	0.03	0.32
		794 - 850	0.00	0.00	0.00	0.03	0.12	0.15
	Decision Consistency	650 - 699	0.16	0.03	0.00	0.00	0.00	0.20
		700 - 724	0.03	0.07	0.05	0.01	0.00	0.15
		725 - 749	0.00	0.04	0.09	0.05	0.00	0.18
		750 - 793	0.00	0.00	0.05	0.21	0.03	0.31
		794 - 850	0.00	0.00	0.00	0.05	0.11	0.16

Table A.8.9 Reliability of Classification: Grade 11 ELA/L

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Summative Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.15	0.02	0.00	0.00	0.00	0.17
		700 - 724	0.03	0.13	0.04	0.00	0.00	0.20
		725 - 749	0.00	0.04	0.16	0.05	0.00	0.24
		750 - 791	0.00	0.00	0.04	0.25	0.02	0.31
		792 - 850	0.00	0.00	0.00	0.02	0.05	0.07
	Decision Consistency	650 - 699	0.15	0.04	0.00	0.00	0.00	0.18
		700 - 724	0.04	0.10	0.05	0.00	0.00	0.20
		725 - 749	0.00	0.05	0.12	0.06	0.00	0.23
		750 - 791	0.00	0.00	0.06	0.22	0.03	0.31
		792 - 850	0.00	0.00	0.00	0.03	0.05	0.08
PBT	Decision Accuracy	650 - 699	0.11	0.02	0.00	0.00	0.00	0.14
		700 - 724	0.03	0.11	0.04	0.00	0.00	0.18
		725 - 749	0.00	0.04	0.14	0.05	0.00	0.23
		750 - 791	0.00	0.00	0.05	0.25	0.03	0.33
		792 - 850	0.00	0.00	0.00	0.02	0.10	0.12
	Decision Consistency	650 - 699	0.11	0.04	0.00	0.00	0.00	0.15
		700 - 724	0.03	0.09	0.05	0.01	0.00	0.18
		725 - 749	0.00	0.05	0.11	0.06	0.00	0.22
		750 - 791	0.00	0.01	0.06	0.22	0.04	0.32
		792 - 850	0.00	0.00	0.00	0.04	0.09	0.13

Table A.8.10 Reliability of Classification: Grade 3 Mathematics

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.10	0.02	0.00	0.00	0.00	0.12
		700 - 724	0.02	0.13	0.03	0.00	0.00	0.19
		725 - 749	0.00	0.04	0.17	0.04	0.00	0.25
		750 - 789	0.00	0.00	0.04	0.28	0.02	0.35
		790 - 850	0.00	0.00	0.00	0.02	0.07	0.09
	Decision Consistency	650 - 699	0.10	0.03	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.11	0.05	0.00	0.00	0.19
		725 - 749	0.00	0.05	0.14	0.06	0.00	0.24
		750 - 789	0.00	0.00	0.06	0.25	0.03	0.34
		790 - 850	0.00	0.00	0.00	0.03	0.07	0.10
PBT	Decision Accuracy	650 - 699	0.10	0.02	0.00	0.00	0.00	0.11
		700 - 724	0.02	0.12	0.04	0.00	0.00	0.19
		725 - 749	0.00	0.04	0.18	0.04	0.00	0.26
		750 - 789	0.00	0.00	0.04	0.28	0.02	0.34
		790 - 850	0.00	0.00	0.00	0.02	0.08	0.10
	Decision Consistency	650 - 699	0.09	0.03	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.10	0.05	0.00	0.00	0.19
		725 - 749	0.00	0.04	0.15	0.06	0.00	0.25
		750 - 789	0.00	0.00	0.06	0.25	0.03	0.34
		790 - 850	0.00	0.00	0.00	0.03	0.08	0.11

Table A.8.11 Reliability of Classification: Grade 4 Mathematics

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.10	0.02	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.17	0.03	0.00	0.00	0.23
		725 - 749	0.00	0.04	0.19	0.04	0.00	0.28
		750 - 795	0.00	0.00	0.04	0.29	0.01	0.34
		796 - 850	0.00	0.00	0.00	0.01	0.02	0.03
	Decision Consistency	650 - 699	0.10	0.03	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.14	0.05	0.00	0.00	0.23
		725 - 749	0.00	0.05	0.16	0.05	0.00	0.27
		750 - 795	0.00	0.00	0.05	0.27	0.01	0.34
		796 - 850	0.00	0.00	0.00	0.02	0.02	0.04
PBT	Decision Accuracy	650 - 699	0.13	0.02	0.00	0.00	0.00	0.15
		700 - 724	0.03	0.17	0.04	0.00	0.00	0.24
		725 - 749	0.00	0.04	0.18	0.04	0.00	0.26
		750 - 795	0.00	0.00	0.04	0.26	0.01	0.31
		796 - 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 - 699	0.12	0.04	0.00	0.00	0.00	0.16
		700 - 724	0.03	0.15	0.05	0.00	0.00	0.23
		725 - 749	0.00	0.05	0.15	0.05	0.00	0.26
		750 - 795	0.00	0.00	0.05	0.24	0.01	0.31
		796 - 850	0.00	0.00	0.00	0.02	0.03	0.05

Table A.8.12 Reliability of Classification: Grade 5 Mathematics

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.08	0.02	0.00	0.00	0.00	0.10
		700 - 724	0.03	0.17	0.04	0.00	0.00	0.23
		725 - 749	0.00	0.05	0.21	0.04	0.00	0.30
		750 - 789	0.00	0.00	0.04	0.26	0.02	0.32
		790 - 850	0.00	0.00	0.00	0.01	0.04	0.05
	Decision Consistency	650 - 699	0.08	0.03	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.14	0.06	0.00	0.00	0.23
		725 - 749	0.00	0.06	0.17	0.06	0.00	0.28
		750 - 789	0.00	0.00	0.06	0.24	0.02	0.31
		790 - 850	0.00	0.00	0.00	0.02	0.04	0.06
PBT	Decision Accuracy	650 - 699	0.08	0.02	0.00	0.00	0.00	0.10
		700 - 724	0.03	0.16	0.04	0.00	0.00	0.22
		725 - 749	0.00	0.04	0.20	0.04	0.00	0.28
		750 - 789	0.00	0.00	0.04	0.27	0.02	0.32
		790 - 850	0.00	0.00	0.00	0.01	0.06	0.07
	Decision Consistency	650 - 699	0.08	0.03	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.14	0.05	0.00	0.00	0.22
		725 - 749	0.00	0.05	0.17	0.06	0.00	0.27
		750 - 789	0.00	0.00	0.05	0.24	0.02	0.32
		790 - 850	0.00	0.00	0.00	0.02	0.05	0.08

Table A.8.13 Reliability of Classification: Grade 6 Mathematics

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.11	0.02	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.18	0.04	0.00	0.00	0.25
		725 - 749	0.00	0.04	0.22	0.04	0.00	0.30
		750 - 787	0.00	0.00	0.04	0.24	0.01	0.29
		788 - 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 - 699	0.10	0.03	0.00	0.00	0.00	0.13
		700 - 724	0.03	0.16	0.05	0.00	0.00	0.24
		725 - 749	0.00	0.05	0.18	0.05	0.00	0.29
		750 - 787	0.00	0.00	0.05	0.22	0.02	0.29
		788 - 850	0.00	0.00	0.00	0.02	0.03	0.05
PBT	Decision Accuracy	650 - 699	0.08	0.02	0.00	0.00	0.00	0.10
		700 - 724	0.02	0.18	0.04	0.00	0.00	0.24
		725 - 749	0.00	0.04	0.20	0.04	0.00	0.29
		750 - 787	0.00	0.00	0.04	0.25	0.02	0.31
		788 - 850	0.00	0.00	0.00	0.01	0.06	0.07
	Decision Consistency	650 - 699	0.08	0.03	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.15	0.05	0.00	0.00	0.23
		725 - 749	0.00	0.06	0.17	0.05	0.00	0.28
		750 - 787	0.00	0.00	0.05	0.23	0.02	0.30
		788 - 850	0.00	0.00	0.00	0.02	0.05	0.08

Table A.8.14 Reliability of Classification: Grade 7 Mathematics

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Summative Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.09	0.02	0.00	0.00	0.00	0.11
		700 - 724	0.02	0.19	0.04	0.00	0.00	0.25
		725 - 749	0.00	0.04	0.25	0.04	0.00	0.33
		750 - 785	0.00	0.00	0.04	0.22	0.02	0.28
		786 - 850	0.00	0.00	0.00	0.01	0.02	0.02
	Decision Consistency	650 - 699	0.09	0.03	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.16	0.06	0.00	0.00	0.25
		725 - 749	0.00	0.06	0.21	0.05	0.00	0.32
		750 - 785	0.00	0.00	0.06	0.20	0.02	0.28
		786 - 850	0.00	0.00	0.00	0.01	0.02	0.03
PBT	Decision Accuracy	650 - 699	0.07	0.02	0.00	0.00	0.00	0.09
		700 - 724	0.02	0.17	0.04	0.00	0.00	0.23
		725 - 749	0.00	0.04	0.22	0.05	0.00	0.31
		750 - 785	0.00	0.00	0.04	0.25	0.02	0.31
		786 - 850	0.00	0.00	0.00	0.01	0.05	0.06
	Decision Consistency	650 - 699	0.07	0.03	0.00	0.00	0.00	0.10
		700 - 724	0.03	0.14	0.06	0.00	0.00	0.23
		725 - 749	0.00	0.05	0.18	0.06	0.00	0.30
		750 - 785	0.00	0.00	0.06	0.23	0.02	0.31
		786 - 850	0.00	0.00	0.00	0.02	0.05	0.07

Table A.8.15 Reliability of Classification: Grade 8 Mathematics

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT		650 - 699	0.20	0.04	0.00	0.00	0.00	0.23
		700 - 724	0.04	0.17	0.05	0.00	0.00	0.26
		725 - 749	0.00	0.05	0.15	0.04	0.00	0.24
		750 - 800	0.00	0.00	0.04	0.20	0.01	0.25
		801 - 850	0.00	0.00	0.00	0.00	0.02	0.02
	Decision	650 - 699	0.19	0.05	0.00	0.00	0.00	0.25
		700 - 724	0.05	0.14	0.06	0.00	0.00	0.24
		725 - 749	0.00	0.06	0.12	0.05	0.00	0.23
		750 - 800	0.00	0.01	0.05	0.18	0.01	0.25
		801 - 850	0.00	0.00	0.00	0.01	0.02	0.03
PBT		650 - 699	0.16	0.03	0.00	0.00	0.00	0.18
		700 - 724	0.04	0.11	0.04	0.00	0.00	0.19
		725 - 749	0.00	0.04	0.13	0.05	0.00	0.22
		750 - 800	0.00	0.00	0.05	0.26	0.03	0.33
		801 - 850	0.00	0.00	0.00	0.02	0.06	0.08
	Decision	650 - 699	0.15	0.04	0.01	0.00	0.00	0.20
		700 - 724	0.04	0.08	0.05	0.01	0.00	0.19
		725 - 749	0.01	0.05	0.10	0.06	0.00	0.20
		750 - 800	0.00	0.01	0.06	0.23	0.03	0.32
		801 - 850	0.00	0.00	0.00	0.03	0.06	0.09

Table A.8.16 Reliability of Classification: Algebra I

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
		Summative Scale Score						
CBT	Decision Accuracy	650 - 699	0.11	0.03	0.00	0.00	0.00	0.14
		700 - 724	0.03	0.18	0.04	0.00	0.00	0.26
		725 - 749	0.00	0.05	0.18	0.05	0.00	0.27
		750 - 804	0.00	0.00	0.04	0.26	0.01	0.30
		805 - 850	0.00	0.00	0.00	0.00	0.01	0.02
	Decision Consistency	650 - 699	0.11	0.05	0.00	0.00	0.00	0.16
		700 - 724	0.04	0.15	0.06	0.00	0.00	0.25
		725 - 749	0.00	0.06	0.14	0.06	0.00	0.26
		750 - 804	0.00	0.00	0.06	0.24	0.01	0.31
		805 - 850	0.00	0.00	0.00	0.01	0.01	0.02
PBT	Decision Accuracy	650 - 699	0.10	0.02	0.00	0.00	0.00	0.12
		700 - 724	0.03	0.15	0.04	0.00	0.00	0.22
		725 - 749	0.00	0.05	0.16	0.04	0.00	0.25
		750 - 804	0.00	0.00	0.04	0.32	0.01	0.37
		805 - 850	0.00	0.00	0.00	0.01	0.03	0.04
	Decision Consistency	650 - 699	0.10	0.04	0.00	0.00	0.00	0.14
		700 - 724	0.03	0.12	0.05	0.00	0.00	0.21
		725 - 749	0.00	0.05	0.13	0.06	0.00	0.24
		750 - 804	0.00	0.00	0.05	0.29	0.02	0.37
		805 - 850	0.00	0.00	0.00	0.02	0.03	0.05

Table A.8.17 Reliability of Classification: Geometry

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Summative Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.08	0.02	0.00	0.00	0.00	0.10
		700 - 724	0.02	0.25	0.04	0.00	0.00	0.31
		725 - 749	0.00	0.05	0.24	0.04	0.00	0.33
		750 - 782	0.00	0.00	0.03	0.19	0.01	0.23
		783 - 850	0.00	0.00	0.00	0.01	0.02	0.03
	Decision Consistency	650 - 699	0.08	0.03	0.00	0.00	0.00	0.11
		700 - 724	0.03	0.22	0.05	0.00	0.00	0.30
		725 - 749	0.00	0.06	0.21	0.05	0.00	0.32
		750 - 782	0.00	0.00	0.05	0.18	0.01	0.23
		783 - 850	0.00	0.00	0.00	0.01	0.02	0.03
PBT	Decision Accuracy	650 - 699	0.08	0.02	0.00	0.00	0.00	0.09
		700 - 724	0.02	0.22	0.04	0.00	0.00	0.27
		725 - 749	0.00	0.04	0.26	0.04	0.00	0.34
		750 - 782	0.00	0.00	0.04	0.22	0.02	0.27
		783 - 850	0.00	0.00	0.00	0.01	0.02	0.03
	Decision Consistency	650 - 699	0.07	0.03	0.00	0.00	0.00	0.10
		700 - 724	0.02	0.19	0.05	0.00	0.00	0.26
		725 - 749	0.00	0.06	0.22	0.05	0.00	0.33
		750 - 782	0.00	0.00	0.05	0.20	0.02	0.27
		783 - 850	0.00	0.00	0.00	0.01	0.02	0.03

Table A.8.18 Reliability of Classification: Algebra II

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Summative Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.29	0.04	0.00	0.00	0.00	0.33
		700 - 724	0.04	0.16	0.04	0.00	0.00	0.24
		725 - 749	0.00	0.04	0.13	0.03	0.00	0.20
		750 - 807	0.00	0.00	0.03	0.18	0.01	0.21
		808 - 850	0.00	0.00	0.00	0.00	0.01	0.01
	Decision Consistency	650 - 699	0.28	0.05	0.00	0.00	0.00	0.34
		700 - 724	0.05	0.13	0.05	0.00	0.00	0.23
		725 - 749	0.00	0.05	0.10	0.04	0.00	0.20
		750 - 807	0.00	0.00	0.04	0.16	0.01	0.22
		808 - 850	0.00	0.00	0.00	0.01	0.01	0.02
PBT	Decision Accuracy	650 - 699	0.24	0.03	0.00	0.00	0.00	0.27
		700 - 724	0.04	0.13	0.04	0.00	0.00	0.21
		725 - 749	0.00	0.04	0.13	0.04	0.00	0.22
		750 - 807	0.00	0.00	0.04	0.24	0.02	0.30
		808 - 850	0.00	0.00	0.00	0.00	0.00	0.00
	Decision Consistency	650 - 699	0.23	0.05	0.01	0.00	0.00	0.28
		700 - 724	0.05	0.10	0.05	0.01	0.00	0.21
		725 - 749	0.00	0.05	0.10	0.05	0.00	0.21
		750 - 807	0.00	0.01	0.05	0.22	0.01	0.29
		808 - 850	0.00	0.00	0.00	0.01	0.00	0.01

Table A.8.19 Reliability of Classification: Integrated Mathematics I

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.14	0.03	0.00	0.00	0.00	0.17
		700 - 724	0.04	0.19	0.04	0.00	0.00	0.27
		725 - 749	0.00	0.06	0.16	0.05	0.00	0.26
		750 - 798	0.00	0.00	0.04	0.23	0.01	0.28
		799 - 850	0.00	0.00	0.00	0.00	0.02	0.02
	Decision Consistency	650 - 699	0.14	0.05	0.00	0.00	0.00	0.19
		700 - 724	0.04	0.15	0.06	0.00	0.00	0.26
		725 - 749	0.00	0.07	0.12	0.06	0.00	0.25
		750 - 798	0.00	0.01	0.05	0.21	0.01	0.28
		799 - 850	0.00	0.00	0.00	0.01	0.02	0.03

2016 Technical Report

Table A.8.20 Reliability of Classification: Integrated Mathematics II

		Full						
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.17	0.05	0.00	0.00	0.00	0.22
		700 - 724	0.05	0.22	0.06	0.00	0.00	0.33
		725 - 749	0.00	0.06	0.15	0.05	0.00	0.26
		750 - 784	0.00	0.00	0.03	0.12	0.02	0.17
		785 - 850	0.00	0.00	0.00	0.01	0.02	0.03
	Decision Consistency	650 - 699	0.16	0.07	0.01	0.00	0.00	0.24
		700 - 724	0.06	0.17	0.07	0.01	0.00	0.30
		725 - 749	0.01	0.07	0.12	0.05	0.00	0.25
		750 - 784	0.00	0.01	0.05	0.10	0.01	0.17
		785 - 850	0.00	0.00	0.00	0.02	0.02	0.04
PBT	Decision Accuracy	650 - 699	0.09	0.02	0.00	0.00	0.00	0.11
		700 - 724	0.06	0.18	0.06	0.00	0.00	0.30
		725 - 749	0.00	0.06	0.23	0.06	0.00	0.34
		750 - 784	0.00	0.00	0.04	0.17	0.02	0.23
		785 - 850	0.00	0.00	0.00	0.00	0.01	0.01
	Decision Consistency	650 - 699	0.09	0.04	0.00	0.00	0.00	0.13
		700 - 724	0.06	0.15	0.08	0.01	0.00	0.29
		725 - 749	0.00	0.07	0.18	0.07	0.00	0.31
		750 - 784	0.00	0.00	0.07	0.15	0.02	0.24
		785 - 850	0.00	0.00	0.00	0.01	0.01	0.02

2016 Technical Report

Table A.8.21 Reliability of Classification: Integrated Mathematics III

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Summative Scale Score						Total
CBT	Decision Accuracy	650 - 699	0.28	0.03	0.00	0.00	0.00	0.31
		700 - 724	0.04	0.13	0.04	0.00	0.00	0.21
		725 - 749	0.00	0.04	0.14	0.04	0.00	0.21
		750 - 803	0.00	0.00	0.04	0.21	0.01	0.26
		804 - 850	0.00	0.00	0.00	0.00	0.01	0.01
	Decision Consistency	650 - 699	0.26	0.04	0.00	0.00	0.00	0.31
		700 - 724	0.05	0.10	0.05	0.00	0.00	0.21
		725 - 749	0.00	0.04	0.11	0.05	0.00	0.20
		750 - 803	0.00	0.00	0.05	0.19	0.01	0.26
		804 - 850	0.00	0.00	0.00	0.01	0.01	0.02
PBT	Decision Accuracy	650 - 699	0.41	0.08	0.00	0.02	0.00	0.52
		700 - 724	0.07	0.20	0.06	0.01	0.00	0.34
		725 - 749	0.00	0.05	0.08	0.00	0.00	0.13
		750 - 803	0.00	0.00	0.00	0.00	0.00	0.00
		804 - 850	0.00	0.00	0.00	0.00	0.00	0.00
	Decision Consistency	650 - 699	0.39	0.11	0.01	0.02	0.00	0.53
		700 - 724	0.08	0.15	0.05	0.01	0.00	0.29
		725 - 749	0.01	0.07	0.06	0.00	0.00	0.15
		750 - 803	0.49	0.33	0.13	0.03	0.00	0.97
		804 - 850	0.49	0.34	0.14	0.04	0.00	1.00

Appendix 10.1: IRT Results for Spring 2016 English Language Arts/Literacy (ELA/L)

Table A.10.1 CBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	E03	All Items	140	64	0.49	1.07	-1.42	3.25	0.55	0.19	0.16	1.07
CBT	E03	Reading	104	52	0.25	1.02	-1.42	3.25	0.5	0.18	0.16	1.07
CBT	E03	Writing	36	12	1.57	0.39	0.99	2.28	0.74	0.09	0.49	0.84
CBT	E04	All Items	189	85	0.46	1.09	-3.55	4.96	0.48	0.22	0.17	0.99
CBT	E04	Reading	142	71	0.38	1.17	-3.55	4.96	0.41	0.15	0.17	0.82
CBT	E04	Writing	47	14	0.83	0.40	0.35	1.47	0.87	0.09	0.66	0.99
CBT	E05	All Items	186	84	0.83	1.38	-1.31	7.91	0.48	0.25	0.05	1.12
CBT	E05	Reading	140	70	0.76	1.49	-1.31	7.91	0.40	0.16	0.05	0.85
CBT	E05	Writing	46	14	1.16	0.59	0.46	2.24	0.92	0.12	0.63	1.12
CBT	E06	All Items	178	80	0.55	0.87	-1.43	3.80	0.45	0.21	0.10	0.98
CBT	E06	Reading	136	68	0.48	0.92	-1.43	3.80	0.39	0.15	0.10	0.81
CBT	E06	Writing	42	12	0.94	0.34	0.45	1.54	0.81	0.10	0.68	0.98
CBT	E07	All Items	194	88	0.31	0.87	-1.81	3.18	0.45	0.21	0.17	1.02
CBT	E07	Reading	152	76	0.26	0.92	-1.81	3.18	0.39	0.12	0.17	0.74
CBT	E07	Writing	42	12	0.64	0.30	0.17	1.19	0.87	0.10	0.63	1.02
CBT	E08	All Items	186	84	0.38	0.95	-1.57	5.15	0.47	0.23	0.10	1.03
CBT	E08	Reading	144	72	0.34	1.02	-1.57	5.15	0.40	0.16	0.10	0.90
CBT	E08	Writing	42	12	0.57	0.34	0.11	1.13	0.89	0.08	0.76	1.03
CBT	E09	All Items	196	89	0.74	1.09	-0.78	6.49	0.48	0.25	0.08	1.16

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
					Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
CBT	E09	Reading	154	77	0.68	1.15	-0.78	6.49	0.40	0.14	0.08	0.69
CBT	E09	Writing	42	12	1.09	0.33	0.57	1.62	1.02	0.08	0.88	1.16
CBT	E10	All Items	225	102	0.70	0.83	-1.15	3.39	0.49	0.19	0.14	0.95
CBT	E10	Reading	176	88	0.69	0.88	-1.15	3.39	0.45	0.15	0.14	0.85
CBT	E10	Writing	49	14	0.78	0.28	0.38	1.19	0.80	0.10	0.59	0.95
CBT	E11	All Items	174	78	1.00	1.37	-1.15	8.68	0.47	0.22	0.09	0.96
CBT	E11	Reading	132	66	1.01	1.49	-1.15	8.68	0.40	0.17	0.09	0.90
CBT	E11	Writing	42	12	0.96	0.30	0.46	1.40	0.82	0.10	0.61	0.96

Table A.10.2 PBT IRT Summary Parameter Estimates for All Items for ELA/L by Grade

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
					Mean	SD	Min	Max	Mean	SD	Min	Max
PBT	E03	All Items	114	52	0.66	0.88	-1.10	3.25	0.52	0.18	0.12	0.82
PBT	E03	Reading	84	42	0.51	0.91	-1.10	3.25	0.47	0.15	0.12	0.82
PBT	E03	Writing	30	10	1.27	0.27	0.99	1.83	0.73	0.08	0.51	0.81
PBT	E04	All Items	111	50	0.73	1.06	-1.13	5.09	0.41	0.17	0.15	0.82
PBT	E04	Reading	84	42	0.73	1.14	-1.13	5.09	0.36	0.11	0.15	0.63
PBT	E04	Writing	27	8	0.72	0.53	0.14	1.37	0.71	0.11	0.51	0.82
PBT	E05	All Items	98	44	0.91	0.83	-0.97	2.93	0.46	0.21	0.13	0.94
PBT	E05	Reading	72	36	0.93	0.89	-0.97	2.93	0.38	0.15	0.13	0.84
PBT	E05	Writing	26	8	0.86	0.47	0.50	1.69	0.79	0.09	0.67	0.94
PBT	E06	All Items	121	56	0.50	0.74	-0.82	2.14	0.43	0.16	0.16	0.78
PBT	E06	Reading	100	50	0.48	0.77	-0.82	2.14	0.41	0.14	0.16	0.70
PBT	E06	Writing	21	6	0.66	0.39	0.20	1.18	0.66	0.10	0.53	0.78
PBT	E07	All Items	159	72	0.28	0.83	-1.99	2.92	0.40	0.15	0.11	0.82
PBT	E07	Reading	124	62	0.28	0.88	-1.99	2.92	0.35	0.11	0.11	0.65
PBT	E07	Writing	35	10	0.27	0.40	-0.26	0.85	0.65	0.09	0.54	0.82
PBT	E08	All Items	124	56	0.37	0.84	-1.17	2.97	0.43	0.18	0.09	0.88
PBT	E08	Reading	96	48	0.41	0.89	-1.17	2.97	0.38	0.15	0.09	0.73
PBT	E08	Writing	28	8	0.14	0.35	-0.32	0.57	0.72	0.12	0.52	0.88

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
					Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
PBT	E09	All Items	144	66	0.47	0.83	-1.12	2.93	0.43	0.16	0.18	0.79
PBT	E09	Reading	116	58	0.39	0.84	-1.12	2.93	0.39	0.13	0.18	0.77
PBT	E09	Writing	28	8	1.01	0.47	0.23	1.72	0.71	0.07	0.62	0.79
PBT	E10	All Items	128	58	0.70	0.76	-0.53	2.59	0.50	0.19	0.18	0.90
PBT	E10	Reading	100	50	0.68	0.81	-0.53	2.59	0.46	0.17	0.18	0.83
PBT	E10	Writing	28	8	0.83	0.34	0.38	1.23	0.78	0.1	0.57	0.90
PBT	E11	All Items	140	64	1.06	0.99	-0.97	3.75	0.41	0.21	0.1	0.89
PBT	E11	Reading	112	56	1.06	1.05	-0.97	3.75	0.37	0.19	0.10	0.89
PBT	E11	Writing	28	8	1.08	0.37	0.60	1.53	0.69	0.06	0.57	0.75

Table A.10.3 CBT IRT Standard Errors of Parameter Estimates for All Items for ELA/L by Grade

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	E03	All Items	140	64	0.010	0.010	0.003	0.057	0.004	0.002	0.002	0.009
CBT	E03	Reading	104	52	0.009	0.010	0.003	0.057	0.004	0.001	0.002	0.006
CBT	E03	Writing	36	12	0.013	0.005	0.008	0.024	0.007	0.002	0.005	0.009
CBT	E04	All Items	189	85	0.013	0.016	0.004	0.107	0.005	0.002	0.002	0.011
CBT	E04	Reading	142	71	0.014	0.017	0.004	0.107	0.004	0.001	0.002	0.007
CBT	E04	Writing	47	14	0.009	0.003	0.005	0.015	0.009	0.002	0.006	0.011
CBT	E05	All Items	186	84	0.018	0.041	0.004	0.316	0.004	0.002	0.002	0.011
CBT	E05	Reading	140	70	0.019	0.045	0.004	0.316	0.004	0.001	0.002	0.007
CBT	E05	Writing	46	14	0.014	0.016	0.006	0.065	0.009	0.002	0.006	0.011
CBT	E06	All Items	178	80	0.010	0.010	0.004	0.071	0.004	0.002	0.002	0.010
CBT	E06	Reading	136	68	0.011	0.010	0.004	0.071	0.003	0.001	0.002	0.007
CBT	E06	Writing	42	12	0.007	0.002	0.004	0.012	0.007	0.002	0.005	0.010
CBT	E07	All Items	194	88	0.008	0.005	0.003	0.043	0.004	0.002	0.002	0.010
CBT	E07	Reading	152	76	0.009	0.006	0.003	0.043	0.003	0.001	0.002	0.006
CBT	E07	Writing	42	12	0.005	0.001	0.003	0.007	0.007	0.002	0.004	0.010
CBT	E08	All Items	186	84	0.009	0.010	0.003	0.093	0.003	0.002	0.002	0.009
CBT	E08	Reading	144	72	0.009	0.011	0.003	0.093	0.003	0.001	0.002	0.007
CBT	E08	Writing	42	12	0.005	0.001	0.003	0.006	0.007	0.001	0.005	0.009
CBT	E09	All Items	196	89	0.014	0.027	0.004	0.238	0.004	0.002	0.002	0.011
CBT	E09	Reading	154	77	0.015	0.029	0.004	0.238	0.003	0.001	0.002	0.007

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	E09	Writing	42	12	0.007	0.003	0.004	0.014	0.009	0.001	0.007	0.011
CBT	E10	All Items	225	102	0.015	0.012	0.006	0.079	0.006	0.002	0.003	0.013
CBT	E10	Reading	176	88	0.016	0.013	0.006	0.079	0.006	0.001	0.003	0.009
CBT	E10	Writing	49	14	0.009	0.002	0.006	0.013	0.011	0.002	0.008	0.013
CBT	E11	All Items	174	78	0.022	0.037	0.005	0.297	0.006	0.003	0.003	0.015
CBT	E11	Reading	132	66	0.024	0.040	0.005	0.297	0.005	0.002	0.003	0.010
CBT	E11	Writing	42	12	0.010	0.003	0.006	0.016	0.011	0.003	0.007	0.015

Table A.10.4 PBT IRT Standard Errors of Parameter Estimates for All Items for ELA/L by Grade

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
PBT	E03	All Items	114	52	0.016	0.012	0.007	0.085	0.007	0.003	0.003	0.014
PBT	E03	Reading	84	42	0.016	0.014	0.007	0.085	0.006	0.002	0.003	0.012
PBT	E03	Writing	30	10	0.018	0.003	0.015	0.022	0.012	0.002	0.01	0.014
PBT	E04	All Items	111	50	0.022	0.027	0.007	0.187	0.007	0.003	0.003	0.015
PBT	E04	Reading	84	42	0.023	0.029	0.007	0.187	0.006	0.001	0.003	0.009
PBT	E04	Writing	27	8	0.014	0.006	0.008	0.026	0.011	0.002	0.009	0.015
PBT	E05	All Items	98	44	0.024	0.016	0.009	0.076	0.009	0.004	0.004	0.020
PBT	E05	Reading	72	36	0.026	0.017	0.009	0.076	0.008	0.002	0.004	0.015
PBT	E05	Writing	26	8	0.017	0.005	0.010	0.027	0.017	0.002	0.013	0.020
PBT	E06	All Items	121	56	0.021	0.012	0.010	0.083	0.009	0.003	0.005	0.015
PBT	E06	Reading	100	50	0.022	0.013	0.010	0.083	0.008	0.002	0.005	0.015
PBT	E06	Writing	21	6	0.014	0.003	0.011	0.018	0.012	0.002	0.010	0.014
PBT	E07	All Items	159	72	0.024	0.018	0.011	0.154	0.009	0.003	0.004	0.020
PBT	E07	Reading	124	62	0.025	0.019	0.013	0.154	0.008	0.002	0.004	0.013
PBT	E07	Writing	35	10	0.016	0.003	0.011	0.021	0.015	0.002	0.012	0.020
PBT	E08	All Items	124	56	0.022	0.015	0.010	0.090	0.009	0.003	0.004	0.020
PBT	E08	Reading	96	48	0.023	0.016	0.010	0.090	0.008	0.002	0.004	0.014

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
PBT	E08	Writing	28	8	0.013	0.003	0.011	0.019	0.014	0.003	0.011	0.02
PBT	E09	All Items	144	66	0.032	0.018	0.015	0.122	0.013	0.005	0.007	0.029
PBT	E09	Reading	116	58	0.033	0.019	0.015	0.122	0.012	0.003	0.007	0.020
PBT	E09	Writing	28	8	0.027	0.009	0.018	0.047	0.022	0.005	0.016	0.029
PBT	E10	All Items	128	58	0.043	0.032	0.019	0.174	0.02	0.007	0.009	0.040
PBT	E10	Reading	100	50	0.046	0.033	0.019	0.174	0.018	0.005	0.009	0.030
PBT	E10	Writing	28	8	0.03	0.006	0.021	0.038	0.033	0.004	0.028	0.040
PBT	E11	All Items	140	64	0.074	0.057	0.026	0.258	0.021	0.008	0.010	0.044
PBT	E11	Reading	112	56	0.079	0.059	0.026	0.258	0.019	0.006	0.010	0.037
PBT	E11	Writing	28	8	0.041	0.009	0.030	0.054	0.035	0.008	0.026	0.044

Table A.10.5 CBT IRT Model Fit for All Items for ELA/L by Grade

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	E03	All Items	140	64	3136.9	2980.0	450.4	17317.0	3004.6	2881.8	418.0	17000.5
CBT	E03	Reading	104	52	3547.7	3162.4	450.4	17317.0	3389.7	3066.4	418.0	17000.5
CBT	E03	Writing	36	12	1357.1	489.4	719.0	2248.3	1335.6	478.8	745.5	2147.6
CBT	E04	All Items	189	85	2181.4	1726.3	155.4	9059.9	2121.4	1741.4	148.9	9757.6
CBT	E04	Reading	142	71	2321.8	1831.2	155.4	9059.9	2254.5	1851.4	148.9	9757.6
CBT	E04	Writing	47	14	1469.5	739.5	678.5	2965.0	1446.7	742.5	643.7	2993.0
CBT	E05	All Items	186	84	2521.2	2016.4	282.2	10665.5	2432.3	1982.7	278.3	10922.5
CBT	E05	Reading	140	70	2733.6	2130.0	282.2	10665.5	2617.2	2102.3	278.3	10922.5
CBT	E05	Writing	46	14	1458.9	648.1	891.8	3328.4	1507.5	730.6	875.9	3507.3
CBT	E06	All Items	178	80	2898.3	2583.5	324.9	15422.4	2813.2	2623.8	320.2	16395.8
CBT	E06	Reading	136	68	3073.9	2736.8	324.9	15422.4	2994.9	2781.4	320.2	16395.8
CBT	E06	Writing	42	12	1903.3	1020.6	861.2	3993.5	1783.4	981.4	809.4	3733.8
CBT	E07	All Items	194	88	3447.8	3089.5	333.1	17360.3	3432.7	3351.0	300.3	21851.3
CBT	E07	Reading	152	76	3612.5	3221.8	333.1	17360.3	3632.7	3515.0	300.3	21851.3
CBT	E07	Writing	42	12	2405.2	1829.0	797.9	7344.1	2165.9	1595.5	737.0	6489.6
CBT	E08	All Items	186	84	2919.8	3105.5	253.1	23884.3	2873.3	3175.0	248.4	24231.3
CBT	E08	Reading	144	72	2988.7	3337.8	253.1	23884.3	2966.9	3410.5	248.4	24231.3
CBT	E08	Writing	42	12	2506.5	803.1	1462.9	3690.6	2311.7	765.4	1357.6	3529.1
CBT	E09	All Items	196	89	2629.2	2164.4	301.4	9393.9	2564.0	2245.2	289.5	10501.1
CBT	E09	Reading	154	77	2683.7	2290.4	301.4	9393.9	2646.2	2379.5	289.5	10501.1
CBT	E09	Writing	42	12	2279.4	1037.7	933.7	3918.3	2036.8	926.2	822.2	3611.7
CBT	E10	All Items	225	102	1382.0	987.7	172.2	5524.0	1309.8	971.0	168.4	5623.9
CBT	E10	Reading	176	88	1378.1	1047.1	172.2	5524.0	1323.7	1032.4	168.4	5623.9
CBT	E10	Writing	49	14	1406.2	491.1	807.8	2310.4	1222.5	427.2	704.5	2047.6
CBT	E11	All Items	174	78	1367.9	1250.8	147.5	5899.0	1350.7	1291.0	143.6	6363.8
CBT	E11	Reading	132	66	1341.0	1314.8	147.5	5899.0	1337.9	1372.4	143.6	6363.8
CBT	E11	Writing	42	12	1516.1	841.5	854.2	3381.0	1421.0	729.2	765.0	2973.6

Table A.10.6 PBT IRT Model Fit for All Items for ELA/L by Grade

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	SD	Min	Max	Mean	SD	Min	Max
PBT	E03	All Items	114	52	868.8	661.6	145.2	3861	813.8	604.8	133.3	3544.3
PBT	E03	Reading	84	42	965.9	699.8	145.2	3861	903.4	639	133.3	3544.3
PBT	E03	Writing	30	10	461.1	144.5	260.8	744	437.4	133.3	254.1	697.7
PBT	E04	All Items	111	50	915.2	583.6	54.6	2234.4	872.1	564.8	56	2600.5
PBT	E04	Reading	84	42	935	604.2	54.6	2234.4	882.5	591.7	56	2600.5
PBT	E04	Writing	27	8	811.2	481.2	348.2	1736.3	817.6	422.4	335.7	1613.5
PBT	E05	All Items	98	44	587.1	471.9	79.1	2011.4	582.3	486.6	73.8	2373.9
PBT	E05	Reading	72	36	620.5	512.2	79.1	2011.4	617.4	527.7	73.8	2373.9
PBT	E05	Writing	26	8	437.1	157.8	275.6	703.1	424	164.1	256.8	673.5
PBT	E06	All Items	121	56	497.7	347.2	69.9	1732.8	482.7	346.1	66.8	1695
PBT	E06	Reading	100	50	487.7	361.2	69.9	1732.8	470.6	354.7	66.8	1695
PBT	E06	Writing	21	6	580.9	195.2	384	935.7	584	266.4	362.1	1096.7
PBT	E07	All Items	159	72	461.8	269	75.7	1484.6	457.6	278.1	72.9	1433.7
PBT	E07	Reading	124	62	473.1	283.2	75.7	1484.6	471.8	292.6	72.9	1433.7
PBT	E07	Writing	35	10	391.9	145.3	211.4	643.6	369.4	140.9	205	641.9
PBT	E08	All Items	124	56	481.9	417.7	71.6	2489	505.8	555.4	68.6	3741.3
PBT	E08	Reading	96	48	502.3	442.9	71.6	2489	531.7	592.6	68.6	3741.3
PBT	E08	Writing	28	8	359.4	183.2	199.2	635.7	350.2	183.1	194.9	648.4
PBT	E09	All Items	144	66	230.2	158.4	42.7	927.9	225.1	159.6	39.5	877.2
PBT	E09	Reading	116	58	225.2	163.8	42.7	927.9	219.7	163.1	39.5	877.2

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
PBT	E09	Writing	28	8	265.9	112.3	154.9	439.8	264.0	134.9	137.1	516.6
PBT	E10	All Items	128	58	146.9	103.7	31.6	550.8	142.8	110.1	26.4	530.0
PBT	E10	Reading	100	50	144.6	108.6	31.6	550.8	142.2	116.2	26.4	530.0
PBT	E10	Writing	28	8	160.8	69.6	73.9	267.6	146.2	63.6	65.2	240.9
PBT	E11	All Items	140	64	92.2	65.0	26.5	376.2	89.1	66.3	24.3	373.3
PBT	E11	Reading	112	56	88.9	67.9	26.5	376.2	85.9	68.9	24.3	373.3
PBT	E11	Writing	28	8	115.1	32.3	68.8	163.2	111.3	41.1	62.0	177.6

Appendix 10.2: IRT Results for Spring 2016 Mathematics

Table A.10.7 CBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	M03	All Items	144	97	-0.23	1.34	-3.20	3.58	0.76	0.25	0.24	1.36
CBT	M03	SSMC	26	26	-1.00	1.07	-2.63	0.89	0.74	0.22	0.35	1.36
CBT	M03	CR	118	71	0.05	1.33	-3.20	3.58	0.77	0.27	0.24	1.36
CBT	M03	Type I	92	83	-0.40	1.37	-3.20	3.58	0.80	0.25	0.24	1.36
CBT	M03	Type II	28	8	0.78	0.56	-0.17	1.35	0.50	0.11	0.40	0.71
CBT	M03	Type III	24	6	0.81	0.30	0.31	1.08	0.62	0.13	0.48	0.79
CBT	M04	All Items	155	101	-0.23	1.29	-2.93	2.34	0.72	0.19	0.34	1.32
CBT	M04	SSMC	33	33	-0.92	0.94	-2.65	0.94	0.69	0.18	0.34	0.98
CBT	M04	CR	122	68	0.10	1.31	-2.93	2.34	0.74	0.20	0.38	1.32
CBT	M04	Type I	102	86	-0.45	1.27	-2.93	2.34	0.73	0.20	0.34	1.32
CBT	M04	Type II	35	10	0.89	0.34	0.44	1.38	0.63	0.10	0.50	0.77
CBT	M04	Type III	18	5	1.20	0.68	0.12	1.85	0.73	0.27	0.40	1.13
CBT	M05	All Items	157	95	0.11	1.16	-3.41	2.60	0.67	0.24	0.24	1.26
CBT	M05	SSMC	24	24	-0.69	1.23	-3.41	1.79	0.54	0.21	0.24	0.98
CBT	M05	CR	133	71	0.38	1.02	-2.44	2.60	0.71	0.24	0.30	1.26
CBT	M05	Type I	98	79	-0.05	1.19	-3.41	2.60	0.68	0.25	0.24	1.26
CBT	M05	Type II	29	8	0.82	0.53	0.11	1.73	0.56	0.13	0.34	0.74

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
					Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
CBT	M05	Type III	30	8	0.95	0.71	0.14	1.89	0.66	0.23	0.45	1.15
CBT	M06	All Items	169	101	0.25	1.27	-4.46	3.74	0.74	0.25	0.23	1.42
CBT	M06	SSMC	27	27	-0.51	1.48	-4.46	3.74	0.68	0.24	0.26	1.13
CBT	M06	CR	142	74	0.53	1.06	-1.90	3.44	0.76	0.24	0.23	1.42
CBT	M06	Type I	109	85	0.09	1.28	-4.46	3.74	0.77	0.25	0.23	1.42
CBT	M06	Type II	36	10	1.07	0.89	-0.48	2.02	0.61	0.16	0.38	0.87
CBT	M06	Type III	24	6	1.27	0.65	0.27	1.86	0.60	0.18	0.41	0.87
CBT	M07	All Items	173	103	0.82	1.21	-2.23	3.24	0.76	0.32	0.23	1.84
CBT	M07	SSMC	19	19	-0.22	1.28	-2.23	2.72	0.56	0.20	0.23	1.00
CBT	M07	CR	154	84	1.05	1.07	-2.15	3.24	0.81	0.32	0.30	1.84
CBT	M07	Type I	108	86	0.76	1.29	-2.23	3.24	0.79	0.34	0.23	1.84
CBT	M07	Type II	32	9	1.00	0.81	-0.83	2.29	0.62	0.14	0.42	0.80
CBT	M07	Type III	33	8	1.27	0.50	0.49	1.72	0.60	0.12	0.45	0.77
CBT	M08	All Items	162	89	0.87	1.24	-1.53	3.48	0.63	0.23	0.23	1.29
CBT	M08	SSMC	22	22	-0.46	0.65	-1.53	1.05	0.57	0.16	0.23	0.86
CBT	M08	CR	140	67	1.31	1.06	-1.40	3.48	0.64	0.25	0.24	1.29
CBT	M08	Type I	104	74	0.65	1.21	-1.53	3.48	0.63	0.23	0.23	1.29
CBT	M08	Type II	28	8	1.94	0.73	1.11	3.26	0.70	0.27	0.24	1.01
CBT	M08	Type III	30	7	1.94	0.68	1.08	2.87	0.50	0.14	0.38	0.79

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	A1	All Items	192	100	1.19	1.08	-1.16	3.90	0.63	0.28	0.10	1.51
CBT	A1	SSMC	31	31	0.67	1.12	-1.16	3.55	0.46	0.20	0.10	0.85
CBT	A1	CR	161	69	1.42	0.99	-0.77	3.90	0.71	0.28	0.22	1.51
CBT	A1	Type I	118	81	1.06	1.11	-1.16	3.90	0.62	0.30	0.10	1.51
CBT	A1	Type II	35	10	2.06	0.53	1.11	2.84	0.68	0.16	0.47	0.99
CBT	A1	Type III	39	9	1.40	0.78	0.10	2.47	0.67	0.18	0.43	0.95
CBT	GO	All Items	203	109	0.87	1.13	-1.98	2.98	0.75	0.31	0.22	1.73
CBT	GO	SSMC	27	27	-0.24	1.01	-1.98	1.99	0.55	0.21	0.25	1.11
CBT	GO	CR	176	82	1.23	0.91	-0.86	2.98	0.82	0.30	0.22	1.73
CBT	GO	Type I	126	90	0.67	1.13	-1.98	2.98	0.73	0.32	0.22	1.73
CBT	GO	Type II	32	9	1.90	0.44	1.40	2.57	0.91	0.16	0.58	1.11
CBT	GO	Type III	45	10	1.67	0.52	0.66	2.31	0.76	0.24	0.44	1.14
CBT	A2	All Items	198	101	1.35	0.81	-1.41	2.80	0.69	0.30	0.12	1.44
CBT	A2	SSMC	17	17	0.90	1.08	-1.39	2.67	0.44	0.22	0.12	1.12
CBT	A2	CR	181	84	1.44	0.72	-1.41	2.80	0.74	0.30	0.18	1.44
CBT	A2	Type I	121	82	1.27	0.85	-1.41	2.67	0.69	0.32	0.12	1.44
CBT	A2	Type II	35	10	1.64	0.56	0.69	2.80	0.76	0.19	0.54	1.05
CBT	A2	Type III	42	9	1.73	0.44	1.12	2.33	0.58	0.19	0.28	0.80
CBT	M1	All Items	81	42	1.20	1.06	-0.95	3.41	0.62	0.33	0.17	1.61

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	M1	SSMC	13	13	0.65	1.26	-0.95	3.41	0.41	0.22	0.17	0.92
CBT	M1	CR	68	29	1.45	0.87	-0.22	2.99	0.71	0.33	0.18	1.61
CBT	M1	Type I	49	34	0.99	1.03	-0.95	3.41	0.59	0.35	0.17	1.61
CBT	M1	Type II	14	4	2.37	0.55	1.70	2.99	0.85	0.13	0.73	1.02
CBT	M1	Type III	18	4	1.84	0.65	0.99	2.46	0.61	0.10	0.47	0.68
CBT	M2	All Items	80	41	1.90	1.45	-0.74	5.99	0.53	0.27	0.12	1.18
CBT	M2	SSMC	14	14	1.01	1.21	-0.74	3.24	0.37	0.13	0.12	0.56
CBT	M2	CR	66	27	2.36	1.37	0.04	5.99	0.62	0.29	0.17	1.18
CBT	M2	Type I	48	33	1.81	1.53	-0.74	5.99	0.52	0.29	0.12	1.18
CBT	M2	Type II	14	4	2.87	0.82	2.04	3.76	0.56	0.11	0.43	0.68
CBT	M2	Type III	18	4	1.65	0.98	0.20	2.27	0.64	0.16	0.47	0.79
CBT	M3	All Items	81	40	1.27	1.08	-2.27	4.28	0.59	0.27	0.16	1.27
CBT	M3	SSMC	12	12	0.84	1.71	-2.27	4.28	0.44	0.15	0.23	0.66
CBT	M3	CR	69	28	1.46	0.61	-0.04	2.79	0.66	0.29	0.16	1.27
CBT	M3	Type I	49	32	1.20	1.17	-2.27	4.28	0.58	0.29	0.16	1.27
CBT	M3	Type II	14	4	1.58	0.61	0.96	2.20	0.74	0.08	0.63	0.83
CBT	M3	Type III	18	4	1.57	0.57	0.90	2.11	0.53	0.20	0.30	0.76

Note: M03 through M08 = mathematics grades 3 through 8, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table A.10.8 PBT IRT Summary Parameter Estimates for All Items for Mathematics by Grade/Subject

Mode	Grade	Item Grouping	No. of		<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
			Score Points	No. of Items	Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
PBT	M03	All Items	111	75	-0.34	1.17	-3.22	2.25	0.72	0.24	0.28	1.32
PBT	M03	SSMC	30	30	-0.88	1.29	-3.22	2.25	0.65	0.19	0.28	1.01
PBT	M03	CR	81	45	0.02	0.95	-2.16	1.98	0.77	0.26	0.29	1.32
PBT	M03	Type I	72	64	-0.53	1.16	-3.22	2.25	0.75	0.24	0.28	1.32
PBT	M03	Type II	21	6	0.70	0.62	-0.32	1.38	0.47	0.10	0.37	0.64
PBT	M03	Type III	18	5	0.75	0.52	0.18	1.23	0.57	0.14	0.37	0.73
PBT	M04	All Items	109	65	-0.36	1.12	-2.69	1.53	0.69	0.22	0.30	1.24
PBT	M04	SSMC	24	24	-1.08	1.10	-2.69	1.21	0.65	0.20	0.35	1.07
PBT	M04	CR	85	41	0.06	0.91	-1.83	1.53	0.71	0.23	0.30	1.24
PBT	M04	Type I	67	54	-0.60	1.08	-2.69	1.53	0.71	0.22	0.35	1.24
PBT	M04	Type II	24	7	0.91	0.24	0.65	1.27	0.65	0.16	0.40	0.86
PBT	M04	Type III	18	4	0.57	0.56	0.16	1.39	0.47	0.14	0.30	0.64
PBT	M05	All Items	115	69	0.06	1.15	-2.45	2.28	0.67	0.20	0.31	1.20
PBT	M05	SSMC	28	28	-0.76	0.96	-2.45	1.09	0.68	0.22	0.41	1.20
PBT	M05	CR	87	41	0.62	0.91	-1.49	2.28	0.66	0.19	0.31	1.20
PBT	M05	Type I	70	57	-0.17	1.11	-2.45	2.28	0.68	0.21	0.31	1.20
PBT	M05	Type II	21	6	0.94	0.70	-0.03	1.76	0.57	0.08	0.48	0.68

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
					Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
PBT	M05	Type III	24	6	1.32	0.50	0.52	1.81	0.63	0.19	0.42	0.92
PBT	M06	All Items	110	63	0.20	1.17	-3.54	2.47	0.67	0.22	0.23	1.40
PBT	M06	SSMC	23	23	-0.49	1.24	-3.54	1.52	0.61	0.21	0.23	0.99
PBT	M06	CR	87	40	0.61	0.91	-1.50	2.47	0.71	0.22	0.36	1.40
PBT	M06	Type I	67	52	0.07	1.22	-3.54	2.47	0.70	0.23	0.23	1.40
PBT	M06	Type II	25	7	0.64	0.47	-0.19	1.14	0.53	0.11	0.36	0.71
PBT	M06	Type III	18	4	1.19	0.67	0.18	1.57	0.52	0.06	0.47	0.61
PBT	M07	All Items	102	65	0.86	1.31	-2.03	3.14	0.62	0.22	0.26	1.39
PBT	M07	SSMC	26	26	-0.19	1.08	-2.03	1.79	0.51	0.15	0.26	0.81
PBT	M07	CR	76	39	1.55	0.93	-0.69	3.14	0.69	0.23	0.34	1.39
PBT	M07	Type I	73	57	0.79	1.36	-2.03	3.14	0.63	0.23	0.26	1.39
PBT	M07	Type II	14	4	1.40	0.46	0.89	2.00	0.59	0.08	0.48	0.66
PBT	M07	Type III	15	4	1.30	0.91	0.10	2.22	0.49	0.07	0.45	0.59
PBT	M08	All Items	114	63	1.09	1.27	-1.37	3.54	0.54	0.22	0.18	1.07
PBT	M08	SSMC	25	25	0.46	1.20	-1.37	2.82	0.44	0.15	0.18	0.68
PBT	M08	CR	89	38	1.51	1.15	-0.85	3.54	0.61	0.24	0.20	1.07
PBT	M08	Type I	75	53	0.88	1.18	-1.37	2.91	0.57	0.23	0.18	1.07
PBT	M08	Type II	18	5	2.44	1.12	1.06	3.54	0.43	0.21	0.20	0.71
PBT	M08	Type III	21	5	1.98	1.35	-0.25	3.17	0.42	0.07	0.35	0.50

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	<i>b</i> Estimates Summary				<i>a</i> Estimates Summary			
					Mean	<i>SD</i>	Min	Max	Mean	<i>SD</i>	Min	Max
PBT	A1	All Items	128	69	1.17	1.15	-0.99	3.36	0.55	0.23	0.13	1.30
PBT	A1	SSMC	26	26	0.49	1.09	-0.99	3.09	0.46	0.19	0.13	0.85
PBT	A1	CR	102	43	1.59	0.99	-0.66	3.36	0.61	0.23	0.20	1.30
PBT	A1	Type I	83	57	1.01	1.14	-0.99	3.09	0.54	0.24	0.13	1.30
PBT	A1	Type II	24	7	2.38	0.69	1.27	3.36	0.62	0.14	0.44	0.84
PBT	A1	Type III	21	5	1.40	0.87	0.07	2.37	0.55	0.15	0.43	0.81
PBT	GO	All Items	139	79	1.23	1.28	-1.51	5.98	0.67	0.30	0.18	1.41
PBT	GO	SSMC	26	26	0.74	1.64	-1.51	5.98	0.45	0.18	0.18	0.82
PBT	GO	CR	113	53	1.47	0.99	-0.49	5.40	0.78	0.29	0.29	1.41
PBT	GO	Type I	91	67	1.13	1.35	-1.51	5.98	0.66	0.32	0.18	1.41
PBT	GO	Type II	21	6	2.04	0.22	1.81	2.28	0.78	0.17	0.54	1.02
PBT	GO	Type III	27	6	1.55	0.61	0.54	2.26	0.66	0.18	0.51	0.96
PBT	A2	All Items	143	72	1.65	1.59	-1.14	9.06	0.60	0.27	0.12	1.41
PBT	A2	SSMC	18	18	1.33	2.32	-1.14	9.06	0.53	0.25	0.12	1.04
PBT	A2	CR	125	54	1.76	1.26	-1.14	6.66	0.63	0.27	0.24	1.41
PBT	A2	Type I	83	57	1.57	1.75	-1.14	9.06	0.60	0.29	0.12	1.41
PBT	A2	Type II	24	7	2.12	0.69	1.35	3.03	0.69	0.14	0.50	0.90
PBT	A2	Type III	36	8	1.83	0.60	1.21	3.01	0.55	0.18	0.29	0.77

Note: M03 through M08 = mathematics grades 3 through 8, A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Table A.10.9 CBT IRT Standard Errors of Parameter Estimates for All Items for Mathematics by Grade/Subject

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	M03	All Items	144	97	0.018	0.015	0.005	0.086	0.011	0.004	0.004	0.023
CBT	M03	SSMC	26	26	0.020	0.011	0.006	0.052	0.011	0.004	0.007	0.022
CBT	M03	CR	118	71	0.017	0.017	0.005	0.086	0.010	0.004	0.004	0.023
CBT	M03	Type I	92	83	0.019	0.016	0.005	0.086	0.011	0.004	0.004	0.023
CBT	M03	Type II	28	8	0.010	0.003	0.006	0.016	0.005	0.001	0.004	0.007
CBT	M03	Type III	24	6	0.007	0.002	0.006	0.010	0.006	0.001	0.004	0.008
CBT	M04	All Items	155	101	0.011	0.008	0.003	0.038	0.007	0.002	0.002	0.013
CBT	M04	SSMC	33	33	0.013	0.007	0.007	0.037	0.007	0.002	0.004	0.013
CBT	M04	CR	122	68	0.010	0.008	0.003	0.038	0.007	0.002	0.002	0.013
CBT	M04	Type I	102	86	0.012	0.008	0.003	0.038	0.007	0.002	0.003	0.013
CBT	M04	Type II	35	10	0.006	0.001	0.003	0.007	0.005	0.001	0.004	0.006
CBT	M04	Type III	18	5	0.006	0.002	0.003	0.009	0.005	0.003	0.002	0.009
CBT	M05	All Items	157	95	0.012	0.010	0.003	0.061	0.007	0.002	0.002	0.013
CBT	M05	SSMC	24	24	0.018	0.012	0.008	0.061	0.007	0.002	0.004	0.012
CBT	M05	CR	133	71	0.011	0.008	0.003	0.055	0.007	0.002	0.002	0.013
CBT	M05	Type I	98	79	0.013	0.010	0.004	0.061	0.007	0.002	0.004	0.013
CBT	M05	Type II	29	8	0.007	0.003	0.004	0.013	0.005	0.001	0.002	0.007

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	M05	Type III	30	8	0.008	0.003	0.003	0.012	0.006	0.003	0.003	0.012
CBT	M06	All Items	169	101	0.011	0.012	0.003	0.104	0.007	0.002	0.003	0.014
CBT	M06	SSMC	27	27	0.016	0.020	0.006	0.104	0.007	0.002	0.005	0.012
CBT	M06	CR	142	74	0.009	0.007	0.003	0.050	0.007	0.002	0.003	0.014
CBT	M06	Type I	109	85	0.012	0.013	0.004	0.104	0.007	0.002	0.003	0.014
CBT	M06	Type II	36	10	0.007	0.003	0.005	0.011	0.005	0.002	0.003	0.008
CBT	M06	Type III	24	6	0.007	0.004	0.003	0.013	0.004	0.001	0.003	0.006
CBT	M07	All Items	173	103	0.015	0.010	0.004	0.051	0.009	0.005	0.003	0.041
CBT	M07	SSMC	19	19	0.019	0.011	0.009	0.047	0.008	0.002	0.005	0.015
CBT	M07	CR	154	84	0.014	0.009	0.004	0.051	0.010	0.005	0.003	0.041
CBT	M07	Type I	108	86	0.016	0.010	0.005	0.051	0.010	0.005	0.003	0.041
CBT	M07	Type II	32	9	0.009	0.006	0.006	0.024	0.006	0.002	0.003	0.011
CBT	M07	Type III	33	8	0.009	0.003	0.004	0.014	0.007	0.003	0.004	0.011
CBT	M08	All Items	162	89	0.015	0.010	0.005	0.045	0.007	0.003	0.003	0.020
CBT	M08	SSMC	22	22	0.013	0.005	0.007	0.022	0.007	0.001	0.006	0.010
CBT	M08	CR	140	67	0.016	0.011	0.005	0.045	0.007	0.003	0.003	0.020
CBT	M08	Type I	104	74	0.015	0.009	0.006	0.045	0.007	0.002	0.003	0.020
CBT	M08	Type II	28	8	0.014	0.013	0.005	0.044	0.007	0.004	0.004	0.013
CBT	M08	Type III	30	7	0.017	0.011	0.007	0.038	0.005	0.001	0.004	0.007

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	A1	All Items	192	100	0.017	0.021	0.003	0.191	0.007	0.003	0.002	0.028
CBT	A1	SSMC	31	31	0.024	0.033	0.007	0.191	0.006	0.001	0.004	0.009
CBT	A1	CR	161	69	0.014	0.013	0.003	0.074	0.008	0.004	0.002	0.028
CBT	A1	Type I	118	81	0.019	0.023	0.005	0.191	0.007	0.004	0.003	0.028
CBT	A1	Type II	35	10	0.014	0.007	0.006	0.028	0.007	0.003	0.004	0.013
CBT	A1	Type III	39	9	0.009	0.005	0.003	0.018	0.006	0.003	0.002	0.012
CBT	GO	All Items	203	109	0.016	0.011	0.005	0.066	0.010	0.004	0.004	0.022
CBT	GO	SSMC	27	27	0.020	0.010	0.009	0.053	0.009	0.003	0.007	0.016
CBT	GO	CR	176	82	0.015	0.011	0.005	0.066	0.011	0.004	0.004	0.022
CBT	GO	Type I	126	90	0.017	0.011	0.005	0.066	0.010	0.004	0.004	0.022
CBT	GO	Type II	32	9	0.014	0.006	0.006	0.022	0.013	0.005	0.005	0.021
CBT	GO	Type III	45	10	0.013	0.006	0.006	0.024	0.010	0.005	0.005	0.021
CBT	A2	All Items	198	101	0.020	0.018	0.005	0.161	0.010	0.004	0.004	0.021
CBT	A2	SSMC	17	17	0.033	0.035	0.008	0.161	0.008	0.002	0.005	0.014
CBT	A2	CR	181	84	0.017	0.010	0.005	0.056	0.010	0.004	0.004	0.021
CBT	A2	Type I	121	82	0.021	0.019	0.005	0.161	0.010	0.004	0.004	0.021
CBT	A2	Type II	35	10	0.014	0.009	0.006	0.037	0.010	0.004	0.005	0.014
CBT	A2	Type III	42	9	0.013	0.005	0.008	0.021	0.007	0.003	0.004	0.013
CBT	M1	All Items	81	42	0.051	0.038	0.021	0.199	0.017	0.007	0.008	0.040

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	M1	SSMC	13	13	0.064	0.056	0.025	0.199	0.013	0.003	0.011	0.022
CBT	M1	CR	68	29	0.045	0.026	0.021	0.152	0.018	0.008	0.008	0.040
CBT	M1	Type I	49	34	0.048	0.037	0.021	0.199	0.016	0.008	0.008	0.040
CBT	M1	Type II	14	4	0.059	0.028	0.030	0.094	0.022	0.005	0.018	0.027
CBT	M1	Type III	18	4	0.066	0.058	0.028	0.152	0.013	0.003	0.009	0.016
CBT	M2	All Items	80	41	0.153	0.180	0.028	0.977	0.033	0.021	0.018	0.123
CBT	M2	SSMC	14	14	0.137	0.149	0.044	0.618	0.025	0.002	0.022	0.029
CBT	M2	CR	66	27	0.162	0.196	0.028	0.977	0.037	0.025	0.018	0.123
CBT	M2	Type I	48	33	0.161	0.192	0.031	0.977	0.034	0.024	0.018	0.123
CBT	M2	Type II	14	4	0.193	0.145	0.052	0.344	0.029	0.006	0.024	0.036
CBT	M2	Type III	18	4	0.051	0.017	0.028	0.066	0.028	0.006	0.022	0.035
CBT	M3	All Items	81	40	0.108	0.107	0.038	0.631	0.041	0.018	0.015	0.097
CBT	M3	SSMC	12	12	0.172	0.176	0.053	0.631	0.036	0.005	0.030	0.043
CBT	M3	CR	69	28	0.081	0.037	0.038	0.190	0.043	0.021	0.015	0.097
CBT	M3	Type I	49	32	0.119	0.117	0.049	0.631	0.042	0.019	0.020	0.097
CBT	M3	Type II	14	4	0.062	0.031	0.038	0.107	0.043	0.008	0.038	0.055
CBT	M3	Type III	18	4	0.066	0.020	0.041	0.090	0.031	0.015	0.015	0.045

Note: M03 through M08 = mathematics grades 3 through 8, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table A.10.10 PBT IRT Standard Errors of Parameter Estimates for All Items for Mathematics by Grade/Subject

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
PBT	M03	All Items	111	75	0.018	0.014	0.005	0.096	0.011	0.003	0.004	0.022
PBT	M03	SSMC	30	30	0.024	0.017	0.009	0.096	0.011	0.003	0.008	0.022
PBT	M03	CR	81	45	0.013	0.009	0.005	0.058	0.010	0.004	0.004	0.017
PBT	M03	Type I	72	64	0.019	0.015	0.008	0.096	0.011	0.003	0.007	0.022
PBT	M03	Type II	21	6	0.011	0.003	0.007	0.016	0.006	0.001	0.004	0.008
PBT	M03	Type III	18	5	0.010	0.002	0.005	0.011	0.007	0.003	0.005	0.011
PBT	M04	All Items	109	65	0.018	0.009	0.008	0.044	0.011	0.004	0.004	0.023
PBT	M04	SSMC	24	24	0.025	0.010	0.008	0.044	0.013	0.004	0.009	0.023
PBT	M04	CR	85	41	0.014	0.005	0.008	0.034	0.011	0.004	0.004	0.020
PBT	M04	Type I	67	54	0.019	0.010	0.008	0.044	0.012	0.004	0.006	0.023
PBT	M04	Type II	24	7	0.013	0.002	0.011	0.016	0.009	0.003	0.005	0.013
PBT	M04	Type III	18	4	0.012	0.002	0.009	0.014	0.006	0.001	0.004	0.007
PBT	M05	All Items	115	69	0.020	0.011	0.007	0.070	0.012	0.003	0.005	0.020
PBT	M05	SSMC	28	28	0.024	0.014	0.009	0.070	0.013	0.003	0.010	0.020
PBT	M05	CR	87	41	0.017	0.008	0.007	0.049	0.011	0.003	0.005	0.017
PBT	M05	Type I	70	57	0.021	0.012	0.009	0.070	0.012	0.003	0.006	0.020
PBT	M05	Type II	21	6	0.012	0.005	0.007	0.021	0.008	0.002	0.005	0.010
PBT	M05	Type III	24	6	0.014	0.002	0.012	0.018	0.011	0.004	0.006	0.017

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
PBT	M06	All Items	110	63	0.022	0.024	0.008	0.156	0.012	0.004	0.006	0.020
PBT	M06	SSMC	23	23	0.033	0.036	0.013	0.156	0.013	0.003	0.008	0.020
PBT	M06	CR	87	40	0.015	0.006	0.008	0.034	0.012	0.004	0.006	0.020
PBT	M06	Type I	67	52	0.024	0.026	0.008	0.156	0.013	0.003	0.007	0.020
PBT	M06	Type II	25	7	0.011	0.002	0.008	0.015	0.008	0.002	0.006	0.011
PBT	M06	Type III	18	4	0.012	0.003	0.008	0.015	0.008	0.001	0.006	0.009
PBT	M07	All Items	102	65	0.032	0.019	0.009	0.100	0.015	0.006	0.006	0.036
PBT	M07	SSMC	26	26	0.035	0.014	0.019	0.073	0.014	0.003	0.008	0.018
PBT	M07	CR	76	39	0.030	0.022	0.009	0.100	0.015	0.007	0.006	0.036
PBT	M07	Type I	73	57	0.034	0.020	0.010	0.100	0.016	0.005	0.008	0.036
PBT	M07	Type II	14	4	0.014	0.003	0.011	0.017	0.008	0.002	0.007	0.011
PBT	M07	Type III	15	4	0.017	0.006	0.009	0.024	0.009	0.003	0.006	0.014
PBT	M08	All Items	114	63	0.029	0.015	0.010	0.074	0.012	0.004	0.005	0.028
PBT	M08	SSMC	25	25	0.036	0.016	0.020	0.074	0.011	0.002	0.008	0.017
PBT	M08	CR	89	38	0.025	0.014	0.010	0.072	0.012	0.005	0.005	0.028
PBT	M08	Type I	75	53	0.029	0.014	0.011	0.074	0.012	0.004	0.006	0.028
PBT	M08	Type II	18	5	0.034	0.028	0.010	0.072	0.007	0.003	0.005	0.012
PBT	M08	Type III	21	5	0.027	0.014	0.011	0.047	0.008	0.002	0.007	0.011
PBT	A1	All Items	128	69	0.047	0.044	0.015	0.293	0.017	0.010	0.007	0.077

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	SE of <i>b</i> Estimates				SE of <i>a</i> Estimates			
					Mean	SD	Min	Max	Mean	SD	Min	Max
PBT	A1	SSMC	26	26	0.057	0.064	0.020	0.293	0.015	0.003	0.011	0.024
PBT	A1	CR	102	43	0.040	0.024	0.015	0.096	0.019	0.012	0.007	0.077
PBT	A1	Type I	83	57	0.049	0.047	0.015	0.293	0.017	0.011	0.007	0.077
PBT	A1	Type II	24	7	0.045	0.026	0.018	0.092	0.019	0.006	0.010	0.028
PBT	A1	Type III	21	5	0.021	0.006	0.015	0.027	0.012	0.004	0.008	0.019
PBT	GO	All Items	139	79	0.100	0.153	0.030	1.088	0.039	0.015	0.014	0.091
PBT	GO	SSMC	26	26	0.148	0.213	0.044	1.088	0.032	0.004	0.026	0.043
PBT	GO	CR	113	53	0.077	0.107	0.030	0.804	0.042	0.018	0.014	0.091
PBT	GO	Type I	91	67	0.109	0.164	0.030	1.088	0.039	0.015	0.014	0.091
PBT	GO	Type II	21	6	0.054	0.009	0.043	0.071	0.045	0.016	0.026	0.072
PBT	GO	Type III	27	6	0.044	0.011	0.031	0.060	0.031	0.017	0.016	0.063
PBT	A2	All Items	143	72	0.112	0.215	0.021	1.482	0.029	0.013	0.011	0.096
PBT	A2	SSMC	18	18	0.166	0.343	0.025	1.482	0.027	0.007	0.017	0.052
PBT	A2	CR	125	54	0.094	0.152	0.021	1.044	0.030	0.015	0.011	0.096
PBT	A2	Type I	83	57	0.124	0.240	0.021	1.482	0.030	0.014	0.014	0.096
PBT	A2	Type II	24	7	0.079	0.068	0.022	0.196	0.030	0.009	0.020	0.044
PBT	A2	Type III	36	8	0.053	0.031	0.029	0.126	0.025	0.013	0.011	0.047

Note: M03 through M08 = mathematics grades 3 through 8, A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Table A.10.11 CBT IRT Model Fit for All Items for Mathematics by Grade/Subject

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	M03	All Items	144	97	292.4	396.5	23.3	2232.4	283.4	387.1	23.8	2184.7
CBT	M03	SSMC	26	26	205.7	318.2	31.0	1518.7	205.3	324.9	28.9	1511.7
CBT	M03	CR	118	71	324.2	419.1	23.3	2232.4	312.1	405.8	23.8	2184.7
CBT	M03	Type I	92	83	231.0	357.7	23.3	2232.4	228.8	356.6	23.8	2184.7
CBT	M03	Type II	28	8	662.3	525.7	200.5	1824.9	628.6	509.0	205.8	1746.0
CBT	M03	Type III	24	6	649.4	314.7	172.2	1013.8	579.3	289.4	145.1	918.3
CBT	M04	All Items	155	101	762.6	1235.7	63.9	9556.2	741.0	1188.3	61.5	8903.2
CBT	M04	SSMC	33	33	423.2	447.4	63.9	1779.1	425.3	479.6	61.5	2033.9
CBT	M04	CR	122	68	927.4	1448.8	88.5	9556.2	894.2	1387.4	85.9	8903.2
CBT	M04	Type I	102	86	583.5	881.5	63.9	7649.1	584.3	892.4	61.5	7623.2
CBT	M04	Type II	35	10	1119.5	553.3	541.8	2277.5	1014.6	485.6	502.3	2067.8
CBT	M04	Type III	18	5	3130.6	3599.5	1328.5	9556.2	2888.7	3371.7	1144.6	8903.2
CBT	M05	All Items	157	95	743.0	742.7	67.4	5094.5	708.8	729.9	63.9	4918.4
CBT	M05	SSMC	24	24	596.1	1009.5	79.1	5094.5	579.8	977.7	78.0	4918.4
CBT	M05	CR	133	71	792.6	629.2	67.4	2628.7	752.4	627.4	63.9	2864.7
CBT	M05	Type I	98	79	655.2	716.5	67.4	5094.5	628.5	711.5	63.9	4918.4
CBT	M05	Type II	29	8	1473.5	830.7	318.5	2628.7	1381.1	810.9	302.3	2704.7
CBT	M05	Type III	30	8	878.9	527.8	227.7	1947.8	829.1	499.4	215.9	1809.6

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	M06	All Items	169	101	925.5	1233.4	90.5	5676.1	906.3	1233.7	87.0	5694.0
CBT	M06	SSMC	27	27	667.3	936.7	90.5	4223.6	696.4	1069.9	87.0	5134.5
CBT	M06	CR	142	74	1019.6	1318.3	106.5	5676.1	982.9	1286.4	103.6	5694.0
CBT	M06	Type I	109	85	742.3	1087.1	90.5	5676.1	735.7	1123.9	87.0	5694.0
CBT	M06	Type II	36	10	1471.1	1102.8	469.8	3778.7	1426.4	1099.6	446.4	3793.5
CBT	M06	Type III	24	6	2611.3	1950.9	358.7	5633.5	2456.5	1766.5	329.6	5226.6
CBT	M07	All Items	173	103	727.0	934.8	29.1	5174.9	688.3	893.0	29.2	4958.0
CBT	M07	SSMC	19	19	431.2	510.6	36.2	2001.6	422.9	509.0	36.5	1999.2
CBT	M07	CR	154	84	793.9	996.4	29.1	5174.9	748.4	950.8	29.2	4958.0
CBT	M07	Type I	108	86	558.4	678.7	29.1	3300.0	537.6	670.4	29.2	3151.7
CBT	M07	Type II	32	9	1576.6	1564.8	210.8	5174.9	1481.9	1524.6	209.7	4958.0
CBT	M07	Type III	33	8	1583.6	1492.0	219.8	4267.1	1415.7	1346.2	168.2	3855.5
CBT	M08	All Items	162	89	739.4	886.9	60.1	6471.5	721.6	926.6	58.8	7013.4
CBT	M08	SSMC	22	22	394.7	206.7	144.8	922.7	376.1	199.8	141.1	882.3
CBT	M08	CR	140	67	852.6	991.2	60.1	6471.5	835.0	1038.9	58.8	7013.4
CBT	M08	Type I	104	74	612.6	856.0	60.1	6471.5	590.1	894.7	58.8	7013.4
CBT	M08	Type II	28	8	1550.0	1019.4	307.0	3541.5	1653.5	1038.1	296.3	3141.0
CBT	M08	Type III	30	7	1153.6	366.9	661.8	1781.9	1046.7	320.9	578.2	1542.2
CBT	A1	All Items	192	100	932.8	1111.0	75.4	8440.3	880.5	1008.2	73.5	7402.2

Mode	Grade	Item Grouping	No. of		G^2				Q_1			
			Score Points	No. of Items	Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	A1	SSMC	31	31	682.3	845.2	104.5	4235.3	663.9	844.1	103.1	4264.5
CBT	A1	CR	161	69	1045.4	1200.2	75.4	8440.3	977.9	1065.2	73.5	7402.2
CBT	A1	Type I	118	81	747.5	742.4	75.4	4235.3	709.4	715.8	73.5	4264.5
CBT	A1	Type II	35	10	1025.4	967.9	331.6	3613.1	1009.5	838.4	284.2	3113.7
CBT	A1	Type III	39	9	2498.0	2365.5	745.3	8440.3	2277.6	2037.7	769.6	7402.2
CBT	GO	All Items	203	109	537.5	509.7	15.4	2455.1	534.1	554.5	15.3	3491.6
CBT	GO	SSMC	27	27	335.4	316.8	76.5	1252.4	328.5	315.0	75.2	1200.9
CBT	GO	CR	176	82	604.1	544.1	15.4	2455.1	601.8	599.5	15.3	3491.6
CBT	GO	Type I	126	90	453.5	439.2	15.4	2041.2	431.9	417.8	15.3	1706.1
CBT	GO	Type II	32	9	703.2	475.8	111.1	1571.1	950.5	1029.7	118.3	3491.6
CBT	GO	Type III	45	10	1144.6	707.9	427.3	2455.1	1079.6	642.3	385.6	2232.5
CBT	A2	All Items	198	101	641.4	738.2	34.0	3247.4	592.6	693.4	32.8	3211.1
CBT	A2	SSMC	17	17	332.4	218.0	97.2	921.1	315.5	207.7	97.2	910.4
CBT	A2	CR	181	84	704.0	789.8	34.0	3247.4	648.7	743.0	32.8	3211.1
CBT	A2	Type I	121	82	516.6	660.8	34.0	3247.4	485.3	627.7	32.8	3211.1
CBT	A2	Type II	35	10	1208.5	914.0	250.6	3184.8	1092.0	850.4	375.5	3054.9
CBT	A2	Type III	42	9	1149.0	773.9	146.9	2828.8	1015.9	759.0	124.1	2759.9
CBT	M1	All Items	81	42	165.2	145.5	18.6	780.0	288.4	852.7	18.1	5611.7
CBT	M1	SSMC	13	13	104.7	85.7	18.6	283.0	99.4	80.1	18.1	260.7

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	SD	Min	Max	Mean	SD	Min	Max
CBT	M1	CR	68	29	192.3	159.3	23.4	780.0	373.2	1018.8	23.0	5611.7
CBT	M1	Type I	49	34	132.5	103.2	18.6	374.9	122.1	91.4	18.1	345.0
CBT	M1	Type II	14	4	199.6	115.1	95.0	357.7	1562.6	2700.9	117.4	5611.7
CBT	M1	Type III	18	4	408.7	256.9	207.4	780.0	427.8	215.3	177.6	696.1
CBT	M2	All Items	80	41	57.3	55.9	9.2	257.6	54.5	56.2	9.1	234.1
CBT	M2	SSMC	14	14	30.5	12.8	9.3	51.9	29.8	12.4	9.1	51.0
CBT	M2	CR	66	27	71.2	64.4	9.2	257.6	67.3	65.5	9.5	234.1
CBT	M2	Type I	48	33	40.5	36.6	9.2	199.0	37.4	34.9	9.1	193.1
CBT	M2	Type II	14	4	84.5	50.8	32.1	151.2	80.7	53.4	28.6	149.7
CBT	M2	Type III	18	4	168.2	66.0	101.4	257.6	169.4	67.9	104.9	234.1
CBT	M3	All Items	81	40	37.2	28.6	7.6	125.8	31.9	23.1	7.0	102.3
CBT	M3	SSMC	12	12	24.5	12.1	8.0	41.5	23.5	11.6	7.9	42.4
CBT	M3	CR	69	28	42.7	32.0	7.6	125.8	35.5	25.9	7.0	102.3
CBT	M3	Type I	49	32	27.0	14.5	7.6	75.2	24.0	13.0	7.0	53.3
CBT	M3	Type II	14	4	67.0	28.3	26.2	91.6	50.6	23.2	20.8	72.3
CBT	M3	Type III	18	4	89.6	41.7	48.6	125.8	76.2	30.2	35.1	102.3

Note: M03 through M08 = mathematics grades 3 through 8, A1 = Algebra I, GO = Geometry, A2 = Algebra II, M1 = Integrated Mathematics I, M2 = Integrated Mathematics II, M3 = Integrated Mathematics III.

Table A.10.12 PBT IRT Model Fit for All Items for Mathematics by Grade/Subject

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	SD	Min	Max	Mean	SD	Min	Max
PBT	M03	All Items	111	75	257.8	265.3	33.1	1417.6	248.6	256.9	32.7	1368.0
PBT	M03	SSMC	30	30	174.5	205.9	33.1	750.8	174.5	212.2	32.7	755.8
PBT	M03	CR	81	45	313.4	287.3	34.0	1417.6	298.0	274.0	32.8	1368.0
PBT	M03	Type I	72	64	227.5	258.3	33.1	1417.6	222.5	254.4	32.7	1368.0
PBT	M03	Type II	21	6	405.1	211.6	245.6	805.1	376.6	188.4	235.2	730.8
PBT	M03	Type III	18	5	469.2	304.1	141.0	859.4	429.4	284.0	128.8	801.1
PBT	M04	All Items	109	65	272.4	302.9	28.0	1383.4	266.9	303.9	26.4	1356.7
PBT	M04	SSMC	24	24	155.0	221.9	29.0	1043.3	160.8	254.2	28.9	1214.0
PBT	M04	CR	85	41	341.2	324.6	28.0	1383.4	329.0	316.0	26.4	1356.7
PBT	M04	Type I	67	54	228.5	288.4	28.0	1383.4	223.2	291.0	26.4	1356.7
PBT	M04	Type II	24	7	493.8	331.1	161.6	1109.4	493.9	327.5	163.3	1078.9
PBT	M04	Type III	18	4	478.7	249.8	317.4	850.2	459.9	233.7	314.4	807.5
PBT	M05	All Items	115	69	211.3	173.7	19.5	742.4	199.3	160.7	19.0	680.1
PBT	M05	SSMC	28	28	140.4	99.0	35.0	450.8	135.9	96.8	33.5	431.6
PBT	M05	CR	87	41	259.7	196.8	19.5	742.4	242.6	181.2	19.0	680.1
PBT	M05	Type I	70	57	183.7	153.2	19.5	669.0	174.7	143.0	19.0	601.8
PBT	M05	Type II	21	6	360.7	226.2	155.8	742.4	333.8	206.3	139.6	680.1
PBT	M05	Type III	24	6	323.7	213.0	78.3	652.3	298.9	197.5	69.5	599.6

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	SD	Min	Max	Mean	SD	Min	Max
PBT	M06	All Items	110	63	268.8	279.8	19.2	1646.4	256.0	277.3	18.8	1648.5
PBT	M06	SSMC	23	23	178.9	200.2	29.9	774.0	177.6	205.4	28.7	818.0
PBT	M06	CR	87	40	320.5	307.2	19.2	1646.4	301.0	304.6	18.8	1648.5
PBT	M06	Type I	67	52	241.1	289.2	19.2	1646.4	231.1	287.9	18.8	1648.5
PBT	M06	Type II	25	7	354.5	207.3	121.2	717.4	335.0	209.3	110.2	705.0
PBT	M06	Type III	18	4	479.3	143.3	286.7	620.6	440.3	151.2	261.4	631.1
PBT	M07	All Items	102	65	188.6	250.0	15.6	1441.4	173.3	223.1	15.4	1282.4
PBT	M07	SSMC	26	26	107.8	104.0	15.6	501.0	104.1	99.6	15.4	480.2
PBT	M07	CR	76	39	242.4	301.2	21.1	1441.4	219.4	268.0	21.3	1282.4
PBT	M07	Type I	73	57	144.7	181.2	15.6	919.9	135.8	164.5	15.4	852.1
PBT	M07	Type II	14	4	428.4	216.2	176.0	657.6	377.3	188.9	159.4	597.2
PBT	M07	Type III	15	4	574.7	601.8	71.2	1441.4	503.0	541.8	52.0	1282.4
PBT	M08	All Items	114	63	242.9	281.6	15.0	1486.4	226.3	261.0	10.7	1309.9
PBT	M08	SSMC	25	25	160.4	185.9	24.7	902.8	164.7	208.6	24.4	998.0
PBT	M08	CR	89	38	297.2	320.7	15.0	1486.4	266.9	285.8	10.7	1309.9
PBT	M08	Type I	75	53	181.4	189.1	15.0	902.8	173.3	189.2	10.7	998.0
PBT	M08	Type II	18	5	620.0	540.7	94.4	1486.4	561.4	482.8	86.8	1309.9
PBT	M08	Type III	21	5	517.2	387.7	161.7	1115.3	452.9	340.7	145.3	988.1
PBT	A1	All Items	128	69	126.5	152.1	10.5	896.8	123.8	147.9	10.4	829.3

Mode	Grade	Item Grouping	No. of Score Points	No. of Items	G^2				Q_1			
					Mean	SD	Min	Max	Mean	SD	Min	Max
PBT	A1	SSMC	26	26	66.7	69.3	11.0	367.3	65.1	69.8	11.0	370.2
PBT	A1	CR	102	43	162.7	176.2	10.5	896.8	159.2	170.5	10.4	829.3
PBT	A1	Type I	83	57	99.4	113.6	10.5	678.3	95.8	113.2	10.4	699.2
PBT	A1	Type II	24	7	134.3	103.9	62.9	359.9	160.6	124.3	52.2	343.1
PBT	A1	Type III	21	5	424.9	273.2	241.7	896.8	390.7	258.3	220.6	829.3
PBT	GO	All Items	139	79	45.9	44.2	4.6	221.6	41.7	39.7	4.3	197.7
PBT	GO	SSMC	26	26	35.8	30.3	6.9	126.2	35.3	30.2	6.8	126.0
PBT	GO	CR	113	53	50.9	49.1	4.6	221.6	44.8	43.5	4.3	197.7
PBT	GO	Type I	91	67	36.9	31.4	4.6	126.2	33.7	29.3	4.3	126.0
PBT	GO	Type II	21	6	68.4	50.8	32.3	170.9	63.2	46.8	27.5	146.8
PBT	GO	Type III	27	6	124.1	76.6	50.3	221.6	109.2	64.5	36.4	197.7
PBT	A2	All Items	143	72	70.9	66.7	6.9	322.2	71.8	99.9	6.3	746.5
PBT	A2	SSMC	18	18	43.3	38.6	13.0	171.6	42.2	38.2	12.7	169.1
PBT	A2	CR	125	54	80.1	71.7	6.9	322.2	81.7	111.8	6.3	746.5
PBT	A2	Type I	83	57	60.5	61.2	6.9	322.2	62.9	102.1	6.3	746.5
PBT	A2	Type II	24	7	97.4	63.3	26.1	170.7	114.2	104.5	19.8	318.6
PBT	A2	Type III	36	8	121.9	84.9	28.5	230.3	98.1	71.6	28.8	193.9

Note: M03 through M08 = mathematics grades 3 through 8, A1 = Algebra I, GO = Geometry, A2 = Algebra II.

Appendix 12.1: Form Composition

Table A.12.1 Form Composition for ELA/L Grade 3

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	10 - 12	20 - 24
	Reading Informational Text	8 - 9	16 - 18
	Vocabulary	5 - 8	10 - 16
	Claim Total	26	52
Writing			
	Written Expression	2	30
	Knowledge of Conventions	1	12
	Claim Total	3	42
SUMMATIVE TOTAL		29	94

Note: This table is identical to Table 12.1 in Section 12.

Table A.12.2 Form Composition for ELA/L Grade 4

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	10 - 11	20 - 22
	Reading Informational Text	10 - 12	20 - 24
	Vocabulary	5 - 8	10 - 16
	Claim Total	28	56
Writing			
	Written Expression	2	38
	Knowledge of Conventions	1	12
	Claim Total	3	50
SUMMATIVE TOTAL		31	106

Table A.12.3 Form Composition for ELA/L Grade 5

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	9 - 11	18 - 22
	Reading Informational Text	9 - 11	18 - 22
	Vocabulary	7 - 8	14 - 16
	Claim Total	28	56
Writing			
	Written Expression	2	38
	Knowledge of Conventions	1	12
	Claim Total	3	50
SUMMATIVE TOTAL		31	106

Table A.12.4 Form Composition for ELA/L Grade 6

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	9 - 11	18 - 22
	Reading Informational Text	14 - 16	28 - 32
	Vocabulary	7 - 9	14 - 18
	Claim Total	34	68
Writing			
	Written Expression	2	38
	Knowledge of Conventions	1	15
	Claim Total	3	53
SUMMATIVE TOTAL		37	121

Table A.12.5 Form Composition for ELA/L Grade 7

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	7 - 13	14 - 26
	Reading Informational Text	13 - 16	26 - 32
	Vocabulary	8 - 11	16 - 22
	Claim Total	34	68
Writing			
	Written Expression	2	38
	Knowledge of Conventions	1	15
	Claim Total	3	53
SUMMATIVE TOTAL		37	121

Table A.12.6 Form Composition for ELA/L Grade 8

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	7 - 12	14 - 24
	Reading Informational Text	12 - 19	24 - 38
	Vocabulary	7 - 10	14 - 20
	Claim Total	34	68
Writing			
	Written Expression	2	38
	Knowledge of Conventions	1	15
	Claim Total	3	53
SUMMATIVE TOTAL		37	121

Table A.12.7 Form Composition for ELA/L Grade 9

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	8 - 15	16 - 30
	Reading Informational Text	12 - 18	24 - 36
	Vocabulary	7 - 9	14 - 18
	Claim Total	34	68
Writing			
	Written Expression	2	38
	Knowledge of Conventions	1	15
	Claim Total	3	53
SUMMATIVE TOTAL		37	121

Table A.12.8 Form Composition for ELA/L Grade 10

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	8 - 14	16 - 28
	Reading Informational Text	13 - 19	26 - 38
	Vocabulary	7 - 9	14 - 18
	Claim Total	34	68
Writing			
	Written Expression	2	38
	Knowledge of Conventions	1	15
	Claim Total	3	53
SUMMATIVE TOTAL		37	121

Table A.12.9 Form Composition for ELA/L Grade 11

Claims	Subclaims	Number of Items	Number of Points
Reading			
	Reading Literary Text	8 - 14	16 - 28
	Reading Informational Text	13 - 19	26 - 38
	Vocabulary	7 - 9	14 - 18
	Claim Total	34	68
Writing			
	Written Expression	2	38
	Knowledge of Conventions	1	15
	Claim Total	3	53
SUMMATIVE TOTAL		37	121

Table A.12.10 Form Composition for Mathematics Grade 3

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	26	30
	Additional & Supporting Content	10	10
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		43	66

Note: This table is identical to Table 12.3 in Section 12.

Table A.12.11 Form Composition for Mathematics Grade 4

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	25	31
	Additional & Supporting Content	8	9
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		40	66

Table A.12.12 Form Composition for Mathematics Grade 5

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	25	30
	Additional & Supporting Content	8	10
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		40	66

Table A.12.13 Form Composition for Mathematics Grade 6

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	20	26
	Additional & Supporting Content	11	14
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		38	66

Table A.12.14 Form Composition for Mathematics Grade 7

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	23	29
	Additional & Supporting Content	8	11
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		38	66

Table A.12.15 Form Composition for Mathematics Grade 8

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	21	27
	Additional & Supporting Content	8	13
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	3	12
TOTAL		36	66

Table A.12.16 Form Composition for Algebra I

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	21	28
	Additional & Supporting Content	13	21
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		42	81

Table A.12.17 Form Composition for Geometry

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	21	30
	Additional & Supporting Content	14	19
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		43	81

Table A.12.18 Form Composition for Algebra II

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	20	29
	Additional & Supporting Content	13	20
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		41	81

Table A.12.19 Form Composition for Integrated Mathematics I

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	21	31
	Additional & Supporting Content	13	18
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		42	81

Table A.12.20 Form Composition for Integrated Mathematics II

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	22	32
	Additional & Supporting Content	12	17
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		42	81

Table A.12.21 Form Composition for Integrated Mathematics III

	Subclaims	Number of Items	Number of Points
Mathematics			
	Major Content	19	26
	Additional & Supporting Content	13	23
	Expressing Mathematical Reasoning	4	14
	Modeling and Applications	4	18
TOTAL		40	81

Appendix 12.2: Scaling Constants and Associated Information

Table A.12.22 Threshold Scores and Scaling Constants for ELA/L Grades 3 to 8

PARCC Assessment	Threshold Cut	Theta	Scale Score	A	B
Grade 3 ELA	Level 2 Cut	-0.9648	700	36.7227	735.4297
	Level 3 Cut	-0.2840	726		
	Level 4 Cut	0.3968	750		
	Level 5 Cut	2.0360	810		
Grade 4 ELA	Level 2 Cut	-1.3004	700	31.5462	741.0214
	Level 3 Cut	-0.5079	725		
	Level 4 Cut	0.2846	750		
	Level 5 Cut	1.5578	790		
Grade 5 ELA	Level 2 Cut	-1.3411	700	29.4580	739.5050
	Level 3 Cut	-0.4924	726		
	Level 4 Cut	0.3563	750		
	Level 5 Cut	2.0224	799		
Grade 6 ELA	Level 2 Cut	-1.3656	700	28.3160	738.6673
	Level 3 Cut	-0.4827	725		
	Level 4 Cut	0.4002	750		
	Level 5 Cut	1.8133	790		
Grade 7 ELA	Level 2 Cut	-1.2488	700	33.9161	742.3542
	Level 3 Cut	-0.5117	725		
	Level 4 Cut	0.2254	750		
	Level 5 Cut	1.2614	785		
Grade 8 ELA	Level 2 Cut	-1.2730	700	34.1183	743.4330
	Level 3 Cut	-0.5402	725		
	Level 4 Cut	0.1925	750		
	Level 5 Cut	1.4696	794		

Table A.12.23 Threshold Scores and Scaling Constants for Mathematics Grades 3 to 8

PARCC Assessment	Threshold Cut	Theta	Scale Score	A	B
Grade 3 Mathematics	Level 2 Cut	-1.4141	700	32.1135	745.4119
	Level 3 Cut	-0.6356	727		
	Level 4 Cut	0.1429	750		
	Level 5 Cut	1.3931	790		
Grade 4 Mathematics	Level 2 Cut	-1.3840	700	29.9167	741.4049
	Level 3 Cut	-0.5484	727		
	Level 4 Cut	0.2873	750		
	Level 5 Cut	1.8323	796		
Grade 5 Mathematics	Level 2 Cut	-1.4571	700	29.0301	742.2997
	Level 3 Cut	-0.5959	725		
	Level 4 Cut	0.2653	750		
	Level 5 Cut	1.6262	790		
Grade 6 Mathematics	Level 2 Cut	-1.3829	700	28.1465	738.9252
	Level 3 Cut	-0.4948	725		
	Level 4 Cut	0.3935	750		
	Level 5 Cut	1.7567	788		
Grade 7 Mathematics	Level 2 Cut	-1.4464	700	25.1033	736.3102
	Level 3 Cut	-0.4505	725		
	Level 4 Cut	0.5453	750		
	Level 5 Cut	1.9919	786		
Grade 8 Mathematics	Level 2 Cut	-0.8851	700	32.9505	729.1640
	Level 3 Cut	-0.1264	728		
	Level 4 Cut	0.6323	750		
	Level 5 Cut	2.1896	801		

Table A.12.24 Threshold Scores and Scaling Constants for High School ELA

PARCC Assessment	Threshold Cut	Theta	Scale Score	<i>A</i>	<i>B</i>
Grade 9 ELA	Level 2 Cut	-1.1635	700	34.2174	739.8124
	Level 3 Cut	-0.4329	726		
	Level 4 Cut	0.2977	750		
	Level 5 Cut	1.5065	791		
Grade 10 ELA	Level 2 Cut	-0.8909	700	43.1280	738.4223
	Level 3 Cut	-0.3112	725		
	Level 4 Cut	0.2684	750		
	Level 5 Cut	1.2858	794		
Grade 11 ELA	Level 2 Cut	-1.1017	700	34.9278	738.4801
	Level 3 Cut	-0.3859	726		
	Level 4 Cut	0.3298	750		
	Level 5 Cut	1.5206	792		

Table A.12.25 Threshold Scores and Scaling Constants for High School Mathematics

PARCC Assessment	Threshold Cut	Theta	Scale Score	<i>A</i>	<i>B</i>
Algebra I	Level 2 Cut	-1.1781	700	31.5325	737.1490
	Level 3 Cut	-0.3853	728		
	Level 4 Cut	0.4075	750		
	Level 5 Cut	2.1651	805		
Algebra II	Level 2 Cut	-0.5759	700	37.7676	721.7509
	Level 3 Cut	0.0860	726		
	Level 4 Cut	0.7480	750		
	Level 5 Cut	2.2728	808		
Geometry	Level 2 Cut	-1.3013	700	25.9775	733.8039
	Level 3 Cut	-0.3389	726		
	Level 4 Cut	0.6235	750		
	Level 5 Cut	1.8940	783		
Integrated Mathematics I	Level 2 Cut	-1.0919	700	32.0043	734.9446
	Level 3 Cut	-0.3107	726		
	Level 4 Cut	0.4704	750		
	Level 5 Cut	1.9934	799		
Integrated Mathematics II	Level 2 Cut	-0.9175	700	29.2865	726.8695
	Level 3 Cut	-0.0638	725		
	Level 4 Cut	0.7898	750		
	Level 5 Cut	1.9817	785		
Integrated Mathematics III	Level 2 Cut	-0.7076	700	37.3549	726.4336
	Level 3 Cut	-0.0384	726		
	Level 4 Cut	0.6309	750		
	Level 5 Cut	2.0689	804		

Table A.12.26 Scaling Constants for Reading and Writing Grades 3 to 8

	Reading		Writing	
	A_R	B_R	A_W	B_W
Grade 3 ELA	14.6891	44.1719	7.3445	32.0859
Grade 4 ELA	12.6184	46.4086	6.3093	33.2043
Grade 5 ELA	11.7832	45.8019	5.8916	32.9010
Grade 6 ELA	11.3264	45.4669	5.6632	32.7335
Grade 7 ELA	13.5664	46.9416	6.7832	33.4708
Grade 8 ELA	13.6472	47.3732	6.8237	33.6866
Grade 9 ELA	13.6870	45.9250	6.8435	32.9625
Grade 10 ELA	17.2512	45.3690	8.6256	32.6845
Grade 11 ELA	13.9712	45.3920	6.9856	32.6961

Appendix 12.3: Raw-to-Scale Conversion Tables for Performance Level Setting (PLS) Forms

Table A.12.27 Conversion Table for Performance Level Setting Form: ELA/L Grade 3

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-6.0000	2.5	518.0432	650	15
1	-5.7004	2.3	528.8959	650	15
2	-4.0919	1.2	587.1622	650	15
3	-3.3588	0.9	613.7180	650	15
4	-2.9022	0.7	630.2579	650	15
5	-2.5750	0.6	642.1104	650	15
6	-2.3214	0.5	651.2968	651	15
7	-2.1144	0.5	658.7952	659	15
8	-1.9395	0.4	665.1308	665	15
9	-1.7878	0.4	670.6259	671	14.8
10	-1.6535	0.4	675.4908	675	14.0
11	-1.5327	0.4	679.8667	680	13.2
12	-1.4227	0.3	683.8513	684	12.6
13	-1.3214	0.3	687.5208	688	12.1
14	-1.2273	0.3	690.9295	691	11.7
15	-1.1391	0.3	694.1244	694	11.3
16	-1.0559	0.3	697.1383	697	11.0
17	-0.9769	0.3	700.0000	700	10.7
18	-0.9016	0.3	702.7276	703	10.5
19	-0.8294	0.3	705.3430	705	10.2
20	-0.7599	0.3	707.8606	708	10.1
21	-0.6926	0.3	710.2985	710	9.9
22	-0.6274	0.3	712.6603	713	9.8
23	-0.5639	0.3	714.9605	715	9.6
24	-0.5018	0.3	717.2100	717	9.5
25	-0.4411	0.3	719.4088	719	9.4
26	-0.3814	0.3	721.5714	722	9.4
27	-0.3227	0.3	723.6977	724	9.3
28	-0.2648	0.3	725.7951	726	9.3
29	-0.2075	0.3	727.8707	728	9.2
30	-0.1509	0.3	729.9210	730	9.2
31	-0.0946	0.3	731.9604	732	9.2
32	-0.0388	0.3	733.9817	734	9.2
33	0.0167	0.3	735.9921	736	9.2
34	0.0721	0.3	737.9990	738	9.2
35	0.1273	0.3	739.9985	740	9.2
36	0.1825	0.3	741.9981	742	9.2
37	0.2376	0.3	743.9940	744	9.3
38	0.2928	0.3	745.9936	746	9.3
39	0.3480	0.3	747.9932	748	9.3
40	0.4034	0.3	750.0000	750	9.4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
41	0.4590	0.3	752.0140	752	9.4
42	0.5149	0.3	754.0389	754	9.5
43	0.5709	0.3	756.0675	756	9.5
44	0.6273	0.3	758.1105	758	9.6
45	0.6840	0.3	760.1644	760	9.7
46	0.7410	0.3	762.2292	762	9.7
47	0.7983	0.3	764.3048	764	9.8
48	0.8561	0.3	766.3986	766	9.9
49	0.9142	0.3	768.5032	769	9.9
50	0.9728	0.3	770.6259	771	10.0
51	1.0318	0.3	772.7631	773	10.1
52	1.0913	0.3	774.9185	775	10.2
53	1.1513	0.3	777.0919	777	10.2
54	1.2118	0.3	779.2834	779	10.3
55	1.2728	0.3	781.4931	781	10.4
56	1.3344	0.3	783.7245	784	10.5
57	1.3966	0.3	785.9776	786	10.6
58	1.4595	0.3	788.2561	788	10.7
59	1.5230	0.3	790.5564	791	10.8
60	1.5873	0.3	792.8856	793	10.9
61	1.6525	0.3	795.2474	795	11.0
62	1.7185	0.3	797.6381	798	11.1
63	1.7855	0.3	800.0652	800	11.2
64	1.8536	0.3	802.5320	803	11.3
65	1.9228	0.3	805.0387	805	11.4
66	1.9933	0.3	807.5925	808	11.6
67	2.0652	0.3	810.1970	810	11.7
68	2.1386	0.3	812.8558	813	11.9
69	2.2137	0.3	815.5763	816	12.0
70	2.2906	0.3	818.3619	818	12.2
71	2.3695	0.3	821.2200	821	12.4
72	2.4505	0.3	824.1541	824	12.6
73	2.5339	0.4	827.1752	827	12.8
74	2.6199	0.4	830.2905	830	13.0
75	2.7088	0.4	833.5108	834	13.3
76	2.8007	0.4	836.8398	837	13.6
77	2.8960	0.4	840.2919	840	13.9
78	2.9952	0.4	843.8853	844	14.2
79	3.0985	0.4	847.6273	848	14.5
80	3.2064	0.4	851.5358	850	14.5
81	3.3197	0.4	855.6400	850	14.5
82	3.4389	0.4	859.9579	850	14.5
83	3.5651	0.5	864.5294	850	14.5
84	3.6992	0.5	869.3870	850	14.5

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
85	3.8427	0.5	874.5852	850	14.5
86	3.9975	0.5	880.1926	850	14.5
87	4.1660	0.5	886.2964	850	14.5
88	4.3514	0.6	893.0123	850	14.5
89	4.5582	0.6	900.5034	850	14.5
90	4.7929	0.7	909.0052	850	14.5
91	5.0649	0.7	918.8581	850	14.5
92	5.3888	0.8	930.5911	850	14.5
93	5.7881	1	945.0553	850	14.5
94	6.3022	1.1	963.6781	850	14.5
95	6.9979	1.4	988.8791	850	14.5
96	7.9842	1.8	1024.6069	850	14.5
97	9.4219	2.4	1076.6861	850	14.5
98	10.000	2.7	1097.6272	850	14.5
99	10.000	2.7	1097.6272	850	14.5
100	10.000	2.7	1097.6272	850	14.5

Table A.12.28 Conversion Table for Performance Level Setting Form: ELA/L Grade 4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-6.0000	1.9	556.1541	650	15
1	-5.5863	1.6	568.8904	650	15
2	-4.4430	1.0	604.0884	650	15
3	-3.8331	0.8	622.8649	650	15
4	-3.4200	0.7	635.5828	650	15
5	-3.1085	0.6	645.1727	650	15
6	-2.8589	0.5	652.8569	653	15
7	-2.6509	0.5	659.2605	659	15
8	-2.4726	0.5	664.7497	665	14.1
9	-2.3165	0.4	669.5554	670	13.2
10	-2.1776	0.4	673.8317	674	12.5
11	-2.0523	0.4	677.6892	678	11.9
12	-1.9381	0.4	681.2050	681	11.4
13	-1.8330	0.4	684.4406	684	10.9
14	-1.7355	0.3	687.4423	687	10.5
15	-1.6443	0.3	690.2500	690	10.2
16	-1.5587	0.3	692.8853	693	9.9
17	-1.4778	0.3	695.3759	695	9.6
18	-1.4009	0.3	697.7434	698	9.4
19	-1.3276	0.3	700.0000	700	9.2
20	-1.2575	0.3	702.1581	702	9.0
21	-1.1901	0.3	704.2331	704	8.8
22	-1.1251	0.3	706.2342	706	8.7
23	-1.0622	0.3	708.1707	708	8.5
24	-1.0013	0.3	710.0456	710	8.4
25	-0.9422	0.3	711.8650	712	8.3
26	-0.8845	0.3	713.6414	714	8.2
27	-0.8282	0.3	715.3747	715	8.1
28	-0.7732	0.3	717.0679	717	8.0
29	-0.7192	0.3	718.7304	719	8.0
30	-0.6662	0.3	720.3621	720	7.9
31	-0.6142	0.3	721.9630	722	7.8
32	-0.5628	0.3	723.5454	724	7.8
33	-0.5122	0.3	725.1032	725	7.7
34	-0.4623	0.2	726.6394	727	7.7
35	-0.4128	0.2	728.1633	728	7.7
36	-0.3639	0.2	729.6688	730	7.6
37	-0.3154	0.2	731.1619	731	7.6
38	-0.2673	0.2	732.6427	733	7.6
39	-0.2194	0.2	734.1174	734	7.6
40	-0.1719	0.2	735.5797	736	7.5
41	-0.1246	0.2	737.0359	737	7.5
42	-0.0775	0.2	738.4860	738	7.5

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	-0.0305	0.2	739.9329	740	7.5
44	0.0163	0.2	741.3737	741	7.5
45	0.0631	0.2	742.8145	743	7.5
46	0.1098	0.2	744.2522	744	7.5
47	0.1564	0.2	745.6869	746	7.5
48	0.2031	0.2	747.1246	747	7.5
49	0.2498	0.2	748.5623	749	7.6
50	0.2965	0.2	750.0000	750	7.6
51	0.3433	0.2	751.4408	751	7.6
52	0.3902	0.2	752.8847	753	7.6
53	0.4372	0.2	754.3317	754	7.6
54	0.4844	0.2	755.7848	756	7.6
55	0.5317	0.2	757.2410	757	7.7
56	0.5791	0.2	758.7002	759	7.7
57	0.6268	0.3	760.1688	760	7.7
58	0.6746	0.3	761.6403	762	7.7
59	0.7227	0.3	763.1212	763	7.7
60	0.7710	0.3	764.6081	765	7.8
61	0.8196	0.3	766.1044	766	7.8
62	0.8685	0.3	767.6098	768	7.8
63	0.9177	0.3	769.1245	769	7.9
64	0.9673	0.3	770.6515	771	7.9
65	1.0172	0.3	772.1877	772	7.9
66	1.0675	0.3	773.7363	774	8.0
67	1.1182	0.3	775.2971	775	8.0
68	1.1694	0.3	776.8734	777	8.1
69	1.2211	0.3	778.4651	778	8.1
70	1.2733	0.3	780.0721	780	8.1
71	1.3262	0.3	781.7007	782	8.2
72	1.3796	0.3	783.3447	783	8.3
73	1.4338	0.3	785.0133	785	8.3
74	1.4887	0.3	786.7035	787	8.4
75	1.5445	0.3	788.4213	788	8.4
76	1.6011	0.3	790.1638	790	8.5
77	1.6588	0.3	791.9402	792	8.6
78	1.7175	0.3	793.7474	794	8.7
79	1.7775	0.3	795.5945	796	8.8
80	1.8388	0.3	797.4817	797	8.9
81	1.9016	0.3	799.4151	799	9.0
82	1.9661	0.3	801.4008	801	9.1
83	2.0324	0.3	803.4420	803	9.3
84	2.1008	0.3	805.5478	806	9.4
85	2.1715	0.3	807.7244	808	9.6
86	2.2448	0.3	809.9810	810	9.8

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	2.3212	0.3	812.3331	812	10.0
88	2.4009	0.3	814.7867	815	10.3
89	2.4846	0.3	817.3635	817	10.6
90	2.5728	0.4	820.0789	820	10.9
91	2.6663	0.4	822.9574	823	11.2
92	2.7659	0.4	826.0237	826	11.6
93	2.8730	0.4	829.3209	829	12.1
94	2.9889	0.4	832.8891	833	12.7
95	3.1157	0.4	836.7928	837	13.3
96	3.2558	0.5	841.1059	841	14.1
97	3.4127	0.5	845.9363	846	15
98	3.5915	0.5	851.4409	850	15
99	3.7993	0.6	857.8383	850	15
100	4.0472	0.6	865.4702	850	15
101	4.3532	0.7	874.8908	850	15
102	4.7494	0.9	887.0884	850	15
103	5.3000	1.1	904.0393	850	15
104	6.1618	1.5	930.5709	850	15
105	7.8929	2.5	983.8651	850	15
106	10.000	4.2	1048.7349	850	15

Table A.12.29 Conversion Table for Performance Level Setting Form: ELA/L Grade 5

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-6.0000	2.2	563.7589	650	15
1	-5.7665	2.0	570.6399	650	15
2	-4.4023	1.3	610.8415	650	15
3	-3.6956	0.9	631.6673	650	15
4	-3.2311	0.8	645.3556	650	15
5	-2.8900	0.7	655.4075	655	15
6	-2.6225	0.6	663.2904	663	15
7	-2.4033	0.5	669.7501	670	15
8	-2.2179	0.5	675.2136	675	14.3
9	-2.0573	0.5	679.9463	680	13.3
10	-1.9155	0.4	684.1250	684	12.5
11	-1.7883	0.4	687.8735	688	11.8
12	-1.6728	0.4	691.2772	691	11.2
13	-1.5667	0.4	694.4038	694	10.7
14	-1.4685	0.4	697.2977	697	10.3
15	-1.3768	0.3	700.0000	700	10.0
16	-1.2906	0.3	702.5402	703	9.7
17	-1.2091	0.3	704.9419	705	9.4
18	-1.1316	0.3	707.2258	707	9.2
19	-1.0576	0.3	709.4065	709	9.0
20	-0.9867	0.3	711.4958	711	8.8
21	-0.9185	0.3	713.5056	714	8.6
22	-0.8526	0.3	715.4476	715	8.5
23	-0.7888	0.3	717.3278	717	8.3
24	-0.7268	0.3	719.1548	719	8.2
25	-0.6665	0.3	720.9318	721	8.1
26	-0.6077	0.3	722.6646	723	8.0
27	-0.5502	0.3	724.3591	724	7.9
28	-0.4939	0.3	726.0182	726	7.9
29	-0.4386	0.3	727.6478	728	7.8
30	-0.3842	0.3	729.2509	729	7.7
31	-0.3307	0.3	730.8275	731	7.7
32	-0.2780	0.3	732.3805	732	7.6
33	-0.2259	0.3	733.9159	734	7.6
34	-0.1744	0.3	735.4335	735	7.6
35	-0.1235	0.3	736.9335	737	7.5
36	-0.0730	0.3	738.4217	738	7.5
37	-0.0229	0.3	739.8981	740	7.5
38	0.0268	0.3	741.3627	741	7.5
39	0.0762	0.3	742.8184	743	7.4
40	0.1253	0.3	744.2654	744	7.4
41	0.1742	0.3	745.7064	746	7.4
42	0.2229	0.3	747.1415	747	7.4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	0.2715	0.3	748.5737	749	7.4
44	0.3199	0.3	750.0000	750	7.4
45	0.3683	0.3	751.4263	751	7.4
46	0.4165	0.3	752.8467	753	7.4
47	0.4648	0.3	754.2701	754	7.4
48	0.5129	0.3	755.6876	756	7.4
49	0.5611	0.3	757.1080	757	7.4
50	0.6093	0.3	758.5284	759	7.4
51	0.6575	0.3	759.9488	760	7.4
52	0.7058	0.3	761.3721	761	7.4
53	0.7541	0.3	762.7955	763	7.4
54	0.8024	0.3	764.2188	764	7.4
55	0.8509	0.3	765.6481	766	7.4
56	0.8995	0.3	767.0803	767	7.4
57	0.9482	0.3	768.5154	769	7.4
58	0.9970	0.3	769.9535	770	7.4
59	1.0459	0.3	771.3945	771	7.5
60	1.0950	0.3	772.8415	773	7.5
61	1.1443	0.3	774.2943	774	7.5
62	1.1939	0.3	775.7559	776	7.5
63	1.2436	0.3	777.2205	777	7.5
64	1.2936	0.3	778.6940	779	7.5
65	1.3439	0.3	780.1763	780	7.6
66	1.3945	0.3	781.6674	782	7.6
67	1.4455	0.3	783.1703	783	7.6
68	1.4968	0.3	784.6821	785	7.7
69	1.5486	0.3	786.2086	786	7.7
70	1.6009	0.3	787.7498	788	7.7
71	1.6537	0.3	789.3058	789	7.8
72	1.7071	0.3	790.8794	791	7.8
73	1.7611	0.3	792.4708	792	7.9
74	1.8159	0.3	794.0857	794	7.9
75	1.8715	0.3	795.7241	796	8.0
76	1.9279	0.3	797.3862	797	8.0
77	1.9854	0.3	799.0807	799	8.1
78	2.0439	0.3	800.8046	801	8.2
79	2.1036	0.3	802.5639	803	8.3
80	2.1647	0.3	804.3644	804	8.4
81	2.2273	0.3	806.2092	806	8.5
82	2.2915	0.3	808.1011	808	8.6
83	2.3576	0.3	810.0490	810	8.8
84	2.4259	0.3	812.0617	812	8.9
85	2.4966	0.3	814.1452	814	9.1
86	2.5701	0.3	816.3112	816	9.3

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	2.6468	0.3	818.5714	819	9.5
88	2.7272	0.3	820.9408	821	9.8
89	2.8118	0.3	823.4338	823	10.1
90	2.9016	0.4	826.0802	826	10.4
91	2.9974	0.4	828.9033	829	10.8
92	3.1004	0.4	831.9386	832	11.3
93	3.2123	0.4	835.2362	835	11.9
94	3.3350	0.4	838.8520	839	12.6
95	3.4714	0.5	842.8716	843	13.4
96	3.6252	0.5	847.4039	847	14.4
97	3.8017	0.5	852.6052	850	14.4
98	4.0085	0.6	858.6994	850	14.4
99	4.2567	0.7	866.0136	850	14.4
100	4.5635	0.8	875.0547	850	14.4
101	4.9556	0.9	886.6095	850	14.4
102	5.4767	1.1	901.9658	850	14.4
103	6.2052	1.4	923.4339	850	14.4
104	7.3100	1.9	955.9913	850	14.4
105	9.3372	3.0	1015.7308	850	14.4
106	10.000	3.5	1035.2629	850	14.4

Table A.12.30 Conversion Table for Performance Level Setting Form: ELA/L Grade 6

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-6.0000	1.3	566.8536	650	14.5
1	-6.0000	1.3	566.8536	650	14.5
2	-5.2704	1.1	587.8119	650	14.5
3	-4.5982	0.9	607.1213	650	14.5
4	-4.1239	0.8	620.7459	650	14.5
5	-3.7628	0.7	631.1187	650	14.5
6	-3.4739	0.6	639.4176	650	14.5
7	-3.2345	0.5	646.2945	650	14.5
8	-3.0306	0.5	652.1517	652	14.5
9	-2.8535	0.5	657.2390	657	13.5
10	-2.6970	0.4	661.7346	662	12.8
11	-2.5568	0.4	665.7619	666	12.1
12	-2.4298	0.4	669.4101	669	11.5
13	-2.3137	0.4	672.7451	673	11.0
14	-2.2067	0.4	675.8188	676	10.6
15	-2.1074	0.4	678.6713	679	10.2
16	-2.0146	0.3	681.3370	681	9.8
17	-1.9274	0.3	683.8419	684	9.5
18	-1.8452	0.3	686.2031	686	9.3
19	-1.7672	0.3	688.4437	688	9.0
20	-1.6929	0.3	690.5781	691	8.8
21	-1.6220	0.3	692.6147	693	8.6
22	-1.5541	0.3	694.5652	695	8.4
23	-1.4887	0.3	696.4439	696	8.3
24	-1.4258	0.3	698.2507	698	8.1
25	-1.3649	0.3	700.0001	700	8.0
26	-1.3060	0.3	701.6920	702	7.9
27	-1.2489	0.3	703.3323	703	7.8
28	-1.1933	0.3	704.9294	705	7.7
29	-1.1391	0.3	706.4864	706	7.6
30	-1.0862	0.3	708.0059	708	7.5
31	-1.0346	0.3	709.4882	709	7.4
32	-0.9840	0.3	710.9417	711	7.3
33	-0.9345	0.3	712.3636	712	7.2
34	-0.8858	0.2	713.7626	714	7.2
35	-0.8380	0.2	715.1357	715	7.1
36	-0.7910	0.2	716.4858	716	7.1
37	-0.7446	0.2	717.8186	718	7.0
38	-0.6990	0.2	719.1285	719	7.0
39	-0.6539	0.2	720.4241	720	6.9
40	-0.6094	0.2	721.7024	722	6.9
41	-0.5654	0.2	722.9663	723	6.9
42	-0.5218	0.2	724.2187	724	6.8

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	-0.4787	0.2	725.4568	725	6.8
44	-0.4360	0.2	726.6834	727	6.8
45	-0.3936	0.2	727.9014	728	6.8
46	-0.3515	0.2	729.1107	729	6.7
47	-0.3098	0.2	730.3086	730	6.7
48	-0.2683	0.2	731.5007	732	6.7
49	-0.2270	0.2	732.6871	733	6.7
50	-0.1860	0.2	733.8648	734	6.7
51	-0.1452	0.2	735.0368	735	6.7
52	-0.1045	0.2	736.2060	736	6.7
53	-0.0640	0.2	737.3694	737	6.7
54	-0.0237	0.2	738.5270	739	6.6
55	0.0166	0.2	739.6846	740	6.6
56	0.0567	0.2	740.8365	741	6.6
57	0.0967	0.2	741.9856	742	6.6
58	0.1367	0.2	743.1346	743	6.6
59	0.1766	0.2	744.2808	744	6.6
60	0.2165	0.2	745.4269	745	6.6
61	0.2563	0.2	746.5702	747	6.6
62	0.2961	0.2	747.7135	748	6.6
63	0.3359	0.2	748.8568	749	6.7
64	0.3757	0.2	750.0000	750	6.7
65	0.4155	0.2	751.1433	751	6.7
66	0.4553	0.2	752.2866	752	6.7
67	0.4951	0.2	753.4299	753	6.7
68	0.5349	0.2	754.5732	755	6.7
69	0.5748	0.2	755.7193	756	6.7
70	0.6147	0.2	756.8655	757	6.7
71	0.6546	0.2	758.0116	758	6.7
72	0.6946	0.2	759.1607	759	6.7
73	0.7347	0.2	760.3126	760	6.7
74	0.7748	0.2	761.4645	761	6.8
75	0.8149	0.2	762.6164	763	6.8
76	0.8552	0.2	763.7740	764	6.8
77	0.8955	0.2	764.9317	765	6.8
78	0.9358	0.2	766.0893	766	6.8
79	0.9763	0.2	767.2527	767	6.8
80	1.0168	0.2	768.4161	768	6.8
81	1.0574	0.2	769.5824	770	6.8
82	1.0981	0.2	770.7515	771	6.9
83	1.1389	0.2	771.9235	772	6.9
84	1.1798	0.2	773.0984	773	6.9
85	1.2207	0.2	774.2733	774	6.9
86	1.2618	0.2	775.4539	775	6.9

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	1.3031	0.2	776.6403	777	6.9
88	1.3444	0.2	777.8266	778	7.0
89	1.3859	0.2	779.0187	779	7.0
90	1.4275	0.2	780.2137	780	7.0
91	1.4694	0.2	781.4173	781	7.0
92	1.5114	0.2	782.6238	783	7.1
93	1.5536	0.2	783.8360	784	7.1
94	1.5960	0.2	785.0540	785	7.1
95	1.6387	0.2	786.2806	786	7.1
96	1.6817	0.2	787.5158	788	7.2
97	1.7249	0.3	788.7568	789	7.2
98	1.7686	0.3	790.0121	790	7.3
99	1.8126	0.3	791.2760	791	7.3
100	1.8570	0.3	792.5514	793	7.3
101	1.9019	0.3	793.8412	794	7.4
102	1.9473	0.3	795.1454	795	7.5
103	1.9932	0.3	796.4639	796	7.5
104	2.0398	0.3	797.8025	798	7.6
105	2.0872	0.3	799.1641	799	7.6
106	2.1353	0.3	800.5458	801	7.7
107	2.1843	0.3	801.9533	802	7.8
108	2.2342	0.3	803.3868	803	7.9
109	2.2853	0.3	804.8546	805	8.0
110	2.3376	0.3	806.3570	806	8.1
111	2.3912	0.3	807.8967	808	8.2
112	2.4464	0.3	809.4824	809	8.4
113	2.5032	0.3	811.1140	811	8.5
114	2.5620	0.3	812.8030	813	8.7
115	2.6229	0.3	814.5524	815	8.9
116	2.6863	0.3	816.3736	816	9.0
117	2.7525	0.3	818.2753	818	9.3
118	2.8218	0.3	820.2660	820	9.5
119	2.8947	0.3	822.3601	822	9.8
120	2.9718	0.4	824.5748	825	10.1
121	3.0537	0.4	826.9275	827	10.4
122	3.1412	0.4	829.4410	829	10.8
123	3.2353	0.4	832.1441	832	11.3
124	3.3373	0.4	835.0741	835	11.8
125	3.4485	0.4	838.2684	838	12.3
126	3.5711	0.5	841.7901	842	13.0
127	3.7077	0.5	845.7141	846	13.8
128	3.8617	0.5	850.1378	850	13.8
129	4.0381	0.6	855.2050	850	13.8
130	4.2439	0.6	861.1168	850	13.8

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
131	4.4898	0.7	868.1804	850	13.8
132	4.7928	0.8	876.8843	850	13.8
133	5.1826	0.9	888.0816	850	13.8
134	5.7162	1.1	903.4096	850	13.8
135	6.5236	1.4	926.6028	850	13.8
136	8.0053	2.1	969.1656	850	13.8
137	10.000	3.4	1026.4648	850	13.8

Table A.12.31 Conversion Table for Performance Level Setting Form: ELA/L Grade 7

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-6.0000	1.9	536.5141	650	15
1	-6.0000	1.9	536.5141	650	15
2	-4.7838	1.2	577.7244	650	15
3	-4.0880	0.9	601.3013	650	15
4	-3.6310	0.8	616.7865	650	15
5	-3.2948	0.7	628.1784	650	15
6	-3.0302	0.6	637.1443	650	15
7	-2.8128	0.5	644.5108	650	15
8	-2.6284	0.5	650.7591	651	15
9	-2.4685	0.5	656.1772	656	15
10	-2.3272	0.4	660.9651	661	14.4
11	-2.2007	0.4	665.2515	665	13.7
12	-2.0860	0.4	669.1380	669	13.0
13	-1.9810	0.4	672.6959	673	12.4
14	-1.8842	0.4	675.9759	676	11.9
15	-1.7942	0.3	679.0255	679	11.5
16	-1.7100	0.3	681.8786	682	11.1
17	-1.6310	0.3	684.5555	685	10.8
18	-1.5563	0.3	687.0867	687	10.5
19	-1.4854	0.3	689.4891	689	10.2
20	-1.4179	0.3	691.7763	692	9.9
21	-1.3535	0.3	693.9584	694	9.7
22	-1.2917	0.3	696.0525	696	9.5
23	-1.2324	0.3	698.0618	698	9.3
24	-1.1752	0.3	700.0000	700	9.1
25	-1.1199	0.3	701.8738	702	9.0
26	-1.0665	0.3	703.6833	704	8.8
27	-1.0146	0.3	705.4419	705	8.7
28	-0.9642	0.3	707.1497	707	8.6
29	-0.9151	0.2	708.8134	709	8.5
30	-0.8673	0.2	710.4331	710	8.4
31	-0.8205	0.2	712.0189	712	8.3
32	-0.7748	0.2	713.5674	714	8.2
33	-0.7300	0.2	715.0854	715	8.1
34	-0.6861	0.2	716.5729	717	8.0
35	-0.6430	0.2	718.0334	718	7.9
36	-0.6007	0.2	719.4667	719	7.9
37	-0.5590	0.2	720.8797	721	7.8
38	-0.5179	0.2	722.2723	722	7.8
39	-0.4774	0.2	723.6446	724	7.7
40	-0.4375	0.2	724.9966	725	7.7
41	-0.3980	0.2	726.3351	726	7.6
42	-0.3590	0.2	727.6566	728	7.6

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	-0.3203	0.2	728.9679	729	7.5
44	-0.2821	0.2	730.2623	730	7.5
45	-0.2442	0.2	731.5465	732	7.5
46	-0.2066	0.2	732.8206	733	7.5
47	-0.1693	0.2	734.0845	734	7.4
48	-0.1322	0.2	735.3416	735	7.4
49	-0.0954	0.2	736.5885	737	7.4
50	-0.0588	0.2	737.8287	738	7.4
51	-0.0223	0.2	739.0655	739	7.4
52	0.0139	0.2	740.2921	740	7.4
53	0.0501	0.2	741.5187	742	7.4
54	0.0861	0.2	742.7386	743	7.4
55	0.1220	0.2	743.9550	744	7.4
56	0.1578	0.2	745.1681	745	7.4
57	0.1935	0.2	746.3778	746	7.4
58	0.2292	0.2	747.5874	748	7.4
59	0.2648	0.2	748.7937	749	7.4
60	0.3004	0.2	750.0000	750	7.4
61	0.3360	0.2	751.2063	751	7.4
62	0.3716	0.2	752.4126	752	7.4
63	0.4072	0.2	753.6189	754	7.4
64	0.4428	0.2	754.8252	755	7.4
65	0.4785	0.2	756.0348	756	7.4
66	0.5142	0.2	757.2445	757	7.4
67	0.5500	0.2	758.4576	758	7.5
68	0.5858	0.2	759.6706	760	7.5
69	0.6218	0.2	760.8905	761	7.5
70	0.6578	0.2	762.1103	762	7.5
71	0.6940	0.2	763.3369	763	7.5
72	0.7302	0.2	764.5636	765	7.6
73	0.7666	0.2	765.7970	766	7.6
74	0.8032	0.2	767.0371	767	7.6
75	0.8399	0.2	768.2807	768	7.6
76	0.8768	0.2	769.5310	770	7.7
77	0.9139	0.2	770.7881	771	7.7
78	0.9511	0.2	772.0486	772	7.7
79	0.9886	0.2	773.3193	773	7.7
80	1.0263	0.2	774.5968	775	7.8
81	1.0642	0.2	775.8810	776	7.8
82	1.1023	0.2	777.1720	777	7.8
83	1.1407	0.2	778.4731	778	7.9
84	1.1794	0.2	779.7845	780	7.9
85	1.2184	0.2	781.1060	781	8.0
86	1.2577	0.2	782.4376	782	8.0

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	1.2973	0.2	783.7795	784	8.0
88	1.3373	0.2	785.1348	785	8.1
89	1.3776	0.2	786.5004	787	8.1
90	1.4183	0.2	787.8795	788	8.2
91	1.4594	0.2	789.2721	789	8.2
92	1.5009	0.2	790.6783	791	8.3
93	1.5429	0.2	792.1015	792	8.3
94	1.5854	0.2	793.5416	794	8.4
95	1.6283	0.2	794.9952	795	8.5
96	1.6718	0.3	796.4692	796	8.5
97	1.7159	0.3	797.9635	798	8.6
98	1.7606	0.3	799.4782	799	8.7
99	1.8059	0.3	801.0131	801	8.7
100	1.8520	0.3	802.5752	803	8.8
101	1.8988	0.3	804.1610	804	8.9
102	1.9464	0.3	805.7739	806	9.0
103	1.9949	0.3	807.4173	807	9.1
104	2.0444	0.3	809.0946	809	9.2
105	2.0950	0.3	810.8091	811	9.3
106	2.1467	0.3	812.5610	813	9.4
107	2.1997	0.3	814.3568	814	9.5
108	2.2540	0.3	816.1968	816	9.7
109	2.3100	0.3	818.0943	818	9.8
110	2.3676	0.3	820.0460	820	10.0
111	2.4272	0.3	822.0656	822	10.2
112	2.4889	0.3	824.1562	824	10.4
113	2.5530	0.3	826.3282	826	10.6
114	2.6199	0.3	828.5951	829	10.9
115	2.6898	0.3	830.9636	831	11.2
116	2.7634	0.3	833.4575	833	11.5
117	2.8410	0.3	836.0870	836	11.8
118	2.9233	0.4	838.8757	839	12.2
119	3.0112	0.4	841.8541	842	12.7
120	3.1056	0.4	845.0528	845	13.2
121	3.2076	0.4	848.5090	849	13.8
122	3.3189	0.4	852.2804	850	13.8
123	3.4413	0.5	856.4278	850	13.8
124	3.5774	0.5	861.0395	850	13.8
125	3.7307	0.5	866.2340	850	13.8
126	3.9058	0.6	872.1672	850	13.8
127	4.1093	0.6	879.0627	850	13.8
128	4.3505	0.7	887.2356	850	13.8
129	4.6440	0.8	897.1807	850	13.8
130	5.0126	0.9	909.6705	850	13.8

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
131	5.4954	1.0	926.0300	850	13.8
132	6.1659	1.3	948.7495	850	13.8
133	7.1859	1.7	983.3117	850	13.8
134	9.0569	2.7	1046.7096	850	13.8
135	10.000	3.3	1078.6661	850	13.8

Table A.12.32 Conversion Table for Performance Level Setting Form: ELA/L Grade 8

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-6.0000	1.8	533.2679	650	15
1	-6.0000	1.8	533.2679	650	15
2	-5.6481	1.6	545.3482	650	15
3	-4.7448	1.2	576.3575	650	15
4	-4.1705	0.9	596.0726	650	15
5	-3.7567	0.8	610.2779	650	15
6	-3.4362	0.7	621.2803	650	15
7	-3.1760	0.6	630.2127	650	15
8	-2.9575	0.6	637.7136	650	15
9	-2.7696	0.5	644.1640	650	15
10	-2.6050	0.5	649.8145	650	15
11	-2.4585	0.5	654.8437	655	15
12	-2.3267	0.4	659.3682	659	14.9
13	-2.2068	0.4	663.4843	663	14.2
14	-2.0969	0.4	667.2570	667	13.6
15	-1.9954	0.4	670.7414	671	13.0
16	-1.9011	0.4	673.9786	674	12.5
17	-1.8129	0.4	677.0064	677	12.1
18	-1.7301	0.3	679.8489	680	11.7
19	-1.6520	0.3	682.5300	683	11.3
20	-1.5780	0.3	685.0703	685	11.0
21	-1.5076	0.3	687.4871	687	10.7
22	-1.4405	0.3	689.7905	690	10.4
23	-1.3763	0.3	691.9944	692	10.2
24	-1.3147	0.3	694.1091	694	10.0
25	-1.2555	0.3	696.1414	696	9.8
26	-1.1983	0.3	698.1050	698	9.6
27	-1.1431	0.3	699.9999	700	9.5
28	-1.0897	0.3	701.8331	702	9.3
29	-1.0378	0.3	703.6148	704	9.2
30	-0.9873	0.3	705.3484	705	9.0
31	-0.9382	0.3	707.0339	707	8.9
32	-0.8903	0.3	708.6783	709	8.8
33	-0.8435	0.3	710.2849	710	8.7
34	-0.7977	0.3	711.8571	712	8.6
35	-0.7529	0.2	713.3951	713	8.5
36	-0.7089	0.2	714.9055	715	8.4
37	-0.6658	0.2	716.3851	716	8.4
38	-0.6233	0.2	717.8441	718	8.3
39	-0.5816	0.2	719.2756	719	8.2
40	-0.5404	0.2	720.6900	721	8.2
41	-0.4999	0.2	722.0803	722	8.1
42	-0.4598	0.2	723.4569	723	8.1

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	-0.4203	0.2	724.8129	725	8.0
44	-0.3812	0.2	726.1551	726	8.0
45	-0.3425	0.2	727.4837	727	8.0
46	-0.3042	0.2	728.7984	729	7.9
47	-0.2663	0.2	730.0995	730	7.9
48	-0.2287	0.2	731.3903	731	7.9
49	-0.1914	0.2	732.6707	733	7.9
50	-0.1543	0.2	733.9444	734	7.8
51	-0.1175	0.2	735.2077	735	7.8
52	-0.0809	0.2	736.4641	736	7.8
53	-0.0445	0.2	737.7137	738	7.8
54	-0.0082	0.2	738.9598	739	7.8
55	0.0279	0.2	740.1991	740	7.8
56	0.0638	0.2	741.4315	741	7.8
57	0.0997	0.2	742.6639	743	7.8
58	0.1354	0.2	743.8894	744	7.8
59	0.1711	0.2	745.1150	745	7.8
60	0.2067	0.2	746.3371	746	7.8
61	0.2423	0.2	747.5592	748	7.8
62	0.2778	0.2	748.7779	749	7.8
63	0.3134	0.2	750.0000	750	7.8
64	0.3490	0.2	751.2221	751	7.8
65	0.3845	0.2	752.4408	752	7.8
66	0.4202	0.2	753.6663	754	7.8
67	0.4559	0.2	754.8918	755	7.8
68	0.4916	0.2	756.1174	756	7.8
69	0.5274	0.2	757.3464	757	7.9
70	0.5634	0.2	758.5822	759	7.9
71	0.5994	0.2	759.8180	760	7.9
72	0.6356	0.2	761.0607	761	7.9
73	0.6719	0.2	762.3069	762	8.0
74	0.7084	0.2	763.5599	764	8.0
75	0.7450	0.2	764.8163	765	8.0
76	0.7818	0.2	766.0796	766	8.0
77	0.8188	0.2	767.3498	767	8.1
78	0.8560	0.2	768.6268	769	8.1
79	0.8935	0.2	769.9142	770	8.1
80	0.9312	0.2	771.2084	771	8.2
81	0.9691	0.2	772.5094	773	8.2
82	1.0074	0.2	773.8242	774	8.2
83	1.0459	0.2	775.1459	775	8.3
84	1.0847	0.2	776.4779	776	8.3
85	1.1238	0.2	777.8201	778	8.4
86	1.1633	0.2	779.1761	779	8.4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	1.2032	0.2	780.5458	781	8.5
88	1.2434	0.2	781.9259	782	8.5
89	1.2841	0.2	783.3230	783	8.6
90	1.3252	0.3	784.7340	785	8.6
91	1.3668	0.3	786.1620	786	8.7
92	1.4088	0.3	787.6039	788	8.7
93	1.4514	0.3	789.0663	789	8.8
94	1.4946	0.3	790.5493	791	8.9
95	1.5383	0.3	792.0494	792	9.0
96	1.5827	0.3	793.5737	794	9.0
97	1.6278	0.3	795.1219	795	9.1
98	1.6736	0.3	796.6941	797	9.2
99	1.7202	0.3	798.2939	798	9.3
100	1.7677	0.3	799.9245	800	9.4
101	1.8162	0.3	801.5894	802	9.5
102	1.8656	0.3	803.2853	803	9.6
103	1.9161	0.3	805.0189	805	9.7
104	1.9678	0.3	806.7937	807	9.8
105	2.0208	0.3	808.6131	809	10.0
106	2.0753	0.3	810.4841	810	10.1
107	2.1312	0.3	812.4031	812	10.3
108	2.1889	0.3	814.3838	814	10.5
109	2.2485	0.3	816.4298	816	10.6
110	2.3100	0.3	818.5411	819	10.8
111	2.3739	0.3	820.7347	821	11.1
112	2.4403	0.3	823.0141	823	11.3
113	2.5094	0.3	825.3862	825	11.6
114	2.5817	0.3	827.8682	828	11.8
115	2.6575	0.4	830.4704	830	12.1
116	2.7371	0.4	833.2029	833	12.5
117	2.8212	0.4	836.0900	836	12.8
118	2.9103	0.4	839.1487	839	13.2
119	3.0050	0.4	842.3996	842	13.7
120	3.1062	0.4	845.8737	846	14.2
121	3.2149	0.4	849.6053	850	14.2
122	3.3322	0.4	853.6321	850	14.2
123	3.4598	0.5	858.0124	850	14.2
124	3.5994	0.5	862.8047	850	14.2
125	3.7534	0.5	868.0914	850	14.2
126	3.9250	0.6	873.9822	850	14.2
127	4.1184	0.6	880.6214	850	14.2
128	4.3395	0.6	888.2116	850	14.2
129	4.5964	0.7	897.0307	850	14.2
130	4.9015	0.8	907.5044	850	14.2

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
131	5.2737	0.9	920.2816	850	14.2
132	5.7436	1.1	936.4128	850	14.2
133	6.3643	1.3	957.7207	850	14.2
134	7.2370	1.7	987.6795	850	14.2
135	8.5901	2.3	1034.1300	850	14.2
136	10.000	3.1	1082.5303	850	14.2
137	10.000	3.1	1082.5303	850	14.2

Table A.12.33 Conversion Table for Performance Level Setting Form: ELA/L Grade 9

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-6.0000	1.9	531.3885	650	15
1	-5.4896	1.6	548.8728	650	15
2	-4.3236	1.1	588.8153	650	15
3	-3.6708	0.9	611.1776	650	15
4	-3.2296	0.7	626.2913	650	15
5	-2.9022	0.6	637.5067	650	15
6	-2.6441	0.6	646.3482	650	15
7	-2.4323	0.5	653.6036	654	15
8	-2.2529	0.5	659.7492	660	15
9	-2.0975	0.4	665.0725	665	15
10	-1.9605	0.4	669.7656	670	14.2
11	-1.8378	0.4	673.9688	674	13.4
12	-1.7266	0.4	677.7781	678	12.8
13	-1.6249	0.4	681.2619	681	12.2
14	-1.5310	0.3	684.4786	684	11.7
15	-1.4437	0.3	687.4691	687	11.3
16	-1.3620	0.3	690.2678	690	10.9
17	-1.2852	0.3	692.8987	693	10.6
18	-1.2125	0.3	695.3891	695	10.3
19	-1.1436	0.3	697.7493	698	10.0
20	-1.0779	0.3	700.0000	700	9.8
21	-1.0150	0.3	702.1547	702	9.5
22	-0.9547	0.3	704.2203	704	9.3
23	-0.8966	0.3	706.2106	706	9.2
24	-0.8406	0.3	708.1289	708	9.0
25	-0.7865	0.3	709.9822	710	8.8
26	-0.7340	0.3	711.7806	712	8.7
27	-0.6830	0.3	713.5277	714	8.6
28	-0.6334	0.2	715.2267	715	8.5
29	-0.5850	0.2	716.8847	717	8.4
30	-0.5378	0.2	718.5016	719	8.3
31	-0.4916	0.2	720.0843	720	8.2
32	-0.4464	0.2	721.6326	722	8.1
33	-0.4020	0.2	723.1536	723	8.0
34	-0.3585	0.2	724.6437	725	7.9
35	-0.3156	0.2	726.1133	726	7.9
36	-0.2734	0.2	727.5589	728	7.8
37	-0.2319	0.2	728.9805	729	7.8
38	-0.1909	0.2	730.3850	730	7.7
39	-0.1504	0.2	731.7724	732	7.7
40	-0.1104	0.2	733.1426	733	7.6
41	-0.0708	0.2	734.4992	734	7.6
42	-0.0316	0.2	735.8420	736	7.6

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	0.0072	0.2	737.1711	737	7.5
44	0.0457	0.2	738.4900	738	7.5
45	0.0839	0.2	739.7986	740	7.5
46	0.1218	0.2	741.0969	741	7.4
47	0.1595	0.2	742.3883	742	7.4
48	0.1970	0.2	743.6729	744	7.4
49	0.2342	0.2	744.9473	745	7.4
50	0.2713	0.2	746.2182	746	7.4
51	0.3083	0.2	747.4856	747	7.4
52	0.3451	0.2	748.7462	749	7.4
53	0.3817	0.2	750.0000	750	7.4
54	0.4183	0.2	751.2538	751	7.3
55	0.4548	0.2	752.5041	753	7.3
56	0.4912	0.2	753.7510	754	7.3
57	0.5276	0.2	754.9980	755	7.3
58	0.5638	0.2	756.2380	756	7.3
59	0.6001	0.2	757.4815	757	7.3
60	0.6363	0.2	758.7216	759	7.3
61	0.6725	0.2	759.9617	760	7.3
62	0.7087	0.2	761.2017	761	7.3
63	0.7449	0.2	762.4418	762	7.3
64	0.7810	0.2	763.6784	764	7.3
65	0.8172	0.2	764.9185	765	7.3
66	0.8534	0.2	766.1586	766	7.4
67	0.8896	0.2	767.3986	767	7.4
68	0.9259	0.2	768.6421	769	7.4
69	0.9622	0.2	769.8856	770	7.4
70	0.9985	0.2	771.1291	771	7.4
71	1.0348	0.2	772.3726	772	7.4
72	1.0712	0.2	773.6195	774	7.4
73	1.1077	0.2	774.8699	775	7.4
74	1.1442	0.2	776.1202	776	7.4
75	1.1808	0.2	777.3740	777	7.4
76	1.2174	0.2	778.6278	779	7.4
77	1.2541	0.2	779.8849	780	7.5
78	1.2909	0.2	781.1456	781	7.5
79	1.3278	0.2	782.4096	782	7.5
80	1.3648	0.2	783.6771	784	7.5
81	1.4019	0.2	784.9480	785	7.5
82	1.4390	0.2	786.2189	786	7.5
83	1.4764	0.2	787.5001	788	7.5
84	1.5138	0.2	788.7812	789	7.6
85	1.5514	0.2	790.0693	790	7.6
86	1.5891	0.2	791.3607	791	7.6

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	1.6271	0.2	792.6624	793	7.6
88	1.6652	0.2	793.9676	794	7.7
89	1.7035	0.2	795.2796	795	7.7
90	1.7420	0.2	796.5985	797	7.7
91	1.7808	0.2	797.9276	798	7.8
92	1.8199	0.2	799.2670	799	7.8
93	1.8593	0.2	800.6167	801	7.9
94	1.8990	0.2	801.9766	802	7.9
95	1.9390	0.2	803.3469	803	7.9
96	1.9795	0.2	804.7343	805	8.0
97	2.0204	0.2	806.1353	806	8.1
98	2.0617	0.2	807.5501	808	8.1
99	2.1036	0.2	808.9854	809	8.2
100	2.1461	0.2	810.4413	810	8.3
101	2.1892	0.2	811.9177	812	8.3
102	2.2330	0.2	813.4181	813	8.4
103	2.2775	0.2	814.9425	815	8.5
104	2.3229	0.3	816.4978	816	8.6
105	2.3693	0.3	818.0872	818	8.7
106	2.4166	0.3	819.7075	820	8.8
107	2.4651	0.3	821.3690	821	9.0
108	2.5149	0.3	823.0749	823	9.1
109	2.5661	0.3	824.8288	825	9.2
110	2.6188	0.3	826.6341	827	9.4
111	2.6733	0.3	828.5011	829	9.6
112	2.7297	0.3	830.4331	830	9.8
113	2.7883	0.3	832.4405	832	10.0
114	2.8494	0.3	834.5335	835	10.2
115	2.9132	0.3	836.7191	837	10.5
116	2.9803	0.3	839.0177	839	10.8
117	3.0509	0.3	841.4361	841	11.1
118	3.1258	0.3	844.0019	844	11.5
119	3.2054	0.3	846.7287	847	11.9
120	3.2907	0.4	849.6507	850	11.9
121	3.3826	0.4	852.7988	850	11.9
122	3.4824	0.4	856.2176	850	11.9
123	3.5917	0.4	859.9618	850	11.9
124	3.7125	0.4	864.0999	850	11.9
125	3.8476	0.5	868.7279	850	11.9
126	4.0008	0.5	873.9759	850	11.9
127	4.1772	0.5	880.0187	850	11.9
128	4.3844	0.6	887.1165	850	11.9
129	4.6337	0.7	895.6565	850	11.9
130	4.9429	0.8	906.2485	850	11.9

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
131	5.3421	0.9	919.9235	850	11.9
132	5.8878	1.1	938.6170	850	11.9
133	6.7060	1.4	966.6452	850	11.9
134	8.2067	2.2	1018.0532	850	11.9
135	10.000	3.5	1079.4845	850	11.9

Table A.12.34 Conversion Table for Performance Level Setting Form: ELA/L Grade 10

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-6.0000	1.9	480.7338	650	15
1	-5.4329	1.6	504.8104	650	15
2	-4.3822	1.0	549.4185	650	15
3	-3.8069	0.8	573.8432	650	15
4	-3.4127	0.7	590.5792	650	15
5	-3.1134	0.6	603.2861	650	15
6	-2.8725	0.5	613.5137	650	15
7	-2.6709	0.5	622.0727	650	15
8	-2.4976	0.5	629.4303	650	15
9	-2.3455	0.4	635.8878	650	15
10	-2.2101	0.4	641.6363	650	15
11	-2.0879	0.4	646.8244	650	15
12	-1.9765	0.4	651.5539	652	15
13	-1.8742	0.4	655.8971	656	15
14	-1.7795	0.3	659.9177	660	14.4
15	-1.6913	0.3	663.6622	664	13.9
16	-1.6088	0.3	667.1648	667	13.5
17	-1.5312	0.3	670.4594	670	13.0
18	-1.4579	0.3	673.5714	674	12.7
19	-1.3884	0.3	676.5220	677	12.3
20	-1.3224	0.3	679.3241	679	12.0
21	-1.2594	0.3	681.9988	682	11.7
22	-1.1991	0.3	684.5589	685	11.5
23	-1.1413	0.3	687.0128	687	11.2
24	-1.0858	0.3	689.3691	689	11.0
25	-1.0323	0.3	691.6405	692	10.8
26	-0.9807	0.3	693.8312	694	10.6
27	-0.9307	0.2	695.9540	696	10.5
28	-0.8823	0.2	698.0088	698	10.3
29	-0.8354	0.2	700.0000	700	10.1
30	-0.7898	0.2	701.9360	702	10.0
31	-0.7454	0.2	703.8210	704	9.9
32	-0.7021	0.2	705.6593	706	9.7
33	-0.6598	0.2	707.4552	707	9.6
34	-0.6185	0.2	709.2086	709	9.5
35	-0.5780	0.2	710.9281	711	9.4
36	-0.5384	0.2	712.6093	713	9.3
37	-0.4996	0.2	714.2566	714	9.3
38	-0.4614	0.2	715.8784	716	9.2
39	-0.4239	0.2	717.4705	717	9.1
40	-0.3870	0.2	719.0371	719	9.0
41	-0.3506	0.2	720.5825	721	9.0
42	-0.3148	0.2	722.1024	722	8.9

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	-0.2794	0.2	723.6053	724	8.9
44	-0.2445	0.2	725.0870	725	8.8
45	-0.2099	0.2	726.5560	727	8.8
46	-0.1758	0.2	728.0037	728	8.7
47	-0.1420	0.2	729.4387	729	8.7
48	-0.1084	0.2	730.8652	731	8.7
49	-0.0752	0.2	732.2747	732	8.6
50	-0.0422	0.2	733.6758	734	8.6
51	-0.0095	0.2	735.0641	735	8.6
52	0.0231	0.2	736.4481	736	8.6
53	0.0554	0.2	737.8194	738	8.6
54	0.0877	0.2	739.1908	739	8.6
55	0.1197	0.2	740.5493	741	8.5
56	0.1517	0.2	741.9079	742	8.5
57	0.1836	0.2	743.2622	743	8.5
58	0.2154	0.2	744.6123	745	8.5
59	0.2471	0.2	745.9582	746	8.5
60	0.2789	0.2	747.3083	747	8.5
61	0.3106	0.2	748.6541	749	8.6
62	0.3423	0.2	750.0000	750	8.6
63	0.3740	0.2	751.3458	751	8.6
64	0.4058	0.2	752.6959	753	8.6
65	0.4376	0.2	754.0460	754	8.6
66	0.4695	0.2	755.4003	755	8.6
67	0.5015	0.2	756.7589	757	8.6
68	0.5336	0.2	758.1217	758	8.7
69	0.5658	0.2	759.4888	759	8.7
70	0.5982	0.2	760.8643	761	8.7
71	0.6307	0.2	762.2441	762	8.8
72	0.6634	0.2	763.6324	764	8.8
73	0.6962	0.2	765.0250	765	8.8
74	0.7293	0.2	766.4303	766	8.9
75	0.7626	0.2	767.8440	768	8.9
76	0.7961	0.2	769.2663	769	8.9
77	0.8298	0.2	770.6971	771	9.0
78	0.8638	0.2	772.1405	772	9.0
79	0.8981	0.2	773.5968	774	9.1
80	0.9327	0.2	775.0657	775	9.1
81	0.9676	0.2	776.5474	777	9.2
82	1.0028	0.2	778.0419	778	9.2
83	1.0383	0.2	779.5490	780	9.3
84	1.0742	0.2	781.0732	781	9.3
85	1.1104	0.2	782.6101	783	9.4
86	1.1470	0.2	784.1640	784	9.4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	1.1840	0.2	785.7348	786	9.5
88	1.2215	0.2	787.3269	787	9.6
89	1.2593	0.2	788.9317	789	9.6
90	1.2977	0.2	790.5620	791	9.7
91	1.3365	0.2	792.2093	792	9.8
92	1.3758	0.2	793.8778	794	9.8
93	1.4156	0.2	795.5675	796	9.9
94	1.4560	0.2	797.2828	797	10.0
95	1.4969	0.2	799.0192	799	10.1
96	1.5385	0.2	800.7853	801	10.2
97	1.5807	0.2	802.5770	803	10.2
98	1.6235	0.2	804.3941	804	10.3
99	1.6671	0.2	806.2451	806	10.4
100	1.7115	0.2	808.1302	808	10.5
101	1.7566	0.3	810.0449	810	10.7
102	1.8027	0.3	812.0021	812	10.8
103	1.8496	0.3	813.9933	814	10.9
104	1.8976	0.3	816.0311	816	11.0
105	1.9466	0.3	818.1115	818	11.2
106	1.9968	0.3	820.2427	820	11.3
107	2.0482	0.3	822.4250	822	11.5
108	2.1010	0.3	824.6666	825	11.6
109	2.1553	0.3	826.9720	827	11.8
110	2.2112	0.3	829.3452	829	12.0
111	2.2688	0.3	831.7907	832	12.2
112	2.3285	0.3	834.3253	834	12.5
113	2.3902	0.3	836.9448	837	12.7
114	2.4544	0.3	839.6704	840	13.0
115	2.5212	0.3	842.5065	843	13.3
116	2.5909	0.3	845.4656	845	13.6
117	2.6640	0.3	848.5691	849	13.9
118	2.7408	0.3	851.8297	850	13.9
119	2.8218	0.3	855.2686	850	13.9
120	2.9077	0.4	858.9155	850	13.9
121	2.9991	0.4	862.7960	850	13.9
122	3.0969	0.4	866.9481	850	13.9
123	3.2021	0.4	871.4145	850	13.9
124	3.3161	0.4	876.2544	850	13.9
125	3.4403	0.4	881.5274	850	13.9
126	3.5769	0.5	887.3268	850	13.9
127	3.7286	0.5	893.7674	850	13.9
128	3.8988	0.5	900.9933	850	13.9
129	4.0924	0.6	909.2127	850	13.9
130	4.3161	0.6	918.7100	850	13.9

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
131	4.5800	0.7	929.9140	850	13.9
132	4.9000	0.8	943.4998	850	13.9
133	5.3031	0.9	960.6137	850	13.9
134	5.8413	1.0	983.4633	850	13.9
135	6.6365	1.3	1017.2240	850	13.9
136	8.1029	2.1	1079.4809	850	13.9
137	10.000	3.6	1160.0234	850	13.9

Table A.12.35 Conversion Table for Performance Level Setting Form: ELA/L Grade 11

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-6.0000	2.4	520.0799	650	15
1	-4.8745	1.5	561.3130	650	15
2	-3.8597	1.0	598.4906	650	15
3	-3.3200	0.7	618.2628	650	15
4	-2.9561	0.6	631.5944	650	15
5	-2.6823	0.5	641.6252	650	15
6	-2.4628	0.5	649.6666	650	15
7	-2.2792	0.4	656.3929	656	15
8	-2.1212	0.4	662.1813	662	15
9	-1.9823	0.4	667.2699	667	14.2
10	-1.8580	0.4	671.8237	672	13.5
11	-1.7454	0.4	675.9489	676	12.9
12	-1.6424	0.3	679.7223	680	12.4
13	-1.5472	0.3	683.2100	683	11.9
14	-1.4588	0.3	686.4486	686	11.5
15	-1.3761	0.3	689.4783	689	11.2
16	-1.2983	0.3	692.3286	692	10.9
17	-1.2248	0.3	695.0213	695	10.6
18	-1.1552	0.3	697.5711	698	10.3
19	-1.0889	0.3	700.0000	700	10.1
20	-1.0257	0.3	702.3154	702	9.9
21	-0.9652	0.3	704.5318	705	9.7
22	-0.9071	0.3	706.6603	707	9.5
23	-0.8512	0.3	708.7082	709	9.3
24	-0.7974	0.3	710.6792	711	9.2
25	-0.7454	0.2	712.5843	713	9.1
26	-0.6951	0.2	714.4270	714	8.9
27	-0.6463	0.2	716.2148	716	8.8
28	-0.5989	0.2	717.9514	718	8.7
29	-0.5528	0.2	719.6403	720	8.6
30	-0.5080	0.2	721.2815	721	8.5
31	-0.4642	0.2	722.8861	723	8.4
32	-0.4214	0.2	724.4541	724	8.3
33	-0.3795	0.2	725.9892	726	8.2
34	-0.3385	0.2	727.4912	727	8.1
35	-0.2983	0.2	728.9640	729	8.1
36	-0.2588	0.2	730.4111	730	8.0
37	-0.2201	0.2	731.8288	732	8.0
38	-0.1819	0.2	733.2283	733	7.9
39	-0.1443	0.2	734.6058	735	7.9
40	-0.1072	0.2	735.9650	736	7.8
41	-0.0706	0.2	737.3058	737	7.8
42	-0.0345	0.2	738.6284	739	7.7

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	0.0012	0.2	739.9363	740	7.7
44	0.0365	0.2	741.2295	741	7.7
45	0.0715	0.2	742.5117	743	7.6
46	0.1062	0.2	743.7830	744	7.6
47	0.1406	0.2	745.0432	745	7.6
48	0.1747	0.2	746.2925	746	7.6
49	0.2087	0.2	747.5381	748	7.5
50	0.2424	0.2	748.7727	749	7.5
51	0.2759	0.2	750.0000	750	7.5
52	0.3092	0.2	751.2200	751	7.5
53	0.3424	0.2	752.4363	752	7.5
54	0.3755	0.2	753.6489	754	7.5
55	0.4085	0.2	754.8579	755	7.5
56	0.4414	0.2	756.0632	756	7.5
57	0.4742	0.2	757.2648	757	7.5
58	0.5070	0.2	758.4664	758	7.5
59	0.5397	0.2	759.6644	760	7.5
60	0.5724	0.2	760.8624	761	7.5
61	0.6051	0.2	762.0604	762	7.5
62	0.6378	0.2	763.2584	763	7.5
63	0.6705	0.2	764.4563	764	7.5
64	0.7032	0.2	765.6543	766	7.5
65	0.7360	0.2	766.8560	767	7.5
66	0.7688	0.2	768.0576	768	7.5
67	0.8017	0.2	769.2629	769	7.5
68	0.8346	0.2	770.4682	770	7.5
69	0.8676	0.2	771.6772	772	7.5
70	0.9007	0.2	772.8898	773	7.5
71	0.9338	0.2	774.1024	774	7.6
72	0.9671	0.2	775.3224	775	7.6
73	1.0004	0.2	776.5424	777	7.6
74	1.0339	0.2	777.7696	778	7.6
75	1.0675	0.2	779.0006	779	7.6
76	1.1012	0.2	780.2352	780	7.6
77	1.1350	0.2	781.4735	781	7.6
78	1.1690	0.2	782.7191	783	7.7
79	1.2031	0.2	783.9683	784	7.7
80	1.2374	0.2	785.2249	785	7.7
81	1.2718	0.2	786.4852	786	7.7
82	1.3064	0.2	787.7528	788	7.7
83	1.3411	0.2	789.0240	789	7.8
84	1.3761	0.2	790.3063	790	7.8
85	1.4112	0.2	791.5922	792	7.8
86	1.4465	0.2	792.8854	793	7.8

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	1.4821	0.2	794.1896	794	7.9
88	1.5179	0.2	795.5012	796	7.9
89	1.5539	0.2	796.8200	797	7.9
90	1.5902	0.2	798.1499	798	7.9
91	1.6268	0.2	799.4908	799	8.0
92	1.6637	0.2	800.8426	801	8.0
93	1.7009	0.2	802.2055	802	8.0
94	1.7385	0.2	803.5829	804	8.1
95	1.7765	0.2	804.9751	805	8.1
96	1.8149	0.2	806.3819	806	8.2
97	1.8537	0.2	807.8033	808	8.2
98	1.8931	0.2	809.2468	809	8.3
99	1.9329	0.2	810.7049	811	8.3
100	1.9733	0.2	812.1849	812	8.4
101	2.0144	0.2	813.6906	814	8.5
102	2.0561	0.2	815.2183	815	8.5
103	2.0986	0.2	816.7754	817	8.6
104	2.1419	0.2	818.3617	818	8.7
105	2.1860	0.2	819.9773	820	8.8
106	2.2311	0.2	821.6295	822	8.9
107	2.2773	0.2	823.3221	823	9.0
108	2.3246	0.2	825.0550	825	9.1
109	2.3732	0.3	826.8354	827	9.3
110	2.4232	0.3	828.6672	829	9.4
111	2.4747	0.3	830.5539	831	9.5
112	2.5279	0.3	832.5029	833	9.7
113	2.5830	0.3	834.5215	835	9.9
114	2.6401	0.3	836.6134	837	10.1
115	2.6995	0.3	838.7896	839	10.3
116	2.7615	0.3	841.0610	841	10.5
117	2.8263	0.3	843.4349	843	10.8
118	2.8944	0.3	845.9298	846	11.1
119	2.9660	0.3	848.5529	849	11.4
120	3.0417	0.3	851.3262	850	11.4
121	3.1221	0.3	854.2717	850	11.4
122	3.2078	0.3	857.4113	850	11.4
123	3.2996	0.4	860.7745	850	11.4
124	3.3985	0.4	864.3977	850	11.4
125	3.5058	0.4	868.3287	850	11.4
126	3.6231	0.4	872.6260	850	11.4
127	3.7523	0.4	877.3593	850	11.4
128	3.8961	0.5	882.6275	850	11.4
129	4.0583	0.5	888.5697	850	11.4
130	4.2439	0.5	895.3693	850	11.4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
131	4.4606	0.6	903.3082	850	11.4
132	4.7207	0.6	912.8370	850	11.4
133	5.0448	0.7	924.7106	850	11.4
134	5.4736	0.9	940.4198	850	11.4
135	6.1022	1.1	963.4488	850	11.4
136	7.2558	1.7	1005.7114	850	11.4
137	10.000	4.3	1106.2463	850	11.4

Table A.12.36 Conversion Table for Performance Level Setting Form: Mathematics Grade 3

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	14.8	421.0105	650	13.4
1	-4.3334	0.9	601.4872	650	13.4
2	-3.6635	0.6	622.8230	650	13.4
3	-3.2563	0.5	635.7920	650	13.4
4	-2.9514	0.5	645.5028	650	13.4
5	-2.7023	0.4	653.4364	653	13.4
6	-2.4888	0.4	660.2362	660	12.4
7	-2.3007	0.4	666.2270	666	11.7
8	-2.1316	0.3	671.6127	672	11.0
9	-1.9774	0.3	676.5239	677	10.5
10	-1.8353	0.3	681.0497	681	10.0
11	-1.7028	0.3	685.2697	685	9.7
12	-1.5784	0.3	689.2317	689	9.3
13	-1.4606	0.3	692.9836	693	9.1
14	-1.3482	0.3	696.5634	697	8.8
15	-1.2403	0.3	699.9999	700	8.6
16	-1.1362	0.3	703.3154	703	8.5
17	-1.0351	0.3	706.5354	707	8.4
18	-0.9365	0.3	709.6757	710	8.3
19	-0.8401	0.3	712.7460	713	8.2
20	-0.7456	0.3	715.7557	716	8.1
21	-0.6527	0.3	718.7145	719	8.1
22	-0.5613	0.3	721.6255	722	8.0
23	-0.4714	0.2	724.4888	724	8.0
24	-0.3830	0.2	727.3043	727	7.9
25	-0.2962	0.2	730.0688	730	7.9
26	-0.2109	0.2	732.7855	733	7.8
27	-0.1275	0.2	735.4417	735	7.7
28	-0.0460	0.2	738.0374	738	7.6
29	0.0335	0.2	740.5694	741	7.6
30	0.1108	0.2	743.0314	743	7.5
31	0.1860	0.2	745.4265	745	7.4
32	0.2589	0.2	747.7483	748	7.3
33	0.3296	0.2	750.0000	750	7.2
34	0.3982	0.2	752.1849	752	7.1
35	0.4646	0.2	754.2996	754	7.0
36	0.5291	0.2	756.3539	756	6.9
37	0.5917	0.2	758.3477	758	6.8
38	0.6525	0.2	760.2841	760	6.7
39	0.7117	0.2	762.1696	762	6.6
40	0.7694	0.2	764.0073	764	6.5
41	0.8256	0.2	765.7972	766	6.4
42	0.8806	0.2	767.5489	768	6.4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	0.9345	0.2	769.2656	769	6.3
44	0.9873	0.2	770.9472	771	6.3
45	1.0393	0.2	772.6034	773	6.2
46	1.0904	0.2	774.2309	774	6.2
47	1.1410	0.2	775.8424	776	6.2
48	1.1910	0.2	777.4349	777	6.1
49	1.2406	0.2	779.0146	779	6.1
50	1.2899	0.2	780.5848	781	6.1
51	1.3392	0.2	782.1549	782	6.1
52	1.3886	0.2	783.7283	784	6.1
53	1.4382	0.2	785.3080	785	6.2
54	1.4882	0.2	786.9005	787	6.2
55	1.5388	0.2	788.5120	789	6.3
56	1.5902	0.2	790.1491	790	6.3
57	1.6427	0.2	791.8212	792	6.4
58	1.6965	0.2	793.5347	794	6.5
59	1.7519	0.2	795.2991	795	6.6
60	1.8092	0.2	797.1241	797	6.7
61	1.8687	0.2	799.0191	799	6.9
62	1.9308	0.2	800.9969	801	7.1
63	1.9960	0.2	803.0735	803	7.3
64	2.0648	0.2	805.2647	805	7.5
65	2.1377	0.2	807.5865	808	7.7
66	2.2154	0.3	810.0612	810	8.0
67	2.2987	0.3	812.7143	813	8.3
68	2.3885	0.3	815.5743	816	8.7
69	2.4860	0.3	818.6796	819	9.1
70	2.5925	0.3	822.0716	822	9.6
71	2.7098	0.3	825.8075	826	10.1
72	2.8403	0.3	829.9638	830	10.8
73	2.9874	0.4	834.6488	835	11.6
74	3.1559	0.4	840.0154	840	12.6
75	3.3529	0.4	846.2897	846	13.9
76	3.5902	0.5	853.8475	850	13.9
77	3.8879	0.6	863.3290	850	13.9
78	4.2830	0.7	875.9126	850	13.9
79	4.8491	0.9	893.9425	850	13.9
80	5.7384	1.3	922.2659	850	13.9
81	7.3154	1.9	972.4921	850	13.9
82	15.000	12.8	1217.2405	850	13.9

Table A.12.37 Conversion Table for Performance Level Setting Form: Mathematics Grade 4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	17.1	442.1466	650	12.8
1	-4.0819	0.9	617.0726	650	12.8
2	-3.3997	0.6	637.2369	650	12.8
3	-3.0042	0.5	648.9271	650	12.8
4	-2.7224	0.4	657.2564	657	12.8
5	-2.5016	0.4	663.7828	664	11.4
6	-2.3182	0.4	669.2037	669	10.4
7	-2.1599	0.3	673.8827	674	9.7
8	-2.0194	0.3	678.0356	678	9.1
9	-1.8920	0.3	681.8012	682	8.7
10	-1.7746	0.3	685.2713	685	8.4
11	-1.6649	0.3	688.5138	689	8.1
12	-1.5613	0.3	691.5760	692	7.9
13	-1.4626	0.3	694.4934	694	7.7
14	-1.3678	0.3	697.2954	697	7.6
15	-1.2763	0.3	700.0000	700	7.5
16	-1.1874	0.3	702.6277	703	7.4
17	-1.1007	0.2	705.1903	705	7.3
18	-1.0158	0.2	707.6998	708	7.3
19	-0.9323	0.2	710.1679	710	7.2
20	-0.8501	0.2	712.5975	713	7.2
21	-0.7688	0.2	715.0006	715	7.2
22	-0.6885	0.2	717.3741	717	7.2
23	-0.6089	0.2	719.7269	720	7.2
24	-0.5301	0.2	722.0560	722	7.2
25	-0.4519	0.2	724.3674	724	7.2
26	-0.3745	0.2	726.6552	727	7.2
27	-0.2979	0.2	728.9193	729	7.1
28	-0.2221	0.2	731.1598	731	7.1
29	-0.1471	0.2	733.3766	733	7.1
30	-0.0731	0.2	735.5639	736	7.1
31	0.0000	0.2	737.7246	738	7.1
32	0.0720	0.2	739.8528	740	7.0
33	0.1429	0.2	741.9484	742	7.0
34	0.2127	0.2	744.0115	744	6.9
35	0.2813	0.2	746.0392	746	6.9
36	0.3489	0.2	748.0373	748	6.8
37	0.4153	0.2	750.0000	750	6.8
38	0.4806	0.2	751.9301	752	6.7
39	0.5448	0.2	753.8277	754	6.7
40	0.6080	0.2	755.6957	756	6.7
41	0.6703	0.2	757.5372	758	6.6
42	0.7316	0.2	759.3491	759	6.6

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	0.7920	0.2	761.1344	761	6.5
44	0.8517	0.2	762.8990	763	6.5
45	0.9107	0.2	764.6429	765	6.5
46	0.9690	0.2	766.3661	766	6.4
47	1.0269	0.2	768.0775	768	6.4
48	1.0843	0.2	769.7741	770	6.4
49	1.1414	0.2	771.4619	771	6.4
50	1.1982	0.2	773.1408	773	6.4
51	1.2550	0.2	774.8196	775	6.4
52	1.3117	0.2	776.4956	776	6.4
53	1.3686	0.2	778.1774	778	6.4
54	1.4257	0.2	779.8652	780	6.4
55	1.4833	0.2	781.5677	782	6.4
56	1.5413	0.2	783.2820	783	6.5
57	1.6002	0.2	785.0230	785	6.5
58	1.6598	0.2	786.7846	787	6.6
59	1.7206	0.2	788.5818	789	6.6
60	1.7827	0.2	790.4173	790	6.7
61	1.8463	0.2	792.2972	792	6.8
62	1.9116	0.2	794.2273	794	6.9
63	1.9791	0.2	796.2224	796	7.0
64	2.0490	0.2	798.2885	798	7.1
65	2.1217	0.2	800.4374	800	7.3
66	2.1976	0.3	802.6808	803	7.5
67	2.2773	0.3	805.0366	805	7.7
68	2.3613	0.3	807.5194	808	7.9
69	2.4504	0.3	810.1530	810	8.2
70	2.5452	0.3	812.9551	813	8.5
71	2.6466	0.3	815.9523	816	8.9
72	2.7556	0.3	819.1741	819	9.3
73	2.8733	0.3	822.6530	823	9.7
74	3.0011	0.3	826.4305	826	10.2
75	3.1411	0.4	830.5686	831	10.8
76	3.2964	0.4	835.1589	835	11.6
77	3.4724	0.4	840.3611	840	12.5
78	3.6792	0.5	846.4737	846	13.9
79	3.9381	0.5	854.1262	850	13.9
80	4.3004	0.7	864.8350	850	13.9
81	4.9439	1.0	883.8554	850	13.9
82	15.000	93.2	1181.0916	850	13.9

Table A.12.38 Conversion Table for Performance Level Setting Form: Mathematics Grade 5

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	9.6	447.0871	650	16.5
1	-4.6046	1.3	603.4485	650	16.5
2	-3.6649	0.8	630.6815	650	16.5
3	-3.1278	0.7	646.2469	650	16.5
4	-2.7508	0.6	657.1725	657	16.5
5	-2.4589	0.5	665.6319	666	14.6
6	-2.2194	0.5	672.5728	673	13.2
7	-2.0152	0.4	678.4906	678	12.2
8	-1.8361	0.4	683.6810	684	11.4
9	-1.6760	0.4	688.3208	688	10.8
10	-1.5306	0.4	692.5345	693	10.3
11	-1.3970	0.3	696.4063	696	9.8
12	-1.2730	0.3	699.9999	700	9.4
13	-1.1571	0.3	703.3588	703	9.1
14	-1.0481	0.3	706.5176	707	8.8
15	-0.9450	0.3	709.5055	710	8.5
16	-0.8471	0.3	712.3427	712	8.3
17	-0.7538	0.3	715.0466	715	8.1
18	-0.6646	0.3	717.6317	718	7.9
19	-0.5789	0.3	720.1153	720	7.7
20	-0.4965	0.3	722.5033	723	7.6
21	-0.4170	0.3	724.8072	725	7.4
22	-0.3401	0.3	727.0358	727	7.3
23	-0.2655	0.2	729.1978	729	7.2
24	-0.1931	0.2	731.2960	731	7.1
25	-0.1225	0.2	733.3420	733	7.0
26	-0.0537	0.2	735.3358	735	6.9
27	0.0136	0.2	737.2862	737	6.8
28	0.0795	0.2	739.1960	739	6.7
29	0.1442	0.2	741.0711	741	6.7
30	0.2077	0.2	742.9113	743	6.6
31	0.2702	0.2	744.7226	745	6.6
32	0.3317	0.2	746.5049	747	6.5
33	0.3924	0.2	748.2640	748	6.5
34	0.4523	0.2	750.0000	750	6.4
35	0.5116	0.2	751.7185	752	6.4
36	0.5703	0.2	753.4197	753	6.4
37	0.6284	0.2	755.1034	755	6.4
38	0.6861	0.2	756.7756	757	6.3
39	0.7433	0.2	758.4333	758	6.3
40	0.8002	0.2	760.0823	760	6.3
41	0.8568	0.2	761.7226	762	6.3
42	0.9131	0.2	763.3542	763	6.2

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	0.9693	0.2	764.9829	765	6.2
44	1.0253	0.2	766.6058	767	6.2
45	1.0812	0.2	768.2258	768	6.2
46	1.1370	0.2	769.8429	770	6.2
47	1.1928	0.2	771.4600	771	6.2
48	1.2487	0.2	773.0801	773	6.2
49	1.3046	0.2	774.7001	775	6.2
50	1.3606	0.2	776.3230	776	6.2
51	1.4168	0.2	777.9517	778	6.2
52	1.4731	0.2	779.5833	780	6.2
53	1.5295	0.2	781.2178	781	6.2
54	1.5862	0.2	782.8610	783	6.2
55	1.6431	0.2	784.5100	785	6.2
56	1.7003	0.2	786.1676	786	6.2
57	1.7577	0.2	787.8311	788	6.2
58	1.8156	0.2	789.5091	790	6.2
59	1.8739	0.2	791.1987	791	6.2
60	1.9328	0.2	792.9056	793	6.2
61	1.9925	0.2	794.6357	795	6.3
62	2.0533	0.2	796.3978	796	6.4
63	2.1154	0.2	798.1974	798	6.4
64	2.1794	0.2	800.0522	800	6.6
65	2.2458	0.2	801.9765	802	6.7
66	2.3151	0.2	803.9849	804	6.9
67	2.3881	0.2	806.1004	806	7.1
68	2.4657	0.3	808.3493	808	7.4
69	2.5490	0.3	810.7634	811	7.7
70	2.6391	0.3	813.3745	813	8.1
71	2.7377	0.3	816.2320	816	8.6
72	2.8464	0.3	819.3822	819	9.1
73	2.9677	0.3	822.8975	823	9.7
74	3.1045	0.4	826.8621	827	10.5
75	3.2610	0.4	831.3975	831	11.4
76	3.4432	0.4	836.6778	837	12.5
77	3.6601	0.5	842.9636	843	13.9
78	3.9269	0.5	850.6956	850	13.9
79	4.2722	0.6	860.7026	850	13.9
80	4.7611	0.8	874.8712	850	13.9
81	5.6079	1.2	899.4118	850	13.9
82	15.000	38.9	1171.5996	850	13.9

Table A.12.39 Conversion Table for Performance Level Setting Form: Mathematics Grade 6

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	15.2	449.8140	650	16.2
1	-3.9636	1.1	622.8064	650	16.2
2	-3.1760	0.7	645.3776	650	16.2
3	-2.7271	0.6	658.2422	658	16.2
4	-2.4090	0.5	667.3584	667	13.9
5	-2.1615	0.4	674.4513	674	12.3
6	-1.9583	0.4	680.2746	680	11.2
7	-1.7856	0.4	685.2239	685	10.3
8	-1.6350	0.3	689.5398	690	9.6
9	-1.5012	0.3	693.3743	693	9.1
10	-1.3804	0.3	696.8362	697	8.6
11	-1.2700	0.3	700.0001	700	8.2
12	-1.1678	0.3	702.9290	703	7.9
13	-1.0725	0.3	705.6601	706	7.6
14	-0.9829	0.3	708.2279	708	7.4
15	-0.8979	0.3	710.6638	711	7.2
16	-0.8170	0.2	712.9823	713	7.0
17	-0.7395	0.2	715.2033	715	6.8
18	-0.6650	0.2	717.3383	717	6.7
19	-0.5931	0.2	719.3988	719	6.6
20	-0.5233	0.2	721.3992	721	6.5
21	-0.4556	0.2	723.3393	723	6.4
22	-0.3896	0.2	725.2308	725	6.3
23	-0.3251	0.2	727.0792	727	6.2
24	-0.2620	0.2	728.8876	729	6.2
25	-0.2001	0.2	730.6615	731	6.1
26	-0.1394	0.2	732.4010	732	6.0
27	-0.0797	0.2	734.1119	734	6.0
28	-0.0209	0.2	735.7970	736	5.9
29	0.0371	0.2	737.4592	737	5.9
30	0.0942	0.2	739.0956	739	5.9
31	0.1506	0.2	740.7119	741	5.8
32	0.2062	0.2	742.3053	742	5.8
33	0.2612	0.2	743.8815	744	5.8
34	0.3155	0.2	745.4377	745	5.7
35	0.3692	0.2	746.9766	747	5.7
36	0.4222	0.2	748.4955	748	5.7
37	0.4747	0.2	750.0000	750	5.6
38	0.5267	0.2	751.4903	751	5.6
39	0.5781	0.2	752.9633	753	5.6
40	0.6290	0.2	754.4220	754	5.6
41	0.6794	0.2	755.8664	756	5.6
42	0.7295	0.2	757.3022	757	5.5

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	0.7791	0.2	758.7236	759	5.5
44	0.8285	0.2	760.1393	760	5.5
45	0.8777	0.2	761.5493	762	5.5
46	0.9267	0.2	762.9536	763	5.5
47	0.9755	0.2	764.3521	764	5.5
48	1.0244	0.2	765.7535	766	5.5
49	1.0733	0.2	767.1548	767	5.5
50	1.1224	0.2	768.5620	769	5.6
51	1.1716	0.2	769.9719	770	5.6
52	1.2212	0.2	771.3934	771	5.6
53	1.2712	0.2	772.8263	773	5.6
54	1.3216	0.2	774.2707	774	5.7
55	1.3726	0.2	775.7322	776	5.7
56	1.4243	0.2	777.2139	777	5.7
57	1.4767	0.2	778.7156	779	5.8
58	1.5300	0.2	780.2430	780	5.8
59	1.5843	0.2	781.7992	782	5.9
60	1.6397	0.2	783.3869	783	5.9
61	1.6963	0.2	785.0089	785	6.0
62	1.7542	0.2	786.6682	787	6.1
63	1.8136	0.2	788.3705	788	6.1
64	1.8746	0.2	790.1187	790	6.2
65	1.9374	0.2	791.9184	792	6.3
66	2.0023	0.2	793.7783	794	6.4
67	2.0695	0.2	795.7041	796	6.5
68	2.1394	0.2	797.7074	798	6.6
69	2.2125	0.2	799.8023	800	6.7
70	2.2896	0.2	802.0118	802	6.9
71	2.3717	0.3	804.3647	804	7.2
72	2.4604	0.3	806.9066	807	7.5
73	2.5579	0.3	809.7008	810	7.9
74	2.6674	0.3	812.8389	813	8.5
75	2.7934	0.3	816.4498	816	9.2
76	2.9430	0.4	820.7371	821	10.2
77	3.1271	0.4	826.0131	826	11.6
78	3.3645	0.5	832.8165	833	13.6
79	3.6916	0.6	842.1906	842	16.7
80	4.1992	0.8	856.7375	850	16.7
81	5.2729	1.4	887.5078	850	16.7
82	15.000	15.8	1166.2690	850	16.7

Table A.12.40 Conversion Table for Performance Level Setting Form: Mathematics Grade 7

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	6.2	477.8425	650	20.0
1	-6.0057	2.9	580.4077	650	20.0
2	-3.8634	1.3	635.4175	650	20.0
3	-2.9716	0.8	658.3171	658	20.0
4	-2.4450	0.6	671.8390	672	16.2
5	-2.0742	0.5	681.3604	681	13.6
6	-1.7867	0.5	688.7428	689	11.9
7	-1.5504	0.4	694.8105	695	10.8
8	-1.3483	0.4	700.0000	700	9.9
9	-1.1706	0.4	704.5630	705	9.3
10	-1.0111	0.3	708.6586	709	8.7
11	-0.8654	0.3	712.3998	712	8.3
12	-0.7307	0.3	715.8587	716	8.0
13	-0.6047	0.3	719.0941	719	7.7
14	-0.4860	0.3	722.1420	722	7.5
15	-0.3733	0.3	725.0359	725	7.3
16	-0.2657	0.3	727.7989	728	7.1
17	-0.1628	0.3	730.4411	730	7.0
18	-0.0640	0.3	732.9781	733	6.8
19	0.0309	0.3	735.4149	735	6.7
20	0.1222	0.3	737.7593	738	6.5
21	0.2099	0.2	740.0113	740	6.4
22	0.2942	0.2	742.1759	742	6.3
23	0.3751	0.2	744.2533	744	6.1
24	0.4528	0.2	746.2485	746	6.0
25	0.5273	0.2	748.1615	748	5.9
26	0.5989	0.2	750.0000	750	5.8
27	0.6675	0.2	751.7615	752	5.6
28	0.7335	0.2	753.4562	753	5.5
29	0.7969	0.2	755.0842	755	5.4
30	0.8581	0.2	756.6557	757	5.3
31	0.9172	0.2	758.1733	758	5.2
32	0.9743	0.2	759.6395	760	5.1
33	1.0297	0.2	761.0620	761	5.0
34	1.0835	0.2	762.4435	762	5.0
35	1.1360	0.2	763.7916	764	4.9
36	1.1871	0.2	765.1037	765	4.8
37	1.2372	0.2	766.3902	766	4.8
38	1.2862	0.2	767.6484	768	4.7
39	1.3343	0.2	768.8835	769	4.7
40	1.3815	0.2	770.0955	770	4.6
41	1.4281	0.2	771.2921	771	4.6
42	1.4740	0.2	772.4707	772	4.6

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	1.5193	0.2	773.6339	774	4.5
44	1.5643	0.2	774.7894	775	4.5
45	1.6088	0.2	775.9321	776	4.5
46	1.6532	0.2	777.0722	777	4.5
47	1.6973	0.2	778.2046	778	4.5
48	1.7415	0.2	779.3396	779	4.5
49	1.7857	0.2	780.4745	780	4.5
50	1.8301	0.2	781.6146	782	4.5
51	1.8750	0.2	782.7676	783	4.5
52	1.9203	0.2	783.9308	784	4.6
53	1.9663	0.2	785.1120	785	4.6
54	2.0131	0.2	786.3137	786	4.7
55	2.0610	0.2	787.5437	788	4.7
56	2.1100	0.2	788.8019	789	4.8
57	2.1605	0.2	790.0986	790	4.9
58	2.2126	0.2	791.4364	791	5.0
59	2.2665	0.2	792.8205	793	5.1
60	2.3226	0.2	794.2610	794	5.2
61	2.3810	0.2	795.7606	796	5.4
62	2.4422	0.2	797.3321	797	5.5
63	2.5065	0.2	798.9832	799	5.7
64	2.5743	0.2	800.7241	801	5.9
65	2.6461	0.2	802.5678	803	6.1
66	2.7225	0.2	804.5296	805	6.3
67	2.8043	0.3	806.6300	807	6.6
68	2.8924	0.3	808.8923	809	6.9
69	2.9878	0.3	811.3419	811	7.3
70	3.0922	0.3	814.0227	814	7.7
71	3.2073	0.3	816.9782	817	8.2
72	3.3360	0.3	820.2830	820	8.7
73	3.4818	0.4	824.0268	824	9.4
74	3.6502	0.4	828.3510	828	10.3
75	3.8494	0.4	833.4660	833	11.4
76	4.0928	0.5	839.7160	840	13.0
77	4.4038	0.6	847.7018	848	15.2
78	4.8289	0.7	858.6175	850	15.2
79	5.4749	1.0	875.2054	850	15.2
80	6.6447	1.6	905.2434	850	15.2
81	9.4242	3.7	976.6152	850	15.2
82	15.000	8.1	1119.7900	850	15.2

Table A.12.41 Conversion Table for Performance Level Setting Form: Mathematics Grade 8

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	9.1	392.6327	650	18.3
1	-4.3997	1.5	580.5877	650	18.3
2	-3.3280	0.9	616.5557	650	18.3
3	-2.7351	0.7	636.4544	650	18.3
4	-2.3239	0.6	650.2549	650	18.3
5	-2.0068	0.5	660.8973	661	18.3
6	-1.7466	0.5	669.6300	670	16.6
7	-1.5240	0.5	677.1008	677	15.4
8	-1.3280	0.4	683.6789	684	14.5
9	-1.1517	0.4	689.5958	690	13.7
10	-0.9906	0.4	695.0026	695	13.1
11	-0.8417	0.4	699.9999	700	12.6
12	-0.7026	0.4	704.6683	705	12.2
13	-0.5720	0.4	709.0515	709	11.8
14	-0.4487	0.3	713.1896	713	11.5
15	-0.3319	0.3	717.1096	717	11.2
16	-0.2209	0.3	720.8349	721	10.9
17	-0.1152	0.3	724.3824	724	10.6
18	-0.0144	0.3	727.7654	728	10.3
19	0.0818	0.3	730.9940	731	10.1
20	0.1737	0.3	734.0783	734	9.8
21	0.2616	0.3	737.0284	737	9.5
22	0.3456	0.3	739.8476	740	9.3
23	0.4261	0.3	742.5493	743	9.1
24	0.5032	0.3	745.1369	745	8.8
25	0.5771	0.3	747.6171	748	8.6
26	0.6481	0.3	750.0000	750	8.4
27	0.7164	0.2	752.2922	752	8.2
28	0.7822	0.2	754.5006	755	8.0
29	0.8456	0.2	756.6284	757	7.9
30	0.9070	0.2	758.6891	759	7.7
31	0.9664	0.2	760.6826	761	7.6
32	1.0241	0.2	762.6191	763	7.4
33	1.0802	0.2	764.5019	765	7.3
34	1.1350	0.2	766.3411	766	7.2
35	1.1885	0.2	768.1367	768	7.1
36	1.2408	0.2	769.8919	770	7.0
37	1.2923	0.2	771.6204	772	6.9
38	1.3429	0.2	773.3186	773	6.8
39	1.3927	0.2	774.9899	775	6.7
40	1.4420	0.2	776.6445	777	6.7
41	1.4908	0.2	778.2823	778	6.6
42	1.5393	0.2	779.9101	780	6.6

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	1.5874	0.2	781.5244	782	6.5
44	1.6354	0.2	783.1353	783	6.5
45	1.6833	0.2	784.7429	785	6.4
46	1.7312	0.2	786.3505	786	6.4
47	1.7792	0.2	787.9615	788	6.4
48	1.8274	0.2	789.5792	790	6.4
49	1.8759	0.2	791.2069	791	6.4
50	1.9247	0.2	792.8447	793	6.4
51	1.9739	0.2	794.4959	794	6.4
52	2.0236	0.2	796.1640	796	6.4
53	2.0740	0.2	797.8555	798	6.4
54	2.1251	0.2	799.5705	800	6.4
55	2.1770	0.2	801.3123	801	6.5
56	2.2299	0.2	803.0877	803	6.5
57	2.2838	0.2	804.8967	805	6.6
58	2.3389	0.2	806.7459	807	6.6
59	2.3955	0.2	808.6455	809	6.7
60	2.4537	0.2	810.5988	811	6.8
61	2.5137	0.2	812.6125	813	6.9
62	2.5758	0.2	814.6967	815	7.1
63	2.6404	0.2	816.8647	817	7.2
64	2.7079	0.2	819.1302	819	7.4
65	2.7788	0.2	821.5097	822	7.6
66	2.8536	0.2	824.0201	824	7.9
67	2.9330	0.2	826.6849	827	8.2
68	3.0180	0.3	829.5376	830	8.6
69	3.1096	0.3	832.6119	833	9.0
70	3.2092	0.3	835.9546	836	9.5
71	3.3185	0.3	839.6229	840	10.1
72	3.4397	0.3	843.6905	844	10.9
73	3.5757	0.4	848.2549	848	11.8
74	3.7303	0.4	853.4435	850	11.8
75	3.9088	0.4	859.4343	850	11.8
76	4.1191	0.5	866.4923	850	11.8
77	4.3729	0.5	875.0102	850	11.8
78	4.6906	0.6	885.6727	850	11.8
79	5.1106	0.8	899.7686	850	11.8
80	5.7222	1.0	920.2949	850	11.8
81	6.8276	1.6	957.3939	850	11.8
82	15.000	17.2	1231.6727	850	11.8

Table A.12.42 Conversion Table for Performance Level Setting Form: Algebra I

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	7.4	430.2271	650	20.0
1	-6.1183	2.7	547.9470	650	20.0
2	-4.0393	1.4	610.9967	650	20.0
3	-3.1329	1.0	638.4850	650	20.0
4	-2.5610	0.8	655.8289	656	20.0
5	-2.1445	0.7	668.4601	668	20.0
6	-1.8162	0.6	678.4164	678	17.7
7	-1.5441	0.5	686.6683	687	16.1
8	-1.3103	0.5	693.7588	694	14.8
9	-1.1045	0.5	700.0000	700	13.8
10	-0.9197	0.4	705.6045	706	13.0
11	-0.7513	0.4	710.7115	711	12.3
12	-0.5962	0.4	715.4152	715	11.8
13	-0.4521	0.4	719.7853	720	11.3
14	-0.3170	0.4	723.8825	724	10.9
15	-0.1899	0.3	727.7370	728	10.5
16	-0.0696	0.3	731.3853	731	10.2
17	0.0447	0.3	734.8517	735	9.9
18	0.1536	0.3	738.1543	738	9.6
19	0.2576	0.3	741.3083	741	9.4
20	0.3572	0.3	744.3289	744	9.2
21	0.4526	0.3	747.2221	747	9.0
22	0.5442	0.3	750.0000	750	8.8
23	0.6322	0.3	752.6688	753	8.6
24	0.7167	0.3	755.2314	755	8.4
25	0.7981	0.3	757.7000	758	8.3
26	0.8763	0.3	760.0716	760	8.1
27	0.9517	0.3	762.3582	762	8.0
28	1.0244	0.3	764.5630	765	7.8
29	1.0946	0.3	766.6919	767	7.7
30	1.1624	0.2	768.7481	769	7.5
31	1.2280	0.2	770.7375	771	7.4
32	1.2916	0.2	772.6663	773	7.3
33	1.3533	0.2	774.5375	775	7.2
34	1.4132	0.2	776.3541	776	7.1
35	1.4717	0.2	778.1282	778	7.0
36	1.5286	0.2	779.8538	780	6.9
37	1.5843	0.2	781.5430	782	6.8
38	1.6388	0.2	783.1958	783	6.7
39	1.6922	0.2	784.8153	785	6.6
40	1.7446	0.2	786.4044	786	6.6
41	1.7962	0.2	787.9693	788	6.5
42	1.8469	0.2	789.5069	790	6.4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	1.8970	0.2	791.0262	791	6.4
44	1.9464	0.2	792.5244	793	6.3
45	1.9952	0.2	794.0043	794	6.3
46	2.0435	0.2	795.4691	795	6.2
47	2.0913	0.2	796.9187	797	6.2
48	2.1388	0.2	798.3593	798	6.1
49	2.1859	0.2	799.7877	800	6.1
50	2.2326	0.2	801.2039	801	6.0
51	2.2792	0.2	802.6172	803	6.0
52	2.3255	0.2	804.0213	804	6.0
53	2.3716	0.2	805.4194	805	6.0
54	2.4176	0.2	806.8144	807	5.9
55	2.4635	0.2	808.2064	808	5.9
56	2.5093	0.2	809.5954	810	5.9
57	2.5552	0.2	810.9874	811	5.9
58	2.6011	0.2	812.3794	812	5.9
59	2.6470	0.2	813.7714	814	5.9
60	2.6931	0.2	815.1695	815	5.9
61	2.7393	0.2	816.5706	817	5.9
62	2.7857	0.2	817.9777	818	5.9
63	2.8323	0.2	819.3910	819	5.9
64	2.8791	0.2	820.8103	821	5.9
65	2.9263	0.2	822.2417	822	6.0
66	2.9737	0.2	823.6792	824	6.0
67	3.0216	0.2	825.1319	825	6.0
68	3.0698	0.2	826.5936	827	6.0
69	3.1184	0.2	828.0675	828	6.1
70	3.1676	0.2	829.5596	830	6.1
71	3.2174	0.2	831.0699	831	6.1
72	3.2679	0.2	832.6014	833	6.2
73	3.3193	0.2	834.1602	834	6.2
74	3.3717	0.2	835.7493	836	6.3
75	3.4254	0.2	837.3779	837	6.3
76	3.4807	0.2	839.0549	839	6.4
77	3.5380	0.2	840.7927	841	6.5
78	3.5978	0.2	842.6062	843	6.7
79	3.6608	0.2	844.5168	845	6.9
80	3.7277	0.2	846.5457	847	7.1
81	3.7993	0.2	848.7171	849	7.4
82	3.8769	0.3	851.0705	850	7.4
83	3.9617	0.3	853.6422	850	7.4
84	4.0553	0.3	856.4808	850	7.4
85	4.1595	0.3	859.6408	850	7.4
86	4.2767	0.3	863.1952	850	7.4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	4.4094	0.4	867.2195	850	7.4
88	4.5613	0.4	871.8262	850	7.4
89	4.7370	0.4	877.1546	850	7.4
90	4.9436	0.5	883.4202	850	7.4
91	5.1915	0.5	890.9382	850	7.4
92	5.4985	0.6	900.2486	850	7.4
93	5.8955	0.7	912.2883	850	7.4
94	6.4424	0.9	928.8741	850	7.4
95	7.2748	1.2	954.1182	850	7.4
96	8.8244	2.0	1001.1128	850	7.4
97	15.000	9.1	1188.3996	850	7.4

Table A.12.43 Conversion Table for Performance Level Setting Form: Algebra II

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	9.4	363.0005	650	20.0
1	-4.3378	1.5	564.6745	650	20.0
2	-3.2325	1.0	604.0426	650	20.0
3	-2.6144	0.8	626.0578	650	20.0
4	-2.1834	0.7	641.4090	650	20.0
5	-1.8513	0.6	653.2376	653	20.0
6	-1.5804	0.5	662.8864	663	19.5
7	-1.3511	0.5	671.0536	671	18.0
8	-1.1519	0.5	678.1486	678	16.8
9	-0.9753	0.4	684.4387	684	15.8
10	-0.8164	0.4	690.0983	690	15.0
11	-0.6716	0.4	695.2557	695	14.3
12	-0.5384	0.4	700.0000	700	13.7
13	-0.4146	0.4	704.4094	704	13.2
14	-0.2988	0.4	708.5340	709	12.8
15	-0.1897	0.3	712.4198	712	12.4
16	-0.0865	0.3	716.0956	716	12.0
17	0.0116	0.3	719.5897	720	11.7
18	0.1053	0.3	722.9270	723	11.4
19	0.1949	0.3	726.1184	726	11.1
20	0.2809	0.3	729.1815	729	10.8
21	0.3635	0.3	732.1235	732	10.5
22	0.4431	0.3	734.9587	735	10.3
23	0.5198	0.3	737.6905	738	10.0
24	0.5937	0.3	740.3227	740	9.8
25	0.6651	0.3	742.8658	743	9.6
26	0.7341	0.3	745.3234	745	9.3
27	0.8008	0.3	747.6991	748	9.1
28	0.8654	0.3	750.0000	750	8.9
29	0.9279	0.2	752.2261	752	8.7
30	0.9886	0.2	754.3881	754	8.5
31	1.0476	0.2	756.4895	756	8.4
32	1.1049	0.2	758.5304	759	8.2
33	1.1607	0.2	760.5178	761	8.1
34	1.2152	0.2	762.4590	762	7.9
35	1.2685	0.2	764.3574	764	7.8
36	1.3206	0.2	766.2131	766	7.7
37	1.3718	0.2	768.0367	768	7.6
38	1.4220	0.2	769.8247	770	7.5
39	1.4714	0.2	771.5842	772	7.4
40	1.5200	0.2	773.3153	773	7.3
41	1.5679	0.2	775.0213	775	7.2
42	1.6151	0.2	776.7025	777	7.1

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	1.6618	0.2	778.3658	778	7.1
44	1.7079	0.2	780.0078	780	7.0
45	1.7535	0.2	781.6320	782	6.9
46	1.7986	0.2	783.2383	783	6.9
47	1.8432	0.2	784.8269	785	6.8
48	1.8874	0.2	786.4012	786	6.8
49	1.9312	0.2	787.9612	788	6.7
50	1.9746	0.2	789.5070	790	6.7
51	2.0177	0.2	791.0421	791	6.6
52	2.0605	0.2	792.5666	793	6.6
53	2.1030	0.2	794.0803	794	6.6
54	2.1453	0.2	795.5869	796	6.5
55	2.1874	0.2	797.0864	797	6.5
56	2.2294	0.2	798.5824	799	6.5
57	2.2714	0.2	800.0783	800	6.5
58	2.3133	0.2	801.5707	802	6.5
59	2.3553	0.2	803.0666	803	6.5
60	2.3974	0.2	804.5661	805	6.5
61	2.4397	0.2	806.0728	806	6.5
62	2.4823	0.2	807.5901	808	6.5
63	2.5253	0.2	809.1216	809	6.5
64	2.5687	0.2	810.6674	811	6.6
65	2.6127	0.2	812.2346	812	6.6
66	2.6574	0.2	813.8267	814	6.6
67	2.7029	0.2	815.4473	815	6.7
68	2.7493	0.2	817.1000	817	6.8
69	2.7967	0.2	818.7882	819	6.8
70	2.8453	0.2	820.5193	821	6.9
71	2.8952	0.2	822.2966	822	7.0
72	2.9465	0.2	824.1238	824	7.2
73	2.9995	0.2	826.0115	826	7.3
74	3.0543	0.2	827.9633	828	7.4
75	3.1111	0.2	829.9864	830	7.6
76	3.1700	0.2	832.0843	832	7.8
77	3.2312	0.2	834.2641	834	8.0
78	3.2949	0.2	836.5329	837	8.2
79	3.3613	0.2	838.8979	839	8.4
80	3.4307	0.2	841.3698	841	8.7
81	3.5033	0.3	843.9556	844	8.9
82	3.5795	0.3	846.6697	847	9.2
83	3.6596	0.3	849.5227	850	9.2
84	3.7442	0.3	852.5359	850	9.2
85	3.8340	0.3	855.7344	850	9.2
86	3.9299	0.3	859.1501	850	9.2

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	4.0328	0.3	862.8152	850	9.2
88	4.1442	0.3	866.7830	850	9.2
89	4.2657	0.4	871.1105	850	9.2
90	4.3992	0.4	875.8654	850	9.2
91	4.5473	0.4	881.1404	850	9.2
92	4.7132	0.4	887.0494	850	9.2
93	4.9013	0.5	893.7490	850	9.2
94	5.1184	0.5	901.4816	850	9.2
95	5.3748	0.6	910.6140	850	9.2
96	5.6881	0.7	921.7730	850	9.2
97	6.0892	0.8	936.0592	850	9.2
98	6.6384	1.0	955.6204	850	9.2
99	7.4770	1.3	985.4893	850	9.2
100	9.0828	2.1	1042.6840	850	9.2
101	15.000	8.1	1253.4405	850	9.2

Table A.12.44 Conversion Table for Performance Level Setting Form: Geometry

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	8.1	479.4306	650	20.0
1	-5.5752	2.1	590.9426	650	20.0
2	-3.9226	1.1	632.5908	650	20.0
3	-3.1489	0.9	652.0893	652	20.0
4	-2.6400	0.7	664.9144	665	17.7
5	-2.2570	0.6	674.5666	675	15.4
6	-1.9465	0.6	682.3917	682	13.9
7	-1.6830	0.5	689.0323	689	12.7
8	-1.4528	0.5	694.8337	695	11.9
9	-1.2478	0.4	700.0000	700	11.1
10	-1.0629	0.4	704.6598	705	10.5
11	-0.8946	0.4	708.9012	709	9.9
12	-0.7403	0.4	712.7899	713	9.4
13	-0.5978	0.4	716.3811	716	8.9
14	-0.4656	0.3	719.7127	720	8.5
15	-0.3422	0.3	722.8226	723	8.2
16	-0.2263	0.3	725.7435	726	7.9
17	-0.1171	0.3	728.4955	728	7.6
18	-0.0136	0.3	731.1039	731	7.4
19	0.0848	0.3	733.5837	734	7.2
20	0.1787	0.3	735.9501	736	7.0
21	0.2685	0.3	738.2132	738	6.8
22	0.3546	0.3	740.3831	740	6.7
23	0.4372	0.3	742.4647	742	6.5
24	0.5165	0.3	744.4632	744	6.4
25	0.5927	0.2	746.3836	746	6.3
26	0.6659	0.2	748.2283	748	6.1
27	0.7362	0.2	750.0000	750	6.0
28	0.8037	0.2	751.7011	752	5.9
29	0.8686	0.2	753.3367	753	5.8
30	0.9310	0.2	754.9093	755	5.6
31	0.9909	0.2	756.4189	756	5.5
32	1.0486	0.2	757.8730	758	5.4
33	1.1042	0.2	759.2742	759	5.3
34	1.1577	0.2	760.6225	761	5.2
35	1.2094	0.2	761.9254	762	5.1
36	1.2594	0.2	763.1855	763	5.0
37	1.3077	0.2	764.4027	764	4.9
38	1.3546	0.2	765.5847	766	4.8
39	1.4001	0.2	766.7314	767	4.8
40	1.4443	0.2	767.8453	768	4.7
41	1.4874	0.2	768.9315	769	4.6
42	1.5294	0.2	769.9899	770	4.5

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	1.5705	0.2	771.0257	771	4.5
44	1.6107	0.2	772.0388	772	4.4
45	1.6500	0.2	773.0292	773	4.4
46	1.6886	0.2	774.0020	774	4.3
47	1.7265	0.2	774.9572	775	4.3
48	1.7638	0.2	775.8972	776	4.2
49	1.8005	0.2	776.8221	777	4.2
50	1.8367	0.2	777.7344	778	4.1
51	1.8724	0.2	778.6341	779	4.1
52	1.9078	0.2	779.5262	780	4.1
53	1.9427	0.2	780.4057	780	4.0
54	1.9774	0.2	781.2802	781	4.0
55	2.0117	0.2	782.1447	782	4.0
56	2.0459	0.2	783.0066	783	4.0
57	2.0798	0.2	783.8609	784	3.9
58	2.1136	0.2	784.7127	785	3.9
59	2.1474	0.2	785.5645	786	3.9
60	2.1811	0.2	786.4138	786	3.9
61	2.2148	0.2	787.2631	787	3.9
62	2.2486	0.2	788.1149	788	3.9
63	2.2825	0.2	788.9693	789	3.9
64	2.3166	0.2	789.8286	790	3.9
65	2.3510	0.2	790.6956	791	3.9
66	2.3857	0.2	791.5701	792	3.9
67	2.4208	0.2	792.4546	792	4.0
68	2.4564	0.2	793.3518	793	4.0
69	2.4925	0.2	794.2616	794	4.0
70	2.5294	0.2	795.1915	795	4.1
71	2.5671	0.2	796.1416	796	4.1
72	2.6057	0.2	797.1144	797	4.2
73	2.6455	0.2	798.1174	798	4.2
74	2.6865	0.2	799.1507	799	4.3
75	2.7291	0.2	800.2243	800	4.4
76	2.7734	0.2	801.3407	801	4.5
77	2.8197	0.2	802.5076	803	4.6
78	2.8683	0.2	803.7323	804	4.7
79	2.9197	0.2	805.0277	805	4.8
80	2.9742	0.2	806.4012	806	5.0
81	3.0325	0.2	807.8705	808	5.2
82	3.0950	0.2	809.4456	809	5.4
83	3.1627	0.2	811.1517	811	5.6
84	3.2364	0.2	813.0091	813	5.9
85	3.3173	0.2	815.0479	815	6.2
86	3.4069	0.3	817.3059	817	6.6

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	3.5070	0.3	819.8286	820	7.0
88	3.6201	0.3	822.6789	823	7.5
89	3.7498	0.3	825.9476	826	8.1
90	3.9014	0.4	829.7681	830	8.9
91	4.0832	0.4	834.3498	834	9.9
92	4.3105	0.5	840.0781	840	11.4
93	4.6138	0.5	847.7217	848	13.8
94	5.0666	0.7	859.1330	850	13.8
95	5.9232	1.2	880.7207	850	13.8
96	15.000	25.0	1109.4706	850	13.8

Table A.12.45 Conversion Table for Performance Level Setting Form: Integrated Mathematics I

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	6.2	397.1990	650	20.0
1	-6.4170	2.8	516.3782	650	20.0
2	-4.3152	1.5	586.2891	650	20.0
3	-3.3198	1.0	619.3985	650	20.0
4	-2.7035	0.8	639.8981	650	20.0
5	-2.2690	0.7	654.3506	654	20.0
6	-1.9356	0.6	665.4403	665	19.0
7	-1.6644	0.5	674.4611	674	17.1
8	-1.4345	0.5	682.1081	682	15.7
9	-1.2341	0.4	688.7739	689	14.7
10	-1.0563	0.4	694.6879	695	13.8
11	-0.8966	0.4	699.9999	700	13.1
12	-0.7522	0.4	704.8030	705	12.5
13	-0.6209	0.4	709.1704	709	11.9
14	-0.5010	0.3	713.1585	713	11.4
15	-0.3911	0.3	716.8141	717	10.9
16	-0.2898	0.3	720.1836	720	10.5
17	-0.1957	0.3	723.3135	723	10.1
18	-0.1078	0.3	726.2373	726	9.8
19	-0.0248	0.3	728.9981	729	9.5
20	0.0540	0.3	731.6192	732	9.3
21	0.1295	0.3	734.1305	734	9.1
22	0.2025	0.3	736.5586	737	8.9
23	0.2733	0.3	738.9136	739	8.7
24	0.3425	0.3	741.2154	741	8.5
25	0.4103	0.3	743.4706	743	8.4
26	0.4768	0.2	745.6825	746	8.3
27	0.5423	0.2	747.8612	748	8.1
28	0.6066	0.2	750.0000	750	8.0
29	0.6698	0.2	752.1022	752	7.9
30	0.7319	0.2	754.1678	754	7.7
31	0.7926	0.2	756.1868	756	7.6
32	0.8521	0.2	758.1659	758	7.5
33	0.9103	0.2	760.1018	760	7.3
34	0.9671	0.2	761.9911	762	7.2
35	1.0227	0.2	763.8405	764	7.1
36	1.0771	0.2	765.6499	766	7.0
37	1.1303	0.2	767.4195	767	6.9
38	1.1824	0.2	769.1525	769	6.8
39	1.2335	0.2	770.8522	771	6.7
40	1.2837	0.2	772.5219	773	6.6
41	1.3330	0.2	774.1618	774	6.6
42	1.3816	0.2	775.7783	776	6.5

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	1.4294	0.2	777.3683	777	6.4
44	1.4767	0.2	778.9416	779	6.4
45	1.5234	0.2	780.4949	780	6.3
46	1.5697	0.2	782.0350	782	6.3
47	1.6155	0.2	783.5584	784	6.3
48	1.6611	0.2	785.0752	785	6.2
49	1.7064	0.2	786.5820	787	6.2
50	1.7516	0.2	788.0854	788	6.2
51	1.7967	0.2	789.5856	790	6.2
52	1.8418	0.2	791.0857	791	6.2
53	1.8872	0.2	792.5958	793	6.2
54	1.9327	0.2	794.1092	794	6.2
55	1.9786	0.2	795.6360	796	6.3
56	2.0250	0.2	797.1794	797	6.3
57	2.0720	0.2	798.7427	799	6.4
58	2.1196	0.2	800.3260	800	6.4
59	2.1681	0.2	801.9392	802	6.5
60	2.2176	0.2	803.5857	804	6.5
61	2.2681	0.2	805.2654	805	6.6
62	2.3199	0.2	806.9884	807	6.7
63	2.3731	0.2	808.7580	809	6.8
64	2.4277	0.2	810.5741	811	6.9
65	2.4841	0.2	812.4501	812	7.0
66	2.5423	0.2	814.3860	814	7.1
67	2.6024	0.2	816.3851	816	7.3
68	2.6647	0.2	818.4573	818	7.4
69	2.7294	0.2	820.6094	821	7.6
70	2.7965	0.2	822.8413	823	7.7
71	2.8663	0.2	825.1630	825	7.9
72	2.9391	0.2	827.5845	828	8.1
73	3.0151	0.3	830.1125	830	8.3
74	3.0947	0.3	832.7601	833	8.6
75	3.1786	0.3	835.5509	836	8.8
76	3.2674	0.3	838.5046	839	9.2
77	3.3623	0.3	841.6612	842	9.6
78	3.4648	0.3	845.0706	845	10.1
79	3.5771	0.3	848.8059	849	10.7
80	3.7018	0.3	852.9538	850	10.7
81	3.8427	0.4	857.6404	850	10.7
82	4.0046	0.4	863.0256	850	10.7
83	4.1939	0.5	869.3222	850	10.7
84	4.4193	0.5	876.8195	850	10.7
85	4.6930	0.6	885.9234	850	10.7
86	5.0335	0.7	897.2493	850	10.7

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	5.4703	0.8	911.7783	850	10.7
88	6.0520	1.0	931.1270	850	10.7
89	6.8669	1.3	958.2326	850	10.7
90	8.1295	1.8	1000.2297	850	10.7
91	10.7743	3.5	1088.2021	850	10.7
92	15.000	7.4	1228.7590	850	10.7

Table A.12.46 Conversion Table for Performance Level Setting Form: Integrated Mathematics II

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	6.6	434.1360	650	20.0
1	-6.0191	2.3	550.8233	650	20.0
2	-4.2559	1.3	602.5059	650	20.0
3	-3.3887	1.0	627.9251	650	20.0
4	-2.8196	0.8	644.6064	650	20.0
5	-2.3961	0.7	657.0200	657	20.0
6	-2.0573	0.6	666.9508	667	18.1
7	-1.7731	0.6	675.2812	675	16.5
8	-1.5267	0.5	682.5037	683	15.3
9	-1.3080	0.5	688.9142	689	14.4
10	-1.1105	0.5	694.7032	695	13.6
11	-0.9298	0.4	699.9999	700	13.0
12	-0.7628	0.4	704.8950	705	12.4
13	-0.6075	0.4	709.4471	709	11.9
14	-0.4622	0.4	713.7061	714	11.5
15	-0.3258	0.4	717.7042	718	11.0
16	-0.1972	0.4	721.4737	721	10.6
17	-0.0759	0.4	725.0292	725	10.3
18	0.0389	0.3	728.3942	728	9.9
19	0.1477	0.3	731.5834	732	9.6
20	0.2510	0.3	734.6113	735	9.3
21	0.3491	0.3	737.4867	737	9.0
22	0.4426	0.3	740.2274	740	8.7
23	0.5316	0.3	742.8362	743	8.5
24	0.6167	0.3	745.3306	745	8.2
25	0.6981	0.3	747.7166	748	8.0
26	0.7760	0.3	750.0000	750	7.8
27	0.8508	0.3	752.1925	752	7.6
28	0.9226	0.3	754.2971	754	7.4
29	0.9918	0.2	756.3254	756	7.3
30	1.0585	0.2	758.2805	758	7.1
31	1.1229	0.2	760.1682	760	7.0
32	1.1852	0.2	761.9943	762	6.9
33	1.2456	0.2	763.7648	764	6.7
34	1.3042	0.2	765.4824	765	6.6
35	1.3612	0.2	767.1532	767	6.5
36	1.4168	0.2	768.7830	769	6.4
37	1.4710	0.2	770.3717	770	6.4
38	1.5240	0.2	771.9252	772	6.3
39	1.5760	0.2	773.4494	773	6.2
40	1.6270	0.2	774.9443	775	6.2
41	1.6772	0.2	776.4158	776	6.1
42	1.7266	0.2	777.8638	778	6.1

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	1.7755	0.2	779.2971	779	6.0
44	1.8239	0.2	780.7158	781	6.0
45	1.8718	0.2	782.1198	782	5.9
46	1.9194	0.2	783.5151	784	5.9
47	1.9669	0.2	784.9074	785	5.9
48	2.0142	0.2	786.2938	786	5.9
49	2.0614	0.2	787.6773	788	5.8
50	2.1087	0.2	789.0638	789	5.8
51	2.1561	0.2	790.4532	790	5.8
52	2.2037	0.2	791.8484	792	5.8
53	2.2515	0.2	793.2495	793	5.8
54	2.2997	0.2	794.6623	795	5.8
55	2.3482	0.2	796.0840	796	5.8
56	2.3973	0.2	797.5232	798	5.8
57	2.4468	0.2	798.9741	799	5.9
58	2.4970	0.2	800.4456	800	5.9
59	2.5479	0.2	801.9375	802	5.9
60	2.5995	0.2	803.4500	803	5.9
61	2.6520	0.2	804.9889	805	6.0
62	2.7054	0.2	806.5541	807	6.0
63	2.7600	0.2	808.1546	808	6.0
64	2.8157	0.2	809.7872	810	6.1
65	2.8729	0.2	811.4639	811	6.2
66	2.9315	0.2	813.1815	813	6.2
67	2.9920	0.2	814.9549	815	6.3
68	3.0546	0.2	816.7898	817	6.4
69	3.1195	0.2	818.6922	819	6.6
70	3.1871	0.2	820.6736	821	6.7
71	3.2580	0.2	822.7518	823	6.9
72	3.3327	0.2	824.9414	825	7.2
73	3.4119	0.3	827.2629	827	7.5
74	3.4963	0.3	829.7368	830	7.8
75	3.5869	0.3	832.3925	832	8.2
76	3.6846	0.3	835.2563	835	8.7
77	3.7906	0.3	838.3633	838	9.3
78	3.9061	0.3	841.7488	842	9.9
79	4.0324	0.4	845.4509	845	10.6
80	4.1707	0.4	849.5047	850	10.6
81	4.3224	0.4	853.9513	850	10.6
82	4.4890	0.4	858.8347	850	10.6
83	4.6717	0.5	864.1899	850	10.6
84	4.8722	0.5	870.0670	850	10.6
85	5.0924	0.5	876.5214	850	10.6
86	5.3349	0.6	883.6295	850	10.6

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	5.6038	0.6	891.5115	850	10.6
88	5.9060	0.7	900.3695	850	10.6
89	6.2543	0.7	910.5788	850	10.6
90	6.6715	0.8	922.8077	850	10.6
91	7.1999	1.0	938.2960	850	10.6
92	7.9228	1.2	959.4855	850	10.6
93	9.0366	1.7	992.1330	850	10.6
94	11.2797	3.0	1057.8823	850	10.6
95	15.000	6.6	1166.9310	850	10.6

Table A.12.47 Conversion Table for Performance Level Setting Form: Integrated Mathematics III

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
0	-10.000	11.0	320.8200	650	20.0
1	-3.9857	1.4	558.5581	650	20.0
2	-2.9620	0.9	599.0237	650	20.0
3	-2.3774	0.7	622.1322	650	20.0
4	-1.9684	0.6	638.2995	650	20.0
5	-1.6530	0.6	650.7669	651	20.0
6	-1.3955	0.5	660.9456	661	19.8
7	-1.1771	0.5	669.5786	670	18.2
8	-0.9874	0.4	677.0773	677	17.0
9	-0.8195	0.4	683.7141	684	16.1
10	-0.6689	0.4	689.6672	690	15.2
11	-0.5324	0.4	695.0629	695	14.5
12	-0.4075	0.4	700.0000	700	13.9
13	-0.2924	0.3	704.5498	705	13.4
14	-0.1855	0.3	708.7754	709	12.9
15	-0.0856	0.3	712.7243	713	12.5
16	0.0083	0.3	716.4361	716	12.1
17	0.0969	0.3	719.9383	720	11.8
18	0.1810	0.3	723.2627	723	11.5
19	0.2611	0.3	726.4290	726	11.2
20	0.3377	0.3	729.4569	729	10.9
21	0.4110	0.3	732.3543	732	10.7
22	0.4816	0.3	735.1451	735	10.5
23	0.5495	0.3	737.8291	738	10.2
24	0.6151	0.3	740.4222	740	10.0
25	0.6786	0.2	742.9322	743	9.9
26	0.7400	0.2	745.3593	745	9.7
27	0.7996	0.2	747.7152	748	9.5
28	0.8574	0.2	750.0000	750	9.3
29	0.9136	0.2	752.2215	752	9.2
30	0.9683	0.2	754.3837	754	9.0
31	1.0215	0.2	756.4867	756	8.8
32	1.0733	0.2	758.5343	759	8.7
33	1.1238	0.2	760.5305	761	8.5
34	1.1730	0.2	762.4753	762	8.4
35	1.2210	0.2	764.3727	764	8.3
36	1.2678	0.2	766.2226	766	8.1
37	1.3135	0.2	768.0291	768	8.0
38	1.3581	0.2	769.7921	770	7.9
39	1.4017	0.2	771.5155	772	7.7
40	1.4444	0.2	773.2034	773	7.6
41	1.4862	0.2	774.8557	775	7.5
42	1.5271	0.2	776.4724	776	7.4

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
43	1.5672	0.2	778.0575	778	7.3
44	1.6066	0.2	779.6150	780	7.2
45	1.6453	0.2	781.1447	781	7.2
46	1.6834	0.2	782.6508	783	7.1
47	1.7210	0.2	784.1371	784	7.0
48	1.7580	0.2	785.5996	786	7.0
49	1.7945	0.2	787.0424	787	6.9
50	1.8307	0.2	788.4734	788	6.8
51	1.8665	0.2	789.8885	790	6.8
52	1.9020	0.2	791.2918	791	6.8
53	1.9373	0.2	792.6871	793	6.7
54	1.9723	0.2	794.0707	794	6.7
55	2.0072	0.2	795.4502	795	6.7
56	2.0420	0.2	796.8258	797	6.7
57	2.0768	0.2	798.2014	798	6.7
58	2.1115	0.2	799.5731	800	6.6
59	2.1463	0.2	800.9487	801	6.6
60	2.1812	0.2	802.3282	802	6.7
61	2.2163	0.2	803.7157	804	6.7
62	2.2516	0.2	805.1110	805	6.7
63	2.2872	0.2	806.5183	807	6.7
64	2.3232	0.2	807.9413	808	6.7
65	2.3597	0.2	809.3841	809	6.8
66	2.3967	0.2	810.8467	811	6.8
67	2.4343	0.2	812.3330	812	6.9
68	2.4727	0.2	813.8509	814	7.0
69	2.5120	0.2	815.4043	815	7.0
70	2.5522	0.2	816.9934	817	7.1
71	2.5935	0.2	818.6259	819	7.2
72	2.6360	0.2	820.3059	820	7.3
73	2.6798	0.2	822.0373	822	7.5
74	2.7252	0.2	823.8319	824	7.6
75	2.7723	0.2	825.6937	826	7.7
76	2.8213	0.2	827.6306	828	7.9
77	2.8723	0.2	829.6466	830	8.1
78	2.9256	0.2	831.7535	832	8.3
79	2.9815	0.2	833.9631	834	8.5
80	3.0402	0.2	836.2835	836	8.7
81	3.1020	0.2	838.7263	839	9.0
82	3.1674	0.2	841.3115	841	9.2
83	3.2367	0.2	844.0509	844	9.6
84	3.3106	0.3	846.9720	847	9.9
85	3.3896	0.3	850.0948	850	9.9
86	3.4747	0.3	853.4587	850	9.9

Raw Score	IRT Theta	Theta CSEM	Unrounded Scale Score	Rounded Scale Score	Scale Score CSEM
87	3.5668	0.3	857.0993	850	9.9
88	3.6674	0.3	861.0759	850	9.9
89	3.7783	0.3	865.4597	850	9.9
90	3.9019	0.3	870.3454	850	9.9
91	4.0417	0.4	875.8716	850	9.9
92	4.2023	0.4	882.2199	850	9.9
93	4.3905	0.4	889.6592	850	9.9
94	4.6167	0.5	898.6006	850	9.9
95	4.8970	0.6	909.6805	850	9.9
96	5.2601	0.7	924.0334	850	9.9
97	5.7624	0.9	943.8888	850	9.9
98	6.5405	1.2	974.6461	850	9.9
99	8.0537	1.9	1034.4611	850	9.9
100	15.000	9.7	1309.0400	850	9.9

Appendix 12.4: IRT Test Characteristic Curves, Information Curves and CSEM Curves

Figure A.12.1 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 3

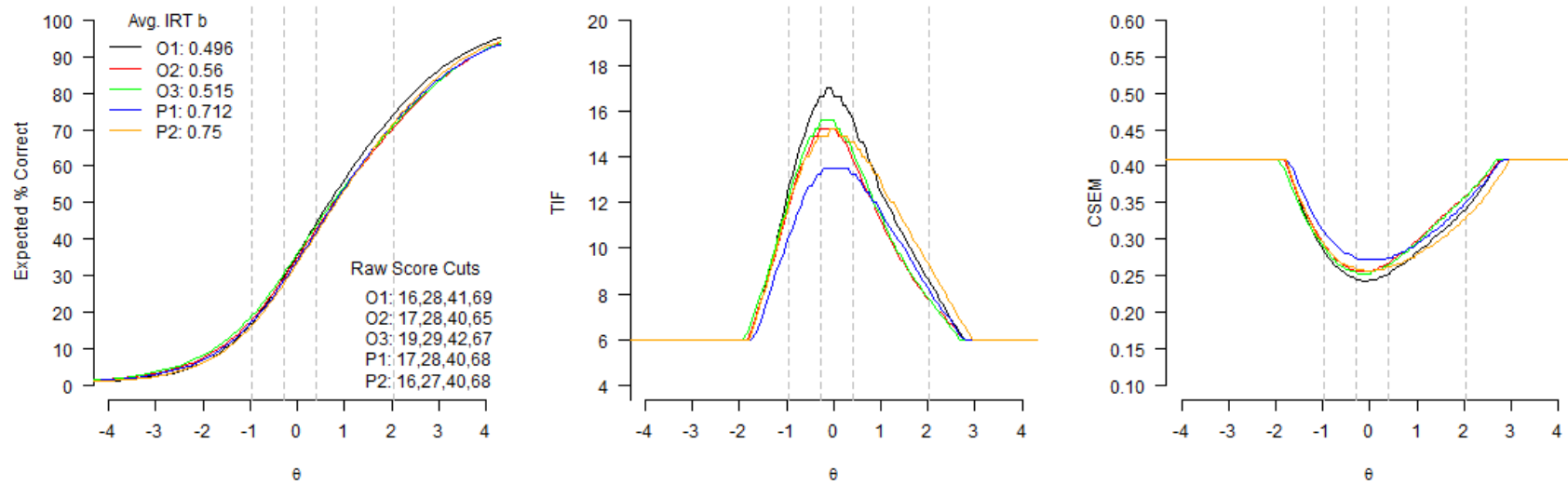


Figure A.12.2 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 4

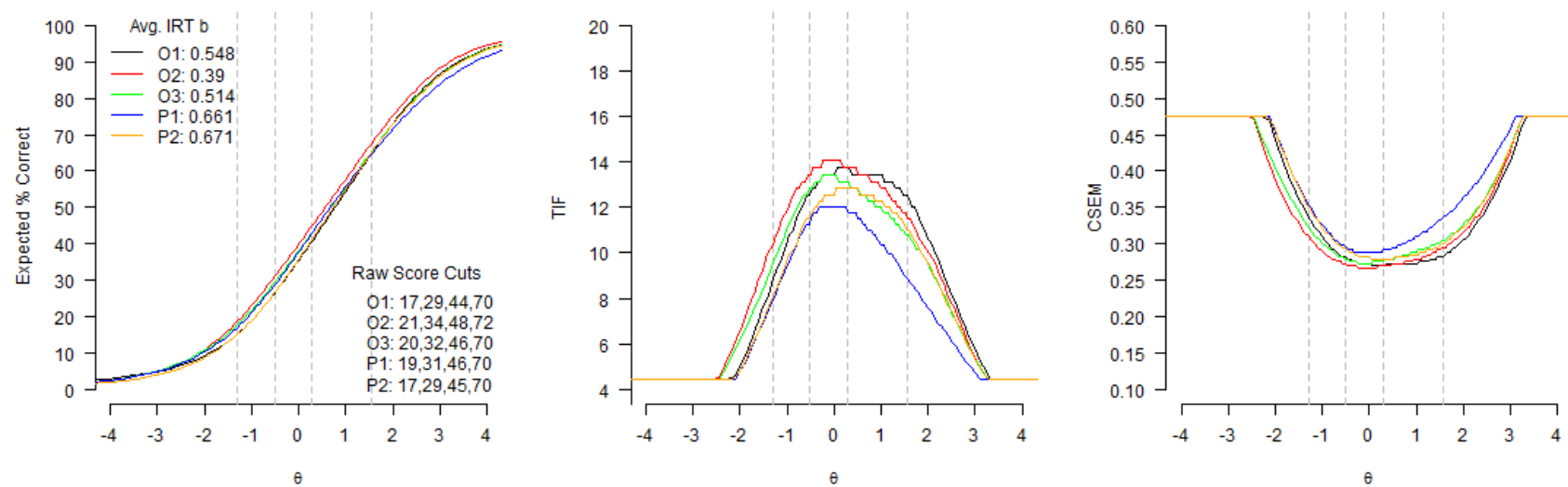


Figure A.12.3 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 5

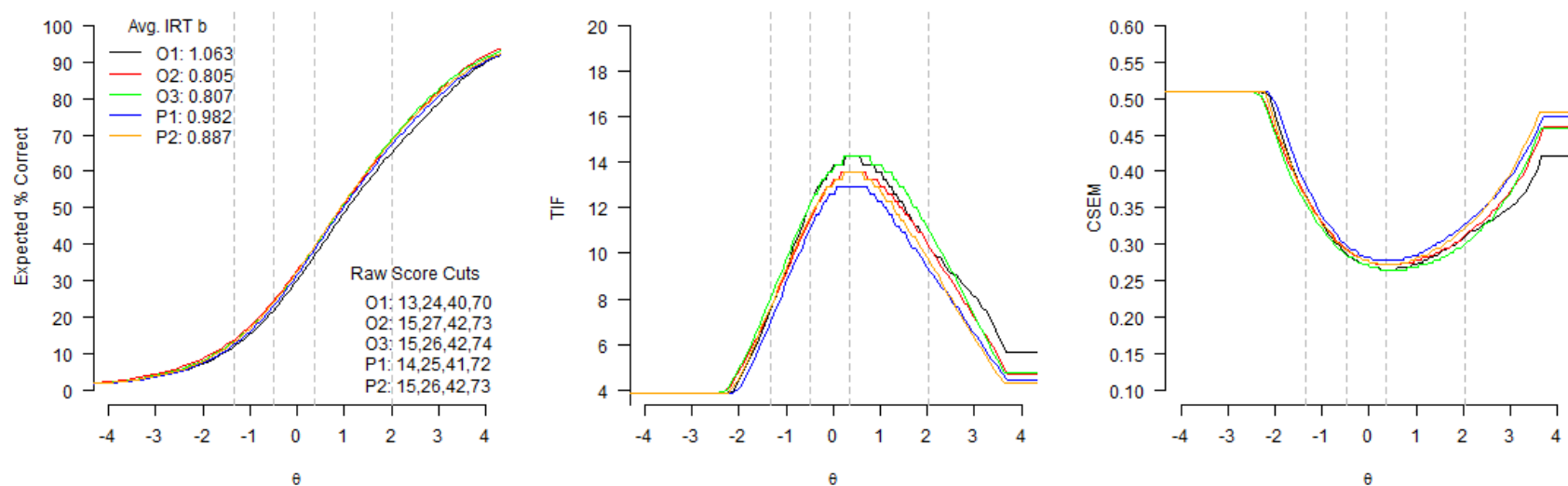


Figure A.12.4 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 6

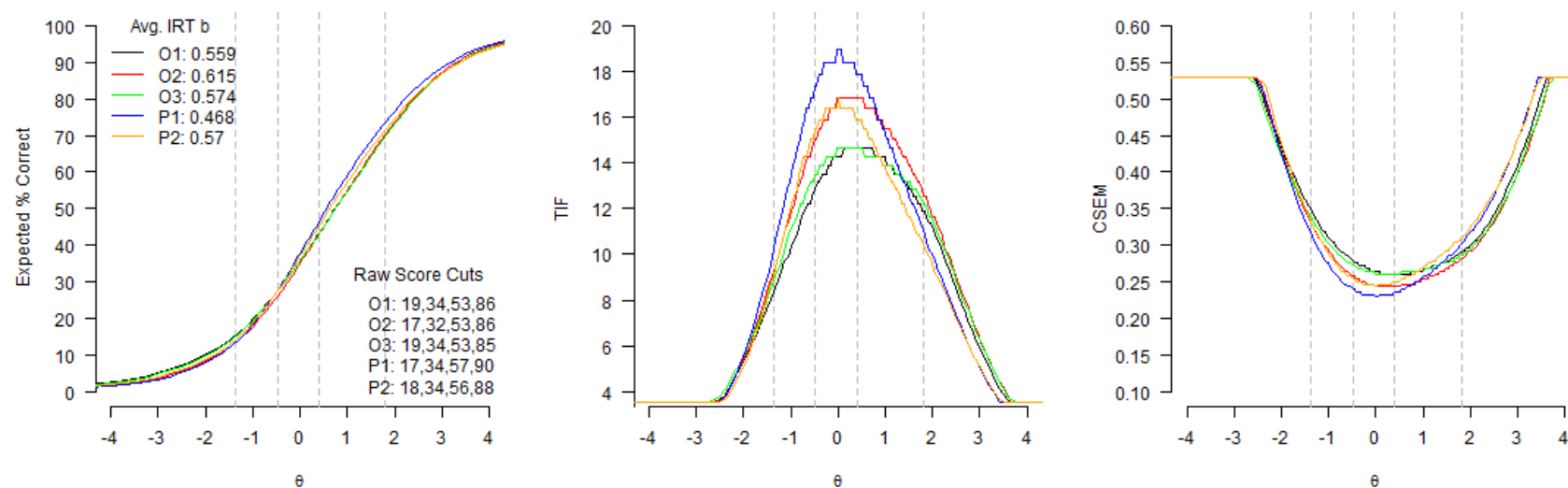


Figure A.12.5 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 7

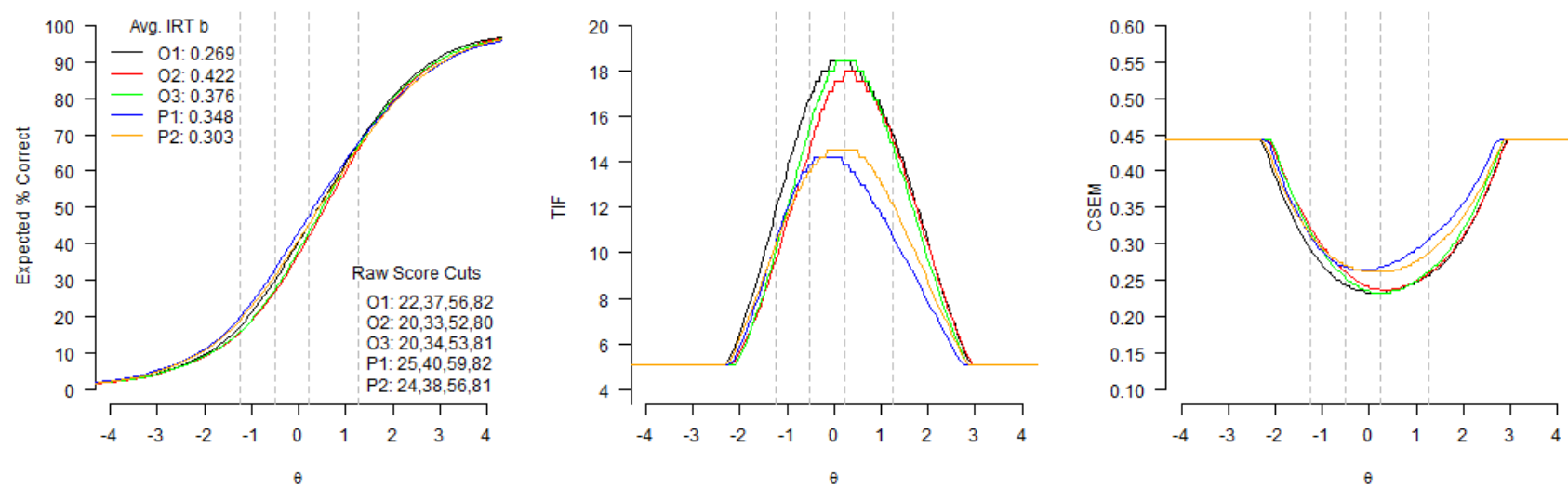


Figure A.12.6 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 8

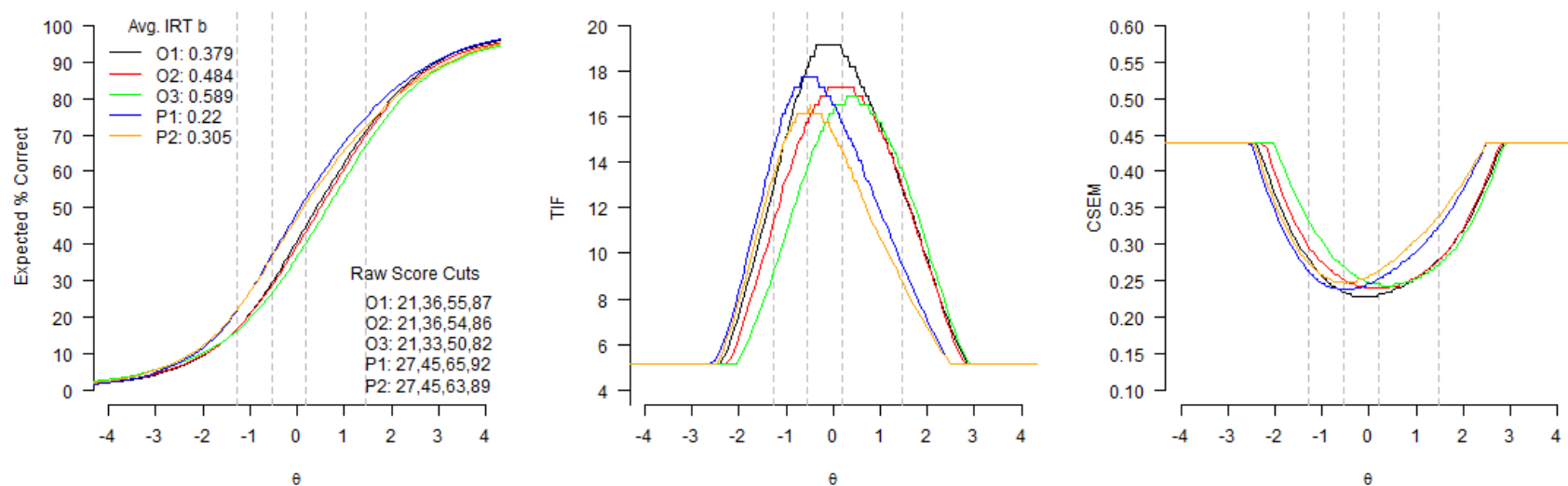


Figure A.12.7 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 9

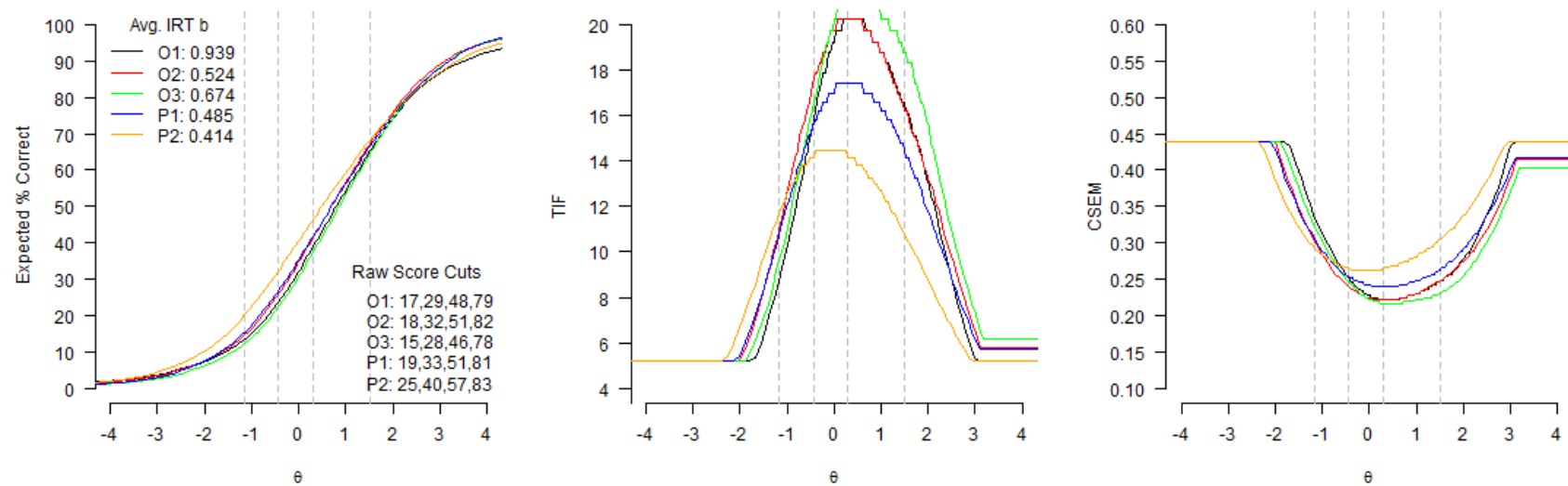


Figure A.12.8 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 10

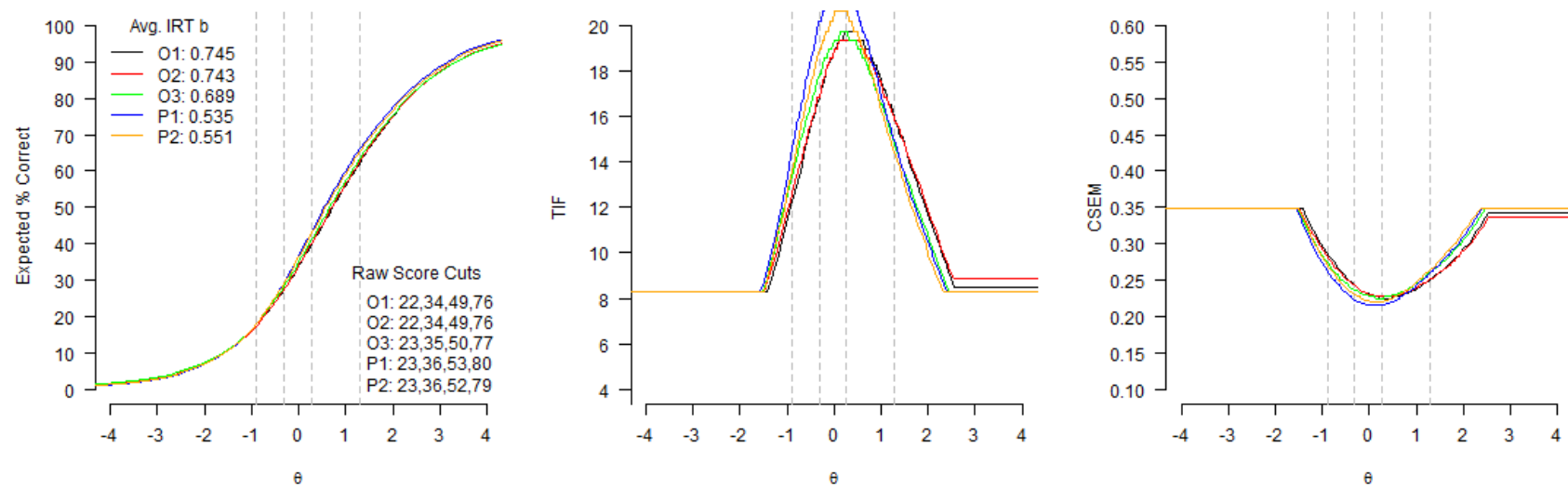


Figure A.12.9 IRT Test Characteristic Curves, Information Curves, and CSEM Curves ELA/L Grade 11

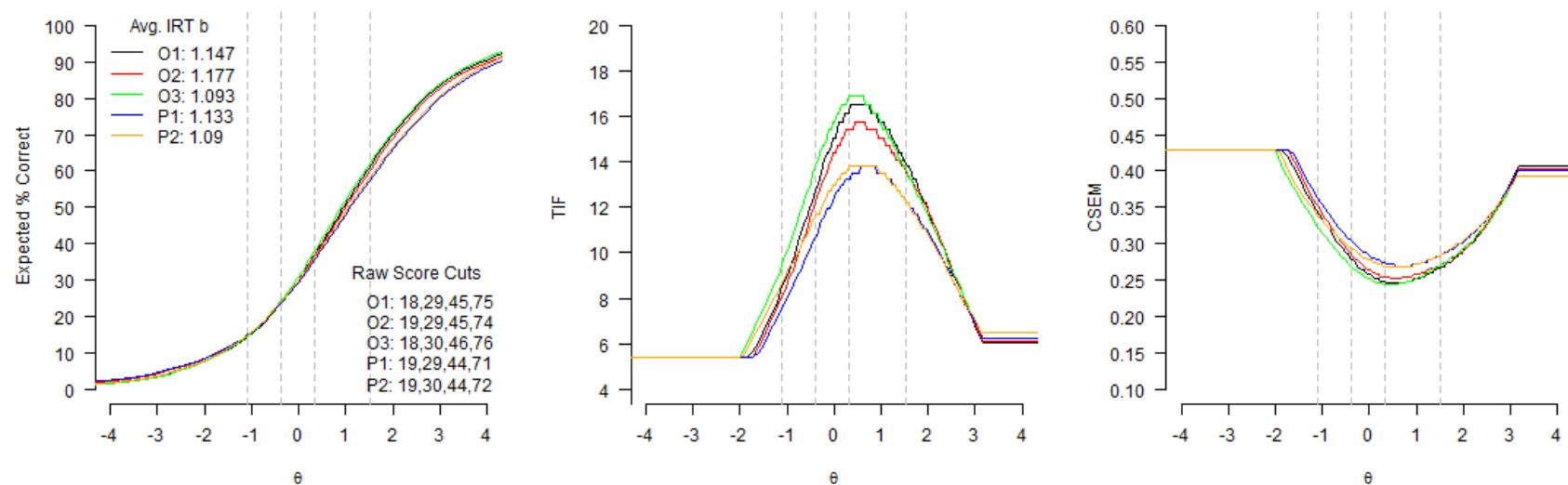


Figure A.12.10 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 3

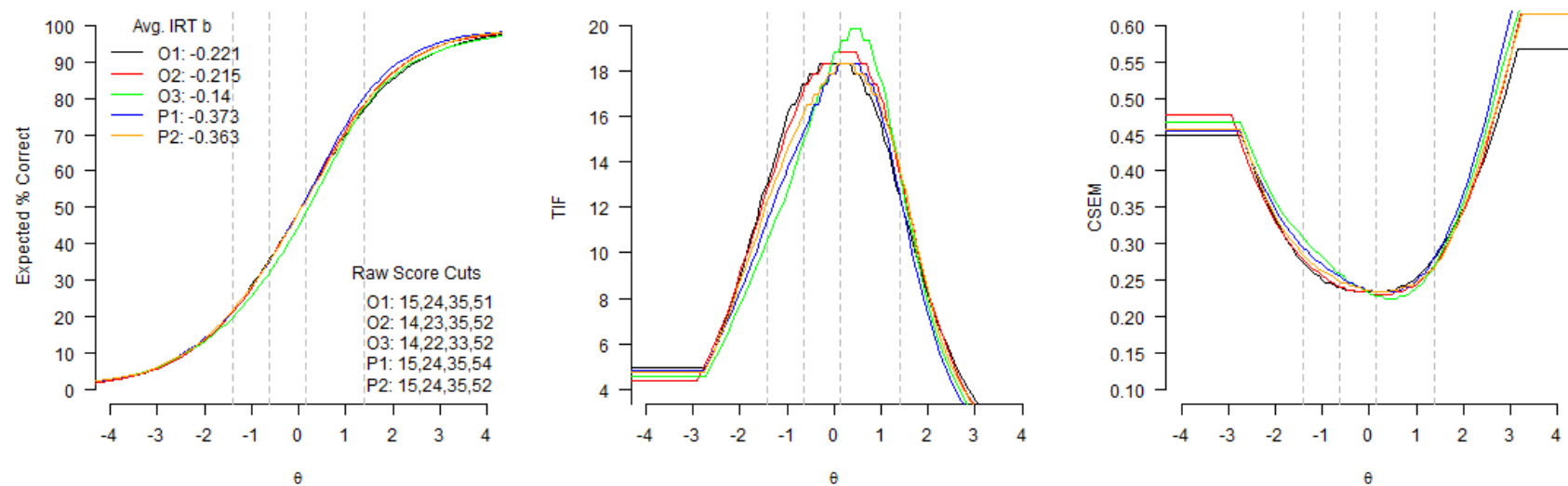


Figure A.12.11 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 4

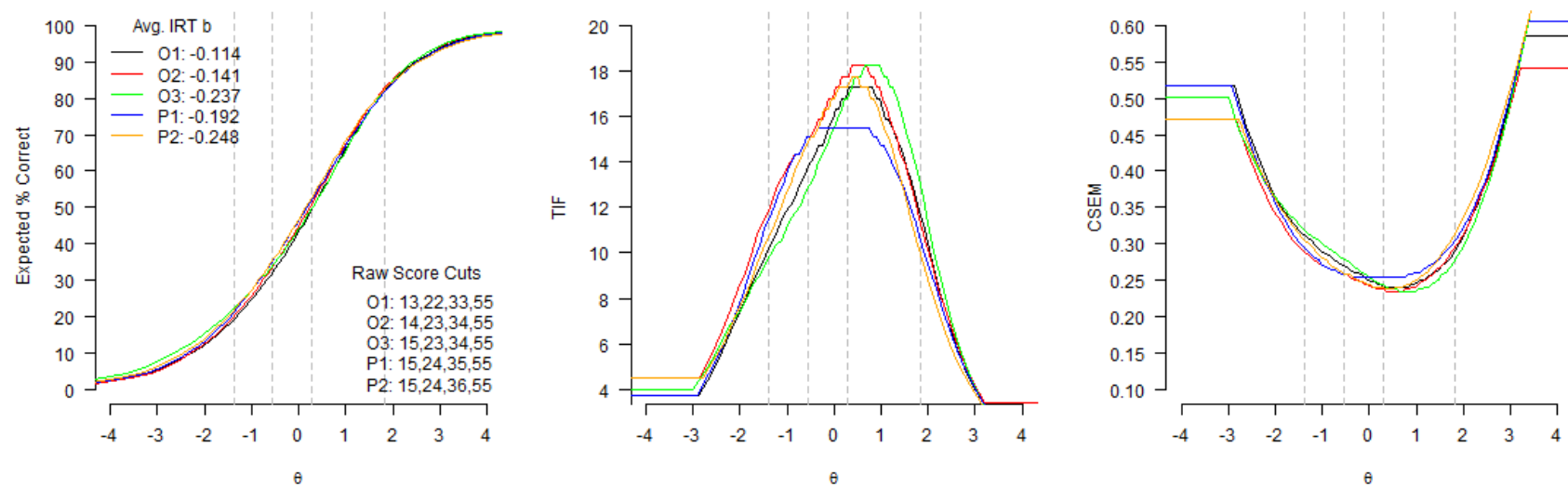


Figure A.12.12 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 5

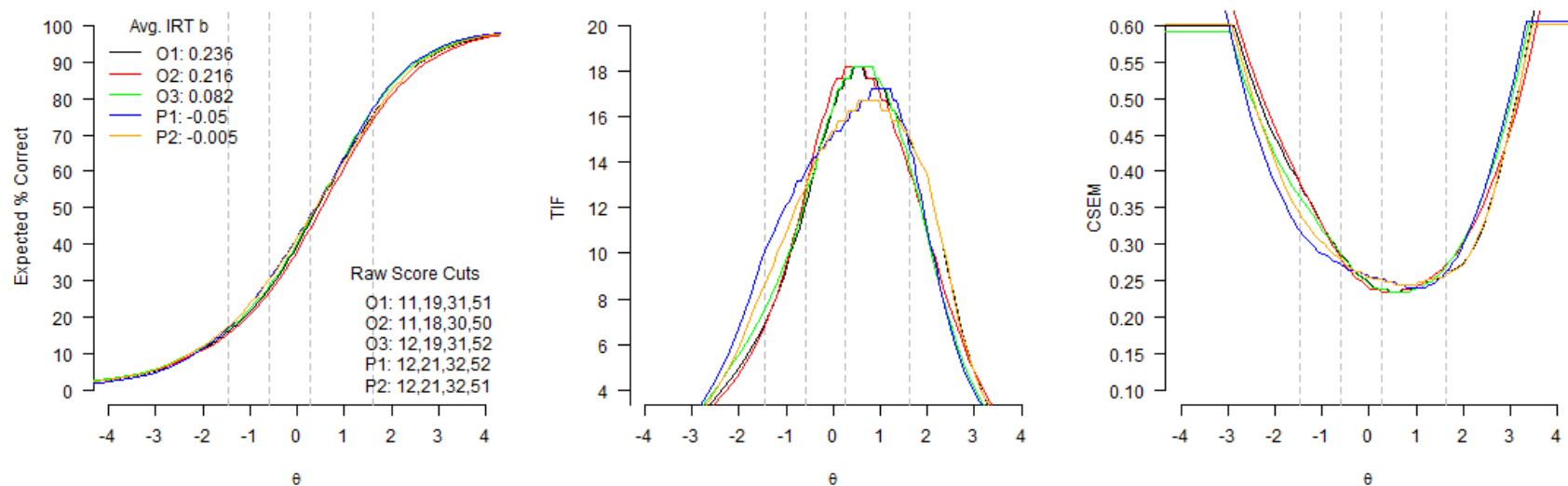


Figure A.12.13 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 6

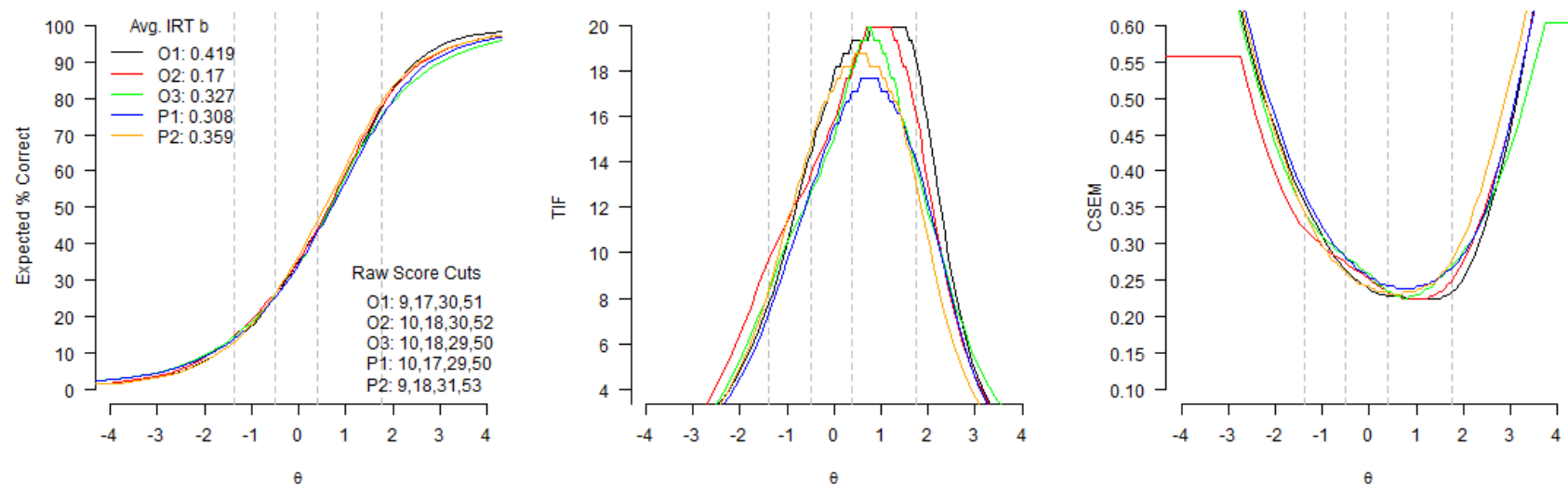


Figure A.12.14 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 7

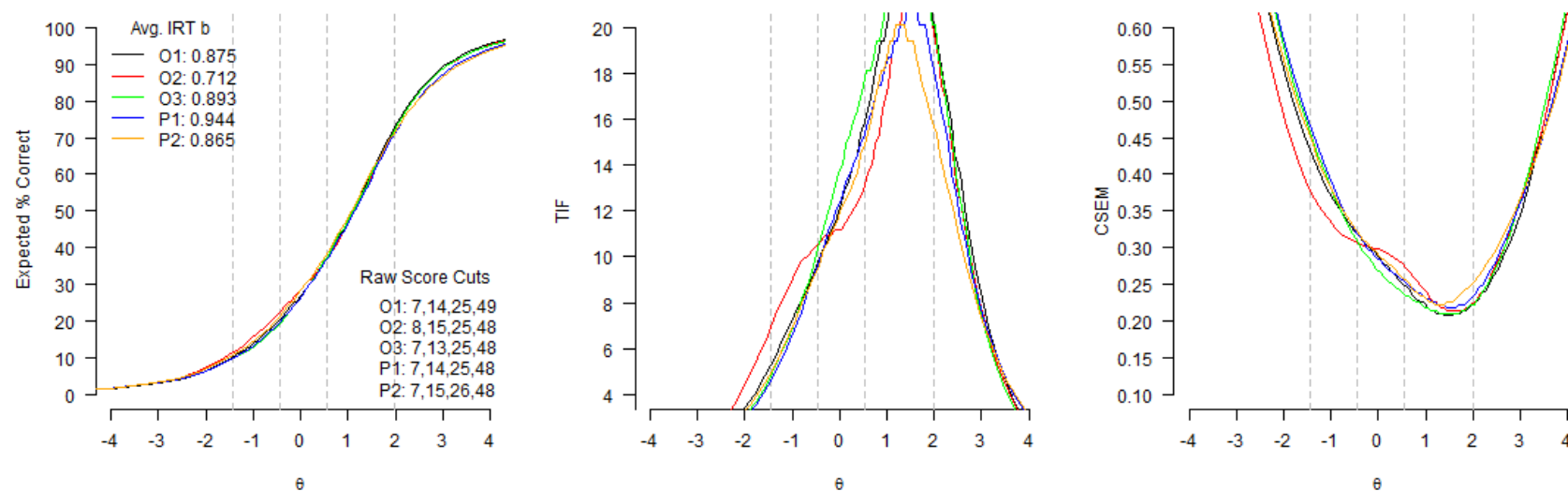


Figure A.12.15 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Mathematics Grade 8

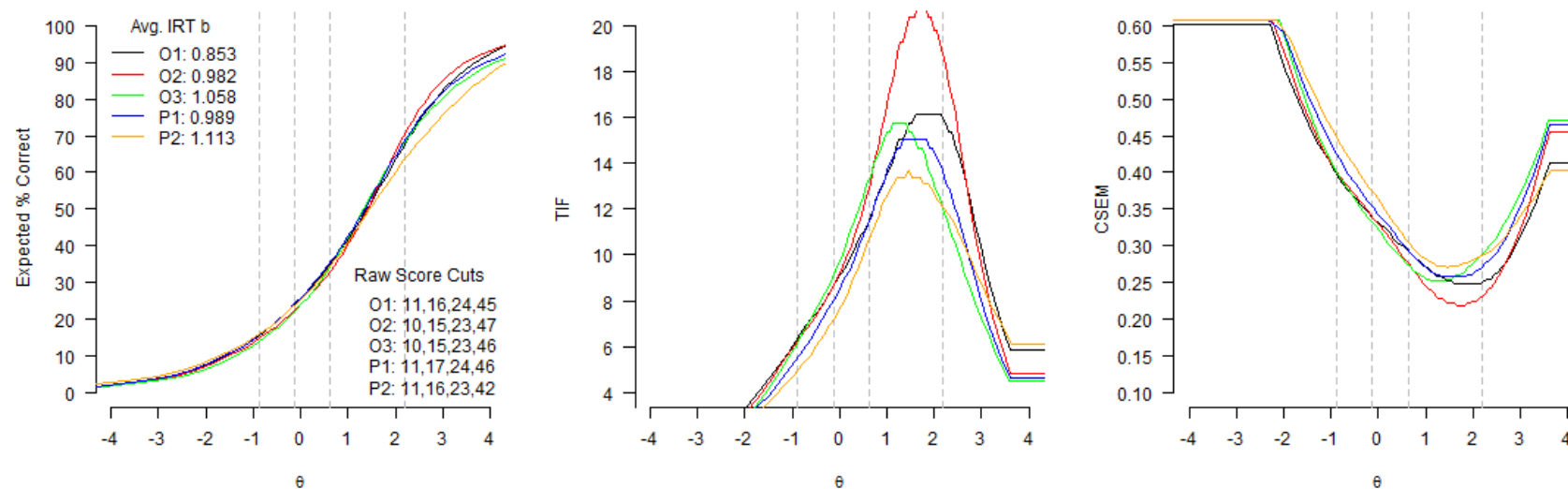


Figure A.12.16 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Algebra I

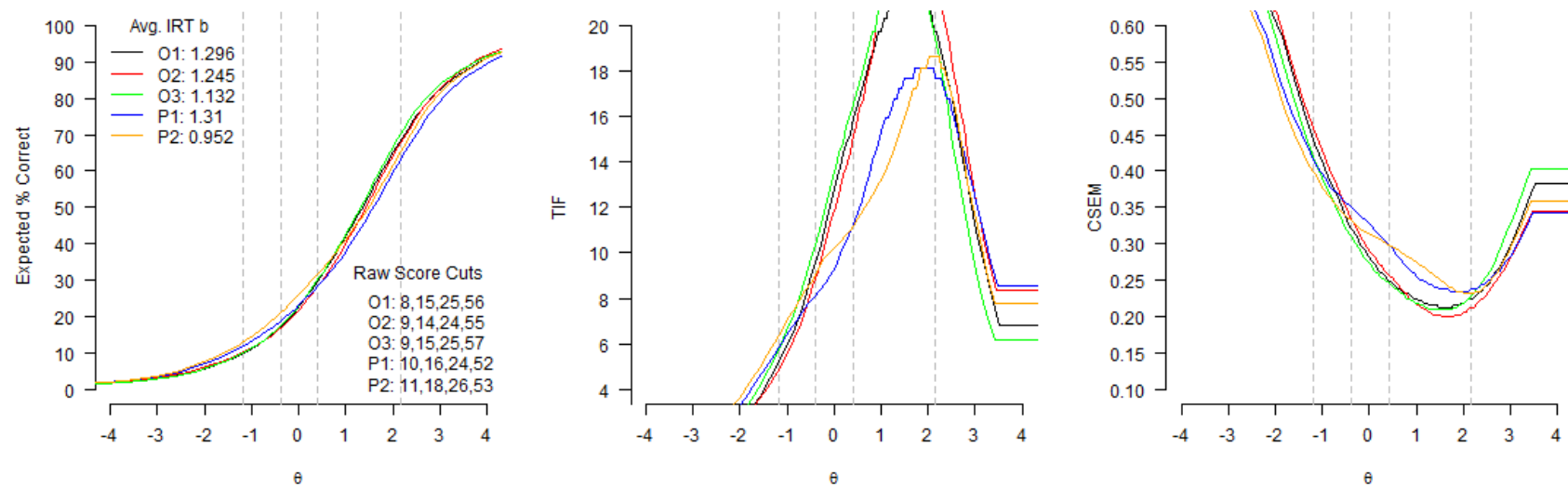


Figure A.12.17 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Geometry

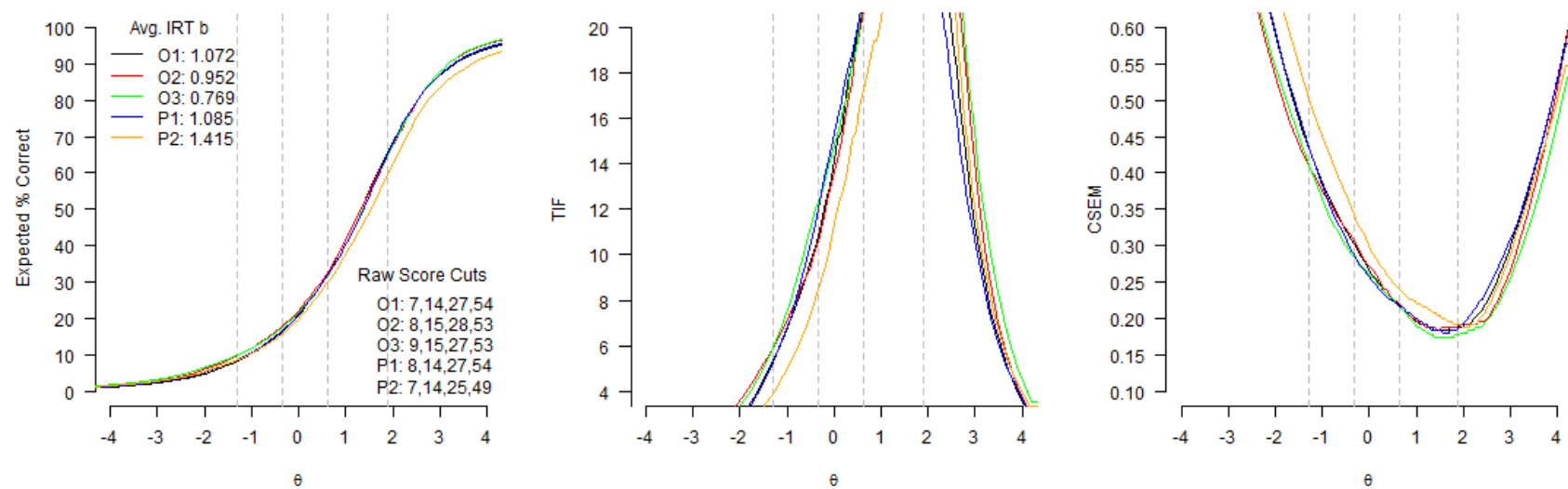


Figure A.12.18 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Algebra II

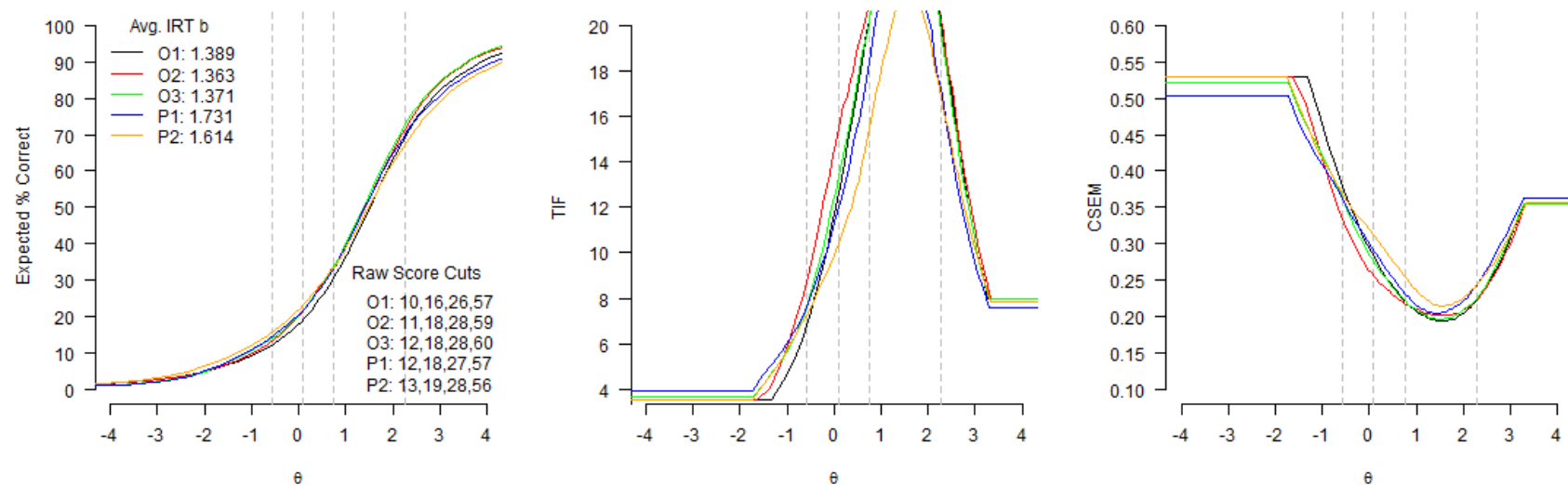


Figure A.12.19 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Integrated Mathematics I

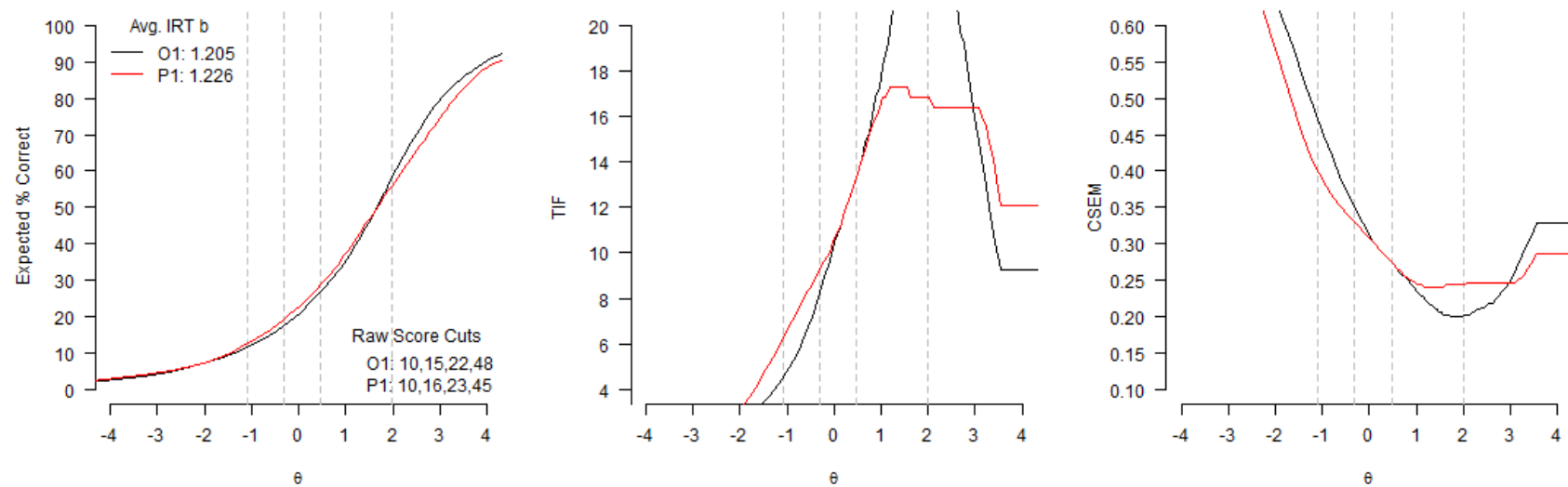


Figure A.12.20 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Integrated Mathematics II

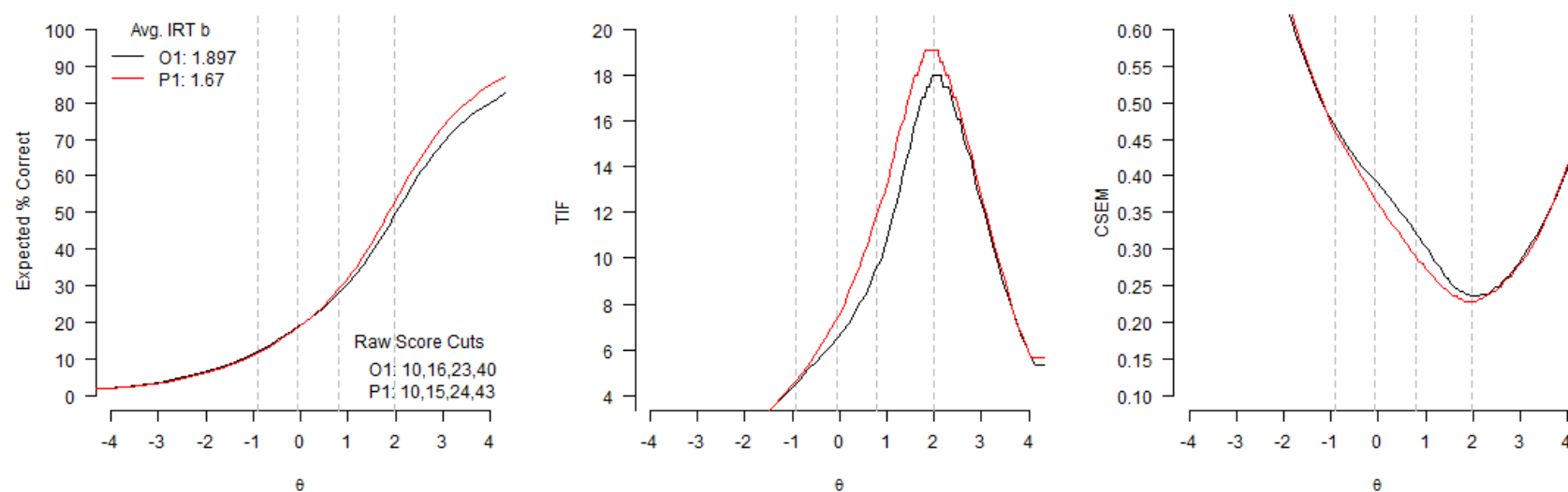
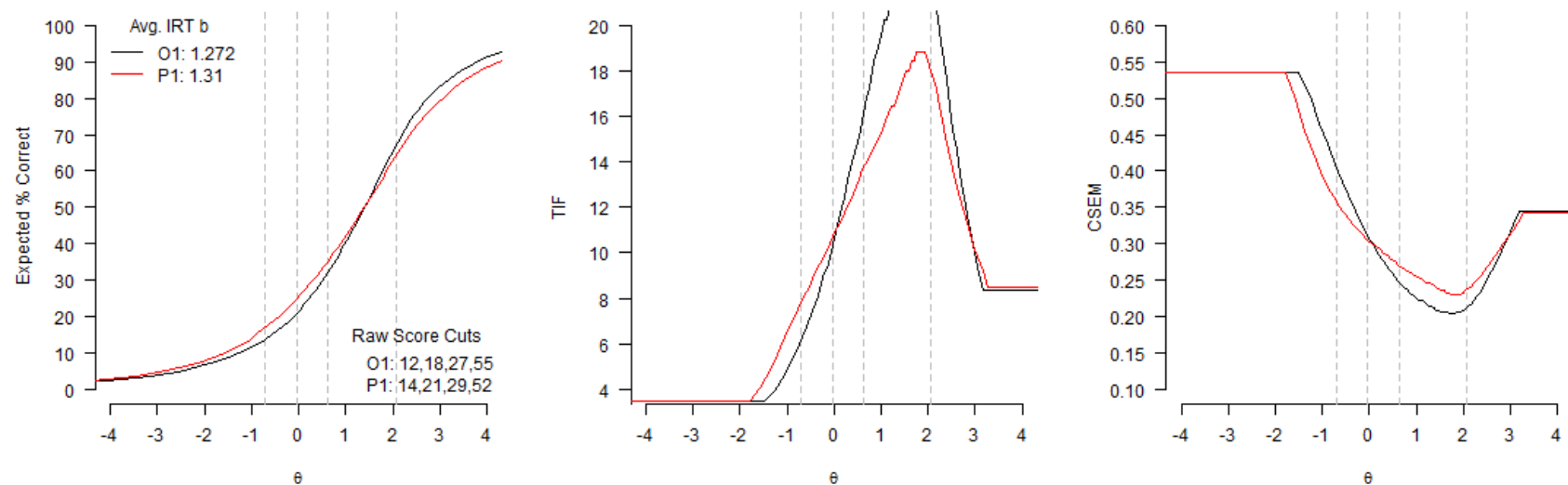


Figure A.12.21 IRT Test Characteristic Curves, Information Curves, and CSEM Curves Integrated Mathematics III



Appendix 12.5: Subgroup Scale Score Performance

Table A.12.48 Subgroup Performance for ELA/L Scale Scores: Grade 3

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		471,801	738.48	40.58	650	850
Gender	Female	231,217	743.26	40.89	650	850
	Male	240,584	733.90	39.74	650	850
Ethnicity	American Indian/Alaska Native	5,431	716.71	34.76	650	850
	Asian	27,059	765.40	39.86	650	850
	Black or African American	78,661	722.26	37.94	650	850
	Hispanic/Latino	131,457	724.68	37.37	650	850
	Native Hawaiian or Pacific Islander	803	749.12	39.26	650	850
	Multiple Race Selected	15,836	744.02	40.43	650	850
	White	212,059	749.76	38.05	650	850
Economic Status*	Economically Disadvantaged	228,865	722.74	36.75	650	850
	Not Economically Disadvantaged	237,879	753.32	38.38	650	850
English Learner Status	English Learner (EL)	68,638	713.99	34.27	650	850
	Non English Learner	396,667	742.77	40.05	650	850
Disabilities	Students with Disabilities (SWD)	63,952	709.97	38.13	650	850
	Students without Disabilities	245,463	745.33	39.39	650	850
Reading Score		471,801	45.19	16.10	10	90
Gender	Female	231,217	46.47	16.18	10	90
	Male	240,584	43.96	15.93	10	90
Ethnicity	American Indian/Alaska Native	5,431	36.37	13.55	10	90
	Asian	27,059	55.21	16.06	10	90
	Black or African American	78,661	38.65	14.58	10	90
	Hispanic/Latino	131,457	39.37	14.46	10	90
	Native Hawaiian or Pacific Islander	803	48.52	15.20	10	90
	Multiple Race Selected	15,836	47.78	16.18	10	90
	White	212,059	49.97	15.31	10	90
Economic Status*	Economically Disadvantaged	228,865	38.82	14.24	10	90
	Not Economically Disadvantaged	237,879	51.18	15.44	10	90
English Learner Status	English Learner (EL)	68,638	35.01	12.95	10	90
	Non English Learner	396,667	46.96	15.93	10	90
Disabilities	Students with Disabilities (SWD)	63,952	34.68	15.11	10	90
	Students without Disabilities	245,463	47.64	15.74	10	90
Writing Score		471,801	31.18	11.90	10	60
Gender	Female	231,217	33.13	11.50	10	60
	Male	240,584	29.32	11.97	10	60

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Ethnicity	American Indian/Alaska Native	5,431	26.48	11.27	10	60
	Asian	27,059	38.17	10.46	10	60
	Black or African American	78,661	27.46	11.95	10	60
	Hispanic/Latino	131,457	28.43	11.73	10	60
	Native Hawaiian or Pacific Islander	803	34.50	11.54	10	60
	Multiple Race Selected	15,836	31.96	11.85	10	60
	White	212,059	33.44	11.24	10	60
Economic Status*	Economically Disadvantaged	228,865	27.62	11.72	10	60
	Not Economically Disadvantaged	237,879	34.55	11.05	10	60
English Learner Status	English Learner (EL)	68,638	26.17	11.51	10	60
	Non English Learner	396,667	32.07	11.74	10	60
Disabilities	Students with Disabilities (SWD)	63,952	23.14	11.94	10	60
	Students without Disabilities	245,463	33.17	11.31	10	60

Note: This table is identical to Table 12.7 in Section 12.

*Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.49 Subgroup Performance for ELA/L Scale Scores: Grade 4

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		461,204	742.44	35.20	650	850
Gender	Female	226,107	747.65	34.80	650	850
	Male	235,097	737.42	34.85	650	850
Ethnicity	American Indian/Alaska Native	5,339	721.85	30.93	650	850
	Asian	26,805	765.70	33.40	650	850
	Black or African American	74,444	727.00	33.01	650	850
	Hispanic/Latino	127,151	730.55	32.35	650	850
	Native Hawaiian or Pacific Islander	756	747.12	34.79	650	850
	Multiple Race Selected	14,693	747.04	35.44	650	850
	White	211,636	752.27	32.89	650	850
Economic Status*	Economically Disadvantaged	218,926	728.08	32.04	650	850
	Not Economically Disadvantaged	237,120	755.47	32.84	650	850
English Learner Status	English Learner (EL)	42,083	712.55	29.08	650	849
	Non English Learner	412,712	745.47	34.28	650	850
Disabilities	Students with Disabilities (SWD)	70,114	715.28	34.14	650	850
	Students without Disabilities	235,751	749.77	33.35	650	850
Reading Score		461,204	46.82	14.03	10	90
Gender	Female	226,107	48.09	13.88	10	90
	Male	235,097	45.59	14.06	10	90
Ethnicity	American Indian/Alaska Native	5,339	38.30	12.20	10	88
	Asian	26,805	55.40	13.57	10	90
	Black or African American	74,444	40.71	12.77	10	90
	Hispanic/Latino	127,151	41.70	12.64	10	90
	Native Hawaiian or Pacific Islander	756	48.15	13.84	10	90
	Multiple Race Selected	14,693	49.03	14.14	10	90
	White	211,636	51.02	13.24	10	90
Economic Status*	Economically Disadvantaged	218,926	40.97	12.48	10	90
	Not Economically Disadvantaged	237,120	52.11	13.24	10	90
English Learner Status	English Learner (EL)	42,083	34.49	10.83	10	90
	Non English Learner	412,712	48.08	13.69	10	90
Disabilities	Students with Disabilities (SWD)	70,114	37.04	13.39	10	90
	Students without Disabilities	235,751	49.41	13.52	10	90
Writing Score		461,204	32.95	9.38	10	60
Gender	Female	226,107	34.84	8.84	10	60
	Male	235,097	31.13	9.52	10	60
Ethnicity	American Indian/Alaska Native	5,339	28.81	8.92	10	60

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
	Asian	26,805	38.58	8.21	10	60
	Black or African American	74,444	29.43	9.57	10	60
	Hispanic/Latino	127,151	30.68	9.20	10	60
	Native Hawaiian or Pacific Islander	756	34.35	9.12	10	60
	Multiple Race Selected	14,693	33.63	9.42	10	60
	White	211,636	34.90	8.69	10	60
Economic Status*	Economically Disadvantaged	218,926	29.86	9.31	10	60
	Not Economically Disadvantaged	237,120	35.77	8.53	10	60
English Learner Status	English Learner (EL)	42,083	26.64	9.42	10	60
	Non English Learner	412,712	33.59	9.12	10	60
Disabilities	Students with Disabilities (SWD)	70,114	25.54	10.27	10	60
	Students without Disabilities	235,751	34.92	8.52	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.50 Subgroup Performance for ELA/L Scale Scores: Grade 5

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		455,980	742.24	32.87	650	850
Gender	Female	223,111	747.29	32.82	650	850
	Male	232,869	737.40	32.19	650	850
Ethnicity	American Indian/Alaska Native	5,107	724.57	28.17	650	832
	Asian	26,923	765.34	31.47	650	850
	Black or African American	72,454	727.16	30.42	650	847
	Hispanic/Latino	124,303	731.00	30.21	650	850
	Native Hawaiian or Pacific Islander	753	749.64	32.85	654	837
	Multiple Race Selected	13,956	746.59	32.80	650	850
	White	212,115	751.18	30.70	650	850
Economic Status*	Economically Disadvantaged	211,469	728.44	29.76	650	849
	Not Economically Disadvantaged	239,390	754.21	30.71	650	850
English Learner Status	English Learner (EL)	32,036	710.31	26.08	650	827
	Non English Learner	417,455	744.72	32.01	650	850
Disabilities	Students with Disabilities (SWD)	72,830	717.23	30.93	650	850
	Students without Disabilities	229,803	750.00	31.31	650	850
Reading Score		455,980	46.67	13.15	10	90
Gender	Female	223,111	47.87	13.08	10	90
	Male	232,869	45.53	13.11	10	90
Ethnicity	American Indian/Alaska Native	5,107	39.48	11.20	10	87
	Asian	26,923	55.20	12.64	10	90
	Black or African American	72,454	40.81	12.03	10	90
	Hispanic/Latino	124,303	42.02	11.95	10	90
	Native Hawaiian or Pacific Islander	753	48.97	12.56	10	81
	Multiple Race Selected	13,956	48.63	13.16	10	90
	White	212,115	50.36	12.42	10	90
Economic Status*	Economically Disadvantaged	211,469	41.20	11.81	10	90
	Not Economically Disadvantaged	239,390	51.41	12.40	10	90
English Learner Status	English Learner (EL)	32,036	33.98	9.91	10	81
	Non English Learner	417,455	47.65	12.85	10	90
Disabilities	Students with Disabilities (SWD)	72,830	37.38	12.39	10	90
	Students without Disabilities	229,803	49.30	12.63	10	90
Writing Score		455,980	31.76	10.57	10	60
Gender	Female	223,111	33.95	9.80	10	60
	Male	232,869	29.65	10.85	10	60
Ethnicity	American Indian/Alaska Native	5,107	27.62	10.21	10	60

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
	Asian	26,923	38.11	8.63	10	60
	Black or African American	72,454	27.48	11.01	10	60
	Hispanic/Latino	124,303	29.06	10.60	10	60
	Native Hawaiian or Pacific Islander	753	34.11	10.25	10	57
	Multiple Race Selected	13,956	32.70	10.35	10	60
	White	212,115	34.03	9.59	10	60
Economic Status*	Economically Disadvantaged	211,469	28.08	10.74	10	60
	Not Economically Disadvantaged	239,390	34.95	9.33	10	60
English Learner Status	English Learner (EL)	32,036	22.89	10.83	10	57
	Non English Learner	417,455	32.45	10.23	10	60
Disabilities	Students with Disabilities (SWD)	72,830	23.67	11.52	10	60
	Students without Disabilities	229,803	34.35	9.40	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.51 Subgroup Performance for ELA/L Scale Scores: Grade 6

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		455,888	741.53	32.22	650	850
Gender	Female	222,704	748.14	31.43	650	850
	Male	233,184	735.23	31.70	650	850
Ethnicity	American Indian/Alaska Native	5,004	726.05	27.59	650	830
	Asian	27,074	764.46	31.51	650	850
	Black or African American	72,633	727.13	29.94	650	850
	Hispanic/Latino	121,345	730.96	29.70	650	850
	Native Hawaiian or Pacific Islander	646	748.69	31.95	664	831
	Multiple Race Selected	13,247	744.07	32.01	650	850
	White	215,559	749.66	30.26	650	850
Economic Status*	Economically Disadvantaged	205,587	728.30	29.29	650	850
	Not Economically Disadvantaged	245,057	752.48	30.42	650	850
English Learner Status	English Learner (EL)	26,361	708.64	25.49	650	821
	Non English Learner	423,384	743.64	31.45	650	850
Disabilities	Students with Disabilities (SWD)	73,617	715.84	30.05	650	850
	Students without Disabilities	227,345	749.22	30.62	650	850
Reading Score		455,888	46.40	12.62	10	90
Gender	Female	222,704	48.01	12.41	10	90
	Male	233,184	44.87	12.62	10	90
Ethnicity	American Indian/Alaska Native	5,004	39.51	10.56	10	80
	Asian	27,074	54.65	12.47	10	90
	Black or African American	72,633	40.85	11.42	10	90
	Hispanic/Latino	121,345	41.91	11.36	10	90
	Native Hawaiian or Pacific Islander	646	48.54	12.32	17	79
	Multiple Race Selected	13,247	47.86	12.70	10	90
	White	215,559	49.83	12.00	10	90
Economic Status*	Economically Disadvantaged	205,587	41.12	11.19	10	90
	Not Economically Disadvantaged	245,057	50.75	12.06	10	90
English Learner Status	English Learner (EL)	26,361	33.21	9.13	10	78
	Non English Learner	423,384	47.24	12.33	10	90
Disabilities	Students with Disabilities (SWD)	73,617	37.10	11.72	10	90
	Students without Disabilities	227,345	48.98	12.12	10	90
Writing Score		455,888	31.77	10.61	10	60
Gender	Female	222,704	34.68	9.36	10	60
	Male	233,184	28.99	10.99	10	60
Ethnicity	American Indian/Alaska Native	5,004	28.94	10.15	10	60

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
	Asian	27,074	38.28	9.08	10	60
	Black or African American	72,633	27.76	10.94	10	60
	Hispanic/Latino	121,345	29.42	10.55	10	60
	Native Hawaiian or Pacific Islander	646	34.28	10.13	10	60
	Multiple Race Selected	13,247	31.95	10.58	10	60
	White	215,559	33.68	9.87	10	60
Economic Status*	Economically Disadvantaged	205,587	28.35	10.72	10	60
	Not Economically Disadvantaged	245,057	34.61	9.66	10	60
English Learner Status	English Learner (EL)	26,361	23.17	10.83	10	60
	Non English Learner	423,384	32.33	10.35	10	60
Disabilities	Students with Disabilities (SWD)	73,617	23.48	11.37	10	60
	Students without Disabilities	227,345	34.30	9.59	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.52 Subgroup Performance for ELA/L Scale Scores: Grade 7

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		449,801	742.56	37.48	650	850
Gender	Female	219,732	750.85	36.51	650	850
	Male	230,069	734.64	36.68	650	850
Ethnicity	American Indian/Alaska Native	4,767	721.31	32.66	650	850
	Asian	26,815	770.44	36.31	650	850
	Black or African American	71,886	726.00	34.45	650	850
	Hispanic/Latino	118,348	729.82	34.40	650	850
	Native Hawaiian or Pacific Islander	691	750.69	37.34	650	850
	Multiple Race Selected	12,408	745.62	37.38	650	850
	White	214,505	751.96	35.22	650	850
Economic Status*	Economically Disadvantaged	197,876	726.92	33.87	650	850
	Not Economically Disadvantaged	246,553	754.91	35.58	650	850
English Learner Status	English Learner (EL)	25,543	703.58	29.04	650	850
	Non English Learner	418,472	745.00	36.57	650	850
Disabilities	Students with Disabilities (SWD)	72,421	712.98	34.54	650	850
	Students without Disabilities	225,948	751.39	36.04	650	850
Reading Score		449,801	46.98	14.85	10	90
Gender	Female	219,732	49.26	14.55	10	90
	Male	230,069	44.80	14.81	10	90
Ethnicity	American Indian/Alaska Native	4,767	38.27	12.89	10	90
	Asian	26,815	57.08	14.65	10	90
	Black or African American	71,886	40.50	13.37	10	90
	Hispanic/Latino	118,348	41.60	13.33	10	90
	Native Hawaiian or Pacific Islander	691	49.26	14.79	10	90
	Multiple Race Selected	12,408	48.83	15.00	10	90
	White	214,505	50.94	14.13	10	90
Economic Status*	Economically Disadvantaged	197,876	40.73	13.15	10	90
	Not Economically Disadvantaged	246,553	51.90	14.27	10	90
English Learner Status	English Learner (EL)	25,543	31.13	10.54	10	90
	Non English Learner	418,472	47.96	14.51	10	90
Disabilities	Students with Disabilities (SWD)	72,421	35.83	13.69	10	90
	Students without Disabilities	225,948	49.94	14.34	10	90
Writing Score		449,801	31.79	11.74	10	60
Gender	Female	219,732	35.00	10.65	10	60
	Male	230,069	28.72	11.91	10	60
Ethnicity	American Indian/Alaska Native	4,767	26.98	11.07	10	60

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
	Asian	26,815	39.63	9.96	10	60
	Black or African American	71,886	27.38	11.71	10	60
	Hispanic/Latino	118,348	28.83	11.61	10	60
	Native Hawaiian or Pacific Islander	691	34.80	11.10	10	60
	Multiple Race Selected	12,408	32.02	11.65	10	60
	White	214,505	34.02	10.97	10	60
Economic Status*	Economically Disadvantaged	197,876	27.74	11.64	10	60
	Not Economically Disadvantaged	246,553	34.99	10.80	10	60
English Learner Status	English Learner (EL)	25,543	22.03	11.24	10	60
	Non English Learner	418,472	32.41	11.49	10	60
Disabilities	Students with Disabilities (SWD)	72,421	23.07	11.93	10	60
	Students without Disabilities	225,948	34.61	10.92	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.53 Subgroup Performance for ELA/L Scale Scores: Grade 8

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		440,160	743.04	37.68	650	850
Gender	Female	214,660	750.93	36.48	650	850
	Male	225,500	735.53	37.27	650	850
Ethnicity	American Indian/Alaska Native	4,784	723.95	31.93	650	850
	Asian	25,536	770.71	36.93	650	850
	Black or African American	71,383	726.25	34.68	650	850
	Hispanic/Latino	114,136	731.42	35.05	650	850
	Native Hawaiian or Pacific Islander	633	751.74	37.63	650	850
	Multiple Race Selected	11,450	745.73	37.60	650	850
	White	211,869	751.92	35.54	650	850
Economic Status*	Economically Disadvantaged	191,396	728.13	34.57	650	850
	Not Economically Disadvantaged	243,253	754.54	35.92	650	850
English Learner Status	English Learner (EL)	25,527	704.42	29.37	650	850
	Non English Learner	409,299	745.51	36.79	650	850
Disabilities	Students with Disabilities (SWD)	69,917	713.45	35.02	650	850
	Students without Disabilities	222,682	751.22	36.41	650	850
Reading Score		440,160	47.33	15.04	10	90
Gender	Female	214,660	49.35	14.65	10	90
	Male	225,500	45.41	15.17	10	90
Ethnicity	American Indian/Alaska Native	4,784	39.05	12.65	10	90
	Asian	25,536	57.49	14.92	10	90
	Black or African American	71,383	40.83	13.69	10	90
	Hispanic/Latino	114,136	42.36	13.81	10	90
	Native Hawaiian or Pacific Islander	633	50.27	14.94	10	90
	Multiple Race Selected	11,450	48.98	15.12	10	90
	White	211,869	51.08	14.30	10	90
Economic Status*	Economically Disadvantaged	191,396	41.33	13.64	10	90
	Not Economically Disadvantaged	243,253	51.97	14.43	10	90
English Learner Status	English Learner (EL)	25,527	31.52	10.93	10	90
	Non English Learner	409,299	48.34	14.70	10	90
Disabilities	Students with Disabilities (SWD)	69,917	36.12	14.01	10	90
	Students without Disabilities	222,682	50.24	14.55	10	90
Writing Score		440,160	32.07	11.16	10	60
Gender	Female	214,660	35.12	10.08	10	60
	Male	225,500	29.16	11.37	10	60
Ethnicity	American Indian/Alaska Native	4,784	28.33	10.22	10	60

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
	Asian	25,536	39.47	9.79	10	60
	Black or African American	71,383	27.67	11.10	10	60
	Hispanic/Latino	114,136	29.56	10.91	10	60
	Native Hawaiian or Pacific Islander	633	34.65	10.62	10	60
	Multiple Race Selected	11,450	32.24	11.13	10	60
	White	211,869	34.08	10.54	10	60
Economic Status*	Economically Disadvantaged	191,396	28.41	11.02	10	60
	Not Economically Disadvantaged	243,253	34.88	10.44	10	60
English Learner Status	English Learner (EL)	25,527	22.88	10.66	10	60
	Non English Learner	409,299	32.66	10.93	10	60
Disabilities	Students with Disabilities (SWD)	69,917	23.88	11.41	10	60
	Students without Disabilities	222,682	34.42	10.54	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.54 Subgroup Performance for ELA/L Scale Scores: Grade 9

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		275,158	738.99	36.84	650	850
Gender	Female	134,144	746.59	35.93	650	850
	Male	141,014	731.76	36.23	650	850
Ethnicity	American Indian/Alaska Native	3,669	724.04	29.81	650	828
	Asian	17,145	767.48	37.14	650	850
	Black or African American	37,389	723.04	33.25	650	850
	Hispanic/Latino	85,831	727.43	33.73	650	850
	Native Hawaiian or Pacific Islander	417	748.40	36.39	650	850
	Multiple Race Selected	5,774	743.50	37.43	650	850
	White	124,609	748.09	34.93	650	850
Economic Status*	Economically Disadvantaged	117,958	724.96	33.12	650	850
	Not Economically Disadvantaged	151,423	749.76	35.89	650	850
English Learner Status	English Learner (EL)	19,031	702.15	26.81	650	846
	Non English Learner	251,285	741.91	35.92	650	850
Disabilities	Students with Disabilities (SWD)	39,611	712.05	32.17	650	850
	Students without Disabilities	127,890	746.64	36.61	650	850
Reading Score		275,158	45.80	14.81	10	90
Gender	Female	134,144	47.91	14.51	10	90
	Male	141,014	43.80	14.82	10	90
Ethnicity	American Indian/Alaska Native	3,669	38.99	11.87	10	79
	Asian	17,145	56.28	15.24	10	90
	Black or African American	37,389	39.61	13.11	10	90
	Hispanic/Latino	85,831	40.85	13.23	10	90
	Native Hawaiian or Pacific Islander	417	48.93	14.59	10	90
	Multiple Race Selected	5,774	48.26	15.36	10	90
	White	124,609	49.73	14.26	10	90
Economic Status*	Economically Disadvantaged	117,958	40.01	12.98	10	90
	Not Economically Disadvantaged	151,423	50.23	14.59	10	90
English Learner Status	English Learner (EL)	19,031	30.82	9.94	10	90
	Non English Learner	251,285	46.97	14.48	10	90
Disabilities	Students with Disabilities (SWD)	39,611	35.61	12.94	10	90
	Students without Disabilities	127,890	48.36	14.66	10	90
Writing Score		275,158	29.76	12.42	10	60
Gender	Female	134,144	33.10	11.31	10	60
	Male	141,014	26.59	12.60	10	60
Ethnicity	American Indian/Alaska Native	3,669	27.23	11.25	10	56

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
	Asian	17,145	38.17	10.53	10	60
	Black or African American	37,389	24.88	12.40	10	60
	Hispanic/Latino	85,831	27.07	12.23	10	60
	Native Hawaiian or Pacific Islander	417	33.25	11.23	10	60
	Multiple Race Selected	5,774	30.18	12.51	10	60
	White	124,609	31.99	11.72	10	60
Economic Status*	Economically Disadvantaged	117,958	26.08	12.27	10	60
	Not Economically Disadvantaged	151,423	32.60	11.77	10	60
English Learner Status	English Learner (EL)	19,031	19.94	11.20	10	60
	Non English Learner	251,285	30.56	12.17	10	60
Disabilities	Students with Disabilities (SWD)	39,611	21.12	11.96	10	60
	Students without Disabilities	127,890	32.39	11.93	10	60

Note: This table is identical to Table 12.8 in Section 12.

*Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.55 Subgroup Performance for ELA/L Scale Scores: Grade 10

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		192,956	737.78	46.78	650	850
Gender	Female	94,310	747.07	45.62	650	850
	Male	98,646	728.90	46.15	650	850
Ethnicity	American Indian/Alaska Native	3,114	721.00	37.08	650	850
	Asian	13,783	768.38	47.04	650	850
	Black or African American	39,758	721.81	42.69	650	850
	Hispanic/Latino	49,081	724.57	42.71	650	850
	Native Hawaiian or Pacific Islander	330	746.32	48.49	650	850
	Multiple Race Selected	3,797	745.72	47.03	650	850
	White	82,836	748.53	45.74	650	850
Economic Status*	Economically Disadvantaged	75,887	721.15	41.74	650	850
	Not Economically Disadvantaged	111,704	749.16	46.54	650	850
English Learner Status	English Learner (EL)	9,873	690.09	30.87	650	850
	Non English Learner	182,464	740.43	46.09	650	850
Disabilities	Students with Disabilities (SWD)	30,079	705.42	40.08	650	850
	Students without Disabilities	152,199	744.85	45.42	650	850
Reading Score		192,956	45.41	18.37	10	90
Gender	Female	94,310	47.80	17.91	10	90
	Male	98,646	43.13	18.51	10	90
Ethnicity	American Indian/Alaska Native	3,114	37.31	14.23	10	90
	Asian	13,783	56.87	19.03	10	90
	Black or African American	39,758	39.18	16.38	10	90
	Hispanic/Latino	49,081	39.81	16.24	10	90
	Native Hawaiian or Pacific Islander	330	48.07	18.61	10	90
	Multiple Race Selected	3,797	49.04	18.74	10	90
	White	82,836	49.99	18.21	10	90
Economic Status*	Economically Disadvantaged	75,887	38.61	15.86	10	90
	Not Economically Disadvantaged	111,704	50.03	18.49	10	90
English Learner Status	English Learner (EL)	9,873	27.16	11.19	10	90
	Non English Learner	182,464	46.42	18.16	10	90
Disabilities	Students with Disabilities (SWD)	30,079	33.45	15.79	10	90
	Students without Disabilities	152,199	47.96	17.94	10	90
Writing Score		192,956	30.60	13.46	10	60
Gender	Female	94,310	34.02	12.77	10	60
	Male	98,646	27.33	13.30	10	60
Ethnicity	American Indian/Alaska Native	3,114	28.30	11.45	10	60

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
	Asian	13,783	38.46	12.65	10	60
	Black or African American	39,758	26.65	12.92	10	60
	Hispanic/Latino	49,081	27.76	12.96	10	60
	Native Hawaiian or Pacific Islander	330	33.41	13.78	10	60
	Multiple Race Selected	3,797	32.14	13.38	10	60
	White	82,836	32.90	13.16	10	60
Economic Status*	Economically Disadvantaged	75,887	26.79	12.81	10	60
	Not Economically Disadvantaged	111,704	33.24	13.24	10	60
English Learner Status	English Learner (EL)	9,873	18.43	10.52	10	60
	Non English Learner	182,464	31.27	13.29	10	60
Disabilities	Students with Disabilities (SWD)	30,079	21.89	12.19	10	60
	Students without Disabilities	152,199	32.56	13.02	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.56 Subgroup Performance for ELA/L Scale Scores: Grade 11

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		136,934	737.42	38.44	650	850
Gender	Female	66,006	745.25	37.33	650	850
	Male	70,928	730.12	38.02	650	850
Ethnicity	American Indian/Alaska Native	2,940	731.82	32.09	650	850
	Asian	6,768	756.06	41.07	650	850
	Black or African American	25,458	725.39	36.42	650	850
	Hispanic/Latino	36,006	734.05	36.69	650	850
	Native Hawaiian or Pacific Islander	216	747.94	36.99	656	838
	Multiple Race Selected	2,233	740.35	39.15	650	850
	White	63,185	742.35	38.49	650	850
Economic Status*	Economically Disadvantaged	57,183	729.57	36.28	650	850
	Not Economically Disadvantaged	79,641	743.11	38.93	650	850
English Learner Status	English Learner (EL)	5,060	705.46	30.70	650	850
	Non English Learner	131,450	738.71	38.17	650	850
Disabilities	Students with Disabilities (SWD)	21,041	712.07	33.93	650	850
	Students without Disabilities	90,767	743.43	37.37	650	850
Reading Score		136,934	45.21	15.19	10	90
Gender	Female	66,006	47.30	14.70	10	90
	Male	70,928	43.27	15.38	10	90
Ethnicity	American Indian/Alaska Native	2,940	41.18	12.27	10	90
	Asian	6,768	52.40	16.74	10	90
	Black or African American	25,458	40.28	13.88	10	90
	Hispanic/Latino	36,006	43.25	14.07	10	90
	Native Hawaiian or Pacific Islander	216	48.40	14.42	14	84
	Multiple Race Selected	2,233	46.66	15.44	10	90
	White	63,185	47.69	15.43	10	90
Economic Status*	Economically Disadvantaged	57,183	41.69	13.88	10	90
	Not Economically Disadvantaged	79,641	47.75	15.58	10	90
English Learner Status	English Learner (EL)	5,060	31.88	10.93	10	87
	Non English Learner	131,450	45.75	15.09	10	90
Disabilities	Students with Disabilities (SWD)	21,041	35.64	13.39	10	90
	Students without Disabilities	90,767	47.33	14.81	10	90
Writing Score		136,934	29.31	12.85	10	60
Gender	Female	66,006	32.46	12.06	10	60
	Male	70,928	26.37	12.87	10	60
Ethnicity	American Indian/Alaska Native	2,940	30.12	11.09	10	60

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
	Asian	6,768	34.43	12.58	10	60
	Black or African American	25,458	26.05	12.88	10	60
	Hispanic/Latino	36,006	29.00	12.62	10	60
	Native Hawaiian or Pacific Islander	216	33.13	12.08	10	60
	Multiple Race Selected	2,233	29.76	13.00	10	60
	White	63,185	30.20	12.75	10	60
Economic Status*	Economically Disadvantaged	57,183	27.52	12.74	10	60
	Not Economically Disadvantaged	79,641	30.61	12.77	10	60
English Learner Status	English Learner (EL)	5,060	21.28	11.84	10	60
	Non English Learner	131,450	29.64	12.78	10	60
Disabilities	Students with Disabilities (SWD)	21,041	21.71	12.15	10	60
	Students without Disabilities	90,767	31.18	12.45	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.57 Subgroup Performance for Mathematics Scale Scores: Grade 3

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		476,620	742.64	36.55	650	850
Gender	Female	233,536	743.30	35.54	650	850
	Male	243,084	742.00	37.49	650	850
Ethnicity	American Indian/Alaska Native	5,436	723.43	31.20	650	845
	Asian	27,498	772.68	35.53	650	850
	Black or African American	78,668	726.06	34.28	650	850
	Hispanic/Latino	135,427	730.79	33.24	650	850
	Native Hawaiian or Pacific Islander	817	748.97	35.46	650	850
	Multiple Race Selected	15,843	746.02	36.86	650	850
	White	212,345	752.70	33.75	650	850
Economic Status*	Economically Disadvantaged	232,719	728.42	33.34	650	850
	Not Economically Disadvantaged	238,816	756.22	34.28	650	850
English Learner Status	English Learner (EL)	73,569	724.77	32.50	650	850
	Non English Learner	396,435	746.02	36.24	650	850
Disabilities	Students with Disabilities (SWD)	64,259	718.66	37.10	650	850
	Students without Disabilities	248,483	748.90	35.18	650	850

Note: This table is identical to Table 12.9 in Section 12.

*Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.58 Subgroup Performance for Mathematics Scale Scores: Grade 4

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		464,485	737.77	33.15	650	850
Gender	Female	227,746	738.39	32.53	650	850
	Male	236,739	737.18	33.73	650	850
Ethnicity	American Indian/Alaska Native	5,330	718.97	28.41	650	850
	Asian	27,281	765.50	32.54	650	850
	Black or African American	74,511	721.56	30.19	650	850
	Hispanic/Latino	129,527	726.69	29.66	650	850
	Native Hawaiian or Pacific Islander	763	742.60	33.06	650	841
	Multiple Race Selected	14,687	740.95	33.83	650	850
	White	211,905	746.96	30.78	650	850
Economic Status*	Economically Disadvantaged	221,350	724.08	29.58	650	850
	Not Economically Disadvantaged	237,943	750.29	31.30	650	850
English Learner Status	English Learner (EL)	45,533	714.30	27.50	650	850
	Non English Learner	412,407	740.42	32.68	650	850
Disabilities	Students with Disabilities (SWD)	70,190	716.84	31.29	650	850
	Students without Disabilities	238,277	744.68	32.12	650	850

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.59 Subgroup Performance for Mathematics Scale Scores: Grade 5

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		458,218	738.98	32.38	650	850
Gender	Female	224,219	740.12	30.98	650	850
	Male	233,999	737.88	33.63	650	850
Ethnicity	American Indian/Alaska Native	5,103	721.97	27.69	650	845
	Asian	27,309	767.66	32.67	650	850
	Black or African American	72,520	723.17	28.65	650	850
	Hispanic/Latino	125,738	728.32	28.79	650	850
	Native Hawaiian or Pacific Islander	754	745.77	31.97	650	850
	Multiple Race Selected	13,965	741.92	33.19	650	850
	White	212,365	747.23	30.36	650	850
Economic Status*	Economically Disadvantaged	212,859	725.60	28.59	650	850
	Not Economically Disadvantaged	240,211	750.66	30.99	650	850
English Learner Status	English Learner (EL)	34,414	713.04	27.00	650	850
	Non English Learner	417,200	741.17	31.80	650	850
Disabilities	Students with Disabilities (SWD)	72,787	717.51	30.54	650	850
	Students without Disabilities	231,784	745.76	31.36	650	850

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.60 Subgroup Performance for Mathematics Scale Scores: Grade 6

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		457,815	735.97	31.56	650	850
Gender	Female	223,647	737.22	30.59	650	850
	Male	234,168	734.77	32.40	650	850
Ethnicity	American Indian/Alaska Native	5,011	719.46	27.12	650	850
	Asian	27,311	763.35	32.49	650	850
	Black or African American	72,593	719.25	27.68	650	850
	Hispanic/Latino	122,894	724.76	27.83	650	850
	Native Hawaiian or Pacific Islander	654	741.56	31.25	653	825
	Multiple Race Selected	13,223	738.29	31.82	650	850
	White	215,695	744.77	29.17	650	850
Economic Status*	Economically Disadvantaged	207,049	722.21	27.61	650	850
	Not Economically Disadvantaged	245,496	747.42	30.09	650	850
English Learner Status	English Learner (EL)	28,725	707.42	25.52	650	850
	Non English Learner	422,886	737.95	30.96	650	850
Disabilities	Students with Disabilities (SWD)	73,609	713.77	29.27	650	850
	Students without Disabilities	228,854	742.52	30.66	650	850

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.61 Subgroup Performance for Mathematics Scale Scores: Grade 7

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		435,545	734.92	29.41	650	850
Gender	Female	212,807	736.75	28.40	650	850
	Male	222,738	733.16	30.24	650	850
Ethnicity	American Indian/Alaska Native	4,711	719.02	25.21	650	815
	Asian	23,568	759.21	29.60	650	850
	Black or African American	69,206	720.27	26.20	650	850
	Hispanic/Latino	118,410	725.33	26.67	650	850
	Native Hawaiian or Pacific Islander	669	740.49	29.13	650	818
	Multiple Race Selected	11,828	737.25	29.49	650	850
	White	206,697	742.80	27.49	650	850
Economic Status*	Economically Disadvantaged	196,370	723.05	26.49	650	850
	Not Economically Disadvantaged	233,786	744.69	28.11	650	850
English Learner Status	English Learner (EL)	27,917	708.59	25.12	650	842
	Non English Learner	402,011	736.80	28.78	650	850
Disabilities	Students with Disabilities (SWD)	71,763	712.92	28.76	650	850
	Students without Disabilities	214,176	740.64	28.15	650	850

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.62 Subgroup Performance for Mathematics Scale Scores: Grade 8

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		359,231	728.32	38.10	650	850
Gender	Female	173,063	731.08	36.92	650	850
	Male	186,168	725.75	39.00	650	850
Ethnicity	American Indian/Alaska Native	4,341	709.83	29.91	650	850
	Asian	15,633	760.34	43.39	650	850
	Black or African American	63,393	711.34	32.69	650	850
	Hispanic/Latino	101,445	718.11	33.95	650	850
	Native Hawaiian or Pacific Islander	461	733.94	37.84	650	850
	Multiple Race Selected	9,242	731.59	40.07	650	850
	White	164,341	738.46	36.91	650	850
Economic Status*	Economically Disadvantaged	172,743	715.58	33.69	650	850
	Not Economically Disadvantaged	182,929	740.37	38.25	650	850
English Learner Status	English Learner (EL)	26,524	699.69	29.72	650	850
	Non English Learner	328,479	730.74	37.77	650	850
Disabilities	Students with Disabilities (SWD)	65,400	703.90	33.92	650	850
	Students without Disabilities	156,965	732.40	36.86	650	850

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.63 Subgroup Performance for Mathematics Scale Scores: Algebra I

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		323,701	735.00	34.43	650	850
Gender	Female	156,671	737.07	33.14	650	850
	Male	167,030	733.05	35.49	650	850
Ethnicity	American Indian/Alaska Native	3,914	717.72	26.42	650	821
	Asian	19,893	765.60	36.37	650	850
	Black or African American	59,226	719.96	29.30	650	850
	Hispanic/Latino	93,396	723.59	29.61	650	850
	Native Hawaiian or Pacific Islander	504	743.90	36.53	650	850
	Multiple Race Selected	7,298	739.34	34.65	650	850
	White	139,104	744.96	33.14	650	850
Economic Status*	Economically Disadvantaged	140,960	721.81	29.38	650	850
	Not Economically Disadvantaged	177,020	745.33	34.58	650	850
English Learner Status	English Learner (EL)	24,388	709.23	25.78	650	850
	Non English Learner	294,669	737.12	34.14	650	850
Disabilities	Students with Disabilities (SWD)	48,956	713.19	29.77	650	850
	Students without Disabilities	194,227	741.88	34.77	650	850

Note: This table is identical to Table 12.10 in Section 12.

*Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.64 Subgroup Performance for Mathematics Scale Scores: Geometry

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		145,270	732.30	27.43	650	850
Gender	Female	70,985	733.33	26.32	650	850
	Male	74,285	731.30	28.42	650	850
Ethnicity	American Indian/Alaska Native	2,810	719.88	20.47	650	803
	Asian	10,525	756.23	28.33	650	850
	Black or African American	20,101	717.09	23.27	650	837
	Hispanic/Latino	41,010	722.26	23.34	650	850
	Native Hawaiian or Pacific Islander	260	737.30	26.90	671	825
	Multiple Race Selected	2,388	736.99	27.96	650	830
	White	67,925	739.55	25.84	650	850
Economic Status*	Economically Disadvantaged	54,565	720.60	23.25	650	830
	Not Economically Disadvantaged	85,320	739.78	27.35	650	850
English Learner Status	English Learner (EL)	7,205	709.32	21.20	650	844
	Non English Learner	136,802	733.40	27.13	650	850
Disabilities	Students with Disabilities (SWD)	21,307	713.19	23.72	650	830
	Students without Disabilities	108,712	734.59	26.69	650	850

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.65 Subgroup Performance for Mathematics Scale Scores: Algebra II

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		139,956	720.47	39.23	650	850
Gender	Female	70,612	721.12	37.40	650	850
	Male	69,344	719.81	41.01	650	850
Ethnicity	American Indian/Alaska Native	2,445	704.86	29.74	650	829
	Asian	10,600	756.53	42.29	650	850
	Black or African American	23,376	699.45	31.43	650	850
	Hispanic/Latino	34,678	707.89	32.94	650	850
	Native Hawaiian or Pacific Islander	269	729.32	38.16	650	825
	Multiple Race Selected	2,323	726.11	39.34	650	850
	White	66,153	729.12	37.73	650	850
Economic Status*	Economically Disadvantaged	51,634	705.22	32.65	650	850
	Not Economically Disadvantaged	88,194	729.38	40.00	650	850
English Learner Status	English Learner (EL)	4,287	691.18	30.76	650	850
	Non English Learner	135,206	721.39	39.07	650	850
Disabilities	Students with Disabilities (SWD)	15,296	694.65	33.68	650	850
	Students without Disabilities	101,616	724.61	39.10	650	850

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.66 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics I

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		16,581	732.30	34.26	650	850
Gender	Female	8,128	734.36	33.01	650	841
	Male	8,453	730.33	35.30	650	850
Ethnicity	American Indian/Alaska Native	166	715.13	28.28	650	790
	Asian	551	749.85	37.96	650	850
	Black or African American	1,891	720.44	30.49	650	821
	Hispanic/Latino	6,552	723.95	30.77	650	835
	Native Hawaiian or Pacific Islander	30	735.00	36.69	673	810
	Multiple Race Selected	416	736.19	35.74	650	850
	White	6,968	742.18	34.65	650	850
Economic Status*	Economically Disadvantaged	8,230	722.92	30.91	650	838
	Not Economically Disadvantaged	8,347	741.57	34.87	650	850
English Learner Status	English Learner (EL)	2,106	710.19	25.55	650	815
	Non English Learner	14,074	736.33	33.95	650	850
Disabilities	Students with Disabilities (SWD)	2,019	709.23	29.76	650	830
	Students without Disabilities	4,216	725.38	31.59	650	848

Note: This table is identical to Table 12.11 in Section 12.

*Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL).

Table A.12.67 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics II

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		4,655	724.69	31.44	650	850
Gender	Female	2,384	724.92	30.84	650	848
	Male	2,271	724.44	32.07	650	850
Ethnicity	American Indian/Alaska Native	407	719.01	25.94	655	803
	Asian	90	751.29	32.69	678	848
	Black or African American	485	708.87	25.35	650	798
	Hispanic/Latino	2,196	716.63	27.17	650	812
	Native Hawaiian or Pacific Islander	n/r	745.60	30.00	702	804
	Multiple Race Selected	84	745.88	32.50	684	828
	White	1,381	741.55	32.18	650	850
Economic Status*	Economically Disadvantaged	3,069	716.19	26.81	650	817
	Not Economically Disadvantaged	1,546	740.33	32.64	650	839
English Learner Status	English Learner (EL)	431	702.49	19.75	650	783
	Non English Learner	4,173	726.57	31.15	650	839
Disabilities	Students with Disabilities (SWD)	410	702.66	26.10	650	822
	Students without Disabilities	1,650	731.82	29.74	650	820

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table A.12.68 Subgroup Performance for Mathematics Scale Scores: Integrated Mathematics III

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		2,371	722.15	39.29	650	850
Gender	Female	1,263	722.68	37.79	650	850
	Male	1,108	721.54	40.94	650	850
Ethnicity	American Indian/Alaska Native	147	705.21	29.51	650	796
	Asian	30	747.03	42.01	650	811
	Black or African American	88	745.05	41.17	650	824
	Hispanic/Latino	575	718.75	41.72	650	850
	Native Hawaiian or Pacific Islander	n/r	699.00	.	699	699
	Multiple Race Selected	32	724.22	46.07	650	809
	White	1,494	723.35	37.89	650	850
Economic Status*	Economically Disadvantaged	1,203	713.04	37.76	650	837
	Not Economically Disadvantaged	1,167	731.55	38.64	650	850
English Learner Status	English Learner (EL)	61	687.05	30.11	650	805
	Non English Learner	2,236	723.64	39.30	650	850
Disabilities	Students with Disabilities (SWD)	245	680.89	25.79	650	805
	Students without Disabilities	711	713.19	34.95	650	821

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Addendum: Statistical Summary of the Fall/Winter Block 2015 Administration

The addendum presents the results of analyses for the Fall/Winter Block 2015 operational administration. These results are reported separately from the Spring 2016 results because fall testing involved a nonrepresentative subset of students testing only for ELA/L grades 9, 10, and 11, as well as Algebra I, Geometry, and Algebra II. In addition, the fall testing reflected the spring 2015 test design with separate administrations of PBA and EOY. Both online and paper test forms were administered for each test.

To organize the addendum, tables are numbered sequentially according to the section represented by the tables. The reader can refer back to the corresponding section in the technical report for related information on the topic. For example, the first addendum table for Section 5 is numbered ADD.5.1, the second addendum table for Section 5 is numbered ADD.5.2, and so on.

Addendum 5: Test Taker Characteristics

Table ADD.5.1 State Participation in ELA/L Fall 2015 Operational Tests, by Grade

State	Category	English Language Arts/Literacy			
		Total	Grade 9	Grade 10	Grade 11
PARCC	N of Students	25,106	4,202	8,097	12,807
	N of CBT	24,741	4,159	7,999	12,583
	% of CBT	99.0	98.9	98.7	98.2
	N of PBT	365	43	98	224
	% of PBT	1.0	1.0	1.2	1.7
IL	N of Students	605	-	-	605
	% of PARCC Data	2.4	-	-	2.4
	N of CBT	605	-	-	605
	% of CBT	100	-	-	100
	N of PBT	-	-	-	-
	% of PBT	-	-	-	-
MD	N of Students	7,759	488	3,879	3,392
	% of PARCC Data	30.9	1.9	15.5	13.5
	N of CBT	7,615	447	3,825	3,343
	% of CBT	98.1	91.6	98.6	98.6
	N of PBT	144	41	54	49
	% of PBT	1.9	8.4	1.4	1.4
NJ	N of Students	10,633	3,331	3,730	3,572
	% of PARCC Data	42.4	13.3	14.9	14.2
	N of CBT	10,620	3,329	3,725	3,566
	% of CBT	99.8	99.9	99.8	99.8
	N of PBT	-	-	-	-

State	Category	English Language Arts/Literacy			
		Total	Grade 9	Grade 10	Grade 11
	% of PBT	0.1	0.1	0.1	0.2
NM	N of Students	5,887	336	313	5,238
	% of PARCC Data	23.4	1.3	1.2	20.9
	N of CBT	5,679	336	274	5,069
	% of CBT	96.5	100	87.5	96.8
	N of PBT	208	-	39	169
	% of PBT	3.5	-	12.5	3.2
RI	N of Students	222	47	175	-
	% of PARCC Data	0.9	0.2	0.7	-
	N of CBT	222	47	175	-
	% of CBT	100	100	100	-
	N of PBT	-	-	-	-
	% of PBT	-	-	-	-

Table ADD.5.2 State Participation in Mathematics Fall 2015 Operational Tests, by Grade

State	Category	Mathematics			
		Total	A1	GO	A2
PARCC	N of Students	24,309	5,733	6,253	12,323
	N of CBT	24,077	5,694	6,133	12,250
	% of CBT	99.0	99.3	98	99.4
	N of PBT	232	39	120	73
	% of PBT	1.0	0.7	1.9	0.6
IL	N of Students	409	-	-	409
	% of PARCC Data	1.7	-	-	1.7
	N of CBT	409	-	-	409
	% of CBT	100	-	-	100
	N of PBT	-	-	-	-
	% of PBT	-	-	-	-
MD	N of Students	5,491	2,603	639	2,249
	% of PARCC Data	22.6	10.7	2.6	9.2
	N of CBT	5,390	2,570	578	2,242
	% of CBT	98.2	98.7	90.5	99.7
	N of PBT	101	33	61	-
	% of PBT	1.8	1.3	9.5	0.3
NJ	N of Students	9,596	2,433	3,769	3,394
	% of PARCC Data	39.5	10	15.5	14
	N of CBT	9,584	2,430	3,765	3,389
	% of CBT	99.8	99.8	99.8	99.7
	N of PBT	-	-	-	-
	% of PBT	0.1	0.1	0.1	0.1
NM	N of Students	8,561	545	1,747	6,269
	% of PARCC Data	35.2	2.2	7.2	25.8
	N of CBT	8,442	542	1,692	6,208
	% of CBT	98.6	99.4	96.9	99
	N of PBT	119	-	55	61
	% of PBT	1.4	0.6	3.1	1
RI	N of Students	252	152	98	-
	% of PARCC Data	1	0.6	0.4	0
	N of CBT	252	152	98	-
	% of CBT	100	100	100	100
	N of PBT	-	-	-	-
	% of PBT	-	-	-	-

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra.

Table ADD.5.3 State Participation in Spanish Mathematics Fall 2015 Operational Tests, by Grade

State	Category	Mathematics			
		Total	A1	GO	A2
PARCC	N of Students	57	-	-	34
	N of CBT	57	6	17	34
	% of CBT	100	100	100	100
	N of PBT	-	-	-	-
	% of PBT	-	-	-	-
NJ	N of Students	-	-	-	-
	% of PARCC Data	14	8.8	3.5	1.8
	N of CBT	-	-	-	-
	% of CBT	100	100	100	100
	N of PBT	-	-	-	-
	% of PBT	-	-	-	-
NM	N of Students	49	-	-	33
	% of PARCC Data	86	1.8	26.3	57.9
	N of CBT	49	-	-	33
	% of CBT	100	100	100	100
	N of PBT	-	-	-	-
	% of PBT	-	-	-	-

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra.

Table ADD.5.4 All States Combined: Fall 2015 ELA/L Test Takers by Grade and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
9	ALL	4,204	2,145	51.0	2,059	49.0
	CBT	4,159	2,126	51.1	2,033	48.9
	PBT	43	19	44.2	24	55.8
10	ALL	8,101	4,009	49.5	4,092	50.5
	CBT	7,999	3,958	49.5	4,041	50.5
	PBT	98	47	48.0	51	52.0
11	ALL	12,808	5,961	46.5	6,847	53.5
	CBT	12,583	5,840	46.4	6,743	53.6
	PBT	224	120	53.6	104	46.4

Table ADD.5.5 All States Combined: Fall 2015 Mathematics Test Takers by Grade and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
A1	ALL	5,734	2,921	50.9	2,813	49.1
	CBT	5,694	2,906	51.0	2,788	49.0
	PBT	39	14	35.9	25	64.1
GO	ALL	6,255	3,181	50.9	3,074	49.1
	CBT	6,133	3,117	50.8	3,016	49.2
	PBT	120	62	51.7	58	48.3
A2	ALL	12,327	6,381	51.8	5,946	48.2
	CBT	12,250	6,330	51.7	5,920	48.3
	PBT	73	50	68.5	23	31.5

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra.

Table ADD.5.6 All States Combined: Fall 2015 Spanish-Language Mathematics Test Takers by Grade and Gender

Grade	Mode	Valid Cases	Gender			
			Female		Male	
		N	N	%	N	%
A1	ALL	6	4	66.7	2	33.3
	CBT	6	4	66.7	2	33.3
	PBT	-	-	-	-	-
GO	ALL	17	11	64.7	6	35.3
	CBT	17	11	64.7	6	35.3
	PBT	-	-	-	-	-
A2	ALL	34	21	61.8	13	38.2
	CBT	34	21	61.8	13	38.2
	PBT	-	-	-	-	-

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra.

Table ADD.5.7 Fall 2015 ELA/L Test Takers by State, Grade, and Gender

State	Grade	Mode	Valid Cases	Gender			
				Female		Male	
			N	N	%	N	%
IL	11	CBT	605	304	50.2	301	49.8
MD	9	CBT	447	228	46.7	219	44.9
		PBT	41	n/a	3.9	n/a	4.5
MD	10	CBT	3,825	1,829	47.1	1,996	51.4
		PBT	54	n/a	0.6	31	0.8
MD	11	CBT	3,343	1,567	46.2	1,776	52.4
		PBT	49	30	0.9	n/a	0.6
NJ	9	CBT	3,329	1,712	51.4	1,617	48.5
NJ	10	CBT	3,725	1,907	51.1	1,818	48.7
NJ	11	CBT	3,566	1,738	48.6	1,828	51.2
NM	9	CBT	336	167	49.7	169	50.3
NM	10	CBT	274	134	42.8	140	44.7
		PBT	39	n/a	6.7	n/a	5.8
NM	11	CBT	5,069	2,231	42.6	2,838	54.2
		PBT	169	85	1.6	84	1.6
RI	9	CBT	47	n/a	40.4	28	59.6
	10	CBT	175	88	50.3	87	49.7

Note: n/a = not applicable.

Table ADD.5.8 Fall 2015 Mathematics Test Takers by State, Content Area, and Gender

State	Grade	Mode	Valid Cases	Gender			
				Female		Male	
			N	N	%	N	%
IL	A2	CBT	409	215	52.6	194	47.4
MD	A1	CBT	2,570	1,258	48.3	1,312	50.4
MD		PBT	33	n/a	0.4	n/a	0.8
MD	GO	CBT	578	286	44.8	292	45.7
MD		PBT	61	31	4.9	30	4.7
MD	A2	CBT	2,242	1,128	50.2	1,114	49.5
NJ	A1	CBT	2,430	1,301	53.5	1,129	46.4
NJ	GO	CBT	3,765	1,956	51.9	1,809	48.0
NJ	A2	CBT	3,389	1,746	51.4	1,643	48.4
NM	A1	CBT	542	267	49.0	275	50.5
NM	GO	CBT	1,692	810	46.4	882	50.5
		PBT	55	29	1.7	26	1.5
NM	A2	CBT	6,208	3,240	51.7	2,968	47.3
		PBT	61	41	0.7	n/a	0.3
RI	A1	CBT	152	80	52.6	72	47.4
RI	GO	CBT	98	65	66.3	33	33.7

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra.

Table ADD.5.9 Demographic Information for Fall 2015 Grade 9 ELA/L, Overall and by State

Demographic	PARCC	IL	MD	NJ	NM	RI
Econ Dis (%)	38.0	n/a	44.5	37.4	28.0	85.1
SWD (%)	15.2	n/a	15.8	16.1	n/r	n/r
EL (%)	2.0	n/a	n/r	0.9	10.4	n/r
Male (%)	49.0	n/a	49.4	48.6	50.3	59.6
Female (%)	51.0	n/a	50.6	51.4	49.7	n/r
AmInd/ANat (%)	0.6	n/a	n/r	n/r	6.0	n/a
Asian (%)	4.7	n/a	n/r	5.4	n/r	n/r
Black/AA (%)	23.1	n/a	16.2	26.4	n/r	n/r
Hisp/Lat (%)	18.5	n/a	8.2	16.8	44.0	66.0
Wh/Caus (%)	49.0	n/a	66.6	49.5	25.3	n/r
NtvHawaii/Pacific (%)	n/r	n/a	n/r	n/r	n/r	n/a
Two Or More (%)	3.1	n/a	6.1	1.5	14.6	n/r
Unknown (%)	0.7	n/a	n/a	n/r	8.3	n/a

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table ADD.5.10 Demographic Information for Fall 2015 Grade 10 ELA/L, Overall and by State

Demographic	PARCC	IL	MD	NJ	NM	RI
Econ Dis (%)	33.3	n/a	27.6	37.7	41.9	49.1
SWD (%)	15.3	n/a	14.3	17.1	7.3	13.7
EL (%)	2.3	n/a	2.1	1.3	16.3	n/r
Male (%)	50.5	n/a	52.2	48.8	50.5	49.7
Female (%)	49.5	n/a	47.8	51.2	49.5	50.3
AmInd/ANat (%)	0.6	n/a	n/r	n/r	8.0	n/r
Asian (%)	3.7	n/a	3.1	4.6	n/r	n/r
Black/AA (%)	20.4	n/a	17.9	24.9	n/r	14.3
Hisp/Lat (%)	15.9	n/a	9.9	17.7	62.6	29.1
Wh/Caus (%)	56.3	n/a	65.2	50.4	19.5	51.4
NtvHawaii/Pacific (%)	n/r	n/a	n/r	n/r	n/a	n/a
Two Or More (%)	2.5	n/a	3.4	1.7	n/r	n/r
Unknown (%)	0.4	n/a	n/a	n/r	n/r	n/a

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table ADD.5.11 Demographic Information for Fall 2015 Grade 11 ELA/L, Overall and by State

Demographic	PARCC	IL	MD	NJ	NM	RI
Econ Dis (%)	42.1	16.2	20.8	36.9	62.6	n/a
SWD (%)	17.9	7.4	16.6	17.9	20.1	n/a
EL (%)	7.2	n/r	0.7	1.4	16.1	n/a
Male (%)	53.5	49.8	52.9	51.2	55.8	n/a
Female (%)	46.5	50.2	47.1	48.8	44.2	n/a
AmInd/ANat (%)	5.8	n/a	n/r	n/r	13.7	n/a
Asian (%)	2.4	n/r	2.3	4.5	1.1	n/a
Black/AA (%)	15.0	4.6	24.4	27.1	1.9	n/a
Hisp/Lat (%)	33.0	13.9	7.7	16.7	62.8	n/a
Wh/Caus (%)	40.4	64.6	62.6	49.5	17.1	n/a
NtvHawaii/Pacific (%)	n/r	n/r	n/r	n/r	n/r	n/a
Two Or More (%)	1.7	n/r	2.7	1.7	0.9	n/a
Unknown (%)	1.6	13.2	n/a	n/r	2.3	n/a

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table ADD.5.12 Demographic Information for Fall 2015 Algebra I, Overall and by State

Demographic	PARCC	IL	MD	NJ	NM	RI
Econ Dis (%)	38.3	n/a	29.8	45.9	39.8	57.2
SWD (%)	13.2	n/a	13.2	13.1	14.5	n/r
EL (%)	2.9	n/a	2.2	1.3	12.8	n/r
Male (%)	49.1	n/a	51.2	46.5	50.6	47.4
Female (%)	50.9	n/a	48.8	53.5	49.4	52.6
AmInd/ANat (%)	1.0	n/a	n/r	n/r	8.4	n/a
Asian (%)	3.1	n/a	3.0	3.8	n/r	n/r
Black/AA (%)	20.7	n/a	17.5	28.9	n/r	n/r
Hisp/Lat (%)	20.5	n/a	10.3	21.2	61.1	38.2
Wh/Caus (%)	51.0	n/a	65.2	43.8	17.6	43.4
NtvHawaii/Pacific (%)	n/r	n/a	n/r	n/r	n/a	n/a
Two Or More (%)	2.7	n/a	3.6	1.7	n/r	n/r
Unknown (%)	0.8	n/a	n/a	n/r	8.3	n/a

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table ADD.5.13 Demographic Information for Fall 2015 Geometry, Overall and by State

Demographic	PARCC	IL	MD	NJ	NM	RI
Econ Dis (%)	46.5	n/a	42.3	40.1	60.8	64.3
SWD (%)	17.8	n/a	16.0	16.6	21.6	n/r
EL (%)	5.7	n/a	n/r	2.1	14.6	n/r
Male (%)	49.1	n/a	50.4	48.0	52.0	33.7
Female (%)	50.9	n/a	49.6	52.0	48.0	66.3
AmInd/ANat (%)	3.6	n/a	n/r	n/r	12.2	n/a
Asian (%)	3.9	n/a	n/r	5.6	n/r	n/r
Black/AA (%)	18.0	n/a	18.2	25.6	1.8	n/r
Hisp/Lat (%)	30.1	n/a	6.7	18.6	62.2	53.1
Wh/Caus (%)	41.3	n/a	69.0	48.0	17.5	28.6
NtvHawaii/Pacific (%)	n/r	n/a	n/a	n/r	n/r	n/a
Two Or More (%)	1.9	n/a	3.6	1.6	1.9	n/r
Unknown (%)	1.1	n/a	n/a	n/r	3.4	n/a

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Table ADD.5.14 Demographic Information for Fall 2015 Algebra II, Overall and by State

Demographic	PARCC	IL	MD	NJ	NM	RI
Econ Dis (%)	44.9	9.5	14.6	38.1	61.7	n/r
SWD (%)	12.3	8.8	11.1	13.3	12.3	n/r
EL (%)	5.7	n/r	n/r	1.0	10.4	n/a
Male (%)	48.2	47.4	49.5	48.5	47.7	n/r
Female (%)	51.8	52.6	50.5	51.5	52.3	n/r
AmInd/ANat (%)	6.1	n/r	n/r	n/r	11.8	n/a
Asian (%)	2.9	n/r	3.9	5.7	1.1	n/a
Black/AA (%)	12.2	6.1	25.6	23.3	1.8	n/a
Hisp/Lat (%)	39.6	15.9	7.2	17.2	65.0	n/r
Wh/Caus (%)	36.6	72.4	60.0	51.5	17.9	n/a
NtvHawaii/Pacific (%)	n/r	n/a	n/r	n/r	n/r	n/a
Two Or More (%)	1.6	n/r	2.9	1.7	1.0	n/r
Unknown (%)	0.9	n/r	n/a	n/r	1.4	n/a

Note: Econ Dis = Economically Disadvantaged; SWD = Student with Disabilities; EL = English learner; AmInd/ANat = American Indian/Alaska Native; Black/AA = Black/African American; Hisp/Lat = Hispanic/Latino; Wh/Caus = White/Caucasian; NtvHawaii/Pacific = Native Hawaiian or Other Pacific Islander; Two or More = two or more races reported; n/r = not reported; n/a = not applicable.

Addendum 8: Reliability

Table ADD.8.1 shows the total group level reliability estimates, raw score SEM, and scale score SEM for the Fall 2015 forms. Tables ADD.8.2 – ADD.8.9 show the subgroup reliability estimates, raw score SEM, and scale score SEM. A minimum sample size of 100 per core form was required for calculating the reliability estimates; therefore, the subgroup totals may not equal the total group sample size. Tables ADD.8.10 – ADD.8.12 provide the claim and subclaim reliability and raw score SEM estimates for the Fall 2015 forms.

Table ADD.8.1 Summary of Test Reliability Estimates for Fall 2015 Total Group

Content	Grade/ Course	Mode	Sample Size	Maximum Possible Score	Reliability	Raw Score SEM	Scale Score SEM	# of Core Operational Forms
ELA/L	9	CBT	4,140	135	0.92	6.47	9.42	3
ELA/L	10	CBT	7,978	137	0.94	6.66	10.88	3
ELA/L	11	CBT	12,502	137	0.94	6.17	9.19	3
ELA/L	11	PBT	224	135	0.93	6.16	9.27	2
Mathematics	A1	CBT	5,672	97	0.89	3.47	10.01	3
Mathematics	GO	CBT	6,089	96	0.89	3.33	7.77	3
Mathematics	GO	PBT	120	96	0.91	3.79	7.95	2
Mathematics	A2	CBT	12,139	107	0.91	3.64	10.44	3

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra. ELA/L grades 9 and 10, Algebra I and Algebra II had insufficient sample sizes for PBT.

Table ADD.8.2 Summary of Test Reliability Estimates for Fall 2015 Subgroups: Grade 9 ELA/L

	CBT			
	Sample Size	Maximum Possible Raw Score	Reliability	Scale Score Standard Error of Measurement
Total Group	4,140	135	0.92	9.42
Gender				
Male	1,869	135	0.92	9.27
Female	2,030	135	0.92	9.48
Unknown/Missing	-	-	-	-
Ethnicity				
White	1,935	135	0.92	9.47
African American	861	135	0.91	9.23
Asian/Pacific Islander	197	135	0.93	9.82
American Indian/Alaska Native	-	-	-	-
Hispanic	731	135	0.91	9.38
Multiple	120	135	0.92	8.92
Special Instructional Needs				
Economically Disadvantaged	1,457	135	0.91	9.43
Not-economically Disadvantaged	2,237	135	0.92	9.41
English Learner (EL)	-	-	-	-
Non English Learner	3,967	135	0.92	9.38
Students with Disabilities (SWD)	585	135	0.89	9.33
Students without Disabilities	3,295	135	0.92	9.37
Students Taking Accommodated Forms				
A: ASL	-	-	-	-
C: Closed-Caption	-	-	-	-
R: Screen Reader	-	-	-	-
T: Text-to-Speech	141	135	0.78	10.21

Table ADD.8.3 Summary of Test Reliability Estimates for Fall 2015 Subgroups: Grade 10 ELA/L

	CBT			
	Sample Size	Maximum Possible Raw Score	Reliability	Scale Score Standard Error of Measurement
Total Group	7,978	137	0.94	10.88
Gender				
Male	3,900	137	0.94	10.58
Female	3,860	137	0.94	11.16
Unknown/Missing	-	-	-	-
Ethnicity				
White	4,387	137	0.94	10.94
African American	1,568	137	0.93	10.48
Asian/Pacific Islander	304	137	0.94	11.71
American Indian/Alaska Native	-	-	-	-
Hispanic	1,225	137	0.93	10.60
Multiple	196	137	0.94	11.09
Special Instructional Needs				
Economically Disadvantaged	2,542	137	0.93	10.70
Not-economically Disadvantaged	5,108	137	0.94	11.02
English Learner (EL)	181	137	0.81	10.29
Non English Learner	7,675	137	0.94	10.91
Students with Disabilities (SWD)	1,195	137	0.93	10.26
Students without Disabilities	6,518	137	0.94	10.99
Students Taking Accommodated Forms				
A: ASL	-	-	-	-
C: Closed-Caption	-	-	-	-
R: Screen Reader	-	-	-	-
T: Text-to-Speech	102	137	0.86	11.05

Table ADD.8.4 Summary of CBT Test Reliability Estimates for Fall 2015 Subgroups: Grade 11 ELA/L

	CBT			
	Sample Size	Maximum Possible Raw Score	Reliability	Scale Score Standard Error of Measurement
Total Group	12,502	137	0.94	9.19
Gender				
Male	6,589	137	0.94	8.92
Female	5,652	137	0.94	9.43
Unknown/Missing	-	-	-	-
Ethnicity				
White	5,047	137	0.94	9.54
African American	1,823	137	0.91	9.42
Asian/Pacific Islander	313	137	0.95	9.70
American Indian/Alaska Native	695	137	0.81	9.83
Hispanic	3,973	137	0.91	9.03
Multiple	200	137	0.94	9.54
Special Instructional Needs				
Economically Disadvantaged	5,150	137	0.90	9.23
Not-economically Disadvantaged	6,638	137	0.94	9.37
English Learner (EL)	823	137	0.75	10.00
Non English Learner	10,971	137	0.94	9.24
Students with Disabilities (SWD)	2,149	137	0.92	8.62
Students without Disabilities	9,719	137	0.94	9.26
Students Taking Accommodated Forms				
A: ASL	-	-	-	-
C: Closed-Caption	-	-	-	-
R: Screen Reader	-	-	-	-
T: Text-to-Speech	199	137	0.86	8.52

Table ADD.8.5 Summary of PBT Test Reliability Estimates for Fall 2015 Subgroups: Grade 11 ELA/L

	PBT			
	Sample Size	Maximum Possible Raw Score	Reliability	Scale Score Standard Error of Measurement
Total Group	224	135	0.93	9.27
Gender				
Male	104	135	0.91	9.17
Female	120	135	0.94	9.46
Unknown/Missing	-	-	-	-
Ethnicity	-	-	-	-
White	-	-	-	-
African American	-	-	-	-
Asian/Pacific Islander	-	-	-	-
American Indian/Alaska Native	-	-	-	-
Hispanic	137	135	0.81	9.85
Multiple	-	-	-	-
Special Instructional Needs				
Economically Disadvantaged	154	135	0.87	9.68
Not-economically Disadvantaged	-	-	-	-
English Learner (EL)	-	-	-	-
Non English Learner	157	135	0.93	9.66
Students with Disabilities (SWD)	-	-	-	-
Students without Disabilities	157	135	0.93	9.05
Students Taking Accommodated Forms				
A: ASL	-	-	-	-
C: Closed-Caption	-	-	-	-
R: Screen Reader	-	-	-	-
T: Text-to-Speech	-	-	-	-

Table ADD.8.6 Summary of Test Reliability Estimates for Fall 2015 Subgroups: Algebra I

	CBT			
	Sample Size	Maximum Possible Raw Score	Reliability	Scale Score Standard Error of Measurement
Total Group	5,672	97	0.89	10.01
Gender				
Male	2,774	97	0.90	9.85
Female	2,898	97	0.88	10.13
Unknown/Missing	-	-	-	-
Ethnicity				
White	2,902	97	0.89	9.82
African American	1,132	97	0.88	10.20
Asian/Pacific Islander	186	97	0.91	9.46
American Indian/Alaska Native	-	-	-	-
Hispanic	1,116	97	0.86	10.50
Multiple	151	97	0.86	10.03
Special Instructional Needs				
Economically Disadvantaged	2,169	97	0.85	10.53
Not-economically Disadvantaged	3,418	97	0.90	9.86
English Learner (EL)	136	97	0.87	9.31
Non English Learner	5,423	97	0.89	10.00
Students with Disabilities (SWD)	738	97	0.84	10.90
Students without Disabilities	4,831	97	0.89	9.82
Students Taking Accommodated Forms				
A: ASL	-	-	-	-
C: Closed-Caption	-	-	-	-
R: Screen Reader	-	-	-	-
T: Text-to-Speech	346	97	0.81	11.73
Students Taking Translated Forms				
Spanish Language Form	-	-	-	-

Table ADD.8.7 Summary of CBT Test Reliability Estimates for Fall 2015 Subgroups: Geometry

	CBT			
	Sample Size	Maximum Possible Raw Score	Reliability	Scale Score Standard Error of Measurement
Total Group	6,089	96	0.89	7.77
Gender				
Male	2,994	96	0.89	7.72
Female	3,095	96	0.89	7.79
Unknown/Missing	-	-	-	-
Ethnicity				
White	2,515	96	0.89	7.55
African American	1,045	96	0.83	8.62
Asian/Pacific Islander	228	96	0.93	7.67
American Indian/Alaska Native	161	96	0.65	9.06
Hispanic	1,831	96	0.81	8.61
Multiple	109	96	0.86	8.46
Special Instructional Needs				
Economically Disadvantaged	2,828	96	0.84	8.40
Not-economically Disadvantaged	3,104	96	0.90	7.61
English Learner (EL)	273	96	0.66	9.99
Non English Learner	5,623	96	0.89	7.74
Students with Disabilities (SWD)	1,080	96	0.82	8.43
Students without Disabilities	4,838	96	0.89	7.62
Students Taking Accommodated Forms				
A: ASL	-	-	-	-
C: Closed-Caption	-	-	-	-
R: Screen Reader	-	-	-	-
T: Text-to-Speech	616	96	0.83	8.87
Students Taking Translated Forms				
Spanish Language Form	-	-	-	-

Table ADD.8.8 Summary of PBT Test Reliability Estimates for Fall 2015 Subgroups: Geometry

	PBT			
	Sample Size	Maximum Possible Raw Score	Reliability	Scale Score Standard Error of Measurement
Total Group	120	96	0.91	7.95
Gender				
Male	-	-	-	-
Female	-	-	-	-
Unknown/Missing	-	-	-	-
Ethnicity				
White	-	-	-	-
African American	-	-	-	-
Asian/Pacific Islander	-	-	-	-
American Indian/Alaska Native	-	-	-	-
Hispanic	-	-	-	-
Multiple	-	-	-	-
Special Instructional Needs				
Economically Disadvantaged	-	-	-	-
Not-economically Disadvantaged	-	-	-	-
English Learner (EL)	-	-	-	-
Non English Learner	111	96	0.90	7.77
Students with Disabilities (SWD)	-	-	-	-
Students without Disabilities	-	-	-	-
Students Taking Accommodated Forms				
A: ASL	-	-	-	-
C: Closed-Caption	-	-	-	-
R: Screen Reader	-	-	-	-
T: Text-to-Speech	-	-	-	-
Students Taking Translated Forms				
Spanish Language Form	-	-	-	-

Table ADD.8.9 Summary of Test Reliability Estimates for Fall 2015 Subgroups: Algebra II

	PBT			
	Sample Size	Maximum Possible Raw Score	Reliability	Scale Score Standard Error of Measurement
Total Group	12,139	107	0.91	10.44
Gender				
Male	5,866	107	0.92	10.22
Female	6,273	107	0.90	10.61
Unknown/Missing	-	-	-	-
Ethnicity				
White	4,488	107	0.92	10.43
African American	1,417	107	0.83	12.03
Asian/Pacific Islander	346	107	0.93	10.34
American Indian/Alaska Native	700	107	0.68	12.54
Hispanic	4,788	107	0.79	11.70
Multiple	190	107	0.89	10.82
Special Instructional Needs				
Economically Disadvantaged	5,439	107	0.80	11.69
Not-economically Disadvantaged	6,147	107	0.92	10.55
English Learner (EL)	601	107	0.58	13.85
Non English Learner	10,961	107	0.91	10.46
Students with Disabilities (SWD)	1,415	107	0.86	10.84
Students without Disabilities	10,213	107	0.91	10.33
Students Taking Accommodated Forms				
A: ASL	-	-	-	-
C: Closed-Caption	-	-	-	-
R: Screen Reader	-	-	-	-
T: Text-to-Speech	745	107	0.90	11.41
Students Taking Translated Forms				
Spanish Language Form	-	-	-	-

Table ADD.8.10 Fall 2015 Average ELA/L Reliability Estimates for Reading Total and Subscores

Grade Level	Mode	Reading: Total		Reading: Literature		Reading: Information		Reading: Vocabulary	
		Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability
9	CBT	90	0.90	28	0.76	42	0.82	20	0.65
10	CBT	92	0.92	28	0.76	44	0.86	20	0.68
11	CBT	92	0.91	40	0.84	34	0.78	18	0.66
11	PBT	90	0.91	42	0.82	32	0.80	16	0.59

Table ADD.8.11 Fall 2015 Average ELA/L Reliability Estimates for Writing Total and Subscores

Grade Level	Mode	Writing: Total		Writing: Written Expression		Writing: Knowledge Language and Conventions	
		Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability
9	CBT	45	0.85	36	0.81	9	0.82
10	CBT	45	0.87	36	0.86	9	0.88
11	CBT	45	0.87	36	0.85	9	0.86
11	PBT	45	0.86	36	0.85	9	0.85

Table ADD.8.12 Fall 2015 Average Mathematics Reliability Estimates for Total Test and Subscores

Grade Level	Mode	Major Content		Additional & Supporting Content		Mathematics Reasoning		Modeling Practice	
		Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability	Max Possible Raw Score	Reliability
A1	CBT	35	0.77	30	0.70	14	0.55	18	0.50
GO	CBT	38	0.78	26	0.69	14	0.42	18	0.52
A2	PBT	39	0.82	25	0.71	14	0.42	18	0.58
A2	CBT	38	0.76	27	0.78	18	0.62	24	0.53

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra.

Tables ADD.8.13 and ADD.8.14 provide information about the accuracy and the consistency of two classifications made on the basis of the scores on the Fall Block 2015 English Language Arts/Literacy and mathematics assessments respectively. The columns labeled “Exact level” provide the classification of the student into one of five achievement levels. The columns labeled “Level 4 or higher vs. 3 or lower” provide the classification of the student as being either in one of the upper two levels (Levels 4 and 5) or in one of the lower three levels (Levels 1, 2, and 3).

Tables ADD.8.15 to ADD.8.20 provide more detailed information about the accuracy and the consistency of the classification of students into proficiency levels for each Fall Block 2015 PARCC assessment. Each cell in the 5-by-5 table shows the estimated proportion of students who would be classified into a particular combination of proficiency levels. The sum of the five ***bold italicized*** values on the diagonal should equal the exact level of decision accuracy or consistency presented in Tables ADD.8.13 or ADD.8.14 for the corresponding PARCC assessment. For “Level 4 and higher vs. 3 and lower” found in Tables ADD.8.13 or ADD.8.14, the sum of the shaded values in Tables ADD.8.13 to ADD.8.18 should equal the level of decision accuracy or consistency for the corresponding PARCC assessment in ADD.8.13 or ADD.8.14. Note that the sums based on values in Tables ADD.8.13 to ADD.8.18 may not match exactly to the values in ADD.8.13 or ADD.8.14 due to truncation and rounding.

Table ADD.8.13 Reliability of Classification: Summary for ELA/L Fall Block 2015

Grade Level	Testing Mode	Decision Accuracy: Proportion Accurately Classified		Decision Consistency: Proportion Consistently Classified	
		Level 4 or higher vs. 3 or lower		Level 4 or higher vs. 3 or lower	
		Exact level		Exact level	
9	CBT	0.75	0.91	0.65	0.88
	PBT	--	--	--	--
10	CBT	0.76	0.93	0.67	0.90
	PBT	--	--	--	--
11	CBT	0.76	0.94	0.67	0.92
	PBT	0.76	0.95	0.66	0.93

Note: “--” means insufficient sample size (< 100 students).

Table ADD.8.14 Reliability of Classification: Summary for Mathematics Fall Block 2015

Grade Level	Testing Mode	Decision Accuracy: Proportion Accurately Classified		Decision Consistency: Proportion Consistently Classified	
		Level 4 or higher vs. 3 or lower		Level 4 or higher vs. 3 or lower	
		Exact Level		Exact Level	
A1	CBT	0.74	0.91	0.64	0.87
	PBT	--	--	--	--

GO	CBT	0.77	0.94	0.67	0.92
	PBT	0.77	0.92	0.68	0.89
A2	CBT	0.77	0.95	0.69	0.93
	PBT	--	--	--	--

Note: A1 = Algebra I, GO = Geometry, A2 = Algebra II. “--” means insufficient sample size (< 100 students).

Table ADD.8.15 Reliability of Classification: Grade 9 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.10	0.02	0.00	0.00	0.00	0.11
		700 – 724	0.03	0.13	0.04	0.00	0.00	0.20
		725 – 749	0.00	0.04	0.19	0.05	0.00	0.27
		750 – 809	0.00	0.00	0.05	0.28	0.02	0.35
		810 – 850	0.00	0.00	0.00	0.01	0.06	0.07
	Decision Consistency	650 – 699	0.09	0.03	0.00	0.00	0.00	0.12
		700 – 724	0.03	0.11	0.06	0.00	0.00	0.20
		725 – 749	0.00	0.05	0.15	0.06	0.00	0.26
		750 – 809	0.00	0.00	0.06	0.25	0.03	0.34
		810 – 850	0.00	0.00	0.00	0.03	0.05	0.08

Table ADD.8.16 Reliability of Classification: Grade 10 ELA/L

		Full Summative Scale Score	Level 1	Level 2	Level 3	Level 4	Level 5	Category Total
CBT	Decision Accuracy	650 – 699	0.21	0.03	0.00	0.00	0.00	0.23
		700 – 724	0.03	0.11	0.04	0.00	0.00	0.18
		725 – 749	0.00	0.03	0.13	0.04	0.00	0.20
		750 – 809	0.00	0.00	0.04	0.23	0.02	0.29
		810 – 850	0.00	0.00	0.00	0.02	0.08	0.10
	Decision Consistency	650 – 699	0.20	0.04	0.00	0.00	0.00	0.24
		700 – 724	0.04	0.09	0.05	0.00	0.00	0.18
		725 – 749	0.00	0.04	0.10	0.05	0.00	0.19
		750 – 809	0.00	0.00	0.05	0.20	0.03	0.28
		810 – 850	0.00	0.00	0.00	0.03	0.08	0.11

Table ADD.8.17 Reliability of Classification: Grade 11 ELA/L

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Summative Scale Score						Total
CBT	Decision Accuracy	650 – 699	0.18	0.04	0.00	0.00	0.00	0.22
		700 – 724	0.04	0.21	0.04	0.00	0.00	0.28
		725 – 749	0.00	0.04	0.14	0.03	0.00	0.21
		750 – 809	0.00	0.00	0.03	0.17	0.01	0.21
		810 – 850	0.00	0.00	0.00	0.01	0.05	0.06
	Decision Consistency	650 – 699	0.17	0.06	0.00	0.00	0.00	0.23
		700 – 724	0.05	0.18	0.05	0.00	0.00	0.27
		725 – 749	0.00	0.06	0.12	0.04	0.00	0.21
		750 – 809	0.00	0.00	0.04	0.16	0.02	0.21
		810 – 850	0.00	0.00	0.00	0.02	0.05	0.07
PBT	Decision Accuracy	650 – 699	0.24	0.06	0.00	0.00	0.00	0.30
		700 – 724	0.04	0.22	0.04	0.00	0.00	0.30
		725 – 749	0.00	0.04	0.14	0.02	0.00	0.20
		750 – 809	0.00	0.00	0.02	0.11	0.01	0.14
		810 – 850	0.00	0.00	0.00	0.01	0.05	0.05
	Decision Consistency	650 – 699	0.22	0.08	0.00	0.00	0.00	0.30
		700 – 724	0.06	0.18	0.05	0.00	0.00	0.30
		725 – 749	0.00	0.05	0.12	0.03	0.00	0.20
		750 – 809	0.00	0.00	0.04	0.10	0.01	0.15
		810 – 850	0.00	0.00	0.00	0.01	0.04	0.05

Table ADD.8.18 Reliability of Classification: Algebra I

		Full	Level 1	Level 2	Level 3	Level 4	Level 5	Category
		Summative Scale Score						Total
CBT	Decision Accuracy	650 – 699	0.06	0.03	0.00	0.00	0.00	0.09
		700 – 724	0.02	0.22	0.04	0.00	0.00	0.29
		725 – 749	0.00	0.07	0.19	0.05	0.00	0.31
		750 – 809	0.00	0.00	0.04	0.26	0.01	0.31
		810 – 850	0.00	0.00	0.00	0.00	0.01	0.01
	Decision Consistency	650 – 699	0.06	0.05	0.00	0.00	0.00	0.11
		700 – 724	0.03	0.18	0.06	0.00	0.00	0.27
		725 – 749	0.00	0.08	0.15	0.06	0.00	0.30
		750 – 809	0.00	0.00	0.06	0.24	0.01	0.31
		810 – 850	0.00	0.00	0.00	0.01	0.01	0.01

Table ADD.8.19 Reliability of Classification: Geometry

		Full						Category
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Total
		Scale Score						
CBT	Decision Accuracy	650 – 699	0.08	0.04	0.00	0.00	0.00	0.12
		700 – 724	0.03	0.34	0.05	0.00	0.00	0.42
		725 – 749	0.00	0.06	0.22	0.04	0.00	0.31
		750 – 809	0.00	0.00	0.02	0.12	0.00	0.14
		810 – 850	0.00	0.00	0.00	0.00	0.00	0.00
	Decision Consistency	650 – 699	0.08	0.06	0.00	0.00	0.00	0.14
		700 – 724	0.04	0.29	0.06	0.00	0.00	0.39
		725 – 749	0.00	0.08	0.19	0.04	0.00	0.31
		750 – 809	0.00	0.00	0.04	0.11	0.00	0.15
		810 – 850	0.00	0.00	0.00	0.00	0.00	0.01
PBT	Decision Accuracy	650 – 699	0.07	0.02	0.00	0.00	0.00	0.09
		700 – 724	0.02	0.24	0.04	0.00	0.00	0.30
		725 – 749	0.00	0.06	0.22	0.05	0.00	0.32
		750 – 809	0.00	0.00	0.03	0.25	0.01	0.28
		810 – 850	0.00	0.00	0.00	0.00	0.00	0.00
	Decision Consistency	650 – 699	0.06	0.04	0.00	0.00	0.00	0.11
		700 – 724	0.03	0.20	0.05	0.00	0.00	0.28
		725 – 749	0.00	0.07	0.18	0.06	0.00	0.32
		750 – 809	0.00	0.00	0.05	0.23	0.01	0.29
		810 – 850	0.00	0.00	0.00	0.01	0.00	0.01

Table ADD.8.20 Reliability of Classification: Algebra II

		Full						Category
		Summative	Level 1	Level 2	Level 3	Level 4	Level 5	Total
		Scale Score						
CBT	Decision Accuracy	650 – 699	0.38	0.05	0.00	0.00	0.00	0.43
		700 – 724	0.05	0.19	0.03	0.00	0.00	0.27
		725 – 749	0.00	0.04	0.09	0.03	0.00	0.16
		750 – 809	0.00	0.00	0.02	0.11	0.00	0.13
		810 – 850	0.00	0.00	0.00	0.00	0.00	0.00
	Decision Consistency	650 – 699	0.37	0.07	0.00	0.00	0.00	0.44
		700 – 724	0.06	0.15	0.04	0.00	0.00	0.25
		725 – 749	0.00	0.06	0.07	0.03	0.00	0.16
		750 – 809	0.00	0.00	0.03	0.10	0.00	0.14
		810 – 850	0.00	0.00	0.00	0.00	0.00	0.01

Addendum 9: Validity

The intercorrelations for the Fall 2015 tests are presented in Tables ADD.9.1 through ADD.9.3 for ELA/L grades 9, 10, and 11 and Tables ADD.9.4 through ADD.9.6 for the traditional mathematics courses (A1, GO, A2). Like the spring intercorrelations, the ELA/L all have moderate to high values with the writing subclaims being highly intercorrelated. The mathematics intercorrelations have moderate values. Tables ADD.9.7 through ADD.9.9 are the correlations between ELA/L and mathematics from the fall block.

Table ADD.9.1 Average Intercorrelations and Reliability between Grade 9 ELA/L Subclaims

CBT								PBT							
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.90	4,140	4,140	4,140	4,140	4,140	4,140	RD							
RL	0.89	0.76	4,140	4,140	4,140	4,140	4,140	RL							
RI	0.95	0.76	0.82	4,140	4,140	4,140	4,140	RI							
RV	0.85	0.68	0.72	0.65	4,140	4,140	4,140	RV							
WR	0.74	0.71	0.71	0.58	0.85	4,140	4,140	WR							
WE	0.74	0.71	0.70	0.57	1.00	0.81	4,140	WE							
WKL	0.74	0.71	0.71	0.58	0.97	0.96	0.82	WKL							

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.2 Average Intercorrelations and Reliability between Grade 10 ELA/L Subclaims

CBT								PBT							
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.92	7,978	7,978	7,978	7,978	7,978	7,978	RD							
RL	0.90	0.76	7,978	7,978	7,978	7,978	7,978	RL							
RI	0.96	0.80	0.86	7,978	7,978	7,978	7,978	RI							
RV	0.87	0.71	0.76	0.68	7,978	7,978	7,978	RV							
WR	0.81	0.77	0.78	0.64	0.87	7,978	7,978	WR							
WE	0.80	0.77	0.77	0.63	1.00	0.86	7,978	WE							
WKL	0.80	0.76	0.78	0.64	0.97	0.95	0.88	WKL							

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.3 Average Intercorrelations and Reliability between Grade 11 ELA/L Subclaims

CBT								PBT							
	RD	RL	RI	RV	WR	WE	WKL		RD	RL	RI	RV	WR	WE	WKL
RD	0.91	12,502	12,502	12,502	12,502	12,502	12,502	RD	0.91	224	224	224	224	224	224
RL	0.94	0.84	12,502	12,502	12,502	12,502	12,502	RL	0.95	0.82	224	224	224	224	224
RI	0.91	0.77	0.78	12,502	12,502	12,502	12,502	RI	0.91	0.79	0.80	224	224	224	224
RV	0.85	0.73	0.69	0.66	12,502	12,502	12,502	RV	0.80	0.71	0.59	0.59	224	224	224
WR	0.83	0.79	0.78	0.66	0.87	12,502	12,502	WR	0.84	0.80	0.78	0.63	0.86	224	224
WE	0.82	0.79	0.77	0.65	1.00	0.85	12,502	WE	0.84	0.80	0.79	0.63	1.00	0.85	224
WKL	0.81	0.78	0.76	0.65	0.97	0.95	0.86	WKL	0.81	0.78	0.76	0.60	0.98	0.96	0.85

Note: RD = Reading, RL = Reading Literature, RI = Reading Information, RV = Reading Vocabulary WR = Writing, WE = Written Expression, and WKL = Writing Knowledge and Conventions. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.4 Average Intercorrelations and Reliability between Algebra I Subclaims

CBT					PBT				
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.77	5,672	5,672	5,672	MC				
ASC	0.76	0.70	5,672	5,672	ASC				
MR	0.67	0.63	0.55	5,672	MR				
MP	0.58	0.55	0.56	0.50	MP				

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average intercorrelations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.5 Average Inter correlations and Reliability between Geometry Subclaims

CBT					PBT				
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.78	6,089	6,089	6,089	MC	0.82	120	120	120
ASC	0.76	0.69	6,089	6,089	ASC	0.79	0.71	120	120
MR	0.57	0.53	0.42	6,089	MR	0.59	0.48	0.42	120
MP	0.66	0.62	0.55	0.52	MP	0.65	0.63	0.44	0.58

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average inter correlations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.6 Average Inter correlations and Reliability between Algebra II Subclaims

CBT					PBT				
	MC	ASC	MR	MP		MC	ASC	MR	MP
MC	0.76	12,139	12,139	12,139	MC				
ASC	0.75	0.78	12,139	12,139	ASC				
MR	0.72	0.72	0.62	12,139	MR				
MP	0.70	0.69	0.71	0.53	MP				

Note: MC = Major Content, ASC = Additional and Supporting Content, MR = Mathematical Reasoning, and MP = Modeling Practice. The shaded values along the diagonal are the reliabilities as reported in Section 8. The average inter correlations are provided in the lower portion of the table and the total sample sizes are provided in the upper portion of the table.

Table ADD.9.7 Average Correlations between ELA/L and Mathematics for High School

ELA/L	CBT		
	A1	GO	A2
9	0.73 (995)	0.71 (387)	
10	0.64 (430)	0.68 (1,037)	0.78 (935)
11	0.57 (231)	0.49 (1,339)	0.69 (4,093)

Note: ELA/L = English language arts/Literacy, A1 = Algebra I, GO = Geometry, A2 = Algebra II. The correlations are provided with the sample sizes, below in parentheses.

Table ADD.9.8 Average Correlations between Reading and Mathematics for High School

RD	CBT		
	A1	GO	A2
9	0.72 (995)	0.70 (387)	
10	0.62 (430)	0.68 (1,037)	0.77 (935)
11	0.54 (231)	0.49 (1,339)	0.67 (4,093)

Note: RD = Reading, A1 = Algebra I, GO = Geometry, A2 = Algebra II. The correlations are provided with the sample sizes, below in parentheses.

Table ADD.9.9 Average Correlations between Writing and Mathematics for High School

WR	CBT		
	A1	GO	A2
9	0.64 (995)	0.60 (387)	
10	0.56 (430)	0.56 (1,037)	0.70 (935)
11	0.48 (231)	0.38 (1,339)	0.60 (4,093)

Note: WR = Writing, A1 = Algebra I, GO = Geometry, A2 = Algebra II. The average correlations are provided with the sample sizes, below in parentheses.

Addendum 12: Scale Scores

Table ADD.12.1 Fall 2015 Subgroup Performance for ELA/L Scale Scores: Grade 9

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		4,204	741.63	34.99	650	850
Gender	Female	2,145	748.71	33.82	650	850
	Male	2,059	734.26	34.67	650	850
Ethnicity	American Indian/Alaska Native	27	726.52	29.56	654	805
	Asian	196	760.77	37.14	654	847
	Black or African American	972	728.78	31.69	650	844
	Hispanic/Latino	779	731.27	32.56	650	829
	Native Hawaiian or Pacific Islander	n/r	744.50	43.22	681	793
	Multiple Race Selected	130	732.58	33.18	650	815
	White	2,061	750.55	34.01	650	850
Economic Status*	Economically Disadvantaged	1,596	729.84	31.98	650	835
	Not Economically Disadvantaged	2,402	750.08	34.83	650	850
English Learner Status	English Learner (EL)	83	693.16	26.08	650	769
	Non English Learner	4,031	742.80	34.52	650	850
Disabilities	Students with Disabilities (SWD)	638	713.57	31.33	650	806
	Students without Disabilities	3,406	747.81	32.73	650	850
Reading Score		4,204	46.85	13.61	10	90
Gender	Female	2,145	48.94	13.23	10	90
	Male	2,059	44.68	13.66	10	90
Ethnicity	American Indian/Alaska Native	27	41.07	12.13	11	78
	Asian	196	53.50	14.26	14	85
	Black or African American	972	42.16	12.32	10	85
	Hispanic/Latino	779	42.96	12.44	10	78
	Native Hawaiian or Pacific Islander	n/r	46.50	14.80	26	68
	Multiple Race Selected	130	43.33	13.62	10	85
	White	2,061	50.20	13.42	10	90
Economic Status*	Economically Disadvantaged	1,596	42.38	12.28	10	85
	Not Economically Disadvantaged	2,402	49.96	13.65	10	90
English Learner Status	English Learner (EL)	83	28.75	9.88	10	56
	Non English Learner	4,031	47.27	13.45	10	90
Disabilities	Students with Disabilities (SWD)	638	36.65	12.13	10	72
	Students without Disabilities	3,406	49.10	12.88	10	90
Writing Score		4,204	32.28	9.54	15	60
Gender	Female	2,145	34.72	8.89	15	57
	Male	2,059	29.73	9.53	15	60

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Ethnicity	American Indian/Alaska Native	27	29.44	7.81	15	43
	Asian	196	37.46	9.41	15	57
	Black or African American	972	28.95	9.05	15	55
	Hispanic/Latino	779	29.78	9.26	15	57
	Native Hawaiian or Pacific Islander	n/r	34.38	13.16	15	53
	Multiple Race Selected	130	30.42	8.82	15	50
	White	2,061	34.45	9.16	15	60
Economic Status*	Economically Disadvantaged	1,596	29.39	9.27	15	57
	Not Economically Disadvantaged	2,402	34.42	9.27	15	60
English Learner Status	English Learner (EL)	83	20.65	7.24	15	40
	Non English Learner	4,031	32.59	9.43	15	60
Disabilities	Students with Disabilities (SWD)	638	24.83	8.94	15	53
	Students without Disabilities	3,406	33.91	8.90	15	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table ADD.12.2 Fall 2015 Subgroup Performance for ELA/L Scale Scores: Grade 10

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		8,101	735.49	45.28	650	850
Gender	Female	4,009	744.36	44.64	650	850
	Male	4,092	726.81	44.22	650	850
Ethnicity	American Indian/Alaska Native	52	721.85	37.59	660	808
	Asian	301	757.58	46.10	650	850
	Black or African American	1,655	711.83	39.78	650	850
	Hispanic/Latino	1,291	720.20	40.26	650	850
	Native Hawaiian or Pacific Islander	n/r	763.67	54.47	664	850
	Multiple Race Selected	203	737.05	43.95	650	849
	White	4,561	746.98	43.72	650	850
Economic Status*	Economically Disadvantaged	2,697	718.07	39.94	650	850
	Not Economically Disadvantaged	5,293	744.37	45.30	650	850
English Learner Status	English Learner (EL)	187	680.47	23.55	650	792
	Non English Learner	7,796	737.07	44.66	650	850
Disabilities	Students with Disabilities (SWD)	1,242	705.70	39.24	650	850
	Students without Disabilities	6,710	741.25	44.03	650	850
Reading Score		8,101	44.70	17.45	10	90
Gender	Female	4,009	47.23	17.08	10	90
	Male	4,092	42.23	17.44	10	90
Ethnicity	American Indian/Alaska Native	52	39.79	14.53	17	72
	Asian	301	52.03	17.60	10	90
	Black or African American	1,655	36.05	15.24	10	90
	Hispanic/Latino	1,291	38.88	15.43	10	90
	Native Hawaiian or Pacific Islander	n/r	56.00	21.59	19	90
	Multiple Race Selected	203	45.85	17.00	10	90
	White	4,561	48.98	17.08	10	90
Economic Status*	Economically Disadvantaged	2,697	38.18	15.37	10	90
	Not Economically Disadvantaged	5,293	48.02	17.52	10	90
English Learner Status	English Learner (EL)	187	24.26	9.36	10	64
	Non English Learner	7,796	45.28	17.24	10	90
Disabilities	Students with Disabilities (SWD)	1,242	34.00	15.67	10	90
	Students without Disabilities	6,710	46.76	16.99	10	90
Writing Score		8,101	30.00	13.04	10	60
Gender	Female	4,009	33.12	12.59	10	60
	Male	4,092	26.94	12.74	10	60
Ethnicity	American Indian/Alaska Native	52	26.90	10.62	10	49

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
	Asian	301	36.61	12.78	10	60
	Black or African American	1,655	23.49	12.07	10	58
	Hispanic/Latino	1,291	26.08	12.17	10	58
	Native Hawaiian or Pacific Islander	n/r	37.22	14.10	10	60
	Multiple Race Selected	203	29.67	13.01	10	58
	White	4,561	33.09	12.39	10	60
Economic Status*	Economically Disadvantaged	2,697	25.43	12.15	10	60
	Not Economically Disadvantaged	5,293	32.33	12.87	10	60
English Learner Status	English Learner (EL)	187	14.79	7.80	10	45
	Non English Learner	7,796	30.45	12.86	10	60
Disabilities	Students with Disabilities (SWD)	1,242	21.68	11.58	10	60
	Students without Disabilities	6,710	31.63	12.65	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table ADD.12.3 Fall 2015 Subgroup Performance for ELA/L Scale Scores: Grade 11

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		12,808	729.53	37.00	650	850
Gender	Female	5,961	736.60	37.39	650	850
	Male	6,847	723.37	35.52	650	850
Ethnicity	American Indian/Alaska Native	739	713.38	22.82	656	823
	Asian	306	750.54	41.54	656	850
	Black or African American	1,925	719.01	31.64	650	849
	Hispanic/Latino	4,229	716.57	29.63	650	850
	Native Hawaiian or Pacific Islander	n/r	729.00	37.49	680	794
	Multiple Race Selected	213	733.68	37.51	650	843
	White	5,179	744.85	39.26	650	850
Economic Status*	Economically Disadvantaged	5,398	716.83	29.45	650	850
	Not Economically Disadvantaged	6,793	738.83	39.53	650	850
English Learner Status	English Learner (EL)	921	700.82	19.62	650	814
	Non English Learner	11,189	731.68	37.22	650	850
Disabilities	Students with Disabilities (SWD)	2,299	708.00	30.35	650	837
	Students without Disabilities	9,879	734.24	36.73	650	850
Reading Score		12,808	43.42	14.17	10	90
Gender	Female	5,961	45.52	14.26	10	90
	Male	6,847	41.59	13.84	10	90
Ethnicity	American Indian/Alaska Native	739	37.09	8.64	15	78
	Asian	306	50.85	16.01	15	90
	Black or African American	1,925	39.56	12.20	10	86
	Hispanic/Latino	4,229	38.58	11.26	10	90
	Native Hawaiian or Pacific Islander	n/r	44.82	14.78	25	71
	Multiple Race Selected	213	45.28	14.14	12	82
	White	5,179	49.14	15.17	10	90
Economic Status*	Economically Disadvantaged	5,398	38.66	11.22	10	90
	Not Economically Disadvantaged	6,793	46.95	15.21	10	90
English Learner Status	English Learner (EL)	921	32.82	7.70	10	80
	Non English Learner	11,189	44.22	14.27	10	90
Disabilities	Students with Disabilities (SWD)	2,299	35.65	11.92	10	86
	Students without Disabilities	9,879	45.14	14.09	10	90
Writing Score		12,808	27.52	11.04	10	60
Gender	Female	5,961	30.12	10.95	10	60
	Male	6,847	25.25	10.62	10	60
Ethnicity	American Indian/Alaska Native	739	23.79	8.19	14	51

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
	Asian	306	33.55	11.62	14	60
	Black or African American	1,925	24.50	9.98	10	60
	Hispanic/Latino	4,229	24.03	9.57	10	60
	Native Hawaiian or Pacific Islander	n/r	25.00	11.42	14	42
	Multiple Race Selected	213	28.32	11.34	10	60
	White	5,179	31.63	11.34	10	60
Economic Status*	Economically Disadvantaged	5,398	24.12	9.56	10	60
	Not Economically Disadvantaged	6,793	29.95	11.46	10	60
English Learner Status	English Learner (EL)	921	19.52	7.14	10	45
	Non English Learner	11,189	28.11	11.07	10	60
Disabilities	Students with Disabilities (SWD)	2,299	21.29	9.22	10	60
	Students without Disabilities	9,879	28.86	10.94	10	60

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table ADD.12.4 Fall 2015 Subgroup Performance for Mathematics Scale Scores: Algebra I

Group Type	Group	N	Mean	SD	Min	Max
Full Summative Score		5,734	735.54	30.74	650	847
Gender	Female	2,921	737.83	29.58	650	821
	Male	2,813	733.16	31.74	650	847
Ethnicity	American Indian/Alaska Native	57	715.74	15.52	678	753
	Asian	180	758.11	32.45	650	841
	Black or African American	1,185	724.87	29.62	650	847
	Hispanic/Latino	1,173	727.29	28.12	650	814
	Native Hawaiian or Pacific Islander	n/r	747.23	29.56	687	775
	Multiple Race Selected	155	738.48	27.70	650	821
	White	2,925	742.03	30.07	650	830
Economic Status*	Economically Disadvantaged	2,195	726.12	27.78	650	815
	Not Economically Disadvantaged	3,450	741.75	31.15	650	847
English Learner Status	English Learner (EL)	167	712.09	23.99	650	821
	Non English Learner	5,473	736.50	30.68	650	847
Disabilities	Students with Disabilities (SWD)	758	715.97	27.46	650	807
	Students without Disabilities	4,869	739.01	30.07	650	847

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table ADD.12.5 Fall 2015 Subgroup Performance for Mathematics Scale Scores: Geometry

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		6,255	724.31	23.80	650	799
Gender	Female	3,181	725.65	23.35	650	798
	Male	3,074	722.92	24.18	650	799
Ethnicity	American Indian/Alaska Native	226	713.82	15.04	665	772
	Asian	242	744.38	28.14	652	799
	Black or African American	1,127	717.52	20.65	650	789
	Hispanic/Latino	1,881	715.83	20.21	650	794
	Native Hawaiian or Pacific Islander	n/r	717.40	31.37	689	758
	Multiple Race Selected	120	725.88	22.52	650	780
	White	2,584	732.53	23.69	650	797
Economic Status*	Economically Disadvantaged	2,907	717.31	20.82	650	792
	Not Economically Disadvantaged	3,184	730.69	24.47	650	799
English Learner Status	English Learner (EL)	356	706.30	17.05	650	773
	Non English Learner	5,750	725.47	23.53	650	799
Disabilities	Students with Disabilities (SWD)	1,115	710.33	20.18	650	797
	Students without Disabilities	4,961	727.53	23.32	650	799

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Table ADD.12.6 Fall 2015 Subgroup Performance for Mathematics Scale Scores: Algebra II

Group Type	Group	<i>N</i>	Mean	<i>SD</i>	Min	Max
Full Summative Score		12,327	710.38	34.85	650	847
Gender	Female	6,381	712.01	33.91	650	840
	Male	5,946	708.64	35.74	650	847
Ethnicity	American Indian/Alaska Native	748	697.11	22.42	650	794
	Asian	357	742.88	39.51	650	822
	Black or African American	1,506	695.58	29.04	650	840
	Hispanic/Latino	4,886	698.14	26.00	650	836
	Native Hawaiian or Pacific Islander	n/r	716.64	49.90	661	816
	Multiple Race Selected	197	717.22	32.76	650	804
	White	4,517	727.66	36.82	650	847
Economic Status*	Economically Disadvantaged	5,535	698.20	26.44	650	822
	Not Economically Disadvantaged	6,228	719.81	37.61	650	847
English Learner Status	English Learner (EL)	702	685.30	21.51	650	761
	Non English Learner	11,096	711.47	34.59	650	847
Disabilities	Students with Disabilities (SWD)	1,512	688.29	29.85	650	836
	Students without Disabilities	10,292	713.08	34.09	650	847

Note: *Economic status was based on participation in National School Lunch Program (NSLP): receipt of free or reduced-price lunch (FRL). n/r = not reported due to n<25.

Addendum 13: Inter-rater Agreement for Prose Constructed Response

This addendum presents the inter-rater agreement for operational results for the online PCR tasks by trait and grade level in spring 2016. For the 2016 administration, the scoring rubrics for the Literary Analysis tasks (LAT) and Research Simulation Tasks (RST) were updated to combine the Reading Comprehension and Written Expression traits. Therefore, the Prose Constructed Response (PCR) task items were scored on two traits instead of three: (1) Reading Comprehension and Written Expression and (2) Knowledge of Language and Conventions. Narrative Writing tasks (NWT) continued to be scored on the same two traits as in 2015: (1) Written Expression and (2) Knowledge of Language and Conventions.

For 10 percent of responses, a second “reliability” score was assigned. The purpose of the reliability score was to provide data for evaluating the consistency of scoring, which is done by evaluating scoring agreement. Inter-rater agreement is the agreement between the first and second scores assigned to student responses and is the measure of how often scorers agree with each other. Pearson scoring staff used inter-rater agreement indices as one factor in determining the needs for continuing training and intervention on both individual and group levels. PARCC inter-rater agreement expectations are provided in Table 4.4 in Section 4.2.4. For ELA/L PCR traits, the expectation for perfect agreement is an inter-rater agreement of 65% or higher between two scorers. When Intelligent Essay Assessor (IEA) provided the first score of record, the second reliability score was a human score. For those states choosing the human scoring option, the second reliability score was assigned by IEA. For a subset of responses, the first and second score were both human scores.

Table ADD.2.1 presents the average across the PCRs for each grade level by trait. The number of prompts included in the analyses is listed for each grade level. The agreement indices (perfect agreement, kappa, quadratic weighted kappa, and Pearson correlation) were calculated separately by PCR for each trait (Written Expression and Conventions). For each grade level, the agreement indices were averaged across the PCRs. The table presents the average count and the average for the agreement indices.

The exact agreement for the PCR traits is above the 65% agreement rate criteria for all PCRs except one ELA03 PCR, one ELA04 PCR, and one ELA11 PCR. The ELA03 PCR has an exact agreement of 61% for the Written Expression trait and 61% for the Conventions trait. The ELA04 PCR has an exact agreement of 64% for the Written Expression trait; however, the Conventions trait met the criteria with an exact agreement of 66%. The ELA11 PCR has an exact agreement of 63% for the Written Expression trait; however, the Conventions trait met the criteria with an exact agreement of 75%. The strength of agreement between raters is moderate to significant agreement as defined by Landis and Koch (1977) for all PCRs except one ELA03 PCR. The quadratic weighted kappa (Kappa QW) distinguishes between differences in ratings that are close to each other versus larger differences. The weighted kappa is substantial to almost perfect for all PCRs except one ELA03 PCR. The Pearson correlations (r) were relatively high for all PCRs except one ELA03 PCR.

During operational scoring, the PCR agreement rates are monitored for quality and items not meeting the criteria are shared with the PARCC handscoring operational working group. After the operational

administration, the performance of all the PCRs is provided to the content team as feedback for re-using PCRs and in order to inform development of future PCRs. This provides evidence for continuous improvement of the testing program.

Table ADD.2.1. PARCC PCR Average Agreement Indices by Test

Test	Number of PCRs	Count	Written Expression				Conventions			
			Exact	Kappa	QW Kappa	r	Exact	Kappa	QW Kappa	r
ELA03	6	29,665	69.26	0.47	0.66	0.67	67.90	0.50	0.72	0.72
ELA04	7	17,860	68.06	0.53	0.78	0.79	67.93	0.53	0.78	0.78
ELA05	7	18,573	72.54	0.58	0.80	0.80	71.23	0.58	0.81	0.81
ELA06	6	21,122	71.53	0.60	0.86	0.86	72.68	0.62	0.84	0.84
ELA07	6	22,593	72.75	0.63	0.89	0.89	73.81	0.64	0.87	0.87
ELA08	6	20,737	71.20	0.61	0.87	0.87	73.00	0.63	0.86	0.86
ELA09	6	17,035	73.81	0.63	0.87	0.87	75.37	0.65	0.86	0.86
ELA10	7	7,480	73.93	0.65	0.89	0.89	74.06	0.65	0.87	0.87
ELA11	6	8,019	71.27	0.61	0.87	0.87	75.42	0.66	0.86	0.86

Reference

Landis, J.R.; Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*. 33 (1): 159–174.