New York State Testing Program 2018: English Language Arts and Mathematics Grades 3–8

Technical Report

Questar Assessment Inc.
2018

Table of Contents

Section 1: Introduction and Overview 1
1.1. Introduction 1
1.2. Test Purpose 1
1.3. Expected Participants 1
1.4. Test Use and Decisions Based on Assessment 1
1.4.1. Scale Scores 1
1.4.2. Statewide Percentile Ranks 2
1.4.3. Performance Level Cut Scores and Classification 2
1.4.4. Subscores 2
1.5. Testing Accommodations 4
1.6. Test Transcriptions 4
1.7. Test Translations 4
Section 2: Test Design and Development 6
2.1. Test Descriptions 6
2.1.1. ELA Tests 6
2.1.2. Mathematics Tests 6
2.2. Test Configuration 7
2.2.1. Test Design 7
2.2.2. Embedded Field-Test Items 7
2.3. New York State Educators' Involvement in Test Development 8
2.4. Test Blueprints 8
2.5. Passage Selection and Item Criteria Documents 9
2.5.1. Principles of Universal Design 10
2.6. Passage Finding 11
2.7. Item Development 11
2.8. Educator Item Review 12
2.9. Field-Testing 13
2.10. Rangefinding 14
2.11. Item Selection and Test Creation (Criteria and Process) 14
2.12. Educator Form Construction 15
2.13. Test Form Production 16
2.14. Final Eyes Committees 16
2.15. Proficiency and Performance Standards 16
Section 3: Validity 18
3.1. Content Validity 18
3.2. Construct (Internal Structure) Validity 19
3.2.1. Internal Consistency 19
3.2.2. Unidimensionality 19
3.2.3. Detection of Bias 22
Section 4: Test Administration and Scoring 24

4.1. Test Administration 24
4.2. Scoring Procedures of Operational Tests 24
4.3. Scoring Models 25
4.4. Scoring of Constructed-Response Items 25
4.5. Scorer Qualifications and Training 26
4.6. Quality Control Process 26
Section 5: Operational Test Data Collection and Classical Analysis 27
5.1. Data Collection 27
5.2. Data Processing 27
5.2.1. Sampling Down for Representativeness 27
5.3. Classical Analysis and Calibration Sample Characteristics 33
5.4. Classical Data Analysis 43
5.4.1. Item Difficulty and Point Biserial Correlation Coefficients 44
5.4.2. Omit Rates 45
5.4.3. Differential Item Functioning (DIF) 45
Section 6: IRT Calibration 49
6.1. IRT Models and Rationale for Use 49
6.2. Calibration Sample 50
6.2.1. Calibration Process 56
6.3. Item-Model Fit 57
6.4. Local Independence 58
6.5. Scaling 59
6.6. Test Characteristic Curves 61
6.7. Scoring Procedure 74
6.7.1. Raw Score-to-Scale Score and SEM Conversion Tables 74
Section 7: Reliability and Standard Error of Measurement 77
7.1. Test Reliability 77
7.1.1. Test Statistics and Reliability for Total Test 77
7.1.2. Reliability of MC Items 79
7.1.3. Reliability of CR Items 79
7.1.4. Test Reliability for Subgroups 80
7.2. Standard Error of Measurement (SEM) 89
7.3. Performance Level Classification Consistency and Accuracy 90
7.3.1. Consistency 90
7.3.2. Accuracy 92
Section 8: Standards Review 94
Section 9: Summary of Operational Test Results 96
9.1. Scale Score Distribution Summary 96
9.1.1. ELA Scale Score and Subscore Distributions 96
9.1.1.1. ELA Grade 3 97
9.1.1.2. ELA Grade 4 98
9.1.1.3. ELA Grade 5 99

9.1.1.4. ELA Grade 6 100
9.1.1.5. ELA Grade 7 101
9.1.1.6. ELA Grade 8 102
9.1.2. Mathematics Scale Score Distributions 104
9.1.2.1. Mathematics Grade 3 105
9.1.2.2. Mathematics Grade 4 106
9.1.2.3. Mathematics Grade 5 107
9.1.2.4. Mathematics Grade 6 109
9.1.2.5. Mathematics Grade 7 110
9.1.2.6. Mathematics Grade 8 111
9.2. Performance Level Distribution Summary 113
9.2.1. ELA Test Performance Level Distributions 113
9.2.1.1. ELA Grade 3 114
9.2.1.2. ELA Grade 4 115
9.2.1.3. ELA Grade 5 116
9.2.1.4. ELA Grade 6 117
9.2.1.5. ELA Grade 7 118
9.2.1.6. ELA Grade 8 119
9.2.2. Mathematics Test Performance Level Distributions 120
9.2.2.1. Mathematics Grade 3 121
9.2.2.2. Mathematics Grade 4 122
9.2.2.3. Mathematics Grade 5 124
9.2.2.4. Mathematics Grade 6 125
9.2.2.5. Mathematics Grade 7 126
9.2.2.6. Mathematics Grade 8 127
Section 10: References 130
Appendix A: ELA and Mathematics Test Configurations 133
Appendix B: ELA and Mathematics Test Blueprints 136
Appendix C: Passage Selection Guidelines for Assessing ELA 138
Appendix D: Universal Design Item Checklist 139
Appendix E: Criteria for Item Acceptability 142
Appendix F: Psychometric Guidelines for Operational Item Selection 144
Appendix G: Operational Item Maps 145
Appendix H: ELA Short-Response Rubric 158
Appendix I: ELA Extended-Response Rubric 159
Appendix J: Mathematics Short-Response Rubric 162
Appendix K: Mathematics Extended-Response Rubric 163
Appendix L: Factor Analysis Results for Select Subgroups 164
Appendix M: Classical Test Theory Statistics 177
Appendix N: Items Flagged for DIF 190
Appendix O: IRT Statistics 195
Appendix P: Derivation and Estimation of Classification Consistency and Accuracy 220
Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables 223
Appendix R: Study of Operational Test Mode Comparability 266

Appendix S: Memo on Operational Test Mode Comparability 286
Appendix T: Standards Review Report 289

List of Tables

Table 1.1. ELA Subscore Categories and Total Possible Score Points 3
Table 1.2. Mathematics Subscore Categories and Total Possible Score Points 3
Table 2.1. Summary of Unique 2017 Field Test Items 13
Table 3.1. ELA Tests Factor Analysis 21
Table 3.2. Mathematics Tests Factor Analysis 22
Table 5.1. ELA Grade 3 Data Cleaning 28
Table 5.2. ELA Grade 4 Data Cleaning 28
Table 5.3. ELA Grade 5 Data Cleaning 29
Table 5.4. ELA Grade 6 Data Cleaning 29
Table 5.5. ELA Grade 7 Data Cleaning 30
Table 5.6. ELA Grade 8 Data Cleaning 30
Table 5.7. Mathematics Grade 3 Data Cleaning 30
Table 5.8. Mathematics Grade 4 Data Cleaning 31
Table 5.9. Mathematics Grade 5 Data Cleaning 31
Table 5.10. Mathematics Grade 6 Data Cleaning 32
Table 5.11. Mathematics Grade 7 Data Cleaning 32
Table 5.12. Mathematics Grade 8 Data Cleaning 33
Table 5.13. ELA Grade 3 Sample Characteristics 33
Table 5.14. ELA Grade 4 Sample Characteristics 34
Table 5.15. ELA Grade 5 Sample Characteristics 35
Table 5.16. ELA Grade 6 Sample Characteristics 36
Table 5.17. ELA Grade 7 Sample Characteristics 37
Table 5.18. ELA Grade 8 Sample Characteristics 37
Table 5.19. Mathematics Grade 3 Sample Characteristics 38
Table 5.20. Mathematics Grade 4 Sample Characteristics 39
Table 5.21. Mathematics Grade 5 Sample Characteristics 40
Table 5.22. Mathematics Grade 6 Sample Characteristics 41
Table 5.23. Mathematics Grade 7 Sample Characteristics 42
Table 5.24. Mathematics Grade 8 Sample Characteristics 43
Table 5.25. ELA Classical DIF Sample N-Counts 47
Table 5.26. Mathematics Classical DIF Sample N-Counts 47
Table 5.27. ELA Items Flagged for DIF 48
Table 5.28. Mathematics Items Flagged for DIF 48
Table 6.1. ELA Grades 3 and 4 Demographic Statistics 50
Table 6.2. ELA Grades 5 and 6 Demographic Statistics 51
Table 6.3. ELA Grades 7 and 8 Demographic Statistics 52
Table 6.4. Mathematics Grades 3 and 4 Demographic Statistics 53
Table 6.5. Mathematics Grades 5 and 6 Demographic Statistics 54
Table 6.6. Mathematics Grades 7 and 8 Demographic Statistics 55

Table 6.7. ELA Calibration Results 57
Table 6.8. Mathematics Calibration Results 57
Table 6.9. 2018 Operational Scaling Coefficients 61
Table 6.10. ELA Scale Score Ranges Associated with Each Performance Level 75
Table 6.11. Mathematics Scale Score Ranges Associated with Each Performance Level 76
Table 7.1. ELA Test Form Statistics 78
Table 7.2. ELA Test Reliability and Standard Error of Measurement 78
Table 7.3. Mathematics Test Form Statistics 78
Table 7.4. Mathematics Test Reliability and Standard Error of Measurement 78
Table 7.5. ELA MC Item Reliability and Standard Error of Measurement 79
Table 7.6. Mathematics MC Item Reliability and Standard Error of Measurement 79
Table 7.7. ELA CR Item Reliability and Standard Error of Measurement 80
Table 7.8. Mathematics CR Item Reliability and Standard Error of Measurement 80
Table 7.9. ELA Grade 3 Test Reliability by Subgroup 81
Table 7.10. ELA Grade 4 Test Reliability by Subgroup 81
Table 7.11. ELA Grade 5 Test Reliability by Subgroup 82
Table 7.12. ELA Grade 6 Test Reliability by Subgroup 83
Table 7.13. ELA Grade 7 Test Reliability by Subgroup 83
Table 7.14. ELA Grade 8 Test Reliability by Subgroup 84
Table 7.15. Mathematics Grade 3 Test Reliability by Subgroup 85
Table 7.16. Mathematics Grade 4 Test Reliability by Subgroup 86
Table 7.17. Mathematics Grade 5 Test Reliability by Subgroup 86
Table 7.18. Mathematics Grade 6 Test Reliability by Subgroup 87
Table 7.19. Mathematics Grade 7 Test Reliability by Subgroup 88
Table 7.20. Mathematics Grade 8 Test Reliability by Subgroup 89
Table 7.21. Decision Consistency (All Cuts)* 91
Table 7.22. Decision Consistency (Level III Cut)* 92
Table 7.23. Decision Agreement (Accuracy) Estimates* 93
Table 8.1. Recommended Cut Points for the English Language Arts Assessments 94
Table 8.2. Recommended Cut Points for the Mathematics Assessments 95
Table 9.1. ELA Scale Score Distribution Summary 96
Table 9.2. ELA Subscore Summary 97
Table 9.3. ELA Grade 3 Scale Score Distribution by Subgroup 97
Table 9.4. ELA Grade 4 Scale Score Distribution by Subgroup 99
Table 9.5. ELA Grade 5 Scale Score Distribution by Subgroup 100
Table 9.6. ELA Grade 6 Scale Score Distribution by Subgroup 101
Table 9.7. ELA Grade 7 Scale Score Distribution by Subgroup 102
Table 9.8. ELA Grade 8 Scale Score Distribution by Subgroup 103
Table 9.9. Mathematics Scale Score Distribution Summary 104
Table 9.10. Mathematics Subscore Summary 104
Table 9.11. Mathematics Grade 3 Scale Score Distribution by Subgroup 105

Table 9.12. Mathematics Grade 4 Scale Score Distribution by Subgroup 106
Table 9.13. Mathematics Grade 5 Scale Score Distribution by Subgroup 108
Table 9.14. Mathematics Grade 6 Scale Score Distribution by Subgroup 109
Table 9.15. Mathematics Grade 7 Scale Score Distribution by Subgroup 111
Table 9.16. Mathematics Grade 8 Scale Score Distribution by Subgroup 112
Table 9.17. ELA Test Performance Level Distributions 113
Table 9.18. ELA Grade 3 Performance Level Distribution by Subgroup 114
Table 9.19. ELA Grade 4 Performance Level Distribution by Subgroup 115
Table 9.20. ELA Grade 5 Performance Level Distribution by Subgroup 116
Table 9.21. ELA Grade 6 Performance Level Distribution by Subgroup 117
Table 9.22. ELA Grade 7 Performance Level Distribution by Subgroup 118
Table 9.23. ELA Grade 8 Performance Level Distribution by Subgroup 120
Table 9.24. Mathematics Test Performance Level Distributions 121
Table 9.25. Mathematics Grade 3 Performance Level Distribution by Subgroup..... 121
Table 9.26. Mathematics Grade 4 Performance Level Distribution by Subgroup..... 123
Table 9.27. Mathematics Grade 5 Performance Level Distribution by Subgroup..... 124
Table 9.28. Mathematics Grade 6 Performance Level Distribution by Subgroup..... 125
Table 9.29. Mathematics Grade 7 Performance Level Distribution by Subgroup..... 127
Table 9.30. Mathematics Grade 8 Performance Level Distribution by Subgroup..... 128
Table A1. ELA Test Configuration 133
Table A2. Mathematics Test Configuration..... 133
Table A3. ELA Estimated Time on Task by Session 134
Table A4. Mathematics Estimated Time on Task by Session 134
Table B1. ELA Test Blueprint 136
Table B2. Mathematics Test Blueprint 136
Table G1. ELA Grade 3 Operational Item Map 145
Table G2. ELA Grade 4 Operational Item Map 146
Table G3. ELA Grade 5 Operational Item Map 147
Table G4. ELA Grade 6 Operational Item Map 148
Table G5. ELA Grade 7 Operational Item Map 149
Table G6. ELA Grade 8 Operational Item Map 150
Table G7. Mathematics Grade 3 Operational Item Map..... 151
Table G8. Mathematics Grade 4 Operational Item Map..... 152
Table G9. Mathematics Grade 5 Operational Item Map..... 153
Table G10. Mathematics Grade 6 Operational Item Map..... 154
Table G11. Mathematics Grade 7 Operational Item Map..... 155
Table G12. Mathematics Grade 8 Operational Item Map..... 156
Table L1. ELA Grade 3 Test Factor Analysis by Subgroup..... 164
Table L2. ELA Grade 4 Test Factor Analysis by Subgroup..... 165
Table L3. ELA Grade 5 Test Factor Analysis by Subgroup..... 166
Table L4. ELA Grade 6 Test Factor Analysis by Subgroup..... 167

Table L5. ELA Grade 7 Test Factor Analysis by Subgroup 168
Table L6. ELA Grade 8 Test Factor Analysis by Subgroup 169
Table L7. Mathematics Grade 3 Test Factor Analysis by Subgroup 171
Table L8. Mathematics Grade 4 Test Factor Analysis by Subgroup 171
Table L9. Mathematics Grade 5 Test Factor Analysis by Subgroup 172
Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup 173
Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup 174
Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup 175
Table M1. ELA Grade 3 Classical Item Analysis 177
Table M2. ELA Grade 4 Classical Item Analysis 177
Table M3. ELA Grade 5 Classical Item Analysis 178
Table M4. ELA Grade 6 Classical Item Analysis 179
Table M5. ELA Grade 7 Classical Item Analysis 180
Table M6. ELA Grade 8 Classical Item Analysis 181
Table M7. Mathematics Grade 3 Classical Item Analysis 182
Table M8. Mathematics Grade 4 Classical Item Analysis 183
Table M9. Mathematics Grade 5 Classical Item Analysis 184
Table M10. Mathematics Grade 6 Classical Item Analysis 185
Table M11. Mathematics Grade 7 Classical Item Analysis 186
Table M12. Mathematics Grade 8 Classical Item Analysis 188
Table N1. ELA MC Item Classical DIF Flags 190
Table N2. ELA CR Item Classical DIF Flags 192
Table N3. Mathematics MC Item Classical DIF Flags 193
Table N4. Mathematics CR Item Classical DIF Flags 194
Table O1. ELA Grade 3 Item Fit Statistics 195
Table O2. ELA Grade 4 Item Fit Statistics 195
Table O3. ELA Grade 5 Item Fit Statistics 196
Table O4. ELA Grade 6 Item Fit Statistics 197
Table O5. ELA Grade 7 Item Fit Statistics 198
Table O6. ELA Grade 8 Item Fit Statistics 199
Table O7. Mathematics Grade 3 Item Fit Statistics 200
Table O8. Mathematics Grade 4 Item Fit Statistics 201
Table O9. Mathematics Grade 5 Item Fit Statistics 202
Table O10. Mathematics Grade 6 Item Fit Statistics 203
Table O11. Mathematics Grade 7 Item Fit Statistics 204
Table O12. Mathematics Grade 8 Item Fit Statistics 205
Table O13. ELA Grade 3 OP Item Parameter Estimates 207
Table O14. ELA Grade 4 OP Item Parameter Estimates 207
Table O15. ELA Grade 5 OP Item Parameter Estimates 208
Table O16. ELA Grade 6 OP Item Parameter Estimates 209
Table O17. ELA Grade 7 OP Item Parameter Estimates 210

Table O18. ELA Grade 8 OP Item Parameter Estimates 211
Table O19. Mathematics Grade 3 OP Item Parameter Estimates 212
Table O20. Mathematics Grade 4 OP Item Parameter Estimates 213
Table O21. Mathematics Grade 5 OP Item Parameter Estimates 214
Table O22. Mathematics Grade 6 OP Item Parameter Estimates 215
Table O23. Mathematics Grade 7 OP Item Parameter Estimates 217
Table O24. Mathematics Grade 8 OP Item Parameter Estimates 218
Table Q1. PBT ELA Grade 3 RSSS Table 223
Table Q2. PBT ELA Grade 4 RSSS Table 223
Table Q3. PBT ELA Grade 5 RSSS Table 224
Table Q4. PBT ELA Grade 6 RSSS Table 225
Table Q5. PBT ELA Grade 7 RSSS Table 225
Table Q6. PBT ELA Grade 8 RSSS Table 226
Table Q7. PBT Mathematics Grade 3 RSSS Table 227
Table Q8. PBT Mathematics Grade 4 RSSS Table 227
Table Q9. PBT Mathematics Grade 5 RSSS Table 228
Table Q10. PBT Mathematics Grade 6 RSSS Table 229
Table Q11. PBT Mathematics Grade 7 RSSS Table 229
Table Q12. PBT Mathematics Grade 8 RSSS Table 230
Table Q13. CBT ELA Grade 3 RSSS Table 231
Table Q14. CBT ELA Grade 4 RSSS Table 232
Table Q15. CBT ELA Grade 5 RSSS Table 232
Table Q16. CBT ELA Grade 6 RSSS Table 233
Table Q17. CBT ELA Grade 7 RSSS Table 234
Table Q18. CBT ELA Grade 8 RSSS Table 234
Table Q19. CBT Mathematics Grade 3 RSSS Table 235
Table Q20. CBT Mathematics Grade 4 RSSS Table 236
Table Q21. CBT Mathematics Grade 5 RSSS Table 236
Table Q22. CBT Mathematics Grade 6 RSSS Table 237
Table Q23. CBT Mathematics Grade 7 RSSS Table 238
Table Q24. CBT Mathematics Grade 8 RSSS Table 239
Table Q25. ELA Grade 3 Scale Score Frequency Distribution 239
Table Q26. ELA Grade 4 Scale Score Frequency Distribution 241
Table Q27. ELA Grade 5 Scale Score Frequency Distribution 243
Table Q28. ELA Grade 6 Scale Score Frequency Distribution 246
Table Q29. ELA Grade 7 Scale Score Frequency Distribution 248
Table Q30. ELA Grade 8 Scale Score Frequency Distribution 251
Table Q31. Mathematics Grade 3 Scale Score Frequency Distribution 254
Table Q32. Mathematics Grade 4 Scale Score Frequency Distribution 256
Table Q33. Mathematics Grade 5 Scale Score Frequency Distribution 258
Table Q34. Mathematics Grade 6 Scale Score Frequency Distribution 259

Table Q35. Mathematics Grade 7 Scale Score Frequency Distribution 262
Table Q36. Mathematics Grade 8 Scale Score Frequency Distribution 263
Table R.1.1. Operational Items Administered in Both CBT and PBT Modes..... 266
Table R.2.1. Sample Sizes Before Matching by Test Mode..... 267
Table R.3.1. Covariate Balance Before and After Matching: ELA Grade 4 271
Table R.3.2. Covariate Balance Before and After Matching: ELA Grade 5 272
Table R.3.3. Covariate Balance Before and After Matching: ELA Grade 6 273
Table R.3.4. Covariate Balance Before and After Matching: ELA Grade 7 274
Table R.3.5. Covariate Balance Before and After Matching: ELA Grade 8 275
Table R.3.6. Covariate Balance Before and After Matching: Mathematics Grade 4 276
Table R.3.7. Covariate Balance Before and After Matching: Mathematics Grade 5 277
Table R.3.8. Covariate Balance Before and After Matching: Mathematics Grade 6 278
Table R.3.9. Covariate Balance Before and After Matching: Mathematics Grade 7 279
Table R.3.10. Covariate Balance Before and After Matching: Mathematics Grade 8 280
Table R.3.11. Test-level Performance between Test Modes Before Matching – ELA 281
Table R.3.12. Test-level Performance between Test Modes After Matching – ELA..... 282
Table R.3.13. Test-level Performance between Test Modes Before Matching – Math..... 282
Table R.3.14. Test-level Performance between Test Modes After Matching – Math 283
Table R.3.15. Item-level Mode DIF Analysis Results..... 283
Table R.4.1. 2018 CBT Scale Score Adjustments 284

List of Figures

Figure 6.1. ELA Grade 3 TCC 62
Figure 6.2. ELA Grade 3 CSEM Curve 62
Figure 6.3. ELA Grade 4 TCC 63
Figure 6.4. ELA Grade 4 CSEM Curve 63
Figure 6.5. ELA Grade 5 TCC 64
Figure 6.6. ELA Grade 5 CSEM Curve 64
Figure 6.7. ELA Grade 6 TCC 65
Figure 6.8. ELA Grade 6 CSEM Curve 65
Figure 6.9. ELA Grade 7 TCC 66
Figure 6.10. ELA Grade 7 CSEM Curve 66
Figure 6.11. ELA Grade 8 TCC 67
Figure 6.12. ELA Grade 8 CSEM Curve 67
Figure 6.13. Mathematics Grade 3 TCC 68
Figure 6.14. Mathematics Grade 3 CSEM Curve 68
Figure 6.15. Mathematics Grade 4 TCC 69
Figure 6.16. Mathematics Grade 4 CSEM Curve 69
Figure 6.17. Mathematics Grade 5 TCC 70
Figure 6.18. Mathematics Grade 5 CSEM Curve 70
Figure 6.19. Mathematics Grade 6 TCC 71
Figure 6.20. Mathematics Grade 6 CSEM Curve 71
Figure 6.21. Mathematics Grade 7 TCC 72
Figure 6.22. Mathematics Grade 7 CSEM Curve 72
Figure 6.23. Mathematics Grade 8 TCC 73
Figure 6.24. Mathematics Grade 8 CSEM Curve 73

Section 1: Introduction and Overview

1.1. Introduction

This technical report provides detailed information regarding the technical, statistical, and measurement attributes of the New York State Testing Program (NYSTP) for the Grades 3–8 English Language Arts (ELA) and Mathematics 2018 Operational Tests. This report includes information about test content and test development, item (i.e., individual test question) and test statistics, validity and reliability, differential item functioning (DIF) studies, test administration, scoring, scaling, and student performance.

1.2. Test Purpose

The 2018 Grades 3–8 ELA and Mathematics NYSTP has been designed to measure student knowledge and skills as defined by grade-level New York State Learning Standards in ELA and Mathematics. The tests are designed to allow the classification of student proficiency into four performance levels (Level I, Level II, Level III, and Level IV). Likewise, the test provides students at each of these performance levels opportunities to demonstrate their knowledge and skills in the Learning Standards. Details about the content standards for ELA and Mathematics are described in Section 2.4: Test Blueprints.

1.3. Expected Participants

Students in New York State public school Grades 3, 4, 5, 6, 7, and 8 (and ungraded students of equivalent chronological ages) are the expected participants for the Grades 3–8 NYSTP. Religious and independent schools may participate in the testing program, but their participation is not mandatory. In 2018, some religious and independent schools participated in the testing program across all grade levels. These schools were included in the data analyses. Public school students were required to take all State assessments administered at their grade level, except for a very small percentage of students with severe cognitive disabilities who took the New York State Alternate Assessment (NYSAA). For more detail on this exemption, please refer to the *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (SAM), available online at http://www.p12.nysed.gov/assessment/sam/ei/eisam18b.pdf.

1.4. Test Use and Decisions Based on Assessment

The NYSTP Grades 3–8 ELA and Mathematics Tests are used to measure the extent to which individual students achieve the New York State Learning Standards in ELA and Mathematics, respectively, in order to determine whether schools, districts, and the State meet the required progress objectives specified in the New York State accountability system. Several types of scores are available from the Grades 3–8 ELA and Mathematics Tests, and they are discussed in this section.

1.4.1. *Scale Scores*

The scale scores are a quantification of the proficiency measured by the Grades 3–8 ELA and Mathematics Tests at each grade level. Scale scores are comparable only within a given subject and grade. Scale scores are not comparable across grades or across subjects. The scale scores are reported at the individual student level, and can be aggregated. Detailed information on the derivation and properties of the scale scores is provided in Section 6: IRT Calibration. The Grades 3–8 ELA and Mathematics Tests' scale scores are the basis for placing students into

performance levels, which are used to determine student progress within schools and districts; support registration of schools and districts; determine eligibility of students for additional educational services; and provide teachers with indicators of a student's need, or lack of need, for remediation in specific content-area knowledge.

1.4.2. *Statewide Percentile Ranks*

Students' scale scores are also presented as percentile ranks in order to indicate student performance relative to the entire testing population on a scale that may be more familiar than the operational test's scale. Such statistics are estimated based on how often each student earned a given scale score, thus presenting similar information as the scale score itself but on an alternate scale.

1.4.3. *Performance Level Cut Scores and Classification*

Student performance is classified as Level I, Level II, Level III, or Level IV for the Grades 3–8 ELA and Mathematics Tests. The definitions of performance levels are as follows:

- **NYS Level I:** Students performing at this level are well below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered insufficient for the expectations at this grade.

- **NYS Level II:** Students performing at this level are below proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered partial but insufficient for the expectations at this grade.

- **NYS Level III:** Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered sufficient for the expectations at this grade.

- **NYS Level IV:** Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered more than sufficient for the expectations at this grade.

The performance level cut scores used to distinguish between Levels I, II, III, and IV were originally established during the process of standard setting in Summer 2013. In July 2018, Questar hosted a standards review meeting to revisit and update the established cut scores given a test design change and a reduced test length in 2018 from 2017. The original standard setting process is described in detail in Section 8 and Appendix P in the 2013 technical report (NYSED, 2013). The *2018 Standards Review Report* is available in Appendix T.

1.4.4. *Subscores*

The Grades 3–8 ELA tests have two subscores: reading (which includes all multiple-choice items assessing both reading and language standards) and writing to sources (which includes all

constructed-response items assessing reading, writing, and language standards). The Grades 3–8 Mathematics tests have three subscores that are the domain-level scores for items measuring the *Major Clusters* in each grade. The New York State Learning Standards are divided into *Major*, *Supporting*, and *Additional Clusters*. Standards within *Major Clusters* are the intended focus of instruction and assessment and account for the majority of the Mathematics test items. The *Supporting* and *Additional Clusters* are Mathematics standards that both introduce and reinforce *Major Clusters*. Tables 1.1 and 1.2 present the reporting subscore categories and the point values that correspond to each on the 2018 tests. In 2018, subscores were reported in two ways:

1. A raw score (i.e., number of points earned) out of the total score on the test
2. The average score at the state level for each subscore category

Table 1.1. ELA Subscore Categories and Total Possible Score Points

	Total Subscore Points	
Grade	**Reading**	**Writing to Sources**
3	18	16
4	18	16
5	28	16
6	28	16
7	28	18
8	28	18

Table 1.2. Mathematics Subscore Categories and Total Possible Score Points

	Reporting Subscores and Total Subscore Points		
Grade	**Subscore 1**	**Subscore 2**	**Subscore 3**
3	Operations and Algebraic Thinking 13	Number and Operations—Fractions 9	Measurement and Data 7
4	Operations and Algebraic Thinking 7	Numbers and Operations in Base 10 6	Number and Operations—Fractions 13
5	Numbers and Operations in Base 10 10	Number and Operations—Fractions 14	Measurement and Data 14
6	Ratios and Proportional Relationships 6	The Number System 7	Expressions and Equations 15
7	Ratios and Proportional Relationships 7	The Number System 10	Expressions and Equations 13

Grade	Reporting Subscores and Total Subscore Points		
	Subscore 1	Subscore 2	Subscore 3
8	Expressions and Equations 16	Functions 13	Geometry 8

1.5. Testing Accommodations

In accordance with federal law under the Americans with Disabilities Act and the section Fairness in Testing and Test Use in the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 2014), accommodations that do not alter the measurement of any construct being tested are allowed for test takers. The allowance is in accordance with a student's Individualized Education Program (IEP) or Section 504 Accommodation Plan (504 Plan). School principals are responsible for ensuring that proper accommodations are provided when necessary, and that staff providing accommodations are properly trained. Details on testing accommodations can be found in the 2018 School Administrator's Manual (SAM).

1.6. Test Transcriptions

For visually impaired students, large-type and Braille editions of the test books are provided. In most cases, the students dictate and/or record their responses, the teachers transcribe student responses to the multiple-choice items onto scannable answer sheets, and the teachers transcribe the responses to the constructed-response items onto the regular test books. Some of the students who use large-type editions will fill in the answer sheets by themselves. The large-type editions are created by Questar Assessment Inc. and printed by SeaChange Print Innovations. SeeWriteHear, LLC, produced the Braille editions. SeeWriteHear employs certified Library of Congress Braille transcribers and delivers Braille in accordance with the Braille Authority of North America (BANA) standards. Camera-ready versions of the regular test books are provided to the Braille vendor, which then produces the Braille editions. Proofs of the Braille editions are submitted to NYSED for review and approval prior to production.

1.7. Test Translations

The NYSTP Grades 3–8 Mathematics Tests are translated into five languages: Chinese (Traditional), Haitian-Creole, Korean, Russian, and Spanish. These tests are translated to provide students the opportunity to demonstrate mathematical proficiency independent of their command of the English language. Sample tests are available in each translated language at the following location: http://www.p12.nysed.gov/assessment/math/samplers/.

English Language Learner/Multilingual Learner (ELL/MLL) students taking the Grades 3–8 Mathematics Tests may be provided with an oral translation of the test when a written translation is not available in the student's native language. The following testing accommodations are also made available to ELLs: separate testing location, bilingual glossaries, simultaneous use of English and alternative-language editions, oral translation for lower-incidence languages, and writing responses in the native language.

The NYSTP Grades 3–8 ELA Tests are not translated into any other language because they are assessments of proficiency in English language arts. The following testing accommodations are

made available to ELL/MLLs taking the ELA Tests: separate testing location and bilingual glossaries.

Section 2: Test Design and Development

2.1. Test Descriptions

The 2018 Grades 3–8 ELA and Mathematics Tests are criterion-referenced tests composed of multiple-choice (MC) and constructed-response (CR) test items based on the New York State P–12 Learning Standards. The tests were administered in New York State classrooms during a three-day period for paper-based tests, and a six day period for computer-based tests from April to May of 2018. Details on the administration and scoring of these tests can be found in Section 4: Test Administration and Scoring. Additional information can be found in the *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (SAM), available at http://www.p12.nysed.gov/assessment/sam/ei/eisam18b.pdf

2.1.1. *ELA Tests*

The 2018 Grade 3–8 ELA Tests were designed to measure student literacy as defined by the New York State Learning Standards. The tests assessed Reading, Writing, and Language standards by using multiple-choice, short-response, and extended-response items. All items were based on close readings of informational, literary, or paired texts. All texts were drawn from authentic, grade-level works.

Multiple-choice items were designed to assess Reading and Language Standards. Multiple-choice items required students to analyze different aspects of a given text, including central idea, style elements, character and plot development, and vocabulary.

Short-response items were designed to assess Reading and Language Standards. These were single items in which students used textual evidence to support their answers to inferential questions. These items asked students to make an inference, state a position, or draw a conclusion based on their analysis of the passage and then provide two pieces of text-based evidence to support their answers. In responding to these items, students were expected to write in complete sentences. Appendix H provides the rubric for the short-response items.

Extended-response items were designed to assess Reading, Writing, and Language Standards, with a focus primarily on the Writing Standard. Extended-response items required comprehension and analysis of either an individual text (Grades 3–8) or paired texts (Grades 4–8). Paired texts required students to read and analyze two related texts. Paired texts were related by theme, genre, tone, time period, or other characteristics. Many extended-response items asked students to express a position and support it with text-based evidence. For paired texts, students were expected to synthesize ideas between and draw evidence from both texts. Extended-response items required students to demonstrate their ability to write a coherent essay, using textual evidence to support their ideas. Appendix I provides the rubric for the extended-response items.

2.1.2. *Mathematics Tests*

The 2018 Grade 3–8 Mathematics Tests were designed to measure student mathematic understanding as defined by the New York State Learning Standards. The tests required that students understand Mathematics conceptually, use prerequisite skills with grade-level mathematical facts, decide which formulas and tools (e.g., protractors and rulers) to use, and solve mathematics problems rooted in the real world. The tests contained multiple-choice, short-

response (2-point), and extended-response (3-point) items. For multiple-choice items, students selected the correct response from four answer choices. For short- and extended-response items, students wrote an answer to an open-ended question. Some items required students to show their work or to explain, in words, how they arrived at their answers.

Mathematics multiple-choice items were used mainly to assess standard algorithms and conceptual standards. Multiple-choice items incorporated the New York State Learning Standards, some in real-world applications. Many multiple-choice items required students to complete multiple steps. Likewise, many of these items were linked to more than one standard, drawing on the simultaneous application of multiple skills and concepts.

Short-response items were used mainly to assess conceptual and application standards. The items required students to complete a task and show their work. Like multiple-choice items, short-response items often required multiple steps and the application of multiple mathematics skills, some in real-world applications. Appendix J provides the rubric for the Mathematics short-response items.

Extended-response items were used mainly to assess students' abilities to show their understanding of mathematical procedures, conceptual understanding, and application of those procedures and concepts. Extended-response items required students to complete two or more tasks, or a more extensive problem, and show their work. Some items also assessed student reasoning and the ability to critique the arguments of others. Appendix K provides the rubric for the Mathematics extended-response items.

2.2. Test Configuration

2.2.1. *Test Design*

The 2018 Grades 3–8 ELA Tests were composed of two sessions per grade and administered over two days. Each day consisted of one session. Session 1 contained literary and informational reading passages and MC items based on the passages. Session 2 contained only reading passages with short-response items and an extended-response item based on those passages.

The 2018 Grades 3–8 Mathematics Tests were composed of two sessions per grade and administered over two days. Each day consisted of one session: Session 1 contained MC items. Session 2 contained MC items as well as short- and extended-response items.

The tables in Appendix A provide information on the numbers and types of items in each session for the Grades 3–8 ELA and Mathematics Tests and the testing times.

2.2.2. *Embedded Field-Test Items*

In 2010, NYSED announced its commitment to embed multiple-choice items for field testing within the Spring 2012 Grades 3–8 ELA and Mathematics Operational Tests. This commitment continued for the Spring 2018 administrations of the tests. Embedding field-test items allows for a better representation of student responses and provides more reliable field-test data on which to build future operational tests. In other words, since the specific locations of the embedded field-test items were not disclosed and they look the same as operational test items, students were unable to differentiate field-test items from operational test items. Therefore, field-test data

derived from embedded items are free of the effects of differential student motivation that may characterize stand-alone field-test designs. Embedding field-test items also reduced the number of stand-alone field-test forms during Spring 2018, although it did not eliminate the need for them.

2.3. New York State Educators' Involvement in Test Development

New York State educators are actively involved in ELA and Mathematics test development. New York State educators provide critical input throughout all stages of the test development process, which include rangefinding, educator item review, operational forms construction, passage selection, item writing, and a *Final Eyes* meeting (a final review of the test materials prior to printing).

NYSED gathers a diverse group of educators to review all test materials, in order to create fair and valid tests. The participants are selected for each testing activity based on:

- Certification and appropriate grade-level experience
- Special population experience
- Geographical region
- Gender
- Ethnicity
- Type of school (urban, suburban, or rural)

The selected participants must be certified and have both teaching and testing experience. Most of the participants are classroom teachers. Specialists such as reading coaches, literacy coaches, and special education and bilingual instructors also participate. Some participants are also recommended by principals, professional organizations, Big Four Cities (i.e., Buffalo, Rochester, Syracuse, and Yonkers), and/or the Staff and Curriculum Development Network (SCDN). A file of participants is maintained and routinely updated with current participant information, as well as the addition of possible future participants as recruitment forms are received. The process of continually updating and adding to this file contributes to NYSED's ability to include many educators in the test development process. Every effort is made to have diverse groups of educators participate in each testing event.

Additionally, Content Advisory Panels (CAPs) meet quarterly to review, vet, and provide comments on curricular and assessment work. CAPs are content-area-specific advisory panels composed of between 15 and 20 New York State P–12 educators whose members are nominated by state professional organizations, institutes of higher education, and educator unions.

2.4. Test Blueprints

After careful consideration of test length and administration constraints (e.g., location of multiple-choice and constructed-response items within test sessions), the representation and distribution of content were determined.

The New York State Learning Standards for ELA are organized into four strands: Reading, Writing, Language, and Speaking/Listening. Due to administration constraints, Speaking/Listening was determined to be best assessed only in the classroom; therefore, the ELA

Tests assess three of the four strands: Reading, Writing, and Language. Content experts reviewed the Reading, Writing, and Language standards and recommended content coverage by standard and item type, based on the depth and breadth of each standard.

The New York State Learning Standards for Mathematics are divided into standards, clusters, and domains. Standards define what students should understand and be able to do and are further articulated into lettered components. Clusters are groups of related standards. Domains are larger groups of related clusters and standards. Content experts reviewed the Mathematics standards and recommended content coverage by standard and item type (MC or CR), based on the emphasis of the cluster (major, supporting, and additional) and depth and breadth of each standard.

Tables B1 and B2 in Appendix B show the test blueprint and actual number of score points in the Grades 3–8 ELA and Mathematics Tests, respectively. The tables include the ranges of allowable points for each ELA strand and Mathematics domain and the actual number of points on the 2018 operational tests.

2.5. Passage Selection and Item Criteria Documents

To guide test item development and to help ensure that New York State tests were measuring the Learning Standards for ELA and Mathematics with fidelity, criteria were established for selecting passages and writing test items, based on the consultation with the groups listed above.

The *Passage Selection Guidelines for Assessing State Standards ELA* were created to provide a framework that allows for the consistent selection of passages that are appropriately complex for the given grade and contain the specific characteristics necessary to measure different standards (see Appendix C). The guidelines describe the quantitative methods used to determine the grade appropriateness of a given text. They also describe the grade-specific text characteristics needed to develop items that measure any particular reading standard. The complete guidelines can be found here: http://www.engageny.org/sites/default/files/resource/attachments/passage_selection_guidelines_for_assessing_ccss_ela.pdf.

Passage Review Criteria documents were created based on the passage selection guidelines and were used to evaluate each potential passage and determine whether it could be used to measure the New York State Learning Standards for ELA. The criteria documents were used to determine whether each passage suggested for testing use was grade appropriate, fair, and possessed the necessary characteristics to assess each standard. Specifically, passages were evaluated for the presence and quality of key ideas and details, craft and structure, and integration of knowledge and ideas. The full passage review criteria can be found here: https://www.engageny.org/resource/new-york-state-passage-selection-resources-for-grade-3-8-assessments

Item Review Criteria for the Grade 3–8 ELA Tests were used to help ensure that each item was clear and fair, measured a specific standard or standards with fidelity, and conformed to the specifications for each item type. Each section of the criteria includes pertinent questions used to determine whether an item was of sufficient quality so that it could move forward in the development process. The first two of the *Item Review Criteria*, clarity and fairness, identify the basic components of quality items. The criteria for clarity are used to help ensure that students understand what is asked in each item and that the language choice in the item does not

negatively affect a student's ability to perform the required task. For example, the criteria include checking to make sure that the vocabulary of test items is at grade level and that items avoid technical terms unrelated to the content. Likewise, the fairness criteria are used to ensure that items are unbiased, non-offensive, and not disadvantageous to any given subgroup. The criteria also address how each item measures a given standard or standards and articulates the aspects of each standard that the items need to address. Finally, the criteria establish key requirements for each item type (e.g., requiring that each two-point constructed-response item asks students to make a clear statement that can be supported with two independent text-based pieces of evidence). The complete ELA criteria documents can be found here: http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-english-language-arts-tests.

Item Review Criteria for the Grade 3–8 Mathematics Tests were used to ensure clarity, language and graphical appropriateness, fairness, freedom from bias, fidelity of measurement to the New York State Learning Standards, and conformity to the expectations for specific item types and formats for each test item. Each section of the criteria includes pertinent questions that determine whether an item is of sufficient quality. The first two criteria, clarity and graphical appropriateness and fairness, identify the basic components of quality test items. The criteria for clarity and graphical appropriateness are used to help ensure that students understand what is asked in each item and that the language in the item does not adversely affect a student's ability to perform the required task. For example, the criteria include checking to make sure that the visual load for any item containing art is reasonable and that interpreting a graphic does not confuse the underlying construct. Likewise, the fairness criteria are used to evaluate whether or not items are unbiased, non-offensive, and not disadvantageous to any given subgroup. The criteria also require documentation of how each item measures the assigned Mathematics standard(s). Finally, the criteria address the specific demands for different item types and formats (making sure that each three-point constructed-response item involves a multi-step process and requires students to show work). The complete Mathematics criteria document can be found here: https://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-mathematics-tests.

The *Multiple Representations for NYS Grade 3–8 Mathematics Tests* document was developed to ensure that the tests measured the deep conceptual understanding that the New York State Learning Standards demand, rather than focusing on predictable Mathematics items that require only algorithmic strategies to be solved correctly. *Multiple Representations* is a broad set of specifications that describes, refers to, and symbolizes the various, but not all, ways that Mathematics standards could be measured within the constraints of the NYSTP. The document specifies three overarching families: procedural skills, conceptual understanding, and application. It also includes information about how to identify standards that might be measured through the use of a particular representation. It identifies types of Mathematics skills (e.g., application of process and explanation of a principle) that are appropriate for assessing different representations. The full document can be found here: https://www.engageny.org/resource/multiple-representations-for-nys-grade-3-8-common-core-mathematics-tests.

2.5.1. *Principles of Universal Design*

To create tests as equitable as possible for students, principles of Universal Design were employed during the creation of the tests and test items. In a report published by the National

Council on Educational Outcomes, "'Universally designed assessments' are designed and developed from the beginning to allow participation of the widest possible range of students, and to result in valid inferences about performance for all students who participate in the assessment" (Thompson, S.J., Johnstone, C.J., & Thurlow, M.L. 2002). The report goes on to describe seven elements of a universally designed assessment. These elements are:

1. Inclusive assessment population
2. Precisely defined constructs
3. Accessible, unbiased items
4. Amenable to accommodations
5. Simple, clear, and intuitive instructions and procedures
6. Maximum readability and comprehensibility
7. Maximum legibility

In accordance with these elements, the Universal Design Item Checklist in Appendix D was developed for use during item development.

2.6. Passage Finding

The goal of passage finding is to obtain high-quality texts from which to generate Learning Standards-aligned test items. To do so, in the 2016–2017 development cycle, independent passage finders were recruited and trained, using passage selection resources such as the passage selection criteria. Passage finders were given assignments based on the test blueprint requirements. Passage finders submitted passages along with completed criteria documents and source information to ELA content specialists, who reviewed the passages against the agreed-upon criteria. Passages that did not meet the criteria were rejected, and passages that did meet the criteria were moved forward in the process, where the text from scanned copies of the original sources was entered into templates. Once in the templates, readability metrics were determined for each text. Passages were then proofread by copyeditors, fact checked by research librarians, reviewed for content issues by Science and Social Studies content specialists, and reviewed for Universal Design issues by specifically trained reviewers. After the passages went through these review steps, ELA content specialists posted the passages and completed criteria documents for NYSED's review and approval for moving forward in the process.

NYSED staff retrieved and reviewed the passages and criteria documents. If NYSED staff determined that a passage did not meet the criteria, the passage was rejected and the NYSED staff provided an explanation for rejection.

In addition to the content reviews performed by NYSED staff and its vendors, executives in both organizations also reviewed the passages. The executive review focused on bias and sensitivity issues particular to New York State. Passages that passed both content and executive reviews were moved forward for item development.

2.7. Item Development

Item development for the 2018 test forms was conducted during the 2016–2017 development cycle. The goal of item development is to develop a sufficient number of high-quality, Learning Standards-aligned items to populate the test forms. Using the criteria documents for both content areas and the multiple-perspective document for Mathematics, content leads trained item writers.

The item writers had teaching or assessment experience in the content area for which they were writing items; experience in writing for large-scale, high-stakes assessments; and, at minimum, a bachelor's degree in either education and/or the content area for which they were assigned. The item writers were given specific assignments, based on the test blueprint. For ELA, the item writers were also provided with the completed passage criteria documents.

Item writers provided items and completed criteria documents to content specialists for review. Two content specialists reviewed each item and its corresponding criteria document. Items that did not meet the criteria were sent back to the writers with specific feedback for revision. Items that did not meet the criteria after an attempted revision were rejected and content specialists replaced them. After the content specialists were satisfied that all of the items met the criteria, the items were reviewed by copyeditors. The Mathematics items were also reviewed by content specialists in Science and Social Studies and by research librarians. The ELA and Mathematics content specialists evaluated the feedback from the different internal groups and edited the items accordingly. The items and criteria documents were then posted for NYSED's review and approval for moving forward in the process.

NYSED content experts retrieved and reviewed the items and criteria documents. If NYSED staff determined that an item did not meet the criteria, the item was rejected and the NYSED staff provided an explanation for rejection. Questar content specialists then replaced the item and completed criteria documents, which were resubmitted to NYSED. If NYSED staff determined that an item met the criteria but could be improved with editing, the staff member recorded notes for the edits. Those notes were reviewed at face-to-face meetings at which content staff and NYSED staff reviewed and edited all of the items to ensure that they met the criteria. All passages and items accepted at that meeting were moved forward for the educator item review.

2.8. Educator Item Review

After being reviewed by NYSED, the items were presented to panels of New York State educators. Based on their expertise, educators were assigned to grade-level and content-specific groups where they reviewed the items. The reviews were facilitated by Questar content specialists and were attended by NYSED staff. For ELA, reviewers first read and then discussed the passages before reviewing items. For Mathematics and ELA, the educators used the following checklist to review each item.

1. Does the item align to the designated standard(s)?
    - The item measures the content standard(s) that it was designed to measure.

2. Does the item meet quality standards?
    - The item is worded clearly.
    - The reading level of the item is grade appropriate.
    - The item has one correct answer.
    - The item has plausible, unambiguous distractors.
    - All of the distractors are mutually exclusive.

3. Is the item fair?
    - The item is free from bias on the basis of students' personal characteristics, such as gender or ethnicity.

As the educators reviewed the items, they discussed their judgments about them. If the educators felt that an item did not align to the standards, did not meet quality standards, or was not fair, they made recommendations for editing the item. NYSED staff and Questar content specialists later reviewed the recommendations and made the appropriate edits.

2.9. Field-Testing

Once the items have been developed and thoroughly reviewed by a variety of stakeholders, they must then be field-tested. Field-testing items is a critically important step in the test development process, as it is only through the gathering of actual student response data that a variety of psychometric characteristics may be evaluated. Table 2.1 provides a summary of the unique items that passed the scrutiny of NYSED and Questar content specialists, as well as that of New York State educators, and were field-tested. More items were field-tested than were needed on the operational forms because that enabled tests to be constructed with items that include the best possible characteristics from both a content and psychometric perspective.

Table 2.1. Summary of Unique 2017 Field Test Items

	Unique ELA Items by Type		Unique Mathematics Items by Type*	
Grade	**MC**	**CR**	**MC**	**CR**
3	84	42	89	25
4	88	42	91	25
5	121	48	91	25
6	139	46	90	25
7	139	46	83	24
8	136	38	79	25

Note. MC = multiple-choice. CR = constructed-response. All CR items were field-tested under stand-alone conditions, while nearly all MC items were administered under the embedded condition only. Twelve MC items were field tested for Math Grade 6 in the stand alone condition with CR items.

Multiple-choice field-test items were administered in Spring 2017 as embedded field-test items within the 2017 operational test forms. A majority of MC items on the forms were embedded field test items; stand alone field-test items were mostly CR items. The use of embedded field-test items yields more reliable field-test data and has nearly eliminated the need for multiple-choice stand-alone field-testing. One additional round of field-testing was administered separately from the 2017 operational forms (i.e., as stand-alone tests) later in Spring 2017, which included CR items and a minimal number of MC items.

A variety of analyses were conducted in order to better understand how the 2017 field-test items may perform on future operational forms. All of the field-test data underwent a series of representativeness checks. Because only a small sample of schools participate for any given content area and grade for stand-alone field-testing, it was necessary to ensure that the stand-

alone field-test samples were representative of the entire State population in terms of student achievement on prior years' tests, student gender, student ethnicity, and school Needs/Resource Capacity Category (NRC). Finally, a variety of psychometric analyses were conducted, including classical item analysis, inter-rater reliability for constructed-response items, differential item functioning (DIF), item response theory (IRT), item calibration, scaling, and fit evaluation. Many of these analyses are described at length below. However, inter-rater reliability analyses were not possible for the operational test, as only a single rater scored each constructed-response.

2.10. Rangefinding

Questar conducted rangefinding for most items included on the 2018 test. Rangefinding occurs after constructed-response items have been field-tested. The purpose of rangefinding is to have New York State educators review student constructed-responses and arrive at consensus scores based on the standards established by NYSED and the scoring rubrics. The consensus scores become the basis for operational rating guides and scoring ancillaries. To arrive at consensus, committees of New York State educators review, discuss, and rate student responses to the constructed-response field-test items. NYSED content experts and Questar Scoring Directors oversaw this process. The first step in the rangefinding process was to have the educator committees review rubrics and a NYSED-approved grounding guide set, previously used for the 2017 field-test rangefinding sessions, to familiarize teachers with the application of NYSED standards and rubrics. The grounding guide sets contain student responses that illustrate the full range of scores on the rubric. The grounding guide sets are composed of student responses that had previously gone through the rangefinding process and been approved by NYSED, and are used to guide the scoring of field-test and operational student responses. Referencing the previously approved guide set papers during the rangefinding sessions ensures consistency in the application of NYSED standards and rubrics from year-to-year.

After the committee reviewed the pre-approved grounding guide set, groups of committee members familiarized themselves with each item type, scoring a small number of responses representative of each of the different score points. After the group-scoring exercise, committee members independently scored other student responses. The committee then reviewed and discussed their results and determined consensus scores for the responses. The rangefinding results were used to build training materials for Questar scorers, who scored the field-test responses to constructed-response items.

2.11. Item Selection and Test Creation (Criteria and Process)

The NYSTP Grades 3–8 ELA and Mathematics Tests were administered from April to May of 2018. The test items were selected from the pools of available ELA and Mathematics items. These items were field-tested either in embedded field-testing or stand-alone field-testing from 2013 through 2017.

The test construction process involved several iterative steps. Three criteria governed the item selection process:

- Meet the ELA and Mathematics content specifications provided by NYSED
- Select items with the best psychometric characteristics from the ELA and Mathematics item pools

- Combine psychometric characteristics of all selected items with the intended psychometric goals for each entire form

Questar content specialists were provided the test designs, blueprints, and psychometric guidelines for item selection. The psychometric guidelines were based on the classical and IRT statistics associated with the test items.

Using the pool of field-tested items, Questar content specialists made preliminary selections for each grade and content area. The selections were then reviewed by the content leads for each content area, to make sure that the items conformed to the different criteria. If the content criteria were not met, new items were selected. After the content leads' review, the item selections were reviewed by Questar psychometricians. If items with undesirable statistics were selected, the psychometricians proposed items with more desirable statistics. The content specialists and their leads then reviewed those items. Once the Questar content teams and the psychometric teams were satisfied that the content and statistics of the selected items and the proposed whole forms met the requirements, the items were given to NYSED staff (including content and assessment experts) to review. Questar content specialists and psychometricians traveled to Albany, New York, in November 2017 to finalize item selection and test creation with NYSED staff (including content and assessment experts) and New York State educators.

2.12. Educator Form Construction

During an educator form construction meeting that took place from November 6–10, 2017 in Saratoga Springs, New York, educators from around the State worked with NYSED and Questar to review the content of the proposed 2018 operational ELA passages, and ELA and Mathematics individual test items. They looked at how those items combine to create entire operational forms, and for quality and appropriateness using their subject matter expertise. The goal was to ensure that all test items and forms are defensible from content and psychometric perspectives. The outcome was test forms that meet psychometric parameters and contain items that meet content criteria.

A different group of educators participated in the review of each subject and grade's test form, so each morning began with training in each room. Once training was complete, participants began the form construction process by independently evaluating the items and passages (for ELA) against the criteria on the provided checklists. Each participant completed his or her own checklist and had access to Questar's Content Management System which displayed the items corresponding to the order of items in the test.

- For ELA, the educators initially reviewed the first passage and a single item from the passage. Once they got used to the process, the educators reviewed the passages and the corresponding items. During this review, educators confirmed that there was only one correct answer for each multiple-choice item, and that the item was aligned to the standard that it purported to address. They also estimated the time that it would take students to read the passage and answer the items.
- For Mathematics, the educators initially reviewed single items and discussed each item as a group. Once they got used to the process, the educators reviewed groups of items (e.g., 4 to 6 items, followed by discussion of each item). During this review, educators

confirmed that there was only one correct answer for each multiple-choice item, and that the item was aligned to the standard that it purported to address. They also estimated the time that it would take students to answer the items.

In both ELA and Mathematics, the educators, in consultation with NYSED and Questar content experts, were permitted to recommend:

- revisions to the stated standard alignment;
- revisions to item sequencing to avoid cueing/clueing; and
- swapping any items and/or passages that they judged as having problems flagged by the above reviews.

Given other constraints, it was not always possible to make every change that educators recommended, but they were given the opportunity to voice any and all concerns that they had and NYSED made the final decision about any educator recommendations.

The facilitators then led a group discussion and helped the group reach consensus. Where time permitted, educators were presented with and approved the items that Questar and NYSED proposed for any necessary replacements. Following each session with educators, NYSED and Questar met to review the content and data of the proposed selections, and explore alternate selections for consideration. NYSED then approved the item selections, including item positions within test books.

2.13. Test Form Production

Once the selection of items for the operational and embedded field-test positions was completed, Questar created test forms. The test forms were reviewed by Questar content specialists and were posted for NYSED to review. NYSED and Questar reviewed the forms to look for any errors in spelling, capitalization, punctuation, grammar, and formatting. They also confirmed that each multiple-choice item had a single correct answer.

2.14. Final Eyes Committees

After NYSED and Questar reviewed copies of the test forms, the test forms were reviewed by the Final Eyes committees. For each content area, the committee consisted of thirty New York State educators from around the State. During that review, the educators were charged with taking the test to make sure that each multiple-choice item had a single correct answer, and to look for errors in spelling, capitalization, punctuation, grammar, and formatting.

After the Final Eyes review and after NYSED approved edits made as a result of the review, the tests were then considered final and produced for the 2018 administration.

2.15. Proficiency and Performance Standards

In July 2018, a standards review meeting occurred in Albany where 56 New York State educators went through a rigorous process, guided by the best practices indicated by this intensely studied process, to recommend updated performance standards. These recommendations were presented to the Commissioner, who, in turn, adopted the recommended

standards set forth by the committees. For additional details on the standards review process, see Appendix T.

Each grade level has four performance levels. Three cut points demarcate the performance levels needed to demonstrate each ascending level of performance. Section 6.7.1 contains the raw score-to-scale score and SEM conversion tables and detailed information related to the performance standards.

Section 3: Validity

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests. Test validation is an ongoing process of gathering evidence from many sources to evaluate the soundness of the desired score interpretation or use. This evidence is acquired from studies of the content of the test and studies involving scores produced by the test. Additionally, reliability has to be considered before considerations of validity are made. A test cannot be valid if the test scores are not first reliable.

The *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 2014) addressed the concept of validity in testing, which refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Validity is the most important consideration in test evaluation. Test validation is the process of accumulating evidence to support any particular inference. Validity, however, is a unitary concept. Although evidence may be accumulated in many ways, validity refers to the degree to which evidence supports the inferences made from test scores.

3.1. Content Validity

Generally, achievement tests are used for student-level outcomes, either for making predictions about students or for describing students' performances (Mehrens and Lehmann, 1991). Tests are now also used for the purposes of accountability and adequate yearly progress (AYP). The NYSED uses various assessment data in reporting AYP. Specific to student-level outcomes, the NYSTP documents student performance in the area of Mathematics as defined by the New York State Mathematics Learning Standards and in the area of ELA as defined by the New York State ELA Learning Standards.

To allow test score interpretations appropriate for this purpose, the content of the test must be carefully matched to the specified standards. The 2014 AERA/APA/NCME standards state that content-related evidence of validity is a central concern during test development. Expert professional judgment should play an integral part in developing the definition of what is to be measured, such as describing the universe of the content, generating or selecting the content sample, and specifying the item format and scoring system.

Expert analysis of test content indicates the degree to which the content of a test covers the domain of content that the test is intended to measure. In the case of the NYSTP, the content is defined by detailed blueprints that describe New York State content standards and define the skills that must be measured to assess these content standards (see Tables B1 and B2 in Appendix B). The NYSTP test development process requires specific attention to content representation and the balance within each test form. New York State educators were involved in test construction in various development stages. For example, during the item review process, they reviewed field-test items for the alignment of the items with the Learning Standards. Educators also participated in a process of establishing scoring rubrics for constructed-response items during rangefinding. Section 2: Test Design and Development contains more information specific to the item review process.

3.2. Construct (Internal Structure) Validity

Construct validity (i.e., what scores mean and what kind of inferences they support) is often considered the most important type of test validity. Construct validity of the NYSTP Grades 3–8 ELA and Mathematics Tests are supported by several types of evidence that can be obtained from the ELA and Mathematics test data.

3.2.1. *Internal Consistency*

Empirical studies of the internal structure of the test provide one type of evidence of construct validity. For example, high internal consistency constitutes evidence of validity. This is because high coefficients imply that the test items are measuring the same domain of skill and are reliable and consistent. Reliability coefficients of the tests for total populations and subgroups of students are presented in Section 7.1: Test Reliability. For the total population, the ELA reliability coefficients (Cronbach's alpha) ranged from 0.87 to 0.89. For all subgroups, the reliability coefficients were greater than or equal to 0.75. For the total population, the Mathematics reliability coefficients (Cronbach's alpha) ranged from 0.91 to 0.94. For all subgroups, the reliability coefficients were greater than or equal to 0.79. Overall, high internal consistency of the NYSTP Grades 3–8 ELA and Mathematics Tests provided sound evidence of construct validity.

3.2.2. *Unidimensionality*

Other validity evidence comes from analyses of the degree to which the test items conform to the requirements of the statistical models. These statistical models are used to scale and link the tests, as well as to generate student scores. The models require that the items fit the model well (item fit) and that the items in a test measure a single domain of skill (unidimensionality).

The first step is to assess the degree to which the items fit the IRT model. The item-model fit for the ELA and Mathematics tests was assessed using Q_1 statistics (Yen, 1981), and the results are described in detail in Section 6: IRT Calibration. Most items demonstrated sound fit across grades and content areas, and only a few items were deemed to have less than ideal fit. This provides solid evidence for the appropriateness of the IRT models used to calibrate and scale the test data.

Additional evidence for the efficacy of the model involves demonstrating that the items on the New York State tests are related to each other, within their respective content areas. This relationship of the items within the ELA or Mathematics tests is the common proficiency acquired by students studying the content area. This "common proficiency," or, more formally, underlying construct, could be labeled as ELA proficiency (using the ELA scores) or Mathematics proficiency (using the mathematics scores), depending on the degree to which the ELA and Mathematics items are related.

Factor analysis of the test data is one way of modeling the common construct. This analysis may show that there is a single or main factor that can account for much of the variability between responses to test items. A large first component in factor analysis would provide evidence of the latent proficiency that students have in common regarding the particular items asked. A large main factor found from a factor analysis of an achievement test would suggest a primary

construct that may be related to what the items were designed to have in common (i.e., Mathematics proficiency or ELA proficiency).

To demonstrate the common factor underlying student responses to the ELA and Mathematics test items, principal component factor analyses were conducted on a correlation matrix of individual items for the ELA and Mathematics tests. Factoring a correlation (i.e., tetrachoric correlation) matrix rather than actual item response data is preferable when dichotomous variables are in the analyzed data set. Because the ELA and Mathematics tests contain both multiple-choice and constructed-response items, the matrices of *polychoric* correlations were used as input for the factor analyses, as polychoric correlations are appropriate with both multiple-choice and constructed-response data. The study was conducted on the New York State public, charter, and religious and independent school students for whom data were available. A large first principal component was evident in each analysis, demonstrating essential unidimensionality of the trait (i.e., proficiency) measured by each test. In other words, statistical evidence indicates that the ELA items are measuring one underlying construct, ELA proficiency, and that the Mathematic items are measuring one underlying construct, Mathematics proficiency.

The factor analyses conducted with the ELA and Mathematics data will show almost as many underlying constructs, or factors, as there are items on the test. Therefore, it is necessary to investigate the factor analysis results further to determine the number of "meaningful" factors. Specifically, more than one factor with an eigenvalue greater than 1.0 present in each dataset would suggest the presence of small additional factors. The magnitude of the ratio of the variance accounted for by the first factor compared to the remaining factors also provides evidence as to the number of meaningful factors. In addition, the total amount of variance accounted for by the main factor was evaluated. According to M. Reckase (1979),

> . . . the 1PL and the 3PL models estimate different abilities when a test measures independent factors, but . . . both estimate the first principal component when it is large relative to the other factors. In this latter case, good ability estimates can be obtained from the models, even when the first factor accounts for less than 10 percent of the test variance, although item calibration results will be unstable. (p. 228)

Factor analyses related to the Grades 3–8 ELA and Mathematics Tests indicated that the ratio of the variance accounted for by the first factor to the remaining factors was sufficiently large to support the claim that the ELA and Mathematics tests were essentially unidimensional; the ELA-related ratios and the Mathematics-related ratios showed that the first eigenvalues were at least five times as large as the second eigenvalues for all of the grades.

All of the Grades 3–8 ELA and Mathematics Tests exhibited first principal component accounting for more than 20% and 25% of the test variance, respectively. Tables 3.1 and 3.2 present the results of factor analyses, including eigenvalues greater than 1.0 and proportions of variance explained by the extracted factors, for ELA and Mathematics, respectively.

The evidence in Table 3.1 supports the claim that one single construct underlies the items/tasks in each ELA test and that scores from each test would represent performance primarily determined by that construct. Construct-irrelevant variance does not appear to create significant nuisance factors. Similarly, Table 3.2 supports the claim that a common construct underlies the

items/tasks in each Mathematics test and that scores from each test would represent performance primarily determined by that construct. Construct-irrelevant variance does not appear to create significant nuisance factors.

Table 3.1. ELA Tests Factor Analysis

Grade	Extracted Factor			
		Initial	Variance Accounted for	
	#	Eigenvalue	%	Cumulative %
3	1	**6.19**	**24.77**	**24.77**
	2	1.53	6.12	30.90
	3	1.03	4.10	35.00
4	1	**6.24**	**24.95**	**24.95**
	2	1.36	5.44	30.39
	3	1.00	4.02	34.41
5	1	**7.05**	**20.14**	**20.14**
	2	1.55	4.43	24.57
	3	1.19	3.40	27.97
	4	1.08	3.07	31.04
	5	1.01	2.89	33.92
6	1	**7.87**	**22.49**	**22.49**
	2	1.39	3.98	26.47
	3	1.11	3.17	29.64
	4	1.05	3.00	32.64
7	1	**8.04**	**22.33**	**22.33**
	2	1.65	4.57	26.90
	3	1.11	3.09	29.99
	4	1.03	2.86	32.85
8	1	**7.87**	**21.87**	**21.87**
	2	1.50	4.16	26.03
	3	1.34	3.71	29.74
	4	1.02	2.85	32.59

Table 3.2. Mathematics Tests Factor Analysis

Grade	Extracted Factor			
		Initial	Variance Accounted for	
	#	Eigenvalue	%	Cumulative %
3	1	**9.40**	**27.64**	**27.64**
	2	1.54	4.53	32.17
	3	1.05	3.09	35.26
4	1	**11.06**	**29.09**	**29.09**
	2	1.35	3.55	32.64
5	1	**11.00**	**28.95**	**28.95**
	2	1.67	4.39	33.35
6	1	**11.35**	**29.10**	**29.10**
	2	1.58	4.04	33.14
	3	1.04	2.67	35.81
7	1	**12.08**	**29.45**	**29.45**
	2	1.40	3.41	32.86
	3	1.08	2.63	35.49
8	1	**10.34**	**25.21**	**25.21**
	2	1.28	3.12	28.33
	3	1.07	2.60	30.93
	4	1.00	2.45	33.38

As additional evidence for construct validity, the same factor analysis procedure was employed to assess the dimensionality of the Mathematics construct for selected subgroups of students in each grade: English language learners/multilingual learners (ELLs/MLLs), students with disabilities (SWD), and students using test accommodations (SUA), as well as ELL/MLL/SUA, and SWD/SUA. The ELL/MLL/SUA subgroup is defined as examinees who are ELLs/MLLs and who use at least one ELL/MLL-related accommodation. The SWD/SUA subgroup includes examinees who are classified as having disabilities and who use at least one disability-related accommodation. The results were comparable to the results obtained from the total population data. Evaluation of eigenvalue magnitude and proportions of variance explained by the main and secondary factors provide evidence of essential unidimensionality of the construct measured by the tests for the analyzed subgroups. Appendix L provides factor analysis results for ELL/MLL, SWD, SUA, ELL/MLL/SUA, and SWD/SUA classifications.

3.2.3. *Detection of Bias*

Minimizing item bias has the goal of minimizing construct-irrelevant variance and helps establish a strong validity argument for the tests. Specifically, bias occurs if items function differentially for key pairs of groups, which may, in turn, cause the test to be differentially valid for certain groups of test takers. The statistical means for flagging items that may exhibit bias is referred to as differential item functioning (DIF). These statistical procedures were designed to be conservative (i.e., they were designed to flag more items for DIF, rather than fewer).

Therefore, it is rare in practice to observe a high-stakes test in which not a single item is flagged for DIF. Since these procedures tend to over-flag items, it is only through review of those flagged items by experts that the items flagged for DIF may be judged to have or be free of bias. If the test involves irrelevant skills or knowledge, the possibility of bias is increased. Thus, preserving content validity is essential.

The developers of the NYSTP tests gave careful attention to items of possible ethnic, gender, socioeconomic status (SES), and—only for the Mathematics tests—translation bias. All materials were written and reviewed to conform to Questar's editorial policies and guidelines for equitable assessment, as well as NYSED's guidelines for item development. All materials were written to NYSED's specifications and carefully checked by groups of trained New York State educators during the item review process. These steps are essential in keeping bias to a minimum. However, current evidence suggests that expertise in this area is no substitute for data; reviewers are sometimes wrong about which items work to the disadvantage of a group, apparently because some of their ideas about how students will react to items may be faulty (Sandoval & Mille, 1979; Jensen, 1980). Thus, empirical studies were conducted.

Statistical methods were used to identify items exhibiting possible DIF. Although items flagged for DIF in the field-test stage were closely examined for content bias and avoided during the operational test construction, DIF analyses were conducted again on operational test data. Different methods were employed to evaluate the amount of DIF in all test items: constructed-response items were evaluated with standardized mean differences, and multiple-choice items were analyzed using Mantel-Haenszel methods (see Section 5: Operational Test Data Collection and Classical Analysis).

In each grade, for both ELA and Mathematics, few items were flagged for DIF. Moreover, the magnitude of DIF for the flagged items was typically small (for more details, see Appendix N). Multiple reviewers carefully reviewed items flagged for statistically significant DIF during the operational test item selection. All such items were deemed by the reviewers to be free of bias (i.e., judged not to adversely affect any demographic subgroup studied) and remained in the tests.

Section 4: Test Administration and Scoring

This section provides summaries of New York State test administration and scoring procedures. For further information, refer to the aforementioned *School Administrator's Manual* and the *New York State Scoring Leader Handbook (2018)* located here: http://www.p12.nysed.gov/assessment/sam/ei/scoringleaderhandbook18.pdf.

4.1. Test Administration

The NYSTP Grades 3–8 ELA and Mathematics Tests were administered to students in a paper-based (PBT) and computer-based (CBT) testing mode in 2018. The PBT testing window was Wednesday, April 11–Friday, April 13 for the Grades 3–8 ELA Tests and Tuesday, May 1–Thursday, May 3 for the Grades 3–8 Mathematics Tests. The CBT testing window was Tuesday, April 10–Tuesday, April 17 for the Grades 3–8 ELA Tests and Tuesday, May 1–Tuesday, May 8 for the Grades 3–8 Mathematics Tests.

The makeup test administration windows allowed students who were ill or otherwise unable to test during the assigned window to take the tests. The makeup test administration window for PBT was Monday, April 16–Wednesday, April 18 for the Grades 3–8 ELA Tests and Friday, May 4–Wednesday, May 9 for the Grades 3–8 Mathematics Tests. The makeup test administration window for CBT was Friday, April 13–Friday, April 20 for the Grades 3–8 ELA Tests and Friday, May 4–Friday, May 11 for the Grades 3–8 Mathematics Tests.

4.2. Scoring Procedures of Operational Tests

Qualified teachers and administrators performed the scoring of the NYSTP 2018 Grades 3–8 ELA and Mathematics Tests at designated sites. The number of personnel at a given site varied, as districts have the option of regional, district-wide, or school-wide scoring (please refer to Section 4.3: Scoring Models for more details). Administrators were responsible for the oversight of scoring operations, including the preparation of the test site, the security of test materials, and the supervision of the scoring process. At each site, designated trainers taught scoring committee members the basic criteria for scoring each item and monitored the scoring sessions in the room. Facilitators or leaders, who also helped in monitoring the sessions and enforced scoring accuracy, assisted the trainers.

The titles for administrators, trainers, and facilitators vary by the scoring model that is selected. At the regional level, a site coordinator conducted oversight. A scoring leader trained the scoring committee members and monitored the sessions, and a table facilitator assisted in monitoring the sessions. For each subject, the oversight was structured in the same way for district- and school-wide models. At the district-wide level, a school district administrator oversaw scoring. A district subject leader trained the scoring committee members and monitored the sessions, and a school subject leader assisted in monitoring the sessions. For school-wide scoring, oversight was provided by the principal; otherwise, titles for the school-wide model were the same as those for the district-wide model. The general title "scoring-committee members" included scorers at every site.

The process for PBT and CBT are the same excluding the following exceptions:

- For CBT, two schools within a district (Scoring Model 4) and one school (Scoring Model 5) are not permitted. Refer to page 15 of the *2018 Grades 3–8 English Language Arts*

and Mathematics Tests School Administrator's Manual for descriptions of all of the scoring models.

- For CBT, scorers use the ScorePoint system to score responses.

4.3. Scoring Models

For the 2017–2018 school year, schools and school districts were able to score Grades 3–8 ELA and/or Mathematics Tests regionally, multi-district, district-wide, or school-wide, based on local need. Schools were required to enter one of the following scoring model codes on student answer sheets:

1. Regional scoring—The scorers for the school's test papers included either staff from three or more school districts or staff from all religious and independent schools in an affiliation group (religious and independent or charter schools may participate in regional scoring with public school districts, and may be counted as one district).
2. Schools from two districts—The scorers for the school's test papers included staff from two school districts, religious and independent schools, charter school districts, or a combination thereof.
3. Three or more schools within a district—The scorers for the school's test papers included staff from all schools administering this test in a district, provided that at least three schools are represented.
4. Two schools within a district—The scorers for the school's test papers included staff from all schools administering this test in a district, provided that two schools are represented (not available for CBT schools).
5. One school, only (local scoring)—The first readers for the school's test papers included staff from the only school in the district administering this test, staff from one charter school, or staff from one religious and independent school (not available for CBT schools).
6. Private contractor—Scored by a private contractor that does not belong to Boards of Cooperative Educational Services (BOCES).

Schools and districts were instructed to carefully analyze their individual needs and capacities to determine their appropriate scoring model. BOCES and the Staff and Curriculum Development Network (SCDN) provided districts with technical support and advice in making this decision.

4.4. Scoring of Constructed-Response Items

The key resource used to train scoring committee members on how to score student responses for constructed response (CR) items were scoring guides. These guides were created by Questar from sets of actual field-test student responses that were consensus scored by NYSED and New York State teachers during Rangefinding sessions. Trainers used these materials to train scoring committee members on the criteria for scoring CR items and rubric application. Additionally, Scoring Leader Handbooks were distributed to provide guidelines, information, and procedures for both the Scorers and Scoring Site Coordinators to facilitate scoring.

Scoring for PBT responses was conducted using pen-and-pencil scoring. For these responses, scoring committee members evaluated the actual student papers rather than electronically scanned images. CBT responses were evaluated electronically.

For three distinct sections of the student tests, three separate scoring committee members scored each constructed response test session. After scoring was completed, the table facilitator or subject (ELA or Mathematics) leader conducted *read behinds* for the Scorers and items assigned to their scoring group.

4.5. Scorer Qualifications and Training

Qualified administrators and teachers conducted the scoring of the 2018 Grades 3–8 ELA and Mathematics Tests. Trainers used the scoring guides to train scoring-committee members on the criteria for scoring constructed-response items. Part of the training process was the administration of a consistency assurance set (CAS) that provided the State's scoring sites with information regarding strengths and weaknesses of their scorers. This tool allowed trainers to retrain their scorers, if necessary. The CAS also acknowledged those scorers who had grasped all aspects of the content area being scored and were well prepared to score student responses.

Regardless of the scoring model used, a minimum of three scorers is necessary to score each student's test. However, to comply with a State requirement, none of the scorers assigned to score a student's test responses may be that student's teacher. This policy is detailed in the *Scoring Leader Handbook* section "Assigning Scorer Numbers and Questions to Scoring Committee Members" on page 21, found online at: http://www.p12.nysed.gov/assessment/sam/ei/scoringleaderhandbook18.pdf.

4.6. Quality Control Process

Test books and electronic responses were randomly distributed throughout each scoring room so that completed tests from each region, district, school, or class were evenly dispersed. Teams were divided into groups of three, in order to ensure that a variety of scorers graded each test. If a scorer and a facilitator could not reach a decision after reviewing the scoring guides, they called the Questar Scoring Helpline. The call center was established to help teachers and administrators during scoring. The helpline staff consisted of trained Questar personnel, who answered questions by phone. When a member of the staff was unable to resolve an issue, it was referred to NYSED for a scoring decision. A quality check was also performed, in order to certify that all of the items were scored and that the scoring-committee members darkened each score on the answer document appropriately. The log of calls received by the scoring helpline was delivered to NYSED twice daily during the scoring window. To affirm that all schools across the state adhered to scoring guidelines and policies, approximately 5% of the schools' results are audited each year by an outside vendor.

Section 5: Operational Test Data Collection and Classical Analysis

5.1. Data Collection

Test data were collected in two phases. During Phase 1, a sample of approximately 95% of the student test records were received from the data warehouse and delivered to Questar, beginning at the end of May 2018. During Phase 2, "straggler files" were submitted to Questar in June 2018.

The "straggler files" contained fewer than about 5% of the total population cases, and were excluded from the classical, IRT, and reliability analyses (as described in Sections 5, 6, and 7, respectively) due to late submission. The analyses described in Section 8, "Summary of Operational Test Results," were based on the data collected from both Phase 1 and Phase 2. Data collected from both public schools and religious and independent schools were included in all data analyses.

5.2. Data Processing

Depending on the nature of the analysis, more student records were included in some analyses than in others. For example, all students with valid test scores were included in the analyses described in Section 8, "Summary of Operational Test Results." For the analyses described in other sections, more stringent data cleaning procedures were applied (see details below).

Data processing here refers to the cleaning and screening procedures used to identify errors (such as out-of-range data), and the decisions made to exclude student cases or to suppress particular items in certain analyses. Questar's psychometric team performed data cleaning to the delivered data, and excluded some student cases, in order to obtain a sample of the utmost integrity. It should be noted that a student case being excluded from certain data analyses did not mean that the student record was invalidated. According to the NYSED's specific instructions, additional procedures were taken to correct or recover these students' records so that their test results were scored properly. As mentioned above, their records were included in later analyses (see Section 8).

The major groups of cases excluded from the data set (used for analyses in Sections 5, 6, and 7) were students with missing school type and those with at least one entirely missing test session. Other deleted cases included students with incorrect or incomplete grade information, duplicate record cases, and no-response record cases. The mathematical data cleaning procedure also excluded records with mismatched form language indicators for translated versions across the two test sessions for a given student.

5.2.1. *Sampling Down for Representativeness*

Historically, after data cleaning, the sample is reviewed for representativeness of the prior year's operational population in terms of key variables such as student gender, racial/ethnic identity, student disability status, English Language Learner/Multilingual Learner (ELL/MLL) status, presence of test accommodation(s), and school Needs/Resource Capacity Category (NRC). At the recommendation of New York State's Assessment Technical Advisory Committee (TAC), Questar shifted the focus from sampling down according to demographic representativeness to instead focus on matching the prior year's population's distribution of ability. Questar and NYSED still reviewed the demographic patterns for 2018 relative to 2017, but they were not

used directly in the sampling down analyses. Comparison results between the final 2018 sample and 2017 operational population are further described in Section 6: IRT Calibration.

The numbers of cases considered for dropping because of sampling down varied across grades and subjects, but the process for all grades was consistent. The cleaned data file for a given subject and grade was the starting point. Questar reviewed the distribution of raw score proportion correct (RSPC) for the 2017 and 2018 operational forms. There were some minor differences in the 2017 and 2018 distributions of RSPC, but overall Questar, NYSED, and its TAC agreed that there was no evidence for a need to sample down in any subject or grade.

The data cleaning procedures and accompanying case counts are represented for ELA and Mathematics in Tables 5.1–5.6 and Tables 5.7–5.12, respectively.

Table 5.1. ELA Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	206,261
Wrong Subject	0	206,261
No Grade	95	206,166
Wrong Grade	42	206,124
Form Code Mismatch	957	205,167
Language or Mismatched Form	0	205,167
School Type	60	205,107
Missing Entire Session	25,732	179,375
Invalid Score	4	179,371
Not Tested Reason	0	179,371
Out-of-Range CR Scores	0	179,371
Duplicated Record	32	179,339
Test Mode Discrepancy	0	179,339

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (ie., refusal or absentee rates).

Table 5.2. ELA Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	213,397
Wrong Subject	0	213,397
No Grade	95	213,302
Wrong Grade	35	213,267
Form Code Mismatch	919	212,348
Language or Mismatched Form	0	212,348
School Type	39	212,309
Missing Entire Session	30,580	181,729
Invalid Score	7	181,722
Not Tested Reason	2	181,720

Exclusion Rule	# Deleted	# Cases Remain
Out-of-Range CR Scores	0	181,720
Duplicated Record	48	181,672
Test Mode Discrepancy	0	181,672

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (ie., refusal or absentee rates).

Table 5.3. ELA Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	211,154
Wrong Subject	0	211,154
No Grade	57	211,097
Wrong Grade	30	211,067
Form Code Mismatch	960	210,107
Language or Mismatched Form	0	210,107
School Type	3	210,104
Missing Entire Session	34,910	175,194
Invalid Score	6	175,188
Not Tested Reason	1	175,187
Out-of-Range CR Scores	0	175,187
Duplicated Record	12	175,175
Test Mode Discrepancy	0	175,175

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (ie., refusal or absentee rates).

Table 5.4. ELA Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	210,253
Wrong Subject	0	210,253
No Grade	76	210,177
Wrong Grade	40	210,137
Form Code Mismatch	1,206	208,931
Language or Mismatched Form	0	208,931
School Type	241	208,690
Missing Entire Session	38,646	170,044
Invalid Score	11	170,033
Not Tested Reason	5	170,028
Out-of-Range CR Scores	0	170,028
Duplicated Record	13	170,015
Test Mode Discrepancy	0	170,015

Table 5.5. ELA Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	201,093
Wrong Subject	0	201,093
No Grade	64	201,029
Wrong Grade	44	200,985
Form Code Mismatch	1,183	199,802
Language or Mismatched Form	0	199,802
School Type	254	199,548
Missing Entire Session	43,587	155,961
Invalid Score	12	155,949
Not Tested Reason	11	155,938
Out-of-Range CR Scores	0	155,938
Duplicated Record	19	155,919
Test Mode Discrepancy	0	155,919

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (ie., refusal or absentee rates).

Table 5.6. ELA Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	207,788
Wrong Subject	0	207,788
No Grade	54	207,734
Wrong Grade	53	207,681
Form Code Mismatch	1,141	206,540
Language or Mismatched Form	0	206,540
School Type	675	205,865
Missing Entire Session	54,317	151,548
Invalid Score	8	151,540
Not Tested Reason	10	151,530
Out-of-Range CR Scores	0	151,530
Duplicated Record	8	151,522
Test Mode Discrepancy	0	151,522

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (ie., refusal or absentee rates).

Table 5.7. Mathematics Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	201,839
Wrong Subject	0	201,839
No Grade	0	201,839
Wrong Grade	30	201,809

Exclusion Rule	# Deleted	# Cases Remain
Form Code Mismatch	2,829	198,980
Language or Mismatched Form	0	198,980
School Type	62	198,918
Missing Entire Session	22,218	176,700
Invalid Score	2	176,698
Not Tested Reason	3	176,695
Out-of-Range CR Scores	0	176,695
Duplicated Record	32	176,663
Test Mode Discrepancy	0	176,663

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (ie., refusal or absentee rates).

Table 5.8. Mathematics Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	207,119
Wrong Subject	0	207,119
No Grade	0	207,119
Wrong Grade	33	207,086
Form Code Mismatch	2,845	204,241
Language or Mismatched Form	0	204,241
School Type	40	204,201
Missing Entire Session	27,249	176,952
Invalid Score	4	176,948
Not Tested Reason	5	176,943
Out-of-Range CR Scores	0	176,943
Duplicated Record	46	176,897
Test Mode Discrepancy	0	176,897

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (ie., refusal or absentee rates).

Table 5.9. Mathematics Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	204,164
Wrong Subject	0	204,164
No Grade	0	204,164
Wrong Grade	19	204,145
Form Code Mismatch	3,528	200,617
Language or Mismatched Form	0	200,617
School Type	4	200,613
Missing Entire Session	32,016	168,597
Invalid Score	3	168,594

Exclusion Rule	# Deleted	# Cases Remain
Not Tested Reason	4	168,590
Out-of-Range CR Scores	0	168,590
Duplicated Record	12	168,578
Test Mode Discrepancy	0	168,578

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (ie., refusal or absentee rates).

Table 5.10. Mathematics Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	205,766
Wrong Subject	0	205,766
No Grade	1	205,765
Wrong Grade	35	205,730
Form Code Mismatch	3,326	202,404
Language or Mismatched Form	0	202,404
School Type	225	202,179
Missing Entire Session	37,706	164,473
Invalid Score	16	164,457
Not Tested Reason	14	164,443
Out-of-Range CR Scores	0	164,443
Duplicated Record	14	164,429
Test Mode Discrepancy	0	164,429

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (ie., refusal or absentee rates).

Table 5.11. Mathematics Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	199,412
Wrong Subject	0	199,412
No Grade	0	199,412
Wrong Grade	35	199,377
Form Code Mismatch	3,123	196,254
Language or Mismatched Form	0	196,254
School Type	150	196,104
Missing Entire Session	44,320	151,784
Invalid Score	3	151,781
Not Tested Reason	4	151,777
Out-of-Range CR Scores	0	151,777
Duplicated Record	28	151,749
Test Mode Discrepancy	0	151,749

Table 5.12. Mathematics Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	155,605
Wrong Subject	0	155,605
No Grade	0	155,605
Wrong Grade	36	155,569
Form Code Mismatch	2,695	152,874
Language or Mismatched Form	0	152,874
School Type	251	152,623
Missing Entire Session	44,188	108,435
Invalid Score	8	108,427
Not Tested Reason	5	108,422
Out-of-Range CR Scores	0	108,422
Duplicated Record	12	108,410
Test Mode Discrepancy	0	108,410

Note. The *Missing Entire Session* n-count includes students who did not participate in testing (ie., refusal or absentee rates).

5.3. Classical Analysis and Calibration Sample Characteristics

The cleaned and sampled-down (if needed) data were used for classical analyses and calibration. The demographic characteristics of students in these data sets are presented in Tables 5.13–5.18 and Tables 5.19–5.24 for ELA and Mathematics, respectively. The Needs/Resource Capacity Category (NRC) is assigned at the district level and is an indicator of district and school socioeconomic status. The ethnicity and gender designations are based on student-level information.

Table 5.13. ELA Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	88,724	49.47
	Male	90,615	50.53
Ethnicity	Asian	17,785	10.03
	Black	31,318	17.66
	Hispanic	50,296	28.37
	American Indian	1,244	0.70
	Multiracial	5,148	2.90
	Pacific Islander	422	0.24
	White	71,082	40.09
NRC	New York	66,509	37.09
	Big 4 Cities	7,735	4.31
	Urban/Suburban	14,213	7.93
	High Needs Rural	9,864	5.50
	Average Needs	42,241	23.55

Demographic Category		N-Count	% of Total N-Count
NRC	Low Needs	18,151	10.12
	Charter School	12,101	6.75
	Religious and Independent	8,525	4.75
SWD	No	154,197	85.98
	Yes	25,142	14.02
SUA	No	156,694	87.37
	Yes	22,645	12.63
ELL/ MLL	No	158,744	88.52
	Yes	20,595	11.48
SWD/ SUA	No	160,362	89.42
	Yes	18,977	10.58
ELL/ MLL/ SUA	No	175,577	97.9
	Yes	3,762	2.1

*The total n-count was 179,339.

Table 5.14. ELA Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	89,676	49.36
	Male	91,996	50.64
Ethnicity	Asian	18,533	10.32
	Black	32,133	17.89
	Hispanic	50,017	27.85
	American Indian	1,258	0.70
	Multiracial	4,731	2.63
	Pacific Islander	501	0.28
	White	72,411	40.32
NRC	New York	66,945	36.85
	Big 4 Cities	7,754	4.27
	Urban/Suburban	13,395	7.37
	High Needs Rural	9,820	5.41
	Average Needs	40,780	22.45
	Low Needs	18,128	9.98
	Charter School	11,288	6.21
	Religious and Independent	13,562	7.47
SWD	No	155,527	85.61
	Yes	26,145	14.39

Demographic Category		N-Count	% of Total N-Count
SUA	No	156,406	86.09
	Yes	25,266	13.91
ELL/ MLL	No	164,175	90.37
	Yes	17,497	9.63
SWD/ SUA	No	160,597	88.4
	Yes	21,075	11.6
ELL/ MLL/ SUA	No	177,980	97.97
	Yes	3,692	2.03

*The total n-count was 181,672.

Table 5.15. ELA Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	86,784	49.54
	Male	88,391	50.46
Ethnicity	Asian	18,643	10.76
	Black	31,523	18.19
	Hispanic	48,692	28.10
	American Indian	1,250	0.72
	Multiracial	4,256	2.46
	Pacific Islander	537	0.31
	White	68,391	39.47
NRC	New York	67,866	38.74
	Big 4 Cities	7,501	4.28
	Urban/Suburban	12,439	7.10
	High Needs Rural	9,295	5.31
	Average Needs	39,116	22.33
	Low Needs	18,282	10.44
	Charter School	11,148	6.36
	Religious and Independent	9,528	5.44
SWD	No	148,648	84.86
	Yes	26,527	15.14
SUA	No	149,255	85.20
	Yes	25,920	14.80
ELL/ MLL	No	160,524	91.64
	Yes	14,651	8.36
SWD/ SUA	No	153,494	87.62

Demographic Category		N-Count	% of Total N-Count
SWD/ SUA	Yes	21,681	12.38
ELL/ MLL/ SUA	No	171,590	97.95
	Yes	3,585	2.05

*The total n-count was 175,175.

Table 5.16. ELA Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	83,617	49.18
	Male	86,398	50.82
Ethnicity	Asian	18,003	10.72
	Black	31,314	18.65
	Hispanic	46,768	27.86
	American Indian	1,141	0.68
	Multiracial	3,714	2.21
	Pacific Islander	614	0.37
	White	66,335	39.51
NRC	New York	64,138	37.72
	Big 4 Cities	6,856	4.03
	Urban/Suburban	11,921	7.01
	High Needs Rural	8,994	5.29
	Average Needs	36,469	21.45
	Low Needs	17,522	10.31
	Charter School	11,389	6.70
	Religious and Independent	12,726	7.49
SWD	No	144,766	85.15
	Yes	25,249	14.85
SUA	No	146,038	85.90
	Yes	23,977	14.10
ELL/ MLL	No	156,512	92.06
	Yes	13,503	7.94
SWD/ SUA	No	150,214	88.35
	Yes	19,801	11.65
ELL/ MLL/ SUA	No	166,934	98.19
	Yes	3,081	1.81

*The total n-count was 170,015.

Table 5.17. ELA Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	75,962	48.72
	Male	79,957	51.28
Ethnicity	Asian	17,046	11.06
	Black	29,642	19.23
	Hispanic	42,405	27.51
	American Indian	1,209	0.78
	Multiracial	2,969	1.93
	Pacific Islander	461	0.30
	White	60,427	39.20
NRC	New York	64,280	41.23
	Big 4 Cities	6,366	4.08
	Urban/Suburban	10,852	6.96
	High Needs Rural	8,368	5.37
	Average Needs	32,952	21.13
	Low Needs	17,060	10.94
	Charter School	10,518	6.75
	Religious and Independent	5,523	3.54
SWD	No	131,616	84.41
	Yes	24,303	15.59
SUA	No	132,824	85.19
	Yes	23,095	14.81
ELL/ MLL	No	144,518	92.69
	Yes	11,401	7.31
SWD/ SUA	No	136,719	87.69
	Yes	19,200	12.31
ELL/ MLL/ SUA	No	153,372	98.37
	Yes	2,547	1.63

*The total n-count was 155,919.

Table 5.18. ELA Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	73,680	48.63
	Male	77,842	51.37
Ethnicity	Asian	17,516	11.67
	Black	29,158	19.43

Demographic Category		N-Count	% of Total N-Count
Ethnicity	Hispanic	41,041	27.35
	American Indian	1,169	0.78
	Multiracial	2,372	1.58
	Pacific Islander	456	0.30
	White	58,332	38.88
NRC	New York	62,273	41.10
	Big 4 Cities	6,205	4.10
	Urban/Suburban	9,428	6.22
	High Needs Rural	7,901	5.21
	Average Needs	29,532	19.49
	Low Needs	15,829	10.45
	Charter School	9,557	6.31
	Religious and Independent	10,797	7.13
SWD	No	129,070	85.18
	Yes	22,452	14.82
SUA	No	130,289	85.99
	Yes	21,233	14.01
ELL/MLL	No	140,581	92.78
	Yes	10,941	7.22
SWD/ SUA	No	133,852	88.34
	Yes	17,670	11.66
ELL/ MLL/ SUA	No	149,219	98.48
	Yes	2,303	1.52

*The total n-count was 151,522.

Table 5.19. Mathematics Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	87,245	49.38
	Male	89,418	50.62
Ethnicity	Asian	18,053	10.31
	Black	30,369	17.35
	Hispanic	50,760	28.99
	American Indian	1,228	0.70
	Multiracial	5,002	2.86
	Pacific Islander	422	0.24
	White	69,246	39.55
NRC	New York	67,143	38.01

Demographic Category		N-Count	% of Total N-Count
NRC	Big 4 Cities	6,623	3.75
	Urban/Suburban	14,296	8.09
	High Needs Rural	9,898	5.60
	Average Needs	42,171	23.87
	Low Needs	18,175	10.29
	Charter School	11,797	6.68
	Religious and Independent	6,560	3.71
SWD	No	152,704	86.44
	Yes	23,959	13.56
SUA	No	154,610	87.52
	Yes	22,053	12.48
ELL/ MLL	No	154,607	87.52
	Yes	22,056	12.48
SWD/ SUA	No	157,977	89.42
	Yes	18,686	10.58
ELL/ MLL/ SUA	No	172,673	97.74
	Yes	3,990	2.26

*The total n-count was 176,663.

Table 5.20. Mathematics Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	86,958	49.16
	Male	89,939	50.84
Ethnicity	Asian	18,806	10.73
	Black	30,994	17.69
	Hispanic	50,271	28.69
	American Indian	1,173	0.67
	Multiracial	4,664	2.66
	Pacific Islander	497	0.28
	White	68,820	39.28
NRC	New York	67,318	38.05
	Big 4 Cities	6,446	3.64
	Urban/Suburban	13,846	7.83
	High Needs Rural	9,841	5.56
	Average Needs	40,750	23.04
	Low Needs	18,169	10.27
	Charter School	10,942	6.19

Demographic Category		N-Count	% of Total N-Count
NRC	Religious and Independent	9,585	5.42
SWD	No	152,114	85.99
	Yes	24,783	14.01
SUA	No	152,611	86.27
	Yes	24,286	13.73
ELL/ MLL	No	158,449	89.57
	Yes	18,448	10.43
SWD/ SUA	No	156,385	88.4
	Yes	20,512	11.6
ELL/ MLL/ SUA	No	173,046	97.82
	Yes	3,851	2.18

*The total n-count was 176,897.

Table 5.21. Mathematics Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	83,182	49.34
	Male	85,396	50.66
Ethnicity	Asian	18,740	11.19
	Black	29,957	17.89
	Hispanic	48,437	28.93
	American Indian	1,223	0.73
	Multiracial	4,069	2.43
	Pacific Islander	544	0.32
	White	64,450	38.50
NRC	New York	67,758	40.19
	Big 4 Cities	5,963	3.54
	Urban/Suburban	12,928	7.67
	High Needs Rural	9,148	5.43
	Average Needs	38,460	22.81
	Low Needs	18,029	10.69
	Charter School	10,601	6.29
	Religious and Independent	5,691	3.38
SWD	No	143,903	85.36
	Yes	24,675	14.64
SUA	No	144,853	85.93
	Yes	23,725	14.07

Demographic Category		N-Count	% of Total N-Count
ELL/ MLL	No	153,662	91.15
	Yes	14,916	8.85
SWD/ SUA	No	147,989	87.79
	Yes	20,589	12.21
ELL/ MLL/ SUA	No	164,974	97.86
	Yes	3,604	2.14

*The total n-count was 168,578.

Table 5.22. Mathematics Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	80,609	49.02
	Male	83,820	50.98
Ethnicity	Asian	18,177	11.14
	Black	30,094	18.44
	Hispanic	46,282	28.36
	American Indian	1,143	0.70
	Multiracial	3,622	2.22
	Pacific Islander	605	0.37
	White	63,244	38.76
NRC	New York	63,931	38.88
	Big 4 Cities	5,499	3.34
	Urban/Suburban	12,070	7.34
	High Needs Rural	8,795	5.35
	Average Needs	36,022	21.91
	Low Needs	17,313	10.53
	Charter School	11,117	6.76
	Religious and Independent	9,682	5.89
SWD	No	140,757	85.60
	Yes	23,672	14.40
SUA	No	141,507	86.06
	Yes	22,922	13.94
ELL/ MLL	No	150,027	91.24
	Yes	14,402	8.76
SWD/ SUA	No	145,204	88.31
	Yes	19,225	11.69

Demographic Category		N-Count	% of Total N-Count
ELL/ MLL/ SUA	No	161,166	98.02
	Yes	3,263	1.98

*The total n-count was 164,429.

Table 5.23. Mathematics Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	73,712	48.57
	Male	78,037	51.43
Ethnicity	Asian	17,142	11.37
	Black	28,558	18.94
	Hispanic	43,042	28.55
	American Indian	1,172	0.78
	Multiracial	2,878	1.91
	Pacific Islander	469	0.31
	White	57,490	38.14
NRC	New York	63,852	42.08
	Big 4 Cities	4,973	3.28
	Urban/Suburban	10,510	6.93
	High Needs Rural	7,972	5.25
	Average Needs	31,694	20.89
	Low Needs	16,444	10.84
	Charter School	10,241	6.75
	Religious and Independent	6,063	4.00
SWD	No	129,303	85.21
	Yes	22,446	14.79
SUA	No	130,570	86.04
	Yes	21,179	13.96
ELL/ MLL	No	139,640	92.02
	Yes	12,109	7.98
SWD/ SUA	No	133,667	88.08
	Yes	18,082	11.92
ELL/ MLL/ SUA	No	149,202	98.32
	Yes	2,547	1.68

*The total n-count was 151,749.

Table 5.24. Mathematics Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	51,682	47.67
	Male	56,728	52.33
Ethnicity	Asian	10,671	9.90
	Black	22,280	20.67
	Hispanic	33,540	31.12
	American Indian	770	0.71
	Multiracial	1,650	1.53
	Pacific Islander	338	0.31
	White	38,542	35.76
NRC	New York	47,927	44.21
	Big 4 Cities	4,497	4.15
	Urban/Suburban	7,670	7.07
	High Needs Rural	6,351	5.86
	Average Needs	19,653	18.13
	Low Needs	8,430	7.78
	Charter School	6,642	6.13
	Religious and Independent	7,240	6.68
SWD	No	89,758	82.79
	Yes	18,652	17.21
SUA	No	90,616	83.59
	Yes	17,794	16.41
ELL/ MLL	No	98,169	90.55
	Yes	10,241	9.45
SWD/ SUA	No	93,287	86.05
	Yes	15,123	13.95
ELL/ MLL/ SUA	No	106,229	97.99
	Yes	2,181	2.01

*The total n-count was 108,410.

5.4. Classical Data Analysis

Classical data analysis of the NYSTP Grades 3–8 ELA and Mathematics Tests consists of several important elements. One element is the analysis of item-level statistical information about student performance. It is important to verify that the items and test forms function as intended. If any serious error were to occur with an item, errors should be flagged and evaluated for rectification (suppression, credit, or other acceptable solution) during item analysis. Analyses of test-level data comprise the second element of classical data analysis. These include examination of the raw score (RS) statistics (mean and standard deviation or "SD") and test

reliability measures Cronbach's alpha (Cronbach, 1951) and Feldt-Raju coefficient (Qualls, 1995). Additionally, classical DIF analysis is conducted at this stage. DIF analysis includes computation of standardized mean differences and Mantel-Haenszel statistics for New York State items to identify potential item bias. All classical data analysis results contribute information on the validity and reliability of the tests (see also Section 3, "Validity," and Section 7, "Reliability and Standard Error of Measurement").

5.4.1. *Item Difficulty and Point Biserial Correlation Coefficients*

Item difficulty is classically measured by the *p*-value statistic. It assesses the proportion of students who responded correctly to each MC item or the average proportion of the maximum score that students earned on each CR item. It is important to have a good range of *p*-values to increase test information and to avoid floor or ceiling effects. *P*-values represent the overall degree of difficulty, but do not account for demonstrated student performance on other test items. Usually, *p*-value information is coupled with point biserial (pbis) statistics, to verify that items are functioning as intended. In Appendix M, Tables M1–M12 illustrate classical test statistics for all items on each grade-level test. Appendix F provides general psychometric guidelines for operational item selection.

Item difficulties (*p*-values) ranged from 0.33 to 0.93 for the ELA tests and 0.21 to 0.94 on the Mathematics tests. These statistics are provided in Appendix M, Tables M1–M12, along with other classical test statistics.

Point-biserial statistics are used to examine item-test correlations, or item discrimination, for MC items. The pbis correlation for the key (i.e., the correct answer) is a measure of internal consistency, while pbis for specific response options aid in flagging possible alternate keys; each is a correlation that ranges between +/–1. It is the correlation of students' responses to an item relative to their performance on the rest of the test and, unless otherwise noted, this discussion will be limited to the point biserial of the correct response with the remainder of the test.

Point-biserial correlations from the operational analyses are presented in Appendix M Tables M1–M12. The column labeled "Pbis Key" contains the point biserial correlation associated with the correct response. The guideline for building the NYSTP Grades 3–8 ELA and Mathematics Tests was that the point-biserial correlation for the key for MC items should be equal to or greater than 0.20, which would indicate that students who responded correctly to that item also tended to do well on the overall test. The few exceptions to this guideline were due to content considerations that required the inclusion of particular items. Decisions to use such items were made very carefully, and no item with a negative point-biserial correlation was allowed on the test.

Point biserials for correct answer options on the ELA tests ranged from 0.10 to 0.71, as shown in Appendix M, Tables M1–M6. For Grade 3, the item pbis values ranged from 0.25 to 0.66, with a mean of 0.42. For Grade 4, the item pbis values ranged from 0.22 to 0.70, with a mean of 0.42. For Grade 5, the item pbis values ranged from 0.14 to 0.63, with a mean of 0.38. For Grade 6, the item pbis values ranged from 0.14 to 0.70, with a mean of 0.41. For Grade 7, the item pbis values ranged from 0.10 to 0.69, with a mean of 0.40. For Grade 8, the item pbis values ranged from 0.19 to 0.71, with a mean of 0.40.

Point biserials for correct answer options on the Mathematics tests ranged from 0.25 to 0.74, as shown in Appendix M, Tables M7–M12. For Grade 3, the item pbis values ranged from 0.29 to 0.69, with a mean of 0.47. For Grade 4, the item pbis values ranged from 0.33 to 0.70, with a mean of 0.50. For Grade 5, the item pbis values ranged from 0.29 to 0.71, with a mean of 0.50. For Grade 6, the item pbis values ranged from 0.34 to 0.68, with a mean of 0.50. For Grade 7, the item pbis values ranged from 0.29 to 0.74, with a mean of 0.50. For Grade 8, the item pbis values ranged from 0.25 to 0.65, with a mean of 0.45.

5.4.2. *Omit Rates*

Omit rates (i.e., percentage of students not answering a given item) are routinely checked, based on test data, after each administration. Tables M1–M12 in Appendix M show the omit rates for items on the Grades 3–8 ELA and Mathematics Tests, respectively. The industry standard general rule of thumb is that omit rates for multiple-choice items should be less than 5%. Omit rates across multiple-choice and constructed-response items on the Grades 3–8 ELA and Mathematics Tests typically ranged from 0% to 3%. As may be expected, omit rates tended to increase for items at the end of the test sessions. That is, omit rates remained within the acceptable range for large-scale achievement tests.

5.4.3. *Differential Item Functioning (DIF)*

Classical differential item functioning (DIF) analyses are statistical methods for identifying items that are estimated to have functioned differently for one group (i.e., the "focal" group) as compared with another group (i.e., the "reference" group). In other words, DIF analysis only *flags* items that may later be judged by content experts to exhibit bias, rather than directly detecting bias. First, the psychometric phenomenon of DIF was extensively investigated and experts' judgments of bias collected when items were field-tested, which reduced the likelihood of including any differentially functioning items on the operational forms for 2018. Turning to the analysis of the 2018 operational data, as discussed in Section 3.2.3. Detection of Bias, items flagged for DIF do not necessarily indicate item bias. For example, DIF may be attributed to true group differences on the content measured by the item or Type I error, which refers to statistically flagging items that have no true DIF. Operational items flagged for DIF are given additional scrutiny by content specialists, above and beyond the existing rounds of reviews by New York State educators, and those content specialists make the final judgment as to whether or not an item is biased for or against the focal group.

DIF was evaluated using two methods, both of which involve checks on statistical and practical significance. First, the Mantel-Haenszel (MH) method is employed for MC items. This non-parametric DIF method partitions the sample of examinees into categories based on total raw test scores. It then compares the log-odds ratio of keyed responses for the focal and reference groups. In terms of statistical significance, the Mantel-Haenszel method has a critical value of 6.63 (degrees of freedom = 1 for MC items; alpha = 0.01) and as far as practical significance is concerned, it is compared to its corresponding delta-value. Delta-values are a commonly used metric in testing that indicates the magnitude of DIF. Typically, delta-values above 1.50 are considered indicative of moderate DIF that should be examined more closely (Zwick, Donoghue, and Grima, 1993). Second, the standardized mean difference (SMD) was computed for CR items. The SMD statistic (Dorans, Schmitt, and Bleistein, 1992) compares the mean scores of reference and focal groups, after adjusting for proficiency differences. The SMD was also evaluated for statistical significance and, in terms of practical significance, a moderate amount of

DIF, for or against the focal group, is represented by an SMD with an absolute value between 0.10 and 0.19, inclusive; a large amount of DIF is represented by an SMD with an absolute value of 0.20 or greater.

Classical DIF analyses were conducted on subgroups of the Needs/Resource Capacity Category (focal group: High Needs; reference group: Low Needs), gender (focal group: Female; reference group: Male), ethnicity (focal groups: Black, Hispanic, and Asian; reference group: White), English language learners (focal group: English language learners; reference group: Non-English language learners), and mode (focal group: PBT students; reference group: CBT students). The DIF analyses were conducted using all cases from the clean data sets. Table 5.25 and Table 5.26 show the numbers of cases for the subgroups for ELA and Mathematics, respectively.

Table 5.25. ELA Classical DIF Sample N-Counts

	Ethnicity				Gender		Needs/Resource Capacity Category		English Language Learners		Mode	
Grade	Black	Hispanic/ Latino	Asian American	White	Female	Male	High	Low	ELL	Non-ELL	CBT	PBT
3	31,318	50,296	17,785	71,082	88,724	90,615	98,321	60,392	20,595	158,744	16,763	162,576
4	32,133	50,017	18,533	72,411	89,676	91,996	97,914	58,908	17,497	164,175	16,282	165,390
5	31,523	48,692	18,643	68,391	86,784	88,391	97,101	57,398	14,651	160,524	15,594	159,581
6	31,314	46,768	18,003	66,335	83,617	86,398	91,909	53,991	13,503	156,512	18,988	151,027
7	29,642	42,405	17,046	60,427	75,962	79,957	89,866	50,012	11,401	144,518	15,907	140,012
8	29,158	41,041	17,516	58,332	73,680	77,842	85,807	45,361	10,941	140,581	14,563	136,959

Table 5.26. Mathematics Classical DIF Sample N-Counts

	Ethnicity				Gender		Needs/Resource Capacity Category		English Language Learners		Mode	
Grade	Black	Hispanic/ Latino	Asian American	White	Female	Male	High	Low	ELL	Non-ELL	CBT	PBT
3	30,369	50,760	18,053	69,246	87,245	89,418	97,960	60,346	22,056	154,607	13,850	162,813
4	30,994	50,271	18,806	68,820	86,958	89,939	97,451	58,919	18,448	158,449	11,872	165,025
5	29,957	48,437	18,740	64,450	83,182	85,396	95,797	56,489	14,916	153,662	10,747	157,831
6	30,094	46,282	18,177	63,244	80,609	83,820	90,295	53,335	14,402	150,027	13,484	150,945
7	28,558	43,042	17,142	57,490	73,712	78,037	87,307	48,138	12,109	139,640	10,663	141,086
8	22,280	33,540	10,671	38,542	51,682	56,728	66,445	28,083	10,241	98,169	7,116	101,294

Table 5.27 (ELA) and Table 5.28 (Mathematics) present the number of items flagged for DIF by either of the classical methods described earlier. Appendix N provides a detailed list of items flagged by either one or both of these classical DIF methods, including DIF direction and associated DIF statistics.

Table 5.27. ELA Items Flagged for DIF

Grade	Flagged Items
3	5
4	4
5	10
6	13
7	9
8	11

Table 5.28. Mathematics Items Flagged for DIF

Grade	Flagged Items
3	3
4	2
5	4
6	2
7	3
8	1

As discussed in Section 3: Validity, items showing statistically significant DIF (flagged as described above for MH statistics on MC items and SMD statistics for CR items) do not necessarily pose bias. The items flagged with DIF were examined further by the content experts; no signs of potential content-based issues were discovered. The items are possibly functioning differently statistically.

Section 6: IRT Calibration

6.1. IRT Models and Rationale for Use

IRT allows for comparisons between items and scale scores, even those from different test forms, by using a common scale for all items and examinees (i.e., as if there were a hypothetical test that contained items from all forms). The three-parameter logistic (3PL) model (Lord and Novick, 1968; Lord, 1980) was used to analyze item responses on the MC items. For analysis of the CR items, the two-parameter partial credit (2PPC) model (Muraki, 1992; Yen, 1993) was used.

IRT is a statistical methodology that takes into account the fact that not all test items are alike and that not all test items provide the same amount of information in determining how much a student knows or can do. Computer programs that implement IRT models use actual student data to estimate the characteristics of the items on a test, called "parameters." The parameter estimation process is called "item calibration."

IRT models typically vary according to the number of parameters estimated. For the New York State tests, three parameters are estimated: the discrimination parameter, the difficulty parameter(s), and, for MC items, the guessing parameter. The discrimination parameter is an index of how well an item differentiates between high-performing and low-performing students. An item that cannot be answered correctly by low-performing students, but can be answered correctly by high-performing students, will have a high-discrimination value. The difficulty parameter is an index of how easy or difficult an item is. The higher the difficulty parameter is, the harder the item is. The guessing parameter is the probability that a student with very low proficiency will answer the item correctly.

Because the characteristics of MC and CR items are different, two IRT models were used in item calibration. The three-parameter logistic (3PL) model was used in the analysis of MC items. In this model, the probability that a student with proficiency θ responds correctly to item i is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]},$$

where

a_i is the item discrimination, b_i is the item difficulty, and c_i is the probability of a correct response from a very low-scoring student.

For analysis of the CR items, the 2PPC model was used. The 2PPC model is a special case of Bock's (1972) nominal model. Bock's model states that the probability of an examinee with proficiency θ having a score (k - 1) at the kth level of the jth item is:

$$P_{jk}(\theta) = P(x_j = k - 1 \mid \theta) = \frac{\exp Z_{jk}}{\sum_{i=1}^{m_j} \exp Z_{ji}}, \; k = 1 \text{K } m_j,$$

where

$$Z_{jk} = A_{jk}\theta + C_{jk},$$

and

k is the item response category (k = 1, 2, …. m_j).

The m_j denotes the number of score levels for the jth item, and, typically, the highest score level is assigned (m_j - 1) score points. For the special case of the 2PPC model used here, the following constraints were used:

$$A_{jk} = \alpha_j(k-1),$$

and

$$C_{jk} = -\sum_{i=0}^{k-1}\gamma_{ji},$$

where

$$\gamma_{j0} = 0,$$

and

α_j and γ_{ji} are the free parameters to be estimated from the data.

Each item has (m_j - 1) independent γ_{ji} parameters and one α_j parameter; a total of m_j parameters are estimated for each item.

6.2. Calibration Sample

The cleaned data were used for calibration of the NYSTP 2018 Grades 3–8 ELA and Mathematics Tests. It should be noted that the sample sizes were adequate, as the calibration was performed using nearly all (96–99%, depending on grade level) of the New York State public and non-public school student population data in each tested grade. As shown in Tables 6.1–6.3 and Tables 6.4–6.6 for ELA and Mathematics, respectively, the 2018 operational test samples were generally comparable to 2017 populations in terms of NRC, student race and ethnicity, proportions of ELL/MLL students, proportions of students with disabilities, and proportions of students using testing accommodations.

Table 6.1. ELA Grades 3 and 4 Demographic Statistics

		Grade 3		Grade 4	
		2017	2018	2017	2018
Demographic Category		Population	Sample	Population	Sample
Gender	Female	49.11	49.47	49.64	49.36
	Male	50.89	50.53	50.36	50.64
Ethnicity	Asian	9.97	10.03	10.32	10.32
	Black	18.05	17.66	18.11	17.89
	Hispanic	28.16	28.37	27.97	27.85
	American Indian	0.66	0.70	0.67	0.70

Demographic Category		Grade 3		Grade 4	
		2017 Population	2018 Sample	2017 Population	2018 Sample
Ethnicity	Multiracial	2.70	2.90	2.44	2.63
	Pacific Islander	0.28	0.24	0.31	0.28
	White	40.17	40.09	40.18	40.32
NRC	New York	37.71	37.09	38.59	36.85
	Big 4 Cities	4.31	4.31	4.20	4.27
	Urban/Suburban	7.91	7.93	7.39	7.37
	High Needs Rural	5.55	5.50	5.29	5.41
	Average Needs	22.79	23.55	21.57	22.45
	Low Needs	9.93	10.12	9.79	9.98
	Charter School	6.51	6.75	5.48	6.21
	Religious and Independent	5.28	4.75	7.71	7.47
SWD	No	85.12	85.98	84.66	85.61
	Yes	14.88	14.02	15.34	14.39
SUA	No	92.93	87.37	92.46	86.09
	Yes	7.07	12.63	7.54	13.91
ELL/ MLL	No	89.22	88.52	91.06	90.37
	Yes	10.78	11.48	8.94	9.63
SWD/ SUA	No	94.14	89.42	94.13	88.40
	Yes	5.86	10.58	5.87	11.60
ELL/ MLL/ SUA	No	99.28	97.90	99.32	97.97
	Yes	0.72	2.10	0.68	2.03

Table 6.2. ELA Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2017 Population	2018 Sample	2017 Population	2018 Sample
Gender	Female	49.25	49.54	48.88	49.18
	Male	50.75	50.46	51.12	50.82
Ethnicity	Asian	10.36	10.76	10.43	10.72
	Black	18.65	18.19	19.28	18.65
	Hispanic	27.85	28.10	27.67	27.86
	American Indian	0.68	0.72	0.70	0.68
	Multiracial	2.30	2.46	1.93	2.21

Demographic Category		Grade 5		Grade 6	
		2017 Population	2018 Sample	2017 Population	2018 Sample
Ethnicity	Pacific Islander	0.37	0.31	0.30	0.37
	White	39.80	39.47	39.69	39.51
NRC	New York	39.50	38.74	39.03	37.72
	Big 4 Cities	4.10	4.28	3.97	4.03
	Urban/Suburban	7.47	7.10	6.86	7.01
	High Needs Rural	5.39	5.31	4.99	5.29
	Average Needs	21.94	22.33	20.87	21.45
	Low Needs	10.19	10.44	9.93	10.31
	Charter School	5.94	6.36	6.56	6.70
	Religious and Independent	5.47	5.44	7.79	7.49
SWD	No	83.66	84.86	83.71	85.15
	Yes	16.34	15.14	16.29	14.85
SUA	No	91.51	85.20	91.50	85.90
	Yes	8.49	14.80	8.50	14.10
ELL/ MLL	No	91.84	91.64	92.51	92.06
	Yes	8.16	8.36	7.49	7.94
SWD/ SUA	No	93.24	87.62	93.42	88.35
	Yes	6.76	12.38	6.58	11.65
ELL/ MLL/ SUA	No	99.25	97.95	99.30	98.19
	Yes	0.75	2.05	0.70	1.81

Table 6.3. ELA Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2017 Population	2018 Sample	2017 Population	2018 Sample
Gender	Female	48.73	48.72	48.44	48.63
	Male	51.27	51.28	51.56	51.37
Ethnicity	Asian	11.01	11.06	11.22	11.67
	Black	19.35	19.23	20.36	19.43
	Hispanic	27.02	27.51	27.45	27.35
	American Indian	0.74	0.78	0.79	0.78
	Multiracial	1.63	1.93	1.36	1.58
	Pacific Islander	0.28	0.30	0.30	0.30
	White	39.97	39.20	38.51	38.88

Demographic Category		Grade 7		Grade 8	
		2017 Population	2018 Sample	2017 Population	2018 Sample
NRC	New York	40.70	41.23	42.57	41.10
	Big 4 Cities	3.99	4.08	4.11	4.10
	Urban/Suburban	6.63	6.96	6.39	6.22
	High Needs Rural	5.13	5.37	5.01	5.21
	Average Needs	20.17	21.13	18.92	19.49
	Low Needs	10.94	10.94	9.74	10.45
	Charter School	6.58	6.75	5.68	6.31
	Religious and Independent	5.86	3.54	7.58	7.13
SWD	No	83.64	84.41	84.38	85.18
	Yes	16.36	15.59	15.62	14.82
SUA	No	91.27	85.19	92.13	85.99
	Yes	8.73	14.81	7.87	14.01
ELL/ MLL	No	92.71	92.69	93.02	92.78
	Yes	7.29	7.31	6.98	7.22
SWD/ SUA	No	93.02	87.69	93.91	88.34
	Yes	6.98	12.31	6.09	11.66
ELL/ MLL/ SUA	No	99.28	98.37	99.51	98.48
	Yes	0.72	1.63	0.49	1.52

Table 6.4. Mathematics Grades 3 and 4 Demographic Statistics

Demographic Category		Grade 3		Grade 4	
		2017 Population	2018 Sample	2017 Population	2018 Sample
Gender	Female	48.93	49.38	49.35	49.16
	Male	51.07	50.62	50.65	50.84
Ethnicity	Asian	10.26	10.31	10.61	10.73
	Black	17.87	17.35	17.94	17.69
	Hispanic	28.42	28.99	28.24	28.69
	American Indian	0.66	0.70	0.67	0.67
	Multiracial	2.66	2.86	2.39	2.66
	Pacific Islander	0.28	0.24	0.31	0.28
	White	39.85	39.55	39.85	39.28
NRC	New York	38.25	38.01	39.03	38.05
	Big 4 Cities	4.36	3.75	4.24	3.64

Demographic Category		Grade 3 2017 Population	Grade 3 2018 Sample	Grade 4 2017 Population	Grade 4 2018 Sample
NRC	Urban/Suburban	7.92	8.09	7.40	7.83
	High Needs Rural	5.47	5.60	5.21	5.56
	Average Needs	22.34	23.87	21.29	23.04
	Low Needs	9.88	10.29	9.80	10.27
	Charter School	6.44	6.68	5.41	6.19
	Religious and Independent	5.34	3.71	7.63	5.42
SWD	No	85.16	86.44	84.79	85.99
	Yes	14.84	13.56	15.21	14.01
SUA	No	93.22	87.52	92.41	86.27
	Yes	6.78	12.48	7.59	13.73
ELL/ MLL	No	87.86	87.52	89.67	89.57
	Yes	12.14	12.48	10.33	10.43
SWD/ SUA	No	94.32	89.42	93.92	88.40
	Yes	5.68	10.58	6.08	11.60
ELL/ MLL/ SUA	No	99.34	97.74	99.26	97.82
	Yes	0.66	2.26	0.74	2.18

Table 6.5. Mathematics Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5 2017 Population	Grade 5 2018 Sample	Grade 6 2017 Population	Grade 6 2018 Sample
Gender	Female	48.99	49.34	48.70	49.02
	Male	51.01	50.66	51.30	50.98
Ethnicity	Asian	10.64	11.19	10.80	11.14
	Black	18.47	17.89	19.07	18.44
	Hispanic	28.09	28.93	27.98	28.36
	American Indian	0.69	0.73	0.70	0.70
	Multiracial	2.15	2.43	1.87	2.22
	Pacific Islander	0.37	0.32	0.30	0.37
	White	39.58	38.50	39.28	38.76
NRC	New York	40.01	40.19	39.82	38.88
	Big 4 Cities	4.13	3.54	3.98	3.34
	Urban/Suburban	7.44	7.67	6.75	7.34

Demographic Category		Grade 5		Grade 6	
		2017 Population	2018 Sample	2017 Population	2018 Sample
NRC	High Needs Rural	5.26	5.43	4.86	5.35
	Average Needs	21.51	22.81	20.42	21.91
	Low Needs	10.15	10.69	9.83	10.53
	Charter School	5.87	6.29	6.49	6.76
	Religious and Independent	5.63	3.38	7.85	5.89
SWD	No	83.92	85.36	84.05	85.60
	Yes	16.08	14.64	15.95	14.40
SUA	No	91.98	85.93	92.11	86.06
	Yes	8.02	14.07	7.89	13.94
ELL/ MLL	No	90.49	91.15	91.10	91.24
	Yes	9.51	8.85	8.90	8.76
SWD/ SUA	No	93.57	87.79	93.79	88.31
	Yes	6.43	12.21	6.21	11.69
ELL/ MLL/ SUA	No	99.28	97.86	99.36	98.02
	Yes	0.72	2.14	0.64	1.98

Table 6.6. Mathematics Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2017 Population	2018 Sample	2017 Population	2018 Sample
Gender	Female	48.54	48.57	47.49	47.67
	Male	51.46	51.43	52.51	52.33
Ethnicity	Asian	11.32	11.37	9.79	9.90
	Black	19.12	18.94	21.65	20.67
	Hispanic	27.44	28.55	30.58	31.12
	American Indian	0.74	0.78	0.81	0.71
	Multiracial	1.56	1.91	1.28	1.53
	Pacific Islander	0.29	0.31	0.30	0.31
	White	39.53	38.14	35.58	35.76
NRC	New York	41.72	42.08	45.59	44.21
	Big 4 Cities	4.01	3.28	4.75	4.15
	Urban/Suburban	6.44	6.93	6.51	7.07
	High Needs Rural	4.96	5.25	5.11	5.86

Demographic Category		Grade 7		Grade 8	
		2017 Population	2018 Sample	2017 Population	2018 Sample
NRC	Average Needs	19.53	20.89	16.10	18.13
	Low Needs	10.62	10.84	6.75	7.78
	Charter School	6.56	6.75	5.63	6.13
	Religious and Independent	6.15	4.00	9.56	6.68
SWD	No	84.00	85.21	81.94	82.79
	Yes	16.00	14.79	18.06	17.21
SUA	No	92.40	86.04	91.63	83.59
	Yes	7.60	13.96	8.37	16.41
ELL/ MLL	No	91.04	92.02	89.45	90.55
	Yes	8.96	7.98	10.55	9.45
SWD/ SUA	No	93.91	88.08	93.34	86.05
	Yes	6.09	11.92	6.66	13.95
ELL/ MLL/ SUA	No	99.43	98.32	99.45	97.99
	Yes	0.57	1.68	0.55	2.01

6.2.1. *Calibration Process*

The item parameters were estimated using Scientific Software International (SSI) Inc.'s IRTPRO Version 2.1 (Cai, Thissen, & du Toit, 2011) package. MC and CR items were calibrated simultaneously, using marginal maximum likelihood procedures.

The calibration of NYSTP 2018 Grades 3–8 ELA and Mathematics Tests did not exhibit any test-level issues. The estimated parameters were on the original theta scale, and all of the items were well within the prescribed parameter ranges. For both the Grades 3–8 ELA and Mathematics Tests, all calibration estimation results were reasonable. Tables 6.7 and 6.8 present the summaries of the calibration results for ELA and Mathematics, respectively. Additional details, including individual item parameter estimates, may be found in Appendix O, in Tables O13–O24. The parameter estimates are expressed on the theta metric and are defined below:

- MC items:
  - *a*-parameter is a discrimination parameter
  - *b*-parameter is a difficulty parameter
  - *c*-parameter is a guessing parameter

- CR items:
  - *alpha* is a discrimination parameter
  - *step* is a difficulty parameter for category m_j

As described above in Section 6.1, m_j denotes the number of score levels for the jth item, and, typically, the highest score level is assigned (m_j - 1) score points. For the 2PPC model, there are m_j - 1 independent steps and one alpha, for a total of m_j independent parameters estimated for each item, while there is one a-parameter and one b-parameter per item in the 3PL model.

Table 6.7. ELA Calibration Results

	Item-Level			Student-Level		
	Largest	Range of b-			Theta Est.*	
Grade	a-Parameter	Parameters		N-Count	Mean	SD
3	1.343	-1.343	0.919	179,354	0.01	0.91
4	1.241	-1.357	0.934	181,683	0.00	0.91
5	1.509	-2.650	2.120	175,187	0.01	0.92
6	1.605	-2.125	2.237	170,026	0.00	0.92
7	1.500	-2.390	1.982	155,927	0.00	0.93
8	1.324	-3.557	0.917	151,530	-0.01	0.93

*Maximum *a posteriori* (MAP) theta estimates.

Table 6.8. Mathematics Calibration Results

	Item-Level			Student-Level		
	Largest	Range of b-			Theta Est.*	
Grade	a-Parameter	Parameters		N-Count	Mean	SD
3	1.635	-2.271	1.218	176,663	0.00	0.92
4	1.592	-1.758	0.795	176,897	0.00	0.92
5	1.808	-1.582	1.102	168,578	0.01	0.92
6	1.876	-1.401	1.385	164,429	0.02	0.91
7	2.254	-1.174	1.210	151,749	0.02	0.90
8	1.750	-0.934	1.411	108,410	0.04	0.89

*Maximum *a posteriori* (MAP) theta estimates.

6.3. Item-Model Fit

Item fit statistics provide evidence of the appropriateness of using an item in the 3PL or 2PPC model. The Q_1 procedure described by Yen (1981) was used to measure fit to the three-parameter model. Students are rank-ordered based on $\hat{\theta}$ values and sorted into ten cells with 10% of the sample in each cell. For each item, the number of students in cell k who answered item i, N_{ik}, and the number of students in that cell who answered item i correctly, R_{ik}, were determined. The observed proportion in cell k passing item i, O_{ik}, is R_{ik}/N_{ik}. The fit index for item i is:

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

with:

$$E_{ik} = \frac{1}{N_{ik}} \sum_{j \varepsilon \text{ cell } k}^{N_{ik}} P_i(\hat{\theta}_j) \quad .$$

A modification of this procedure was used to measure fit to the 2PPC model. For the 2PPC model, Q_{1j} was assumed to have an approximate chi-square distribution with the following degrees of freedom (*df*):

$$df = I(m_j - 1) - m_j \quad ,$$

where I is the total number of cells (usually 10) and m_j is the possible number of score levels for item j.

To adjust for differences in degrees of freedom among items, Q_1 was transformed to Z_{Q_1} where:

$$Z_{Q_1} = (Q_1 - df)/(2df)^{1/2} \quad .$$

The value of Z increases with sample size, when all else is equal. To use this standardized statistic to flag items for potential poor fit, it has been a common practice to vary the critical value for Z as a function of sample size. For the tests that have large calibration sample sizes, the criterion $Z_{Q_1}Crit$ was used to flag items and was calculated using the expression

$$Z_{Q_1}Crit = \left(\frac{N}{1500}\right) * 4 \quad ,$$

where N is the calibration sample size.

To compute the Q_1 and related statistics, a stratified sampling procedure was implemented in a way that a representative sample with the size of approximately 70,000 students was drawn at each grade level. Items were considered to have poor fit if the value of the obtained Z_{Q_1} was greater than the value of Z_{Q_1} critical. If the obtained Z_{Q_1} was less than Z_{Q_1} critical, the items were rated as having acceptable fit.

Any item flagged with extreme item parameters or significant mis-fit was reviewed by both content and psychometric teams. Interventions were applied as needed to improve the parameter estimates and model fit.The fact that the majority of the items in the NYSTP 2018 Grades 3–8 ELA and Mathematics Tests demonstrated good model fit further supports the use of the chosen models. Item fit statistics are presented in Tables O1–O12 in Appendix O.

6.4. Local Independence

In using IRT models, one of the assumptions made is that the items are locally independent; that a student's response to one item is not dependent upon his or her response to another item. In

other words, when a student's proficiency is accounted for, his or her response to each item is statistically independent.

One way to measure the statistical independence of items within a test is via the Q_3 statistic (Yen, 1984). This statistic was obtained by correlating differences between students' observed and expected responses for pairs of items after taking into account overall test performance. The Q_3 statistic for binary items was computed as

$$d_{ij} \equiv u_{ij} - P_j\left(\hat{\theta}_i\right)$$

where $\hat{\theta}_i$ is the estimated trait value (i.e., proficiency) for the *i*th examinee; u_{ij} is the observed probability for the *i*th examinee to get the *j*th item correct and P_j is estimated probability for the *i*th examinee to get the *j*th item correct, and

$$Q_{3,jj'} = r\left(d_j, d_{j'}\right).$$

The generalization to items with multiple response categories uses

$$d_{ij} \equiv x_{ij} - E_{ij},$$

where

$$E_{ij} \equiv E\left(x \middle| \hat{\theta}_i\right) = \sum_{k=1}^{m_j} k P_{jk}\left(\hat{\theta}_i\right).$$

If a substantial number of items in the test demonstrate local dependence, these items may need to be calibrated separately. All pairs of items with Q_3 values greater than 0.20 were classified as significant for local dependency. The maximum value for this index is 1.00. When item pairs are flagged by Q_3, the content of the flagged items is examined to identify possible sources of the local dependence. The primary concern about locally dependent items is that they contribute less psychometric information about examinee proficiency than do locally independent items, and therefore inflate score reliability estimates. After reviewing the results and the content of the pairs of items, there was not sufficient evidence to warrant further concern or action regarding the IRT calibration.

6.5. Scaling

A new reporting scale was established following the Standards Review meeting in Summer 2018. The reporting scale was developed to quantify the information captured by the assessment. Because the theta score scale used in the psychometric modeling and the IRT calibration do not appeal to the public, the reporting scale was developed to interpret changes, make comparisons, facilitate inferences, and inform educational decisions.

The scaling process was used to determine the transformation from the theta scale to the reporting scale. The following analysis steps were involved in the scaling process:

1. All operational items in the 2018 Grades 3–8 ELA and Mathematics tests were calibrated using IRT models.
2. The raw-to-theta score conversion tables were built up using the test characteristic curve (TCC) approach, based on which each student receives a theta score estimate corresponding to their raw score.
3. For raw scores below the chance level or near the perfect score, the following adjustment and interpolation was conducted to derive the adjusted theta scores:
    - At the lower end of the scale, for any theta estimates that were lower than -2.5, 0.25 was subtracted from the preceding adjusted theta value that was within the range.
    - At the higher end of the scale, for any theta estimates that were higher than 3.0, 0.25 was added to the previous theta value that was within the range.
    - See the table below for an example in the lower end of the scale.

Raw score	Theta	Adjusted theta
7	-3.66491	-3.07129
8	-3.03055	-2.82129
9	-2.62458	-2.57129
10	-2.32129	-2.32129

4. The mean and SD of the theta scores were computed from the 2018 Grades 3–8 ELA and Mathematics calibration population. They are summarized below.

Test	$\bar{\theta}$	SD_{Θ}
ELA3	-0.02	1.09
ELA4	-0.01	1.09
ELA5	-0.01	1.10
ELA6	0.00	1.09
ELA7	-0.01	1.09
ELA8	-0.01	1.09
MATH3	0.00	1.07
MATH4	0.00	1.08
MATH5	0.00	1.09
MATH6	-0.02	1.10
MATH7	-0.03	1.08
MATH8	-0.04	1.10

5. The scaling linear transformation slope (M_1^S) and intercept (M_2^S) were obtained using the formula below. They are summarized in Table 6.9.

$$M_1^S = SD_{SS} / SD_{\Theta}$$

$$M_2^S = \overline{SS} - M_1^S * \bar{\theta}$$

where $\bar{\theta}$ and SD_Θ are the mean and SD of the theta scores; $\overline{SS}$ and SD_{SS} are the mean and SD of the scale scores, which equal 600 and 20, respectively.

6. The M_1^S and M_2^S were applied to derive the scale score of each student from their theta score estimate as follows

$$ScaleScore = (M_1^S \cdot \theta) + M_2^S ,$$

Table 6.9. 2018 Operational Scaling Coefficients

Grade	Slope (M_1^S)	Intercept (M_2^S)
ELA		
3	18.310914	600.340994
4	18.276716	600.101132
5	18.212931	600.127742
6	18.309278	600.006654
7	18.318571	600.223246
8	18.308395	600.129092
Mathematics		
3	18.635919	600.082128
4	18.485491	600.009369
5	18.404109	600.040856
6	18.191784	600.432302
7	18.559827	600.499091
8	18.115200	600.640639

6.6. Test Characteristic Curves

Test Characteristic Curves (TCCs) provide an overview of the tests in the IRT scale score metric. The 2018 TCCs were generated using final item parameters for all reporting test items administered in Spring 2018. TCCs are the summation of all the item characteristic curves (ICCs) for items that contribute to the scale score. Conditional standard error of measurement (CSEM) curves graphically show the amount of measurement error at different performance levels. The TCCs and CSEM curves are presented in Figures 6.1–6.24.

Figure 6.1. ELA Grade 3 TCC

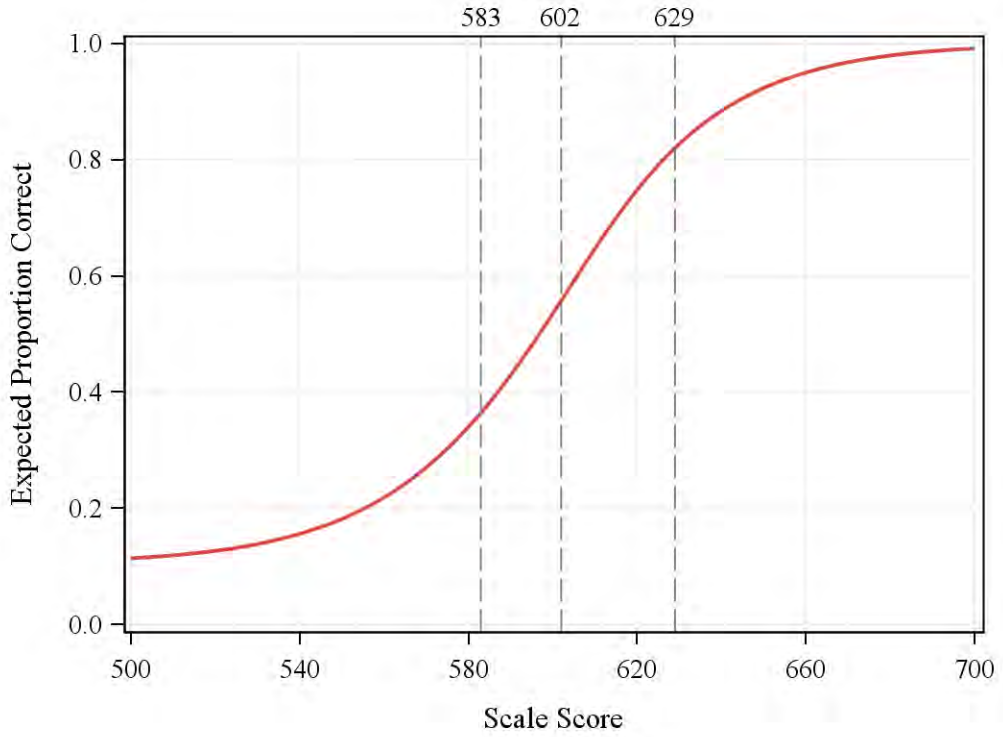


Figure 6.2. ELA Grade 3 CSEM Curve

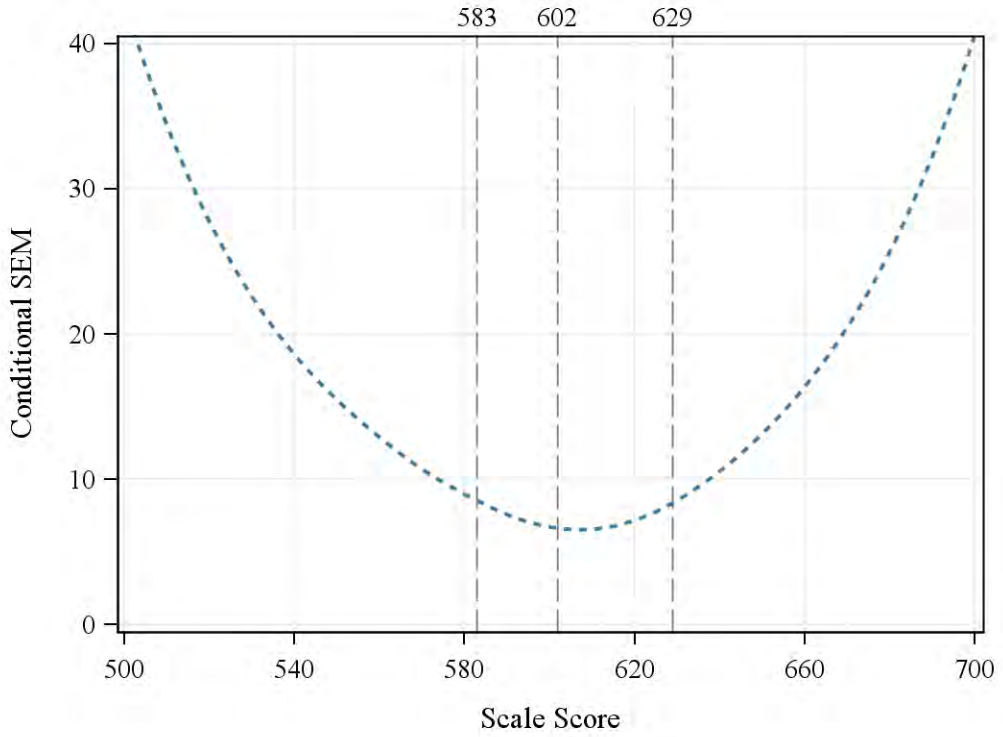


Figure 6.3. ELA Grade 4 TCC

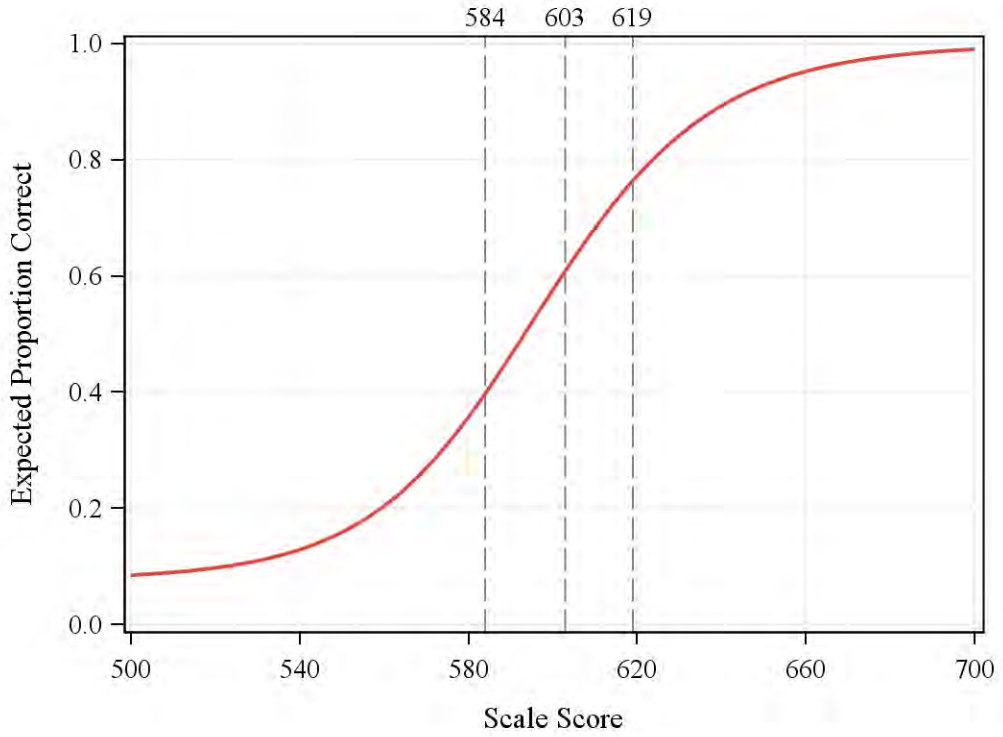


Figure 6.4. ELA Grade 4 CSEM Curve

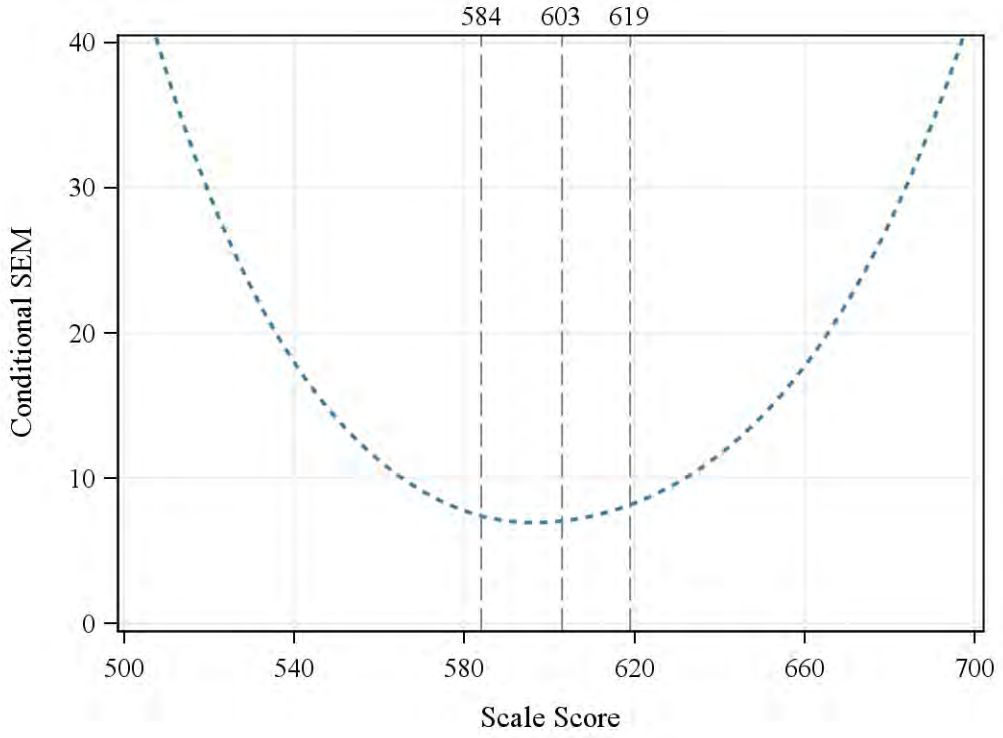


Figure 6.5. ELA Grade 5 TCC

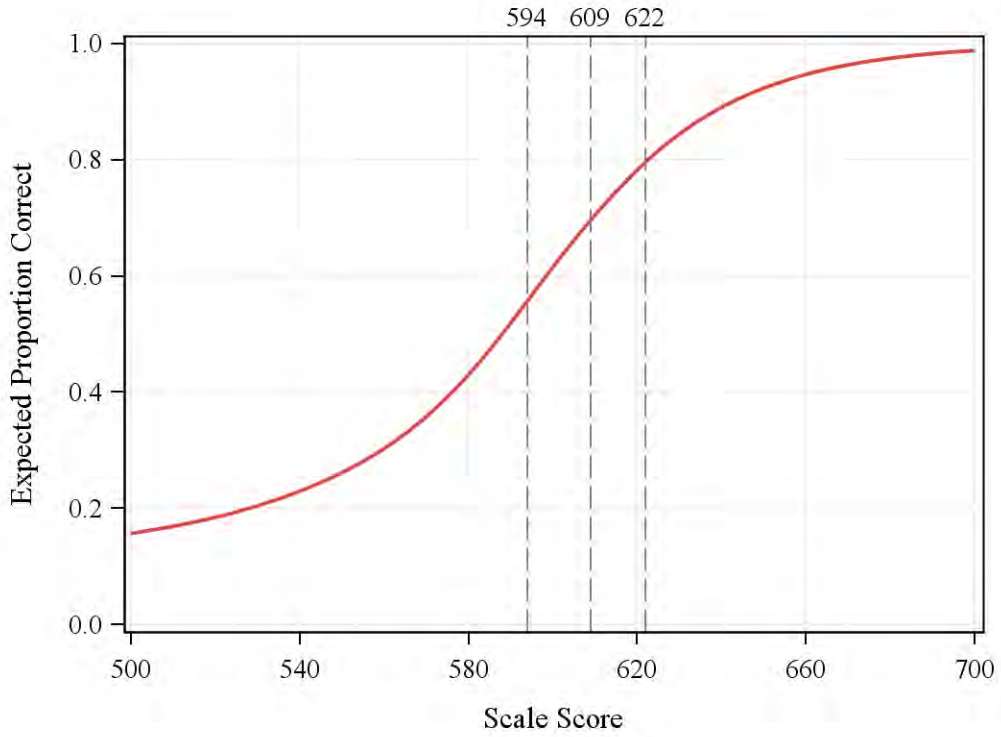


Figure 6.6. ELA Grade 5 CSEM Curve

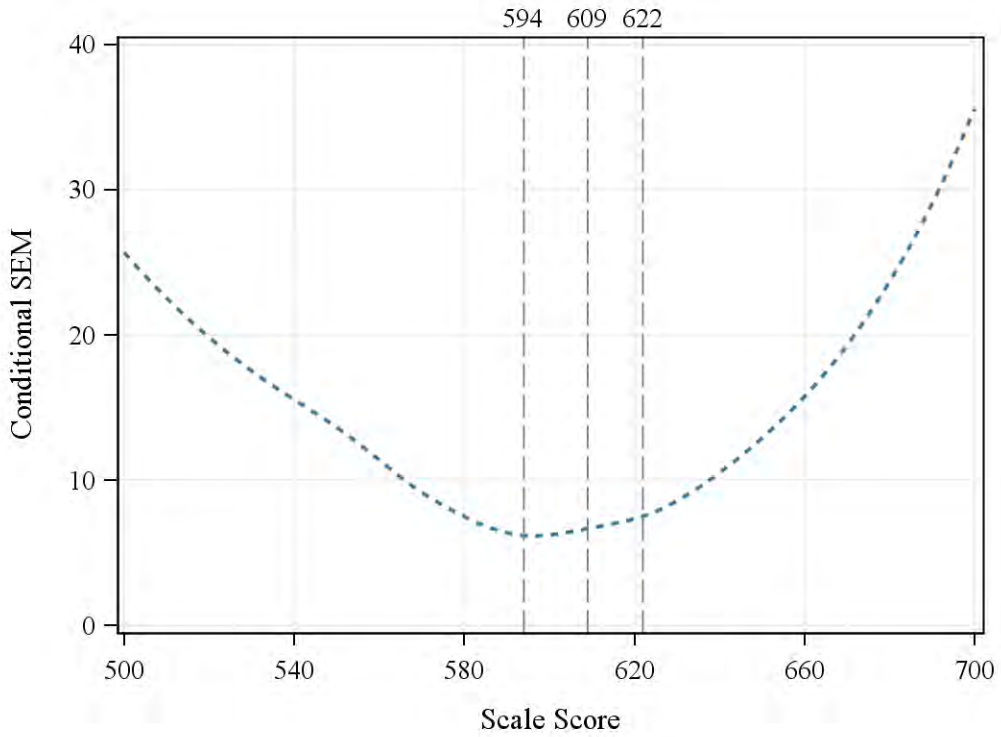


Figure 6.7. ELA Grade 6 TCC

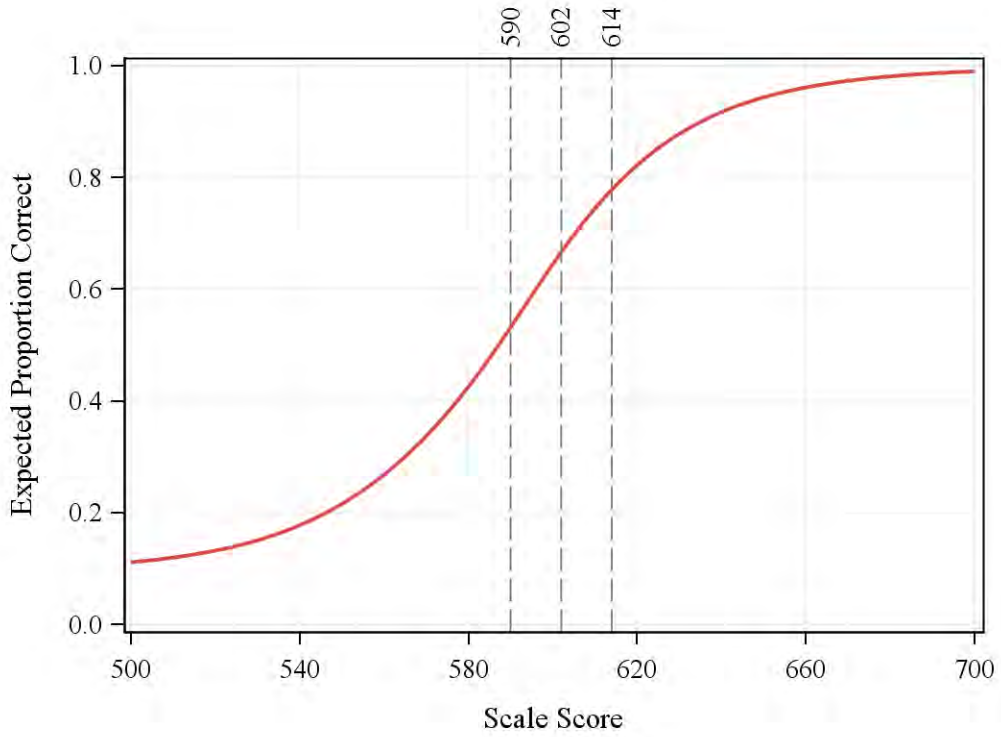


Figure 6.8. ELA Grade 6 CSEM Curve

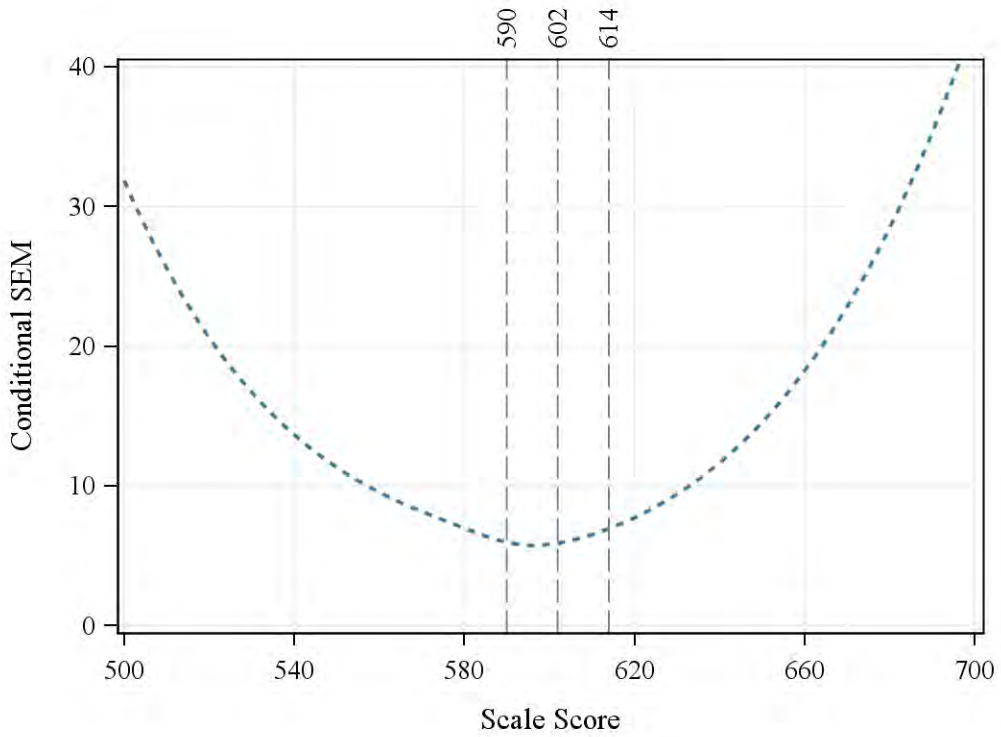


Figure 6.9. ELA Grade 7 TCC

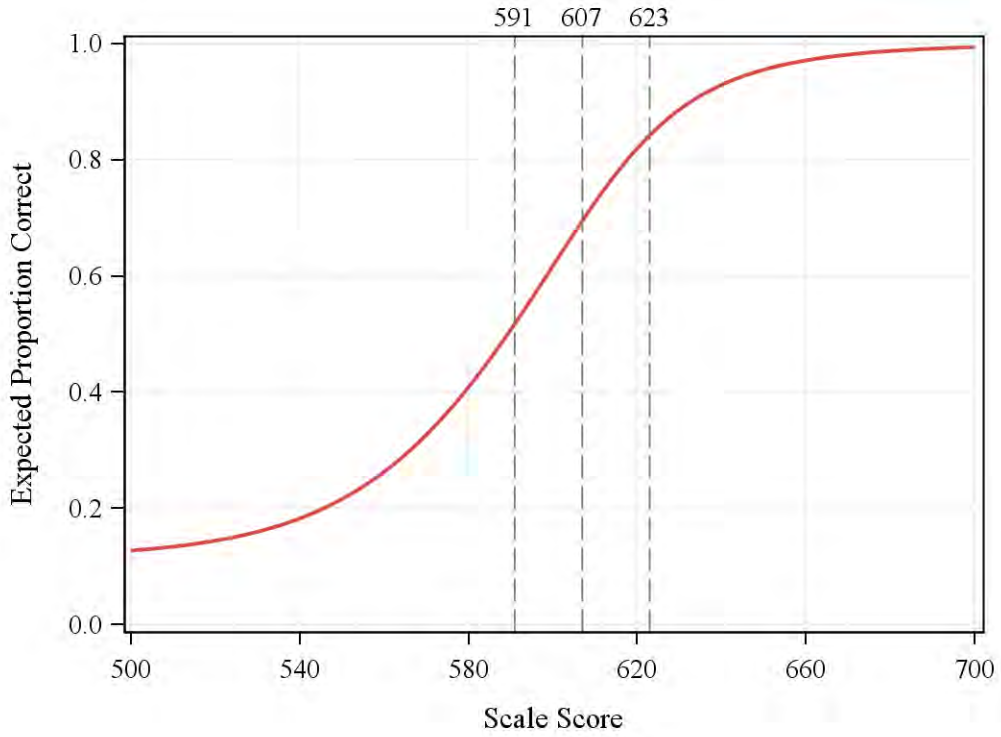


Figure 6.10. ELA Grade 7 CSEM Curve

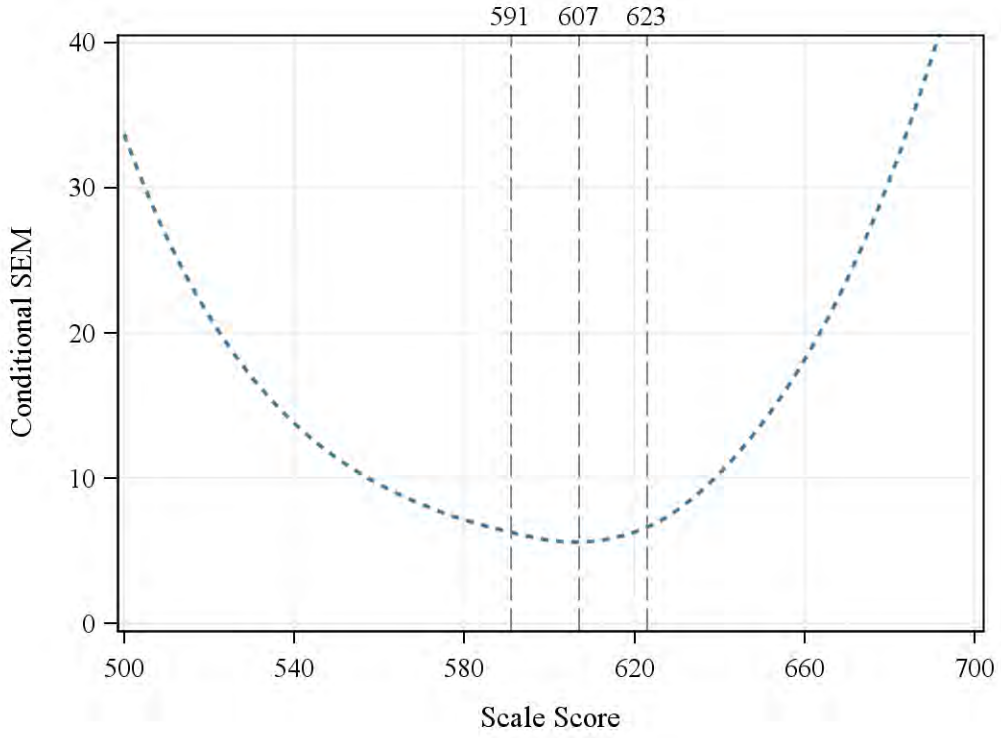


Figure 6.11. ELA Grade 8 TCC

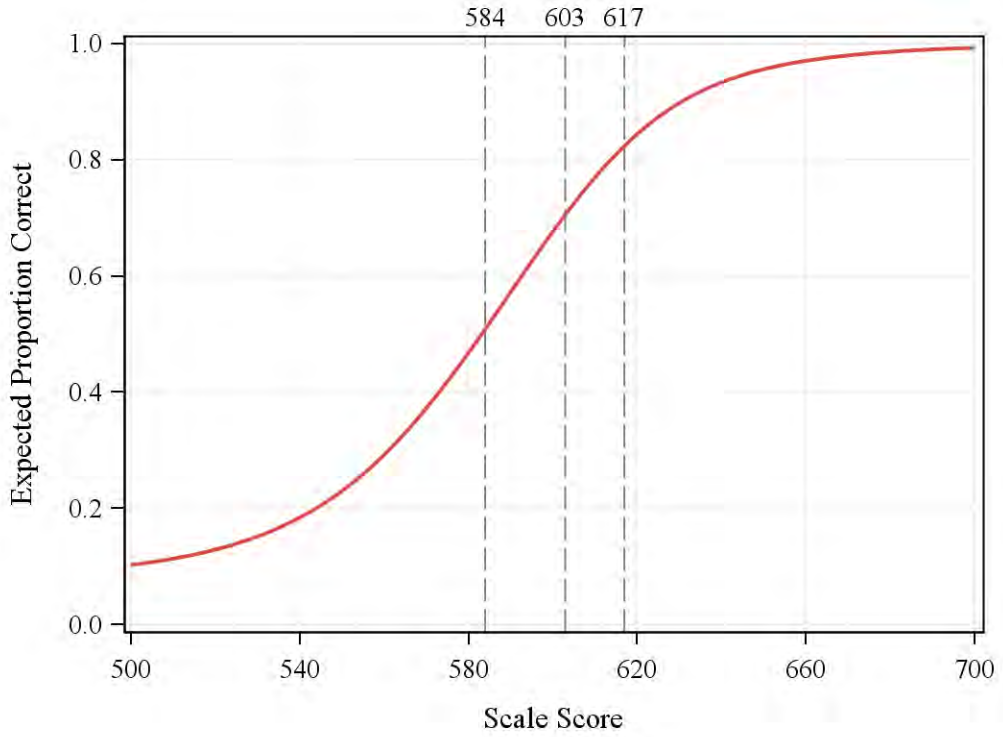


Figure 6.12. ELA Grade 8 CSEM Curve

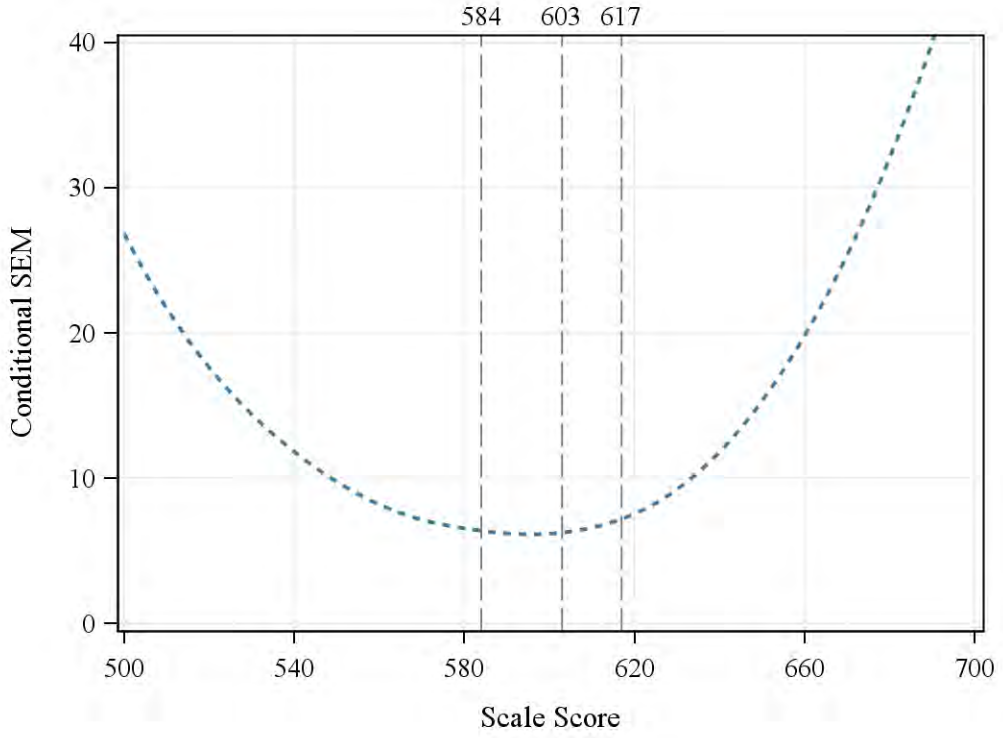


Figure 6.13. Mathematics Grade 3 TCC

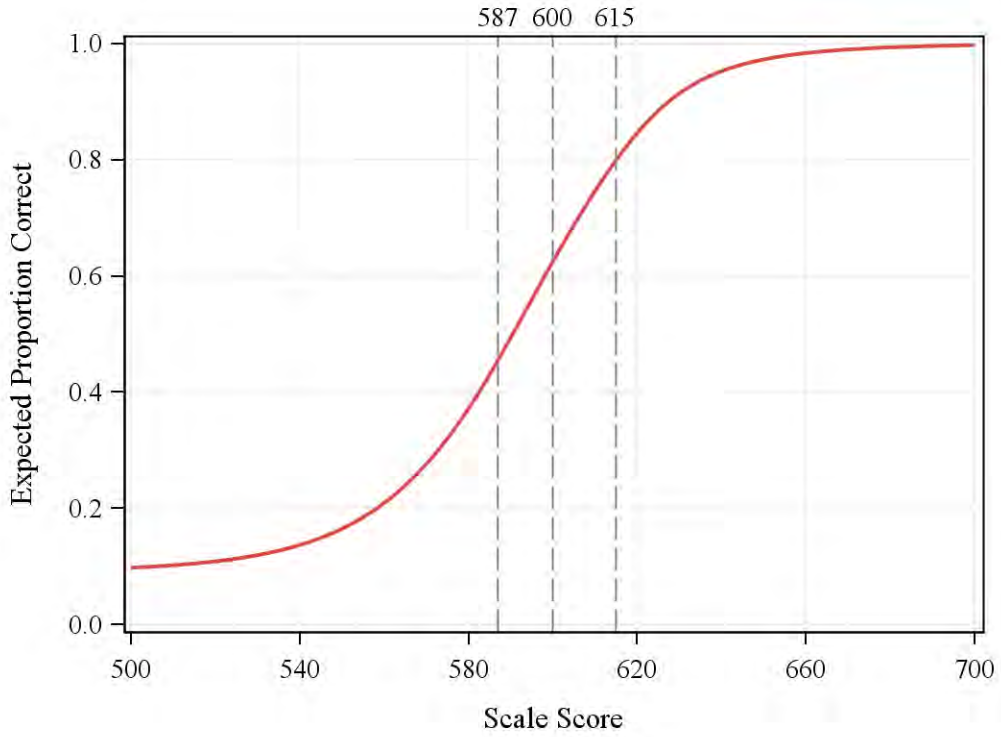


Figure 6.14. Mathematics Grade 3 CSEM Curve

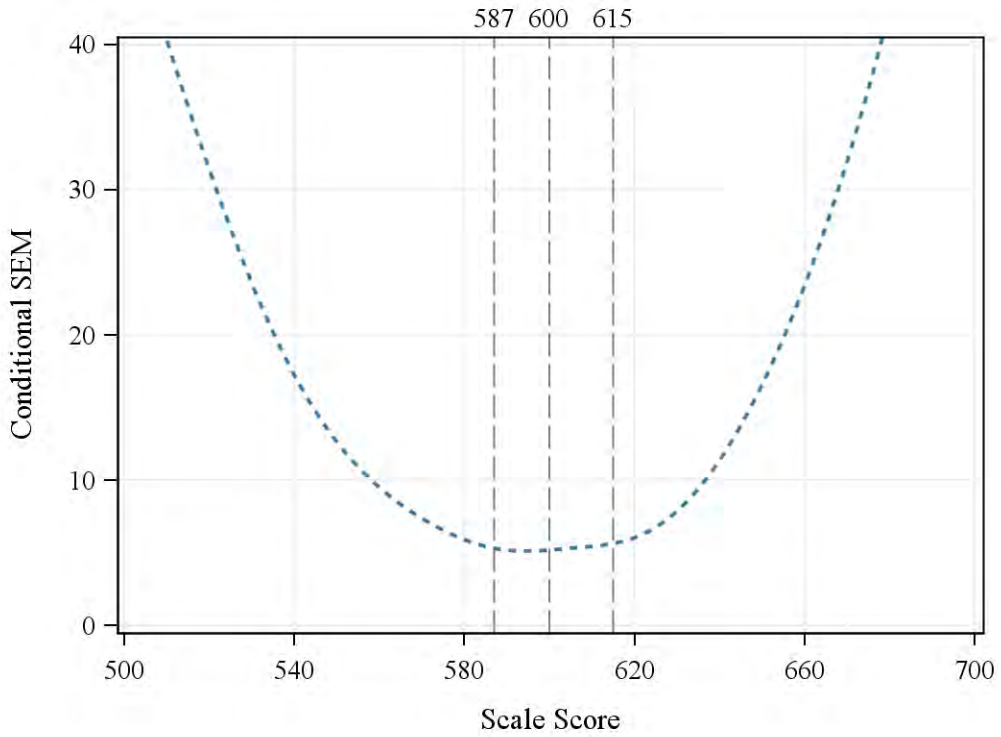


Figure 6.15. Mathematics Grade 4 TCC

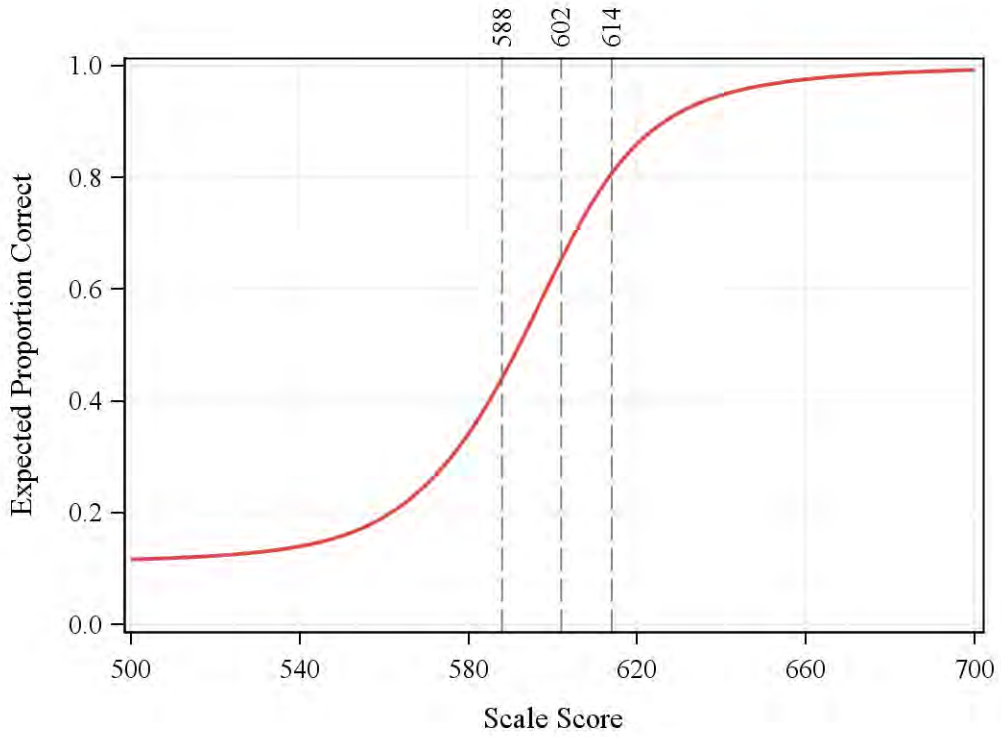


Figure 6.16. Mathematics Grade 4 CSEM Curve

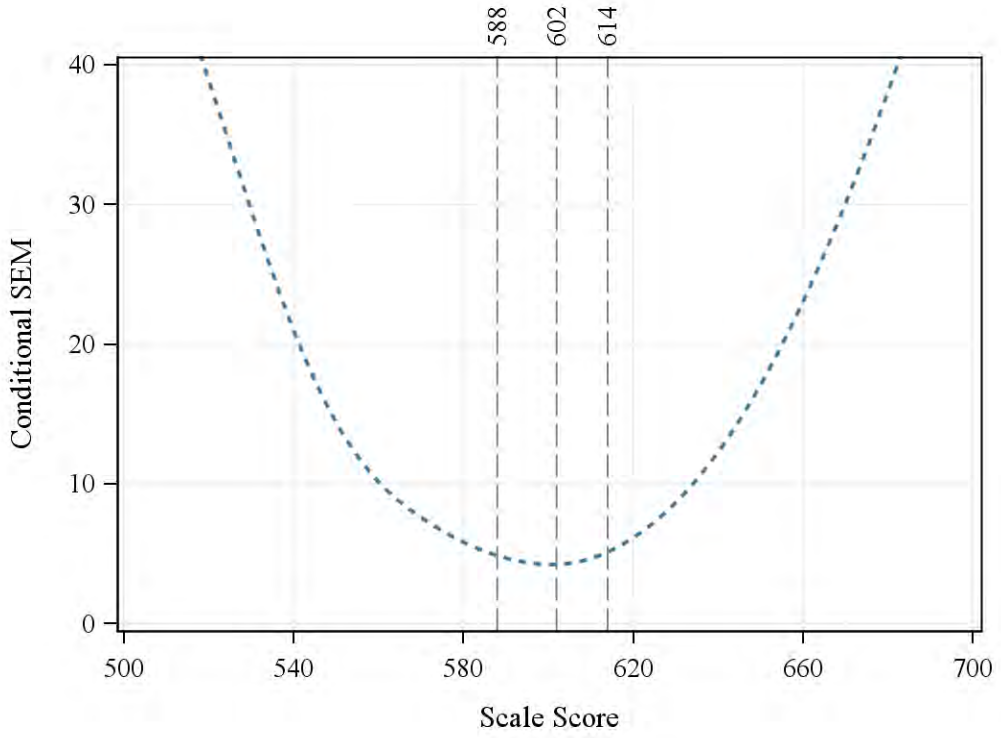


Figure 6.17. Mathematics Grade 5 TCC

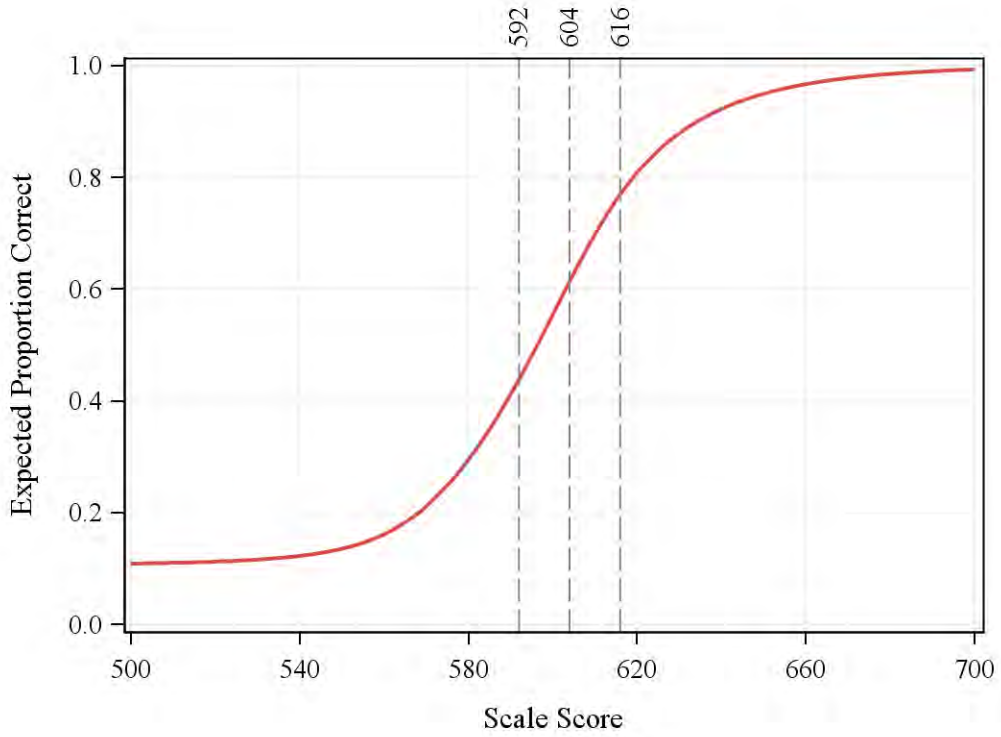


Figure 6.18. Mathematics Grade 5 CSEM Curve

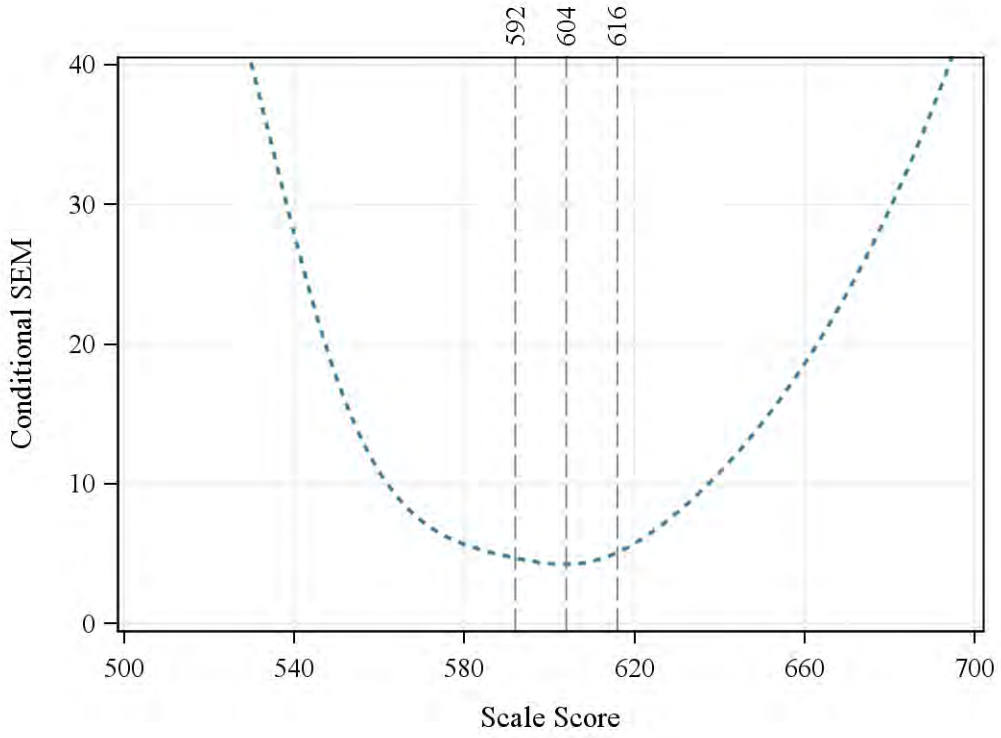


Figure 6.19. Mathematics Grade 6 TCC

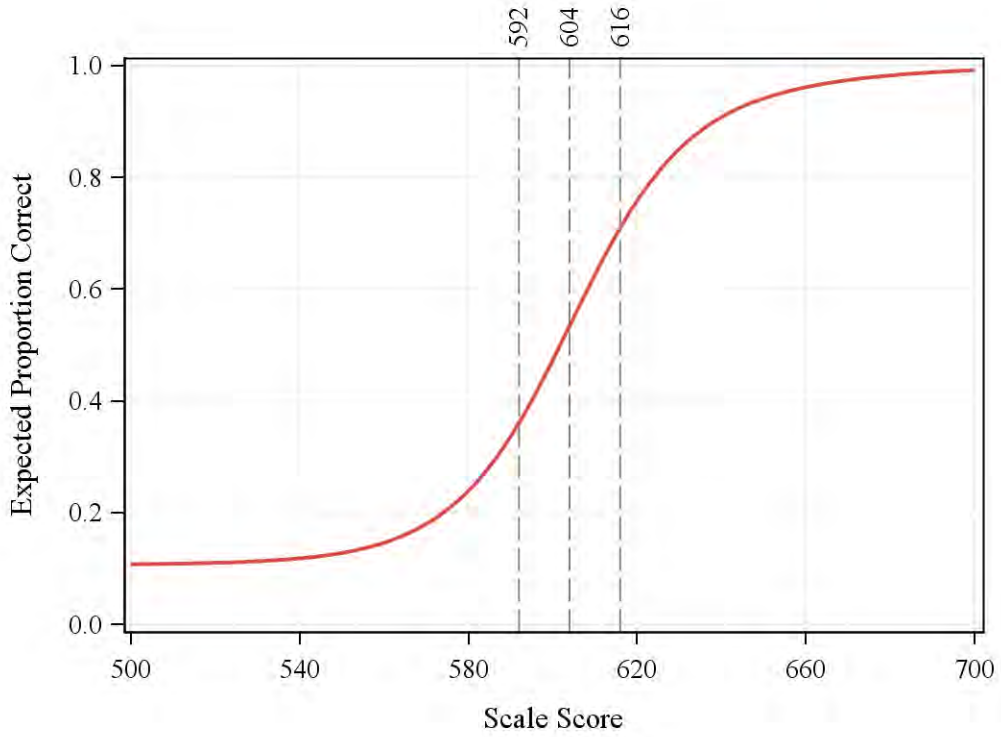


Figure 6.20. Mathematics Grade 6 CSEM Curve

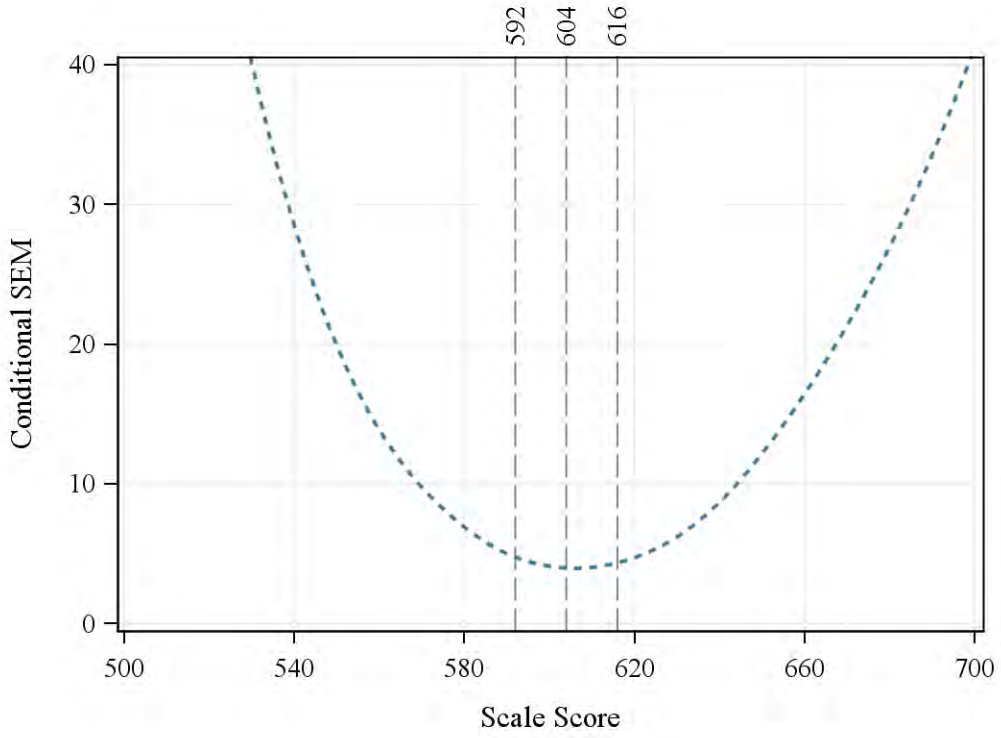


Figure 6.21. Mathematics Grade 7 TCC

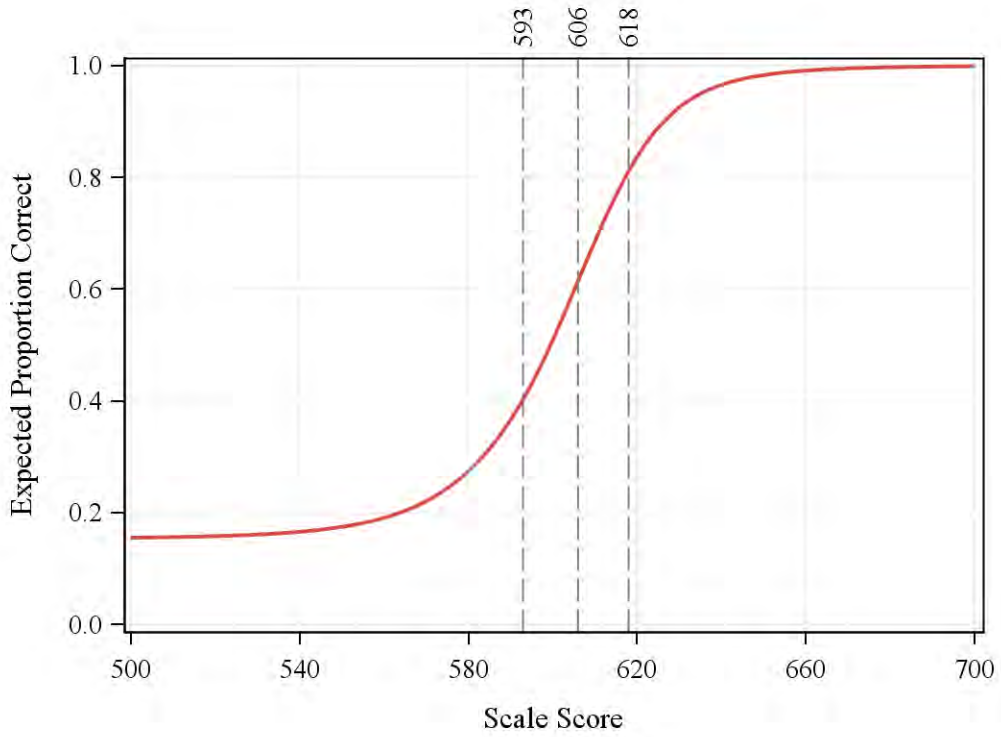


Figure 6.22. Mathematics Grade 7 CSEM Curve

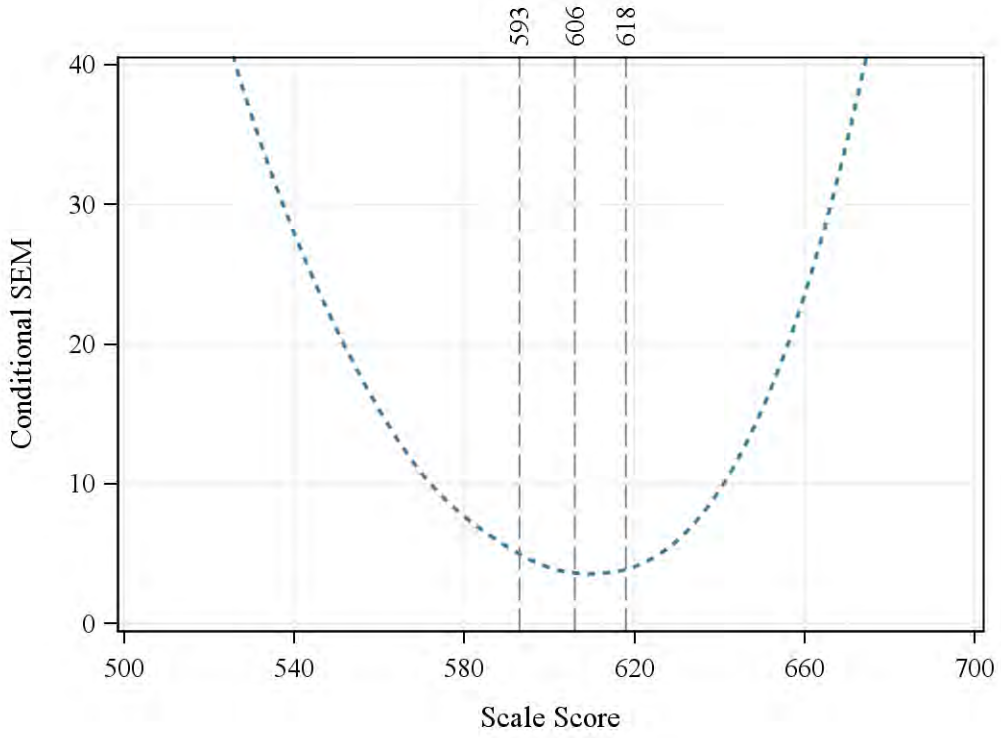


Figure 6.23. Mathematics Grade 8 TCC

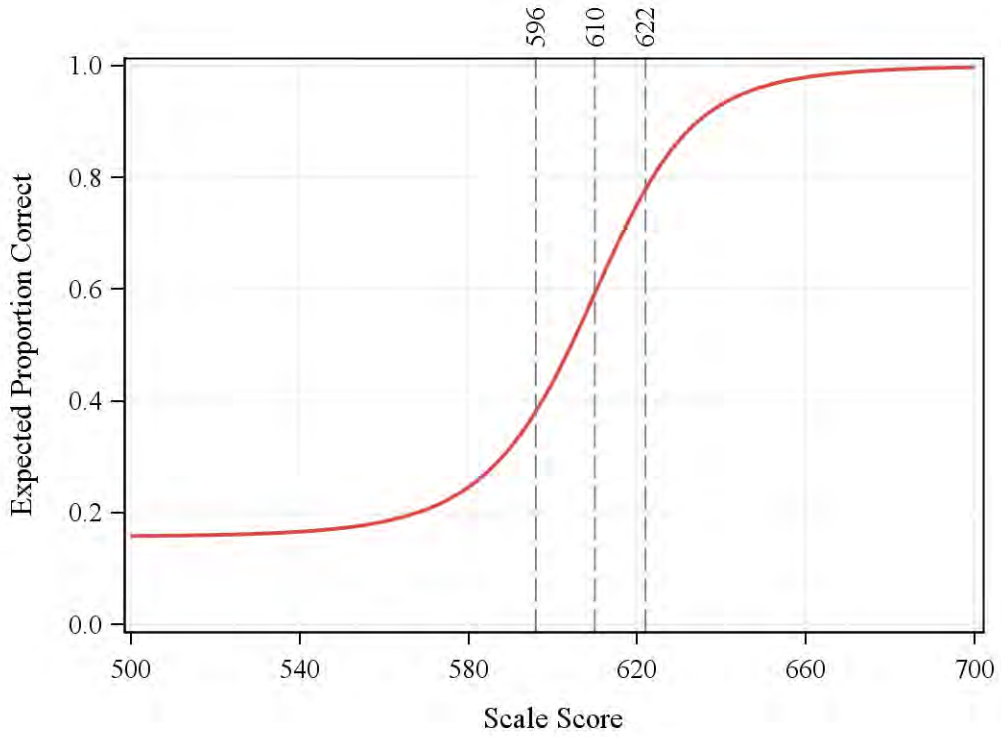
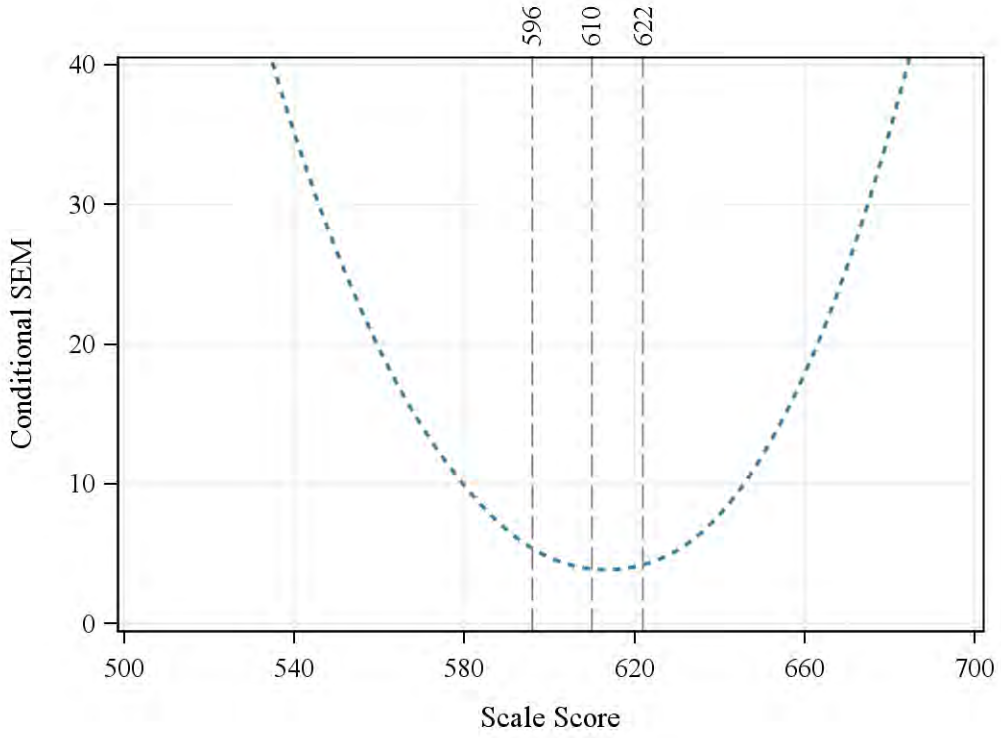


Figure 6.24. Mathematics Grade 8 CSEM Curve



6.7. Scoring Procedure

New York State student examinations were scored using the number correct (NC) scoring method. This method considers how many score points that a student obtained on a test in determining his or her scale score. That is, two students with the same number of score points on the test will receive the same scale score, regardless of which items they answered correctly. In this method, the number correct (or raw) score on the test is converted to a scale score by means of a conversion table. This traditional scoring method is often preferred for its conceptual simplicity and familiarity.

The final item parameters were used to calculate the raw-score-to-theta tables, using a TCC method (see the details provided below). The obtained scaling transformation intercept and slope (M_1^S and M_2^S) were then applied to the theta values to produce raw score-to-scale score-conversion tables for the Grades 3–8 ELA and mathematics Tests.

An inverse TCC method was employed using POLYEQUATE (Kolen & Cui, 2004). The inverse of the TCC procedure produces trait values (i.e., proficiency) based on unweighted raw scores. These estimates show negligible statistical bias (defined in statistics as the difference between an estimator's expected value and the true value of the parameter being estimated) for tests with maximum possible raw scores of at least 30 points. All NYSTP ELA and mathematics tests have a maximum raw score higher than 30 points. In the inverse TCC method, a student's trait (i.e., proficiency) estimate is taken to be the trait value that has an expected raw score equal to the student's observed raw score. It was found that, for tests containing only MC items, the inverse of the TCC is an excellent first-order approximation of the number of correct maximum likelihood estimates (MLE) showing negligible bias for tests of at least 30 items. For tests with a mixture of MC and CR items, the MLE and TCC estimates are even more similar (Yen, 1984).

The inverse of the TCC method relies on the following equation:

$$\sum_{i=1}^{n} v_i x_i = \sum_{i=1}^{n} v_i E(X_i | \tilde{\theta}),$$

where:

x_i is a student's observed raw score on item *i,*

v_i is a non-optimal weight specified in a scoring process (v_i = 1 if no weights are specified), and

$\tilde{\theta}$ is a trait estimate.

Potential differences in test form difficulty at different performance levels are accounted for in the resulting raw score-to-scale score conversion tables, so that students of the same proficiency are expected to obtain the same scale score, regardless of which form they took.

6.7.1. *Raw Score-to-Scale Score and SEM Conversion Tables*

The scale score is the basic score for the NYSTP. Raw score-to-scale score (RSSS) conversion tables based on the total number correct are presented in Appendix Q, Tables Q1–Q12.

The standard error (SE) of a scale score indicates the precision with which the proficiency is estimated, and it inversely is related to the amount of information provided by the test at each performance level. The SE is estimated as follows:

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}},$$

where

$SE(\hat{\theta})$ is the standard error of the scale score (theta).

$I\hat{\theta}$ is the amount of information provided by the test at a given performance level.

The information is estimated based on thetas in the scale score metric; therefore, the SE is also expressed in the scale score metric. The SE value varies across performance levels and is the highest at the extreme ends of the scale where the amount of test information is typically the lowest. The final element of the raw score-to-scale score tables is the application of the performance level cut scores.

New scale score cuts were set this summer in 2018 and therefore, it was not necessary to perform any linking to the previous scale. See Section 8 and Appendix T for more information on the standards review process.

Table 6.10 and Table 6.11 present scale score ranges associated with each performance level for ELA and Mathematics, respectively.

Table 6.10. ELA Scale Score Ranges Associated with Each Performance Level

Testing Platform	Grade	NYS Level 1	NYS Level 2	NYS Level 3	NYS Level 4
PBT	3	530-580	581-600	601-626	627-655
	4	532-581	582-601	602-616	617-654
	5	509-592	593-606	607-619	620-661
	6	514-588	589-601	602-611	612-657
	7	511-589	590-605	606-620	621-654
	8	507-582	583-601	602-614	615-651
CBT	3	531-581	582-601	602-627	628-655
	4	533-582	583-602	603-617	618-654
	5	510-593	594-607	608-620	621-661
	6	515-589	590-600	601-612	613-657
	7	512-590	591-606	607-621	622-654
	8	508-583	584-602	603-615	616-651

Table 6.10. Mathematics Scale Score Ranges Associated with Each Performance Level

Testing Platform	Grade	NYS Level 1	NYS Level 2	NYS Level 3	NYS Level 4
PBT	3	526-586	587-598	599-613	614-646
	4	525-587	588-601	602-612	613-650
	5	527-590	591-602	603-615	616-654
	6	528-590	591-603	604-615	616-656
	7	524-591	592-605	606-617	618-644
	8	527-594	595-609	610-620	621-651
CBT	3	527-585	586-599	600-614	615-646
	4	526-586	587-600	601-613	614-650
	5	527-590	591-602	603-615	616-654
	6	530-590	591-603	604-614	615-656
	7	524-591	592-605	606-617	618-644
	8	528-595	596-609	610-621	622-651

A mode comparability study was completed to identify whether or not there were any differences in student performance that could be attributed to the mode of test administration (i.e. PBT versus CBT). The main inference to be drawn from the mode comparability study is whether scores that arise from students testing on paper or on computer are interchangeable. A propensity score matching approach was conducted to generate the CBT and PBT samples that were comparable on covariates that may affect student performance, aside from the test mode itself (e.g., gender, school-type, previous performance). The difference in students' test scores were computed between the matched CBT and PBT samples to evaluate test-level mode comparability, and mode adjustments were made accordingly. Please see Appendix R (the mode comparability report) and Appendix S (the NYSED memorandum on the mode comparability results) for more details.

Section 7: Reliability and Standard Error of Measurement

This section presents specific information on various test reliability statistics and standard error of measurement (SEM), as well as the results from a study of performance level classification accuracy and consistency. The data set for these studies includes all tested New York State students who received valid scores.

7.1. Test Reliability

Test reliability is directly related to score stability and standard error and, as such, is an essential element of fairness and validity. Test reliability can be directly measured with an alpha statistic, or the alpha statistic can be used to derive the SEM. For the Grades 3–8 ELA and Mathematics Tests, Questar calculated two types of reliability statistics: Cronbach's alpha (Cronbach, 1951) and Feldt-Raju coefficient (Qualls, 1995). These two measures are appropriate for assessment of a test's internal consistency when a single test is administered to a group of examinees on one occasion. The reliability of the test is then estimated by considering how well the items that reflect the same construct yield similar results (or how consistent the results are for different items that reflect the same construct measured by the test). Both Cronbach's alpha and Feldt-Raju coefficient measures are appropriate for tests of multiple-item formats (MC and CR items).

7.1.1. *Test Statistics and Reliability for Total Test*

Table 7.1 and Table 7.3 present the test statistics including raw-score (RS) means and raw-score standard deviations (SDs) for ELA and Mathematics, respectively. These statistics give the necessary context for Table 7.2 and Table 7.4, which present the case counts (n-count), number of test items (# Items), Cronbach's alpha and associated SEM, and Feldt-Raju coefficient and associated SEM obtained for the total ELA and Mathematics tests. Reliability coefficients provide measures of internal consistency that range from zero to one. High reliability indicates that scores are consistent and not unduly influenced by random error. Overall test reliability is a very good indication of each test's internal consistency.

Grades 3–8 ELA reliability estimates (Cronbach's alpha and Feldt-Raju) ranged from 0.87 to 0.89. Grades 3–8 Mathematics reliability estimates (Cronbach's alpha and Feldt-Raju) ranged from 0.91 to 0.94. The reliabilities are similar across grades and slightly higher for the Mathematics tests than for the ELA tests. All reliabilities were at least 0.87 across all grades and both subjects, which is a good indication that the NYSTP Grades 3–8 ELA and Mathematics Tests are acceptably reliable.

Table 7.1. ELA Test Form Statistics

	Item-Level			Student-Level			
	p-value				Raw Score		
Grade	Mean	Min.	Max.	N-Count	Max.	Mean	SD
3	0.56	0.38	0.91	179,339	34	18.54	6.84
4	0.59	0.39	0.81	181,672	34	19.46	7.06
5	0.61	0.33	0.93	175,175	44	26.82	7.94
6	0.64	0.36	0.87	170,015	44	27.91	8.70
7	0.62	0.37	0.87	155,919	46	28.85	9.15
8	0.67	0.41	0.90	151,522	46	30.74	8.75

Table 7.2. ELA Test Reliability and Standard Error of Measurement

			Raw Score	Cronbach's Alpha		Feldt-Raju Coefficient	
Grade	N-Count	Items	Points	Est.	SEM	Est.	SEM
3	179,339	25	34	0.87	2.48	0.88	2.38
4	181,672	25	34	0.87	2.56	0.88	2.43
5	175,175	35	44	0.87	2.84	0.88	2.75
6	170,015	35	44	0.89	2.87	0.90	2.73
7	155,919	36	46	0.89	2.98	0.90	2.83
8	151,522	36	46	0.89	2.89	0.90	2.73

Table 7.3. Mathematics Test Form Statistics

	Item-Level			Student-Level			
	p-value				Raw Score		
Grade	Mean	Min.	Max.	N-Count	Max.	Mean	SD
3	0.63	0.31	0.94	176,663	42	25.73	9.94
4	0.62	0.35	0.90	176,897	46	27.79	11.26
5	0.59	0.33	0.90	168,578	46	25.46	11.15
6	0.53	0.21	0.82	164,429	48	23.56	11.50
7	0.55	0.33	0.87	151,749	50	27.28	12.91
8	0.51	0.22	0.79	108,410	50	24.27	12.05

Table 7.4. Mathematics Test Reliability and Standard Error of Measurement

			Raw Score	Cronbach's Alpha		Feldt-Raju Coefficient	
Grade	N-Count	Items	Points	Est.	SEM	Est.	SEM
3	176,663	34	42	0.91	2.93	0.92	2.74
4	176,897	38	46	0.93	3.02	0.94	2.86
5	168,578	38	46	0.93	2.99	0.94	2.84

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
6	164,429	39	48	0.93	2.99	0.94	2.87
7	151,749	41	50	0.94	3.24	0.94	3.04
8	108,410	41	50	0.92	3.39	0.93	3.22

7.1.2. *Reliability of MC Items*

In addition to overall test reliability, Cronbach's alpha and Feldt-Raju coefficient were computed separately for MC and CR item sets. It is important to recognize that reliability is directly affected by test length; therefore, reliability estimates for tests by item type will always be lower than reliability estimates for the overall test form. Table 7.5 and Table 7.6 present reliabilities for the subsets of MC items.

Table 7.5. ELA MC Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
3	179,339	18	0.79	1.85	0.79	1.84
4	181,672	18	0.77	1.86	0.77	1.86
5	175,175	28	0.83	2.25	0.83	2.24
6	170,015	28	0.83	2.25	0.84	2.23
7	155,919	28	0.82	2.33	0.82	2.31
8	151,522	28	0.83	2.19	0.83	2.18

Table 7.6. Mathematics MC Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
3	176,663	27	0.88	2.07	0.88	2.05
4	176,897	31	0.91	2.25	0.91	2.24
5	168,578	31	0.91	2.24	0.91	2.23
6	164,429	31	0.91	2.29	0.91	2.28
7	151,749	33	0.90	2.44	0.91	2.43
8	108,410	33	0.88	2.52	0.88	2.51

7.1.3. *Reliability of CR Items*

Reliability coefficients were also computed for the subsets of CR items. The results are presented in Table 7.7 and Table 7.8.

Table 7.7. ELA CR Item Reliability and Standard Error of Measurement

			Raw Score	Cronbach's Alpha		Feldt-Raju Coefficient	
Grade	N-Count	Items	Points	Est.	SEM	Est.	SEM
3	179,339	7	16	0.83	1.47	0.85	1.40
4	181,672	7	16	0.84	1.54	0.86	1.47
5	175,175	7	16	0.78	1.56	0.81	1.49
6	170,015	7	16	0.84	1.54	0.86	1.47
7	155,919	8	18	0.87	1.58	0.89	1.50
8	151,522	8	18	0.85	1.64	0.87	1.54

Note. Results should be interpreted with caution because the number of items is low.

Table 7.8. Mathematics CR Item Reliability and Standard Error of Measurement

			Raw Score	Cronbach's Alpha		Feldt-Raju Coefficient	
Grade	N-Count	Items	Points	Est.	SEM	Est.	SEM
3	176,663	7	15	0.83	1.86	0.84	1.81
4	176,897	7	15	0.82	1.84	0.83	1.77
5	168,578	7	15	0.83	1.78	0.84	1.73
6	164,429	8	17	0.85	1.74	0.86	1.72
7	151,749	8	17	0.89	1.82	0.89	1.79
8	108,410	8	17	0.85	2.02	0.85	1.99

Note. Results should be interpreted with caution because the number of items is low.

7.1.4. *Test Reliability for Subgroups*

In this section, reliability coefficients that were estimated for the population and subgroups are presented. The reporting subgroups include the following: gender, ethnicity, NRC, ELL/MLL, all SWD, all SUA, SWD/SUA (includes examinees who are classified as having a disability and who use at least one disability-related accommodation), and English language learners/Multi Language Learners using accommodations specific to their ELL/MLL status (ELL/MLL/SUA). Accommodations available to students include the following: Flexibility in Scheduling/Timing, Flexibility in Setting, Method of Presentation (excluding Braille), Method of Response, Braille and Large-type, and others. Accommodations available to English language learners are Separate Location, and Bilingual Dictionaries and Glossaries.

As shown in Tables 7.9–7.14 and Tables 7.15–7.20 for ELA and Mathematics, respectively, the estimated reliabilities for subgroups were close in magnitude to the test reliability estimates of the population. Cronbach's alpha reliability coefficients were all at least 0.75. Feldt-Raju reliability coefficients, which tend to be larger than the Cronbach's alpha estimates for the same group, were at least 0.76. These indicate a very good test internal consistency (reliability) for analyzed subgroups of examinees.

Table 7.9. ELA Grade 3 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
State	All Items	179,339	0.87	2.48	0.88	2.38
Gender	Female	88,724	0.87	2.47	0.88	2.37
	Male	90,615	0.87	2.49	0.88	2.38
Ethnicity	Asian	17,785	0.87	2.42	0.88	2.30
	Black	31,318	0.87	2.51	0.88	2.40
	Hispanic	50,296	0.86	2.50	0.87	2.40
	American Indian	1,244	0.87	2.51	0.88	2.39
	Multiracial	5,148	0.87	2.45	0.88	2.34
	Pacific Islander	422	0.87	2.49	0.88	2.37
	White	71,082	0.86	2.46	0.87	2.36
NRC	New York	66,509	0.88	2.49	0.89	2.37
	Big 4 Cities	7,735	0.86	2.48	0.87	2.38
	Urban/Suburban	14,213	0.85	2.48	0.86	2.40
	Rural	9,864	0.85	2.48	0.85	2.40
	Average Needs	42,241	0.85	2.45	0.86	2.37
	Low Needs	18,151	0.84	2.37	0.85	2.29
	Charter School	12,101	0.86	2.49	0.87	2.39
	Religious and Independent	8,525	0.86	2.54	0.88	2.40
SWD	All Codes	25,142	0.84	2.49	0.85	2.40
SUA	All Codes	22,645	0.82	2.47	0.83	2.40
ELL/MLL	ELL=Y	20,595	0.80	2.51	0.82	2.41
SWD/SUA	SWD & SUA codes	18,977	0.82	2.48	0.83	2.40
ELL/MLL/SUA	SUA & ELL codes	3,762	0.77	2.46	0.78	2.38

Table 7.10. ELA Grade 4 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
State	All Items	181,672	0.87	2.56	0.88	2.43
Gender	Female	89,676	0.86	2.56	0.88	2.43
	Male	91,996	0.87	2.55	0.88	2.43
Ethnicity	Asian	18,533	0.87	2.45	0.88	2.33
	Black	32,133	0.86	2.59	0.88	2.46
	Hispanic	50,017	0.86	2.58	0.87	2.47
	American Indian	1,258	0.86	2.61	0.88	2.46
	Multiracial	4,731	0.88	2.53	0.89	2.39
	Pacific Islander	501	0.87	2.54	0.88	2.41

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
Ethnicity	White	72,411	0.86	2.53	0.87	2.41
NRC	New York	66,945	0.88	2.55	0.89	2.40
	Big 4 Cities	7,754	0.86	2.56	0.87	2.44
	Urban/Suburban	13,395	0.85	2.59	0.86	2.48
	Rural	9,820	0.85	2.55	0.86	2.46
	Average Needs	40,780	0.85	2.54	0.86	2.44
	Low Needs	18,128	0.82	2.43	0.84	2.34
	Charter School	11,288	0.84	2.51	0.85	2.42
	Religious and Independent	13,562	0.87	2.63	0.88	2.46
SWD	All Codes	26,145	0.84	2.55	0.85	2.46
SUA	All Codes	25,266	0.83	2.54	0.84	2.46
ELL/MLL	ELL=Y	17,497	0.80	2.60	0.81	2.50
SWD/SUA	SWD & SUA codes	21,075	0.83	2.54	0.84	2.46
ELL/MLL/SUA	SUA & ELL codes	3,692	0.76	2.51	0.77	2.45

Table 7.11. ELA Grade 5 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	175,175	0.87	2.84	0.88	2.75
Gender	Female	86,784	0.86	2.80	0.87	2.72
	Male	88,391	0.88	2.86	0.88	2.76
Ethnicity	Asian	18,643	0.87	2.68	0.88	2.60
	Black	31,523	0.87	2.89	0.87	2.80
	Hispanic	48,692	0.86	2.88	0.87	2.79
	American Indian	1,250	0.85	2.90	0.86	2.81
	Multiracial	4,256	0.88	2.81	0.89	2.70
	Pacific Islander	537	0.86	2.77	0.87	2.67
	White	68,391	0.87	2.81	0.88	2.71
NRC	New York	67,866	0.88	2.83	0.89	2.73
	Big 4 Cities	7,501	0.87	2.93	0.88	2.83
	Urban/Suburban	12,439	0.86	2.88	0.87	2.80
	Rural	9,295	0.85	2.86	0.86	2.79
	Average Needs	39,116	0.85	2.81	0.86	2.73
	Low Needs	18,282	0.83	2.67	0.83	2.61
	Charter School	11,148	0.85	2.79	0.85	2.73
	Religious and Independent	9,528	0.89	2.91	0.90	2.76
SWD	All Codes	26,527	0.84	2.92	0.85	2.85

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
SUA	All Codes	25,920	0.84	2.92	0.85	2.85
ELL/MLL	ELL=Y	14,651	0.79	2.96	0.80	2.87
SWD/SUA	SWD & SUA codes	21,681	0.83	2.92	0.84	2.85
ELL/MLL/SUA	SUA & ELL codes	3,585	0.75	2.92	0.76	2.85

Table 7.12. ELA Grade 6 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	170,015	0.89	2.87	0.90	2.73
Gender	Female	83,617	0.88	2.81	0.89	2.68
	Male	86,398	0.89	2.90	0.90	2.76
Ethnicity	Asian	18,003	0.89	2.66	0.90	2.54
	Black	31,314	0.88	2.94	0.89	2.81
	Hispanic	46,768	0.88	2.92	0.89	2.80
	American Indian	1,141	0.89	2.91	0.90	2.77
	Multiracial	3,714	0.89	2.82	0.90	2.66
	Pacific Islander	614	0.89	2.79	0.90	2.66
	White	66,335	0.89	2.82	0.90	2.67
NRC	New York	64,138	0.90	2.84	0.91	2.70
	Big 4 Cities	6,856	0.89	2.95	0.90	2.83
	Urban/Suburban	11,921	0.88	2.97	0.89	2.84
	Rural	8,994	0.88	2.92	0.89	2.79
	Average Needs	36,469	0.88	2.85	0.89	2.72
	Low Needs	17,522	0.85	2.66	0.86	2.55
	Charter School	11,389	0.86	2.84	0.87	2.76
	Religious and Independent	12,726	0.90	2.96	0.91	2.76
SWD	All Codes	25,249	0.86	2.98	0.87	2.88
SUA	All Codes	23,977	0.86	2.98	0.87	2.88
ELL/MLL	ELL=Y	13,503	0.82	3.01	0.84	2.90
SWD/SUA	SWD & SUA codes	19,801	0.85	2.98	0.86	2.88
ELL/MLL/SUA	SUA & ELL codes	3,081	0.79	2.95	0.80	2.87

Table 7.13. ELA Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	155,919	0.89	2.98	0.90	2.83
Gender	Female	75,962	0.88	2.90	0.89	2.78

Demographic Category		N-Count	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
Gender	Male	79,957	0.90	3.02	0.91	2.86
Ethnicity	Asian	17,046	0.90	2.72	0.91	2.59
	Black	29,642	0.88	3.06	0.89	2.90
	Hispanic	42,405	0.88	3.03	0.89	2.88
	American Indian	1,209	0.89	3.03	0.90	2.85
	Multiracial	2,969	0.90	2.95	0.91	2.78
	Pacific Islander	461	0.89	2.97	0.91	2.77
	White	60,427	0.89	2.95	0.90	2.79
NRC	New York	64,280	0.90	2.91	0.91	2.76
	Big 4 Cities	6,366	0.89	3.12	0.90	2.93
	Urban/Suburban	10,852	0.88	3.10	0.89	2.94
	Rural	8,368	0.88	3.05	0.89	2.91
	Average Needs	32,952	0.88	2.99	0.89	2.85
	Low Needs	17,060	0.87	2.80	0.88	2.69
	Charter School	10,518	0.85	2.92	0.86	2.84
	Religious and Independent	5,523	0.92	3.14	0.93	2.88
SWD	All Codes	24,303	0.85	3.09	0.87	2.96
SUA	All Codes	23,095	0.86	3.09	0.87	2.96
ELL/MLL	ELL=Y	11,401	0.82	3.11	0.84	2.95
SWD/SUA	SWD & SUA codes	19,200	0.85	3.09	0.86	2.96
ELL/MLL/SUA	SUA & ELL codes	2,547	0.77	3.05	0.78	2.93

Table 7.14. ELA Grade 8 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
State	All Items	151,522	0.89	2.89	0.90	2.73
Gender	Female	73,680	0.88	2.81	0.89	2.67
	Male	77,842	0.90	2.93	0.91	2.77
Ethnicity	Asian	17,516	0.89	2.64	0.90	2.49
	Black	29,158	0.88	2.95	0.89	2.80
	Hispanic	41,041	0.88	2.93	0.89	2.79
	American Indian	1,169	0.88	2.93	0.89	2.77
	Multiracial	2,372	0.90	2.86	0.91	2.68
	Pacific Islander	456	0.89	2.83	0.91	2.65
	White	58,332	0.89	2.86	0.90	2.69
NRC	New York	62,273	0.89	2.83	0.90	2.68
	Big 4 Cities	6,205	0.89	3.02	0.90	2.87

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
NRC	Urban/Suburban	9,428	0.89	3.00	0.90	2.84
	Rural	7,901	0.88	2.94	0.89	2.81
	Average Needs	29,532	0.89	2.89	0.90	2.74
	Low Needs	15,829	0.87	2.73	0.88	2.59
	Charter School	9,557	0.84	2.81	0.85	2.72
	Religious and Independent	10,797	0.90	2.97	0.91	2.76
SWD	All Codes	22,452	0.86	3.01	0.87	2.90
SUA	All Codes	21,233	0.86	3.01	0.87	2.90
ELL/MLL	ELL=Y	10,941	0.84	3.02	0.85	2.90
SWD/SUA	SWD & SUA codes	17,670	0.86	3.01	0.86	2.91
ELL/MLL/SUA	SUA & ELL codes	2,303	0.79	2.99	0.81	2.90

Table 7.15. Mathematics Grade 3 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	176,663	0.91	2.93	0.92	2.74
Gender	Female	87,245	0.91	2.92	0.92	2.74
	Male	89,418	0.92	2.94	0.93	2.74
Ethnicity	Asian	18,053	0.90	2.68	0.91	2.49
	Black	30,369	0.92	2.95	0.93	2.77
	Hispanic	50,760	0.91	2.97	0.92	2.80
	American Indian	1,228	0.91	2.95	0.92	2.78
	Multiracial	5,002	0.91	2.91	0.93	2.71
	Pacific Islander	422	0.91	2.90	0.92	2.70
	White	69,246	0.90	2.91	0.91	2.73
NRC	New York	67,143	0.92	2.93	0.93	2.72
	Big 4 Cities	6,623	0.91	2.94	0.92	2.78
	Urban/Suburban	14,296	0.91	2.98	0.92	2.82
	Rural	9,898	0.90	2.98	0.91	2.82
	Average Needs	42,171	0.90	2.95	0.91	2.77
	Low Needs	18,175	0.88	2.80	0.90	2.63
	Charter School	11,797	0.91	2.70	0.92	2.50
	Religious and Independent	6,560	0.89	3.03	0.91	2.82
SWD	All Codes	23,959	0.91	2.94	0.92	2.79
SUA	All Codes	22,053	0.90	2.93	0.91	2.80
ELL/MLL	ELL=Y	22,056	0.90	2.96	0.91	2.82
SWD/SUA	SWD & SUA codes	18,686	0.90	2.92	0.91	2.80

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
ELL/MLL/SUA	SUA & ELL codes	3,990	0.89	2.89	0.90	2.78

Table 7.16. Mathematics Grade 4 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	176,897	0.93	3.02	0.94	2.86
Gender	Female	86,958	0.93	3.04	0.93	2.87
	Male	89,939	0.93	3.00	0.94	2.84
Ethnicity	Asian	18,806	0.93	2.67	0.93	2.53
	Black	30,994	0.93	3.10	0.93	2.94
	Hispanic	50,271	0.92	3.10	0.93	2.95
	American Indian	1,173	0.92	3.08	0.93	2.91
	Multiracial	4,664	0.93	2.99	0.94	2.81
	Pacific Islander	497	0.92	3.00	0.93	2.83
	White	68,820	0.92	2.96	0.93	2.82
NRC	New York	67,318	0.93	3.02	0.94	2.84
	Big 4 Cities	6,446	0.92	3.06	0.93	2.92
	Urban/Suburban	13,846	0.92	3.10	0.93	2.96
	Rural	9,841	0.92	3.09	0.93	2.95
	Average Needs	40,750	0.92	3.03	0.92	2.88
	Low Needs	18,169	0.91	2.80	0.92	2.67
	Charter School	10,942	0.93	2.81	0.94	2.64
	Religious and Independent	9,585	0.91	3.13	0.92	2.98
SWD	All Codes	24,783	0.91	3.06	0.92	2.94
SUA	All Codes	24,286	0.90	3.06	0.91	2.95
ELL/MLL	ELL=Y	18,448	0.90	3.10	0.91	2.98
SWD/SUA	SWD & SUA codes	20,512	0.90	3.05	0.91	2.94
ELL/MLL/SUA	SUA & ELL codes	3,851	0.88	3.03	0.89	2.94

Table 7.17. Mathematics Grade 5 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	168,578	0.93	2.99	0.94	2.84
Gender	Female	83,182	0.92	3.00	0.93	2.85
	Male	85,396	0.93	2.98	0.94	2.83
Ethnicity	Asian	18,740	0.92	2.84	0.93	2.62
	Black	29,957	0.92	2.97	0.93	2.87

Demographic Category		N-Count	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
Ethnicity	Hispanic	48,437	0.92	2.99	0.92	2.88
	American Indian	1,223	0.92	3.02	0.93	2.89
	Multiracial	4,069	0.93	2.98	0.94	2.82
	Pacific Islander	544	0.92	2.98	0.93	2.81
	White	64,450	0.92	2.99	0.93	2.84
NRC	New York	67,758	0.93	2.99	0.94	2.82
	Big 4 Cities	5,963	0.92	2.88	0.92	2.81
	Urban/Suburban	12,928	0.92	2.96	0.92	2.86
	Rural	9,148	0.91	3.01	0.92	2.90
	Average Needs	38,460	0.92	3.01	0.92	2.87
	Low Needs	18,029	0.90	2.92	0.91	2.76
	Charter School	10,601	0.92	2.98	0.93	2.81
	Religious and Independent	5,691	0.91	3.05	0.92	2.91
SWD	All Codes	24,675	0.91	2.86	0.91	2.80
SUA	All Codes	23,725	0.90	2.86	0.90	2.81
ELL/MLL	ELL=Y	14,916	0.90	2.83	0.90	2.79
SWD/SUA	SWD & SUA codes	20,589	0.89	2.83	0.90	2.79
ELL/MLL/SUA	SUA & ELL codes	3,604	0.86	2.74	0.86	2.71

Table 7.18. Mathematics Grade 6 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
State	All Items	164,429	0.93	2.99	0.94	2.87
Gender	Female	80,609	0.93	3.01	0.93	2.89
	Male	83,820	0.94	2.96	0.94	2.85
Ethnicity	Asian	18,177	0.93	2.94	0.94	2.75
	Black	30,094	0.92	2.90	0.92	2.84
	Hispanic	46,282	0.92	2.93	0.92	2.87
	American Indian	1,143	0.93	2.96	0.93	2.87
	Multiracial	3,622	0.94	3.00	0.94	2.87
	Pacific Islander	605	0.93	3.02	0.94	2.87
	White	63,244	0.92	3.02	0.93	2.90
NRC	New York	63,931	0.94	2.97	0.94	2.85
	Big 4 Cities	5,499	0.92	2.85	0.92	2.79
	Urban/Suburban	12,070	0.92	2.88	0.92	2.82
	Rural	8,795	0.92	2.98	0.92	2.90
	Average Needs	36,022	0.92	3.02	0.93	2.91

Demographic Category		N-Count	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
NRC	Low Needs	17,313	0.92	2.98	0.92	2.84
	Charter School	11,117	0.93	2.95	0.93	2.85
	Religious and Independent	9,682	0.92	3.06	0.92	2.97
SWD	All Codes	23,672	0.90	2.76	0.90	2.73
SUA	All Codes	22,922	0.89	2.77	0.90	2.74
ELL/MLL	ELL=Y	14,402	0.89	2.77	0.89	2.75
SWD/SUA	SWD & SUA codes	19,225	0.88	2.74	0.88	2.71
ELL/MLL/SUA	SUA & ELL codes	3,263	0.83	2.65	0.83	2.63

Table 7.19. Mathematics Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha Est.	Cronbach's Alpha SEM	Feldt-Raju Coefficient Est.	Feldt-Raju Coefficient SEM
State	All Items	151,749	0.94	3.24	0.94	3.04
Gender	Female	73,712	0.94	3.24	0.94	3.03
	Male	78,037	0.94	3.24	0.95	3.04
Ethnicity	Asian	17,142	0.94	2.89	0.95	2.70
	Black	28,558	0.92	3.29	0.93	3.11
	Hispanic	43,042	0.92	3.30	0.93	3.12
	American Indian	1,172	0.93	3.30	0.94	3.09
	Multiracial	2,878	0.94	3.20	0.95	2.99
	Pacific Islander	469	0.94	3.22	0.95	3.01
	White	57,490	0.93	3.19	0.94	3.01
NRC	New York	63,852	0.94	3.22	0.95	3.00
	Big 4 Cities	4,973	0.92	3.24	0.93	3.05
	Urban/Suburban	10,510	0.90	3.28	0.91	3.12
	Rural	7,972	0.91	3.32	0.92	3.15
	Average Needs	31,694	0.92	3.26	0.93	3.09
	Low Needs	16,444	0.92	3.05	0.93	2.89
	Charter School	10,241	0.93	3.21	0.94	3.03
	Religious and Independent	6,063	0.92	3.30	0.93	3.12
SWD	All Codes	22,446	0.89	3.15	0.90	3.03
SUA	All Codes	21,179	0.89	3.16	0.90	3.04
ELL/MLL	ELL=Y	12,109	0.89	3.13	0.90	3.03
SWD/SUA	SWD & SUA codes	18,082	0.88	3.13	0.89	3.02
ELL/MLL/SUA	SUA & ELL codes	2,547	0.80	2.99	0.81	2.93

Table 7.20. Mathematics Grade 8 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	108,410	0.92	3.39	0.93	3.22
Gender	Female	51,682	0.92	3.40	0.93	3.23
	Male	56,728	0.92	3.36	0.93	3.21
Ethnicity	Asian	10,671	0.94	3.14	0.95	2.94
	Black	22,280	0.91	3.35	0.92	3.21
	Hispanic	33,540	0.91	3.39	0.92	3.23
	American Indian	770	0.91	3.38	0.92	3.24
	Multiracial	1,650	0.92	3.36	0.93	3.20
	Pacific Islander	338	0.93	3.39	0.94	3.18
	White	38,542	0.91	3.42	0.92	3.28
NRC	New York	47,927	0.93	3.36	0.94	3.17
	Big 4 Cities	4,497	0.91	3.20	0.91	3.07
	Urban/Suburban	7,670	0.88	3.29	0.88	3.18
	Rural	6,351	0.89	3.41	0.90	3.29
	Average Needs	19,653	0.89	3.44	0.90	3.32
	Low Needs	8,430	0.91	3.40	0.91	3.26
	Charter School	6,642	0.93	3.31	0.94	3.15
	Religious and Independent	7,240	0.91	3.40	0.92	3.25
SWD	All Codes	18,652	0.87	3.16	0.88	3.08
SUA	All Codes	17,794	0.88	3.18	0.88	3.09
ELL/MLL	ELL=Y	10,241	0.89	3.17	0.90	3.08
SWD/SUA	SWD & SUA codes	15,123	0.86	3.14	0.87	3.07
ELL/MLL/SUA	SUA & ELL codes	2,181	0.79	3.00	0.79	2.97

7.2. Standard Error of Measurement (SEM)

Table 7.2 and Table 7.4 present the SEMs, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, for ELA and Mathematics, respectively. The SEMs ranged from 2.38 to 3.39 across subjects, grades, and the two methods of estimation, which is reasonable and small. The SEMs are directly related to reliability: the higher the reliability, the lower the standard error. As discussed, the reliability of these tests is relatively high, so it was expected that the SEMs would be very low.

The SEMs for the subpopulations, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, are presented in Tables 7.9–7.14 and Tables 7.15–7.20. The SEMs associated with all reliability estimates for all subjects, grades, methods of estimation, and subpopulations ranged from 2.29 to 3.44, which is acceptably close to those for the entire population. This narrow range indicates that across the Grades 3–8 ELA and Mathematics Tests, all students' test scores are reasonably reliable with minimal error.

7.3. Performance Level Classification Consistency and Accuracy

This subsection describes the analyses conducted to estimate performance level classification consistency and accuracy for the Grades 3–8 ELA and Mathematics Tests. In other words, this provides statistical information on the classification of students into the four performance categories. Classification consistency refers to the estimated degree of agreement between examinees' performance classification from two independent administrations of the same test (or from two parallel forms of the test). Because obtaining test scores from two independent administrations of New York State tests was not feasible due to item release after each administration, a psychometric model was used to obtain the estimated classification consistency indices, using test scores from a single administration. Classification accuracy can be defined as the agreement between the actual classifications using observed cut scores and true classifications based on known true cut scores (Livingston and Lewis, 1995).

In conjunction with measures of internal consistency, classification consistency is an important type of reliability and is particularly relevant to high-stakes pass/fail tests. As a form of reliability, classification consistency represents how reliably students can be classified into performance categories.

Classification consistency is most relevant for students whose proficiency is near the pass/fail cut score. For example, consider the cut score delineating Levels II and III or simply the "Level III Cut." Students whose proficiency is far above or far below that cut score are unlikely to be misclassified because repeated administration of the test will nearly always result in the same classification. Examinees whose true scores are close to the cut score are a more serious concern. These students' true scores will likely lie within the SEM of the cut score. For this reason, the measurement error at the cut scores should be considered when evaluating the classification consistency of a test. Furthermore, the number of students near the cut scores should also be considered when evaluating classification consistency; these numbers show the number of students who are at risk of being misclassified. Scoring tables with SEMs are located in Section 6: IRT Calibration and Scaling, and student scale score frequency distributions are located in Appendix Q. Classification consistency and accuracy were estimated using the IRT procedure suggested by Lee, Hanson, and Brennan (2002) and Wang, Kolen, and Harris (2000). Appendix P includes a description of the calculations and procedure based on the paper by Lee et al. (2002).

7.3.1. *Consistency*

The results for classifying students into four performance levels are separated from the results based solely on the Level III cut. Table 7.21 and Table 7.22 include case counts (n-count), classification consistency (Agreement), classification inconsistency (Inconsistency), and Cohen's kappa (Kappa). Consistency indicates the rate at which that a second administration would yield the same performance category designation (or a different designation for the inconsistency rate). The agreement index is a sum of the diagonal element in the contingency table. Kappa is similar, but corrects for chance agreement. The inconsistency index is equal to the "1 - agreement index."

Table 7.21 depicts the ELA and Mathematics consistency study results, based on the range of performance levels for all grades. For ELA, 66–69% of students were estimated to be classified consistently to one of the four performance categories with a hypothetical second administration. Kappa—that corrects for chance agreement—ranged from 0.53 to 0.58. These are between

"moderate" and "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For Mathematics, 71–77% of students were estimated to be classified consistently to one of the four performance categories, and kappa ranged from 0.62 to 0.68. These are all considered "substantial" agreement, by Landis and Koch's (1977) rules of thumb for the kappa statistic. As mentioned above and for all tests, there is an acceptable amount of measurement error that all scores contain. By random chance, students testing twice may be classified first, for example, as a Level III and second as a Level IV. This is expected to occur more often for students scoring around the selected cut score, and less often for students closer to the middle of the performance level (i.e., close to the mid-point of two adjacent cut scores).

Table 7.21. Decision Consistency (All Cuts)*

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	179,339	69%	31%	0.55
4	181,672	66%	34%	0.53
5	175,175	66%	34%	0.53
6	170,015	66%	34%	0.55
7	155,919	69%	31%	0.58
8	151,522	68%	32%	0.56
Mathematics				
3	176,663	71%	29%	0.62
4	176,897	73%	27%	0.64
5	168,578	73%	27%	0.64
6	164,429	74%	26%	0.65
7	151,749	77%	23%	0.68
8	108,410	75%	25%	0.65

**Note*. Decision consistency was calculated for PBT students only as item parameters were disproportionally based on PBT.

Table 7.22 depicts the ELA and Mathematics consistency study results based on two performance levels (NYS Level II and NYS Level III) as defined by the Level III cut. For ELA, 88–94% of the classifications of individual students were estimated to remain stable with a second administration. Kappa coefficients for ELA classification consistency ranged from 0.61 to 0.69. These are considered "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For Mathematics, 91–95% of the classifications were estimated consistently, and kappa coefficients ranged from 0.74 to 0.81. As with ELA, these statistics indicate at least "substantial" agreement (where kappa > 0.60) and some indicating "almost perfect" agreement (where kappa > 0.80), as per Landis and Koch's (1977) rules of thumb for kappa.

Table 7.22. Decision Consistency (Level III Cut)*

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	179,339	94%	6%	0.61
4	181,672	90%	10%	0.65
5	175,175	91%	9%	0.64
6	170,015	88%	12%	0.69
7	155,919	92%	8%	0.65
8	151,522	89%	11%	0.67
Mathematics				
3	176,663	91%	9%	0.74
4	176,897	91%	9%	0.77
5	168,578	92%	8%	0.76
6	164,429	93%	7%	0.80
7	151,749	94%	6%	0.81
8	108,410	95%	5%	0.78

**Note.* Decision consistency was calculated for PBT students only as item parameters were disproportionally based on PBT.

7.3.2. *Accuracy*

Table 7.23 presents the results of classification accuracy for ELA and Mathematics across all grades. Included in the table are case counts (n-count) and classification accuracy (Accuracy) for all performance levels (All Cuts) and for the Level III cut score. By definition, accuracy associated with the Level III cut is at least as great as that with the entire set of cut scores because there are only two categories for the former, as opposed to the latter, which has four.

For ELA, the estimated accuracy rates indicate that the categorization of a student's observed performance is in agreement with the location of his or her underlying proficiency from 74% to 77% of the time across all performance levels and 91% to 96% of the time in regard to the Level III cut score. For mathematics, the estimated accuracy rates indicate that the categorization of a student's observed performance is in agreement with the location of his or her true proficiency from 79% to 83% of the time across all performance levels and 93% to 96% of the time in regard to the Level III cut score.

Table 7.23. Decision Agreement (Accuracy) Estimates*

Grade	N-Count	Accuracy	
		All Cuts	Level III Cut
ELA			
3	179,339	77%	96%
4	181,672	75%	93%
5	175,175	74%	93%
6	170,015	74%	91%
7	155,919	77%	94%
8	151,522	76%	92%
Mathematics			
3	176,663	79%	93%
4	176,897	80%	93%
5	168,578	81%	95%
6	164,429	81%	95%
7	151,749	83%	96%
8	108,410	81%	96%

**Note*. Decision agreement was calculated for PBT students only as item parameters were disproportionally based on PBT.

Section 8: Standards Review

Given a test design change and a reduction in test length in 2018 from 2017, a Standards Review meeting was held in the summer of 2018. The review was done to ensure that the knowledge, skills, and abilities specified in New York's performance level descriptions (PLDs) remain relevant and that the operational cut scores appropriately separate the four performance levels from both content and psychometric perspectives.

During the week of July 9, 2018, 56 educators from the state of New York participated as panelists to review and recommend cut score points for the Grades 3–8 English Language Arts (ELA) and Mathematics tests. The following steps were used in the Standards Review process:

1. Convene Standards Review Committees
2. Identify equated cut score points on the 2018 test that were comparable to those from the 2017 test
3. Panelists review the current PLDs and develop threshold PLDs
4. Panelists review and recommend cut score points on the 2018 test following the Bookmark Standard Setting methodology (2 rounds of judgements)
5. Conduct vertical articulation
6. Panelists complete the evaluation survey

The recommended cut score points from the panelists and impact data were discussed during the vertical articulation. Slight changes to 2 out of 36 cut score points were recommended based upon this review, and consensus was reached on the most appropriate cut score for each test. The cut score recommendations were then approved by the Commissioner of Education without any further changes. The recommended raw score cuts, along with the corresponding scale score cuts, are shown in Tables 8.1 and 8.2 for ELA and Mathematics, respectively.

Table 8.1. Recommended Cut Points for the English Language Arts Assessments

Performance Level	Cut Scores/ % Students	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level II	Raw score cut	12	13	24	23	24	23
	Scale score cut	583	584	594	590	591	584
	% students	31.7%	33.1%	30.5%	23.7%	31.5%	33.5%
NYS Level III	Raw score cut	19	21	31	30	33	33
	Scale score cut	602	603	609	602	607	603
	% students	43.7%	29.8%	22.5%	22.0%	28.1%	27.5%
NYS Level IV	Raw score cut	29	27	36	35	40	39
	Scale score cut	629	619	622	614	623	617
	% students	7.3%	18.1%	14.3%	27.2%	12.1%	20.9%

Table 8.2. Recommended Cut Points for the Mathematics Assessments

Performance Level	Cut Scores/ % Students	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level II	Raw score cut	12	13	24	23	24	23
	Scale score cut	583	584	594	590	591	584
	% students	31.7%	33.1%	30.5%	23.7%	31.5%	33.5%
NYS Level III	Raw score cut	19	21	31	30	33	33
	Scale score cut	602	603	609	602	607	603
	% students	43.7%	29.8%	22.5%	22.0%	28.1%	27.5%
NYS Level IV	Raw score cut	29	27	36	35	40	39
	Scale score cut	629	619	622	614	623	617
	% students	7.3%	18.1%	14.3%	27.2%	12.1%	20.9%

Appendix T presents the full Standards Review report that describes the process, composition of the committees, ratings from the rounds, evaluation forms, results, and other materials.

Section 9: Summary of Operational Test Results

This section summarizes the distribution of scale score results on the NYSTP 2018 Grades 3–8 ELA and Mathematics Tests. These include the scale score means, standard deviations, percentile ranks, and performance level distributions for each grade's population and specific subgroups. Gender, ethnic identification, NRC, ELL/MLL, SWD, and SUA variables were used to calculate the results of subgroups required for federal reporting and test equity purposes for both the ELA and mathematics tests. Additionally, the ELL/MLL/SUA subgroup is defined as English language learners/Multilingual Learner who use one or more ELL-related accommodations. The SWD/SUA subgroup is defined as examinees with disabilities who use one or more disability-related accommodation(s). For the mathematics analyses, the test translation language is also indicated. (Recall that the ELA tests are not translated, as they are a measure of mastery of the English language.) ELA and mathematics data include examinees with valid scores from all public, non-public, and charter schools. Complete scale score frequency distribution tables for ELA and mathematics are located in Appendix Q.

9.1. Scale Score Distribution Summary

Scale score distribution summary tables for ELA and mathematics are presented and discussed. ELA scale score distributions are described first, followed by mathematics. In the following two subsections, ELA and mathematics scale score and subscore statistics are presented for all grades, and across selected subgroups in each grade level. Use caution when interpreting the statistics for subgroups with small number counts that are included in the scale score summaries.

9.1.1. *ELA Scale Score and Subscore Distributions*

Table 9.1 shows some key statistics characterizing the distribution of ELA scale scores, while Table 9.2 summarizes the ELA subscores derived from the test in each grade. Tables 9.3–9.8 break down the scale scores by selected subgroups. Some general observations from these tables include: Females outperformed Males; Asian and White students outperformed their peers from other reported ethnic groups; students from Low Needs (as identified by NRC) districts outperformed students from other districts (New York City, Big 4 Cities, Urban/Suburban, Rural, Average Needs, and Charter); and ELL/MLL students, SWD, SUA, and SWD/SUA tended to under-perform the State population (All Students). This pattern of achievement was consistent across all grades.

Table 9.1. ELA Scale Score Distribution Summary

		Scale Score		Percentile Ranks				
Grade	N-Count	Mean	SD	10th	25th	50th	75th	90th
3	182,885	599.79	20.22	573	586	602	614	626
4	184,266	599.77	20.17	572	586	601	614	624
5	177,609	599.88	20.27	573	587	602	614	625
6	173,183	599.74	20.30	574	587	601	614	623
7	161,958	599.74	20.26	574	587	601	613	623
8	154,663	599.59	20.50	574	588	601	614	624

Table 9.2. ELA Subscore Summary

Grade	Subscore	N-Count	Subscore Max	Subscore Mean	Subscore SD
3	Reading	182,885	18	10.29	4.01
	Writing	182,885	16	8.14	3.62
4	Reading	184,266	18	10.92	3.93
	Writing	184,266	16	8.45	3.90
5	Reading	177,609	28	16.88	5.42
	Writing	177,609	16	9.83	3.42
6	Reading	173,183	28	17.63	5.58
	Writing	173,183	16	10.12	3.96
7	Reading	161,958	28	16.60	5.55
	Writing	161,958	18	12.14	4.53
8	Reading	154,663	28	18.64	5.34
	Writing	154,663	18	11.89	4.36

9.1.1.1. ELA Grade 3

Table 9.3 presents the scale score statistics and n-counts of demographic subgroups for Grade 3. The population scale score mean was 599.79 with a standard deviation of 20.22. Female students tended to outperform male students by around five scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from New York City, Average Needs and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (609.44). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score— by about two-thirds of a standard deviation below the population mean. The students with disabilities (SWD), students tested under accommodations (SUA), and English language learners /Multilingual Learner (ELL/MLL) subgroups scored, on average, about one standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 38 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (602): Female (604), Asian (612), Multiracial (603), Pacific Islander (604), and White (604) students, those attending schools in Low Needs districts (609), and students attending Charter schools (609).

Table 9.3. ELA Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
State	All Students	182,885	599.79	20.22	573	586	602	614	626
Gender	Female	90,155	602.42	19.81	577	589	604	617	626
	Male	92,730	597.23	20.28	569	584	600	612	622
Ethnicity	Asian	17,913	609.44	19.42	583	597	612	622	634
	Black	31,910	595.64	20.63	569	583	597	609	622

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
Ethnicity	Hispanic	51,096	595.46	19.61	569	583	597	609	619
	American Indian	1,263	596.90	20.86	569	584	597	612	622
	Multiracial	5,223	601.67	20.32	577	589	603	615	626
	Pacific Islander	425	602.73	20.93	574	592	604	617	626
	White	72,941	602.63	19.15	577	592	604	615	626
NRC	New York	67,325	600.26	20.87	573	586	602	614	626
	Big 4 Cities	7,898	586.44	21.13	558	573	586	602	614
	Urban/Suburban	14,389	592.01	19.35	565	580	592	604	617
	Rural	10,027	595.14	18.69	570	583	597	608	618
	Average Needs	42,841	599.91	18.45	577	589	602	612	622
	Low Needs	18,448	608.44	16.82	586	600	609	619	629
	Charter	12,276	606.64	18.87	583	595	609	619	629
	Religious and Independent	9,624	597.96	21.19	569	586	600	612	622
SWD	All Codes	26,715	584.77	19.72	558	573	586	597	609
SUA	All Codes	12,177	583.18	18.93	558	570	583	597	607
ELL/MLL	ELL=Y	21,353	584.81	18.18	558	573	586	597	607
SWD/SUA	SWD & SUA codes	8,706	581.93	18.89	553	569	583	595	604
ELL/MLL/SUA	SUA & ELL codes	1,068	578.43	17.32	553	565	580	589	600

9.1.1.2. ELA Grade 4

Table 9.4 contains Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 599.77 with a standard deviation of 20.17. Female students tended to outperform male students by around 10 scale score points. Asian, Multiracial, Pacific Islander and White students' scale score means exceeded the state mean scale score, as did those of students from New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (610.08). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, about three-quarters deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 17 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601): Female (603), Asian (611), Multiracial (602), Pacific Islander (606), and White (603) students, those from Low Needs districts (608), and those enrolled at Charter 608) schools.

Table 9.4. ELA Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
State	All Students	184,266	599.77	20.17	572	586	601	614	624
Gender	Female	90,669	602.34	19.80	575	589	603	616	627
	Male	93,597	597.28	20.23	570	584	598	611	623
Ethnicity	Asian	18,629	610.08	19.90	584	598	611	623	637
	Black	32,693	595.76	19.93	569	582	596	611	619
	Hispanic	50,723	595.64	19.33	570	584	596	608	619
	American Indian	1,279	596.48	19.71	572	584	596	611	619
	Multiracial	4,851	601.05	20.73	575	586	602	616	627
	Pacific Islander	502	603.49	19.96	575	590	606	616	628
	White	73,442	602.23	19.46	576	591	603	616	627
NRC	New York	67,657	601.07	21.32	572	586	601	616	627
	Big 4 Cities	7,874	585.36	19.68	561	572	585	598	611
	Urban/Suburban	13,903	591.92	18.81	569	578	591	606	616
	Rural	9,972	594.01	18.44	569	581	595	606	616
	Average Needs	41,234	599.50	18.16	575	589	601	612	623
	Low Needs	18,378	607.83	16.70	586	598	608	619	627
	Charter	11,436	605.71	17.41	581	596	608	616	627
	Religious and Independent	13,774	598.84	21.10	569	586	601	614	623
SWD	All Codes	27,585	584.04	18.62	561	572	584	596	608
SUA	All Codes	13,737	582.51	17.55	561	572	582	594	606
ELL/MLL	ELL=Y	17,775	582.21	16.99	561	572	584	594	603
SWD/SUA	SWD & SUA codes	9,762	580.64	17.20	561	569	581	591	603
ELL/MLL/SUA	SUA & ELL codes	1,038	577.79	14.53	561	569	578	587	596

9.1.1.3. ELA Grade 5

Table 9.5 provides the scale score summary statistics by key demographic subgroups for Grade 5 students. The population scale score mean was 599.88 with a standard deviation of 20.27. Female students tended to outperform male students by around six scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average Needs and Low Needs districts, and Charter schools. Across all ethnic groups, Asian students earned the highest mean score (609.76). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, one standard deviation below the mean scale score for the population. English language learners /Multilingual Learner tested under accommodations were the lowest performing subgroup analyzed, scoring about 22 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (602):

Female (604), Asian (611), Pacific Islander (606), and White (604) students, as well as those from Low Needs districts (610), and Charter schools (604).

Table 9.5. ELA Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
State	All Students	177,609	599.88	20.27	573	587	602	614	625
Gender	Female	87,756	602.85	19.33	578	592	604	616	625
	Male	89,853	596.99	20.74	570	585	598	611	622
Ethnicity	Asian	18,755	609.76	19.61	585	598	611	622	633
	Black	32,041	595.05	19.96	570	583	596	609	619
	Hispanic	49,253	595.82	19.13	571	584	598	609	619
	American Indian	1,264	597.19	18.53	573	585	598	609	619
	Multiracial	4,345	601.86	20.81	575	590	602	616	629
	Pacific Islander	541	606.03	19.19	581	596	606	619	629
	White	69,491	602.67	19.79	578	592	604	616	625
NRC	New York	68,524	600.67	20.65	573	587	602	614	625
	Big 4 Cities	7,610	584.99	21.63	555	571	585	600	612
	Urban/Suburban	13,174	592.74	19.12	567	581	594	606	616
	Rural	9,433	594.41	18.78	570	583	596	606	617
	Average Needs	39,498	600.54	18.32	578	590	602	614	622
	Low Needs	18,404	609.09	16.60	590	600	610	619	629
	Charter	11,234	604.10	17.96	581	594	604	616	625
	Religious and Independent	9,729	596.02	23.77	563	583	600	611	622
SWD	All Codes	27,838	583.43	19.32	559	573	585	596	606
SUA	All Codes	13,986	582.13	19.17	555	570	583	595	605
ELL/MLL	ELL=Y	14,867	578.19	18.15	554	566	581	592	600
SWD/SUA	SWD & SUA codes	10,046	579.87	18.74	554	570	581	592	602
ELL/MLL/SUA	SUA & ELL codes	1,064	575.15	16.20	554	566	575	587	594

9.1.1.4. ELA Grade 6

Table 9.6 contains Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 599.74 with a standard deviation of 20.30. Female students tended to outperform male students by around seven scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average Needs and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (609.93). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, one standard deviation below the mean scale score for the population. English language learners /Multilingual Learner tested under accommodations

were the lowest-performing subgroup analyzed, scoring about 22 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601): Female (604), Asian (611), Multiracial (604), Pacific Islander (607), and White (604) students, and those enrolled in Average (602) and Low (611) Needs districts, and Charter (602) and Religious and Independent (602) schools.

Table 9.6. ELA Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	173,183	599.74	20.30	574	587	601	614	623
Gender	Female	84,878	603.56	19.37	578	592	604	616	627
	Male	88,305	596.07	20.49	569	582	597	611	619
Ethnicity	Asian	18,129	609.93	19.92	584	599	611	623	632
	Black	32,033	594.41	19.38	569	582	595	608	619
	Hispanic	47,601	595.51	19.12	571	584	597	609	619
	American Indian	1,167	597.46	20.09	571	584	599	611	620
	Multiracial	3,822	602.02	20.55	575	590	604	616	627
	Pacific Islander	620	604.73	19.96	579	593	607	616	627
	White	67,598	602.88	19.90	577	592	604	616	627
NRC	New York	65,208	600.36	20.74	574	586	601	614	627
	Big 4 Cities	6,993	585.68	20.22	559	571	586	600	611
	Urban/Suburban	12,437	591.29	19.76	566	578	592	604	616
	Rural	9,182	595.03	18.67	571	583	596	608	617
	Average Needs	37,022	600.34	18.88	576	589	602	614	623
	Low Needs	17,655	609.40	17.26	588	599	611	619	632
	Charter	11,540	601.75	17.22	580	592	602	614	623
	Religious and Independent	12,939	598.99	21.95	569	588	602	614	623
SWD	All Codes	26,971	582.64	18.05	559	571	582	595	604
SUA	All Codes	13,408	581.55	18.28	559	570	582	593	604
ELL/MLL	ELL=Y	13,822	577.99	17.16	555	566	579	590	599
SWD/SUA	SWD & SUA codes	9,431	578.90	17.79	555	569	580	592	601
ELL/MLL/SUA	SUA & ELL codes	975	575.84	16.07	555	566	578	586	595

9.1.1.5. ELA Grade 7

Table 9.7 presents the Grade 7 scale score statistics and n-counts of demographic subgroups. The population scale score mean was 599.74 with a standard deviation of 20.26. Female students tended to outperform male students by around seven scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students from New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (610.88). Across NRC subgroups,

students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, about one standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 23 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601): Female (605), Asian (613), Multiracial (603), Pacific Islander (607), and White (605) students as well as those enrolled in New York City (603), Low Needs districts (609), and Charter (603) schools.

Table 9.7. ELA Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	161,958	599.74	20.26	574	587	601	613	623
Gender	Female	78,711	603.51	18.78	579	593	605	615	627
	Male	83,247	596.18	20.95	570	583	598	611	620
Ethnicity	Asian	17,363	610.88	19.96	585	600	613	623	637
	Black	30,819	594.56	19.05	570	583	596	607	618
	Hispanic	44,352	595.70	18.86	573	585	598	609	618
	American Indian	1,236	596.77	20.23	570	584	598	611	620
	Multiracial	3,097	601.63	21.28	574	589	603	618	627
	Pacific Islander	482	604.02	20.40	575	593	607	618	627
	White	62,775	602.48	20.02	577	591	605	615	627
NRC	New York	65,334	601.60	20.15	575	589	603	615	627
	Big 4 Cities	6,554	585.01	21.24	557	571	586	600	612
	Urban/Suburban	11,075	589.47	20.05	564	577	591	603	613
	Rural	8,494	594.40	18.81	570	583	596	607	618
	Average Needs	33,387	599.29	19.13	575	588	601	613	621
	Low Needs	17,179	608.27	17.41	586	598	609	620	627
	Charter	10,617	601.81	16.07	581	593	603	613	620
	Religious and Independent	9,118	597.65	23.24	564	587	601	613	623
SWD	All Codes	25,931	583.66	18.38	561	573	585	596	605
SUA	All Codes	12,802	581.62	18.87	557	570	582	594	605
ELL/MLL	ELL=Y	11,704	576.65	17.96	552	565	579	589	598
SWD/SUA	SWD & SUA codes	9,017	579.04	18.34	553	567	579	592	601
ELL/MLL/SUA	SUA & ELL codes	787	574.59	15.52	552	564	577	585	593

9.1.1.6. ELA Grade 8

Table 9.8 presents the Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 599.59 with a standard deviation of 20.50. Female students tended to outperform male students by around eight scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean

scale score, as did those of students enrolled in New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (610.52). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, one standard deviation below the mean scale score for the population. English language learners/Multilingual Learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 24 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601), Female (605), Asian (612), Multiracial (603), Pacific Islander (607), and White (603) students, as well as those enrolled in New York City (603) and Low Needs (607) districts, Charter (603), and Religious and Independent (603) schools.

Table 9.8. ELA Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	154,663	599.59	20.50	574	588	601	614	624
Gender	Female	74,920	603.57	18.97	580	591	605	617	628
	Male	79,743	595.85	21.17	568	584	597	609	620
Ethnicity	Asian	17,639	610.52	20.41	584	599	612	624	634
	Black	29,862	595.17	19.27	570	584	597	607	617
	Hispanic	41,867	596.02	19.20	571	584	597	609	620
	American Indian	1,203	596.97	20.00	571	586	597	609	620
	Multiracial	2,443	600.98	21.93	572	588	603	617	628
	Pacific Islander	457	604.47	20.41	576	591	607	617	628
	White	59,660	601.40	20.45	576	590	603	614	624
NRC	New York	63,216	601.74	20.06	576	590	603	614	628
	Big 4 Cities	6,388	584.78	21.04	557	572	586	599	612
	Urban/Suburban	10,000	589.61	20.82	563	576	591	603	614
	Rural	8,047	593.93	19.28	569	582	595	607	617
	Average Needs	30,019	598.52	19.92	573	587	600	612	621
	Low Needs	15,970	606.64	18.48	584	597	607	618	628
	Charter	9,644	602.90	16.07	582	593	603	614	620
	Religious and Independent	10,988	599.02	22.42	570	590	603	614	624
SWD	All Codes	23,999	583.53	18.09	560	572	584	595	605
SUA	All Codes	11,628	581.80	18.82	557	570	582	594	605
ELL/MLL	ELL=Y	11,224	575.85	17.72	553	566	577	588	597
SWD/SUA	SWD & SUA codes	8,365	579.11	18.17	557	568	580	591	601
ELL/MLL/SUA	SUA & ELL codes	683	573.52	17.04	548	566	574	585	593

9.1.2. *Mathematics Scale Score Distributions*

Table 9.9 shows some key statistics characterizing the distribution of mathematics scale scores, while Table 9.10 summarizes the mathematics subscores derived from the test in each grade. Tables 9.11–9.16 break down the scale scores by selected subgroups. Some general observations from the mathematics data are as follows: Female and Male students performed fairly consistently; Asian students scored considerably higher than other reported ethnic groups; schools belonging to Low Needs districts (as identified by the NRC code) and Charter schools outperformed most other school types (New York City, Big 4 Cities, High Needs Urban/Suburban, and Rural and Average Needs districts). Students taking the Chinese and Korean translations tended to outperform the other translation subgroups (Haitian-Creole, Spanish, and Russian); and ELL/MLLs, SWDs, and/or SUAs achieved below the State mean in most percentile ranks. This pattern of achievement was fairly consistent across all grades.

Table 9.9. Mathematics Scale Score Distribution Summary

		Scale Score		Percentile Ranks				
Grade	N-Count	Mean	SD	10th	25th	50th	75th	90th
3	184,970	599.48	20.19	574	587	601	613	623
4	186,331	599.38	20.23	573	587	600	612	624
5	178,875	599.09	20.39	574	587	600	613	625
6	173,731	599.36	20.36	575	586	600	613	624
7	160,487	599.16	20.42	572	587	601	613	623
8	116,534	598.98	20.47	568	586	601	612	623

Table 9.10. Mathematics Subscore Summary

			Subscore		
Grade	Subscore	N-Count	Max	Mean	SD
3	Operations and Algebraic Thinking	184,970	19	12.09	4.59
	Number and Operations-Fractions	184,970	8	4.31	2.21
	Measurement and Data	184,970	10	5.30	3.02
4	Operations and Algebraic Thinking	186,331	9	5.30	2.40
	Number and Operations in Base 10	186,331	12	7.16	3.57
	Number and Operations-Fractions	186,331	12	7.51	3.12
5	Number and Operations in Base 10	178,875	12	6.95	3.27
	Number and Operations-Fractions	178,875	16	8.27	4.09
	Measurement and Data	178,875	14	7.75	3.81
6	Ratios and Proportional Relationships	173,731	12	6.37	3.26
	The Number System	173,731	10	5.22	2.83
	Expressions and Equations	173,731	19	8.74	4.59
7	Ratios and Proportional Relationships	160,487	13	6.98	3.76
	The Number System	160,487	10	6.44	2.80

Grade	Subscore	N-Count	Subscore Max	Subscore Mean	Subscore SD
7	Expressions and Equations	160,487	15	7.24	4.19
8	Expressions and Equations	116,534	21	10.63	5.88
	Functions	116,534	14	5.77	3.65
	Geometry	116,534	10	4.59	2.49

9.1.2.1. Mathematics Grade 3

Table 9.11 presents the Grade 3 scale score statistics and n-counts of demographic subgroups. The population scale score mean was 599.48 with a standard deviation of 20.19. Female and Male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (611.28). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored about three-quarters of a standard deviation below the mean scale score for the population. SUA students tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 18 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601): Asian (613), Pacific Islander (604), and White (604) students, as well as those enrolled at Low Needs (609) districts and Charter schools (611). In terms of the 50th-percentile ranks for students using translated forms, they ranged from 582 (Haitian-Creole, n = 34 and Spanish (n = 1,768) to 613 (Chinese, n = 102).

Table 9.11. Mathematics Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
State	All Students	184,970	599.48	20.19	574	587	601	613	623
Gender	Female	90,724	599.69	19.50	574	587	601	613	623
	Male	94,246	599.28	20.84	571	586	601	613	623
Ethnicity	Asian	18,468	611.28	18.60	587	600	613	623	633
	Black	32,048	594.41	20.95	568	580	595	609	620
	Hispanic	52,069	594.56	19.50	568	582	595	608	618
	American Indian	1,304	596.78	20.15	571	585	598	609	620
	Multiracial	5,226	600.59	20.23	574	587	601	614	624
	Pacific Islander	436	602.80	19.29	576	592	604	615	627
	White	73,811	602.54	18.70	578	592	604	615	623
NRC	New York	68,732	599.38	20.51	574	586	600	613	623
	Big 4 Cities	8,089	585.74	20.93	560	571	586	601	612
	Urban/Suburban	14,507	591.14	19.61	564	578	592	604	615
	Rural	9,917	595.76	19.22	571	584	597	609	620
	Average Needs	42,715	600.00	18.52	576	589	601	612	623

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
NRC	Low Needs	18,513	608.22	16.80	587	598	609	620	627
	Charter	12,230	609.83	19.30	584	597	611	623	633
	Religious and Independent	10,103	596.12	19.20	571	584	597	609	620
SWD	All Codes	27,469	584.94	20.73	560	571	584	600	611
SUA	All Codes	12,974	581.83	20.34	554	568	582	596	608
ELL/MLL	ELL=Y	23,775	586.94	19.26	560	574	587	600	611
SWD/SUA	SWD & SUA codes	9,625	580.02	20.22	554	564	580	594	606
ELL/MLL/SUA	SUA & ELL codes	1,289	578.07	19.61	554	564	577	591	604
ELL/MLL Test Language	Chinese	102	612.57	18.21	587	604	613	623	633
	English	183,028	599.65	20.13	574	587	601	613	623
	Haitian-Creole	34	578.65	17.88	549	568	582	592	598
	Korean	25	596.68	18.84	574	583	598	608	623
	Russian	13	591.54	19.99	568	580	595	609	609
	Spanish	1,768	581.39	18.11	555	568	582	594	604
	All Translations	1,942	583.24	19.48	560	571	584	597	608

9.1.2.2. Mathematics Grade 4

Table 9.12 presents the Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 599.38 with a standard deviation of 20.23. Female and Male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students enrolled in New York City, Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (612.43). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored about three-quarters of a standard deviation below the mean scale score for the population. Students tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 18 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (600): Male (601), Asian (614), Multiracial (601), Pacific Islander (603), and White (603) students, and those enrolled in Average (602) and Low (609) Needs districts and Charter schools (609). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 580 (Haitian-Creole, n = 31, and Spanish, n = 1,543) to 607 (Chinese, n = 93).

Table 9.12. Mathematics Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
State	All Students	186,331	599.38	20.23	573	587	600	612	624
Gender	Female	91,167	599.52	19.93	573	587	599	612	624

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
Gender	Male	95,164	599.25	20.51	573	587	601	612	624
Ethnicity	Asian	19,181	612.43	19.44	588	601	614	624	635
	Black	32,747	593.61	20.29	567	580	593	607	619
	Hispanic	51,671	594.34	19.09	571	582	595	607	618
	American Indian	1,309	596.14	20.41	571	584	597	609	621
	Multiracial	4,880	600.14	20.78	573	587	601	614	624
	Pacific Islander	521	602.88	19.49	578	591	603	616	628
	White	74,330	602.41	18.87	578	592	603	614	624
NRC	New York	68,895	599.54	20.92	573	585	599	614	628
	Big 4 Cities	8,027	584.55	20.14	557	571	585	598	610
	Urban/Suburban	14,040	590.82	19.38	564	578	592	603	614
	Rural	9,845	595.09	18.71	571	584	596	607	618
	Average Needs	41,365	600.42	18.06	578	589	602	612	621
	Low Needs	18,564	608.81	17.02	588	599	609	618	628
	Charter	11,413	608.26	19.31	584	596	609	621	635
	Religious and Independent	14,040	596.02	19.31	571	585	597	609	618
SWD	All Codes	28,137	583.55	19.61	557	571	584	597	609
SUA	All Codes	14,755	581.74	19.53	557	567	582	595	606
ELL/MLL	ELL=Y	20,144	584.50	18.05	563	573	585	596	607
SWD/SUA	SWD & SUA codes	10,824	579.23	19.37	552	567	580	592	603
ELL/MLL/SUA	SUA & ELL codes	1,297	576.31	17.17	552	563	578	588	598
ELL/MLL Test Language	Chinese	93	607.83	15.79	591	598	607	616	628
	English	184,631	599.55	20.18	573	587	601	613	624
	Haitian-Creole	31	579.10	13.15	563	571	580	588	595
	Korean	21	610.33	17.36	593	601	606	616	628
	Russian	12	595.42	25.38	557	580	599	612	628
	Spanish	1,543	579.51	16.64	557	571	580	591	599
	All Translations	1,700	581.54	18.13	557	571	582	593	604

9.1.2.3. Mathematics Grade 5

Table 9.13 presents the Grade 5 demographic subgroup n-counts and scale score statistics. The population scale score mean was 599.09 with a standard deviation of 20.39. Female and male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students from New York City, Average and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (613.09). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation

below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, about 0.82 standard deviations below the mean scale score for the population. Students tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 17 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (600): Asian (613), Multiracial (601), Pacific Islander (605), and White (604) students, as well as those enrolled at Average (602) and Low (610) Needs districts, and Charter schools (602). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 574 (Haitian-Creole, n = 25) to 613 (Chinese, n = 68, Korean, n = 15).

Table 9.13. Mathematics Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	178,875	599.09	20.39	574	587	600	613	625
Gender	Female	87,953	599.36	19.61	574	587	600	611	622
	Male	90,922	598.82	21.12	571	585	600	613	625
Ethnicity	Asian	19,173	613.09	20.09	589	601	613	625	639
	Black	31,917	591.59	19.36	567	579	592	604	616
	Hispanic	49,798	593.83	18.85	571	581	595	606	616
	American Indian	1,294	594.94	19.71	571	583	595	607	618
	Multiracial	4,271	600.89	21.18	574	587	601	615	629
	Pacific Islander	558	604.43	19.08	579	592	605	616	629
	White	70,297	602.69	19.02	579	592	604	615	625
NRC	New York	69,433	599.20	21.16	574	585	599	613	625
	Big 4 Cities	7,659	584.16	20.02	554	571	583	597	610
	Urban/Suburban	13,167	590.47	19.09	567	579	590	604	615
	Rural	9,297	595.03	18.13	571	583	596	607	616
	Average Needs	38,729	600.83	18.37	577	590	602	613	622
	Low Needs	18,457	609.46	17.15	589	600	610	620	629
	Charter	11,167	603.38	19.20	579	590	602	615	629
	Religious and Independent	10,376	594.53	20.16	567	581	596	609	618
SWD	All Codes	27,878	583.04	18.94	554	571	583	595	607
SUA	All Codes	14,352	581.73	18.93	554	571	581	595	606
ELL/MLL	ELL=Y	16,975	582.53	17.82	562	571	583	593	604
SWD/SUA	SWD & SUA codes	10,526	579.22	18.22	554	567	579	592	602
ELL/MLL/SUA	SUA & ELL codes	1,184	574.94	16.50	550	562	574	587	596
ELL/MLL Test Language	Chinese	68	611.53	20.22	589	602	613	622	633
	English	177,539	599.23	20.34	574	587	600	613	625
	Haitian-Creole	25	577.60	17.27	554	567	574	587	602
	Korean	15	609.33	24.06	587	592	613	622	639

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
ELL/MLL Test Language	Russian	18	587.50	19.36	554	574	587	602	611
	Spanish	1,210	578.08	16.69	554	567	577	589	599
	All Translations	1,336	580.25	18.80	554	567	579	592	604

9.1.2.4. Mathematics Grade 6

Table 9.14 presents the Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 599.36 with a standard deviation of 20.36. Female and male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students enrolled in Average and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (612.79). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, 0.80 standard deviations below the mean scale score for the population. Students with disabilities were the lowest-performing subgroup analyzed for English forms, scoring about 17 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (600): Asian (614), Multiracial (603), Pacific Islander (603), and White (605) students, as well as those enrolled in Average (604) and Low (612) Needs districts and Charter schools (603). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 579 (Haitian-Creole, n = 27, Spanish, n = 1,421) to 611 (Chinese, n = 47).

Table 9.14. Mathematics Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
State	All Students	173,731	599.36	20.36	575	586	600	613	624
Gender	Female	84,789	599.70	19.84	575	588	600	612	624
	Male	88,942	599.04	20.84	571	586	600	614	624
Ethnicity	Asian	18,551	612.79	19.93	588	600	614	625	637
	Black	31,805	591.44	19.33	565	579	592	604	615
	Hispanic	47,969	593.34	19.02	571	582	595	606	616
	American Indian	1,219	595.58	20.17	571	584	596	609	620
	Multiracial	3,757	601.94	20.57	575	588	603	616	626
	Pacific Islander	634	603.76	19.88	579	592	603	617	630
	White	68,127	603.85	18.65	581	593	605	616	625
NRC	New York	66,041	598.32	21.35	571	584	598	612	625
	Big 4 Cities	7,072	585.63	20.90	555	571	586	600	612
	Urban/Suburban	12,292	589.81	19.28	565	579	590	603	614
	Rural	9,015	596.35	18.56	573	586	598	609	618
	Average Needs	36,269	602.22	18.18	579	592	604	614	624

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
NRC	Low Needs	17,487	610.66	16.88	590	601	612	620	630
	Charter	11,454	602.34	18.61	579	592	603	615	625
	Religious and Independent	13,389	597.37	18.70	575	586	599	610	619
SWD	All Codes	26,769	581.90	19.15	555	571	582	595	605
SUA	All Codes	14,218	582.75	19.59	555	571	584	596	608
ELL/MLL	ELL=Y	15,906	582.47	18.55	555	571	584	595	605
SWD/SUA	SWD & SUA codes	10,038	579.42	19.07	551	565	582	592	603
ELL/MLL/SUA	SUA & ELL codes	1,238	575.70	17.49	551	565	579	588	596
ELL/MLL Test Language	Chinese	47	610.66	15.54	590	603	611	620	628
	English	172,211	599.55	20.28	575	588	600	614	624
	Haitian-Creole	27	575.33	16.15	551	565	579	586	596
	Korean	11	601.45	16.60	584	593	605	616	619
	Russian	14	587.29	21.51	555	571	584	603	617
	Spanish	1,421	577.13	17.15	551	565	579	588	598
	All Translations	1,520	578.40	18.19	551	565	579	590	600

9.1.2.5. Mathematics Grade 7

Table 9.15 presents the Grade 7 n-counts and scale score statistics for key demographic subgroups. The population scale score mean was 599.16 with a standard deviation of 20.42. Female students tended to outperform male students by around three scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students from Average and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (613.20). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, 0.85 standard deviations below the mean scale score for the population. English language learners/Multilingual Learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 18 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601): Female (602), Asian (615), Multiracial (604), Pacific Islander (605), and White (606) students, those enrolled in Average (603) and Low (612) Needs districts, and Charter schools (604). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 572 (Russia, n = 45) to 618 (Korean, n = 9).

Table 9.15. Mathematics Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
State	All Students	160,487	599.16	20.42	572	587	601	613	623
Gender	Female	77,750	600.57	19.64	577	588	602	614	623
	Male	82,737	597.84	21.03	566	585	600	613	623
Ethnicity	Asian	17,451	613.20	18.74	588	603	615	626	639
	Black	30,218	591.16	20.07	561	580	593	605	615
	Hispanic	44,513	593.61	19.11	566	582	595	607	617
	American Indian	1,253	594.95	20.21	566	582	595	610	620
	Multiracial	3,005	601.29	20.44	572	588	604	615	626
	Pacific Islander	483	602.00	20.20	572	590	605	617	626
	White	62,214	603.43	18.60	580	593	606	615	626
NRC	New York	65,783	599.12	21.34	572	585	600	614	626
	Big 4 Cities	6,491	583.64	20.81	556	566	585	599	611
	Urban/Suburban	10,759	587.89	18.90	561	577	590	601	611
	Rural	8,190	594.97	17.99	572	585	596	608	615
	Average Needs	31,907	601.06	18.01	577	591	603	613	621
	Low Needs	16,621	609.94	16.17	590	602	612	620	628
	Charter	10,548	603.14	18.60	580	591	604	615	626
	Religious and Independent	9,627	596.84	19.79	566	587	600	610	620
SWD	All Codes	25,434	581.77	18.84	556	566	582	594	606
SUA	All Codes	12,705	582.05	19.27	556	566	582	595	607
ELL/MLL	ELL=Y	13,657	581.41	18.87	556	566	582	594	605
SWD/SUA	SWD & SUA codes	9,268	579.37	18.68	556	566	580	593	603
ELL/MLL/SUA	SUA & ELL codes	881	574.73	16.29	552	561	577	587	595
ELL/MLL Test Language	Chinese	70	610.17	21.43	584	607	614	623	630
	English	159,040	599.36	20.32	572	587	601	613	623
	Haitian-Creole	33	576.15	17.46	556	566	580	588	598
	Korean	9	616.11	16.41	587	607	618	626	644
	Russian	45	577.73	23.32	552	561	572	596	608
	Spanish	1,290	575.08	17.16	552	561	577	587	595
	All Translations	1,447	577.14	19.37	552	561	580	588	602

9.1.2.6. Mathematics Grade 8

Table 9.16 presents the Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 598.98 with a standard deviation of 20.47. Female students tended to outperform male students by around four scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean

scale score, as did those of students enrolled in New York City, Low Needs districts, Charter and Religious and Independent schools. Across ethnic groups, Asian students earned the highest mean score (614.35). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—by three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL/MLL subgroups scored, on average, 0.69 standard deviations below the mean scale score for the population. Students with disabilities were the lowest performing subgroup analyzed for English forms, scoring about 15 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (601): Female (603), Asian (616), Pacific Islander (607), and White (604) students, as well as those enrolled in New York City (602), Low Needs (609) districts, and Charter (606) and Religious and Independent (603) schools. In terms of the 50th percentile ranks for students using translated forms, they ranged from: 572 (Haitian-Creole, n = 28) to 609 (Chinese, n = 32).

Table 9.16. Mathematics Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10th	25th	50th	75th	90th
State	All Students	116,534	598.98	20.47	568	586	601	612	623
Gender	Female	55,578	601.04	19.57	575	591	603	613	625
	Male	60,956	597.10	21.09	568	584	599	611	622
Ethnicity	Asian	10,984	614.35	20.53	589	603	616	629	639
	Black	23,918	592.99	20.09	563	580	594	606	618
	Hispanic	34,909	595.69	19.48	568	584	597	609	619
	American Indian	847	594.96	20.08	568	584	597	608	619
	Multiracial	1,718	599.11	21.14	568	586	601	613	625
	Pacific Islander	354	604.41	20.63	580	592	607	618	627
	White	42,911	601.44	18.95	575	592	604	613	623
NRC	New York	49,766	600.58	21.25	568	589	602	614	627
	Big 4 Cities	5,684	583.91	20.67	559	568	584	598	611
	Urban/Suburban	7,896	588.18	18.76	563	575	591	602	611
	Rural	6,523	595.10	17.81	568	585	597	607	616
	Average Needs	19,925	598.46	17.17	575	590	601	610	618
	Low Needs	8,588	607.07	17.13	586	598	609	618	627
	Charter	6,898	605.12	19.63	580	594	606	619	629
	Religious and Independent	10,847	600.75	20.55	568	591	603	614	625
SWD	All Codes	21,321	584.32	18.59	559	568	586	597	607
SUA	All Codes	10,993	584.46	18.91	559	568	586	598	608
ELL/MLL	ELL=Y	12,091	586.07	19.63	559	575	586	599	610
SWD/SUA	SWD & SUA codes	8,099	582.00	18.28	559	568	584	594	605
ELL/MLL/SUA	SUA & ELL codes	761	577.63	16.26	559	563	580	589	598

Demographic Category		N-Count	Scale Score Mean	Scale Score SD	Percentile Ranks 10th	25th	50th	75th	90th
ELL/MLL Test Language	Chinese	32	606.38	19.54	575	598	609	619	625
	English	115,328	599.20	20.39	568	586	601	612	623
	Haitian-Creole	28	574.11	19.39	549	559	572	590	604
	Korean	9	607.33	19.59	568	598	607	622	634
	Russian	55	572.98	19.04	549	559	575	586	601
	Spanish	1,082	577.02	16.03	559	563	580	589	597
	All Translations	1,206	577.77	17.25	555	563	580	589	599

9.2. Performance Level Distribution Summary

Students are classified as NYS Level I, NYS Level II, NYS Level III, or NYS Level IV. The cut scores were established in 2013 during the standard-setting. It is inappropriate to compare scale scores across grades because they neither measure the same content, nor are they on the same scale. During the standards review process, the established cut scores were revisited and updated separately for different grades within a subject, additional care was taken to vertically articulate performance levels; see *2018 Standards Review Report* in Appendix T for details. While vertical articulation helps to build consistent meaning to the performance levels, the very nature of grade-specific content, differing performance expectations, and panel-set cut scores result in cut score differences across grades.

9.2.1. *ELA Test Performance Level Distributions*

Table 9.17 shows the performance level distribution for all examinees from public, charter, and non-public schools with valid ELA scores. Performance level data for selected subgroups of students were also examined. In general, these distributions reflect the same achievement trends in the scale score summary discussion. Across Tables 9.18 through 9.23, more Female students were classified in Level III and above subgroups than were Male students. Similarly, more Asian and White students were classified in Level III and above subgroups than were their peers from other reported ethnic groups. Consistent with the pattern shown in scale score distribution across the subgroups, students from Low and Average Needs districts outperformed students from High Needs districts (New York City, Big 4 Cities, Urban/Suburban, and Rural). The Level III and above rates for students in the ELL/MLL, SWD, and SUA subgroups were low, compared to the total population of examinees.

Table 9.17. ELA Test Performance Level Distributions

Grade	N-Count	Performance Levels: Level I	Level II	Level III	Level IV	Level III & IV
3	182,885	17.99	31.61	43.23	7.17	50.40
4	184,266	19.56	33.04	29.51	17.89	47.40
5	177,609	33.28	30.30	22.31	14.10	36.41
6	173,183	27.84	23.07	22.23	26.87	49.09
7	161,958	28.67	31.36	28.02	11.95	39.97
8	154,663	19.01	33.33	27.11	20.56	47.66

9.2.1.1. ELA Grade 3

Table 9.18 presents the ELA Grade 3 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 50% of students achieved Level III and Level IV. About 56% of Female students were at Level III or above, as compared to 45% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (70%) students and students from Low Needs districts (71%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 25–42% of students in those same performance categories. Only about 19% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (50%), Female (56%), Asian (70%), Multiracial (54%), Pacific Islander (58%), White (57%) students, and those enrolled in Low Needs (71%) Needs districts and Charter (65%) schools.

Table 9.18. ELA Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	182,885	17.99	31.61	43.23	7.17	50.40
Gender	Female	90,155	14.76	29.54	46.70	9.00	55.70
	Male	92,730	21.13	33.62	39.86	5.39	45.25
Ethnicity	Asian	17,913	8.21	21.57	53.81	16.41	70.22
	Black	31,910	24.08	34.24	36.60	5.09	41.69
	Hispanic	51,096	23.11	36.32	36.32	4.25	40.57
	American Indian	1,263	21.06	35.63	36.82	6.49	43.31
	Multiracial	5,223	15.95	29.91	45.01	9.13	54.15
	Pacific Islander	425	14.59	27.29	48.94	9.18	58.12
	White	72,941	13.41	29.55	49.14	7.89	57.03
NRC	New York	67,325	18.21	31.20	41.70	8.88	50.59
	Big 4 Cities	7,898	40.53	33.98	23.54	1.95	25.49
	Urban/Suburban	14,389	28.65	37.82	31.02	2.51	33.53
	Rural	10,027	22.42	38.15	36.34	3.09	39.43
	Average Needs	42,841	15.64	34.13	45.02	5.21	50.23
	Low Needs	18,448	6.30	22.84	59.87	10.99	70.86
	Charter	12,276	9.81	25.18	53.06	11.95	65.01
	Religious and Independent	9,624	20.69	30.25	43.15	5.91	49.06
SWD	All Codes	26,715	42.33	36.92	19.44	1.32	20.76
SUA	All Codes	12,177	44.88	38.06	16.54	0.52	17.06
ELL/MLL	ELL=Y	21,353	40.91	40.35	18.10	0.64	18.74
SWD/ SUA	SWD & SUA codes	8,706	47.12	37.33	15.22	0.33	15.55

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
ELL/MLL /SUA	SUA & ELL codes	1,068	55.62	35.77	8.43	0.19	8.61

9.2.1.2. ELA Grade 4

Table 9.19 presents the ELA Grade 4 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 47% of students achieved Level III and Level IV. About 52% of Female students were at Level III or above, as compared to 43% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (69%) students and students from Low Needs districts (66%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 20–39% of students in those same performance categories. Only about 14% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (47%): Female (52%), Asian (69%), Multiracial (50%), Pacific Islander (55%), and White (53%) students as well as those enrolled in New York City (49%) and Low (66%) Needs districts, and Charter schools (62%).

Table 9.19. ELA Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	184,266	19.56	33.04	29.51	17.89	47.40
Gender	Female	90,669	16.14	31.49	31.01	21.36	52.37
	Male	93,597	22.87	34.54	28.06	14.52	42.58
Ethnicity	Asian	18,629	9.09	22.19	32.95	35.77	68.72
	Black	32,693	25.15	35.55	26.26	13.04	39.30
	Hispanic	50,723	24.54	37.15	26.46	11.85	38.31
	American Indian	1,279	24.32	35.97	25.57	14.15	39.72
	Multiracial	4,851	18.86	31.15	29.66	20.33	49.99
	Pacific Islander	502	15.74	28.88	31.67	23.71	55.38
	White	73,442	15.33	31.90	32.79	19.98	52.77
NRC	New York	67,657	19.33	31.36	27.45	21.87	49.32
	Big 4 Cities	7,874	45.28	35.04	14.85	4.84	19.69
	Urban/Suburban	13,903	30.89	38.78	22.40	7.93	30.32
	Rural	9,972	26.72	39.53	24.85	8.89	33.74
	Average Needs	41,234	17.48	36.59	31.66	14.27	45.93
	Low Needs	18,378	6.98	26.68	40.06	26.28	66.34
NRC	Charter	11,436	10.20	27.37	38.27	24.17	62.43
	Religious and Independent	13,774	20.23	32.29	30.80	16.69	47.49
SWD	All Codes	27,585	47.41	36.03	12.80	3.76	16.56

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
SUA	All Codes	13,737	50.38	36.57	10.88	2.18	13.05
ELL/MLL	ELL=Y	17,775	49.13	39.27	10.12	1.48	11.60
SWD/ SUA	SWD & SUA codes	9,762	54.29	35.13	9.10	1.49	10.58
ELL/MLL /SUA	SUA & ELL codes	1,038	62.62	33.24	3.56	0.58	4.14

9.2.1.3. ELA Grade 5

Table 9.20 presents the ELA Grade 5 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 36% of students achieved Level III and Level IV. About 41% of Female students were at Level III or above, as compared to 32% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (58%) students and students from Low Needs districts (56%). The Big 4 Cities, High Needs/Urban/ Suburban, Black, and Hispanic students had a range of 14–27% of students in those same performance categories. Only about 3% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (36%): Female (41%), Asian (58%), Multiracial (40%), Pacific Islander (47%), and White (42%) students, as well as those enrolled in New York City (38%), Low Needs (56%) districts and Charter schools (44%).

Table 9.20. ELA Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	177,609	33.28	30.30	22.31	14.10	36.41
Gender	Female	87,756	27.80	30.84	24.29	17.06	41.35
	Male	89,853	38.64	29.78	20.37	11.21	31.58
Ethnicity	Asian	18,755	17.00	24.75	29.10	29.15	58.26
	Black	32,041	42.63	30.64	18.01	8.72	26.73
	Hispanic	49,253	40.72	31.94	19.06	8.28	27.34
	American Indian	1,264	40.19	30.78	19.46	9.57	29.03
	Multiracial	4,345	31.02	28.63	22.67	17.68	40.35
	Pacific Islander	541	22.00	31.05	23.84	23.11	46.95
	White	69,491	27.15	30.85	25.24	16.76	42.00
NRC	New York	68,524	32.87	29.16	21.84	16.12	37.97
	Big 4 Cities	7,610	62.73	23.48	9.83	3.96	13.78
	Urban/Suburban	13,174	47.67	30.71	15.85	5.77	21.62
	Rural	9,433	44.19	32.21	16.70	6.91	23.61
	Average Needs	39,498	31.32	33.29	23.16	12.23	35.39
	Low Needs	18,404	14.48	29.40	32.37	23.75	56.12

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	Charter	11,234	24.09	31.92	27.06	16.93	43.99
	Religious and Independent	9,729	37.28	29.03	21.54	12.15	33.69
SWD	All Codes	27,838	68.26	22.28	7.23	2.23	9.46
SUA	All Codes	13,986	71.01	21.27	6.03	1.69	7.72
ELL/MLL	ELL=Y	14,867	78.93	17.82	3.01	0.24	3.25
SWD/ SUA	SWD & SUA codes	10,046	75.55	19.07	4.46	0.92	5.38
ELL/MLL /SUA	SUA & ELL codes	1,064	87.22	11.75	1.03	0.00	1.03

9.2.1.4. ELA Grade 6

Table 9.21 presents the ELA Grade 6 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 49% of students achieved Level III and Level IV. About 57% of Female students were at Level III or above, as compared to 42% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (70%) students and students from Low Needs districts (71%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 23–39% of students in those same performance categories. Only about 12% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (49%): Female (57%), Asian (70%), Multiracial (54%), Pacific Islander (58%), and White (57%) students, as well as those from Average (50%) and Low (71%) Needs districts, and Charter (53%) and Religious and Independent (51%) schools.

Table 9.21. ELA Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	173,183	27.84	23.07	22.23	26.87	49.09
Gender	Female	84,878	21.20	22.18	23.71	32.91	56.61
	Male	88,305	34.21	23.92	20.81	21.06	41.87
Ethnicity	Asian	18,129	13.67	15.95	22.68	47.70	70.38
	Black	32,033	36.98	26.13	19.46	17.43	36.89
	Hispanic	47,601	34.51	26.31	20.97	18.22	39.19
	American Indian	1,167	32.48	23.39	21.59	22.54	44.13
	Multiracial	3,822	24.44	21.82	21.95	31.79	53.74
	Pacific Islander	620	17.42	24.19	21.13	37.26	58.39
	White	67,598	21.71	21.41	24.64	32.23	56.87

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	New York	65,208	28.00	23.08	20.20	28.72	48.92
	Big 4 Cities	6,993	55.96	21.44	13.81	8.79	22.61
	Urban/Suburban	12,437	43.76	25.17	17.68	13.38	31.06
	Rural	9,182	36.05	25.14	22.17	16.64	38.82
	Average Needs	37,022	25.48	24.09	24.80	25.63	50.42
	Low Needs	17,655	10.80	17.88	26.93	44.40	71.33
	Charter	11,540	21.25	25.39	26.61	26.75	53.36
	Religious and Independent	12,939	26.55	22.56	23.63	27.27	50.89
SWD	All Codes	26,971	63.03	22.86	9.63	4.48	14.11
SUA	All Codes	13,408	65.21	21.39	9.40	4.01	13.40
ELL/MLL	ELL=Y	13,822	72.54	20.32	5.81	1.32	7.13
SWD/ SUA	SWD & SUA codes	9,431	70.72	19.87	6.93	2.47	9.41
ELL/MLL /SUA	SUA & ELL codes	975	80.00	15.69	3.69	0.62	4.31

9.2.1.5. ELA Grade 7

Table 9.22 presents the ELA Grade 7 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 40% of students achieved Level III and Level IV. About 47% of Female students were at Level III or above, as compared to 33% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (65%) students and students from Low Needs (60%) districts. The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 15–30% of students in those same performance categories. Only about 7% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (40%): Female (47%), Asian (65%), Multiracial (45%), Pacific Islander (51%), and White (46%) students, as well as those enrolled in New York City (43%), Low Needs (62%) districts, and Charter (42%) schools.

Table 9.22. ELA Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	161,958	28.67	31.36	28.02	11.95	39.97
Gender	Female	78,711	21.54	31.59	32.03	14.85	46.88
	Male	83,247	35.42	31.15	24.22	9.21	33.43
Ethnicity	Asian	17,363	13.65	21.52	35.61	29.22	64.83
	Black	30,819	37.34	34.26	22.84	5.56	28.39
	Hispanic	44,352	34.56	35.41	23.59	6.44	30.03

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
Ethnicity	American Indian	1,236	35.52	30.74	24.03	9.71	33.74
	Multiracial	3,097	27.99	27.06	28.54	16.40	44.95
	Pacific Islander	482	21.78	27.18	34.65	16.39	51.04
	White	62,775	23.28	30.26	32.16	14.30	46.46
NRC	New York	65,334	26.39	31.00	27.60	15.00	42.61
	Big 4 Cities	6,554	58.10	26.67	12.37	2.85	15.23
	Urban/Suburban	11,075	48.62	31.08	16.35	3.95	20.30
	Rural	8,494	38.31	34.08	22.30	5.31	27.61
	Average Needs	33,387	28.62	33.05	28.81	9.52	38.33
	Low Needs	17,179	13.37	26.71	40.00	19.91	59.92
	Charter	10,617	21.21	36.73	33.43	8.63	42.06
	Religious and Independent	9,118	28.37	31.38	29.85	10.40	40.25
SWD	All Codes	25,931	61.78	28.33	8.58	1.32	9.90
SUA	All Codes	12,802	66.27	24.89	7.66	1.18	8.83
ELL/MLL	ELL=Y	11,704	76.40	20.45	2.91	0.24	3.15
SWD/ SUA	SWD & SUA codes	9,017	71.35	22.83	5.32	0.49	5.81
ELL/MLL /SUA	SUA & ELL codes	787	83.74	15.50	0.76	0.00	0.76

9.2.1.6. ELA Grade 8

Table 9.23 presents the ELA Grade 8 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 48% of students achieved Level III and Level IV. About 55% of Female students were at Level III or above, as compared to 41% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (71%) students and students from Low Needs (64%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 20–39% of students in those same performance categories. Only about 11% of the SWD, SUA, and ELL/MLL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (48%): Female (55%), Asian (71%), Multiracial (52%), Pacific Islander (61%), and White (52%) students, as well as those attending New York City (51%) and Low Needs (64%) districts, those enrolled in Charter (55%), and Religious and Independent (50%) schools.

Table 9.23. ELA Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	154,663	19.01	33.33	27.11	20.56	47.66
Gender	Female	74,920	12.88	31.88	30.08	25.17	55.24
	Male	79,743	24.77	34.69	24.32	16.23	40.54
Ethnicity	Asian	17,639	9.40	20.09	28.57	41.94	70.51
	Black	29,862	23.55	38.83	25.17	12.45	37.62
	Hispanic	41,867	22.56	38.10	25.61	13.72	39.34
	American Indian	1,203	21.45	36.82	25.94	15.79	41.73
	Multiracial	2,443	19.32	28.94	26.65	25.09	51.74
	Pacific Islander	457	14.66	24.51	31.73	29.10	60.83
	White	59,660	16.34	31.24	29.09	23.33	52.42
NRC	New York	63,216	16.47	32.80	26.92	23.81	50.73
	Big 4 Cities	6,388	45.95	34.19	13.87	6.00	19.87
	Urban/Suburban	10,000	34.30	38.17	18.44	9.09	27.53
	Rural	8,047	26.31	38.70	23.70	11.30	34.99
	Average Needs	30,019	19.91	34.58	27.57	17.94	45.50
	Low Needs	15,970	9.13	27.04	32.98	30.85	63.83
	Charter	9,644	10.16	34.58	35.60	19.66	55.26
	Religious and Independent	10,988	18.23	31.76	29.13	20.88	50.01
SWD	All Codes	23,999	46.52	39.59	11.03	2.86	13.89
SUA	All Codes	11,628	50.94	36.08	10.30	2.68	12.99
ELL/MLL	ELL=Y	11,224	63.59	31.20	4.75	0.46	5.21
SWD/ SUA	SWD & SUA codes	8,365	56.71	34.18	7.64	1.47	9.11
ELL/MLL SUA	SUA & ELL codes	683	71.30	25.62	2.78	0.29	3.07

9.2.2. *Mathematics Test Performance Level Distributions*

Table 9.24 shows the performance level distributions for all examinees from public, charter, and non-public schools with valid scores, and presents Mathematics performance level data for total populations of students in Grades 3–8. Performance level data for selected subgroups of students were also examined. In general, these summaries reflect the same achievement trends as in the scale score summary discussion. Across Table 9.25 through Table 9.30, Male and Female students performed similarly across grades. More White, Pacific Islander, and Asian students were classified in Level III and above, as compared to their peers from other ethnic subgroups. Students from Low and Average Needs districts and Charter schools outperformed students from High Needs districts (New York City, Big 4 Cities, High Needs Urban/Suburban, and High Needs Rural), and Religious and Independent schools. The subgroups that used the Korean or Chinese

translations outperformed other test translation subgroups. The Level III and above rates for SWD and SUA subgroups were low, compared to the total population of examinees. The n-counts for the Haitian-Creole, Korean, and Russian translation subgroups were very low, and the results might have been heavily influenced by very high and/or very low achieving individual students.

Table 9.24. Mathematics Test Performance Level Distributions

		Performance Levels				
Grade	**N-Count**	**Level I**	**Level II**	**Level III**	**Level IV**	**Level III & IV**
3	184,970	24.52	22.15	30.47	22.86	53.33
4	186,331	26.09	26.33	22.96	24.61	47.58
5	178,875	33.20	23.78	22.86	20.16	43.02
6	173,731	31.33	24.94	22.72	21.01	43.73
7	160,487	33.38	25.70	22.97	17.96	40.93
8	116,534	38.18	31.47	18.14	12.21	30.36

9.2.2.1. Mathematics Grade 3

Table 9.25 presents the Mathematics Grade 3 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 53% of students achieved Level III and Level IV. About 53% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (77%) students and students from Low Needs (74%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 27–42% of students in those same performance categories. Only about 24% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (53%): Asian (77%), Multiracial (56%), Pacific Islander (61%), and White (61%) students, as well as those enrolled at Average (55%) and Low (74%) Needs districts and Charter schools (72%). For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 9% (Haitian-Creole) to 85% (Chinese).

Table 9.25. Mathematics Grade 3 Performance Level Distribution by Subgroup

			Performance Levels				
Demographic Category		**N-Count**	**Level I**	**Level II**	**Level III**	**Level IV**	**Level III & IV**
State	All Students	184,970	24.52	22.15	30.47	22.86	53.33
Gender	Female	90,724	23.70	22.91	31.17	22.21	53.38
	Male	94,246	25.30	21.41	29.79	23.49	53.29
Ethnicity	Asian	18,468	9.07	14.06	31.41	45.46	76.87
	Black	32,048	34.42	23.51	25.23	16.83	42.07
	Hispanic	52,069	32.55	25.18	27.58	14.70	42.28
	American Indian	1,304	27.68	24.69	29.29	18.33	47.62
	Multiracial	5,226	23.08	21.03	31.13	24.76	55.89

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
Ethnicity	Pacific Islander	436	18.35	20.87	35.32	25.46	60.78
	White	73,811	17.79	21.44	34.86	25.91	60.77
NRC	New York	68,732	25.71	22.08	28.90	23.31	52.21
	Big 4 Cities	8,089	50.71	22.01	19.04	8.25	27.28
	Urban/Suburban	14,507	38.77	25.32	25.04	10.87	35.91
	Rural	9,917	28.68	25.49	30.53	15.30	45.83
	Average Needs	42,715	21.15	23.77	34.10	20.98	55.08
	Low Needs	18,513	9.00	17.24	37.72	36.03	73.75
	Charter	12,230	11.75	15.77	29.82	42.66	72.48
	Religious and Independent	10,103	29.01	24.58	30.15	16.26	46.41
SWD	All Codes	27,469	53.49	21.48	17.01	8.03	25.04
SUA	All Codes	12,974	58.70	20.66	15.06	5.58	20.64
ELL/MLL	ELL=Y	23,775	48.39	24.96	19.37	7.28	26.65
SWD/ SUA	SWD & SUA codes	9,625	62.88	18.84	13.61	4.68	18.29
ELL/MLL/SUA	SUA & ELL codes	1,289	67.18	16.99	11.87	3.96	15.83
ELL/MLL Test Language	Chinese	102	8.82	5.88	36.27	49.02	85.29
	English	183,028	24.18	22.15	30.63	23.05	53.67
	Haitian-Creole	34	67.65	23.53	5.88	2.94	8.82
	Korean	25	32.00	20.00	32.00	16.00	48.00
	Russian	13	38.46	23.08	30.77	7.69	38.46
	Spanish	1,768	59.84	22.85	14.31	3.00	17.31
	All Translations	184,970	24.52	22.15	30.47	22.86	53.33

9.2.2.2. Mathematics Grade 4

Table 9.26 presents the Mathematics Grade 4 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 48% of students achieved Level III and Level IV. About 48% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (74%) students and students from Low Needs (70%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 20–36% of students in those same performance categories. Only about 17% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (48%): Asian (74%), Multiracial (50%), Pacific Islander (54%), and White (55%) students, as well as students enrolled in Average (51%) and Low (70%) Needs and Charter schools (65%). For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 3% (Haitian-Creole) to 71% (Korean).

Table 9.26. Mathematics Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	186,331	26.09	26.33	22.96	24.61	47.58
Gender	Female	91,167	25.86	26.86	22.80	24.47	47.28
	Male	95,164	26.31	25.82	23.12	24.75	47.87
Ethnicity	Asian	19,181	9.58	16.76	23.50	50.16	73.66
	Black	32,747	37.66	27.72	17.77	16.85	34.62
	Hispanic	51,671	34.50	29.79	20.07	15.64	35.71
	American Indian	1,309	32.31	27.58	20.55	19.56	40.11
	Multiracial	4,880	25.47	24.96	23.03	26.54	49.57
	Pacific Islander	521	20.54	25.91	23.61	29.94	53.55
	White	74,330	18.69	25.87	27.39	28.05	55.44
NRC	New York	68,895	27.68	25.97	20.57	25.78	46.36
	Big 4 Cities	8,027	55.62	24.22	12.38	7.77	20.16
	Urban/Suburban	14,040	41.16	29.16	17.89	11.79	29.68
	Rural	9,845	30.61	30.55	23.18	15.65	38.83
	Average Needs	41,365	20.93	28.36	27.36	23.35	50.71
	Low Needs	18,564	9.49	20.36	30.74	39.41	70.16
	Charter	11,413	14.36	20.60	23.06	41.98	65.04
	Religious and Independent	14,040	29.84	30.00	22.42	17.74	40.16
SWD	All Codes	28,137	57.72	24.57	11.02	6.69	17.71
SUA	All Codes	14,755	60.02	24.24	10.70	5.04	15.74
ELL/MLL	ELL=Y	20,144	55.82	27.51	11.37	5.30	16.67
SWD/ SUA	SWD & SUA codes	10,824	65.65	21.78	8.56	4.01	12.56
ELL/MLL/SUA	SUA & ELL codes	1,297	74.63	18.20	5.09	2.08	7.17
ELL/MLL Test Language	Chinese	93	4.30	25.81	29.03	40.86	69.89
	English	184,631	25.75	26.35	23.10	24.80	47.90
	Haitian-Creole	31	74.19	22.58	3.23	0.00	3.23
	Korean	21	4.76	23.81	38.10	33.33	71.43
	Russian	12	33.33	25.00	16.67	25.00	41.67
	Spanish	1,543	67.92	23.66	6.48	1.94	8.43
	All Translations	186,331	26.09	26.33	22.96	24.61	47.58

9.2.2.3. Mathematics Grade 5

Table 9.27 presents the Mathematics Grade 5 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 43% of students achieved Level III and Level IV. About 43% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (71%) students and students from Low Needs districts (68%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 18–31% of students in those same performance categories. Only about 13% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (43%): Asian (71%), Multiracial (47%), Pacific Islander (54%), and White (52%) students, as well as those enrolled in Average (47%) and Low (68%) Needs districts and Charter schools (50%). For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 6% (Spanish) to 72% (Chinese).

Table 9.27. Mathematics Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	178,875	33.20	23.78	22.86	20.16	43.02
Gender	Female	87,953	32.30	25.25	23.06	19.39	42.45
	Male	90,922	34.07	22.36	22.66	20.90	43.56
Ethnicity	Asian	19,173	12.41	16.78	25.39	45.42	70.81
	Black	31,917	48.38	24.63	16.78	10.21	26.99
	Hispanic	49,798	43.00	26.22	19.18	11.60	30.79
	American Indian	1,294	42.35	25.04	19.24	13.37	32.61
	Multiracial	4,271	31.47	21.85	22.59	24.09	46.69
	Pacific Islander	558	23.66	22.04	27.60	26.70	54.30
	White	70,297	24.30	23.81	27.87	24.01	51.89
NRC	New York	69,433	34.76	23.56	20.38	21.30	41.68
	Big 4 Cities	7,659	64.13	18.27	11.61	5.99	17.60
	Urban/Suburban	13,167	50.30	23.65	17.22	8.83	26.05
	Rural	9,297	38.86	27.30	22.24	11.60	33.84
	Average Needs	38,729	27.39	25.47	26.86	20.28	47.13
	Low Needs	18,457	12.30	20.04	32.63	35.03	67.66
	Charter	11,167	25.12	25.16	24.79	24.93	49.72
	Religious and Independent	10,376	40.51	24.97	20.97	13.55	34.52
SWD	All Codes	27,878	67.22	18.63	9.62	4.53	14.15
SUA	All Codes	14,352	68.50	18.28	9.61	3.61	13.22
ELL/MLL	ELL=Y	16,975	69.02	19.75	8.08	3.15	11.23

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
SWD/ SUA	SWD & SUA codes	10,526	74.11	16.18	7.34	2.37	9.71
ELL//MLL/SUA	SUA & ELL codes	1,184	83.70	12.33	3.46	0.51	3.97
ELL/MLL Test Language	Chinese	68	10.29	17.65	26.47	45.59	72.06
	English	177,539	32.89	23.85	22.98	20.28	43.26
	Haitian-Creole	25	76.00	16.00	8.00	0.00	8.00
	Korean	15	20.00	20.00	20.00	40.00	60.00
	Russian	18	55.56	22.22	16.67	5.56	22.22
	Spanish	1,210	79.09	14.46	5.12	1.32	6.45
	All Translations	178,875	33.20	23.78	22.86	20.16	43.02

9.2.2.4. Mathematics Grade 6

Table 9.28 presents the Mathematics Grade 6 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 44% of students achieved Level III and Level IV. About 44% of Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (71%) students and students from Low Needs districts (71%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 20–30% of students in those same performance categories. Only about 13% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (44%): Asian (71%), Multiracial (49%), Pacific Islander (50%), and White (55%) students, as well as those enrolled in Average (51%) and Low (71%) Needs districts and Charter schools (50%). For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 4% (Haitian-Creole) to 72% (Chinese).

Table 9.28. Mathematics Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	173,731	31.33	24.94	22.72	21.01	43.73
Gender	Female	84,789	30.22	26.16	23.06	20.55	43.61
	Male	88,942	32.39	23.77	22.39	21.45	43.84
Ethnicity	Asian	18,551	12.62	16.81	24.10	46.47	70.57
	Black	31,805	46.84	26.80	16.75	9.61	26.35
	Hispanic	47,969	41.85	28.20	19.10	10.86	29.95
	American Indian	1,219	38.88	24.86	21.00	15.26	36.26
	Multiracial	3,757	27.81	22.86	23.29	26.03	49.32
	Pacific Islander	634	23.19	26.97	21.45	28.39	49.84
	White	68,127	21.29	24.07	27.97	26.67	54.64

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	New York	66,041	35.28	24.88	19.31	20.53	39.84
	Big 4 Cities	7,072	58.77	21.63	12.26	7.34	19.60
	Urban/Suburban	12,292	50.11	25.43	16.32	8.14	24.46
	Rural	9,015	35.17	27.89	22.84	14.10	36.94
	Average Needs	36,269	23.19	26.31	27.60	22.90	50.50
	Low Needs	17,487	10.42	18.52	31.03	40.02	71.06
	Charter	11,454	24.59	25.94	26.51	22.97	49.48
	Religious and Independent	13,389	33.06	28.43	23.24	15.26	38.50
SWD	All Codes	26,769	68.41	19.47	8.32	3.80	12.11
SUA	All Codes	14,218	65.68	20.40	9.61	4.30	13.92
ELL/MLL	ELL=Y	15,906	67.72	20.63	8.20	3.46	11.66
SWD/ SUA	SWD & SUA codes	10,038	72.81	17.35	7.18	2.65	9.83
ELL/MLL/SUA	SUA & ELL codes	1,238	81.42	13.81	3.47	1.29	4.77
ELL/MLL Test Language	Chinese	47	10.64	17.02	27.66	44.68	72.34
	English	172,211	30.93	25.02	22.87	21.18	44.05
	Haitian-Creole	27	88.89	7.41	3.70	0.00	3.70
	Korean	11	18.18	27.27	27.27	27.27	54.55
	Russian	14	57.14	21.43	7.14	14.29	21.43
	Spanish	1,421	79.59	15.48	4.15	0.77	4.93
	All Translations	173,731	31.33	24.94	22.72	21.01	43.73

9.2.2.5. Mathematics Grade 7

Table 9.29 presents the Mathematics Grade 7 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 41% of students achieved Level III and Level IV. About 43% of Female students were at Level III or above, as compared to 39% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (71%) students and students from Low Needs districts (67%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 16–28% of students in those same performance categories. Only about 10% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (41%): Female (43%), Asian (71%), Multiracial (46%), Pacific Islander (49%), and White (51%) students, as well as those enrolled in Average (45%) and Low (67%) Needs districts and Charter schools (48%). For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 3% (Spanish) to 78% (Korean).

Table 9.29. Mathematics Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	160,487	33.38	25.70	22.97	17.96	40.93
Gender	Female	77,750	30.39	26.58	24.24	18.79	43.03
	Male	82,737	36.19	24.87	21.77	17.18	38.95
Ethnicity	Asian	17,451	12.47	16.77	26.21	44.55	70.76
	Black	30,218	49.67	25.72	16.05	8.55	24.60
	Hispanic	44,513	43.86	28.61	18.32	9.21	27.54
	American Indian	1,253	43.58	24.58	20.03	11.81	31.84
	Multiracial	3,005	30.05	23.66	24.29	22.00	46.29
	Pacific Islander	483	28.16	23.19	26.09	22.57	48.65
	White	62,214	23.01	26.43	29.00	21.56	50.56
NRC	New York	65,783	35.74	24.47	19.88	19.91	39.79
	Big 4 Cities	6,491	65.37	19.00	10.74	4.90	15.64
	Urban/Suburban	10,759	55.95	26.86	12.91	4.28	17.19
	Rural	8,190	38.75	31.47	21.77	8.01	29.78
	Average Needs	31,907	26.50	28.32	28.78	16.40	45.18
	Low Needs	16,621	11.86	21.38	33.82	32.93	66.76
	Charter	10,548	25.68	26.76	25.56	21.99	47.55
	Religious and Independent	9,627	34.43	29.97	23.22	12.38	35.60
SWD	All Codes	25,434	70.40	19.43	7.48	2.69	10.17
SUA	All Codes	12,705	69.17	19.28	8.69	2.87	11.55
ELL/MLL	ELL=Y	13,657	71.39	18.99	7.01	2.61	9.62
SWD/ SUA	SWD & SUA codes	9,268	74.65	17.12	6.46	1.76	8.22
ELL/MLL/SUA	SUA & ELL codes	881	86.04	10.78	3.18	0.00	3.18
ELL/MLL Test Language	Chinese	70	18.57	5.71	34.29	41.43	75.71
	English	159,040	32.94	25.83	23.13	18.09	41.23
	Haitian-Creole	33	75.76	18.18	6.06	0.00	6.06
	Korean	9	11.11	11.11	22.22	55.56	77.78
	Russian	45	71.11	15.56	6.67	6.67	13.33
	Spanish	1,290	85.89	10.70	2.87	0.54	3.41
	All Translations	160,487	33.38	25.70	22.97	17.96	40.93

9.2.2.6. Mathematics Grade 8

Table 9.30 presents the Mathematics Grade 8 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 30% of students achieved Level III and Level IV. About 33% of Female students were at Level III or above, as compared to 28% of Male

students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (62%) students and students from Low Needs districts (47%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 11–23% of students in those same performance categories. Only about 9% of the SWD, SUA, and ELL/MLL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (30%): Female (33%), Asian (62%), Multiracial (31%), Pacific Islander (42%), and White (34%) students, as well as those enrolled in New York City (33%) and Low Needs districts (47%), and Charter (43%) and Religious and Independent (35%) schools. For ELL/MLL students who used translated test forms, the percentages of students earning at least a Level III ranged from 0% (Haitian-Creole) to 47% (Chinese).

Table 9.30. Mathematics Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	116,534	38.18	31.47	18.14	12.21	30.36
Gender	Female	55,578	33.88	32.94	19.85	13.33	33.18
	Male	60,956	42.09	30.12	16.59	11.19	27.78
Ethnicity	Asian	10,984	15.63	22.21	23.25	38.90	62.15
	Black	23,918	50.41	29.61	13.21	6.76	19.97
	Hispanic	34,909	45.00	31.76	14.96	8.28	23.24
	American Indian	847	45.22	33.06	14.29	7.44	21.72
	Multiracial	1,718	38.24	31.02	17.05	13.68	30.73
	Pacific Islander	354	27.97	30.23	24.01	17.80	41.81
	White	42,911	30.78	34.86	22.52	11.84	34.36
NRC	New York	49,766	37.33	29.44	17.08	16.14	33.22
	Big 4 Cities	5,684	70.13	18.77	7.42	3.68	11.10
	Urban/Suburban	7,896	60.25	28.53	8.71	2.51	11.22
	Rural	6,523	44.40	35.29	16.08	4.23	20.31
	Average Needs	19,925	35.30	38.59	20.32	5.79	26.11
	Low Needs	8,588	18.77	34.51	29.45	17.27	46.72
	Charter	6,898	26.92	29.79	22.92	20.37	43.29
	Religious and Independent	10,847	32.86	32.45	21.13	13.56	34.69
SWD	All Codes	21,321	70.95	21.34	5.48	2.23	7.71
SUA	All Codes	10,993	69.51	21.89	6.37	2.24	8.61
ELL/MLL	ELL=Y	12,091	67.19	22.25	6.83	3.73	10.56
SWD/ SUA	SWD & SUA codes	8,099	75.29	18.83	4.37	1.51	5.88
ELL/MLL/SUA	SUA & ELL codes	761	85.15	13.14	1.31	0.39	1.71

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
ELL/MLL Test Language	Chinese	32	21.88	31.25	25.00	21.88	46.88
	English	115,328	37.69	31.67	18.31	12.33	30.64
	Haitian-Creole	28	82.14	17.86	0.00	0.00	0.00
	Korean	9	22.22	33.33	11.11	33.33	44.44
	Russian	55	89.09	7.27	1.82	1.82	3.64
	Spanish	1,082	87.25	10.91	1.39	0.46	1.85
	All Translations	116,534	38.18	31.47	18.14	12.21	30.36

Section 10: References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* 37: 29–51.

Bock, R.D. & M. Aitkin (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika* 46: 443–459.

Cattell, R.B. (1966). The Screen Test for the Number of Factors. *Multivariate Behavioral Research* 1:245–276.

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

Dorans, N.J., A.P. Schmitt & C.A. Bleistein (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement* 29:309–319.

Dorans, N.J. & P. W. Holland (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum.

Fleiss J.L. & J. Cohen (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33: 613–619.

Green, D.R., W.M. Yen & G.R. Burket (1989). Experiences in the application of item response theory in test construction. *Applied Measurement in Education* 2: 297–312.

Huynh, H. & C. Schneider (2004). *Vertically moderated standards as an alternative to vertical scaling: assumptions, practices, and an odyssey through NAEP*. Paper presented at the National Conference on Large-Scale Assessment. Boston, MA, June 21.

Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.

Johnson, N.L. & S. Kotz (1970). *Distributions in Statistics: Continuous Univariate Distributions,* Vol. 2. New York: John Wiley.

Kim, S. & M. J. Kolen (2004). *STUIRT: A computer program for scale transformation under unidimensional item response theory models*. Iowa City, IA: Iowa Testing Programs, The University of Iowa.

Kolen, M.J. & Z. Cui (2004). *POLYEQUATE*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

Kolen, M.J. & R.L. Brennan (1995). *Test Equating: Methods and Practices.* New York: Springer-Verlag.

Landis, J. R. & G. G. Koch. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174.

Lee, W. C., B.A. Hanson & R.L. Brennan (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement* 26:412–432.

Lee, W. C. (2008). *Classification consistency and accuracy for complex assessments using item response theory.* (CASMA Research Report No. 27). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

Lee, W. C. & M. J. Kolen (2006, Revised 2008). IRT-CLASS (Version 2.0). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

Linn, R.L. (1991). Linking results of distinct assessments. *Applied Measurement in Education* 6(1): 83–102.

Linn, R.L. & D. Harnisch (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurem*ent 18: 109–118.

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

Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Lawrence Erlbaum.

Lord, F.M. & M.R. Novick (1968). *Statistical Theories of Mental Test Scores.* Menlo Park, CA: Addison-Wesley.

Mehrens, W.A. & I.J. Lehmann (1991). *Measurement and Evaluation in Education and Psychology, 3rd ed*. New York: Holt, Rinehart, and Winston.

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

Muraki, E. & R.D. Bock (1991). *PARSCALE*: *Parameter Scaling of Rating Data* [Computer program]. Chicago, IL: Scientific Software, Inc.

Novick, M.R. & P.H. Jackson (1974). *Statistical Methods for Educational and Psychological Research.* New York: McGraw-Hill.

NYSED. (2013) New York State Testing Program 2013: English Language Arts and Mathematics Grades 3–8 Technical Report. Albany, NY: New York State Education Department (NYSED). Retrieved from: http://www.p12.nysed.gov/assessment/reports/2013/ela-math-tr13.pdf

Qualls, A.L. (1995). Estimating the reliability of a test containing multiple-item formats. *Applied Measurement in Education* 8: 111–120.

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

Sandoval, J.H. & M.P. Mille (1979) *Accuracy of judgments of WISC-R item difficulty for minority groups*. Paper presented at the annual meeting of the American Psychological Association, New York. August.

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

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* 47: 175–186.

Cai, L., Thissen, D. J., & du Toit, S. (2011). IRTPRO (Version 2.1). Skokie, IL: Scientific Software International, Inc.

Thompson, S.J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal Design Applied to Large Scale Assessments (NCEO Synthesis Report 44)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from: http://www.cehd.umn.edu/nceo/onlinepubs/Synthesis44.html.

Wang, T.M., J. Kolen, & D.J. Harris (2000). Psychometric properties of scale scores and performance levels for performance assessment using polytomous IRT. *Journal of Educational Measurement* 37: 141–162.

Yen, W.M. (1997). The technical quality of performance assessments: Standard errors of percents of students reaching standards. *Educational Measurement: Issues and Practice*: 5–15.

Yen, W.M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement* 30: 187–213.

Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number correct scores for the three-parameter logistic model. *Journal of Educational Measurement* 21: 93–111.

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

Yen, W.M., R.C. Sykes, K. Ito & M. Julian (1997). *A Bayesian/IRT index of objective performance for tests with mixed-item types*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago: March.

Zwick, R., J.R. Donoghue & A. Grima, (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement* 36: 225–33.

Appendix A: ELA and Mathematics Test Configurations

Table A1. ELA Test Configuration

			Number of Items				
			Multiple-Choice		Constructed-Response		
Grade	Day	Session	Operational	Embedded	Operational	Embedded	Total
3	1	1	18	6	0	0	24
	2	2	0	0	7	0	7
		Total	18	6	7	0	31
4	1	1	18	6	0	0	24
	2	2	0	0	7	0	7
		Total	18	6	7	0	31
5	1	1	28	7	0	0	35
	2	2	0	0	7	0	7
		Total	28	7	7	0	42
6	1	1	28	7	0	0	35
	2	2	0	0	7	0	7
		Total	28	7	7	0	42
7	1	1	28	7	0	0	35
	2	2	0	0	8	0	8
		Total	28	7	8	0	43
8	1	1	28	7	0	0	35
	2	2	0	0	8	0	8
		Total	28	7	8	0	43

Table A2. Mathematics Test Configuration

			Number of Items				
			Multiple-Choice		Constructed-Response		
Grade	Day	Session	Operational	Embedded	Operational	Embedded	Total
3	1	1	19	6	0	0	25
	2	2	8	0	7	0	15
		Total	27	6	7	0	40
4	1	1	23	7	0	0	30
	2	2	8	0	7	0	15
		Total	31	7	7	0	45
5	1	1	23	7	0	0	30
	2	2	8	0	7	0	15
		Total	31	7	7	0	45
6	1	1	24	7	0	0	31
	2	2	7	0	8	0	15
		Total	31	7	8	0	46
7	1	1	26	7	0	0	33
	2	2	7	0	8	0	15
		Total	33	7	8	0	48
8	1	1	26	7	0	0	33
	2	2	7	0	8	0	15
		Total	33	7	8	0	48

Table A3. ELA Estimated Time on Task by Session

Grade	Day	Session	Estimated Time on Task (min.)	Previous Session Time (min.)
3	1	1	70	70
	2	2	70	70
		Total	**140**	**140**
4	1	1	70	70
	2	2	70	70
		Total	**140**	**140**
5	1	1	90	90
	2	2	90	90
		Total	**180**	**180**
6	1	1	90	90
	2	2	90	90
		Total	**180**	**180**
7	1	1	90	90
	2	2	90	90
		Total	**180**	**180**
8	1	1	90	90
	2	2	90	90
		Total	**180**	**180**

Source: *2018 ELA and Mathematics Test Guides.*

The ELA estimated times on task were based on the following rules of thumb:

- Average time to read a passage—5 minutes
- Average time to respond to a multiple-choice question—1 minute
- Average time to respond to a two-point constructed response question—3 minutes
- Average time to respond to a four-point constructed response question—20 minutes

Table A4. Mathematics Estimated Time on Task by Session

Grade	Day	Session	Estimated Time on Task (min.)	Previous Session Time (min.)
3	1	1	65	60
	2	2	70	60
		Total	**135**	**120**
4	1	1	75	60
	2	2	75	60
		Total	**150**	**120**
5	1	1	90	80
	2	2	80	80
		Total	**170**	**160**
6	1	1	90	80
	2	2	85	80
		Total	**175**	**160**
7	1	1	90	80
	2	2	85	80
		Total	**175**	**160**

Grade	Day	Session	Estimated Time on Task (min.)	Previous Session Time (min.)
8	1	1	90	80
	2	2	85	80
		Total	**175**	**160**

Source: *2018 ELA and Mathematics Test Guides.*

The Mathematics estimated times on task were based on the following rules of thumb:

- Average time to respond to a multiple-choice question—1.5 minutes
- Average time to respond to a two-point constructed response question—5 minutes
- Average time to respond to a three-point constructed response question—9 minutes

The testing times listed above do not include approximately 10 minutes reserved for preparation at the beginning of each session for handing out materials and reading directions. Additional details on security, scheduling, classroom organization and preparation, test materials, and administration can be found in the *2018 Teacher's Directions* and the *School Administrator's Manual*, which are accessible online:

- *2018 ELA Teacher's Directions*
  - Grades 3–5: http://www.p12.nysed.gov/assessment/sam/ei/td-35ela18.pdf
  - Grades 6–8: http://www.p12.nysed.gov/assessment/sam/ei/td-68ela18.pdf
- *2018 Mathematics Teacher's Directions*
  - Grades 3–5: http://www.p12.nysed.gov/assessment/sam/ei/td-35math18.pdf
  - Grades 6–8: http://www.p12.nysed.gov/assessment/sam/ei/td-68math18.pdf
- *2018 ELA and Mathematics Tests School Administrator's Manual*
  - http://www.p12.nysed.gov/assessment/sam/ei/eisam18b.pdf
- *2018 ELA and Mathematics Test Guides*
  - https://www.engageny.org/resource/test-guides-english-language-arts-and-mathematics

Appendix B: ELA and Mathematics Test Blueprints

Table B1. ELA Test Blueprint

Grade	Total Points on OP Test	Strand	Point Range		% of Test	
			Target	Actual	Target	Actual
3	34	Literature	18	18	53%	53%
		Informational Text	16	16	47%	47%
4	34	Literature	16-18	18	47%-53%	53%
		Informational Text	16-18	16	47%-53%	47%
5	44	Literature	20-24	24	45%-55%	55%
		Informational Text	20-24	20	45%-55%	45%
6	44	Literature	20-24	24	45%-55%	55%
		Informational Text	20-24	20	45%-55%	45%
7	46	Literature	20-26	20	43%-57%	43%
		Informational Text	20-26	26	43%-57%	57%
8	46	Literature	20-26	20	43%-57%	43%
		Informational Text	20-26	26	43%-57%	57%

Table B2. Mathematics Test Blueprint

Grade	Total Points on OP Test	Domain	Point Range		% of Test	
			Target	Actual	Target	Actual
3	42	Operations and Algebraic Thinking	17-21	19	40%-50%	45%
		Number and Operations in Base Ten	2-4	3	5%-10%	7%
		Number and Operations – Fractions	6-10	8	14%-24%	19%
		Measurement and Data	9-13	11	21%-31%	26%
		Geometry*	1-2	1	2%-5%	2%
4	42	Operations and Algebraic Thinking	7-11	9	15%-26%	20%
		Number and Operations in Base Ten	10-14	12	22%-30%	26%
		Number and Operations – Fractions	10-14	12	22%-30%	26%
		Measurement and Data	7-11	9	15%-26%	20%
		Geometry	4-6	4	9%-13%	9%
5	46	Operations and Algebraic Thinking	2-4	2	4%-9%	4%
		Number and Operations in Base Ten	10-14	13	22%-30%	28%
		Number and Operations – Fractions	16-20	18	35%-43%	39%
		Measurement and Data	10-14	12	22%-30%	26%
		Geometry*	1-2	1	2%-4%	2%
6	48	Ratios and Proportional Relationships	10-14	12	21%-29%	25%
		The Number System	9-13	8	19%-27%	17%

Grade	Total Points on OP Test	Domain	Point Range		% of Test	
			Target	Actual	Target	Actual
6	48	Expressions and Equations	16-22	22	33%-46%	46%
		Geometry	5-9	6	10%-19%	12%
7	50	Ratios and Proportional Relationships	12-16	14	24%-32%	28%
		The Number System	8-12	9	16%-24%	18%
		Expressions and Equations	13-19	17	26%-38%	34%
		Geometry	3-5	2	6%-10%	4%
		Statistics and Probability	6-10	8	12%-20%	16%
8	50	Expressions and Equations	18-24	18	36%-48%	36%
		Functions	11-15	15	22%-30%	30%
		Geometry	10-14	12	20%-28%	24%
		Statistics and Probability	3-5	5	6%-10%	10%

*There is a slight difference between the "Target% of Test" shown in these tables and the tables presented in the Guides to the 2018 Mathematics Tests. The guides were intended to provide general guidance regarding content coverage of mathematics domains so that classroom instruction would continue to cover the depth and breadth of the mathematics standards.

Appendix C: Passage Selection Guidelines for Assessing ELA

General Guidelines

The New York State Learning Standards for ELA devote considerable attention to the types and nature of texts used in instruction and assessment. The foundation for preparing students for the linguistic rigors of college and of the workplace lies in the texts with which they interact. By the time that they graduate, students should be prepared to successfully read and analyze the types of complex texts that they will encounter after high school. Selecting passages of appropriate type and complexity for use in assessment is integral to this preparation.

The New York State Learning Standards for ELA emphasize developing skills for comprehending and analyzing both literary and informational texts. Increased exposure to informational texts better prepares students for the various types of texts that they will encounter in college and in the workplace. The array of passages selected for assessment from K–12 should support the development of the necessary skills to handle a range of literary and informational texts.

In addition to the usual fairness and sensitivity guidelines when selecting passages for assessment, attention should be dedicated to three additional considerations:

- *Text Complexity*
- *Text Types*
- *Text Suitability for Specific Standards*

These guidelines should inform the training of passage finders, in order to ensure a pool of acceptable passages that can support assessment of all the Reading Informational Texts standards. They should also alert form assemblers as they construct forms that will assess the complete range of skills.

Appendix D: Universal Design Item Checklist

Universal Design Item Checklist	
A.	**Precisely Designed Constructs**
Definition	The item construct is clearly defined so that all irrelevant cognitive, sensory, emotional, and physical barriers are removed.
√	The item does not add skills to those being measured (no extraneous skills tested).
B.	**Language Appropriateness**
Definition	The item avoids words or phrases that are sexist, racist, or otherwise offensive, inappropriate, or negative to any subgroup. Language should be simple and clear.
√	The item uses commonly used words—simpler is better.
√	The item uses vocabulary appropriate for the grade level.
√	Idiomatic speech and figurative language are avoided unless being measured.
√	The item avoids technical terms unrelated to the content.
√	The item contains no unnecessary words.
√	The sentence complexity contained in the item is appropriate for the grade level.
√	The item avoids ambiguous or multiple-meaning words (e.g., crane—the bird—can easily be confused with crane—heavy machinery).
√	All pronouns have clear referents.
√	The item avoids the use of proper names. (Such names may be unfamiliar or difficult for cultural subgroups.)
√	The item avoids irregularly spelled words.
C.	**Gender Stereotypes**
Definition	The item avoids stereotyping as results of associating genders with certain professions or activities. All groups of society should be portrayed accurately and fairly regarding gender.
√	The item is free of content that might offend a gender subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a gender subgroup.
D.	**Ethnic Stereotypes**
Definition	The item avoids unnecessary references to and uses the proper reference for ethnic, racial, or cultural groups.
√	The item is free of content that might offend an ethnic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage an ethnic subgroup.
√	The artwork included in an item adequately reflects the diversity of the student population.
E.	**Cultural Familiarity**
Definition	Does not rely on an assumed shared experience that is class oriented or native English speaking oriented. Presentations of cultural or ethnic differences should neither explicitly nor implicitly rely on stereotypes nor make moral judgments.
√	The item does not rely on an assumed shared experience that is class oriented or native English speaking oriented.
√	The item is free from content that might offend a socioeconomic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a socioeconomic subgroup.

Universal Design Item Checklist	
√	The item is free from unnecessary cultural references.
√	The item is free from religious references.
F.	**Geographic Bias**
Definition	All groups of society should be portrayed accurately and fairly regarding geographic setting. A particular geographic setting shouldn't be used repeatedly, and urban, suburban, and rural settings should be represented across items.
√	The item is free of content that might offend a geographic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a geographic subgroup.
G.	**Disability Bias**
Definition	All groups of society should be portrayed accurately and fairly regarding disability. Stereotypes related to any particular disability should be avoided. No undue restrictions should exist in the item that would interfere with the ability of a student to comprehend or respond to the item.
√	The item is free of content that might offend a disability subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a disability subgroup.
√	A graphic representation is used in the items, as appropriate. The complexity of the graphic is appropriate to the purpose—simpler is better.
√	The item avoids content that depends on sensory knowledge (such as references to movement, sound, smell, etc.) unless this is crucial to the overall item.
√	The item could be put into Braille.
√	The item avoids using both O and Q.
√	Letter pairs can be easily distinguished when read. (S and T are okay; S and X are not).
H.	**Art Supports Text**
Definition	The art is related to the item and supports the reader when possible. The item text and art are legible and accessible, and the art is appropriately placed in the item to support the reader. The art does not distract the test taker, but instead provides a scaffold to overall comprehension.
√	All pictures relate to items.
√	The item is free from pictorial clutter: All pictures are needed to answer the item.
√	Graphics are clear and non-fuzzy.
√	Any symbols used are highly distinguishable.
√	Visual load requirements are reasonable for the grade level.
√	Multi-dimensional graphics and complex shading are avoided.
√	Tables have replaced any cluttered graphs.
√	Labels read clockwise (as is easier for Braille readers).
I.	**Special Populations Considerations**
Definition	Consideration must be given for maximum accessibility to all students including, but not limited to, English language learners, limited sight, hearing impaired, cognitively challenged, etc. These considerations will assist all students.
√	The item contains scaffolding techniques to support student understanding of what is being asked in the item.
√	Text is replaced with graphic representations, when appropriate.
√	The item is written with simplified text load.
√	The item is written with simplified sentences.

Universal Design Item Checklist	
√	The item has as little extraneous information as possible.
√	The item provides context, but it is simplified.
√	The item uses smaller or less complicated numbers or expressions where not otherwise required.
√	The item avoids negative phrasing or questions; for example, questions are not asked in the negative.

Appendix E: Criteria for Item Acceptability

The following criteria represent best practices in item development, and were implemented during the creation and review of the New York State 3–8 test questions; however, these criteria are not a substitute for the full, detailed criteria documents, which are available online at the following links:

- http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-english-language-arts-tests; and
- http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-mathematics-tests.

For Multiple-Choice Items:

Check that the content of each item:

- is targeted to assess only one objective or skill (unless specifications indicate otherwise)
- deals with material that is important in testing the targeted performance indicator
- uses grade-appropriate content and thinking skills
- is presented at a reading level suitable for the grade level being tested
- has a stem that facilitates answering the question or completing the statement without looking at the answer choices
- has a stem that does **not** present clues to the correct answer choice
- has answer choices that are plausible and attractive to the student who has not mastered the objective or skill
- has mutually exclusive distractors
- has one and only one correct answer choice
- is free of cultural, racial, ethnic, age, gender, disability, regional, or other apparent bias

Check that the format of each item:

- is worded in the positive unless it is absolutely necessary to use the negative form
- is free of extraneous words or expressions in both the stem and the answer choices (e.g., the same word or phrase does not begin each answer choice)
- indicates emphasis on key words, such as best, first, least, not, and others that are important and might be overlooked
- places the interrogative word at the **beginning** of a stem in the form of a question, or places the omitted portion of an incomplete statement at the **end** of the statement
- indicates the correct answer choice
- provides the rationale for all distractors
- is conceptually, grammatically, and syntactically consistent–between the stem and answer choices, and among the answer choices
- has answer choices balanced in length, or contains two long and two short answer choices
- clearly identifies the passage or other stimulus material associated with the item
- clearly identifies a need of for art, if applicable, and the art is conceptualized and sketched, with important considerations explicated

Also check that:

- one item does not present clues to the correct answer choice for any other item
- any item based on a passage is answerable from the information given in the passage and is not dependent on skills related to other content areas
- any item based on a passage is truly passage-dependent; that is, **not** answerable without reference to the passage
- there is a balance of reasonable, non-stereotypical representation of economic classes, races, cultures, ages, genders, and persons with disabilities in context and art

For Constructed-Response Items:

Check that the content of each item is:

- designed to assess the targeted performance indicator
- appropriate for the grade level being tested
- presented at a reading level suitable for the grade level being tested
- appropriate in context
- written so that a student possessing knowledge or skill being tested can construct a response that can be scored with the specified rubric or scoring tool; that is, the range of possible correct responses must be wide enough to allow for a diversity of responses, but narrow enough so that students who do not clearly show their grasp of the objective or skill being assessed cannot obtain the maximum score
- presented without clues to the correct response
- checked for accuracy and documented against reliable, up-to-date sources (including rubrics)
- free of cultural, racial, ethnic, age, gender, disability, or other apparent bias

Check that the format of each item is:

- appropriate for the question being asked and the intended response
- worded clearly and concisely, using simple vocabulary and sentence structure
- precise and unambiguous in its directions for the desired response
- free of extraneous words or expressions
- worded in the positive form rather than in the negative form
- conceptually, grammatically, and syntactically consistent
- marked with emphasis on key words, such as best, first, least, and others that are important and might be overlooked
- clearly identified as needing art, if applicable, and the art is conceptualized and sketched, with important considerations explicated

Also check that:

- one item does not present clues to the correct response to any other item
- there is a balance of reasonable, non-stereotypical representation of economic classes, races, cultures, ages, genders, and persons with disabilities in context and art
- for each set of items related to a reading passage, each item is designed to elicit a unique and independent response
- items designed to assess reading do not depend on prior knowledge of the subject matter used in the prompt/question

Appendix F: Psychometric Guidelines for Operational Item Selection

It is primarily up to the content development department to select items for the 2018 Operational Test. The psychometrics department will provide support, as necessary, and will review the final item selection. The psychometrics department will provide data files with parameters for all FT items eligible for the item pool. The pools of items eligible for 2018 item selection included 2013, 2014, 2015, 2016, and 2017 embedded and stand-alone field-test items.

Here are the general guidelines for item selection:

- Satisfy the content specifications in terms of objective coverage and the number and percentage of MC and CR items on the test. An often-used criterion for objective coverage is within 5% of the percentages of score points and items per objective.
- To the extent possible, select both easy and difficult items to provide good measurement information at both ends of the performance scale.
- Avoid selecting items with too high/low *p*-values, items with flagged point biserials, and poorly fitting items.
- Minimize the number of items flagged for DIF (gender, ethnic, and High/Low Needs schools). Flagged items should be reviewed for content again. It needs to be remembered that some items may be flagged for DIF by chance only, and that their content may not necessarily be biased against any of the analyzed subgroups. The psychometrics department will provide DIF information for each item. It is also possible to get "significant" DIF, but not bias, if the content is a necessary part of the construct that is measured. That is, there may be some non-false positive DIF flags on items that do not exhibit bias.
- Provide the NYSED with the following summary information:
    - Overview of the statistical properties of the tests
    - Blueprint comparison between the test build and the target. The focus is on the total number of points on the test
    - Raw score proportion correct comparison between the test build and the reference (i.e., Spring 2017 test)
    - Vertical linked average difficulty parameter (MC items only) across all grades
    - Vertically linked TCC based on the constructed test
    - TCC, Test Information Curves and Conditional SEM Curves for each subject and grade, again using the Spring 2017 operational test as a reference.

Appendix G: Operational Item Maps

The following tables show the operational item maps for the 2018 NYSTP Grades 3–8 ELA and Mathematics Tests. Field test items that do not contribute to students' scores have been omitted. Additional detail on the standards to which these items align may be found at: http://www.engageny.org/resource/new-york-state-p-12-common-core-learning-standards.

Table G1. ELA Grade 3 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.3.4
2	MC	1	CCSS.ELA-Literacy.RI.3.3
3	MC	1	CCSS.ELA-Literacy.RI.3.4
4	MC	1	CCSS.ELA-Literacy.RI.3.7
5	MC	1	CCSS.ELA-Literacy.RI.3.5
6	MC	1	CCSS.ELA-Literacy.RI.3.2
13	MC	1	CCSS.ELA-Literacy.RL.3.4
14	MC	1	CCSS.ELA-Literacy.RL.3.3
15	MC	1	CCSS.ELA-Literacy.RL.3.3
16	MC	1	CCSS.ELA-Literacy.RL.3.3
17	MC	1	CCSS.ELA-Literacy.RL.3.6
18	MC	1	CCSS.ELA-Literacy.RL.3.2
19	MC	1	CCSS.ELA-Literacy.RL.3.4
20	MC	1	CCSS.ELA-Literacy.RL.3.6
21	MC	1	CCSS.ELA-Literacy.L.3.4
22	MC	1	CCSS.ELA-Literacy.RL.3.2
23	MC	1	CCSS.ELA-Literacy.RL.3.3
24	MC	1	CCSS.ELA-Literacy.RL.3.3
25	CR	2	CCSS.ELA-Literacy.RI.3.3
26	CR	2	CCSS.ELA-Literacy.RI.3.2
27	CR	2	CCSS.ELA-Literacy.RI.3.7
28	CR	2	CCSS.ELA-Literacy.RI.3.8
29	CR	2	CCSS.ELA-Literacy.RI.3.3
30	CR	2	CCSS.ELA-Literacy.RL.3.5
31	CR	4	CCSS.ELA-Literacy.RL.3.3

Table G2. ELA Grade 4 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.4.4
2	MC	1	CCSS.ELA-Literacy.L.4.4
3	MC	1	CCSS.ELA-Literacy.RL.4.6
4	MC	1	CCSS.ELA-Literacy.RL.4.3
5	MC	1	CCSS.ELA-Literacy.RL.4.3
6	MC	1	CCSS.ELA-Literacy.RL.4.2
7	MC	1	CCSS.ELA-Literacy.RI.4.4
8	MC	1	CCSS.ELA-Literacy.RI.4.3
9	MC	1	CCSS.ELA-Literacy.RI.4.3
10	MC	1	CCSS.ELA-Literacy.RI.4.2
11	MC	1	CCSS.ELA-Literacy.RI.4.5
12	MC	1	CCSS.ELA-Literacy.RI.4.7
19	MC	1	CCSS.ELA-Literacy.RL.4.2
20	MC	1	CCSS.ELA-Literacy.RL.4.4
21	MC	1	CCSS.ELA-Literacy.L.4.4
22	MC	1	CCSS.ELA-Literacy.RL.4.3
23	MC	1	CCSS.ELA-Literacy.RL.4.3
24	MC	1	CCSS.ELA-Literacy.RL.4.2
25	CR	2	CCSS.ELA-Literacy.RL.4.4
26	CR	2	CCSS.ELA-Literacy.RL.4.2
27	CR	2	CCSS.ELA-Literacy.RL.4.6
28	CR	2	CCSS.ELA-Literacy.RI.4.3
29	CR	2	CCSS.ELA-Literacy.RI.4.2
30	CR	2	CCSS.ELA-Literacy.RI.4.3
31	CR	4	CCSS.ELA-Literacy.RI.4.3

Table G3. ELA Grade 5 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.5.4
2	MC	1	CCSS.ELA-Literacy.L.5.4
3	MC	1	CCSS.ELA-Literacy.RL.5.3
4	MC	1	CCSS.ELA-Literacy.RL.5.3
5	MC	1	CCSS.ELA-Literacy.RL.5.3
6	MC	1	CCSS.ELA-Literacy.RL.5.4
7	MC	1	CCSS.ELA-Literacy.RL.5.2
15	MC	1	CCSS.ELA-Literacy.RI.5.5
16	MC	1	CCSS.ELA-Literacy.RI.5.3
17	MC	1	CCSS.ELA-Literacy.RI.5.3
18	MC	1	CCSS.ELA-Literacy.RI.5.4
19	MC	1	CCSS.ELA-Literacy.RI.5.2
20	MC	1	CCSS.ELA-Literacy.RI.5.2
21	MC	1	CCSS.ELA-Literacy.RI.5.6
22	MC	1	CCSS.ELA-Literacy.RL.5.4
23	MC	1	CCSS.ELA-Literacy.RL.5.5
24	MC	1	CCSS.ELA-Literacy.RL.5.3
25	MC	1	CCSS.ELA-Literacy.RL.5.2
26	MC	1	CCSS.ELA-Literacy.RL.5.3
27	MC	1	CCSS.ELA-Literacy.RL.5.3
28	MC	1	CCSS.ELA-Literacy.RL.5.6
29	MC	1	CCSS.ELA-Literacy.RI.5.2
30	MC	1	CCSS.ELA-Literacy.RI.5.4
31	MC	1	CCSS.ELA-Literacy.RI.5.7
32	MC	1	CCSS.ELA-Literacy.RI.5.3
33	MC	1	CCSS.ELA-Literacy.RI.5.3
34	MC	1	CCSS.ELA-Literacy.RI.5.3
35	MC	1	CCSS.ELA-Literacy.RI.5.2
36	CR	2	CCSS.ELA-Literacy.RL.5.3
37	CR	2	CCSS.ELA-Literacy.RL.5.6
38	CR	2	CCSS.ELA-Literacy.RL.5.2
39	CR	2	CCSS.ELA-Literacy.RI.5.8
40	CR	2	CCSS.ELA-Literacy.RI.5.7
41	CR	2	CCSS.ELA-Literacy.RI.5.5
42	CR	4	CCSS.ELA-Literacy.RI.5.3

Table G4. ELA Grade 6 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.6.3
2	MC	1	CCSS.ELA-Literacy.RL.6.5
3	MC	1	CCSS.ELA-Literacy.RL.6.4
4	MC	1	CCSS.ELA-Literacy.RL.6.3
5	MC	1	CCSS.ELA-Literacy.RL.6.6
6	MC	1	CCSS.ELA-Literacy.RL.6.2
7	MC	1	CCSS.ELA-Literacy.RL.6.2
15	MC	1	CCSS.ELA-Literacy.RI.6.3
16	MC	1	CCSS.ELA-Literacy.RI.6.8
17	MC	1	CCSS.ELA-Literacy.L.6.4
18	MC	1	CCSS.ELA-Literacy.RI.6.2
19	MC	1	CCSS.ELA-Literacy.RI.6.7
20	MC	1	CCSS.ELA-Literacy.RI.6.5
21	MC	1	CCSS.ELA-Literacy.RI.6.2
22	MC	1	CCSS.ELA-Literacy.RL.6.3
23	MC	1	CCSS.ELA-Literacy.RL.6.2
24	MC	1	CCSS.ELA-Literacy.RL.6.3
25	MC	1	CCSS.ELA-Literacy.RL.6.4
26	MC	1	CCSS.ELA-Literacy.RL.6.3
27	MC	1	CCSS.ELA-Literacy.RL.6.2
28	MC	1	CCSS.ELA-Literacy.RL.6.6
29	MC	1	CCSS.ELA-Literacy.RI.6.3
30	MC	1	CCSS.ELA-Literacy.RI.6.4
31	MC	1	CCSS.ELA-Literacy.RI.6.7
32	MC	1	CCSS.ELA-Literacy.RI.6.5
33	MC	1	CCSS.ELA-Literacy.RI.6.2
34	MC	1	CCSS.ELA-Literacy.RI.6.6
35	MC	1	CCSS.ELA-Literacy.RI.6.2
36	CR	2	CCSS.ELA-Literacy.RL.6.4
37	CR	2	CCSS.ELA-Literacy.RL.6.6
38	CR	2	CCSS.ELA-Literacy.RL.6.5
39	CR	2	CCSS.ELA-Literacy.RI.6.2
40	CR	2	CCSS.ELA-Literacy.RI.6.3
41	CR	2	CCSS.ELA-Literacy.RI.6.2
42	CR	4	CCSS.ELA-Literacy.RI.6.6

Table G5. ELA Grade 7 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.L.7.4
2	MC	1	CCSS.ELA-Literacy.RL.7.4
3	MC	1	CCSS.ELA-Literacy.RL.7.3
4	MC	1	CCSS.ELA-Literacy.RL.7.3
5	MC	1	CCSS.ELA-Literacy.RL.7.3
6	MC	1	CCSS.ELA-Literacy.RL.7.4
7	MC	1	CCSS.ELA-Literacy.RL.7.2
15	MC	1	CCSS.ELA-Literacy.RI.7.3
16	MC	1	CCSS.ELA-Literacy.RI.7.4
17	MC	1	CCSS.ELA-Literacy.RI.7.5
18	MC	1	CCSS.ELA-Literacy.RI.7.2
19	MC	1	CCSS.ELA-Literacy.RI.7.8
20	MC	1	CCSS.ELA-Literacy.RI.7.6
21	MC	1	CCSS.ELA-Literacy.RI.7.2
22	MC	1	CCSS.ELA-Literacy.RL.7.4
23	MC	1	CCSS.ELA-Literacy.RL.7.3
24	MC	1	CCSS.ELA-Literacy.RL.7.2
25	MC	1	CCSS.ELA-Literacy.RL.7.2
26	MC	1	CCSS.ELA-Literacy.RL.7.2
27	MC	1	CCSS.ELA-Literacy.RL.7.3
28	MC	1	CCSS.ELA-Literacy.RL.7.6
29	MC	1	CCSS.ELA-Literacy.RI.7.2
30	MC	1	CCSS.ELA-Literacy.RI.7.4
31	MC	1	CCSS.ELA-Literacy.RI.7.4
32	MC	1	CCSS.ELA-Literacy.RI.7.3
33	MC	1	CCSS.ELA-Literacy.RI.7.8
34	MC	1	CCSS.ELA-Literacy.RI.7.3
35	MC	1	CCSS.ELA-Literacy.RI.7.3
36	CR	2	CCSS.ELA-Literacy.RI.7.6
37	CR	2	CCSS.ELA-Literacy.RI.7.5
38	CR	2	CCSS.ELA-Literacy.RI.7.5
39	CR	2	CCSS.ELA-Literacy.RL.7.3
40	CR	2	CCSS.ELA-Literacy.RL.7.2
41	CR	2	CCSS.ELA-Literacy.RL.7.4
42	CR	2	CCSS.ELA-Literacy.RL.7.6
43	CR	4	CCSS.ELA-Literacy.RL.7.3

Table G6. ELA Grade 8 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.8.3
2	MC	1	CCSS.ELA-Literacy.RL.8.4
3	MC	1	CCSS.ELA-Literacy.RL.8.3
4	MC	1	CCSS.ELA-Literacy.RL.8.2
5	MC	1	CCSS.ELA-Literacy.RL.8.4
6	MC	1	CCSS.ELA-Literacy.L.8.4
7	MC	1	CCSS.ELA-Literacy.RL.8.6
8	MC	1	CCSS.ELA-Literacy.RI.8.4
9	MC	1	CCSS.ELA-Literacy.RI.8.6
10	MC	1	CCSS.ELA-Literacy.RI.8.2
11	MC	1	CCSS.ELA-Literacy.RI.8.3
12	MC	1	CCSS.ELA-Literacy.RI.8.5
13	MC	1	CCSS.ELA-Literacy.RI.8.2
14	MC	1	CCSS.ELA-Literacy.RI.8.3
15	MC	1	CCSS.ELA-Literacy.RL.8.3
16	MC	1	CCSS.ELA-Literacy.RL.8.4
17	MC	1	CCSS.ELA-Literacy.RL.8.2
18	MC	1	CCSS.ELA-Literacy.RL.8.3
19	MC	1	CCSS.ELA-Literacy.RL.8.3
20	MC	1	CCSS.ELA-Literacy.RL.8.2
21	MC	1	CCSS.ELA-Literacy.RL.8.6
29	MC	1	CCSS.ELA-Literacy.RI.8.3
30	MC	1	CCSS.ELA-Literacy.RI.8.4
31	MC	1	CCSS.ELA-Literacy.RI.8.5
32	MC	1	CCSS.ELA-Literacy.RI.8.3
33	MC	1	CCSS.ELA-Literacy.RI.8.2
34	MC	1	CCSS.ELA-Literacy.RI.8.6
35	MC	1	CCSS.ELA-Literacy.RI.8.2
36	CR	2	CCSS.ELA-Literacy.RI.8.5
37	CR	2	CCSS.ELA-Literacy.RI.8.5
38	CR	2	CCSS.ELA-Literacy.RI.8.6
39	CR	2	CCSS.ELA-Literacy.RL.8.4
40	CR	2	CCSS.ELA-Literacy.RL.8.2
41	CR	2	CCSS.ELA-Literacy.RL.8.2
42	CR	2	CCSS.ELA-Literacy.RL.8.3
43	CR	4	CCSS.ELA-Literacy.RL.8.6

Table G7. Mathematics Grade 3 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.3.OA.B.5
2	MC	1	CCSS.Math.Content.3.NBT.A.1
3	MC	1	CCSS.Math.Content.3.OA.A.4
4	MC	1	CCSS.Math.Content.3.NF.A.3b
6	MC	1	CCSS.Math.Content.3.OA.A.1
7	MC	1	CCSS.Math.Content.3.MD.C.5b
9	MC	1	CCSS.Math.Content.3.OA.D.8
11	MC	1	CCSS.Math.Content.3.NBT.A.3
12	MC	1	CCSS.Math.Content.3.OA.B.6
14	MC	1	CCSS.Math.Content.3.OA.A.2
15	MC	1	CCSS.Math.Content.3.G.A.2
17	MC	1	CCSS.Math.Content.3.MD.A.1
18	MC	1	CCSS.Math.Content.3.NF.A.2b
20	MC	1	CCSS.Math.Content.3.OA.A.3
21	MC	1	CCSS.Math.Content.3.OA.D.9
22	MC	1	CCSS.Math.Content.3.NF.A.2a
23	MC	1	CCSS.Math.Content.3.MD.A.2
24	MC	1	CCSS.Math.Content.3.OA.B.5
25	MC	1	CCSS.Math.Content.3.NF.A.3d
26	MC	1	CCSS.Math.Content.3.OA.A.4
27	MC	1	CCSS.Math.Content.3.NBT.A.3
28	MC	1	CCSS.Math.Content.3.NF.A.3a
29	MC	1	CCSS.Math.Content.3.OA.A.2
30	MC	1	CCSS.Math.Content.3.OA.D.9
31	MC	1	CCSS.Math.Content.3.MD.C.7d
32	MC	1	CCSS.Math.Content.3.G.A.2
33	MC	1	CCSS.Math.Content.3.NF.A.3c
34	CR	2	CCSS.Math.Content.3.MD.A.1
35	CR	2	CCSS.Math.Content.3.OA.A.1
36	CR	2	CCSS.Math.Content.3.NF.A.1
37	CR	2	CCSS.Math.Content.3.MD.B.3
38	CR	2	CCSS.Math.Content.3.OA.A.3
39	CR	2	CCSS.Math.Content.3.MD.C.7b
40	CR	3	CCSS.Math.Content.3.OA.D.8

Table G8. Mathematics Grade 4 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.4.OA.A.1
2	MC	1	CCSS.Math.Content.4.NF.B.3d
3	MC	1	CCSS.Math.Content.4.NBT.A.3
4	MC	1	CCSS.Math.Content.3.MD.B.4
6	MC	1	CCSS.Math.Content.4.NF.B.3a
7	MC	1	CCSS.Math.Content.4.MD.A.3
9	MC	1	CCSS.Math.Content.4.OA.A.3
10	MC	1	CCSS.Math.Content.4.NBT.B.5
12	MC	1	CCSS.Math.Content.4.MD.C.5a
13	MC	1	CCSS.Math.Content.4.NBT.A.1
14	MC	1	CCSS.Math.Content.4.OA.C.5
16	MC	1	CCSS.Math.Content.4.OA.A.2
17	MC	1	CCSS.Math.Content.4.MD.C.6
18	MC	1	CCSS.Math.Content.4.NF.B.3b
20	MC	1	CCSS.Math.Content.4.G.A.2
21	MC	1	CCSS.Math.Content.4.NF.A.1
23	MC	1	CCSS.Math.Content.4.NBT.B.6
24	MC	1	CCSS.Math.Content.4.NBT.A.2
25	MC	1	CCSS.Math.Content.4.NF.B.4a
27	MC	1	CCSS.Math.Content.3.MD.D.8
28	MC	1	CCSS.Math.Content.4.OA.A.3
29	MC	1	CCSS.Math.Content.4.MD.B.4
30	MC	1	CCSS.Math.Content.4.NF.A.2
31	MC	1	CCSS.Math.Content.4.G.A.3
32	MC	1	CCSS.Math.Content.4.OA.B.4
33	MC	1	CCSS.Math.Content.4.NBT.A.2
34	MC	1	CCSS.Math.Content.4.NF.A.2
35	MC	1	CCSS.Math.Content.4.OA.A.1
36	MC	1	CCSS.Math.Content.4.MD.C.5b
37	MC	1	CCSS.Math.Content.4.NBT.A.1
38	MC	1	CCSS.Math.Content.4.NF.B.4b
39	CR	2	CCSS.Math.Content.4.G.A.1
40	CR	2	CCSS.Math.Content.4.NBT.B.6
41	CR	2	CCSS.Math.Content.4.NF.B.4c
42	CR	2	CCSS.Math.Content.4.OA.A.2
43	CR	2	CCSS.Math.Content.4.NF.B.3d
44	CR	2	CCSS.Math.Content.4.MD.C.7

Item	Type	Points	Standard
45	CR	3	CCSS.Math.Content.4.NBT.B.5

Table G9. Mathematics Grade 5 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.5.NBT.B.6
2	MC	1	CCSS.Math.Content.4.NF.C.5
3	MC	1	CCSS.Math.Content.5.MD.C.5a
4	MC	1	CCSS.Math.Content.5.NF.B.3
6	MC	1	CCSS.Math.Content.5.NBT.B.7
7	MC	1	CCSS.Math.Content.4.MD.A.2
8	MC	1	CCSS.Math.Content.5.OA.A.1
10	MC	1	CCSS.Math.Content.5.NF.A.1
11	MC	1	CCSS.Math.Content.5.NF.B.4
13	MC	1	CCSS.Math.Content.5.G.B.4
14	MC	1	CCSS.Math.Content.5.NF.B.5
16	MC	1	CCSS.Math.Content.5.NBT.B.6
17	MC	1	CCSS.Math.Content.5.MD.C.3
18	MC	1	CCSS.Math.Content.5.NBT.A.3a
20	MC	1	CCSS.Math.Content.5.MD.C.5c
21	MC	1	CCSS.Math.Content.5.NF.B.7
22	MC	1	CCSS.Math.Content.5.NBT.A.2
24	MC	1	CCSS.Math.Content.5.MD.A.1
25	MC	1	CCSS.Math.Content.5.OA.A.2
27	MC	1	CCSS.Math.Content.5.NBT.A.3b
28	MC	1	CCSS.Math.Content.5.MD.C.3b
29	MC	1	CCSS.Math.Content.5.NF.B.7a
30	MC	1	CCSS.Math.Content.5.MD.C.5a
31	MC	1	CCSS.Math.Content.5.NF.B.7c
32	MC	1	CCSS.Math.Content.5.NF.A.1
33	MC	1	CCSS.Math.Content.4.NF.C.6
34	MC	1	CCSS.Math.Content.5.NF.B.6
35	MC	1	CCSS.Math.Content.5.MD.C.5b
36	MC	1	CCSS.Math.Content.5.MD.A.1
37	MC	1	CCSS.Math.Content.5.G.B.4
38	MC	1	CCSS.Math.Content.5.MD.C.4
39	CR	2	CCSS.Math.Content.5.MD.A.1
40	CR	2	CCSS.Math.Content.5.NBT.A.1
41	CR	2	CCSS.Math.Content.5.NF.A.2

Item	Type	Points	Standard
42	CR	2	CCSS.Math.Content.5.NF.B.4b
43	CR	2	CCSS.Math.Content.5.MD.B.2
44	CR	2	CCSS.Math.Content.5.NF.B.6
45	CR	3	CCSS.Math.Content.5.NBT.B.7

Table G10. Mathematics Grade 6 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.6.EE.B.5
2	MC	1	CCSS.Math.Content.6.RP.A.3c
3	MC	1	CCSS.Math.Content.6.NS.B.4
4	MC	1	CCSS.Math.Content.6.EE.C.9
6	MC	1	CCSS.Math.Content.6.EE.A.2c
7	MC	1	CCSS.Math.Content.6.RP.A.1
9	MC	1	CCSS.Math.Content.6.G.A.3
10	MC	1	CCSS.Math.Content.6.EE.B.7
12	MC	1	CCSS.Math.Content.6.NS.C.6c
13	MC	1	CCSS.Math.Content.6.G.A.2
15	MC	1	CCSS.Math.Content.6.RP.A.3c
16	MC	1	CCSS.Math.Content.6.EE.C.9
18	MC	1	CCSS.Math.Content.6.EE.A.1
19	MC	1	CCSS.Math.Content.6.RP.A.3b
21	MC	1	CCSS.Math.Content.6.EE.B.5
22	MC	1	CCSS.Math.Content.6.G.A.1
24	MC	1	CCSS.Math.Content.6.NS.C.6
25	MC	1	CCSS.Math.Content.6.RP.A.3b
26	MC	1	CCSS.Math.Content.6.EE.A.3
27	MC	1	CCSS.Math.Content.5.G.A.2
28	MC	1	CCSS.Math.Content.6.G.A.4
29	MC	1	CCSS.Math.Content.6.EE.A.4
30	MC	1	CCSS.Math.Content.6.RP.A.3a
31	MC	1	CCSS.Math.Content.6.EE.B.6
32	MC	1	CCSS.Math.Content.6.NS.C.5
33	MC	1	CCSS.Math.Content.6.RP.A.1
34	MC	1	CCSS.Math.Content.6.EE.B.8
35	MC	1	CCSS.Math.Content.6.G.A.3
36	MC	1	CCSS.Math.Content.6.NS.A.1
37	MC	1	CCSS.Math.Content.6.EE.A.3
38	MC	1	CCSS.Math.Content.6.RP.A.2

Item	Type	Points	Standard
39	CR	2	CCSS.Math.Content.6.NS.A.1
40	CR	2	CCSS.Math.Content.6.EE.A.2a
41	CR	2	CCSS.Math.Content.6.RP.A.3d
42	CR	2	CCSS.Math.Content.6.EE.A.1
43	CR	2	CCSS.Math.Content.6.NS.C.6b
44	CR	2	CCSS.Math.Content.6.RP.A.2
45	CR	2	CCSS.Math.Content.6.G.A.2
46	CR	3	CCSS.Math.Content.6.EE.B.7

Table G11. Mathematics Grade 7 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.7.NS.A.2d
2	MC	1	CCSS.Math.Content.7.G.B.4
3	MC	1	CCSS.Math.Content.7.EE.B.4a
4	MC	1	CCSS.Math.Content.7.RP.A.1
6	MC	1	CCSS.Math.Content.7.SP.C.7b
7	MC	1	CCSS.Math.Content.7.NS.A.3
8	MC	1	CCSS.Math.Content.7.SP.A.2
10	MC	1	CCSS.Math.Content.7.EE.A.1
11	MC	1	CCSS.Math.Content.7.RP.A.2a
13	MC	1	CCSS.Math.Content.7.EE.B.3
14	MC	1	CCSS.Math.Content.7.SP.A.1
16	MC	1	CCSS.Math.Content.7.RP.A.3
17	MC	1	CCSS.Math.Content.7.NS.A.1a
19	MC	1	CCSS.Math.Content.7.EE.B.4b
20	MC	1	CCSS.Math.Content.7.RP.A.3
22	MC	1	CCSS.Math.Content.7.NS.A.3
23	MC	1	CCSS.Math.Content.7.EE.B.3
24	MC	1	CCSS.Math.Content.7.RP.A.2b
26	MC	1	CCSS.Math.Content.7.RP.A.3
27	MC	1	CCSS.Math.Content.7.NS.A.1d
28	MC	1	CCSS.Math.Content.7.SP.B.3
29	MC	1	CCSS.Math.Content.7.NS.A.1c
30	MC	1	CCSS.Math.Content.7.SP.C.6
31	MC	1	CCSS.Math.Content.6.SP.B.4
32	MC	1	CCSS.Math.Content.7.RP.A.2d
33	MC	1	CCSS.Math.Content.7.SP.C.7b
34	MC	1	CCSS.Math.Content.7.SP.C.5

Item	Type	Points	Standard
35	MC	1	CCSS.Math.Content.7.EE.A.1
36	MC	1	CCSS.Math.Content.7.RP.A.1
37	MC	1	CCSS.Math.Content.7.EE.A.1
38	MC	1	CCSS.Math.Content.7.EE.B.4b
39	MC	1	CCSS.Math.Content.7.EE.A.2
40	MC	1	CCSS.Math.Content.7.G.A.1
41	CR	2	CCSS.Math.Content.7.G.A.1
42	CR	2	CCSS.Math.Content.7.NS.A.3
43	CR	2	CCSS.Math.Content.7.EE.B.4b
44	CR	2	CCSS.Math.Content.7.EE.B.3
45	CR	2	CCSS.Math.Content.7.NS.A.3
46	CR	2	CCSS.Math.Content.7.EE.B.4a
47	CR	2	CCSS.Math.Content.7.RP.A.3
48	CR	3	CCSS.Math.Content.7.RP.A.2b

Table G12. Mathematics Grade 8 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.8.F.A.1
2	MC	1	CCSS.Math.Content.8.EE.A.3
3	MC	1	CCSS.Math.Content.8.F.A.3
4	MC	1	CCSS.Math.Content.8.EE.C.8b
6	MC	1	CCSS.Math.Content.8.EE.B.5
7	MC	1	CCSS.Math.Content.7.G.B.6
8	MC	1	CCSS.Math.Content.8.SP.A.2
10	MC	1	CCSS.Math.Content.8.EE.C.7
11	MC	1	CCSS.Math.Content.8.G.A.2
12	MC	1	CCSS.Math.Content.8.EE.A.4
14	MC	1	CCSS.Math.Content.8.F.A.3
15	MC	1	CCSS.Math.Content.7.G.A.2
16	MC	1	CCSS.Math.Content.8.SP.A.3
18	MC	1	CCSS.Math.Content.8.EE.A.1
19	MC	1	CCSS.Math.Content.8.F.A.2
20	MC	1	CCSS.Math.Content.8.G.A.3
22	MC	1	CCSS.Math.Content.8.F.B.4
23	MC	1	CCSS.Math.Content.8.G.A.4
24	MC	1	CCSS.Math.Content.8.EE.C.8c
26	MC	1	CCSS.Math.Content.8.G.C.9
27	MC	1	CCSS.Math.Content.8.EE.A.1

Item	Type	Points	Standard
28	MC	1	CCSS.Math.Content.8.F.B.4
30	MC	1	CCSS.Math.Content.8.F.A.2
31	MC	1	CCSS.Math.Content.8.SP.A.1
32	MC	1	CCSS.Math.Content.8.SP.A.3
33	MC	1	CCSS.Math.Content.8.G.C.9
34	MC	1	CCSS.Math.Content.8.EE.B.5
35	MC	1	CCSS.Math.Content.8.EE.B.6
36	MC	1	CCSS.Math.Content.8.G.A.5
37	MC	1	CCSS.Math.Content.7.G.A.3
38	MC	1	CCSS.Math.Content.8.SP.A.4
39	MC	1	CCSS.Math.Content.8.EE.A.4
40	MC	1	CCSS.Math.Content.8.F.B.5
41	CR	2	CCSS.Math.Content.8.EE.C.7a
42	CR	2	CCSS.Math.Content.8.F.B.4
43	CR	2	CCSS.Math.Content.8.G.A.3
44	CR	2	CCSS.Math.Content.8.EE.B.5
45	CR	2	CCSS.Math.Content.8.F.A.2
46	CR	2	CCSS.Math.Content.8.EE.A.4
47	CR	2	CCSS.Math.Content.8.F.A.3
48	CR	3	CCSS.Math.Content.8.EE.C.8c

Appendix H: ELA Short-Response Rubric

2-Point Rubric–Short Response

Score	Response Features
2 Point	The features of a 2-point response are: • Valid inferences and/or claims from the text where required by the prompt • Evidence of analysis of the text where required by the prompt • Relevant facts, definitions, concrete details, and/or other information from the text to develop response according to the requirements of the prompt • Sufficient number of facts, definitions, concrete details, and/or other information from the text as required by the prompt • Complete sentences where errors do not affect readability
1 Point	The features of a 1-point response are: • A mostly literal recounting of events or details from the text as required by the prompt • Some relevant facts, definitions, concrete details, and/or other information from the text to develop response according to the requirements of the prompt • Incomplete sentences or bullets
0 Point*	The features of a 0-point response are: • A response that does not address any of the requirements of the prompt or is totally inaccurate • A response that is not written in English • A response that is unintelligible or indecipherable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 1.

Appendix I: ELA Extended-Response Rubric

New York State Grade 3 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 Essays at this level:	3 Essays at this level:	2 Essays at this level:	1 Essays at this level:	0* Essays at this level:
CONTENT AND ANALYSIS: the extent to which the essay conveys ideas and information clearly and accurately in order to support analysis of topics or text	W.2, R.1–9	–clearly introduce a topic in a manner that follows logically from the task and purpose –demonstrate comprehension and analysis of the text	–clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate comprehension of the text	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a confused comprehension of the text	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text	–demonstrate a lack of comprehension of the text or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided text to support analysis and reflection	W.2 R.1–8	–develop the topic with relevant, well-chosen facts, definitions, and details throughout the essay	–develop the topic with relevant facts, definitions, and details throughout the essay	–partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant	–demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	–provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2 L.3 L.6	–clearly and consistently group related information together –skillfully connect ideas within categories of information using linking words and phrases – provide a concluding statement that follows clearly from the topic and information presented	–generally group related information together –connect ideas within categories of information using linking words and phrases –provide a concluding statement that follows from the topic and information presented	–exhibit some attempt to group related information together –inconsistently connect ideas using some linking words and phrases –provide a concluding statement that follows generally from the topic and information presented	–exhibit little attempt at organization –lack the use of linking words and phrases –provide a concluding statement that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –do not provide a concluding statement
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2 L.1 L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

New York State Grade 4-5 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 **Essays at this level:**	**3** **Essays at this level:**	**2** **Essays at this level:**	**1** **Essays at this level**	**0*** **Essays at this level:**
CONTENT AND ANALYSIS: the extent to which the essay conveys ideas and information clearly and accurately in order to support an analysis of topics or texts	W.2 R.1–9	– clearly introduce a topic in a manner that follows logically from the task and purpose –demonstrate insightful comprehension and analysis of the text(s)	– clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate comprehension and analysis of the text(s)	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a literal comprehension of the text(s)	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text(s)	–demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided texts to support analysis and reflection	W.2 W.9 R.1–9	–develop the topic with relevant, well-chosen facts, definitions, concrete details, quotations, or other information and examples from the text(s) –sustain the use of varied, relevant evidence	–develop the topic with relevant facts, definitions, details, quotations, or other information and examples from the text(s) –sustain the use of relevant evidence, with some lack of variety	–partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant –use relevant evidence with inconsistency	–demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	–provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2 L.3 L.6	–exhibit clear, purposeful organization –skillfully link ideas using grade-appropriate words and phrases –use grade-appropriate, stylistically sophisticated language and domain-specific vocabulary –provide a concluding statement that follows clearly from the topic and information presented	–exhibit clear organization –link ideas using grade-appropriate words and phrases –use grade-appropriate precise language and domain-specific vocabulary –provide a concluding statement that follows from the topic and information presented	–exhibit some attempt at organization –inconsistently link ideas using words and phrases –inconsistently use appropriate language and domain-specific vocabulary –provide a concluding statement that follows generally from the topic and information presented	–exhibit little attempt at organization, or attempts to organize are irrelevant to the task –lack the use of linking words and phrases –use language that is imprecise or inappropriate for the text(s) and task –provide a concluding statement that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –exhibit no use of linking words and phrases –use language that is predominantly incoherent or copied directly from the text(s) –do not provide a concluding statement
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2 L.1 L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 2.
- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

New York State Grade 6-8 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 **Essays at this level:**	**3** **Essays at this level:**	**2** **Essays at this level:**	**1** **Essays at this level:**	**0*** **Essays at this level:**
CONTENT AND ANALYSIS: the extent to which the essay conveys complex ideas and information clearly and accurately in order to support claims in an analysis of topics or texts	W.2, R.1–9	–clearly introduce a topic in a manner that is compelling and follows logically from the task and purpose –demonstrate insightful analysis of the text(s)	– clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate analysis of the text(s)	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a literal comprehension of the text(s)	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text(s)	–demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided texts to support analysis and reflection	W.9, R.1–9	–develop the topic with relevant, well-chosen facts, definitions, concrete details, quotations, or other information and examples from the text(s) –sustain the use of varied, relevant evidence	–develop the topic with relevant facts, definitions, details, quotations, or other information and examples from the text(s) –sustain the use of relevant evidence, with some lack of variety	–partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant –use relevant evidence with inconsistency	–demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	–provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2, L.3, L.6	–exhibit clear organization, with the skillful use of appropriate and varied transitions to create a unified whole and enhance meaning –establish and maintain a formal style, using grade-appropriate, stylistically sophisticated language and domain-specific vocabulary with a notable sense of voice –provide a concluding statement or section that is compelling and follows clearly from the topic and information presented	–exhibit clear organization, with the use of appropriate transitions to create a unified whole –establish and maintain a formal style using precise language and domain-specific vocabulary –provide a concluding statement or section that follows from the topic and information presented	–exhibit some attempt at organization, with inconsistent use of transitions –establish but fail to maintain a formal style, with inconsistent use of language and domain-specific vocabulary –provide a concluding statement or section that follows generally from the topic and information presented	–exhibit little attempt at organization, or attempts to organize are irrelevant to the task –lack a formal style, using language that is imprecise or inappropriate for the text(s) and task –provide a concluding statement or section that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –use language that is predominantly incoherent or copied directly from the text(s) –do not provide a concluding statement or section
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2, L.1, L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 2.
- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

Appendix J: Mathematics Short-Response Rubric

2-Point Holistic Rubric

2 Points	A two-point response includes the correct solution to the question and demonstrates a thorough understanding of the mathematical concepts and/or procedures in the task. This response: • indicates that the student has completed the task correctly, using mathematically sound procedures • contains sufficient work to demonstrate a thorough understanding of the mathematical concepts and/or procedures • may contain inconsequential errors that do not detract from the correct solution and the demonstration of a thorough understanding
1 Point	A one-point response demonstrates only a partial understanding of the mathematical concepts and/or procedures in the task. This response: • correctly addresses only some elements of the task • may contain an incorrect solution but applies a mathematically appropriate process • may contain the correct solution but required work is incomplete
0 Points*	A zero-point response is incorrect, irrelevant, incoherent, or contains a correct solution obtained using an obviously incorrect procedure. Although some elements may contain correct mathematical procedures, holistically they are not sufficient to demonstrate even a limited understanding of the mathematical concepts embodied in the task.

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

Appendix K: Mathematics Extended-Response Rubric

3-Point Holistic Rubric

3 Points	A three-point response includes the correct solution(s) to the question and demonstrates a thorough understanding of the mathematical concepts and/or procedures in the task. This response: • indicates that the student has completed the task correctly, using mathematically sound procedures • contains sufficient work to demonstrate a thorough understanding of the mathematical concepts and/or procedures • may contain inconsequential errors that do not detract from the correct solution(s) and the demonstration of a thorough understanding
2 Points	A two-point response demonstrates a partial understanding of the mathematical concepts and/or procedures in the task. This response: • appropriately addresses most, but not all, aspects of the task using mathematically sound procedures • may contain an incorrect solution but provides sound procedures, reasoning, and/or explanations • may reflect some minor misunderstanding of the underlying mathematical concepts and/or procedures
1 Point	A one-point response demonstrates only a limited understanding of the mathematical concepts and/or procedures in the task. This response: • may address some elements of the task correctly but reaches an inadequate solution and/or provides reasoning that is faulty or incomplete • exhibits multiple flaws related to misunderstanding of important aspects of the task, misuse of mathematical procedures, or faulty mathematical reasoning • reflects a lack of essential understanding of the underlying mathematical concepts • may contain the correct solution(s) but required work is limited
0 Points*	A zero-point response is incorrect, irrelevant, incoherent, or contains a correct solution obtained using an obviously incorrect procedure. Although some elements may contain correct mathematical procedures, holistically they are not sufficient to demonstrate even a limited understanding of the mathematical concepts embodied in the task.

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

Appendix L: Factor Analysis Results for Select Subgroups

As described in Section 3: Validity, a principal components factor analysis was conducted on the Grades 3–8 ELA and Mathematics Tests data. The analyses were conducted for the total population of students and select subgroups: ELL/MLL, SWD, SUA, SWD/SUA students using disability accommodations, and ELL/MLL students using ELL-related accommodations (ELL & SUA). Tables L1 and L2 contain the results of factor analysis on the subpopulation data for the Grades 3–8 ELA and Mathematics Tests, respectively.

Table L1. ELA Grade 3 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**4.71**	**18.84**	**18.84**
		1.52	6.06	24.90
		1.10	4.39	29.29
		1.02	4.08	33.37
		1.00	4.01	37.38
SWD	All Codes	**5.30**	**21.19**	**21.19**
		1.65	6.59	27.78
		1.05	4.22	31.99
SUA	All Codes	**5.00**	**19.99**	**19.99**
		1.69	6.78	26.77
		1.06	4.25	31.02
SWD/SUA	SUA=504 plan codes	**4.94**	**19.77**	**19.77**
		1.71	6.85	26.61
		1.07	4.28	30.89
ELL/MLL/ SUA	SUA & ELL Codes	**4.25**	**17.02**	**17.02**
		1.55	6.22	23.23
		1.14	4.55	27.79
		1.09	4.35	32.13
		1.04	4.18	36.31
		1.01	4.02	40.33

Table L2. ELA Grade 4 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**4.58**	**18.34**	**18.34**
		1.41	5.65	23.99
		1.07	4.29	28.27
		1.04	4.17	32.44
		1.03	4.12	36.57
SWD	All Codes	**5.41**	**21.66**	**21.66**
		1.42	5.66	27.32
		1.06	4.25	31.57
		1.01	4.03	35.60
SUA	All Codes	**5.21**	**20.83**	**20.83**
		1.44	5.74	26.57
		1.07	4.27	30.85
		1.01	4.04	34.88
SWD/SUA	SUA=504 plan codes	**5.10**	**20.40**	**20.40**
		1.45	5.79	26.19
		1.08	4.31	30.50
		1.01	4.06	34.56
		1.00	4.00	38.56
ELL/MLL/ SUA	SUA & ELL Codes	**4.10**	**16.42**	**16.42**
		1.45	5.79	22.20
		1.13	4.52	26.73
		1.08	4.31	31.03
		1.05	4.19	35.22
		1.03	4.13	39.34
		1.02	4.09	43.43

Table L3. ELA Grade 5 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**4.94**	**14.10**	**14.10**
		1.62	4.63	18.73
		1.18	3.38	22.11
		1.10	3.13	25.24
		1.09	3.11	28.35
		1.06	3.02	31.37
		1.04	2.97	34.34
		1.03	2.94	37.29
		1.00	2.87	40.16
SWD	All Codes	**5.83**	**16.67**	**16.67**
		1.70	4.85	21.52
		1.18	3.37	24.89
		1.12	3.20	28.10
		1.05	2.99	31.09
		1.01	2.88	33.97
SUA	All Codes	**5.78**	**16.51**	**16.51**
		1.73	4.93	21.44
		1.19	3.39	24.83
		1.11	3.17	28.00
		1.04	2.97	30.97
		1.00	2.86	33.84
SWD/SUA	SUA=504 plan codes	**5.57**	**15.92**	**15.92**
		1.74	4.96	20.87
		1.20	3.43	24.31
		1.11	3.17	27.48
		1.05	3.01	30.49
		1.01	2.88	33.37
		1.00	2.86	36.23
ELL/MLL/ SUA	SUA & ELL Codes	**4.26**	**12.16**	**12.16**
		1.63	4.66	16.82
		1.24	3.55	20.37
		1.13	3.22	23.59
		1.13	3.22	26.81
		1.11	3.17	29.98
		1.08	3.09	33.07

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL/ SUA	SUA & ELL Codes	1.04	2.98	36.05
		1.03	2.95	39.00
		1.03	2.93	41.93
		1.01	2.87	44.80

Table L4. ELA Grade 6 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**5.33**	**15.24**	**15.24**
		1.59	4.54	19.78
		1.15	3.30	23.08
		1.10	3.16	26.23
		1.06	3.03	29.27
		1.04	2.98	32.25
		1.02	2.93	35.18
		1.01	2.90	38.08
		1.00	2.87	40.94
SWD	All Codes	**6.21**	**17.76**	**17.76**
		1.61	4.60	22.36
		1.13	3.22	25.58
		1.08	3.09	28.67
		1.04	2.96	31.63
		1.02	2.91	34.54
SUA	All Codes	**6.20**	**17.72**	**17.72**
		1.62	4.63	22.35
		1.13	3.22	25.57
		1.08	3.07	28.64
		1.04	2.96	31.60
		1.02	2.90	34.51
SWD/SUA	SUA=504 plan codes	**5.97**	**17.06**	**17.06**
		1.64	4.68	21.74
		1.13	3.24	24.98
		1.08	3.09	28.08
		1.05	2.99	31.07
		1.03	2.93	34.00

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SWD/SUA	SUA=504 plan codes	1.01	2.88	36.88
		1.00	2.86	39.74
ELL/MLL/ SUA	SUA & ELL Codes	**4.62**	**13.21**	**13.21**
		1.64	4.70	17.91
		1.17	3.35	21.25
		1.13	3.24	24.49
		1.10	3.15	27.64
		1.09	3.11	30.75
		1.06	3.04	33.78
		1.05	2.99	36.78
		1.03	2.94	39.72
		1.03	2.93	42.65
		1.01	2.88	45.53

Table L5. ELA Grade 7 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**5.60**	**15.56**	**15.56**
		1.58	4.39	19.95
		1.20	3.34	23.28
		1.13	3.13	26.41
		1.09	3.03	29.44
		1.06	2.94	32.38
		1.04	2.90	35.28
		1.02	2.84	38.12
		1.01	2.81	40.93
SWD	All Codes	**6.32**	**17.56**	**17.56**
		1.79	4.98	22.53
		1.18	3.27	25.80
		1.10	3.06	28.86
		1.04	2.88	31.73
		1.03	2.87	34.60
SUA	All Codes	**6.38**	**17.73**	**17.73**
		1.79	4.97	22.70
		1.19	3.29	26.00

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SUA	All Codes	1.10	3.06	29.06
		1.03	2.87	31.93
		1.02	2.84	34.77
SWD/SUA	SUA=504 plan codes	**6.14**	**17.06**	**17.06**
		1.79	4.96	22.02
		1.19	3.30	25.32
		1.11	3.08	28.40
		1.04	2.89	31.30
		1.04	2.88	34.18
		1.01	2.80	36.98
ELL/MLL/ SUA	SUA & ELL Codes	**4.75**	**13.18**	**13.18**
		1.51	4.19	17.37
		1.25	3.46	20.83
		1.19	3.31	24.14
		1.15	3.20	27.35
		1.11	3.09	30.44
		1.10	3.05	33.49
		1.07	2.99	36.48
		1.05	2.91	39.39
		1.04	2.88	42.26
		1.02	2.83	45.10
		1.00	2.79	47.88

Table L6. ELA Grade 8 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**5.88**	**16.32**	**16.32**
		1.63	4.54	20.86
		1.21	3.37	24.24
		1.10	3.05	27.29
		1.07	2.97	30.26
		1.04	2.89	33.15
		1.02	2.82	35.98
		1.01	2.79	38.77

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SWD	All Codes	**6.38**	**17.71**	**17.71**
		1.66	4.60	22.31
		1.28	3.55	25.86
		1.04	2.88	28.75
		1.03	2.85	31.60
		1.00	2.78	34.38
SUA	All Codes	**6.49**	**18.04**	**18.04**
		1.65	4.57	22.61
		1.29	3.59	26.20
		1.03	2.85	29.05
		1.02	2.83	31.88
		1.00	2.78	34.66
SWD/SUA	SUA=504 plan codes	**6.21**	**17.26**	**17.26**
		1.65	4.59	21.85
		1.26	3.51	25.36
		1.05	2.92	28.28
		1.03	2.86	31.14
		1.01	2.80	33.94
ELL/MLL/ SUA	SUA & ELL Codes	**4.91**	**13.64**	**13.64**
		1.62	4.49	18.13
		1.24	3.45	21.58
		1.16	3.24	24.82
		1.13	3.13	27.94
		1.09	3.03	30.97
		1.07	2.98	33.95
		1.07	2.97	36.92
		1.04	2.88	39.80
		1.03	2.86	42.66
		1.01	2.80	45.45

Table L7. Mathematics Grade 3 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**8.34**	**24.52**	**24.52**
		1.61	4.74	29.26
		1.09	3.22	32.48
SWD	All Codes	**8.99**	**26.45**	**26.45**
		1.53	4.51	30.96
		1.06	3.13	34.09
		1.01	2.96	37.05
SUA	All Codes	**8.38**	**24.65**	**24.65**
		1.50	4.41	29.05
		1.08	3.17	32.23
		1.02	3.00	35.22
SWD/SUA	SUA=504 plan codes	**8.28**	**24.34**	**24.34**
		1.49	4.40	28.73
		1.09	3.20	31.93
		1.03	3.02	34.95
ELL/MLL/ SUA	SUA & ELL Codes	**7.83**	**23.04**	**23.04**
		1.54	4.53	27.56
		1.11	3.27	30.84
		1.05	3.09	33.93
		1.01	2.96	36.88
		1.00	2.95	39.83

Table L8. Mathematics Grade 4 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**8.45**	**22.24**	**22.24**
		1.40	3.69	25.93
		1.11	2.91	28.84
		1.04	2.74	31.58
SWD	All Codes	**9.33**	**24.56**	**24.56**
		1.39	3.65	28.21
		1.06	2.80	31.01
		1.00	2.64	33.65

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SUA	All Codes	**8.74**	**23.01**	**23.01**
		1.37	3.60	26.62
		1.07	2.83	29.44
		1.03	2.70	32.14
SWD/SUA	SUA=504 plan codes	**8.51**	**22.39**	**22.39**
		1.37	3.61	26.00
		1.09	2.86	28.87
		1.04	2.73	31.60
ELL/MLL/ SUA	SUA & ELL Codes	**7.40**	**19.48**	**19.48**
		1.38	3.63	23.11
		1.18	3.10	26.20
		1.08	2.85	29.06
		1.03	2.72	31.78
		1.02	2.68	34.46
		1.01	2.67	37.13

Table L9. Mathematics Grade 5 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**8.06**	**21.21**	**21.21**
		1.89	4.96	26.17
		1.02	2.67	28.84
SWD	All Codes	**8.70**	**22.90**	**22.90**
		1.71	4.50	27.40
		1.02	2.68	30.09
SUA	All Codes	**8.35**	**21.97**	**21.97**
		1.71	4.49	26.46
		1.03	2.71	29.17
SWD/SUA	SUA=504 plan codes	**7.93**	**20.86**	**20.86**
		1.71	4.50	25.36
		1.04	2.75	28.11
ELL/SUA	SUA & ELL Codes	**6.43**	**16.91**	**16.91**
		1.84	4.85	21.76
		1.13	2.97	24.73
		1.10	2.91	27.64

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL/ SUA	SUA & ELL Codes	1.04	2.73	30.37
		1.03	2.72	33.09
		1.01	2.66	35.75

Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**7.79**	**19.98**	**19.98**
		1.61	4.14	24.12
		1.10	2.81	26.93
		1.07	2.73	29.67
		1.01	2.59	32.26
SWD	All Codes	**8.20**	**21.03**	**21.03**
		1.55	3.98	25.01
		1.11	2.84	27.85
		1.04	2.65	30.50
SUA	All Codes	**8.15**	**20.90**	**20.90**
		1.53	3.93	24.83
		1.12	2.87	27.70
		1.04	2.67	30.37
SWD/SUA	SUA=504 plan codes	**7.43**	**19.06**	**19.06**
		1.54	3.95	23.01
		1.14	2.92	25.93
		1.06	2.71	28.64
		1.00	2.57	31.21
ELL/MLL/ SUA	SUA & ELL Codes	**5.80**	**14.88**	**14.88**
		1.55	3.98	18.86
		1.23	3.15	22.01
		1.17	3.00	25.01
		1.06	2.71	27.72
		1.05	2.69	30.41
		1.04	2.67	33.09
		1.02	2.60	35.69
		1.01	2.60	38.28
		1.01	2.58	40.86

Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**8.08**	**19.71**	**19.71**
		1.34	3.27	22.98
		1.23	2.99	25.97
		1.03	2.52	28.50
		1.02	2.49	30.99
		1.00	2.45	33.44
SWD	All Codes	**8.08**	**19.71**	**19.71**
		1.34	3.27	22.97
		1.19	2.91	25.89
		1.03	2.51	28.40
SUA	All Codes	**8.22**	**20.05**	**20.05**
		1.33	3.24	23.30
		1.19	2.90	26.20
		1.03	2.52	28.71
SWD/SUA	SUA=504 plan codes	**7.49**	**18.27**	**18.27**
		1.34	3.26	21.54
		1.21	2.94	24.48
		1.04	2.54	27.02
		1.01	2.47	29.49
		1.01	2.45	31.94
ELL/MLL/ SUA	SUA & ELL Codes	**5.17**	**12.62**	**12.62**
		1.39	3.40	16.02
		1.25	3.06	19.08
		1.18	2.87	21.95
		1.14	2.78	24.73
		1.13	2.76	27.49
		1.09	2.67	30.16
		1.08	2.64	32.80
		1.05	2.57	35.37
		1.05	2.56	37.94
		1.02	2.50	40.43
		1.00	2.45	42.88
		1.00	2.44	45.32

Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL	ELL=Y	**8.19**	**19.98**	**19.98**
		1.23	3.01	22.99
		1.13	2.77	25.76
		1.08	2.62	28.38
		1.04	2.54	30.92
		1.00	2.44	33.36
SWD	All Codes	**6.98**	**17.04**	**17.04**
		1.21	2.95	19.99
		1.16	2.82	22.81
		1.08	2.62	25.43
		1.05	2.56	28.00
		1.01	2.46	30.46
		1.01	2.46	32.92
SUA	All Codes	**7.25**	**17.68**	**17.68**
		1.22	2.97	20.65
		1.16	2.82	23.47
		1.07	2.60	26.07
		1.04	2.54	28.61
		1.01	2.47	31.08
		1.00	2.44	33.52
SWD/SUA	SUA=504 plan codes	**6.61**	**16.13**	**16.13**
		1.21	2.94	19.07
		1.17	2.86	21.93
		1.08	2.63	24.56
		1.06	2.58	27.14
		1.03	2.50	29.64
		1.01	2.47	32.12
		1.00	2.45	34.56
ELL/MLL/ SUA	SUA & ELL Codes	**4.95**	**12.07**	**12.07**
		1.27	3.09	15.16
		1.23	3.01	18.17
		1.19	2.89	21.06
		1.17	2.84	23.91
		1.14	2.78	26.69
		1.14	2.77	29.46

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/MLL/ SUA	SUA & ELL Codes	1.11	2.70	32.16
		1.09	2.66	34.82
		1.08	2.62	37.45
		1.06	2.57	40.02
		1.04	2.53	42.55
		1.01	2.46	45.01

Appendix M: Classical Test Theory Statistics

These tables support the classical test theory analyses described in Section 5, "Operational Test Data Collection and Classical Analysis." They include item type, sample size, *p*-value, percent of omitted responses and the point-biserial of the key. Field test items that do not contribute to students' scores have been omitted.

Table M1. ELA Grade 3 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	179,299	0.91	0.02	0.36
2	MC	179,213	0.56	0.07	0.33
3	MC	179,157	0.74	0.10	0.37
4	MC	179,169	0.71	0.09	0.31
5	MC	179,125	0.56	0.12	0.33
6	MC	179,109	0.48	0.13	0.25
13	MC	179,049	0.56	0.16	0.29
14	MC	178,998	0.46	0.19	0.40
15	MC	178,984	0.65	0.20	0.48
16	MC	178,979	0.44	0.20	0.46
17	MC	179,033	0.42	0.17	0.33
18	MC	178,998	0.56	0.19	0.45
19	MC	179,038	0.50	0.17	0.29
20	MC	178,990	0.63	0.19	0.39
21	MC	179,021	0.45	0.18	0.37
22	MC	178,891	0.49	0.25	0.39
23	MC	178,750	0.72	0.33	0.49
24	MC	178,601	0.51	0.41	0.31
25	CR	178,821	0.62	0.29	0.54
26	CR	178,178	0.63	0.65	0.50
27	CR	178,052	0.43	0.72	0.56
28	CR	177,389	0.59	1.09	0.59
29	CR	176,830	0.63	1.40	0.60
30	CR	176,480	0.48	1.59	0.53
31	CR	175,897	0.38	1.92	0.66

Table M2. ELA Grade 4 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	181,617	0.63	0.03	0.45
2	MC	181,623	0.58	0.03	0.33
3	MC	181,585	0.75	0.05	0.48
4	MC	181,577	0.72	0.05	0.38

Item	Type	N-Count	*p*-value	% Omit	PBis Key
5	MC	181,588	0.81	0.05	0.39
6	MC	181,551	0.50	0.07	0.22
7	MC	181,553	0.59	0.07	0.41
8	MC	181,571	0.49	0.06	0.33
9	MC	181,563	0.53	0.06	0.35
10	MC	181,535	0.71	0.08	0.39
11	MC	181,540	0.52	0.07	0.26
12	MC	181,549	0.68	0.07	0.41
19	MC	181,460	0.56	0.12	0.38
20	MC	181,471	0.51	0.11	0.25
21	MC	181,467	0.64	0.11	0.41
22	MC	181,380	0.57	0.16	0.32
23	MC	181,330	0.50	0.19	0.27
24	MC	181,178	0.68	0.27	0.38
25	CR	181,336	0.59	0.18	0.57
26	CR	180,723	0.58	0.52	0.57
27	CR	180,312	0.39	0.75	0.50
28	CR	180,979	0.56	0.38	0.58
29	CR	180,139	0.57	0.84	0.58
30	CR	180,122	0.63	0.85	0.64
31	CR	179,504	0.48	1.19	0.70

Table M3. ELA Grade 5 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	175,158	0.93	0.01	0.26
2	MC	175,120	0.56	0.03	0.24
3	MC	175,112	0.61	0.04	0.42
4	MC	175,117	0.84	0.03	0.42
5	MC	175,013	0.62	0.09	0.36
6	MC	175,115	0.90	0.03	0.42
7	MC	175,073	0.55	0.06	0.31
15	MC	175,105	0.70	0.04	0.52
16	MC	175,077	0.43	0.06	0.30
17	MC	175,048	0.48	0.07	0.36
18	MC	175,055	0.80	0.07	0.39
19	MC	175,031	0.44	0.08	0.26
20	MC	175,041	0.59	0.08	0.35
21	MC	175,062	0.43	0.06	0.25

Item	Type	N-Count	*p*-value	% Omit	PBis Key
22	MC	175,050	0.80	0.07	0.46
23	MC	175,002	0.35	0.10	0.24
24	MC	175,014	0.39	0.09	0.18
25	MC	174,970	0.66	0.12	0.39
26	MC	174,990	0.71	0.11	0.51
27	MC	175,001	0.57	0.10	0.38
28	MC	174,938	0.73	0.14	0.46
29	MC	174,924	0.65	0.14	0.35
30	MC	174,937	0.73	0.14	0.50
31	MC	174,986	0.33	0.11	0.14
32	MC	174,913	0.44	0.15	0.35
33	MC	174,882	0.45	0.17	0.23
34	MC	174,863	0.52	0.18	0.36
35	MC	174,673	0.76	0.29	0.42
36	CR	175,118	0.70	0.03	0.43
37	CR	174,760	0.75	0.24	0.49
38	CR	174,489	0.64	0.39	0.47
39	CR	174,681	0.69	0.28	0.53
40	CR	174,011	0.63	0.66	0.49
41	CR	173,596	0.60	0.90	0.47
42	CR	173,456	0.49	0.98	0.63

Table M4. ELA Grade 6 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	169,974	0.85	0.02	0.31
2	MC	169,946	0.66	0.04	0.39
3	MC	169,926	0.87	0.05	0.31
4	MC	169,912	0.68	0.06	0.37
5	MC	169,915	0.70	0.06	0.34
6	MC	169,942	0.87	0.04	0.33
7	MC	169,915	0.56	0.06	0.32
15	MC	169,930	0.48	0.05	0.39
16	MC	169,901	0.43	0.07	0.37
17	MC	169,910	0.69	0.06	0.51
18	MC	169,884	0.65	0.08	0.26
19	MC	169,874	0.57	0.08	0.37
20	MC	169,864	0.45	0.09	0.16
21	MC	169,886	0.36	0.08	0.20

Item	Type	N-Count	*p*-value	% Omit	PBis Key
22	MC	169,888	0.68	0.07	0.53
23	MC	169,856	0.70	0.09	0.36
24	MC	169,856	0.73	0.09	0.54
25	MC	169,858	0.73	0.09	0.42
26	MC	169,845	0.68	0.10	0.42
27	MC	169,828	0.69	0.11	0.35
28	MC	169,797	0.38	0.13	0.14
29	MC	169,784	0.40	0.14	0.24
30	MC	169,805	0.78	0.12	0.42
31	MC	169,818	0.52	0.12	0.34
32	MC	169,787	0.58	0.13	0.41
33	MC	169,789	0.76	0.13	0.46
34	MC	169,719	0.59	0.17	0.46
35	MC	169,667	0.69	0.20	0.45
36	CR	169,258	0.62	0.45	0.55
37	CR	169,315	0.76	0.41	0.55
38	CR	168,711	0.67	0.77	0.52
39	CR	169,395	0.75	0.36	0.57
40	CR	168,772	0.57	0.73	0.63
41	CR	168,856	0.67	0.68	0.61
42	CR	168,313	0.55	1.00	0.70

Table M5. ELA Grade 7 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	155,849	0.87	0.04	0.28
2	MC	155,868	0.75	0.03	0.29
3	MC	155,873	0.73	0.03	0.39
4	MC	155,810	0.56	0.07	0.10
5	MC	155,845	0.73	0.05	0.41
6	MC	155,797	0.68	0.08	0.43
7	MC	155,844	0.76	0.05	0.23
15	MC	155,812	0.48	0.07	0.29
16	MC	155,817	0.43	0.07	0.24
17	MC	155,761	0.54	0.10	0.33
18	MC	155,797	0.64	0.08	0.31
19	MC	155,795	0.67	0.08	0.26
20	MC	155,766	0.44	0.10	0.27
21	MC	155,787	0.41	0.08	0.24

Item	Type	N-Count	*p*-value	% Omit	PBis Key
22	MC	155,797	0.81	0.08	0.48
23	MC	155,782	0.75	0.09	0.38
24	MC	155,761	0.59	0.10	0.40
25	MC	155,720	0.60	0.13	0.34
26	MC	155,717	0.60	0.13	0.35
27	MC	155,724	0.60	0.13	0.53
28	MC	155,696	0.50	0.14	0.28
29	MC	155,738	0.56	0.12	0.44
30	MC	155,731	0.60	0.12	0.43
31	MC	155,753	0.49	0.11	0.39
32	MC	155,727	0.50	0.12	0.48
33	MC	155,689	0.42	0.15	0.39
34	MC	155,656	0.37	0.17	0.28
35	MC	155,597	0.56	0.21	0.38
36	CR	155,741	0.79	0.11	0.58
37	CR	155,189	0.80	0.47	0.60
38	CR	154,362	0.72	1.00	0.62
39	CR	154,884	0.74	0.66	0.60
40	CR	154,070	0.68	1.19	0.64
41	CR	154,049	0.67	1.20	0.54
42	CR	153,275	0.67	1.70	0.65
43	CR	152,688	0.56	2.07	0.69

Table M6. ELA Grade 8 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	151,504	0.90	0.01	0.19
2	MC	151,495	0.81	0.02	0.40
3	MC	151,452	0.41	0.05	0.32
4	MC	151,404	0.85	0.08	0.39
5	MC	151,443	0.60	0.05	0.28
6	MC	151,482	0.78	0.03	0.36
7	MC	151,446	0.68	0.05	0.42
8	MC	151,440	0.87	0.05	0.34
9	MC	151,429	0.66	0.06	0.43
10	MC	151,438	0.70	0.06	0.30
11	MC	151,437	0.78	0.06	0.48
12	MC	151,370	0.47	0.10	0.29
13	MC	151,425	0.89	0.06	0.44

Item	Type	N-Count	*p*-value	% Omit	PBis Key
14	MC	151,413	0.66	0.07	0.35
15	MC	151,454	0.81	0.04	0.31
16	MC	151,443	0.62	0.05	0.37
17	MC	151,404	0.83	0.08	0.38
18	MC	151,378	0.50	0.10	0.21
19	MC	151,436	0.76	0.06	0.39
20	MC	151,415	0.66	0.07	0.25
21	MC	151,406	0.51	0.08	0.20
29	MC	151,373	0.71	0.10	0.49
30	MC	151,351	0.59	0.11	0.37
31	MC	151,378	0.52	0.10	0.44
32	MC	151,257	0.56	0.17	0.37
33	MC	151,334	0.51	0.12	0.33
34	MC	151,308	0.41	0.14	0.31
35	MC	151,288	0.73	0.15	0.44
36	CR	151,327	0.78	0.13	0.55
37	CR	150,953	0.76	0.38	0.56
38	CR	150,556	0.55	0.64	0.51
39	CR	149,738	0.69	1.18	0.60
40	CR	149,190	0.65	1.54	0.62
41	CR	149,788	0.73	1.14	0.57
42	CR	149,341	0.72	1.44	0.57
43	CR	148,704	0.59	1.86	0.71

Table M7. Mathematics Grade 3 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	176,586	0.78	0.04	0.42
2	MC	176,545	0.85	0.07	0.44
3	MC	176,383	0.75	0.16	0.47
4	MC	176,387	0.57	0.16	0.48
6	MC	176,485	0.88	0.10	0.37
7	MC	176,428	0.61	0.13	0.29
9	MC	176,391	0.55	0.15	0.52
11	MC	176,452	0.74	0.12	0.55
12	MC	176,437	0.87	0.13	0.46
14	MC	176,452	0.69	0.12	0.56
15	MC	176,450	0.56	0.12	0.43
17	MC	176,393	0.66	0.15	0.41

Item	Type	N-Count	*p*-value	% Omit	PBis Key
18	MC	176,422	0.70	0.14	0.47
20	MC	176,399	0.74	0.15	0.58
21	MC	176,402	0.38	0.15	0.30
22	MC	176,421	0.62	0.14	0.52
23	MC	176,273	0.43	0.22	0.46
24	MC	176,216	0.31	0.25	0.29
25	MC	175,476	0.66	0.67	0.44
26	MC	176,608	0.87	0.03	0.38
27	MC	176,567	0.71	0.05	0.58
28	MC	176,499	0.51	0.09	0.36
29	MC	176,506	0.60	0.09	0.43
30	MC	176,510	0.77	0.09	0.46
31	MC	176,269	0.36	0.22	0.44
32	MC	176,510	0.94	0.09	0.31
33	MC	176,342	0.32	0.18	0.40
34	CR	175,843	0.57	0.46	0.63
35	CR	176,058	0.81	0.34	0.46
36	CR	175,968	0.50	0.39	0.57
37	CR	176,082	0.56	0.33	0.64
38	CR	175,840	0.63	0.47	0.69
39	CR	175,819	0.55	0.48	0.61
40	CR	175,698	0.39	0.55	0.69

Table M8. Mathematics Grade 4 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	176,864	0.87	0.02	0.41
2	MC	176,825	0.78	0.04	0.58
3	MC	176,780	0.73	0.07	0.52
4	MC	176,761	0.54	0.08	0.33
6	MC	176,743	0.70	0.09	0.47
7	MC	176,600	0.74	0.17	0.49
9	MC	176,674	0.55	0.13	0.55
10	MC	176,689	0.65	0.12	0.61
12	MC	176,654	0.55	0.14	0.50
13	MC	176,435	0.37	0.26	0.42
14	MC	176,693	0.81	0.12	0.47
16	MC	176,654	0.48	0.14	0.44
17	MC	176,744	0.69	0.09	0.42

Item	Type	N-Count	*p*-value	% Omit	PBis Key
18	MC	176,747	0.90	0.08	0.40
20	MC	176,688	0.66	0.12	0.43
21	MC	176,690	0.52	0.12	0.50
23	MC	176,478	0.62	0.24	0.49
24	MC	176,638	0.68	0.15	0.50
25	MC	176,622	0.57	0.16	0.54
27	MC	176,602	0.52	0.17	0.54
28	MC	176,601	0.69	0.17	0.49
29	MC	176,524	0.52	0.21	0.53
30	MC	176,251	0.61	0.37	0.52
31	MC	176,850	0.81	0.03	0.37
32	MC	176,810	0.57	0.05	0.37
33	MC	176,824	0.87	0.04	0.42
34	MC	176,798	0.65	0.06	0.53
35	MC	176,786	0.52	0.06	0.38
36	MC	176,780	0.65	0.07	0.47
37	MC	176,718	0.52	0.10	0.49
38	MC	176,637	0.62	0.15	0.40
39	CR	176,417	0.59	0.27	0.56
40	CR	176,393	0.51	0.28	0.70
41	CR	176,295	0.48	0.34	0.43
42	CR	176,305	0.44	0.33	0.64
43	CR	176,248	0.66	0.37	0.63
44	CR	176,079	0.35	0.46	0.66
45	CR	176,080	0.60	0.46	0.67

Table M9. Mathematics Grade 5 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	168,389	0.74	0.11	0.49
2	MC	168,503	0.70	0.04	0.54
3	MC	168,486	0.90	0.05	0.33
4	MC	168,327	0.61	0.15	0.44
6	MC	168,460	0.89	0.07	0.41
7	MC	168,236	0.49	0.20	0.52
8	MC	168,431	0.66	0.09	0.46
10	MC	168,394	0.54	0.11	0.59
11	MC	168,346	0.56	0.14	0.37
13	MC	168,376	0.36	0.12	0.32

Item	Type	N-Count	*p*-value	% Omit	PBis Key
14	MC	168,398	0.59	0.11	0.50
16	MC	168,331	0.74	0.15	0.49
17	MC	168,440	0.47	0.08	0.51
18	MC	168,386	0.57	0.11	0.46
20	MC	168,361	0.61	0.13	0.45
21	MC	168,438	0.65	0.08	0.50
22	MC	168,361	0.47	0.13	0.47
24	MC	168,323	0.66	0.15	0.57
25	MC	168,389	0.55	0.11	0.47
27	MC	168,388	0.74	0.11	0.48
28	MC	168,336	0.87	0.14	0.43
29	MC	168,329	0.73	0.15	0.40
30	MC	168,007	0.70	0.34	0.50
31	MC	168,508	0.41	0.04	0.50
32	MC	168,496	0.78	0.05	0.52
33	MC	168,494	0.86	0.05	0.45
34	MC	168,507	0.75	0.04	0.51
35	MC	168,406	0.44	0.10	0.56
36	MC	168,340	0.58	0.14	0.51
37	MC	168,442	0.49	0.08	0.29
38	MC	168,230	0.62	0.21	0.47
39	CR	167,718	0.34	0.51	0.59
40	CR	167,640	0.46	0.56	0.64
41	CR	167,906	0.33	0.40	0.62
42	CR	168,022	0.38	0.33	0.52
43	CR	168,209	0.44	0.22	0.71
44	CR	168,044	0.34	0.32	0.59
45	CR	167,948	0.39	0.37	0.64

Table M10. Mathematics Grade 6 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	164,397	0.81	0.02	0.38
2	MC	164,076	0.63	0.21	0.59
3	MC	164,252	0.63	0.11	0.41
4	MC	164,308	0.58	0.07	0.42
6	MC	164,262	0.70	0.10	0.57
7	MC	164,301	0.71	0.08	0.42
9	MC	164,251	0.61	0.11	0.50

Item	Type	N-Count	*p*-value	% Omit	PBis Key
10	MC	164,228	0.73	0.12	0.47
12	MC	164,306	0.42	0.07	0.52
13	MC	164,014	0.42	0.25	0.61
15	MC	164,079	0.42	0.21	0.46
16	MC	164,247	0.28	0.11	0.41
18	MC	164,201	0.61	0.14	0.56
19	MC	164,132	0.67	0.18	0.55
21	MC	164,206	0.46	0.14	0.62
22	MC	164,211	0.40	0.13	0.43
24	MC	164,223	0.58	0.13	0.45
25	MC	164,216	0.69	0.13	0.45
26	MC	164,111	0.55	0.19	0.37
27	MC	164,203	0.49	0.14	0.53
28	MC	164,078	0.55	0.21	0.57
29	MC	163,999	0.55	0.26	0.45
30	MC	164,133	0.62	0.18	0.53
31	MC	163,829	0.77	0.36	0.41
32	MC	164,377	0.82	0.03	0.43
33	MC	164,327	0.73	0.06	0.38
34	MC	164,315	0.68	0.07	0.40
35	MC	164,216	0.48	0.13	0.40
36	MC	164,164	0.61	0.16	0.41
37	MC	164,151	0.28	0.17	0.34
38	MC	163,882	0.80	0.33	0.42
39	CR	163,711	0.54	0.44	0.65
40	CR	163,234	0.38	0.73	0.58
41	CR	163,110	0.30	0.80	0.64
42	CR	163,116	0.25	0.80	0.68
43	CR	162,761	0.35	1.01	0.59
44	CR	163,543	0.30	0.54	0.64
45	CR	163,160	0.22	0.77	0.47
46	CR	163,007	0.21	0.86	0.65

Table M11. Mathematics Grade 7 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	151,686	0.83	0.04	0.41
2	MC	151,318	0.43	0.28	0.41
3	MC	151,639	0.75	0.07	0.47

Item	Type	N-Count	*p*-value	% Omit	PBis Key
4	MC	151,609	0.56	0.09	0.50
6	MC	151,586	0.51	0.11	0.56
7	MC	151,524	0.68	0.15	0.47
8	MC	151,287	0.58	0.30	0.53
10	MC	151,619	0.57	0.09	0.38
11	MC	151,559	0.53	0.13	0.46
13	MC	151,449	0.55	0.20	0.59
14	MC	151,590	0.58	0.10	0.46
16	MC	151,454	0.54	0.19	0.38
17	MC	151,571	0.71	0.12	0.50
19	MC	151,579	0.57	0.11	0.47
20	MC	151,464	0.56	0.19	0.56
22	MC	151,550	0.60	0.13	0.61
23	MC	151,457	0.35	0.19	0.29
24	MC	151,574	0.87	0.12	0.41
26	MC	151,434	0.49	0.21	0.45
27	MC	151,557	0.68	0.13	0.43
28	MC	151,450	0.57	0.20	0.43
29	MC	151,466	0.33	0.19	0.44
30	MC	151,440	0.58	0.20	0.47
31	MC	151,481	0.62	0.18	0.45
32	MC	151,465	0.58	0.19	0.45
33	MC	151,103	0.54	0.43	0.39
34	MC	151,634	0.46	0.08	0.35
35	MC	151,665	0.40	0.06	0.45
36	MC	151,586	0.53	0.11	0.52
37	MC	151,615	0.61	0.09	0.40
38	MC	151,649	0.49	0.07	0.43
39	MC	151,534	0.39	0.14	0.45
40	MC	151,430	0.51	0.21	0.48
41	CR	150,824	0.42	0.61	0.70
42	CR	150,567	0.72	0.78	0.55
43	CR	150,123	0.43	1.07	0.73
44	CR	150,693	0.60	0.70	0.68
45	CR	150,758	0.66	0.65	0.66
46	CR	149,536	0.34	1.46	0.74
47	CR	150,540	0.55	0.80	0.67
48	CR	150,401	0.46	0.89	0.72

Table M12. Mathematics Grade 8 Classical Item Analysis

Item	Type	N-Count	*p*-value	% Omit	PBis Key
1	MC	108,335	0.72	0.07	0.29
2	MC	108,343	0.64	0.06	0.45
3	MC	108,272	0.68	0.13	0.36
4	MC	108,195	0.67	0.20	0.48
6	MC	108,260	0.45	0.14	0.55
7	MC	108,044	0.39	0.34	0.33
8	MC	108,170	0.48	0.22	0.47
10	MC	108,277	0.54	0.12	0.45
11	MC	108,313	0.79	0.09	0.39
12	MC	108,308	0.64	0.09	0.48
14	MC	108,206	0.44	0.19	0.38
15	MC	108,297	0.49	0.10	0.45
16	MC	108,273	0.69	0.13	0.43
18	MC	108,246	0.49	0.15	0.40
19	MC	108,168	0.53	0.22	0.43
20	MC	108,240	0.63	0.16	0.45
22	MC	108,215	0.37	0.18	0.43
23	MC	108,241	0.38	0.16	0.35
24	MC	108,176	0.60	0.22	0.42
26	MC	108,139	0.34	0.25	0.25
27	MC	108,235	0.52	0.16	0.44
28	MC	108,161	0.40	0.23	0.43
30	MC	108,213	0.52	0.18	0.44
31	MC	108,195	0.62	0.20	0.34
32	MC	108,092	0.48	0.29	0.34
33	MC	107,873	0.47	0.50	0.34
34	MC	108,352	0.61	0.05	0.53
35	MC	108,326	0.56	0.08	0.39
36	MC	108,322	0.45	0.08	0.46
37	MC	108,324	0.43	0.08	0.31
38	MC	108,320	0.52	0.08	0.49
39	MC	108,268	0.50	0.13	0.52
40	MC	108,268	0.32	0.13	0.38
41	CR	107,081	0.43	1.23	0.64
42	CR	105,046	0.22	3.10	0.61
43	CR	105,691	0.41	2.51	0.65
44	CR	106,777	0.58	1.51	0.65

Item	Type	N-Count	*p*-value	% Omit	PBis Key
45	CR	106,599	0.49	1.67	0.61
46	CR	106,591	0.54	1.68	0.61
47	CR	104,494	0.28	3.61	0.65
48	CR	106,163	0.39	2.07	0.56

Appendix N: Items Flagged for DIF

These tables support the DIF information in Section 5, "Operational Test Data Collection and Classical Analysis." They include item numbers, focal group, and directions of DIF and DIF statistics. Tables N1–N4 show items flagged by the SMD, or Mantel-Haenszel methods. Positive values of SMD and Delta in Tables N1–N4 indicate DIF in favor of a focal group, and negative values of SMD and Delta indicate DIF against a focal group. Field test items that do not contribute to students' scores have been omitted.

Table N1. ELA MC Item Classical DIF Flags

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
3	13	Asian	Against	1.62	713.4	-1.13
3	21	ELL/MLL	Against	1.56	620.0	-1.04
3	23	CBT	In Favor	0.58	647.5	1.27
4	3	Hispanic	Against	1.54	786.0	-1.02
4	3	Asian	Against	1.54	290.7	-1.02
4	3	ELL/MLL	Against	1.80	1086.9	-1.39
4	12	ELL/MLL	Against	1.70	947.6	-1.24
4	20	Hispanic	Against	1.60	1504.0	-1.10
5	2	Hispanic	Against	1.55	1233.5	-1.02
5	2	Asian	Against	1.73	980.7	-1.29
5	2	ELL/MLL	Against	1.76	856.7	-1.33
5	3	ELL/MLL	Against	1.60	559.5	-1.11
5	4	Black	Against	1.56	464.0	-1.04
5	4	Hispanic	Against	1.64	714.7	-1.16
5	4	Asian	Against	1.88	480.5	-1.49
5	4	High Needs	Against	1.65	809.0	-1.17
5	4	ELL/MLL	Against	1.81	902.2	-1.40
5	6	Hispanic	Against	1.98	870.0	-1.60
5	6	Asian	Against	2.66	735.7	-2.30
5	6	High Needs	Against	1.99	900.0	-1.62
5	6	ELL/MLL	Against	3.17	3051.0	-2.71
5	6	CBT	In Favor	0.59	247.0	1.26
5	23	ELL/MLL	Against	1.56	357.0	-1.04
5	30	ELL/MLL	Against	1.64	645.5	-1.16
6	1	ELL/MLL	Against	2.02	1174.2	-1.65
6	2	ELL/MLL	Against	1.54	458.4	-1.02
6	3	Asian	Against	1.76	387.2	-1.33
6	3	ELL/MLL	Against	2.18	1335.8	-1.83
6	4	Asian	Against	1.57	466.0	-1.06
6	4	ELL/MLL	Against	1.71	741.2	-1.26

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
6	6	Black	Against	1.72	726.0	-1.28
6	6	Asian	Against	1.55	193.9	-1.03
6	6	High Needs	Against	1.55	558.3	-1.03
6	17	Hispanic	Against	1.63	1054.1	-1.15
6	17	High Needs	Against	1.69	1334.1	-1.23
6	17	ELL/MLL	Against	1.71	706.0	-1.26
6	17	CBT	In Favor	0.62	547.8	1.11
6	25	Asian	Against	1.82	737.9	-1.40
6	25	High Needs	Against	1.58	1050.1	-1.08
6	25	ELL/MLL	Against	1.67	663.6	-1.21
6	34	Black	Against	1.78	1390.4	-1.36
6	34	High Needs	Against	1.62	1404.5	-1.13
6	35	High Needs	Against	1.55	1022.1	-1.03
7	1	Asian	Against	1.59	278.8	-1.08
7	1	ELL/MLL	Against	2.13	1092.3	-1.78
7	3	Black	Against	1.64	829.9	-1.16
7	3	Hispanic	Against	1.77	1318.7	-1.34
7	3	Asian	Against	2.58	1694.3	-2.23
7	3	High Needs	Against	1.76	1527.3	-1.33
7	3	ELL/MLL	Against	2.29	1499.8	-1.95
7	3	CBT	In Favor	0.57	644.2	1.30
7	6	Hispanic	Against	1.72	1282.7	-1.27
7	6	Asian	Against	2.09	1086.5	-1.73
7	6	High Needs	Against	1.60	1180.3	-1.11
7	6	ELL/MLL	Against	2.20	1274.4	-1.85
7	18	ELL/MLL	Against	1.57	431.4	-1.05
7	31	Hispanic	Against	1.53	955.2	-1.01
7	31	ELL/MLL	Against	1.55	327.3	-1.04
7	32	Black	Against	1.88	1472.5	-1.48
7	32	Hispanic	Against	1.91	1945.8	-1.52
7	32	High Needs	Against	1.76	1868.0	-1.33
8	2	Hispanic	Against	1.84	1176.3	-1.43
8	2	Asian	Against	1.75	409.3	-1.31
8	2	High Needs	Against	1.83	1220.3	-1.42
8	2	ELL/MLL	Against	2.77	2061.0	-2.39
8	5	ELL/MLL	Against	1.72	580.3	-1.28
8	9	Asian	Against	1.63	489.3	-1.14
8	9	ELL/MLL	Against	1.62	444.7	-1.13

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
8	11	ELL/MLL	Against	1.70	538.2	-1.24
8	14	Asian	Against	1.68	662.3	-1.22
8	15	Asian	In Favor	0.63	287.8	1.07
8	16	Black	Against	1.71	1174.4	-1.27
8	16	Hispanic	Against	1.68	1298.0	-1.22
8	16	Asian	Against	1.70	666.3	-1.25
8	16	ELL/MLL	Against	1.56	377.0	-1.04
8	18	Asian	In Favor	0.57	935.9	1.31
8	30	Black	Against	1.79	1404.5	-1.37
8	30	Asian	Against	1.62	558.8	-1.13
8	30	High Needs	Against	1.68	1587.6	-1.22

Table N2. ELA CR Item Classical DIF Flags

Grade	Item	Subgroup	DIF	SMD	Effect
3	26	Asian	In Favor	0.12	0.2
3	31	Asian	In Favor	0.25	0.2
3	31	CBT	Against	0.02	0.0
4	27	High Needs	In Favor	0.16	0.2
4	27	CBT	Against	0.03	0.0
5	36	Black	In Favor	0.12	0.2
5	36	Hispanic	In Favor	0.12	0.2
5	36	Asian	In Favor	0.10	0.2
5	36	High Needs	In Favor	0.12	0.2
5	36	CBT	Against	0.08	0.1
5	38	Black	In Favor	0.12	0.2
5	38	Hispanic	In Favor	0.11	0.2
5	38	Asian	In Favor	0.13	0.2
5	38	High Needs	In Favor	0.14	0.2
5	39	Hispanic	In Favor	0.10	0.2
5	39	Asian	In Favor	0.11	0.2
5	39	High Needs	In Favor	0.12	0.2
5	39	CBT	Against	0.06	0.1
5	42	CBT	Against	0.12	0.1
6	37	High Needs	In Favor	0.11	0.2
6	38	Black	In Favor	0.13	0.2
6	38	Hispanic	In Favor	0.13	0.2
6	38	Asian	In Favor	0.14	0.2
6	38	High Needs	In Favor	0.18	0.3

Grade	Item	Subgroup	DIF	SMD	Effect
6	41	Black	In Favor	0.12	0.2
6	41	High Needs	In Favor	0.13	0.2
6	42	Asian	In Favor	0.19	0.2
6	42	CBT	Against	0.12	0.1
7	39	Hispanic	In Favor	0.12	0.2
7	39	High Needs	In Favor	0.14	0.2
7	42	High Needs	In Favor	0.14	0.2
7	43	Asian	In Favor	0.20	0.2
7	43	CBT	Against	0.19	0.2
8	36	Black	In Favor	0.11	0.2
8	36	Hispanic	In Favor	0.11	0.2
8	36	High Needs	In Favor	0.15	0.3
8	37	Black	In Favor	0.14	0.2
8	37	Hispanic	In Favor	0.14	0.2
8	37	Asian	In Favor	0.12	0.2
8	37	High Needs	In Favor	0.18	0.3
8	37	ELL/MLL	In Favor	0.11	0.2

Table N3. Mathematics MC Item Classical DIF Flags

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
3	3	Asian	In Favor	0.55	458.7	1.42
3	4	Asian	In Favor	0.65	416.9	1.01
3	12	Asian	In Favor	0.61	130.1	1.17
4	2	Black	Against	1.81	870.7	-1.40
4	2	Hispanic	Against	1.81	1071.8	-1.39
4	2	Asian	Against	1.60	203.8	-1.11
4	2	High Needs	Against	1.59	732.3	-1.09
4	2	ELL/MLL	Against	1.78	892.6	-1.35
4	29	Black	Against	1.64	891.4	-1.16
4	29	Hispanic	Against	1.68	1348.4	-1.22
4	29	Asian	Against	1.60	464.4	-1.11
4	29	High Needs	Against	1.59	1334.5	-1.09
5	1	Asian	In Favor	0.65	232.6	1.02
5	16	Asian	In Favor	0.59	337.9	1.26
5	20	Asian	In Favor	0.64	423.1	1.05
5	33	CBT	In Favor	0.60	188.7	1.20
6	27	Black	Against	1.62	791.5	-1.13
6	32	ELL/MLL	Against	1.54	434.0	-1.01

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
7	1	Female	In Favor	0.65	765.5	1.00
7	22	ELL/MLL	Against	1.98	693.5	-1.61

Table N4. Mathematics CR Item Classical DIF Flags

Grade	Item	Subgroup	DIF	SMD	Effect
7	42	ELL/MLL	Against	-0.13	0.2
8	43	CBT	Against	0.05	0.1

Appendix O: IRT Statistics

Field test items that do not contribute to students' scores have been omitted.

Table O1. ELA Grade 3 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	736.39	8	182.10	478.13	Y
2	3PL	425.32	8	104.33	477.90	Y
3	3PL	468.04	8	115.01	477.75	Y
4	3PL	359.00	8	87.75	477.78	Y
5	3PL	603.30	8	148.82	477.67	Y
6	3PL	262.60	8	63.65	477.62	Y
13	3PL	295.99	8	72.00	477.46	Y
14	3PL	926.48	8	229.62	477.33	Y
15	3PL	935.38	8	231.84	477.29	Y
16	3PL	1266.10	8	314.52	477.28	Y
17	3PL	630.24	8	155.56	477.42	Y
18	3PL	916.71	8	227.18	477.33	Y
19	3PL	345.12	8	84.28	477.43	Y
20	3PL	623.16	8	153.79	477.31	Y
21	3PL	721.39	8	178.35	477.39	Y
22	3PL	852.20	8	211.05	477.04	Y
23	3PL	978.70	8	242.68	476.67	Y
24	3PL	418.57	8	102.64	476.27	Y
25	2PPC	1140.60	17	192.69	476.86	Y
26	2PPC	1071.10	17	180.77	475.14	Y
27	2PPC	1294.00	17	219.01	474.81	Y
28	2PPC	1628.70	17	276.40	473.04	Y
29	2PPC	1278.50	17	216.35	471.55	Y
30	2PPC	1138.50	17	192.34	470.61	Y
31	2PPC	908.86	35	104.45	469.06	Y

Table O2. ELA Grade 4 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	602.68	8	148.67	484.31	Y
2	3PL	428.00	8	105.00	484.33	Y
3	3PL	749.62	8	185.40	484.23	Y
4	3PL	509.41	8	125.35	484.21	Y
5	3PL	661.45	8	163.36	484.23	Y
6	3PL	427.35	8	104.84	484.14	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
7	3PL	597.02	8	147.26	484.14	Y
8	3PL	520.56	8	128.14	484.19	Y
9	3PL	534.71	8	131.68	484.17	Y
10	3PL	445.73	8	109.43	484.09	Y
11	3PL	250.44	8	60.61	484.11	Y
12	3PL	492.57	8	121.14	484.13	Y
19	3PL	627.17	8	154.79	483.89	Y
20	3PL	379.48	8	92.87	483.92	Y
21	3PL	586.27	8	144.57	483.91	Y
22	3PL	797.21	8	197.30	483.68	Y
23	3PL	288.65	8	70.16	483.55	Y
24	3PL	510.11	8	125.53	483.14	Y
25	2PPC	1052.60	17	177.61	483.56	Y
26	2PPC	1108.90	17	187.27	481.93	Y
27	2PPC	1412.10	17	239.25	480.83	Y
28	2PPC	2379.40	17	405.15	482.61	Y
29	2PPC	1717.80	17	291.68	480.37	Y
30	2PPC	1628.60	17	276.38	480.33	Y
31	2PPC	1591.10	35	185.99	478.68	Y

Table O3. ELA Grade 5 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	311.13	8	75.78	467.09	Y
2	3PL	270.69	8	65.67	466.99	Y
3	3PL	603.80	8	148.95	466.97	Y
4	3PL	641.25	8	158.31	466.98	Y
5	3PL	592.05	8	146.01	466.70	Y
6	3PL	671.14	8	165.78	466.97	Y
7	3PL	592.69	8	146.17	466.86	Y
15	3PL	989.34	8	245.34	466.95	Y
16	3PL	516.16	8	127.04	466.87	Y
17	3PL	766.61	8	189.65	466.79	Y
18	3PL	531.49	8	130.87	466.81	Y
19	3PL	1333.20	8	331.29	466.75	Y
20	3PL	721.10	8	178.27	466.78	Y
21	3PL	448.54	8	110.13	466.83	Y
22	3PL	615.07	8	151.77	466.80	Y
23	3PL	437.97	8	107.49	466.67	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
24	3PL	199.92	8	47.98	466.70	Y
25	3PL	474.67	8	116.67	466.59	Y
26	3PL	947.26	8	234.81	466.64	Y
27	3PL	574.61	8	141.65	466.67	Y
28	3PL	1070.00	8	265.50	466.50	Y
29	3PL	436.60	8	107.15	466.46	Y
30	3PL	809.35	8	200.34	466.50	Y
31	3PL	344.72	8	84.18	466.63	Y
32	3PL	856.88	8	212.22	466.43	Y
33	3PL	290.16	8	70.54	466.35	Y
34	3PL	780.77	8	193.19	466.30	Y
35	3PL	502.33	8	123.58	465.79	Y
36	2PPC	407.52	17	66.97	466.98	Y
37	2PPC	506.11	17	83.88	466.03	Y
38	2PPC	533.94	17	88.65	465.30	Y
39	2PPC	638.87	17	106.65	465.82	Y
40	2PPC	566.44	17	94.23	464.03	Y
41	2PPC	817.66	17	137.31	462.92	Y
42	2PPC	759.09	35	86.55	462.55	Y

Table O4. ELA Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	249.55	8	60.39	453.26	Y
2	3PL	419.10	8	102.78	453.19	Y
3	3PL	309.60	8	75.40	453.14	Y
4	3PL	347.18	8	84.80	453.10	Y
5	3PL	375.29	8	91.82	453.11	Y
6	3PL	286.31	8	69.58	453.18	Y
7	3PL	313.61	8	76.40	453.11	Y
15	3PL	669.95	8	165.49	453.15	Y
16	3PL	719.40	8	177.85	453.07	Y
17	3PL	611.92	8	150.98	453.09	Y
18	3PL	897.76	8	222.44	453.02	Y
19	3PL	660.76	8	163.19	453.00	Y
20	3PL	585.47	8	144.37	452.97	Y
21	3PL	310.75	8	75.69	453.03	Y
22	3PL	586.52	8	144.63	453.03	Y
23	3PL	315.74	8	76.93	452.95	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
24	3PL	646.23	8	159.56	452.95	Y
25	3PL	548.51	8	135.13	452.95	Y
26	3PL	857.47	8	212.37	452.92	Y
27	3PL	553.49	8	136.37	452.87	Y
28	3PL	133.63	8	31.41	452.79	Y
29	3PL	490.55	8	120.64	452.76	Y
30	3PL	514.07	8	126.52	452.81	Y
31	3PL	507.59	8	124.90	452.85	Y
32	3PL	624.18	8	154.05	452.77	Y
33	3PL	447.68	8	109.92	452.77	Y
34	3PL	656.93	8	162.23	452.58	Y
35	3PL	418.19	8	102.55	452.45	Y
36	2PPC	1474.40	17	249.94	451.35	Y
37	2PPC	553.90	17	92.08	451.51	Y
38	2PPC	815.07	17	136.87	449.90	Y
39	2PPC	593.94	17	98.95	451.72	Y
40	2PPC	2045.80	17	347.93	450.06	Y
41	2PPC	1168.10	17	197.41	450.28	Y
42	2PPC	1207.30	35	140.12	448.83	Y

Table O5. ELA Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	221.67	8	53.42	415.60	Y
2	3PL	142.95	8	33.74	415.65	Y
3	3PL	312.45	8	76.11	415.66	Y
4	3PL	127.17	8	29.79	415.49	Y
5	3PL	258.15	8	62.54	415.59	Y
6	3PL	358.28	8	87.57	415.46	Y
7	3PL	613.59	8	151.40	415.58	Y
15	3PL	236.03	8	57.01	415.50	Y
16	3PL	329.30	8	80.33	415.51	Y
17	3PL	384.71	8	94.18	415.36	Y
18	3PL	242.49	8	58.62	415.46	Y
19	3PL	520.82	8	128.20	415.45	Y
20	3PL	267.34	8	64.84	415.38	Y
21	3PL	387.83	8	94.96	415.43	Y
22	3PL	291.97	8	70.99	415.46	Y
23	3PL	218.91	8	52.73	415.42	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
24	3PL	446.05	8	109.51	415.36	Y
25	3PL	207.82	8	49.95	415.25	Y
26	3PL	261.24	8	63.31	415.25	Y
27	3PL	684.49	8	169.12	415.26	Y
28	3PL	256.54	8	62.13	415.19	Y
29	3PL	374.57	8	91.64	415.30	Y
30	3PL	430.75	8	105.69	415.28	Y
31	3PL	480.60	8	118.15	415.34	Y
32	3PL	822.00	8	203.50	415.27	Y
33	3PL	797.22	8	197.31	415.17	Y
34	3PL	1171.40	8	290.86	415.08	Y
35	3PL	379.11	8	92.78	414.93	Y
36	2PPC	590.66	17	98.38	415.31	Y
37	2PPC	521.11	17	86.45	413.84	Y
38	2PPC	474.90	17	78.53	411.63	Y
39	2PPC	670.19	17	112.02	413.02	Y
40	2PPC	704.47	17	117.90	410.85	Y
41	2PPC	693.82	17	116.07	410.80	Y
42	2PPC	1084.50	17	183.07	408.73	Y
43	2PPC	878.41	35	100.81	407.17	Y

Table O6. ELA Grade 8 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	340.16	8	83.04	404.01	Y
2	3PL	250.01	8	60.50	403.99	Y
3	3PL	444.05	8	109.01	403.87	Y
4	3PL	240.44	8	58.11	403.74	Y
5	3PL	179.16	8	42.79	403.85	Y
6	3PL	366.31	8	89.58	403.95	Y
7	3PL	379.92	8	92.98	403.86	Y
8	3PL	518.51	8	127.63	403.84	Y
9	3PL	406.18	8	99.54	403.81	Y
10	3PL	1055.50	8	261.86	403.83	Y
11	3PL	349.50	8	85.38	403.83	Y
12	3PL	601.69	8	148.42	403.65	Y
13	3PL	343.99	8	84.00	403.80	Y
14	3PL	279.17	8	67.79	403.77	Y
15	3PL	428.24	8	105.06	403.88	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
16	3PL	325.67	8	79.42	403.85	Y
17	3PL	280.60	8	68.15	403.74	Y
18	3PL	190.30	8	45.58	403.67	Y
19	3PL	268.08	8	65.02	403.83	Y
20	3PL	936.03	8	232.01	403.77	Y
21	3PL	802.22	8	198.55	403.75	Y
29	3PL	560.57	8	138.14	403.66	Y
30	3PL	450.37	8	110.59	403.60	Y
31	3PL	719.82	8	177.95	403.67	Y
32	3PL	630.93	8	155.73	403.35	Y
33	3PL	403.69	8	98.92	403.56	Y
34	3PL	945.28	8	234.32	403.49	Y
35	3PL	355.87	8	86.97	403.43	Y
36	2PPC	314.07	17	50.95	403.54	Y
37	2PPC	304.45	17	49.30	402.54	Y
38	2PPC	3348.10	17	571.28	401.48	N
39	2PPC	1054.00	17	177.85	399.30	Y
40	2PPC	947.15	17	159.52	397.84	Y
41	2PPC	916.45	17	154.25	399.43	Y
42	2PPC	1092.80	17	184.50	398.24	Y
43	2PPC	1363.60	35	158.79	396.54	Y

Table O7. Mathematics Grade 3 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	358.73	8	87.68	470.90	Y
2	3PL	368.59	8	90.15	470.79	Y
3	3PL	534.30	8	131.57	470.35	Y
4	3PL	504.57	8	124.14	470.37	Y
6	3PL	339.46	8	82.87	470.63	Y
7	3PL	267.02	8	64.75	470.47	Y
9	3PL	508.84	8	125.21	470.38	Y
11	3PL	526.80	8	129.70	470.54	Y
12	3PL	809.85	8	200.46	470.50	Y
14	3PL	423.66	8	103.91	470.54	Y
15	3PL	374.50	8	91.63	470.53	Y
17	3PL	364.37	8	89.09	470.38	Y
18	3PL	340.29	8	83.07	470.46	Y
20	3PL	546.22	8	134.55	470.40	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
21	3PL	483.29	8	118.82	470.41	Y
22	3PL	490.65	8	120.66	470.46	Y
23	3PL	600.95	8	148.24	470.06	Y
24	3PL	1139.80	8	282.95	469.91	Y
25	3PL	439.05	8	107.76	467.94	Y
26	3PL	423.10	8	103.78	470.95	Y
27	3PL	571.27	8	140.82	470.85	Y
28	3PL	489.60	8	120.40	470.66	Y
29	3PL	393.13	8	96.28	470.68	Y
30	3PL	264.56	8	64.14	470.69	Y
31	3PL	1245.70	8	309.42	470.05	Y
32	3PL	514.96	8	126.74	470.69	Y
33	3PL	1279.50	8	317.86	470.25	Y
34	2PPC	541.25	17	89.91	468.91	Y
35	2PPC	349.52	17	57.03	469.49	Y
36	2PPC	423.10	17	69.65	469.25	Y
37	2PPC	547.48	17	90.98	469.55	Y
38	2PPC	2151.90	17	366.13	468.91	Y
39	2PPC	1024.20	17	172.74	468.85	Y
40	2PPC	582.80	26	77.21	468.53	Y

Table O8. Mathematics Grade 4 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	506.87	8	124.72	471.64	Y
2	3PL	639.18	8	157.79	471.53	Y
3	3PL	352.78	8	86.19	471.41	Y
4	3PL	299.29	8	72.82	471.36	Y
6	3PL	324.16	8	79.04	471.31	Y
7	3PL	589.62	8	145.41	470.93	Y
9	3PL	412.65	8	101.16	471.13	Y
10	3PL	490.33	8	120.58	471.17	Y
12	3PL	444.32	8	109.08	471.08	Y
13	3PL	743.24	8	183.81	470.49	Y
14	3PL	646.89	8	159.72	471.18	Y
16	3PL	362.83	8	88.71	471.08	Y
17	3PL	210.60	8	50.65	471.32	Y
18	3PL	1461.30	8	363.33	471.33	Y
20	3PL	455.31	8	111.83	471.17	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
21	3PL	544.44	8	134.11	471.17	Y
23	3PL	290.05	8	70.51	470.61	Y
24	3PL	256.71	8	62.18	471.03	Y
25	3PL	416.92	8	102.23	470.99	Y
27	3PL	528.57	8	130.14	470.94	Y
28	3PL	258.33	8	62.58	470.94	Y
29	3PL	603.72	8	148.93	470.73	Y
30	3PL	353.94	8	86.49	470.00	Y
31	3PL	302.92	8	73.73	471.60	Y
32	3PL	446.51	8	109.63	471.49	Y
33	3PL	927.50	8	229.87	471.53	Y
34	3PL	294.62	8	71.66	471.46	Y
35	3PL	476.33	8	117.08	471.43	Y
36	3PL	344.50	8	84.12	471.41	Y
37	3PL	474.71	8	116.68	471.25	Y
38	3PL	493.02	8	121.26	471.03	Y
39	2PPC	2014.50	17	342.57	470.45	Y
40	2PPC	582.37	17	96.96	470.38	Y
41	2PPC	1687.80	17	286.53	470.12	Y
42	2PPC	1341.30	17	227.11	470.15	Y
43	2PPC	6437.10	17	1101.00	469.99	N
44	2PPC	618.38	17	103.14	469.54	Y
45	2PPC	142.93	26	16.22	469.55	Y

Table O9. Mathematics Grade 5 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	790.17	8	195.54	449.04	Y
2	3PL	346.08	8	84.52	449.34	Y
3	3PL	278.96	8	67.74	449.30	Y
4	3PL	1044.70	8	259.18	448.87	Y
6	3PL	1417.10	8	352.28	449.23	Y
7	3PL	441.39	8	108.35	448.63	Y
8	3PL	582.30	8	143.58	449.15	Y
10	3PL	813.74	8	201.43	449.05	Y
11	3PL	645.93	8	159.48	448.92	Y
13	3PL	313.70	8	76.43	449.00	Y
14	3PL	337.98	8	82.49	449.06	Y
16	3PL	1053.90	8	261.47	448.88	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
17	3PL	615.02	8	151.76	449.17	Y
18	3PL	299.83	8	72.96	449.03	Y
20	3PL	263.64	8	63.91	448.96	Y
21	3PL	273.84	8	66.46	449.17	Y
22	3PL	344.57	8	84.14	448.96	Y
24	3PL	486.02	8	119.51	448.86	Y
25	3PL	321.51	8	78.38	449.04	Y
27	3PL	285.22	8	69.30	449.03	Y
28	3PL	964.81	8	239.20	448.90	Y
29	3PL	535.46	8	131.86	448.88	Y
30	3PL	390.37	8	95.59	448.02	Y
31	3PL	504.67	8	124.17	449.35	Y
32	3PL	860.32	8	213.08	449.32	Y
33	3PL	663.04	8	163.76	449.32	Y
34	3PL	408.76	8	100.19	449.35	Y
35	3PL	904.32	8	224.08	449.08	Y
36	3PL	824.05	8	204.01	448.91	Y
37	3PL	188.16	8	45.04	449.18	Y
38	3PL	244.84	8	59.21	448.61	Y
39	2PPC	882.89	17	148.50	447.25	Y
40	2PPC	166.93	17	25.71	447.04	Y
41	2PPC	525.21	17	87.16	447.75	Y
42	2PPC	1454.70	17	246.57	448.06	Y
43	2PPC	481.78	17	79.71	448.56	Y
44	2PPC	415.01	17	68.26	448.12	Y
45	2PPC	591.46	26	78.41	447.86	Y

Table O10. Mathematics Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	1451.80	8	360.96	438.39	Y
2	3PL	786.99	8	194.75	437.54	Y
3	3PL	551.15	8	135.79	438.01	Y
4	3PL	197.73	8	47.43	438.15	Y
6	3PL	628.70	8	155.17	438.03	Y
7	3PL	663.46	8	163.86	438.14	Y
9	3PL	352.35	8	86.09	438.00	Y
10	3PL	325.90	8	79.47	437.94	Y
12	3PL	280.76	8	68.19	438.15	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
13	3PL	313.14	8	76.28	437.37	Y
15	3PL	365.36	8	89.34	437.54	Y
16	3PL	454.91	8	111.73	437.99	Y
18	3PL	1441.20	8	358.29	437.87	Y
19	3PL	315.94	8	76.98	437.69	Y
21	3PL	447.06	8	109.77	437.88	Y
22	3PL	344.01	8	84.00	437.90	Y
24	3PL	201.17	8	48.29	437.93	Y
25	3PL	257.45	8	62.36	437.91	Y
26	3PL	311.74	8	75.94	437.63	Y
27	3PL	546.17	8	134.54	437.87	Y
28	3PL	270.47	8	65.62	437.54	Y
29	3PL	305.57	8	74.39	437.33	Y
30	3PL	527.58	8	129.90	437.69	Y
31	3PL	491.38	8	120.85	436.88	Y
32	3PL	1131.50	8	280.87	438.34	Y
33	3PL	331.75	8	80.94	438.21	Y
34	3PL	246.77	8	59.69	438.17	Y
35	3PL	394.34	8	96.58	437.91	Y
36	3PL	306.23	8	74.56	437.77	Y
37	3PL	308.59	8	75.15	437.74	Y
38	3PL	2304.20	8	574.04	437.02	N
39	2PPC	288.62	17	46.58	436.56	Y
40	2PPC	178.00	17	27.61	435.29	Y
41	2PPC	87.78	17	12.14	434.96	Y
42	2PPC	142.33	17	21.49	434.98	Y
43	2PPC	55.80	17	6.65	434.03	Y
44	2PPC	322.48	17	52.39	436.11	Y
45	2PPC	3796.50	17	648.18	435.09	N
46	2PPC	202.18	26	24.43	434.69	Y

Table O11. Mathematics Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	980.31	8	243.08	404.50	Y
2	3PL	323.67	8	78.92	403.51	Y
3	3PL	410.78	8	100.70	404.37	Y
4	3PL	394.66	8	96.66	404.29	Y
6	3PL	198.50	8	47.62	404.23	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
7	3PL	119.63	8	27.91	404.06	Y
8	3PL	147.23	8	34.81	403.43	Y
10	3PL	491.63	8	120.91	404.32	Y
11	3PL	306.80	8	74.70	404.16	Y
13	3PL	217.16	8	52.29	403.86	Y
14	3PL	155.19	8	36.80	404.24	Y
16	3PL	354.65	8	86.66	403.88	Y
17	3PL	340.16	8	83.04	404.19	Y
19	3PL	176.41	8	42.10	404.21	Y
20	3PL	184.32	8	44.08	403.90	Y
22	3PL	541.33	8	133.33	404.13	Y
23	3PL	459.66	8	112.92	403.89	Y
24	3PL	1073.80	8	266.45	404.20	Y
26	3PL	233.01	8	56.25	403.82	Y
27	3PL	228.21	8	55.05	404.15	Y
28	3PL	153.54	8	36.39	403.87	Y
29	3PL	450.78	8	110.70	403.91	Y
30	3PL	268.39	8	65.10	403.84	Y
31	3PL	294.94	8	71.73	403.95	Y
32	3PL	123.52	8	28.88	403.91	Y
33	3PL	129.40	8	30.35	402.94	Y
34	3PL	122.96	8	28.74	404.36	Y
35	3PL	394.86	8	96.72	404.44	Y
36	3PL	254.10	8	61.53	404.23	Y
37	3PL	525.70	8	129.43	404.31	Y
38	3PL	240.45	8	58.11	404.40	Y
39	3PL	426.22	8	104.56	404.09	Y
40	3PL	262.17	8	63.54	403.81	Y
41	2PPC	59.32	17	7.26	402.20	Y
42	2PPC	436.57	17	71.96	401.51	Y
43	2PPC	386.26	17	63.33	400.33	Y
44	2PPC	303.20	17	49.08	401.85	Y
45	2PPC	384.80	17	63.08	402.02	Y
46	2PPC	250.85	17	40.10	398.76	Y
47	2PPC	660.25	17	110.32	401.44	Y
48	2PPC	219.22	26	26.80	401.07	Y

Table O12. Mathematics Grade 8 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	414.37	8	101.59	288.89	Y
2	3PL	263.40	8	63.85	288.91	Y
3	3PL	134.27	8	31.57	288.73	Y
4	3PL	369.01	8	90.25	288.52	Y
6	3PL	210.43	8	50.61	288.69	Y
7	3PL	143.06	8	33.77	288.12	Y
8	3PL	143.58	8	33.90	288.45	Y
10	3PL	223.62	8	53.90	288.74	Y
11	3PL	1258.80	8	312.71	288.83	N
12	3PL	419.43	8	102.86	288.82	Y
14	3PL	125.48	8	29.37	288.55	Y
15	3PL	228.45	8	55.11	288.79	Y
16	3PL	629.24	8	155.31	288.73	Y
18	3PL	185.36	8	44.34	288.66	Y
19	3PL	107.19	8	24.80	288.45	Y
20	3PL	213.42	8	51.35	288.64	Y
22	3PL	295.23	8	71.81	288.57	Y
23	3PL	113.38	8	26.35	288.64	Y
24	3PL	270.03	8	65.51	288.47	Y
26	3PL	181.43	8	43.36	288.37	Y
27	3PL	152.86	8	36.22	288.63	Y
28	3PL	190.12	8	45.53	288.43	Y
30	3PL	168.74	8	40.19	288.57	Y
31	3PL	211.31	8	50.83	288.52	Y
32	3PL	77.77	8	17.44	288.25	Y
33	3PL	106.46	8	24.61	287.66	Y
34	3PL	262.49	8	63.62	288.94	Y
35	3PL	115.58	8	26.90	288.87	Y
36	3PL	214.63	8	51.66	288.86	Y
37	3PL	78.14	8	17.54	288.86	Y
38	3PL	171.32	8	40.83	288.85	Y
39	3PL	238.35	8	57.59	288.71	Y
40	3PL	149.91	8	35.48	288.71	Y
41	2PPC	359.18	17	58.68	285.55	Y
42	2PPC	284.02	17	45.79	280.12	Y
43	2PPC	202.57	17	31.83	281.84	Y
44	2PPC	143.54	17	21.70	284.74	Y
45	2PPC	67.52	17	8.66	284.26	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
46	2PPC	137.27	17	20.63	284.24	Y
47	2PPC	66.71	17	8.53	278.65	Y
48	2PPC	150.88	26	17.32	283.10	Y

Table O13. ELA Grade 3 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	1.169	-1.343	0.391		
2	1	0.773	0.422	0.273		
3	1	0.758	-0.596	0.276		
4	1	0.571	-0.599	0.237		
5	1	0.479	-0.273	0.031		
6	1	0.430	0.711	0.148		
7	1	0.436	-0.097	0.092		
8	1	0.920	0.562	0.159		
9	1	1.128	-0.169	0.203		
10	1	1.226	0.556	0.152		
11	1	0.945	0.919	0.212		
12	1	1.118	0.228	0.209		
13	1	0.506	0.472	0.139		
14	1	0.746	-0.204	0.182		
15	1	0.838	0.672	0.169		
16	1	1.066	0.551	0.218		
17	1	1.343	-0.362	0.259		
18	1	0.637	0.595	0.211		
19	2	0.876	-0.646	1.207		
20	2	0.777	-0.737	1.319		
21	2	0.853	0.335	1.049		
22	2	0.969	-0.384	0.896		
23	2	1.100	-0.592	1.078		
24	2	0.918	0.157	1.408		
25	4	0.835	0.588	1.661	1.185	2.172

Table O14. ELA Grade 4 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.814	-0.263	0.134		
2	1	0.593	0.136	0.198		
3	1	1.052	-0.733	0.168		
4	1	0.626	-0.878	0.105		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
5	1	0.749	-1.357	0.103		
6	1	0.291	0.126	0.033		
7	1	0.827	0.067	0.204		
8	1	0.673	0.629	0.192		
9	1	0.757	0.428	0.213		
10	1	0.797	-0.419	0.263		
11	1	0.364	0.103	0.075		
12	1	0.789	-0.394	0.192		
13	1	0.586	-0.185	0.045		
14	1	0.607	0.934	0.285		
15	1	0.686	-0.443	0.104		
16	1	0.447	-0.317	0.025		
17	1	0.448	0.509	0.146		
18	1	0.618	-0.713	0.089		
19	2	0.915	-0.449	1.067		
20	2	0.886	-0.358	1.007		
21	2	0.648	0.540	0.769		
22	2	0.913	-0.283	0.965		
23	2	0.988	-0.302	1.033		
24	2	1.241	-0.552	0.961		
25	4	0.947	0.128	1.443	0.554	1.688

Table O15. ELA Grade 5 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.686	-2.650	0.091		
2	1	0.372	-0.036	0.105		
3	1	0.848	-0.095	0.172		
4	1	0.934	-1.300	0.144		
5	1	0.937	0.224	0.313		
6	1	1.191	-1.516	0.204		
7	1	0.543	0.149	0.140		
8	1	1.268	-0.453	0.153		
9	1	0.610	0.803	0.137		
10	1	0.951	0.589	0.208		
11	1	0.729	-1.241	0.110		
12	1	1.397	1.078	0.301		
13	1	0.525	-0.350	0.043		
14	1	0.684	1.119	0.230		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
15	1	1.084	-0.885	0.209		
16	1	0.616	1.465	0.162		
17	1	0.451	1.766	0.206		
18	1	0.760	-0.327	0.193		
19	1	1.509	-0.317	0.252		
20	1	0.733	0.042	0.151		
21	1	1.000	-0.613	0.172		
22	1	0.661	-0.222	0.206		
23	1	1.363	-0.418	0.240		
24	1	0.698	2.120	0.245		
25	1	0.972	0.753	0.192		
26	1	0.523	1.092	0.207		
27	1	1.012	0.535	0.251		
28	1	0.957	-0.605	0.266		
29	2	0.560	-1.284	1.385		
30	2	0.641	-1.338	0.895		
31	2	0.640	-0.861	1.443		
32	2	0.801	-1.064	1.300		
33	2	0.703	-0.780	1.457		
34	2	0.537	-0.546	0.972		
35	4	0.679	0.085	1.724	0.650	1.813

Table O16. ELA Grade 6 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.567	-1.998	0.054		
2	1	0.603	-0.689	0.048		
3	1	0.618	-2.125	0.061		
4	1	0.691	-0.397	0.214		
5	1	0.637	-0.482	0.226		
6	1	0.695	-1.590	0.266		
7	1	0.488	-0.077	0.092		
8	1	0.823	0.423	0.142		
9	1	0.789	0.639	0.129		
10	1	1.319	-0.281	0.238		
11	1	0.352	-1.051	0.017		
12	1	0.919	0.337	0.258		
13	1	0.203	0.780	0.025		
14	1	0.555	1.676	0.195		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
15	1	1.203	-0.393	0.152		
16	1	0.567	-0.878	0.092		
17	1	1.605	-0.398	0.254		
18	1	0.719	-0.858	0.105		
19	1	0.702	-0.640	0.095		
20	1	0.519	-1.009	0.022		
21	1	0.354	2.237	0.208		
22	1	0.731	1.283	0.222		
23	1	0.752	-1.188	0.081		
24	1	0.722	0.456	0.197		
25	1	0.881	0.132	0.205		
26	1	0.893	-0.856	0.149		
27	1	1.117	0.061	0.199		
28	1	0.909	-0.422	0.191		
29	2	0.719	-0.554	0.871		
30	2	0.850	-1.298	0.959		
31	2	0.701	-0.861	1.030		
32	2	0.859	-1.150	0.853		
33	2	0.888	-0.221	0.395		
34	2	0.944	-0.739	0.802		
35	4	0.881	-0.216	1.562	0.293	1.382

Table O17. ELA Grade 7 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.512	-2.390	0.045		
2	1	0.481	-1.164	0.184		
3	1	0.737	-0.646	0.206		
4	1	0.256	1.982	0.369		
5	1	0.757	-0.750	0.164		
6	1	0.847	-0.332	0.219		
7	1	0.343	-2.069	0.025		
8	1	0.407	0.263	0.042		
9	1	0.715	1.219	0.256		
10	1	0.912	0.572	0.282		
11	1	0.431	-0.739	0.041		
12	1	0.351	-1.192	0.023		
13	1	0.627	1.037	0.206		
14	1	0.745	1.261	0.241		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
15	1	1.201	-0.820	0.273		
16	1	0.692	-0.839	0.194		
17	1	1.115	0.298	0.300		
18	1	0.538	-0.246	0.106		
19	1	0.763	0.189	0.257		
20	1	1.471	0.029	0.196		
21	1	0.605	0.708	0.228		
22	1	0.886	0.065	0.146		
23	1	1.087	0.147	0.254		
24	1	1.005	0.542	0.213		
25	1	1.486	0.393	0.191		
26	1	1.094	0.704	0.160		
27	1	1.500	1.104	0.225		
28	1	0.712	0.137	0.171		
29	2	0.979	-1.327	0.852		
30	2	1.111	-1.363	0.832		
31	2	0.986	-0.862	0.673		
32	2	0.978	-0.980	0.669		
33	2	1.043	-0.675	0.641		
34	2	0.728	-0.761	0.802		
35	2	1.038	-0.601	0.582		
36	4	0.861	-0.243	1.439	0.307	1.215

Table O18. ELA Grade 8 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.372	-3.557	0.048		
2	1	0.754	-1.242	0.145		
3	1	0.671	0.802	0.129		
4	1	0.768	-1.548	0.131		
5	1	0.396	-0.433	0.068		
6	1	0.566	-1.475	0.032		
7	1	0.817	-0.389	0.189		
8	1	0.650	-2.069	0.027		
9	1	0.869	-0.247	0.199		
10	1	0.424	-1.279	0.014		
11	1	0.989	-0.909	0.149		
12	1	0.874	0.880	0.261		
13	1	1.038	-1.636	0.096		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
14	1	0.525	-0.717	0.064		
15	1	0.486	-1.910	0.025		
16	1	0.640	-0.277	0.137		
17	1	0.711	-1.562	0.089		
18	1	0.286	0.150	0.029		
19	1	0.710	-0.911	0.174		
20	1	0.329	-1.221	0.017		
21	1	0.251	0.026	0.021		
22	1	1.324	-0.286	0.279		
23	1	0.883	0.231	0.256		
24	1	1.017	0.259	0.152		
25	1	1.032	0.361	0.268		
26	1	0.709	0.458	0.188		
27	1	1.073	0.917	0.204		
28	1	0.940	-0.527	0.229		
29	2	0.863	-1.445	0.996		
30	2	0.918	-1.373	1.110		
31	2	0.624	-0.213	0.694		
32	2	0.954	-0.847	0.906		
33	2	1.018	-0.604	0.812		
34	2	0.870	-1.007	0.805		
35	2	0.825	-0.949	0.701		
36	4	0.936	-0.358	1.310	0.022	1.116

Table O19. Mathematics Grade 3 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.786	-1.118	0.090	
2	1	1.029	-1.274	0.161	
3	1	0.905	-0.922	0.079	
4	1	0.895	-0.064	0.095	
5	1	0.917	-1.435	0.264	
6	1	0.612	0.235	0.291	
7	1	1.268	0.118	0.159	
8	1	1.270	-0.673	0.115	
9	1	1.181	-1.488	0.039	
10	1	1.306	-0.444	0.132	
11	1	0.841	0.091	0.147	
12	1	0.656	-0.656	0.054	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
13	1	0.899	-0.608	0.114	
14	1	1.442	-0.680	0.096	
15	1	0.842	1.106	0.187	
16	1	1.279	-0.058	0.193	
17	1	0.878	0.407	0.053	
18	1	1.398	1.218	0.170	
19	1	0.976	-0.154	0.245	
20	1	0.832	-1.761	0.044	
21	1	1.370	-0.551	0.096	
22	1	0.634	0.304	0.126	
23	1	1.135	0.180	0.261	
24	1	1.007	-0.727	0.227	
25	1	1.635	0.777	0.133	
26	1	0.943	-2.271	0.061	
27	1	1.532	0.938	0.130	
28	2	0.714	-0.199	-0.551	
29	2	0.518	-1.387	-1.003	
30	2	0.662	0.011	0.581	
31	2	0.752	-0.166	-0.336	
32	2	0.967	-0.407	-0.039	
33	2	0.685	-0.134	-0.463	
34	3	0.875	0.358	0.372	-0.139

Table O20. Mathematics Grade 4 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	1.036	-1.432	0.163	
2	1	1.565	-0.816	0.101	
3	1	1.184	-0.576	0.164	
4	1	0.559	0.204	0.144	
5	1	0.967	-0.510	0.174	
6	1	0.982	-0.841	0.077	
7	1	1.262	0.020	0.113	
8	1	1.478	-0.358	0.088	
9	1	1.178	0.159	0.171	
10	1	1.156	0.795	0.128	
11	1	1.032	-1.123	0.078	
12	1	0.754	0.201	0.055	
13	1	0.880	-0.346	0.243	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
14	1	1.134	-1.758	0.015	
15	1	0.921	-0.209	0.233	
16	1	1.150	0.207	0.144	
17	1	1.057	-0.130	0.175	
18	1	1.252	-0.239	0.247	
19	1	1.196	-0.012	0.129	
20	1	1.589	0.247	0.164	
21	1	1.347	-0.185	0.280	
22	1	1.555	0.249	0.173	
23	1	1.592	0.057	0.241	
24	1	0.821	-0.904	0.319	
25	1	0.701	0.139	0.183	
26	1	0.985	-1.583	0.024	
27	1	1.309	-0.213	0.193	
28	1	0.916	0.458	0.231	
29	1	0.919	-0.315	0.151	
30	1	1.027	0.182	0.121	
31	1	0.911	0.107	0.268	
32	2	0.591	-0.380	0.064	
33	2	1.032	-0.028	0.335	
34	2	0.465	0.135	1.286	
35	2	0.907	0.234	0.485	
36	2	0.780	-0.565	0.041	
37	2	1.054	0.553	0.402	
38	3	0.587	-0.307	-1.472	-0.943

Table O21. Mathematics Grade 5 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.981	-0.869	0.041	
2	1	1.440	-0.407	0.192	
3	1	0.998	-1.358	0.397	
4	1	0.746	-0.313	0.078	
5	1	1.242	-1.582	0.021	
6	1	1.334	0.330	0.151	
7	1	0.973	-0.287	0.200	
8	1	1.808	0.146	0.158	
9	1	0.764	0.256	0.218	
10	1	0.778	1.102	0.154	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
11	1	1.538	0.189	0.263	
12	1	0.986	-0.879	0.028	
13	1	1.437	0.425	0.159	
14	1	0.956	0.049	0.165	
15	1	0.822	-0.233	0.110	
16	1	1.282	-0.139	0.229	
17	1	1.060	0.420	0.149	
18	1	1.373	-0.361	0.120	
19	1	1.235	0.272	0.229	
20	1	1.166	-0.546	0.219	
21	1	1.236	-1.401	0.064	
22	1	0.721	-0.858	0.107	
23	1	1.071	-0.524	0.138	
24	1	1.253	0.547	0.116	
25	1	1.275	-0.930	0.050	
26	1	1.430	-1.179	0.201	
27	1	1.234	-0.700	0.152	
28	1	1.669	0.442	0.123	
29	1	1.021	-0.078	0.123	
30	1	0.689	0.867	0.256	
31	1	1.231	0.056	0.270	
32	2	0.732	0.594	0.022	
33	2	0.740	0.168	-0.285	
34	2	0.825	0.627	0.127	
35	2	0.685	0.646	1.160	
36	2	1.050	0.199	0.118	
37	2	0.771	0.645	0.422	
38	3	0.569	0.399	-0.138	-0.065

Table O22. Mathematics Grade 6 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.813	-1.401	0.022	
2	1	1.329	-0.364	0.039	
3	1	0.732	-0.311	0.120	
4	1	1.024	0.237	0.257	
5	1	1.611	-0.460	0.145	
6	1	0.800	-0.687	0.122	
7	1	1.079	-0.130	0.155	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
8	1	1.280	-0.431	0.273	
9	1	1.590	0.585	0.146	
10	1	1.621	0.402	0.074	
11	1	1.116	0.636	0.148	
12	1	1.349	1.108	0.115	
13	1	1.145	-0.332	0.032	
14	1	1.717	-0.196	0.223	
15	1	1.876	0.325	0.107	
16	1	1.450	0.814	0.199	
17	1	1.084	0.192	0.239	
18	1	1.103	-0.245	0.275	
19	1	0.719	0.252	0.188	
20	1	1.078	0.193	0.080	
21	1	1.629	0.145	0.165	
22	1	1.147	0.312	0.234	
23	1	1.192	-0.213	0.130	
24	1	0.987	-0.713	0.256	
25	1	1.042	-1.153	0.098	
26	1	0.846	-0.473	0.292	
27	1	0.875	-0.224	0.270	
28	1	1.212	0.682	0.257	
29	1	0.942	0.106	0.264	
30	1	1.334	1.304	0.149	
31	1	0.928	-1.251	0.011	
32	2	0.842	-0.129	0.093	
33	2	0.735	0.531	0.672	
34	2	0.806	0.675	-0.705	
35	2	1.254	0.873	0.203	
36	2	0.685	0.585	-0.117	
37	2	0.945	0.777	0.482	
38	2	0.598	1.385	0.350	
39	3	0.845	1.072	0.149	1.301

Table O23. Mathematics Grade 7 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.961	-1.174	0.174	
2	1	1.161	0.746	0.199	
3	1	1.187	-0.547	0.247	
4	1	1.031	0.049	0.131	
5	1	1.633	0.285	0.164	
6	1	1.414	-0.020	0.334	
7	1	1.558	0.189	0.230	
8	1	1.436	0.581	0.363	
9	1	1.441	0.468	0.264	
10	1	1.931	0.175	0.183	
11	1	1.227	0.244	0.266	
12	1	1.197	0.576	0.303	
13	1	1.278	-0.376	0.236	
14	1	1.085	0.143	0.200	
15	1	1.876	0.202	0.218	
16	1	1.494	-0.143	0.090	
17	1	1.679	1.210	0.240	
18	1	1.352	-1.153	0.297	
20	1	1.260	0.519	0.218	
21	1	0.945	-0.240	0.266	
22	1	1.141	0.347	0.276	
23	1	1.830	0.925	0.149	
24	1	1.344	0.261	0.276	
25	1	1.159	0.136	0.286	
26	1	1.146	0.239	0.246	
27	1	0.838	0.386	0.221	
28	1	1.001	0.813	0.244	
29	1	1.434	0.756	0.171	
30	1	1.510	0.305	0.210	
31	1	1.034	0.278	0.320	
32	1	1.088	0.514	0.212	
33	1	2.254	0.804	0.205	
34	1	1.375	0.430	0.222	
35	2	0.952	0.275	-0.439	
36	2	0.745	-0.857	0.048	
37	2	1.214	0.270	0.397	
38	2	0.880	-0.267	-0.976	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
39	2	0.989	-0.496	0.059	
40	2	1.502	0.569	0.426	
41	2	0.881	-0.145	-0.141	
42	3	0.753	0.157	0.299	0.136

Table O24. Mathematics Grade 8 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.624	-0.357	0.340	
2	1	1.120	-0.036	0.250	
3	1	1.303	0.304	0.457	
4	1	1.220	-0.200	0.235	
5	1	1.607	0.434	0.143	
6	1	1.537	1.101	0.253	
7	1	1.599	0.570	0.238	
8	1	1.037	0.233	0.191	
9	1	0.927	-0.934	0.191	
10	1	1.101	-0.177	0.186	
11	1	1.044	0.786	0.218	
12	1	0.914	0.359	0.136	
13	1	1.006	-0.321	0.249	
14	1	1.097	0.629	0.246	
15	1	1.287	0.470	0.268	
16	1	1.224	0.075	0.283	
17	1	1.323	0.885	0.162	
18	1	1.035	1.048	0.194	
19	1	1.094	0.204	0.279	
20	1	1.668	1.411	0.255	
21	1	1.129	0.429	0.223	
22	1	1.750	0.851	0.221	
23	1	1.473	0.496	0.274	
24	1	0.714	0.064	0.260	
25	1	1.034	0.841	0.285	
26	1	1.202	0.870	0.296	
27	1	1.524	-0.015	0.210	
28	1	1.529	0.577	0.356	
29	1	1.400	0.628	0.207	
30	1	0.778	1.015	0.225	
31	1	1.298	0.319	0.196	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
32	1	1.220	0.256	0.129	
33	1	1.412	1.120	0.172	
34	2	0.842	0.279	0.192	
35	2	0.920	0.978	-0.463	
36	2	0.861	0.332	0.168	
37	2	0.950	-0.226	0.070	
38	2	0.715	0.072	-0.257	
39	2	0.752	-0.109	-0.293	
40	2	0.893	0.750	-0.603	
41	3	0.398	0.436	-2.086	-1.242

Appendix P: Derivation and Estimation of Classification Consistency and Accuracy

Classification Consistency

Assume that θ is a single latent trait measured by a test and denote Φ as a latent random variable. When a test X consists of *K* items and its maximum number correct score is *N*, the marginal probability of the number correct (NC) score *x* is

$$P(X = x) = \int P(X = x \mid \Phi = \theta) g(\theta) d\theta, \quad x = 0,1,..., N$$

where

$g(\theta)$ is the density of θ.

In this report, the marginal distribution $P(X = x)$ is denoted as $f(x)$, and the conditional error distribution $P(X = x \mid \Phi = \theta)$ is denoted as $f(x \mid \theta)$. It is assumed that examinees are classified into one of H mutually exclusive categories on the basis of predetermined *H* - 1 observed score cutoffs, C_1, C_2, …, C_{H-1}. Let L_h represent the hth category into which examinees with $C_{h-1} \le X < C_h$ are classified. $C_0 = 0$ and C_H = the maximum number-correct score plus one. Then, the conditional and marginal probabilities of each category classification are as follows:

$$P(X \in L_h \mid \theta) = \sum_{x=C_{h-1}}^{C_h - 1} f(x \mid \theta), \; h = 1, 2, \ldots, \mathrm{H}$$

$$P(X \in L_h) = \int \sum_{x=C_{h-1}}^{C_h - 1} f(x \mid \theta) g(\theta) d\theta, \; h = 1, 2, \ldots, \mathrm{H}$$

Because obtaining test scores from two independent administrations of New York State tests was not feasible due to item release after each OP administration, a psychometric model was used to obtain the estimated classification consistency indices using test scores from a single administration. Based on the psychometric model, a symmetric *H*-by-*H* contingency table can be constructed. The elements of the *H*-by-*H* contingency table consist of the joint probabilities of the row and column observed category classifications.

That two administrations are independent implies that if X_1 and X_2 represent the raw score random variables on the two administrations, then, conditioned on θ, X_1 and X_2 are independent and identically distributed. Consequently, the conditional bivariate distribution of X_1 and X_2 is

$$f(x_1, \; x_2 \mid \theta) = f(x_1 \mid \theta) f(x_2 \mid \theta)$$

The marginal bivariate distribution of X_1 and X_2 can be expressed as follows:

$$f(x_1, \; x_2) = \int f(x_1, x_2 \mid \theta) f(\theta) d\theta$$

Consistent classification means that both X_1 and X_2 fall in the same category. The conditional probability of falling in the same category on the two administrations is

$$P(X_1 \in L_h,\ X_2 \in L_h \mid \theta) = \left[\sum_{x_1 = C_{h-1}}^{C_{h-1}} f(x_1 \mid \theta) \right]^2,\ h = 1, 2, \ldots, \mathrm{H}$$

The agreement index P, conditional on theta, is obtained by

$$P(\theta) = \sum_{h=1}^{H} P(X_1 \in L_h,\ X_2 \in L_h \mid \theta)$$

The agreement index (classification consistency) can be computed as

$$P = \int P(\theta) g(\theta) d(\theta)$$

The probability of consistent classification by chance, P_C, is the sum of squared marginal probabilities of each category classification.

$$P_C = \sum_{h=1}^{H} P(X_1 \in L_h) P(X_2 \in L_h) = \sum_{h=1}^{H} \left[P(X_1 \in L_h) \right]^2$$

Then, Kappa (Cohen, 1960) is

$$k = \frac{P - P_C}{1 - P_C}$$

Classification Accuracy

Let Γ_w denote true category. When an examinee has an observed score, $x \in L_h$ (h =1, 2,…, H), and a latent score, $\theta \in \Gamma_w$ (w =1, 2,…, H), an accurate classification is made when $h = w$. The conditional probability of accurate classification is

$$\gamma(\theta) = P(X \in L_w \mid \theta),$$

where

w is the category such that $\theta \in \Gamma_w$

Lee (2008) thoroughly discusses this IRT method for estimating decision indices, including the computational method used to estimate the results when integrating across the latent variable, θ.

Estimating Classification Indices

The classification consistency and accuracy estimates were obtained using an open-source software program, IRT-CLASS v2.0 (Lee & Kolen, 2006). Below is a brief description of the files that are used and their purpose. (See the IRT-CLASS v2.0 manual for complete instructions.)

Files needed:

- Raw-to-Scale score conversion file
    a. Contains the raw-to-scale score conversions
    b. This is used to provide both raw and scale score classification estimates, which is useful when the raw-to-scale score transformation is not one-to-one.

- Cut score file
    a. Contains the cut scores to be used
    b. Results are provided for all cut scores simultaneously (all performance levels), as well as the estimates based on each of the cut scores separately (Level III only).

- Item parameter file
    a. This contains the IRT model used and item parameter estimates.
    b. This information is used when calculating the classification indices.

- Theta file
    a. Contains the theta distribution in terms of quadrature points
    b. The theta and the item parameter files are used to solve the integrals mentioned above.

- Control card
    a. This is used to run the program.
    b. Identifies the names of the four files above and gives a name to the output file

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Tables Q1–Q12 show the PBT raw-to-scale score conversion tables, while Tables Q13-Q24 show the CBT raw-to-scale score conversion tables. Tables Q25–Q36 show the scale score distributions that include all students with valid scores, by frequency (n-count), percent, cumulative frequency, and cumulative percent.

Table Q1. PBT ELA Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	530	26	18	600	6
1	535	22	19	602	6
2	540	19	20	604	6
3	544	16	21	607	6
4	549	14	22	609	6
5	553	12	23	612	6
6	558	10	24	614	6
7	564	9	25	617	6
8	569	8	26	619	6
9	573	8	27	622	6
10	577	7	28	626	7
11	580	7	29	629	7
12	583	7	30	634	8
13	586	7	31	639	9
14	589	6	32	646	11
15	592	6	33	650	12
16	595	6	34	655	14
17	597	6			

Table Q2. PBT ELA Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	532	26	18	596	6
1	536	22	19	598	6
2	541	18	20	601	6
3	546	15	21	603	6
4	550	13	22	606	6
5	555	11	23	608	6
6	561	9	24	611	6
7	565	8	25	614	6

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
8	569	7	26	616	6
9	572	7	27	619	7
10	575	7	28	623	7
11	578	6	29	627	8
12	581	6	30	631	8
13	584	6	31	637	10
14	586	6	32	645	12
15	589	6	33	649	14
16	591	6	34	654	16
17	594	6			

Table Q3. PBT ELA Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	509	29	23	592	6
1	513	26	24	594	6
2	518	23	25	596	6
3	523	20	26	598	6
4	527	18	27	600	6
5	532	16	28	602	6
6	536	14	29	604	6
7	541	13	30	606	6
8	545	12	31	609	6
9	550	11	32	611	6
10	554	10	33	614	6
11	559	9	34	616	6
12	563	9	35	619	7
13	566	8	36	622	7
14	570	8	37	625	7
15	573	7	38	629	8
16	575	7	39	633	8
17	578	7	40	637	9
18	581	6	41	644	11
19	583	6	42	652	13
20	585	6	43	657	15
21	587	6	44	661	16
22	590	6			

Table Q4. PBT ELA Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	514	31	23	590	5
1	518	26	24	592	5
2	523	22	25	593	5
3	528	19	26	595	5
4	532	17	27	597	5
5	537	14	28	599	5
6	541	13	29	601	5
7	546	11	30	602	5
8	550	10	31	604	5
9	555	9	32	607	5
10	559	8	33	609	6
11	563	8	34	611	6
12	566	7	35	614	6
13	569	7	36	616	6
14	571	7	37	619	7
15	574	6	38	623	7
16	576	6	39	627	8
17	578	6	40	632	9
18	580	6	41	638	11
19	582	6	42	648	14
20	584	5	43	652	16
21	586	5	44	657	18
22	588	5			

Table Q5. PBT ELA Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	511	35	24	591	5
1	516	30	25	593	5
2	520	26	26	594	5
3	525	22	27	596	5
4	529	19	28	598	5
5	534	16	29	600	5
6	539	14	30	601	5
7	543	12	31	603	5
8	548	10	32	605	5

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
9	552	9	33	607	5
10	557	8	34	609	5
11	561	7	35	611	5
12	564	7	36	613	5
13	567	6	37	615	5
14	570	6	38	618	6
15	573	6	39	620	6
16	575	6	40	623	6
17	577	6	41	627	7
18	579	6	42	631	8
19	581	5	43	637	9
20	583	5	44	644	12
21	585	5	45	649	14
22	587	5	46	654	16
23	589	5			

Table Q6. PBT ELA Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	507	30	24	586	5
1	511	26	25	588	5
2	516	23	26	590	5
3	521	20	27	591	5
4	525	17	28	593	5
5	530	15	29	595	5
6	534	13	30	597	5
7	539	11	31	599	5
8	543	10	32	601	5
9	548	9	33	603	5
10	553	8	34	605	5
11	557	7	35	607	6
12	560	7	36	609	6
13	563	6	37	612	6
14	566	6	38	614	6
15	568	6	39	617	6
16	570	6	40	620	7
17	572	6	41	624	7

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
18	574	6	42	628	8
19	576	6	43	634	10
20	578	5	44	642	12
21	580	5	45	646	14
22	582	5	46	651	17
23	584	5			

Table Q7. PBT Mathematics Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	530	26	18	600	6
1	535	22	19	602	6
2	540	19	20	604	6
3	544	16	21	607	6
4	549	14	22	609	6
5	553	12	23	612	6
6	558	10	24	614	6
7	564	9	25	617	6
8	569	8	26	619	6
9	573	8	27	622	6
10	577	7	28	626	7
11	580	7	29	629	7
12	583	7	30	634	8
13	586	7	31	639	9
14	589	6	32	646	11
15	592	6	33	650	12
16	595	6	34	655	14
17	597	6			

Table Q8. PBT Mathematics Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	532	26	18	596	6
1	536	22	19	598	6
2	541	18	20	601	6
3	546	15	21	603	6
4	550	13	22	606	6

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
5	555	11	23	608	6
6	561	9	24	611	6
7	565	8	25	614	6
8	569	7	26	616	6
9	572	7	27	619	7
10	575	7	28	623	7
11	578	6	29	627	8
12	581	6	30	631	8
13	584	6	31	637	10
14	586	6	32	645	12
15	589	6	33	649	14
16	591	6	34	654	16
17	594	6			

Table Q9. PBT Mathematics Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	509	29	23	592	6
1	513	26	24	594	6
2	518	23	25	596	6
3	523	20	26	598	6
4	527	18	27	600	6
5	532	16	28	602	6
6	536	14	29	604	6
7	541	13	30	606	6
8	545	12	31	609	6
9	550	11	32	611	6
10	554	10	33	614	6
11	559	9	34	616	6
12	563	9	35	619	7
13	566	8	36	622	7
14	570	8	37	625	7
15	573	7	38	629	8
16	575	7	39	633	8
17	578	7	40	637	9
18	581	6	41	644	11
19	583	6	42	652	13

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
20	585	6	43	657	15
21	587	6	44	661	16
22	590	6			

Table Q10. PBT Mathematics Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	514	31	23	590	5
1	518	26	24	592	5
2	523	22	25	593	5
3	528	19	26	595	5
4	532	17	27	597	5
5	537	14	28	599	5
6	541	13	29	601	5
7	546	11	30	602	5
8	550	10	31	604	5
9	555	9	32	607	5
10	559	8	33	609	6
11	563	8	34	611	6
12	566	7	35	614	6
13	569	7	36	616	6
14	571	7	37	619	7
15	574	6	38	623	7
16	576	6	39	627	8
17	578	6	40	632	9
18	580	6	41	638	11
19	582	6	42	648	14
20	584	5	43	652	16
21	586	5	44	657	18
22	588	5			

Table Q11. PBT Mathematics Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	511	35	24	591	5
1	516	30	25	593	5
2	520	26	26	594	5

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
3	525	22	27	596	5
4	529	19	28	598	5
5	534	16	29	600	5
6	539	14	30	601	5
7	543	12	31	603	5
8	548	10	32	605	5
9	552	9	33	607	5
10	557	8	34	609	5
11	561	7	35	611	5
12	564	7	36	613	5
13	567	6	37	615	5
14	570	6	38	618	6
15	573	6	39	620	6
16	575	6	40	623	6
17	577	6	41	627	7
18	579	6	42	631	8
19	581	5	43	637	9
20	583	5	44	644	12
21	585	5	45	649	14
22	587	5	46	654	16
23	589	5			

Table Q12. PBT Mathematics Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	507	30	24	586	5
1	511	26	25	588	5
2	516	23	26	590	5
3	521	20	27	591	5
4	525	17	28	593	5
5	530	15	29	595	5
6	534	13	30	597	5
7	539	11	31	599	5
8	543	10	32	601	5
9	548	9	33	603	5
10	553	8	34	605	5
11	557	7	35	607	6

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
12	560	7	36	609	6
13	563	6	37	612	6
14	566	6	38	614	6
15	568	6	39	617	6
16	570	6	40	620	7
17	572	6	41	624	7
18	574	6	42	628	8
19	576	6	43	634	10
20	578	5	44	642	12
21	580	5	45	646	14
22	582	5	46	651	17
23	584	5			

Table Q13. CBT ELA Grade 3 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	531	25	18	601	6
1	536	21	19	603	6
2	541	18	20	605	6
3	545	15	21	608	6
4	550	13	22	610	6
5	554	12	23	613	6
6	559	10	24	615	6
7	565	9	25	618	6
8	570	8	26	620	6
9	574	8	27	623	6
10	578	7	28	627	7
11	581	7	29	630	7
12	584	7	30	635	8
13	587	7	31	640	9
14	590	6	32	647	11
15	593	6	33	651	12
16	596	6	34	655	14
17	598	6			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q14. CBT ELA Grade 4 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	533	25	18	597	6
1	537	21	19	599	6
2	542	18	20	602	6
3	547	14	21	604	6
4	551	12	22	607	6
5	556	10	23	609	6
6	562	9	24	612	6
7	566	8	25	615	6
8	570	7	26	617	6
9	573	7	27	620	7
10	576	7	28	624	7
11	579	6	29	628	8
12	582	6	30	632	9
13	585	6	31	638	10
14	587	6	32	646	12
15	590	6	33	650	14
16	592	6	34	654	16
17	595	6			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q15. CBT ELA Grade 5 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	510	28	23	593	6
1	514	25	24	595	6
2	519	22	25	597	6
3	524	19	26	599	6
4	528	17	27	601	6
5	533	15	28	603	6
6	537	14	29	605	6
7	542	12	30	607	6
8	546	12	31	610	6
9	551	11	32	612	6
10	555	10	33	615	6
11	560	9	34	617	6
12	564	8	35	620	7

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
13	567	8	36	623	7
14	571	7	37	626	7
15	574	7	38	630	8
16	576	7	39	634	9
17	579	7	40	638	9
18	582	6	41	645	11
19	584	6	42	653	13
20	586	6	43	658	15
21	588	6	44	661	16
22	591	6			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q16. CBT ELA Grade 6 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	515	29	23	591	5
1	519	26	24	593	5
2	524	22	25	594	5
3	529	18	26	596	5
4	533	16	27	598	5
5	538	14	28	600	5
6	542	12	29	602	5
7	547	11	30	603	5
8	551	10	31	605	5
9	556	9	32	608	6
10	560	8	33	610	6
11	564	7	34	612	6
12	567	7	35	615	6
13	570	7	36	617	6
14	572	6	37	620	7
15	575	6	38	624	8
16	577	6	39	628	8
17	579	6	40	633	9
18	581	6	41	639	11
19	583	6	42	649	15
20	585	5	43	653	16
21	587	5	44	657	18

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
22	589	5			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q17. CBT ELA Grade 7 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	512	34	24	592	5
1	517	29	25	594	5
2	521	25	26	595	5
3	526	21	27	597	5
4	530	19	28	599	5
5	535	16	29	601	5
6	540	13	30	602	5
7	544	12	31	604	5
8	549	10	32	606	5
9	553	9	33	608	5
10	558	8	34	610	5
11	562	7	35	612	5
12	565	7	36	614	5
13	568	6	37	616	6
14	571	6	38	619	6
15	574	6	39	621	6
16	576	6	40	624	6
17	578	6	41	628	7
18	580	5	42	632	8
19	582	5	43	638	10
20	584	5	44	645	12
21	586	5	45	650	15
22	588	5	46	654	16
23	590	5			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q18. CBT ELA Grade 8 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	508	29	24	587	5
1	512	26	25	589	5
2	517	22	26	591	5

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
3	522	19	27	592	5
4	526	17	28	594	5
5	531	14	29	596	5
6	535	13	30	598	5
7	540	11	31	600	5
8	544	10	32	602	5
9	549	9	33	604	5
10	554	8	34	606	5
11	558	7	35	608	6
12	561	7	36	610	6
13	564	6	37	613	6
14	567	6	38	615	6
15	569	6	39	618	7
16	571	6	40	621	7
17	573	6	41	625	8
18	575	6	42	629	8
19	577	5	43	635	10
20	579	5	44	643	13
21	581	5	45	647	15
22	583	5	46	651	17
23	585	5			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q19. CBT Mathematics Grade 3 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	531	25	18	601	6
1	536	21	19	603	6
2	541	18	20	605	6
3	545	15	21	608	6
4	550	13	22	610	6
5	554	12	23	613	6
6	559	10	24	615	6
7	565	9	25	618	6
8	570	8	26	620	6
9	574	8	27	623	6
10	578	7	28	627	7

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
11	581	7	29	630	7
12	584	7	30	635	8
13	587	7	31	640	9
14	590	6	32	647	11
15	593	6	33	651	12
16	596	6	34	655	14
17	598	6			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q20. CBT Mathematics Grade 4 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	533	25	18	597	6
1	537	21	19	599	6
2	542	18	20	602	6
3	547	14	21	604	6
4	551	12	22	607	6
5	556	10	23	609	6
6	562	9	24	612	6
7	566	8	25	615	6
8	570	7	26	617	6
9	573	7	27	620	7
10	576	7	28	624	7
11	579	6	29	628	8
12	582	6	30	632	9
13	585	6	31	638	10
14	587	6	32	646	12
15	590	6	33	650	14
16	592	6	34	654	16
17	595	6			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q21. CBT Mathematics Grade 5 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	510	28	23	593	6
1	514	25	24	595	6
2	519	22	25	597	6

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
3	524	19	26	599	6
4	528	17	27	601	6
5	533	15	28	603	6
6	537	14	29	605	6
7	542	12	30	607	6
8	546	12	31	610	6
9	551	11	32	612	6
10	555	10	33	615	6
11	560	9	34	617	6
12	564	8	35	620	7
13	567	8	36	623	7
14	571	7	37	626	7
15	574	7	38	630	8
16	576	7	39	634	9
17	579	7	40	638	9
18	582	6	41	645	11
19	584	6	42	653	13
20	586	6	43	658	15
21	588	6	44	661	16
22	591	6			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q22. CBT Mathematics Grade 6 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	515	29	23	591	5
1	519	26	24	593	5
2	524	22	25	594	5
3	529	18	26	596	5
4	533	16	27	598	5
5	538	14	28	600	5
6	542	12	29	602	5
7	547	11	30	603	5
8	551	10	31	605	5
9	556	9	32	608	6
10	560	8	33	610	6
11	564	7	34	612	6

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
12	567	7	35	615	6
13	570	7	36	617	6
14	572	6	37	620	7
15	575	6	38	624	8
16	577	6	39	628	8
17	579	6	40	633	9
18	581	6	41	639	11
19	583	6	42	649	15
20	585	5	43	653	16
21	587	5	44	657	18
22	589	5			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q23. CBT Mathematics Grade 7 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	512	34	24	592	5
1	517	29	25	594	5
2	521	25	26	595	5
3	526	21	27	597	5
4	530	19	28	599	5
5	535	16	29	601	5
6	540	13	30	602	5
7	544	12	31	604	5
8	549	10	32	606	5
9	553	9	33	608	5
10	558	8	34	610	5
11	562	7	35	612	5
12	565	7	36	614	5
13	568	6	37	616	6
14	571	6	38	619	6
15	574	6	39	621	6
16	576	6	40	624	6
17	578	6	41	628	7
18	580	5	42	632	8
19	582	5	43	638	10
20	584	5	44	645	12

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
21	586	5	45	650	15
22	588	5	46	654	16
23	590	5			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q24. CBT Mathematics Grade 8 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	508	29	24	587	5
1	512	26	25	589	5
2	517	22	26	591	5
3	522	19	27	592	5
4	526	17	28	594	5
5	531	14	29	596	5
6	535	13	30	598	5
7	540	11	31	600	5
8	544	10	32	602	5
9	549	9	33	604	5
10	554	8	34	606	5
11	558	7	35	608	6
12	561	7	36	610	6
13	564	6	37	613	6
14	567	6	38	615	6
15	569	6	39	618	7
16	571	6	40	621	7
17	573	6	41	625	8
18	575	6	42	629	8
19	577	5	43	635	10
20	579	5	44	643	13
21	581	5	45	647	15
22	583	5	46	651	17
23	585	5			

* A CBT mode adjustment has been taken into account for these scale scores

Table Q25. ELA Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
530	60	0.03%	60	0.03%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
535	140	0.08%	200	0.11%
536	11	0.01%	211	0.12%
540	365	0.20%	576	0.31%
541	33	0.02%	609	0.33%
544	789	0.43%	1,398	0.76%
545	101	0.06%	1,499	0.82%
549	1,280	0.70%	2,779	1.52%
550	158	0.09%	2,937	1.61%
553	1,892	1.03%	4,829	2.64%
554	205	0.11%	5,034	2.75%
558	2,423	1.32%	7,457	4.08%
559	318	0.17%	7,775	4.25%
564	3,066	1.68%	10,841	5.93%
565	359	0.20%	11,200	6.12%
569	3,724	2.04%	14,924	8.16%
570	489	0.27%	15,413	8.43%
573	4,445	2.43%	19,858	10.90%
574	606	0.33%	20,464	11.20%
577	5,166	2.82%	25,630	14.00%
578	694	0.38%	26,324	14.40%
580	5,855	3.20%	32,179	17.60%
581	718	0.39%	32,897	18.00%
583	6,354	3.47%	39,251	21.50%
584	876	0.48%	40,127	21.90%
586	6,957	3.80%	47,084	25.70%
587	857	0.47%	47,941	26.20%
589	7,231	3.95%	55,172	30.20%
590	939	0.51%	56,111	30.70%
592	7,446	4.07%	63,557	34.80%
593	957	0.52%	64,514	35.30%
595	7,635	4.17%	72,149	39.50%
596	929	0.51%	73,078	40.00%
597	7,762	4.24%	80,840	44.20%
598	925	0.51%	81,765	44.70%
600	7,989	4.37%	89,754	49.10%
601	952	0.52%	90,706	49.60%
602	7,917	4.33%	98,623	53.90%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
603	879	0.48%	99,502	54.40%
604	8,173	4.47%	107,675	58.90%
605	856	0.47%	108,531	59.30%
607	8,004	4.38%	116,535	63.70%
608	860	0.47%	117,395	64.20%
609	7,870	4.30%	125,265	68.50%
610	822	0.45%	126,087	68.90%
612	7,976	4.36%	134,063	73.30%
613	763	0.42%	134,826	73.70%
614	7,601	4.16%	142,427	77.90%
615	667	0.36%	143,094	78.20%
617	6,985	3.82%	150,079	82.10%
618	589	0.32%	150,668	82.40%
619	6,643	3.63%	157,311	86.00%
620	485	0.27%	157,796	86.30%
622	5,908	3.23%	163,704	89.50%
623	425	0.23%	164,129	89.70%
626	5,332	2.92%	169,461	92.70%
627	312	0.17%	169,773	92.80%
629	4,405	2.41%	174,178	95.20%
630	258	0.14%	174,436	95.40%
634	3,400	1.86%	177,836	97.20%
635	154	0.08%	177,990	97.30%
639	2,375	1.30%	180,365	98.60%
640	90	0.05%	180,455	98.70%
646	1,509	0.83%	181,964	99.50%
647	44	0.02%	182,008	99.50%
650	674	0.37%	182,682	99.90%
651	16	0.01%	182,698	99.90%
655	187	0.10%	182,885	100.00%

Table Q26. ELA Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
532	58	0.03%	58	0.03%
536	113	0.06%	171	0.09%
537	8	0.00%	179	0.10%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
541	386	0.21%	565	0.31%
542	19	0.01%	584	0.32%
546	775	0.42%	1,359	0.74%
547	64	0.03%	1,423	0.77%
550	1,300	0.71%	2,723	1.48%
551	90	0.05%	2,813	1.53%
555	1,858	1.01%	4,671	2.53%
556	181	0.10%	4,852	2.63%
561	2,463	1.34%	7,315	3.97%
562	256	0.14%	7,571	4.11%
565	2,939	1.59%	10,510	5.70%
566	331	0.18%	10,841	5.88%
569	3,621	1.97%	14,462	7.85%
570	424	0.23%	14,886	8.08%
572	4,058	2.20%	18,944	10.30%
573	462	0.25%	19,406	10.50%
575	4,573	2.48%	23,979	13.00%
576	555	0.30%	24,534	13.30%
578	4,824	2.62%	29,358	15.90%
579	593	0.32%	29,951	16.30%
581	5,431	2.95%	35,382	19.20%
582	660	0.36%	36,042	19.60%
584	5,659	3.07%	41,701	22.60%
585	652	0.35%	42,353	23.00%
586	5,841	3.17%	48,194	26.20%
587	707	0.38%	48,901	26.50%
589	6,491	3.52%	55,392	30.10%
590	775	0.42%	56,167	30.50%
591	6,791	3.69%	62,958	34.20%
592	796	0.43%	63,754	34.60%
594	6,921	3.76%	70,675	38.40%
595	823	0.45%	71,498	38.80%
596	7,156	3.88%	78,654	42.70%
597	849	0.46%	79,503	43.10%
598	7,659	4.16%	87,162	47.30%
599	828	0.45%	87,990	47.80%
601	8,059	4.37%	96,049	52.10%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
602	881	0.48%	96,930	52.60%
603	8,062	4.38%	104,992	57.00%
604	880	0.48%	105,872	57.50%
606	8,575	4.65%	114,447	62.10%
607	873	0.47%	115,320	62.60%
608	8,301	4.50%	123,621	67.10%
609	841	0.46%	124,462	67.50%
611	8,356	4.53%	132,818	72.10%
612	783	0.42%	133,601	72.50%
614	8,321	4.52%	141,922	77.00%
615	763	0.41%	142,685	77.40%
616	7,946	4.31%	150,631	81.70%
617	679	0.37%	151,310	82.10%
619	7,194	3.90%	158,504	86.00%
620	585	0.32%	159,089	86.30%
623	6,691	3.63%	165,780	90.00%
624	470	0.26%	166,250	90.20%
627	5,730	3.11%	171,980	93.30%
628	355	0.19%	172,335	93.50%
631	4,626	2.51%	176,961	96.00%
632	248	0.13%	177,209	96.20%
637	3,372	1.83%	180,581	98.00%
638	182	0.10%	180,763	98.10%
645	2,066	1.12%	182,829	99.20%
646	78	0.04%	182,907	99.30%
649	1,037	0.56%	183,944	99.80%
650	29	0.02%	183,973	99.80%
654	293	0.16%	184,266	100.00%

Table Q27. ELA Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
509	31	0.02%	31	0.02%
510	1	0.00%	32	0.02%
513	8	0.00%	40	0.02%
514	3	0.00%	43	0.02%
518	34	0.02%	77	0.04%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
519	1	0.00%	78	0.04%
523	59	0.03%	137	0.08%
524	1	0.00%	138	0.08%
527	128	0.07%	266	0.15%
528	10	0.01%	276	0.16%
532	223	0.13%	499	0.28%
533	18	0.01%	517	0.29%
536	402	0.23%	919	0.52%
537	35	0.02%	954	0.54%
541	620	0.35%	1,574	0.89%
542	48	0.03%	1,622	0.91%
545	796	0.45%	2,418	1.36%
546	65	0.04%	2,483	1.40%
550	1,078	0.61%	3,561	2.00%
551	101	0.06%	3,662	2.06%
554	1,344	0.76%	5,006	2.82%
555	141	0.08%	5,147	2.90%
559	1,713	0.96%	6,860	3.86%
560	176	0.10%	7,036	3.96%
563	2,049	1.15%	9,085	5.12%
564	212	0.12%	9,297	5.23%
566	2,312	1.30%	11,609	6.54%
567	262	0.15%	11,871	6.68%
570	2,719	1.53%	14,590	8.21%
571	321	0.18%	14,911	8.40%
573	3,220	1.81%	18,131	10.20%
574	337	0.19%	18,468	10.40%
575	3,335	1.88%	21,803	12.30%
576	421	0.24%	22,224	12.50%
578	3,722	2.10%	25,946	14.60%
579	435	0.24%	26,381	14.90%
581	3,972	2.24%	30,353	17.10%
582	493	0.28%	30,846	17.40%
583	4,411	2.48%	35,257	19.90%
584	517	0.29%	35,774	20.10%
585	4,779	2.69%	40,553	22.80%
586	555	0.31%	41,108	23.10%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
587	5,126	2.89%	46,234	26.00%
588	555	0.31%	46,789	26.30%
590	5,439	3.06%	52,228	29.40%
591	648	0.36%	52,876	29.80%
592	5,553	3.13%	58,429	32.90%
593	688	0.39%	59,117	33.30%
594	6,054	3.41%	65,171	36.70%
595	686	0.39%	65,857	37.10%
596	6,355	3.58%	72,212	40.70%
597	711	0.40%	72,923	41.10%
598	6,692	3.77%	79,615	44.80%
599	760	0.43%	80,375	45.30%
600	7,062	3.98%	87,437	49.20%
601	781	0.44%	88,218	49.70%
602	7,253	4.08%	95,471	53.80%
603	791	0.45%	96,262	54.20%
604	7,544	4.25%	103,806	58.40%
605	777	0.44%	104,583	58.90%
606	7,583	4.27%	112,166	63.20%
607	775	0.44%	112,941	63.60%
609	7,661	4.31%	120,602	67.90%
610	714	0.40%	121,316	68.30%
611	7,595	4.28%	128,911	72.60%
612	694	0.39%	129,605	73.00%
614	7,464	4.20%	137,069	77.20%
615	655	0.37%	137,724	77.50%
616	7,272	4.09%	144,996	81.60%
617	577	0.32%	145,573	82.00%
619	6,490	3.65%	152,063	85.60%
620	501	0.28%	152,564	85.90%
622	6,019	3.39%	158,583	89.30%
623	443	0.25%	159,026	89.50%
625	5,100	2.87%	164,126	92.40%
626	365	0.21%	164,491	92.60%
629	4,375	2.46%	168,866	95.10%
630	237	0.13%	169,103	95.20%
633	3,310	1.86%	172,413	97.10%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
634	170	0.10%	172,583	97.20%
637	2,269	1.28%	174,852	98.40%
638	112	0.06%	174,964	98.50%
644	1,464	0.82%	176,428	99.30%
645	59	0.03%	176,487	99.40%
652	748	0.42%	177,235	99.80%
653	18	0.01%	177,253	99.80%
657	306	0.17%	177,559	100.00%
658	2	0.00%	177,561	100.00%
661	48	0.03%	177,609	100.00%

Table Q28. ELA Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
514	48	0.03%	48	0.03%
515	1	0.00%	49	0.03%
518	21	0.01%	70	0.04%
519	1	0.00%	71	0.04%
523	40	0.02%	111	0.06%
524	3	0.00%	114	0.07%
528	77	0.04%	191	0.11%
529	3	0.00%	194	0.11%
532	172	0.10%	366	0.21%
533	18	0.01%	384	0.22%
537	284	0.16%	668	0.39%
538	26	0.02%	694	0.40%
541	504	0.29%	1,198	0.69%
542	40	0.02%	1,238	0.71%
546	723	0.42%	1,961	1.13%
547	61	0.04%	2,022	1.17%
550	1,007	0.58%	3,029	1.75%
551	90	0.05%	3,119	1.80%
555	1,251	0.72%	4,370	2.52%
556	140	0.08%	4,510	2.60%
559	1,601	0.92%	6,111	3.53%
560	176	0.10%	6,287	3.63%
563	1,861	1.07%	8,148	4.70%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
564	237	0.14%	8,385	4.84%
566	2,153	1.24%	10,538	6.08%
567	273	0.16%	10,811	6.24%
569	2,432	1.40%	13,243	7.65%
570	333	0.19%	13,576	7.84%
571	2,629	1.52%	16,205	9.36%
572	330	0.19%	16,535	9.55%
574	2,849	1.65%	19,384	11.20%
575	398	0.23%	19,782	11.40%
576	2,969	1.71%	22,751	13.10%
577	426	0.25%	23,177	13.40%
578	3,213	1.86%	26,390	15.20%
579	467	0.27%	26,857	15.50%
580	3,380	1.95%	30,237	17.50%
581	495	0.29%	30,732	17.70%
582	3,626	2.09%	34,358	19.80%
583	498	0.29%	34,856	20.10%
584	3,696	2.13%	38,552	22.30%
585	549	0.32%	39,101	22.60%
586	3,895	2.25%	42,996	24.80%
587	551	0.32%	43,547	25.10%
588	4,080	2.36%	47,627	27.50%
589	581	0.34%	48,208	27.80%
590	4,382	2.53%	52,590	30.40%
591	601	0.35%	53,191	30.70%
592	4,676	2.70%	57,867	33.40%
593	5,476	3.16%	63,343	36.60%
594	754	0.44%	64,097	37.00%
595	5,130	2.96%	69,227	40.00%
596	746	0.43%	69,973	40.40%
597	5,289	3.05%	75,262	43.50%
598	747	0.43%	76,009	43.90%
599	5,530	3.19%	81,539	47.10%
600	820	0.47%	82,359	47.60%
601	5,803	3.35%	88,162	50.90%
602	6,996	4.04%	95,158	54.90%
603	859	0.50%	96,017	55.40%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
604	6,421	3.71%	102,438	59.20%
605	845	0.49%	103,283	59.60%
607	6,731	3.89%	110,014	63.50%
608	926	0.53%	110,940	64.10%
609	6,921	4.00%	117,861	68.10%
610	893	0.52%	118,754	68.60%
611	6,984	4.03%	125,738	72.60%
612	919	0.53%	126,657	73.10%
614	7,315	4.22%	133,972	77.40%
615	883	0.51%	134,855	77.90%
616	7,066	4.08%	141,921	81.90%
617	914	0.53%	142,835	82.50%
619	6,875	3.97%	149,710	86.40%
620	738	0.43%	150,448	86.90%
623	6,112	3.53%	156,560	90.40%
624	632	0.36%	157,192	90.80%
627	5,327	3.08%	162,519	93.80%
628	483	0.28%	163,002	94.10%
632	4,179	2.41%	167,181	96.50%
633	310	0.18%	167,491	96.70%
638	2,850	1.65%	170,341	98.40%
639	243	0.14%	170,584	98.50%
648	1,618	0.93%	172,202	99.40%
649	104	0.06%	172,306	99.50%
652	694	0.40%	173,000	99.90%
653	26	0.02%	173,026	99.90%
657	157	0.09%	173,183	100.00%

Table Q29. ELA Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
511	55	0.03%	55	0.03%
512	1	0.00%	56	0.03%
516	17	0.01%	73	0.05%
517	2	0.00%	75	0.05%
520	42	0.03%	117	0.07%
521	2	0.00%	119	0.07%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
525	74	0.05%	193	0.12%
526	6	0.00%	199	0.12%
529	175	0.11%	374	0.23%
530	12	0.01%	386	0.24%
534	291	0.18%	677	0.42%
535	27	0.02%	704	0.43%
539	503	0.31%	1,207	0.75%
540	35	0.02%	1,242	0.77%
543	664	0.41%	1,906	1.18%
544	56	0.03%	1,962	1.21%
548	822	0.51%	2,784	1.72%
549	79	0.05%	2,863	1.77%
552	1,166	0.72%	4,029	2.49%
553	109	0.07%	4,138	2.55%
557	1,294	0.80%	5,432	3.35%
558	150	0.09%	5,582	3.45%
561	1,521	0.94%	7,103	4.39%
562	192	0.12%	7,295	4.50%
564	1,715	1.06%	9,010	5.56%
565	223	0.14%	9,233	5.70%
567	1,919	1.18%	11,152	6.89%
568	252	0.16%	11,404	7.04%
570	2,055	1.27%	13,459	8.31%
571	238	0.15%	13,697	8.46%
573	2,371	1.46%	16,068	9.92%
574	303	0.19%	16,371	10.10%
575	2,547	1.57%	18,918	11.70%
576	374	0.23%	19,292	11.90%
577	2,808	1.73%	22,100	13.60%
578	364	0.22%	22,464	13.90%
579	3,012	1.86%	25,476	15.70%
580	357	0.22%	25,833	16.00%
581	3,237	2.00%	29,070	17.90%
582	426	0.26%	29,496	18.20%
583	3,495	2.16%	32,991	20.40%
584	484	0.30%	33,475	20.70%
585	3,640	2.25%	37,115	22.90%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
586	454	0.28%	37,569	23.20%
587	3,798	2.35%	41,367	25.50%
588	528	0.33%	41,895	25.90%
589	4,026	2.49%	45,921	28.40%
590	518	0.32%	46,439	28.70%
591	4,239	2.62%	50,678	31.30%
592	535	0.33%	51,213	31.60%
593	4,441	2.74%	55,654	34.40%
594	5,170	3.19%	60,824	37.60%
595	667	0.41%	61,491	38.00%
596	4,879	3.01%	66,370	41.00%
597	630	0.39%	67,000	41.40%
598	5,070	3.13%	72,070	44.50%
599	641	0.40%	72,711	44.90%
600	5,368	3.31%	78,079	48.20%
601	5,945	3.67%	84,024	51.90%
602	646	0.40%	84,670	52.30%
603	5,585	3.45%	90,255	55.70%
604	652	0.40%	90,907	56.10%
605	5,606	3.46%	96,513	59.60%
606	717	0.44%	97,230	60.00%
607	5,869	3.62%	103,099	63.70%
608	715	0.44%	103,814	64.10%
609	5,877	3.63%	109,691	67.70%
610	654	0.40%	110,345	68.10%
611	6,054	3.74%	116,399	71.90%
612	661	0.41%	117,060	72.30%
613	6,012	3.71%	123,072	76.00%
614	658	0.41%	123,730	76.40%
615	5,870	3.62%	129,600	80.00%
616	581	0.36%	130,181	80.40%
618	5,894	3.64%	136,075	84.00%
619	518	0.32%	136,593	84.30%
620	5,515	3.41%	142,108	87.70%
621	497	0.31%	142,605	88.10%
623	5,098	3.15%	147,703	91.20%
624	381	0.24%	148,084	91.40%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
627	4,317	2.67%	152,401	94.10%
628	301	0.19%	152,702	94.30%
631	3,502	2.16%	156,204	96.40%
632	220	0.14%	156,424	96.60%
637	2,680	1.65%	159,104	98.20%
638	129	0.08%	159,233	98.30%
644	1,646	1.02%	160,879	99.30%
645	77	0.05%	160,956	99.40%
649	770	0.48%	161,726	99.90%
650	28	0.02%	161,754	99.90%
654	204	0.13%	161,958	100.00%

Table Q30. ELA Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
507	79	0.05%	79	0.05%
508	4	0.00%	83	0.05%
511	18	0.01%	101	0.07%
516	37	0.02%	138	0.09%
517	1	0.00%	139	0.09%
521	59	0.04%	198	0.13%
522	7	0.00%	205	0.13%
525	139	0.09%	344	0.22%
526	9	0.01%	353	0.23%
530	214	0.14%	567	0.37%
531	14	0.01%	581	0.38%
534	318	0.21%	899	0.58%
535	22	0.01%	921	0.60%
539	484	0.31%	1,405	0.91%
540	58	0.04%	1,463	0.95%
543	612	0.40%	2,075	1.34%
544	44	0.03%	2,119	1.37%
548	754	0.49%	2,873	1.86%
549	75	0.05%	2,948	1.91%
553	936	0.61%	3,884	2.51%
554	94	0.06%	3,978	2.57%
557	1,009	0.65%	4,987	3.22%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
558	125	0.08%	5,112	3.31%
560	1,135	0.73%	6,247	4.04%
561	142	0.09%	6,389	4.13%
563	1,244	0.80%	7,633	4.94%
564	176	0.11%	7,809	5.05%
566	1,501	0.97%	9,310	6.02%
567	168	0.11%	9,478	6.13%
568	1,597	1.03%	11,075	7.16%
569	185	0.12%	11,260	7.28%
570	1,761	1.14%	13,021	8.42%
571	237	0.15%	13,258	8.57%
572	1,838	1.19%	15,096	9.76%
573	217	0.14%	15,313	9.90%
574	2,139	1.38%	17,452	11.30%
575	276	0.18%	17,728	11.50%
576	2,213	1.43%	19,941	12.90%
577	299	0.19%	20,240	13.10%
578	2,424	1.57%	22,664	14.70%
579	314	0.20%	22,978	14.90%
580	2,703	1.75%	25,681	16.60%
581	352	0.23%	26,033	16.80%
582	2,962	1.92%	28,995	18.70%
583	403	0.26%	29,398	19.00%
584	3,312	2.14%	32,710	21.10%
585	444	0.29%	33,154	21.40%
586	3,552	2.30%	36,706	23.70%
587	504	0.33%	37,210	24.10%
588	3,857	2.49%	41,067	26.60%
589	534	0.35%	41,601	26.90%
590	4,263	2.76%	45,864	29.70%
591	5,027	3.25%	50,891	32.90%
592	565	0.37%	51,456	33.30%
593	4,578	2.96%	56,034	36.20%
594	591	0.38%	56,625	36.60%
595	4,989	3.23%	61,614	39.80%
596	618	0.40%	62,232	40.20%
597	5,296	3.42%	67,528	43.70%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
598	608	0.39%	68,136	44.10%
599	5,644	3.65%	73,780	47.70%
600	650	0.42%	74,430	48.10%
601	5,804	3.75%	80,234	51.90%
602	709	0.46%	80,943	52.30%
603	6,042	3.91%	86,985	56.20%
604	672	0.43%	87,657	56.70%
605	6,178	3.99%	93,835	60.70%
606	648	0.42%	94,483	61.10%
607	6,323	4.09%	100,806	65.20%
608	664	0.43%	101,470	65.60%
609	6,486	4.19%	107,956	69.80%
610	675	0.44%	108,631	70.20%
612	6,490	4.20%	115,121	74.40%
613	657	0.42%	115,778	74.90%
614	6,482	4.19%	122,260	79.00%
615	610	0.39%	122,870	79.40%
617	6,206	4.01%	129,076	83.50%
618	589	0.38%	129,665	83.80%
620	5,881	3.80%	135,546	87.60%
621	515	0.33%	136,061	88.00%
624	5,237	3.39%	141,298	91.40%
625	443	0.29%	141,741	91.60%
628	4,549	2.94%	146,290	94.60%
629	360	0.23%	146,650	94.80%
634	3,624	2.34%	150,274	97.20%
635	236	0.15%	150,510	97.30%
642	2,429	1.57%	152,939	98.90%
643	123	0.08%	153,062	99.00%
646	1,169	0.76%	154,231	99.70%
647	48	0.03%	154,279	99.80%
651	384	0.25%	154,663	100.00%

Table Q31. Mathematics Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
526	14	0.01%	14	0.01%
531	58	0.03%	72	0.04%
532	5	0.00%	77	0.04%
535	160	0.09%	237	0.13%
536	14	0.01%	251	0.14%
540	416	0.22%	667	0.36%
541	46	0.02%	713	0.39%
545	780	0.42%	1,493	0.81%
546	80	0.04%	1,573	0.85%
549	1,421	0.77%	2,994	1.62%
550	133	0.07%	3,127	1.69%
554	1,998	1.08%	5,125	2.77%
555	192	0.10%	5,317	2.87%
560	2,411	1.30%	7,728	4.18%
561	217	0.12%	7,945	4.30%
564	2,776	1.50%	10,721	5.80%
565	256	0.14%	10,977	5.93%
568	3,020	1.63%	13,997	7.57%
569	256	0.14%	14,253	7.71%
571	3,183	1.72%	17,436	9.43%
572	297	0.16%	17,733	9.59%
574	3,319	1.79%	21,052	11.40%
575	306	0.17%	21,358	11.50%
576	3,448	1.86%	24,806	13.40%
577	284	0.15%	25,090	13.60%
578	3,473	1.88%	28,563	15.40%
579	299	0.16%	28,862	15.60%
580	3,705	2.00%	32,567	17.60%
581	314	0.17%	32,881	17.80%
582	3,732	2.02%	36,613	19.80%
583	358	0.19%	36,971	20.00%
584	3,986	2.15%	40,957	22.10%
585	367	0.20%	41,324	22.30%
586	4,027	2.18%	45,351	24.50%
587	4,516	2.44%	49,867	27.00%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
588	390	0.21%	50,257	27.20%
589	4,357	2.36%	54,614	29.50%
590	408	0.22%	55,022	29.70%
591	4,430	2.39%	59,452	32.10%
592	4,981	2.69%	64,433	34.80%
593	438	0.24%	64,871	35.10%
594	4,583	2.48%	69,454	37.50%
595	5,377	2.91%	74,831	40.50%
596	476	0.26%	75,307	40.70%
597	5,003	2.70%	80,310	43.40%
598	5,531	2.99%	85,841	46.40%
599	478	0.26%	86,319	46.70%
600	5,472	2.96%	91,791	49.60%
601	5,950	3.22%	97,741	52.80%
602	500	0.27%	98,241	53.10%
603	5,560	3.01%	103,801	56.10%
604	6,106	3.30%	109,907	59.40%
605	482	0.26%	110,389	59.70%
606	5,753	3.11%	116,142	62.80%
607	444	0.24%	116,586	63.00%
608	5,909	3.19%	122,495	66.20%
609	6,476	3.50%	128,971	69.70%
610	504	0.27%	129,475	70.00%
611	6,174	3.34%	135,649	73.30%
612	471	0.25%	136,120	73.60%
613	6,118	3.31%	142,238	76.90%
614	442	0.24%	142,680	77.10%
615	6,234	3.37%	148,914	80.50%
616	480	0.26%	149,394	80.80%
618	5,961	3.22%	155,355	84.00%
619	451	0.24%	155,806	84.20%
620	6,019	3.25%	161,825	87.50%
621	367	0.20%	162,192	87.70%
623	5,845	3.16%	168,037	90.80%
624	358	0.19%	168,395	91.00%
627	5,467	2.96%	173,862	94.00%
628	304	0.16%	174,166	94.20%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
633	4,723	2.55%	178,889	96.70%
634	215	0.12%	179,104	96.80%
642	3,662	1.98%	182,766	98.80%
643	135	0.07%	182,901	98.90%
646	2,069	1.12%	184,970	100.00%

Table Q32. Mathematics Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
525	11	0.01%	11	0.01%
526	2	0.00%	13	0.01%
529	56	0.03%	69	0.04%
530	1	0.00%	70	0.04%
534	113	0.06%	183	0.10%
535	7	0.00%	190	0.10%
538	305	0.16%	495	0.27%
539	14	0.01%	509	0.27%
543	672	0.36%	1,181	0.63%
544	42	0.02%	1,223	0.66%
548	1,124	0.60%	2,347	1.26%
549	59	0.03%	2,406	1.29%
552	1,693	0.91%	4,099	2.20%
553	127	0.07%	4,226	2.27%
557	2,327	1.25%	6,553	3.52%
558	140	0.08%	6,693	3.59%
563	2,814	1.51%	9,507	5.10%
564	187	0.10%	9,694	5.20%
567	3,109	1.67%	12,803	6.87%
568	188	0.10%	12,991	6.97%
571	3,443	1.85%	16,434	8.82%
572	233	0.13%	16,667	8.94%
573	3,541	1.90%	20,208	10.80%
574	216	0.12%	20,424	11.00%
576	3,615	1.94%	24,039	12.90%
577	246	0.13%	24,285	13.00%
578	3,797	2.04%	28,082	15.10%
579	246	0.13%	28,328	15.20%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
580	3,801	2.04%	32,129	17.20%
581	273	0.15%	32,402	17.40%
582	3,712	1.99%	36,114	19.40%
583	292	0.16%	36,406	19.50%
584	3,874	2.08%	40,280	21.60%
585	4,152	2.23%	44,432	23.80%
586	298	0.16%	44,730	24.00%
587	3,889	2.09%	48,619	26.10%
588	4,223	2.27%	52,842	28.40%
589	4,200	2.25%	57,042	30.60%
590	292	0.16%	57,334	30.80%
591	4,061	2.18%	61,395	32.90%
592	4,407	2.37%	65,802	35.30%
593	4,397	2.36%	70,199	37.70%
594	323	0.17%	70,522	37.80%
595	4,167	2.24%	74,689	40.10%
596	4,564	2.45%	79,253	42.50%
597	4,492	2.41%	83,745	44.90%
598	4,592	2.46%	88,337	47.40%
599	4,616	2.48%	92,953	49.90%
600	337	0.18%	93,290	50.10%
601	4,389	2.36%	97,679	52.40%
602	4,934	2.65%	102,613	55.10%
603	4,980	2.67%	107,593	57.70%
604	5,076	2.72%	112,669	60.50%
605	376	0.20%	113,045	60.70%
606	4,800	2.58%	117,845	63.20%
607	5,317	2.85%	123,162	66.10%
608	349	0.19%	123,511	66.30%
609	5,058	2.71%	128,569	69.00%
610	5,610	3.01%	134,179	72.00%
611	405	0.22%	134,584	72.20%
612	5,493	2.95%	140,077	75.20%
613	389	0.21%	140,466	75.40%
614	5,521	2.96%	145,987	78.30%
615	390	0.21%	146,377	78.60%
616	5,748	3.08%	152,125	81.60%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
617	410	0.22%	152,535	81.90%
618	5,783	3.10%	158,318	85.00%
619	359	0.19%	158,677	85.20%
621	5,657	3.04%	164,334	88.20%
622	361	0.19%	164,695	88.40%
624	5,813	3.12%	170,508	91.50%
625	283	0.15%	170,791	91.70%
628	5,364	2.88%	176,155	94.50%
629	237	0.13%	176,392	94.70%
635	4,849	2.60%	181,241	97.30%
636	176	0.09%	181,417	97.40%
645	3,552	1.91%	184,969	99.30%
646	113	0.06%	185,082	99.30%
650	1,249	0.67%	186,331	100%

Table Q33. Mathematics Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
527	9	0.01%	9	0.01%
531	54	0.03%	63	0.04%
536	147	0.08%	210	0.12%
541	432	0.24%	642	0.36%
545	936	0.52%	1,578	0.88%
550	1,620	0.91%	3,198	1.79%
554	2,420	1.35%	5,618	3.14%
562	3,208	1.79%	8,826	4.93%
567	3,764	2.10%	12,590	7.04%
571	4,052	2.27%	16,642	9.30%
574	4,374	2.45%	21,016	11.70%
577	4,603	2.57%	25,619	14.30%
579	4,615	2.58%	30,234	16.90%
581	4,611	2.58%	34,845	19.50%
583	4,839	2.71%	39,684	22.20%
585	4,799	2.68%	44,483	24.90%
587	4,837	2.70%	49,320	27.60%
589	5,002	2.80%	54,322	30.40%
590	5,068	2.83%	59,390	33.20%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
592	4,921	2.75%	64,311	36.00%
593	4,957	2.77%	69,268	38.70%
595	4,895	2.74%	74,163	41.50%
596	4,760	2.66%	78,923	44.10%
597	4,686	2.62%	83,609	46.70%
599	4,572	2.56%	88,181	49.30%
600	4,661	2.61%	92,842	51.90%
601	4,568	2.55%	97,410	54.50%
602	4,520	2.53%	101,930	57.00%
604	4,501	2.52%	106,431	59.50%
605	4,501	2.52%	110,932	62.00%
606	4,542	2.54%	115,474	64.60%
607	4,537	2.54%	120,011	67.10%
609	4,533	2.53%	124,544	69.60%
610	4,547	2.54%	129,091	72.20%
611	4,584	2.56%	133,675	74.70%
613	4,590	2.57%	138,265	77.30%
615	4,552	2.54%	142,817	79.80%
616	4,540	2.54%	147,357	82.40%
618	4,593	2.57%	151,950	84.90%
620	4,461	2.49%	156,411	87.40%
622	4,414	2.47%	160,825	89.90%
625	4,219	2.36%	165,044	92.30%
629	3,939	2.20%	168,983	94.50%
633	3,470	1.94%	172,453	96.40%
639	3,012	1.68%	175,465	98.10%
650	2,216	1.24%	177,681	99.30%
654	1,194	0.67%	178,875	100.00%

Table Q34. Mathematics Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
528	23	0.01%	23	0.01%
530	2	0.00%	25	0.01%
533	71	0.04%	96	0.06%
535	9	0.01%	105	0.06%
537	225	0.13%	330	0.19%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
539	13	0.01%	343	0.20%
542	603	0.35%	946	0.54%
544	42	0.02%	988	0.57%
546	1,283	0.74%	2,271	1.31%
548	78	0.04%	2,349	1.35%
551	2,125	1.22%	4,474	2.58%
553	146	0.08%	4,620	2.66%
555	3,133	1.80%	7,753	4.46%
557	242	0.14%	7,995	4.60%
565	3,835	2.21%	11,830	6.81%
567	273	0.16%	12,103	6.97%
571	4,535	2.61%	16,638	9.58%
573	310	0.18%	16,948	9.76%
575	4,957	2.85%	21,905	12.60%
577	377	0.22%	22,282	12.80%
579	5,157	2.97%	27,439	15.80%
581	424	0.24%	27,863	16.00%
582	5,207	3.00%	33,070	19.00%
584	5,510	3.17%	38,580	22.20%
586	5,430	3.13%	44,010	25.30%
588	5,183	2.98%	49,193	28.30%
590	5,237	3.01%	54,430	31.30%
592	5,082	2.93%	59,512	34.30%
593	4,708	2.71%	64,220	37.00%
594	453	0.26%	64,673	37.20%
595	4,988	2.87%	69,661	40.10%
596	4,460	2.57%	74,121	42.70%
597	460	0.26%	74,581	42.90%
598	4,866	2.80%	79,447	45.70%
599	4,340	2.50%	83,787	48.20%
600	4,667	2.69%	88,454	50.90%
601	4,487	2.58%	92,941	53.50%
602	431	0.25%	93,372	53.70%
603	4,385	2.52%	97,757	56.30%
604	3,964	2.28%	101,721	58.60%
605	4,296	2.47%	106,017	61.00%
606	4,228	2.43%	110,245	63.50%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
607	396	0.23%	110,641	63.70%
608	4,131	2.38%	114,772	66.10%
609	3,677	2.12%	118,449	68.20%
610	3,897	2.24%	122,346	70.40%
611	3,833	2.21%	126,179	72.60%
612	3,799	2.19%	129,978	74.80%
613	381	0.22%	130,359	75.00%
614	3,666	2.11%	134,025	77.10%
615	3,200	1.84%	137,225	79.00%
616	3,546	2.04%	140,771	81.00%
617	3,546	2.04%	144,317	83.10%
618	315	0.18%	144,632	83.30%
619	3,501	2.02%	148,133	85.30%
620	3,044	1.75%	151,177	87.00%
621	300	0.17%	151,477	87.20%
622	3,169	1.82%	154,646	89.00%
624	3,020	1.74%	157,666	90.80%
625	2,672	1.54%	160,338	92.30%
626	199	0.11%	160,537	92.40%
627	245	0.14%	160,782	92.50%
628	2,505	1.44%	163,287	94.00%
630	2,569	1.48%	165,856	95.50%
632	155	0.09%	166,011	95.60%
633	2,176	1.25%	168,187	96.80%
635	125	0.07%	168,312	96.90%
637	1,864	1.07%	170,176	98.00%
639	97	0.06%	170,273	98.00%
642	1,636	0.94%	171,909	99.00%
644	55	0.03%	171,964	99.00%
651	1,093	0.63%	173,057	99.60%
653	43	0.02%	173,100	99.60%
656	631	0.36%	173,731	100.00%

Table Q35. Mathematics Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
524	22	0.01%	22	0.01%
529	52	0.03%	74	0.05%
533	114	0.07%	188	0.12%
538	248	0.15%	436	0.27%
542	546	0.34%	982	0.61%
547	1,114	0.69%	2,096	1.31%
552	1,966	1.23%	4,062	2.53%
556	2,743	1.71%	6,805	4.24%
561	3,547	2.21%	10,352	6.45%
566	4,180	2.60%	14,532	9.05%
572	4,594	2.86%	19,126	11.90%
577	4,662	2.90%	23,788	14.80%
580	4,795	2.99%	28,583	17.80%
582	4,490	2.80%	33,073	20.60%
585	4,457	2.78%	37,530	23.40%
587	4,216	2.63%	41,746	26.00%
588	4,030	2.51%	45,776	28.50%
590	3,947	2.46%	49,723	31.00%
591	3,845	2.40%	53,568	33.40%
593	3,768	2.35%	57,336	35.70%
594	3,642	2.27%	60,978	38.00%
595	3,539	2.21%	64,517	40.20%
596	3,501	2.18%	68,018	42.40%
598	3,544	2.21%	71,562	44.60%
599	3,333	2.08%	74,895	46.70%
600	3,335	2.08%	78,230	48.70%
601	3,377	2.10%	81,607	50.80%
602	3,366	2.10%	84,973	52.90%
603	3,345	2.08%	88,318	55.00%
604	3,259	2.03%	91,577	57.10%
605	3,230	2.01%	94,807	59.10%
606	3,244	2.02%	98,051	61.10%
607	3,183	1.98%	101,234	63.10%
608	3,226	2.01%	104,460	65.10%
609	3,295	2.05%	107,755	67.10%
610	3,322	2.07%	111,077	69.20%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
611	3,383	2.11%	114,460	71.30%
612	3,330	2.07%	117,790	73.40%
613	3,444	2.15%	121,234	75.50%
614	3,527	2.20%	124,761	77.70%
615	3,370	2.10%	128,131	79.80%
617	3,534	2.20%	131,665	82.00%
618	3,514	2.19%	135,179	84.20%
620	3,569	2.22%	138,748	86.50%
621	3,601	2.24%	142,349	88.70%
623	3,558	2.22%	145,907	90.90%
626	3,518	2.19%	149,425	93.10%
628	3,329	2.07%	152,754	95.20%
632	3,161	1.97%	155,915	97.20%
639	2,700	1.68%	158,615	98.80%
644	1,872	1.17%	160,487	100.00%

Table Q36. Mathematics Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
527	21	0.02%	21	0.02%
528	1	0.00%	22	0.02%
531	51	0.04%	73	0.06%
532	6	0.01%	79	0.07%
536	85	0.07%	164	0.14%
537	4	0.00%	168	0.14%
540	167	0.14%	335	0.29%
541	9	0.01%	344	0.30%
545	393	0.34%	737	0.63%
546	31	0.03%	768	0.66%
549	815	0.70%	1,583	1.36%
550	54	0.05%	1,637	1.40%
554	1,560	1.34%	3,197	2.74%
555	133	0.11%	3,330	2.86%
559	2,315	1.99%	5,645	4.84%
560	178	0.15%	5,823	5.00%
563	3,101	2.66%	8,924	7.66%
564	247	0.21%	9,171	7.87%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
568	3,544	3.04%	12,715	10.90%
569	368	0.32%	13,083	11.20%
575	3,903	3.35%	16,986	14.60%
576	353	0.30%	17,339	14.90%
580	4,057	3.48%	21,396	18.40%
581	334	0.29%	21,730	18.60%
584	3,967	3.40%	25,697	22.10%
585	332	0.28%	26,029	22.30%
586	3,726	3.20%	29,755	25.50%
587	348	0.30%	30,103	25.80%
589	3,469	2.98%	33,572	28.80%
590	285	0.24%	33,857	29.10%
591	3,426	2.94%	37,283	32.00%
592	3,597	3.09%	40,880	35.10%
593	249	0.21%	41,129	35.30%
594	3,121	2.68%	44,250	38.00%
595	240	0.21%	44,490	38.20%
596	3,065	2.63%	47,555	40.80%
597	3,155	2.71%	50,710	43.50%
598	3,084	2.65%	53,794	46.20%
599	3,017	2.59%	56,811	48.80%
600	233	0.20%	57,044	49.00%
601	2,698	2.32%	59,742	51.30%
602	2,935	2.52%	62,677	53.80%
603	2,750	2.36%	65,427	56.10%
604	2,817	2.42%	68,244	58.60%
605	2,762	2.37%	71,006	60.90%
606	2,697	2.31%	73,703	63.20%
607	2,619	2.25%	76,322	65.50%
608	2,384	2.05%	78,706	67.50%
609	2,452	2.10%	81,158	69.60%
610	2,363	2.03%	83,521	71.70%
611	2,375	2.04%	85,896	73.70%
612	2,298	1.97%	88,194	75.70%
613	2,217	1.90%	90,411	77.60%
614	2,135	1.83%	92,546	79.40%
615	125	0.11%	92,671	79.50%

Scale Score	Freq.	Pct.	Cumulative Freq.	Cumulative Pct.
616	1,955	1.68%	94,626	81.20%
617	2,013	1.73%	96,639	82.90%
618	1,952	1.68%	98,591	84.60%
619	1,867	1.60%	100,458	86.20%
620	1,770	1.52%	102,228	87.70%
621	74	0.06%	102,302	87.80%
622	1,649	1.42%	103,951	89.20%
623	1,699	1.46%	105,650	90.70%
624	86	0.07%	105,736	90.70%
625	1,599	1.37%	107,335	92.10%
626	57	0.05%	107,392	92.20%
627	1,511	1.30%	108,903	93.50%
628	35	0.03%	108,938	93.50%
629	1,393	1.20%	110,331	94.70%
630	37	0.03%	110,368	94.70%
631	1,474	1.26%	111,842	96.00%
632	30	0.03%	111,872	96.00%
634	1,320	1.13%	113,192	97.10%
635	34	0.03%	113,226	97.20%
639	1,265	1.09%	114,491	98.20%
640	8	0.01%	114,499	98.30%
646	1,202	1.03%	115,701	99.30%
647	8	0.01%	115,709	99.30%
651	825	0.71%	116,534	100.00%

Appendix R: Study of Operational Test Mode Comparability

Section R.1. Introduction

R.1.1. Overview

Following the 2017 administration, the New York State Education Department (NYSED) continued to offer its operational test (OP) in a computer-based testing (CBT) environment for the Grades 3–8 English Language Arts (ELA) and Mathematics tests in 2018. The schools had the option to administer the tests via paper-based testing (PBT) or computer-based testing (CBT). This study is to evaluate differences in test-level student performance that may be attributable to the mode in which a student tested.

In 2018, all the operational items administered in both PBT and CBT modes. The number of operational items and score points are summarized in Table R.1.1.

Table R.1.1. Operational Items Administered in Both CBT and PBT Modes

Grade	ELA				Math			
	MC Items	CR Items 2-Point	CR Items 4-Point	Total Points	MC Items	CR Items 2-Point	CR Items 3-Point	Total Points
3	18	6	1	34	27	6	1	42
4	18	6	1	34	31	6	1	46
5	28	6	1	44	31	6	1	46
6	28	6	1	44	31	7	1	48
7	28	7	1	46	33	7	1	50
8	28	7	1	46	33	7	1	50

The current study consists of three phases:

1. A propensity score matching approach was conducted to generate the CBT and PBT samples that were comparable on selected covariates that may affect student performance, aside from the test mode itself.
2. The difference in students' test scores were computed between the matched CBT and PBT samples to evaluate the test-level mode comparability.
3. An item-level differential item functioning (DIF) analysis was performed to facilitate the understanding the item- and test-level mode effects.

Section R.2. Method

R.2.1. Preparing Balanced Samples

R.2.1.1 Overview

While the ideal conditions under which to investigate test mode comparability would necessitate random assignment of schools to test in either the CBT or PBT modes, the practical constraints and resources of individual districts and schools preclude such designs. The next best solution is often referred to as a quasi-experimental design. Given that the student population was not randomly equivalent between test modes, the propensity score matching (PSM) methodology

(Austin, 2011a; Rosenbaum, 2010) was applied to draw matched samples of PBT and CBT students who were considered comparable on average in their test performance. In other words, effective propensity score matching produces samples of PBT and CBT students that are on average otherwise comparable, with the only observed difference being that each sample tested in differing modes.

Table R.2.1 shows the number of students in the clean datasets by test mode prior to propensity score matching. In 2018, the CBT participation rates ranged from 9-11% for ELA and 6-8% for Math, an increase of 1-2% from 2017. This study used the same data-cleaning procedures that have been used for operational psychometric analyses, with the following additional rules:

- For Grades 4–8, students without prior year scale scores in the same subject on the adjacent lower grade were removed.
- Because of sample size concerns and concerns about effects unrelated to test mode interfering with the study's inferences, students testing with the Braille or large print forms were dropped. Students who used a non-English language translation (i.e., Chinese, Korean, Haitian-Creole, Russian, Spanish) of a Mathematics form were also removed.

Table R.2.1. Sample Sizes Before Matching by Test Mode

Grade	ELA				Math				2017 CBT %	
	PBT_N	CBT_N	PBT %	CBT %	PBT_N	CBT_N	PBT %	CBT %	ELA	Math
3	162,591	16,763	90.65	9.35	162,813	13,850	92.16	7.84	2.27	1.46
4	165,390	16,282	91.04	8.96	165,025	11,872	93.29	6.71	1.55	0.83
5	159,581	15,594	91.10	8.90	157,831	10,747	93.62	6.38	1.48	0.95
6	151,027	18,988	88.83	11.17	150,945	13,484	91.80	8.20	1.32	1.28
7	140,012	15,907	89.80	10.20	141,086	10,663	92.97	7.03	2.01	1.41
8	136,959	14,563	90.39	9.61	101,294	7,116	93.44	6.56	1.34	0.86

Note. Sample sizes indicate the number of students who took at least one item administered in the test after the data cleaning used for operational psychometric analyses.

R.2.1.2. Propensity Score Models and Matching

In discussion with New York State's Assessment Technical Advisory Committee (TAC), the decision was made to model the propensity score at the student level for CBT testing. The decision to adopt CBT was a school-level decision and modeling it at the student level violates one part of the assumption of strong ignorability (Rosenbaum and Rubin, 1983), meaning that some students had probabilities of assignment to CBT that equaled zero or one. By conditioning on student-level and school-level covariates, Questar was able to best approximate the selection process one might observe if students were able to self-select and therefore treat school assignment as something that was ignorable.

The propensity score matching process used a within-caliper matching approach; the caliper width was defined as 0.02 times the standard deviation of the propensity scores. This fine caliper was chosen because it did not cause a reduction in the number of matches while it provided a good balance between the matched samples. The matching procedure was a one-to-one match without replacement (Austin 2011a).

R.2.1.3. Matching Covariates

The propensity scores were calculated using logistic regression based on a list of selected matching covariates. The covariates included students' prior year scale score, which is the most predictive of current year's test performance. In addition, some key student-, school-, and district-level variables were selected.

The following covariates were used for estimating the propensity scores, and directly balanced throughout the process of propensity score matching:

- student prior year (grade $n - 1$) scale score;
- student gender;
- student racial/ethnic category;
- student English language learner (ELL) status;
- student disability (SWD) status;
- school-type (i.e., public, charter, non-public);
- district-level needs/resource capacity (NRC) code; and
- district-level region as specified by the joint management team definitions (JMT).

R.2.1.4. Judging Covariate Balance

The covariate balance between the matched PBT and CBT samples was evaluated after propensity score matching. The standardized difference (d) for each covariate between matched CBT and PBT groups was computed to evaluate the balance and effectiveness of propensity score matching. To the extent that the standardized differences approach zero, balance can be said to be reasonably achieved on the selected covariates.

The formulae of d are different for continuous and discrete variables; there are minor modifications for estimating covariate balance before and after matching samples (Rosenbaum, 2010). The traditional experimental design is still a useful framework for this comparability study, so CBT can be considered the "treatment" and PBT can be considered the "control" condition. The analysis of covariate balance for discrete variables differs in that it uses the unbiased variance estimator for a proportion. (See page 174 of Austin (2011a) for examples of a similar but not identical formula.)

For variable k being treated as **continuous**:

1. Estimate the means and variances for the treatment ($\bar{x}_{tk}$ and s^2_{tk}) and control groups ($\bar{x}_{ck}$ and s^2_{ck}) before matching.
2. Estimate the means only for the treatment ($\bar{x}_{tmk}$) and control groups ($\bar{x}_{cmk}$) after matching.
3. Estimate the standardized difference for variable k **before** matching as:

$$d_{bk} = (\bar{x}_{tk} - \bar{x}_{ck}) \Big/ \sqrt{\left(s^2_{tk} + s^2_{ck} \right)/2} \qquad (1)$$

4. Estimate the standardized difference for variable k—note the use of the pre-matched pooled standard deviation in the denominator—**after** matching as:

$$d_{mk} = (\bar{x}_{tmk} - \bar{x}_{cmk}) \Big/ \sqrt{\left(s^2_{tk} + s^2_{ck} \right)/2} \qquad (2)$$

For variable k being treated as **discrete**:

1. Estimate the proportions for the treatment (p_{tk}) and control groups (p_{ck}) before matching.
2. Estimate the proportions for the treatment (p_{tmk}) and control groups (p_{cmk}) after matching.
3. Estimate the standardized difference for variable k **before** matching as:

$$d_{bk} = (p_{tk} - p_{ck}) / \sqrt{[p_{tk}(1 - p_{tk}) + p_{ck}(1 - p_{ck})]/2} \quad (3)$$

4. Estimate the standardized difference for variable k—note the use of the pre-matched pooled standard deviation in the denominator—**after** matching as:

$$d_{mk} = (p_{tmk} - p_{cmk}) / \sqrt{[p_{tk}(1 - p_{tk}) + p_{ck}(1 - p_{ck})]/2} \quad (4)$$

R.2.2. Evaluating Test-level Mode Comparability

In order to evaluate and detect test-level mode effects, two things were examined after propensity score matching CBT and PBT students. Given that the items were the same between the test modes, the distribution of raw scores for the matched PBT and CBT samples was first reviewed. This enabled a direct means of detecting possible mode effects.

Next, the distribution of scale scores for the matched PBT and CBT samples were reviewed. The scale scores were derived using the single operational raw-score-to-scale-score (RSSS) conversion table, which was estimated based on all students in the operational calibration sample. The mode treatment effect was simply calculated as the difference in scale score means for the matched PBT and CBT samples.

R.2.3. Evaluating Item-level DIF

In addition to test-level mode analyses, item-level mode DIF analyses were performed to evaluate the consistency of item performance across modes. These analyses did not change or impact the test-level mode comparability. Instead, the mode DIF results were available to content specialists and test developers for informing improved item writing and test construction. The item-level mode DIF analyses also facilitate a better understanding of observed mode differences at the test level. It is noteworthy that only those items with a mode effect bigger than average over the full set will be flagged in the mode DIF analysis.

Classical DIF analyses are statistical methods for identifying items that are estimated to have functioned differently for one group as compared with another group (e.g., PBT vs. CBT students). First, the Mantel-Haenszel (MH) method (Holland & Thayer, 1988) was employed for multiple choice (MC) items. This non-parametric DIF method partitions the sample of examinees into categories based on total raw test scores. It compares the log-odds ratio of keyed responses for the focal and reference groups. The log-odd ratio was then transformed onto the delta-value metric to evaluate the practical significance. Second, the standardized mean difference (SMD) was computed for constructed response (CR) items. The SMD statistic compares the mean scores of reference and focal groups, after adjusting for proficiency differences. The SMD statistic was also evaluated for both statistical as well as practical significance.

Section R.3. Results

R.3.1. Propensity Score Matching

R.3.1.1. Covariate Balance

This study summarizes the covariate balance before and after matching. Standardized differences (*d*s) greater than 0.05 in absolute value after matching were bolded. The covariate balance was summarized in Tables R.3.1–R.3.10.

Very few covariates were flagged for having a standardized difference greater than 0.05; only two were flagged with a value greater than 0.20. The standardized difference for the covariate of prior year scale score between matched samples were all at or below 0.05 across the grades and subjects. In general, the propensity score matching generated well-matched PBT and CBT groups.

Table R.3.1. Covariate Balance Before and After Matching: ELA Grade 4

Variable	Value	Before Matching				After Matching			
		CBT	PBT	Δ	*d*	CBT	PBT	Δ	*d*
2017 OP Scale Score	(Mean)	307.80	309.65	-1.85	-0.06	311.41	310.21	1.20	0.04
Disability (%)	No	95.77	84.61	11.17	0.38	94.84	94.28	0.56	0.02
	Yes	4.23	15.39	-11.17	-0.38	5.16	5.72	-0.56	-0.02
ELL/MLL (%)	No	97.80	89.64	8.16	0.34	97.38	96.99	0.39	0.02
	Yes	2.20	10.36	-8.16	-0.34	2.62	3.01	-0.39	-0.02
Ethnicity (%)	Missing	12.82	0.00	12.82	0.54	0.00	0.00	0.00	0.00
	Asian	3.11	10.90	-7.79	-0.31	3.59	4.04	-0.45	-0.02
	Black	5.77	18.86	-13.09	-0.41	7.09	8.60	-1.51	-0.05
	Hispanic or Latino	12.11	29.05	-16.94	-0.43	14.40	14.22	0.18	0.00
	American Indian or Alaska Native	0.34	0.73	-0.39	-0.05	0.37	0.43	-0.06	-0.01
	Multi-racial	2.85	2.58	0.27	0.02	3.47	3.50	-0.03	-0.00
	Native Hawaiian or Other Pacific Islander	0.08	0.30	-0.22	-0.05	0.11	0.08	0.03	0.01
	White	62.92	37.59	25.33	0.52	70.97	69.13	1.84	0.04
Gender (%)	Female	48.94	49.40	-0.47	-0.01	50.51	50.42	0.09	0.00
	Male	51.06	50.60	0.47	0.01	49.49	49.58	-0.09	-0.00
Joint Management Team Region (JMT)%	New York City	0.87	46.01	-45.15	-1.26	1.02	1.55	-0.53	-0.01
	Long Island	4.81	10.57	-5.76	-0.22	4.37	4.44	-0.07	-0.00
	Lower Hudson Valley	13.17	6.52	6.65	0.22	12.10	10.54	1.56	0.05
	Mid-Hudson	10.51	3.60	6.92	0.27	10.80	10.73	0.07	0.00
	Capital District / North Country	19.05	6.14	12.9	0.40	22.21	20.90	1.31	0.04
	Central Region	0.64	2.60	-1.96	-0.16	0.81	1.07	-0.25	-0.02
	Mid-State	23.65	2.51	21.13	0.66	17.83	21.16	-3.34	**-0.10**
	Mid-South	3.95	2.65	1.30	0.07	4.48	4.37	0.11	0.01
	Mid-West	14.15	5.06	9.09	0.31	16.50	16.05	0.45	0.02
	West	7.45	6.31	1.14	0.05	8.71	7.72	0.99	0.04
	Missing	1.76	8.03	-6.26	-0.29	1.17	1.47	-0.30	-0.01
Needs/Resource Category (NRC) (%)	New York	0.01	40.48	-40.47	-1.17	0.00	0.01	0.00	0.00
	Big 4 Cities	13.09	3.40	9.69	0.36	11.74	11.81	-0.07	-0.00
	Urban/Suburban	6.12	7.50	-1.38	-0.05	6.17	6.43	-0.26	-0.01
	High Needs Rural	10.89	4.87	6.02	0.23	11.9	13.03	-1.14	-0.04
	Average Needs	61.45	18.61	42.85	0.97	61.59	60.33	1.27	0.03
	Low Needs	5.82	10.39	-4.57	-0.17	6.40	5.37	1.03	0.04
	Charter School	0.86	6.74	-5.88	-0.31	1.02	1.54	-0.52	-0.03
	Religious and Independent Schools	1.76	8.03	-6.26	-0.29	1.17	1.47	-0.30	-0.01
School Type (%)	Public	97.38	85.23	12.14	0.44	97.81	96.98	0.82	0.03
	Charter	0.86	6.74	-5.88	-0.31	1.02	1.54	-0.52	-0.03
	Non-Public	1.76	8.03	-6.26	-0.29	1.17	1.47	-0.30	-0.01

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

Table R.3.2. Covariate Balance Before and After Matching: ELA Grade 5

Variable	Value	Before Matching				After Matching			
		CBT	PBT	Δ	*d*	CBT	PBT	Δ	*d*
2017 OP Scale Score	(Mean)	304.96	308.18	-3.22	-0.09	308.69	307.99	0.70	0.02
Disability (%)	No	94.51	83.91	10.60	0.35	93.90	93.59	0.32	0.01
	Yes	5.49	16.09	-10.60	-0.35	6.10	6.41	-0.32	-0.01
ELL/MLL (%)	No	97.74	91.04	6.70	0.29	97.61	97.74	-0.13	-0.01
	Yes	2.26	8.96	-6.70	-0.29	2.39	2.26	0.13	0.01
Ethnicity (%)	Missing	12.08	0.00	12.08	0.52	0.00	0.00	0.00	0.00
	Asian	2.89	11.40	-8.51	-0.34	3.34	3.45	-0.11	-0.00
	Black	7.00	19.07	-12.07	-0.36	7.93	7.84	0.08	0.00
	Hispanic or Latino	13.54	29.19	-15.65	-0.39	14.6	13.19	1.41	0.04
	American Indian or Alaska Native	0.45	0.74	-0.29	-0.04	0.50	0.63	-0.13	-0.02
	Multi-racial	3.28	2.35	0.93	0.06	3.84	3.52	0.32	0.02
	Native Hawaiian or Other Pacific Islander	0.03	0.33	-0.30	-0.07	0.04	0.05	-0.01	-0.00
	White	60.74	36.92	23.82	0.49	69.74	71.32	-1.57	-0.03
Gender (%)	Female	49.09	49.59	-0.50	-0.01	50.63	49.85	0.78	0.02
	Male	50.91	50.41	0.50	0.01	49.37	50.15	-0.78	-0.02
Joint Management Team Region (JMT)%	New York City	0.74	48.12	-47.38	-1.32	0.84	1.13	-0.29	-0.01
	Long Island	4.19	10.65	-6.46	-0.25	4.02	4.10	-0.08	-0.00
	Lower Hudson Valley	13.97	6.73	7.24	0.24	12.19	9.31	2.88	**0.10**
	Mid-Hudson	11.57	3.50	8.07	0.31	11.52	11.05	0.47	0.02
	Capital District / North Country	15.33	6.46	8.87	0.29	16.42	17.43	-1.01	-0.03
	Central Region	0.99	2.30	-1.30	-0.10	1.01	0.84	0.18	0.01
	Mid-State	17.95	3.06	14.89	0.50	17.17	20.42	-3.25	**-0.11**
	Mid-South	4.70	2.61	2.09	0.11	4.68	4.44	0.24	0.01
	Mid-West	19.37	4.58	14.79	0.47	20.06	18.39	1.67	0.05
	West	9.96	6.16	3.80	0.14	10.7	11.41	-0.71	-0.03
	Missing	1.23	5.85	-4.62	-0.25	1.39	1.47	-0.08	-0.00
Needs/Resource Category (NRC) (%)	New York	0.00	42.53	-42.53	-1.22	0.00	0.01	0.00	0.00
	Big 4 Cities	14.02	3.33	10.69	0.39	11.84	10.55	1.29	0.05
	Urban/Suburban	6.68	7.14	-0.47	-0.02	6.24	5.35	0.89	0.03
	High Needs Rural	12.89	4.57	8.32	0.30	13.43	14.72	-1.29	-0.05
	Average Needs	59.35	18.71	40.64	0.92	61.08	61.75	-0.67	-0.02
	Low Needs	5.09	10.96	-5.87	-0.22	5.19	5.03	0.16	0.01
	Charter School	0.74	6.91	-6.17	-0.33	0.84	1.12	-0.28	-0.02
	Religious and Independent Schools	1.23	5.85	-4.62	-0.25	1.39	1.47	-0.08	-0.00
School Type (%)	Public	98.02	87.24	10.79	0.42	97.77	97.4	0.37	0.01
	Charter	0.74	6.91	-6.17	-0.33	0.84	1.12	-0.28	-0.02
	Non-Public	1.23	5.85	-4.62	-0.25	1.39	1.47	-0.08	-0.00

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

Table R.3.3. Covariate Balance Before and After Matching: ELA Grade 6

Variable	Value	Before Matching				After Matching			
		CBT	PBT	Δ	*d*	CBT	PBT	Δ	*d*
2017 OP Scale Score	(Mean)	302.36	303.23	-0.87	-0.02	305.63	305.66	-0.03	-0.00
Disability (%)	No	94.28	84.00	10.27	0.33	93.85	93.17	0.68	0.02
	Yes	5.72	16.00	-10.27	-0.33	6.15	6.83	-0.68	-0.02
ELL/MLL (%)	No	98.03	91.31	6.72	0.30	97.92	97.99	-0.06	-0.00
	Yes	1.97	8.69	-6.72	-0.30	2.08	2.01	0.06	0.00
Ethnicity (%)	Missing	11.20	0.00	11.20	0.50	0.00	0.00	0.00	0.00
	Asian	3.90	11.43	-7.53	-0.29	4.55	4.07	0.48	0.02
	Black	6.67	19.90	-13.22	-0.40	7.24	7.21	0.03	0.00
	Hispanic or Latino	12.78	29.36	-16.58	-0.42	14.42	12.55	1.87	0.05
	American Indian or Alaska Native	0.74	0.66	0.07	0.01	0.72	0.44	0.27	0.03
	Multi-racial	2.80	2.11	0.69	0.04	3.21	3.13	0.08	0.00
	Native Hawaiian or Other Pacific Islander	0.10	0.39	-0.29	-0.06	0.12	0.12	0.00	0.00
	White	61.82	36.15	25.67	0.53	69.74	72.47	-2.73	**-0.06**
Gender (%)	Female	48.64	49.25	-0.61	-0.01	49.67	49.72	-0.05	-0.00
	Male	51.36	50.75	0.61	0.01	50.33	50.28	0.05	0.00
Joint Management Team Region (JMT)%	New York City	0.36	48.61	-48.25	-1.36	0.36	0.38	-0.03	-0.00
	Long Island	7.76	9.99	-2.23	-0.08	7.42	8.30	-0.88	-0.03
	Lower Hudson Valley	11.90	6.66	5.24	0.18	12.67	9.13	3.54	**0.12**
	Mid-Hudson	12.41	3.07	9.34	0.35	10.83	12.43	-1.60	**-0.06**
	Capital District / North Country	15.27	5.89	9.38	0.31	16.43	14.16	2.27	**0.07**
	Central Region	1.90	2.24	-0.34	-0.02	1.81	1.45	0.36	0.03
	Mid-State	13.50	3.11	10.39	0.38	14.47	15.39	-0.92	-0.03
	Mid-South	6.65	2.25	4.40	0.21	6.68	6.01	0.68	0.03
	Mid-West	19.01	4.22	14.79	0.47	19.34	21.62	-2.29	**-0.07**
	West	8.23	5.92	2.31	0.09	7.97	8.82	-0.85	-0.03
	Missing	3.01	8.05	-5.04	-0.22	2.03	2.31	-0.27	-0.01
Needs/Resource Category (NRC) (%)	New York	0.01	42.47	-42.46	-1.21	0.00	0.01	0.00	0.00
	Big 4 Cities	11.48	3.10	8.38	0.33	11.79	8.91	2.88	**0.11**
	Urban/Suburban	5.45	7.21	-1.76	-0.07	4.77	4.31	0.45	0.02
	High Needs Rural	14.74	4.10	10.63	0.37	14.70	15.03	-0.33	-0.01
	Average Needs	56.40	17.06	39.34	0.89	57.53	60.89	-3.36	**-0.08**
	Low Needs	8.57	10.52	-1.96	-0.07	8.83	8.17	0.66	0.02
	Charter School	0.35	7.50	-7.14	-0.37	0.36	0.37	-0.01	-0.00
	Religious and Independent Schools	3.01	8.05	-5.04	-0.22	2.03	2.31	-0.27	-0.01
School Type (%)	Public	96.63	84.46	12.18	0.43	97.61	97.32	0.29	0.01
	Charter	0.35	7.50	-7.14	-0.37	0.36	0.37	-0.01	-0.00
	Non-Public	3.01	8.05	-5.04	-0.22	2.03	2.31	-0.27	-0.01

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

Table R.3.4. Covariate Balance Before and After Matching: ELA Grade 7

Variable	Value	Before Matching				After Matching			
		CBT	PBT	Δ	*d*	CBT	PBT	Δ	*d*
2017 OP Scale Score	(Mean)	301.77	300.45	1.32	0.04	305.05	304.22	0.83	0.02
Disability (%)	No	94.22	83.30	10.92	0.35	93.32	93.00	0.32	0.01
	Yes	5.78	16.70	-10.92	-0.35	6.68	7.00	-0.32	-0.01
ELL/MLL (%)	No	97.82	92.10	5.72	0.26	97.57	97.49	0.09	0.00
	Yes	2.18	7.90	-5.72	-0.26	2.43	2.51	-0.09	-0.00
Ethnicity (%)	Missing	11.06	0.00	11.06	0.50	0.00	0.00	0.00	0.00
	Asian	4.00	11.72	-7.72	-0.29	4.64	4.81	-0.17	-0.01
	Black	7.02	20.37	-13.35	-0.40	8.21	8.14	0.07	0.00
	Hispanic or Latino	14.01	28.69	-14.68	-0.36	12.64	14.92	-2.28	**-0.06**
	American Indian or Alaska Native	0.40	0.82	-0.42	-0.05	0.44	0.38	0.06	0.01
	Multi-racial	2.37	1.85	0.52	0.04	2.58	2.51	0.08	0.01
	Native Hawaiian or Other Pacific Islander	0.06	0.32	-0.26	-0.06	0.07	0.07	0.00	0.00
	White	61.06	36.22	24.84	0.51	71.43	69.18	2.25	0.05
Gender (%)	Female	48.26	48.77	-0.51	-0.01	49.55	49.26	0.29	0.01
	Male	51.74	51.23	0.51	0.01	50.45	50.74	-0.29	-0.01
Joint Management Team Region (JMT)%	New York City	0.42	51.98	-51.56	-1.45	0.57	0.62	-0.05	-0.00
	Long Island	8.00	9.44	-1.44	-0.05	8.13	8.33	-0.20	-0.01
	Lower Hudson Valley	13.74	6.57	7.17	0.24	8.65	10.51	-1.86	**-0.06**
	Mid-Hudson	12.65	3.02	9.63	0.36	12.29	11.54	0.75	0.03
	Capital District / North Country	17.65	6.37	11.28	0.35	19.23	17.72	1.51	0.05
	Central Region	2.44	2.18	0.26	0.02	2.69	2.61	0.09	0.01
	Mid-State	13.67	3.33	10.33	0.38	14.81	15.24	-0.43	-0.02
	Mid-South	6.46	2.73	3.74	0.18	7.28	6.01	1.27	**0.06**
	Mid-West	15.17	4.69	10.48	0.36	15.76	15.54	0.22	0.01
	West	7.69	5.98	1.71	0.07	8.20	9.48	-1.29	-0.05
	Missing	2.11	3.71	-1.60	-0.10	2.39	2.40	-0.01	-0.00
Needs/Resource Category (NRC) (%)	New York	0.00	45.91	-45.91	-1.30	0.00	0.00	0.00	0.00
	Big 4 Cities	14.09	2.95	11.15	0.41	8.33	11.28	-2.95	**-0.11**
	Urban/Suburban	5.66	7.11	-1.44	-0.06	5.98	6.56	-0.58	-0.02
	High Needs Rural	11.71	4.65	7.07	0.26	12.25	11.91	0.34	0.01
	Average Needs	59.09	16.82	42.27	0.97	63.33	59.88	3.45	**0.08**
	Low Needs	6.92	11.40	-4.48	-0.16	7.14	7.34	-0.20	-0.01
	Charter School	0.42	7.46	-7.04	-0.37	0.57	0.62	-0.05	-0.00
	Religious and Independent Schools	2.11	3.71	-1.60	-0.10	2.39	2.40	-0.01	-0.00
School Type (%)	Public	97.47	88.83	8.64	0.35	97.03	96.98	0.06	0.00
	Charter	0.42	7.46	-7.04	-0.37	0.57	0.62	-0.05	-0.00
	Non-Public	2.11	3.71	-1.60	-0.10	2.39	2.40	-0.01	-0.00

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

Table R.3.5. Covariate Balance Before and After Matching: ELA Grade 8

Variable	Value	Before Matching				After Matching			
		CBT	PBT	Δ	*d*	CBT	PBT	Δ	*d*
2017 OP Scale Score	(Mean)	308.06	309.54	-1.47	-0.05	310.90	310.87	0.03	0.00
Disability (%)	No	92.96	84.36	8.61	0.27	92.12	92.40	-0.29	-0.01
	Yes	7.04	15.64	-8.61	-0.27	7.88	7.60	0.29	0.01
ELL/MLL (%)	No	97.62	92.26	5.36	0.25	97.56	97.69	-0.13	-0.01
	Yes	2.38	7.74	-5.36	-0.25	2.44	2.31	0.13	0.01
Ethnicity (%)	Missing	10.15	0.00	10.15	0.48	0.00	0.00	0.00	0.00
	Asian	4.39	12.32	-7.93	-0.29	5.05	4.78	0.27	0.01
	Black	6.79	20.57	-13.78	-0.41	7.71	7.04	0.66	0.02
	Hispanic or Latino	14.32	28.44	-14.12	-0.35	12.60	13.81	-1.21	-0.03
	American Indian or Alaska Native	0.46	0.80	-0.34	-0.04	0.49	0.46	0.03	0.00
	Multi-racial	2.13	1.51	0.62	0.05	2.38	2.29	0.09	0.01
	Native Hawaiian or Other Pacific Islander	0.06	0.33	-0.26	-0.06	0.06	0.06	0.01	0.00
	White	61.70	36.03	25.67	0.53	71.71	71.56	0.16	0.00
Gender (%)	Female	48.26	48.67	-0.41	-0.01	48.91	49.55	-0.64	-0.01
	Male	51.74	51.33	0.41	0.01	51.09	50.45	0.64	0.01
Joint Management Team Region (JMT)%	New York City	0.34	51.09	-50.76	-1.43	0.39	0.39	0.00	0.00
	Long Island	8.02	8.41	-0.39	-0.01	7.20	7.87	-0.67	-0.02
	Lower Hudson Valley	13.42	6.44	6.97	0.23	10.32	10.04	0.28	0.01
	Mid-Hudson	12.66	2.81	9.85	0.38	12.27	11.08	1.19	0.05
	Capital District / North Country	18.81	5.91	12.90	0.40	20.66	19.88	0.78	0.02
	Central Region	1.27	2.14	-0.87	-0.07	1.37	1.43	-0.06	-0.00
	Mid-State	13.18	3.37	9.80	0.36	14.40	14.10	0.30	0.01
	Mid-South	6.98	2.28	4.71	0.23	7.44	6.90	0.54	0.03
	Mid-West	16.13	4.19	11.94	0.40	16.59	18.02	-1.43	-0.05
	West	6.78	5.72	1.05	0.04	6.90	7.52	-0.63	-0.03
	Missing	2.42	7.63	-5.21	-0.24	2.47	2.78	-0.30	-0.01
Needs/Resource Category (NRC) (%)	New York	0.00	45.47	-45.47	-1.29	0.00	0.00	0.00	0.00
	Big 4 Cities	13.90	3.05	10.85	0.40	9.77	11.45	-1.67	**-0.06**
	Urban/Suburban	7.12	6.13	0.99	0.04	6.71	7.12	-0.40	-0.02
	High Needs Rural	11.07	4.59	6.48	0.24	11.41	10.81	0.60	0.02
	Average Needs	56.27	15.58	40.69	0.94	59.82	57.80	2.01	0.05
	Low Needs	8.89	10.61	-1.73	-0.06	9.43	9.66	-0.23	-0.01
	Charter School	0.34	6.94	-6.61	-0.36	0.39	0.39	0.00	0.00
	Religious and Independent Schools	2.42	7.63	-5.21	-0.24	2.47	2.78	-0.30	-0.01
School Type (%)	Public	97.25	85.43	11.82	0.43	97.14	96.84	0.30	0.01
	Charter	0.34	6.94	-6.61	-0.36	0.39	0.39	0.00	0.00
	Non-Public	2.42	7.63	-5.21	-0.24	2.47	2.78	-0.30	-0.01

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

Table R.3.6. Covariate Balance Before and After Matching: Mathematics Grade 4

Variable	Value	Before Matching				After Matching			
		CBT	PBT	Δ	*d*	CBT	PBT	Δ	*d*
2017 OP Scale Score	(Mean)	308.65	310.43	-1.78	-0.05	312.88	313.60	-0.73	-0.02
Disability (%)	No	96.19	85.26	10.94	0.38	95.34	94.96	0.38	0.01
	Yes	3.81	14.74	-10.94	-0.38	4.66	5.04	-0.38	-0.01
ELL/MLL (%)	No	97.63	88.99	8.64	0.35	97.10	97.39	-0.30	-0.01
	Yes	2.37	11.01	-8.64	-0.35	2.90	2.61	0.30	0.01
Ethnicity (%)	Missing	14.08	0.00	14.08	0.57	0.00	0.00	0.00	0.00
	Asian	2.77	11.20	-8.43	-0.34	2.92	2.86	0.06	0.00
	Black	6.15	18.34	-12.19	-0.38	7.40	7.23	0.17	0.01
	Hispanic or Latino	14.15	29.44	-15.29	-0.38	10.17	11.63	-1.46	-0.04
	American Indian or Alaska Native	0.34	0.69	-0.35	-0.05	0.35	0.38	-0.04	-0.01
	Multi-racial	2.78	2.63	0.15	0.01	3.28	2.99	0.28	0.02
	Native Hawaiian or Other Pacific Islander	0.08	0.30	-0.22	-0.05	0.09	0.14	-0.05	-0.01
	White	59.65	37.41	22.24	0.46	75.80	74.77	1.03	0.02
Gender (%)	Female	48.69	49.19	-0.50	-0.01	50.25	49.70	0.56	0.01
	Male	51.31	50.81	0.50	0.01	49.75	50.30	-0.56	-0.01
Joint Management Team Region (JMT)%	New York City	1.20	46.15	-44.95	-1.25	1.50	1.72	-0.22	-0.01
	Long Island	3.59	10.95	-7.37	-0.29	3.61	4.71	-1.10	-0.04
	Lower Hudson Valley	18.81	6.43	12.38	0.38	7.68	8.34	-0.67	-0.02
	Mid-Hudson	13.23	3.65	9.58	0.35	13.94	14.10	-0.16	-0.01
	Capital District / North Country	16.23	6.86	9.37	0.30	19.61	18.43	1.19	0.04
	Central Region	0.87	2.61	-1.75	-0.13	1.15	0.99	0.16	0.01
	Mid-State	22.28	2.39	19.89	0.63	25.76	24.97	0.79	0.03
	Mid-South	4.62	2.95	1.68	0.09	5.35	5.07	0.28	0.01
	Mid-West	10.43	5.71	4.72	0.17	12.08	12.05	0.02	0.00
	West	6.31	6.66	-0.35	-0.01	7.56	8.37	-0.80	-0.03
	Missing	2.43	5.63	-3.20	-0.16	1.77	1.26	0.51	0.03
Needs/Resource Category (NRC) (%)	New York	0.01	40.79	-40.78	-1.17	0.01	0.12	-0.11	-0.00
	Big 4 Cities	17.71	2.63	15.08	0.52	5.61	8.16	-2.55	**-0.09**
	Urban/Suburban	7.00	7.89	-0.89	-0.03	7.23	7.40	-0.17	-0.01
	High Needs Rural	10.88	5.18	5.70	0.21	12.77	13.30	-0.53	-0.02
	Average Needs	57.87	20.53	37.34	0.83	68.13	64.21	3.92	**0.09**
	Low Needs	2.91	10.80	-7.89	-0.32	3.00	3.94	-0.94	-0.04
	Charter School	1.19	6.55	-5.36	-0.28	1.48	1.61	-0.12	-0.01
	Religious and Independent Schools	2.43	5.63	-3.20	-0.16	1.77	1.26	0.51	0.03
School Type (%)	Public	96.38	87.82	8.56	0.32	96.75	97.13	-0.38	-0.01
	Charter	1.19	6.55	-5.36	-0.28	1.48	1.61	-0.12	-0.01
	Non-Public	2.43	5.63	-3.20	-0.16	1.77	1.26	0.51	0.03

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

Table R.3.7. Covariate Balance Before and After Matching: Mathematics Grade 5

Variable	Value	Before Matching				After Matching			
		CBT	PBT	Δ	*d*	CBT	PBT	Δ	*d*
2017 OP Scale Score	(Mean)	307.95	307.08	0.88	0.02	311.93	309.91	2.02	0.05
Disability (%)	No	95.60	84.67	10.93	0.37	94.79	94.29	0.50	0.02
	Yes	4.40	15.33	-10.93	-0.37	5.21	5.71	-0.50	-0.02
ELL/MLL (%)	No	97.50	90.72	6.78	0.29	97.52	97.07	0.46	0.02
	Yes	2.50	9.28	-6.78	-0.29	2.48	2.93	-0.46	-0.02
Ethnicity (%)	Missing	10.78	0.00	10.78	0.49	0.00	0.00	0.00	0.00
	Asian	3.45	11.64	-8.19	-0.31	3.54	3.81	-0.27	-0.01
	Black	7.92	18.44	-10.52	-0.31	7.79	8.37	-0.58	-0.02
	Hispanic or Latino	15.99	29.60	-13.61	-0.33	11.16	14.35	-3.19	**-0.08**
	American Indian or Alaska Native	0.45	0.74	-0.30	-0.04	0.43	0.51	-0.08	-0.01
	Multi-racial	3.28	2.35	0.93	0.06	4.11	3.66	0.44	0.03
	Native Hawaiian or Other Pacific Islander	0.06	0.34	-0.29	-0.06	0.05	0.08	-0.03	-0.01
	White	58.08	36.88	21.20	0.43	72.91	69.21	3.70	**0.08**
Gender (%)	Female	48.83	49.38	-0.55	-0.01	49.57	49.03	0.54	0.01
	Male	51.17	50.62	0.55	0.01	50.43	50.97	-0.54	-0.01
Joint Management Team Region (JMT)%	New York City	1.07	48.37	-47.30	-1.31	1.36	1.64	-0.28	-0.01
	Long Island	4.60	10.72	-6.12	-0.23	5.01	5.72	-0.71	-0.03
	Lower Hudson Valley	18.03	6.67	11.36	0.35	4.13	10.34	-6.21	**-0.19**
	Mid-Hudson	13.96	3.57	10.39	0.37	15.90	12.96	2.93	**0.11**
	Capital District / North Country	17.70	6.74	10.96	0.34	21.49	19.60	1.88	**0.06**
	Central Region	1.54	2.48	-0.94	-0.07	1.91	2.21	-0.30	-0.02
	Mid-State	16.49	2.76	13.73	0.48	19.32	15.24	4.08	**0.14**
	Mid-South	6.04	2.85	3.18	0.15	7.27	6.13	1.14	**0.06**
	Mid-West	11.99	5.70	6.29	0.22	13.37	15.11	-1.74	**-0.06**
	West	6.98	6.64	0.34	0.01	8.32	8.71	-0.39	-0.02
	Missing	1.60	3.50	-1.90	-0.12	1.93	2.34	-0.42	-0.03
Needs/Resource Category (NRC) (%)	New York	0.00	42.93	-42.93	-1.23	0.00	0.00	0.00	0.00
	Big 4 Cities	18.12	2.54	15.57	0.53	3.62	9.69	-6.07	**-0.21**
	Urban/Suburban	8.79	7.59	1.20	0.04	9.64	9.95	-0.31	-0.01
	High Needs Rural	11.98	4.98	7.00	0.25	14.30	13.18	1.12	0.04
	Average Needs	53.14	20.75	32.39	0.71	62.84	55.80	7.04	**0.15**
	Low Needs	5.29	11.06	-5.77	-0.21	6.31	7.39	-1.08	-0.04
	Charter School	1.07	6.64	-5.57	-0.29	1.36	1.64	-0.28	-0.01
	Religious and Independent Schools	1.60	3.50	-1.90	-0.12	1.93	2.34	-0.42	-0.03
School Type (%)	Public	97.33	89.86	7.47	0.31	96.72	96.02	0.70	0.03
	Charter	1.07	6.64	-5.57	-0.29	1.36	1.64	-0.28	-0.01
	Non-Public	1.60	3.50	-1.90	-0.12	1.93	2.34	-0.42	-0.03

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

Table R.3.8. Covariate Balance Before and After Matching: Mathematics Grade 6

Variable	Value	Before Matching				After Matching			
		CBT	PBT	Δ	*d*	CBT	PBT	Δ	*d*
2017 OP Scale Score	(Mean)	312.37	309.69	2.68	0.07	314.89	315.41	-0.53	-0.01
Disability (%)	No	95.72	84.70	11.02	0.38	95.16	95.36	-0.19	-0.01
	Yes	4.28	15.30	-11.02	-0.38	4.84	4.64	0.19	0.01
ELL/MLL (%)	No	98.29	90.61	7.68	0.34	97.86	98.49	-0.63	-0.03
	Yes	1.71	9.39	-7.68	-0.34	2.14	1.51	0.63	0.03
Ethnicity (%)	Missing	9.36	0.00	9.36	0.45	0.00	0.00	0.00	0.00
	Asian	3.87	11.70	-7.83	-0.30	4.21	4.75	-0.54	-0.02
	Black	6.47	19.36	-12.88	-0.39	7.50	7.22	0.28	0.01
	Hispanic or Latino	13.96	29.41	-15.46	-0.38	11.2	12.21	-1.01	-0.02
	American Indian or Alaska Native	0.72	0.69	0.03	0.00	0.75	0.76	-0.01	-0.00
	Multi-racial	3.05	2.13	0.92	0.06	3.52	3.33	0.18	0.01
	Native Hawaiian or Other Pacific Islander	0.08	0.39	-0.31	-0.06	0.10	0.09	0.01	0.00
	White	62.49	36.32	26.17	0.54	72.72	71.65	1.08	0.02
Gender (%)	Female	48.84	49.04	-0.20	-0.00	49.36	49.37	-0.01	-0.00
	Male	51.16	50.96	0.20	0.00	50.64	50.63	0.01	0.00
Joint Management Team Region (JMT)%	New York City	0.47	48.35	-47.88	-1.34	0.59	0.68	-0.09	-0.00
	Long Island	7.13	10.16	-3.04	-0.11	7.50	7.91	-0.41	-0.01
	Lower Hudson Valley	14.48	6.69	7.79	0.26	7.67	8.96	-1.29	-0.04
	Mid-Hudson	12.33	3.41	8.93	0.34	11.57	9.92	1.65	**0.06**
	Capital District / North Country	14.27	6.67	7.60	0.25	16.55	16.85	-0.29	-0.01
	Central Region	2.48	2.25	0.22	0.01	2.66	2.18	0.48	0.03
	Mid-State	14.73	2.50	12.23	0.45	17.29	18.31	-1.03	-0.04
	Mid-South	10.24	2.31	7.93	0.33	11.55	9.51	2.04	**0.09**
	Mid-West	16.20	5.09	11.12	0.37	17.35	17.92	-0.57	-0.02
	West	5.02	6.39	-1.37	-0.06	5.53	5.67	-0.14	-0.01
	Missing	2.65	6.18	-3.53	-0.17	1.74	2.08	-0.35	-0.02
Needs/Resource Category (NRC) (%)	New York	0.01	42.35	-42.35	-1.21	0.00	0.02	0.00	0.00
	Big 4 Cities	13.94	2.40	11.54	0.43	6.41	7.66	-1.25	-0.05
	Urban/Suburban	5.54	7.50	-1.96	-0.08	5.14	5.27	-0.13	-0.01
	High Needs Rural	14.81	4.50	10.31	0.35	16.87	16.20	0.67	0.02
	Average Needs	55.96	18.87	37.09	0.83	61.74	59.80	1.94	0.04
	Low Needs	6.64	10.88	-4.24	-0.15	7.51	8.30	-0.79	-0.03
	Charter School	0.47	7.32	-6.86	-0.36	0.59	0.66	-0.07	-0.00
	Religious and Independent Schools	2.65	6.18	-3.53	-0.17	1.74	2.08	-0.35	-0.02
School Type (%)	Public	96.89	86.50	10.39	0.38	97.67	97.26	0.42	0.02
	Charter	0.47	7.32	-6.86	-0.36	0.59	0.66	-0.07	-0.00
	Non-Public	2.65	6.18	-3.53	-0.17	1.74	2.08	-0.35	-0.02

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

Table R.3.9. Covariate Balance Before and After Matching: Mathematics Grade 7

Variable	Value	Before Matching				After Matching			
		CBT	PBT	Δ	*d*	CBT	PBT	Δ	*d*
2017 OP Scale Score	(Mean)	309.29	305.74	3.55	0.09	313.62	311.96	1.66	0.04
Disability (%)	No	96.19	84.38	11.81	0.41	95.38	94.78	0.61	0.02
	Yes	3.81	15.62	-11.81	-0.41	4.62	5.22	-0.61	-0.02
ELL/MLL (%)	No	97.33	91.62	5.71	0.25	97.58	97.28	0.30	0.01
	Yes	2.67	8.38	-5.71	-0.25	2.42	2.72	-0.30	-0.01
Ethnicity (%)	Missing	9.36	0.00	9.36	0.45	0.00	0.00	0.00	0.00
	Asian	4.00	11.85	-7.85	-0.29	4.23	4.39	-0.17	-0.01
	Black	8.10	19.63	-11.53	-0.34	7.66	9.01	-1.35	-0.04
	Hispanic or Latino	16.96	29.23	-12.27	-0.29	11.90	14.56	-2.67	**-0.06**
	American Indian or Alaska Native	0.43	0.80	-0.37	-0.05	0.46	0.44	0.01	0.00
	Multi-racial	2.23	1.87	0.36	0.03	2.40	2.28	0.12	0.01
	Native Hawaiian or Other Pacific Islander	0.06	0.33	-0.27	-0.06	0.08	0.11	-0.03	-0.01
	White	58.87	36.30	22.57	0.46	73.28	69.20	4.08	**0.08**
Gender (%)	Female	48.20	48.60	-0.40	-0.01	48.82	48.83	-0.01	-0.00
	Male	51.80	51.40	0.40	0.01	51.18	51.17	0.01	0.00
Joint Management Team Region (JMT)%	New York City	0.60	51.11	-50.51	-1.41	0.88	1.01	-0.12	-0.00
	Long Island	8.40	9.24	-0.84	-0.03	9.35	10.67	-1.31	-0.05
	Lower Hudson Valley	18.45	6.39	12.06	0.37	4.64	9.62	-4.97	**-0.15**
	Mid-Hudson	13.02	3.23	9.79	0.36	14.07	10.52	3.55	**0.13**
	Capital District / North Country	14.14	6.93	7.21	0.24	17.01	16.08	0.93	0.03
	Central Region	3.09	2.12	0.96	0.06	3.80	4.30	-0.50	-0.03
	Mid-State	15.26	2.70	12.56	0.45	18.74	14.14	4.60	**0.16**
	Mid-South	6.34	2.82	3.52	0.17	7.74	7.59	0.15	0.01
	Mid-West	12.90	5.12	7.77	0.27	14.20	14.90	-0.69	-0.02
	West	5.99	6.17	-0.18	-0.01	7.34	8.43	-1.09	-0.05
	Missing	1.82	4.16	-2.34	-0.14	2.22	2.76	-0.54	-0.03
Needs/Resource Category (NRC) (%)	New York	0.01	45.26	-45.25	-1.29	0.01	0.03	-0.01	-0.00
	Big 4 Cities	19.48	2.05	17.43	0.59	4.50	10.02	-5.51	**-0.19**
	Urban/Suburban	6.37	6.97	-0.60	-0.02	7.67	7.53	0.14	0.01
	High Needs Rural	12.40	4.71	7.68	0.28	14.67	14.12	0.55	0.02
	Average Needs	53.82	18.40	35.43	0.79	63.73	57.12	6.60	**0.15**
	Low Needs	5.51	11.24	-5.73	-0.21	6.31	7.43	-1.12	-0.04
	Charter School	0.60	7.21	-6.61	-0.35	0.87	0.98	-0.11	-0.01
	Religious and Independent Schools	1.82	4.16	-2.34	-0.14	2.22	2.76	-0.54	-0.03
School Type (%)	Public	97.58	88.63	8.95	0.36	96.90	96.26	0.65	0.03
	Charter	0.60	7.21	-6.61	-0.35	0.87	0.98	-0.11	-0.01
	Non-Public	1.82	4.16	-2.34	-0.14	2.22	2.76	-0.54	-0.03

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

Table R.3.10. Covariate Balance Before and After Matching: Mathematics Grade 8

Variable	Value	Before Matching				After Matching			
		CBT	PBT	Δ	*d*	CBT	PBT	Δ	*d*
2017 OP Scale Score	(Mean)	297.92	300.63	-2.72	-0.08	302.15	300.74	1.40	0.04
Disability (%)	No	93.56	82.04	11.53	0.36	92.29	91.76	0.53	0.02
	Yes	6.44	17.96	-11.53	-0.36	7.71	8.24	-0.53	-0.02
ELL/MLL (%)	No	95.64	90.20	5.45	0.21	97.53	96.83	0.70	0.03
	Yes	4.36	9.80	-5.45	-0.21	2.47	3.17	-0.70	-0.03
Ethnicity (%)	Missing	8.70	0.00	8.70	0.44	0.00	0.00	0.00	0.00
	Asian	2.97	10.33	-7.36	-0.30	2.86	3.65	-0.79	-0.03
	Black	9.54	21.33	-11.78	-0.33	8.31	9.27	-0.96	-0.03
	Hispanic or Latino	22.53	31.53	-9.00	-0.20	13.84	16.19	-2.35	-0.05
	American Indian or Alaska Native	0.53	0.72	-0.19	-0.02	0.48	0.50	-0.02	-0.00
	Multi-racial	1.66	1.51	0.15	0.01	1.95	2.02	-0.07	-0.01
	Native Hawaiian or Other Pacific Islander	0.06	0.33	-0.27	-0.06	0.05	0.14	-0.10	-0.02
	White	54.02	34.25	19.76	0.41	72.52	68.22	4.30	**0.09**
Gender (%)	Female	46.78	47.74	-0.95	-0.02	46.34	45.47	0.86	0.02
	Male	53.22	52.26	0.95	0.02	53.66	54.53	-0.86	-0.02
Joint Management Team Region (JMT)%	New York City	0.18	52.65	-52.47	-1.48	0.26	0.36	-0.10	-0.00
	Long Island	5.26	5.10	0.16	0.01	5.67	6.70	-1.03	-0.05
	Lower Hudson Valley	24.82	6.18	18.64	0.53	5.89	11.51	-5.62	**-0.16**
	Mid-Hudson	12.77	3.02	9.75	0.37	15.16	11.22	3.94	**0.15**
	Capital District / North Country	17.06	6.63	10.43	0.33	22.96	22.41	0.55	0.02
	Central Region	1.49	2.33	-0.84	-0.06	1.97	2.38	-0.41	-0.03
	Mid-State	13.86	2.86	11.00	0.41	17.87	12.25	5.62	**0.21**
	Mid-South	5.19	3.07	2.12	0.11	7.11	8.00	-0.89	-0.04
	Mid-West	11.62	4.37	7.25	0.27	14.12	13.36	0.77	0.03
	West	5.04	6.84	-1.80	-0.08	6.03	7.90	-1.87	**-0.08**
	Missing	2.71	6.96	-4.24	-0.20	2.95	3.92	-0.96	-0.05
Needs/Resource Category (NRC) (%)	New York	0.00	47.31	-47.31	-1.34	0.00	0.19	0.00	0.00
	Big 4 Cities	25.28	2.66	22.62	0.69	3.68	10.04	-6.37	**-0.19**
	Urban/Suburban	8.21	7.00	1.21	0.05	10.59	10.16	0.43	0.02
	High Needs Rural	11.89	5.43	6.45	0.23	15.83	16.00	-0.17	-0.01
	Average Needs	49.09	15.95	33.13	0.76	63.06	54.65	8.41	**0.19**
	Low Needs	2.64	8.14	-5.49	-0.25	3.63	4.88	-1.25	**-0.06**
	Charter School	0.18	6.54	-6.36	-0.36	0.26	0.17	0.10	0.01
	Religious and Independent Schools	2.71	6.96	-4.24	-0.20	2.95	3.92	-0.96	-0.05
School Type (%)	Public	97.11	86.50	10.61	0.39	96.78	95.92	0.86	0.03
	Charter	0.18	6.54	-6.36	-0.36	0.26	0.17	0.10	0.01
	Non-Public	2.71	6.96	-4.24	-0.20	2.95	3.92	-0.96	-0.05

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

R.3.2. Test-level Mode Comparability

After having achieved a reasonably good covariate balance between the matched CBT and PBT samples, the test-level mode comparability was evaluated. Questar calculated the sample means for each matched sample and their standardized differences before and after matching for the following variables:

- 2017 Scale Score (SS): the prior year (grade n – 1) scale score, which was the proxy for prior ability that was entered as a key predictor into the propensity score model
- 2018 Raw Score (RS): the current year operational raw score
- 2018 Scale Score (SS): the current year scale score

The test-level performance before and after matching is summarized in Tables R.3.11–R.3.14. The results show that after matching, the PBT group had slightly higher test scores than the CBT group across the tests. However, the mode effects were small; only three tests were flagged having a standardized difference with an absolute value greater than 0.05: ELA Grade 5 (delta = -1.49, *d*= -0.08), Math Grade 6 (delta= -1.60, *d*= -0.09), and Math Grade 8 (delta = -0.90, *d*= -0.05). None had *d* greater than 0.10.

Table R.3.11. Test-level Performance between Test Modes Before Matching – ELA

		PBT			CBT				
Test	**Variable**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**Delta**	**d**
	2017 SS	146350	309.65	34.44	14581	307.80	31.94	-1.85	-0.06
ELA4	2018 RS	165390	19.57	7.09	16282	18.34	6.60	-1.23	-0.18
	2018 SS	165390	600.26	20.16	16282	596.77	18.21	-3.50	-0.18
	2017 SS	144369	308.18	34.98	13927	304.96	32.79	-3.22	-0.10
ELA5	2018 RS	159581	26.94	7.98	15594	25.57	7.49	-1.37	-0.18
	2018 SS	159581	600.40	20.17	15594	596.97	18.43	-3.42	-0.18
	2017 SS	132281	303.23	38.68	16385	302.36	35.46	-0.87	-0.02
ELA6	2018 RS	151027	28.00	8.75	18988	27.22	8.27	-0.78	-0.09
	2018 SS	151027	600.25	20.19	18988	598.27	18.39	-1.98	-0.10
	2017 SS	125455	300.45	35.89	13900	301.77	33.82	1.32	0.04
ELA7	2018 RS	140012	28.98	9.20	15907	27.66	8.60	-1.32	-0.15
	2018 SS	140012	600.21	20.18	15907	597.27	18.18	-2.94	-0.15
	2017 SS	121037	309.54	32.86	12596	308.06	32.05	-1.47	-0.05
ELA8	2018 RS	136959	30.86	8.77	14563	29.59	8.47	-1.27	-0.15
	2018 SS	136959	600.28	20.17	14563	597.26	18.76	-3.02	-0.16

Table R.3.12. Test-level Performance between Test Modes After Matching – ELA

		PBT			CBT				
Test	**Variable**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**Delta**	**d**
ELA4	2017 SS	11539	310.21	32.36	11539	311.41	30.69	1.20	0.04
	2018 RS	11539	19.39	6.67	11539	19.14	6.32	-0.25	-0.04
	2018 SS	11539	599.63	18.64	11539	598.99	17.34	-0.64	-0.04
ELA5	2017 SS	11944	307.99	32.32	11944	308.69	31.15	0.70	0.02
	2018 RS	11944	27.25	7.43	11944	26.64	7.13	**-0.62**	**-0.08**
	2018 SS	11944	601.09	18.42	11944	599.60	17.32	**-1.49**	**-0.08**
ELA6	2017 SS	14647	305.66	36.41	14647	305.63	34.04	-0.03	0.00
	2018 RS	14647	28.58	8.17	14647	28.27	7.84	-0.30	-0.04
	2018 SS	14647	601.35	18.57	14647	600.57	17.50	-0.78	-0.04
ELA7	2017 SS	11736	304.22	34.24	11736	305.05	32.26	0.83	0.02
	2018 RS	11736	29.30	8.78	11736	28.86	8.18	-0.44	-0.05
	2018 SS	11736	600.77	19.03	11736	599.82	17.19	-0.95	-0.05
ELA8	2017 SS	10875	310.87	32.06	10875	310.90	31.14	0.03	0.00
	2018 RS	10875	30.97	8.58	10875	30.74	8.11	-0.23	-0.03
	2018 SS	10875	600.35	19.49	10875	599.75	18.13	-0.60	-0.03

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

Table R.3.13. Test-level Performance between Test Modes Before Matching – Math

		PBT			CBT				
Test	**Variable**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**Delta**	**d**
Math4	2017 SS	146825	310.43	39.48	10657	308.65	37.03	-1.78	-0.05
	2018 RS	165025	27.86	11.29	11872	26.73	10.71	-1.13	-0.10
	2018 SS	165025	600.16	20.07	11872	597.86	18.27	-2.30	-0.12
Math5	2017 SS	143439	307.08	40.98	9625	307.95	36.49	0.88	0.02
	2018 RS	157831	25.51	11.19	10747	24.83	10.58	-0.68	-0.06
	2018 SS	157831	600.05	20.16	10747	598.79	18.35	-1.26	-0.07
Math6	2017 SS	132143	309.69	38.51	11735	312.37	34.21	2.68	0.07
	2018 RS	150945	23.55	11.60	13484	23.67	10.41	0.12	0.01
	2018 SS	150945	599.92	20.28	13484	600.46	17.27	0.54	0.03
Math7	2017 SS	126604	305.74	42.51	9353	309.29	37.83	3.55	0.09
	2018 RS	141086	27.35	12.99	10663	26.36	11.79	-0.99	-0.08
	2018 SS	141086	600.14	20.17	10663	598.85	17.55	-1.29	-0.07
Math8	2017 SS	87887	300.63	36.84	5992	297.92	31.26	-2.72	-0.08
	2018 RS	101294	24.46	12.14	7116	21.66	10.31	-2.79	-0.25
	2018 SS	101294	600.26	20.10	7116	596.19	17.35	-4.07	-0.22

Table R.3.14. Test-level Performance between Test Modes After Matching – Math

		PBT			CBT				
Test	**Variable**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**Delta**	**d**
Math4	2017 SS	8091	313.60	35.95	8091	312.88	34.32	-0.73	-0.02
	2018 RS	8091	28.92	10.34	8091	28.47	10.00	-0.45	-0.04
	2018 SS	8091	601.56	17.74	8091	600.81	16.60	-0.75	-0.04
Math5	2017 SS	7428	309.91	37.40	7428	311.93	34.37	2.02	0.05
	2018 RS	7428	26.41	10.55	7428	26.48	10.18	0.07	0.01
	2018 SS	7428	601.37	18.27	7428	601.69	17.23	0.32	0.02
Math6	2017 SS	9840	315.41	34.04	9840	314.89	33.00	-0.53	-0.01
	2018 RS	9840	25.97	10.68	9840	24.87	10.29	**-1.10**	**-0.10**
	2018 SS	9840	604.07	17.45	9840	602.47	16.65	**-1.60**	**-0.09**
Math7	2017 SS	7237	311.96	37.74	7237	313.62	35.88	1.66	0.04
	2018 RS	7237	28.72	12.02	7237	28.28	11.50	-0.44	-0.04
	2018 SS	7237	602.14	17.89	7237	601.75	16.43	-0.39	-0.02
Math8	2017 SS	4163	300.74	31.54	4163	302.15	28.79	1.40	0.04
	2018 RS	4163	23.95	10.84	4163	23.14	10.08	**-0.81**	**-0.07**
	2018 SS	4163	599.76	17.73	4163	598.86	16.03	**-0.90**	**-0.05**

Note. The standardized difference (*d*) with an absolute value greater than 0.05 after matching was bolded.

R.3.3. Item-level Mode DIF Analysis

The item-level mode DIF analysis is summarized in Table R.3.15. It presents the numbers of items that were classified as A (negligible), B (moderate), and C (large) level of DIF category in each test. Positive values favor a focal group (the CBT group), and negative values favor a reference group (the PBT group).

As can be seen, very few items were flagged as either B or C-level DIF items across the tests. ELA Grade 5 has most number of items flagged–four flagged as B DIF (1 B+, and 3 B-). Whereas the other tests had from zero to two items flagged.

Table R.3.15. Item-level Mode DIF Analysis Results

Test		A	B+	B-	C+	C-	Total # of Items
ELA	3	23	1	0	0	1	25
	4	24	0	1	0	0	25
	5	31	1	3	0	0	35
	6	33	1	1	0	0	35
	7	34	1	0	0	1	36
	8	36	0	0	0	0	36
Math	3	34	0	0	0	0	34
	4	38	0	0	0	0	38
	5	37	1	0	0	0	38
	6	39	0	0	0	0	39
	7	41	0	0	0	0	41
	8	40	0	1	0	0	41

Section R.4. Discussion and Conclusions

R.4.1. Discussion

Based on the analyses described above, NYSED—in consultation with New York State's Assessment TAC and Questar—decided to apply an additive adjustment to CBT students' scale scores because it best balanced concerns about fairness, interpretability and face validity. NYSED also chose to set a ceiling above which the CBT students' scale scores would not be adjusted–namely the maximum observed scale score available to PBT students. In other words, the highest scale score on CBT was constrained to be equal to the highest scale score for PBT students.

The differences in the 2018 scale score means between the matched samples were computed. The differences were rounded to the nearest whole numbers, which were used as the uniform additive adjustment applied to the CBT students within each test.

For Grade 3 students who do not have prior year scores, there was no propensity score matched samples being generated. Alternatively, the average of mode adjustments of Grades 4 and 5 was computed and used as the adjustment for Grade 3 CBT students. The CBT adjustments in all grade/subject were summarized in Table R.4.1.

Table R.4.1. 2018 CBT Scale Score Adjustments

		2018 After Matching									2017 CBT Adjustment
		PBT			CBT					CBT	
Subject	Grade	*n*	*M*	*SD*	*n*	*M*	*SD*	Δ	*d*	Adjustment	
ELA	3	-	-	-	-	-	-	-	-	+1	+4
	4	11539	599.63	18.64	11539	598.99	17.34	-0.64	-0.04	+1	+5
	5	11944	601.09	18.42	11944	599.60	17.32	-1.49	-0.08	+1	+2
	6	14647	601.35	18.57	14647	600.57	17.50	-0.78	-0.04	+1	+5
	7	11736	600.77	19.03	11736	599.82	17.19	-0.95	-0.05	+1	+2
	8	10875	600.35	19.49	10875	599.75	18.13	-0.60	-0.03	+1	0
Math	3	-	-	-	-	-	-	-	-	+1	+4
	4	8091	601.56	17.74	8091	600.81	16.60	-0.75	-0.04	+1	+5
	5	7428	601.37	18.27	7428	601.69	17.23	0.32	0.02	0	+2
	6	9840	604.07	17.45	9840	602.47	16.65	-1.60	-0.09	+2	+3
	7	7237	602.14	17.89	7237	601.75	16.43	-0.39	-0.02	0	0
	8	4163	599.76	17.73	4163	598.86	16.03	-0.90	-0.05	+1	+8

Note. CBT scale scores were only adjusted up to the maximum observed PBT scale score value.

R.4.2. Conclusions

Following 2017, two administration modes were offered in the Grades 3–8 ELA and Mathematics tests in Spring 2018. The decision to offer PBT vs. CBT was optional. Only a proportion of schools (< 10% of students) chose to administer the tests via CBT. The population of students who tested via CBT were not assumed equivalent to the population of students who tested via PBT. The propensity score matching was conducted select a sample of PBT students that could be comparable to the population of CBT students. The test-level performance was then evaluated between the matched CBT and PBT samples. The results revealed from no to small differences between CBT and PBT group across the tests, with mode effects slightly favoring

PBT groups in some tests. The observed mode difference was smaller across the tests in 2018 than 2017. The observed differences were applied as the adjustments to CBT students to ensure that students received comparable test scores regardless of the test mode.

Appendix S: Memo on Operational Test Mode Comparability

THE STATE EDUCATION DEPARTMENT / THE UNIVERSITY OF THE STATE OF NEW YORK / ALBANY, NY 12234

Assistant Commissioner
Office of State Assessment

September 2018

TO: District Superintendents
Superintendents of Schools
Principals of Public, Religious, and Independent Schools
Charter School Leaders

FROM: Steven E. Katz

SUBJECT: Comparability of Spring 2018 Grades 3-8 English Language Arts and Mathematics Paper-based and Computer-based Tests

The purpose of this memorandum is to provide information about the results of the comparability study that was conducted for the Spring 2018 Grades 3-8 English Language Arts (ELA) and Mathematics paper-based and computer-based tests.

Background

In Spring 2018, the Department offered the Grades 3-8 ELA and Mathematics Tests in two administration modes: paper-based testing (PBT) and computer-based testing (CBT). Administering these tests via CBT was optional for schools and those schools that chose to offer CBT made this decision independently for each subject and grade. The Department provided readiness verification tools to help those schools selecting CBT ensure they were well equipped and prepared to provide a successful CBT experience for their students. Additionally, several CBT practice test sessions were made available to CBT schools to familiarize students and teachers with the CBT delivery system. Each of the CBT practice test sessions featured examples of all of the types of test questions included on the tests. This provided the opportunity for students to practice answering both multiple-choice and constructed-response questions on the computer devices they would be using for the actual test.

To further ensure fairness, the Department's contractor, Questar Assessment Inc., conducted a comparability study to identify whether or not there were any differences in student performance that could be attributed to the mode of test administration (i.e., PBT versus CBT). The comparability study methodology and results are summarized below. The findings of this study were used to ensure that students received a score that was representative of their knowledge and skills, regardless of whether they took the tests on paper or computer.

Comparability Study Methodology

Only some schools chose to administer the tests via CBT (representing approximately eight percent of all ELA test takers and six percent of all Math test takers). Therefore, the population of students who tested via CBT were not assumed equivalent to the population of students who tested via PBT. In order to select a sample of students who tested via PBT that could be compared to those students who tested via CBT, a method called propensity score matching was employed. Propensity score matching allowed for the identification of groups of students who tested via PBT that was similar to the groups of students who tested via CBT on a number of school and student characteristics, including achievement on the prior year's test.

Using these characteristics, Questar selected a group of PBT students that matched the group of CBT students for each grade and subject. This allowed for a direct comparison of student results between the two groups. For comparison, the mean scale scores were calculated for each grade and subject by mode of testing. The results are shown in the section below.

Results of Comparability Study

Table 1 shows the scale score means for the PBT and CBT groups on the 2018 English Language Arts Tests by grade as well as the differences in mean scale scores between the matched groups. Table 2 shows these same data for the 2018 Mathematics Tests.

Table 1. *PBT and CBT Means and Differences for Grades 3-8 ELA*

	PBT Scale Score Mean	**CBT Scale Score Mean**	**Difference** (Rounded to nearest whole number)
Grade 3	See footnote*		n/a
Grade 4	599.6	599.0	+1
Grade 5	601.1	599.6	+1
Grade 6	601.3	600.6	+1
Grade 7	600.8	599.8	+1
Grade 8	600.4	599.8	+1

* Because Grade 3 students have no prior test results on which to match PBT to CBT students, a PBT comparison group was not created and group means were not calculated for this grade level.

Table 2. *PBT and CBT Means and Differences for Grades 3-8 Math*

	PBT Scale Score Mean	CBT Scale Score Mean	Difference (Rounded to nearest whole number)
Grade 3	See footnote*		n/a
Grade 4	601.6	600.8	+1
Grade 5	601.4	601.7	0
Grade 6	604.1	602.5	+2
Grade 7	602.1	601.8	0
Grade 8	599.8	598.9	+1

* Because Grade 3 students have no prior test results on which to match PBT to CBT students, a PBT comparison group was not created and group means were not calculated for this grade level.

Adjustments to Scores

For those tests in which no difference in mean scale scores between the matched PBT and CBT groups was observed, no adjustment was made to any students' scale scores. For those tests in which a difference in mean scale scores between the two comparable groups was observed, the scale scores for all students who took the test in that grade via CBT, (which was the lower scoring mode in all such instances during this administration), were adjusted by adding the number of scale score points shown in the "Difference" columns of Tables 1 and 2 to the CBT students' scale scores, up to the maximum attainable scale score. Thus, the scale score adjustments for students who tested via CBT, shown in Table 3 below, reflect the differences between the PBT and CBT groups found in the comparability study. These slight adjustments ensured that students who demonstrated comparable proficiencies in their knowledge and skills received comparable scores whether they tested on paper or on computer.

Table 3. *Summary of Scale Score Adjustments for CBT*

	ELA Scale Score Adjustment	Math Scale Score Adjustment
Grade 3	+1*	+1*
Grade 4	+1	+1
Grade 5	+1	0
Grade 6	+1	+2
Grade 7	+1	0
Grade 8	+1	+1

* Because Grade 3 students have no prior test results on which to match PBT to CBT students, a PBT comparison group was not created and group means were not calculated for this grade level. Instead, the mean adjustment for the other elementary grades for which a comparison was possible (i.e., Grades 4 & 5) was applied to the scores of Grade 3 students who tested via CBT.

For questions concerning the Grades 3-8 ELA or Mathematics Tests, please email the Office of State Assessment at call 518-474-5902. For questions concerning CBT, please email CBT Support.

Appendix T: Standards Review Report

A report on the standards review of the New York State Education Department Grades 3 to 8 assessments

September 25, 2018

Updated: August 5, 2019

Contents

Introduction and purpose 3
Assessment Design 3
Panelists 4
Process 5
Results 9
Round 2 9
Vertical Articulation 10
Final recommendations 11
Results across rounds 12
Conditional Standard Error of Measurement 13
Comparison to equated cut scores 14
Validity of the standards review/Panelist Surveys 18
Conclusion 19
References 20
Appendix A: Demographic information for standards review panelists 21
Appendix B: Agenda 22
Appendix C: Threshold Performance Level Descriptors for all panels 25
Appendix D: Example of standards review item map 50
Appendix E: Sample of feedback provided after Round 1 and Round 2 52
Appendix F: Example of vertical articulation Impact data 54
Appendix G: Round 2 and Vertical Articulation Cut Point Recommendations 55
Appendix H: Total group recommendations and standard error of the median by round 59
Appendix I: Standard error bands for ratings 61
Appendix J: Survey results for all panels 65

Introduction and purpose

The New York State Department of Education (NYSED) has developed the New York state assessment program, designed to measure the current standing of students on the New York State standards and assess the progression of New York students towards college readiness. In order to complete the reporting of test scores, it is necessary to utilize cut points that will be used to classify student performance on the NYSED assessments into categories. The NYSED contracted with Questar to develop and administer all tests for the 2017-18 academic year; included within that work is the work required to review the existing cut points for the assessments, given a test design change and a reduced test length in 2018 from 2017.

During the week of July 9, 2018, Questar convened panels of New York educators in Albany, NY to review the cut points for the NYSED assessments. The educators were focused on the English Language Arts (ELA) and Mathematics (Math) examinations administered in grades 3 to 8. For each assessment, student performance will be reported using four performance categories which require reviewing three cut points. This report provides an overview of the workshop activities along with the results of the panelists review of the current cut points.

Assessment Design

In 2018, the test length for both ELA and Math were reduced from three sessions down to two sessions. The number of total items and score points was also reduced in 2018. In addition, there was also a change in the calculator policy for Grades 7 and 8, where calculators were permitted throughout the test, versus in the past the calculators were prohibited in Session 1 and permitted in later sessions. A high-level overview of test design comparing 2017 and 2018 is presented below:

Table 1: ELA Test Design Grades 3-8

Grade	Session	2017					2018				
		Strand	MC	CR2	CR4	Total Points	Strand	MC	CR2	CR4	Total Points
3 – 4	1	Reading	18	0	0	47	Reading	18	0	0	34
	2	R & W	7	2	1		Writing	0	6	1	
	3	Writing	0	5	1						
5 – 6	1	Reading	28	0	0	57	Reading	28	0	0	44
	2	R & W	7	2	1		Writing	0	6	1	
	3	Writing	0	5	1						
7 – 8	1	Reading	28	0	0	57	Reading	28	0	0	46
	2	R & W	7	2	1		Writing	0	7	1	
	3	Writing	0	5	1						

Table 2: Math Test Design Grades 3-8

Grade	Session	2017					2018				
		Calculator	MC	CR2	CR3	Total Points	Calculator	MC	CR2	CR3	Total Points
3	1	N	15	0	0	56	N	19	0	0	42
	2	N	22	0	0		N	8	6	1	
	3	N	0	5	3						
4 - 5	1	N	15	0	0	62	N	23	0	0	46
	2	N	23	0	0		N	8	6	1	
	3	N	0	6	4						
6	1	N	19	0	0	68	N	24	0	0	48
	2	Y	25	0	0		Y	7	7	1	
	3	Y	0	6	4						
7 - 8	1	N	19	0	0	68	Y	26	0	0	50
	2	Y	25	0	0		Y	7	7	1	
	3	Y	0	6	4						

Panelists

A total of 56 educators from the state of New York participated as panelists in the workshop. The panelists were recruited for participation starting in the spring of 2018 with the intent to represent the diversity of educators in New York. The panelists were organized into six panels to complete the work (see Table 3).

Table 3: Number of panelists and tests assigned to each panel

Panel	# of panelists	Assessments
1	8	ELA Grades 3-4
2	10	ELA Grades 5-6
3	10	ELA Grades 7-8
4	9	Math Grades 3-4
5	10	Math Grades 5-6
6	9	Math Grades 7-8

During the recruitment process, panelists provided information about themselves as well as the school or organization they were affiliated with. A summary of the key information that was collected can be seen in Appendix A. As can be seen in the tables, the panelists for the standards review panels had a wide range of roles and experience. Across all panelists six panels, between 67% and 100% of the panelists were female, while between 78% and 100% of the panelists in each panel were identified as white. Across all 6 of the panels, approximately 9% of the panelists identified as black/African American and approximately 4% identified as Asian/Asian American.

The panel also represented a wide range of roles within education. Across all panels, approximately 80% of the panelists identified as current teachers in New York. The panelists also came from a wide variety of locations in the state of New York, including New York City,

Central New York, and Western New York. More complete descriptions of the panelists and their role are provided in Appendix A.

Process

In order to complete the review of the existing cut point recommendations, the Bookmark standard setting process was followed (Lewis, Mitzel, Mercado, & Schulz, 2012). The Bookmark process is an item-mapping procedure that is one of the most widely used standard setting methods in statewide assessments. As will be described below, the duration of the workshop was 3 days, including a vertical articulation process that was completed on the last day of the workshop. The agenda for the workshop is provided in Appendix B.

The workshop began on Monday July 9th with a general session for all panelists. This session was conducted by the lead facilitator who provided an overview of the goals and procedures that would be followed throughout the week. During this orientation, Dr. Angélica Infante-Green from the NYSED discussed the current state of the New York assessments, the reporting plan for these assessments, and the key goals and milestones for the NYSED assessments over the next year. The lead facilitator then provided an overview of the Bookmark procedure and how panelists would complete their work.

During this orientation, the lead facilitator described the goal of the meeting that was to review the existing standards set for the exam. The lead facilitator reviewed how the current content blueprints for the examinations had been slightly modified from the 2017 exams to the 2018 exams. The facilitator further explained that while they had not yet observed any evidence that the existing cut points were not appropriate, in the interest of due diligence, the state had determined that convening a panel of New York educators was appropriate. The meeting was designed to evaluate whether the current cut points were still considered to be appropriate from a content perspective.

After this orientation, panelists split into their respective panels to begin working on their subject/grade level assessments. The facilitator assigned to each of the panels led their panel through the remainder of the workshop. Each panel was organized into table groups of 4-5 panelists. The members of the table groups remained as small working teams for the remainder of the study.

The first phase of the work allowed the panelists the opportunity to review a test form to become familiar with the knowledge and skills measured. For all panels, this process began with the lower grade level: Grade 3, Grade 5, and Grade 7. Panelists were not required to answer the items but were encouraged to review and consider the expectations for items and how well prepared their students were for the assessment.

After reviewing the assessments, panelists were provided copies of the current Performance Level Descriptors (PLDs) that described the knowledge, skills, and abilities expected from students within each of the four performance categories. Panelists reviewed and discussed the PLDs with their table groups. The panelists were then charged with developing threshold PLDs that would describe the knowledge, skills, and abilities expected from students who were just barely within each performance category. A copy of all threshold PLDs developed by the panelists are provided in Appendix C.

After developing the PLDs, the facilitators discussed the role of a standards review panel and how it would work. Prior to the standards review workshop, the three-parameter logistic model (3PL; Lord and Novick, 1968; Lord, 1980) was used to estimate the parameters for all MC items. The partial credit model (2PPC; Muraki, 1992; Yen, 1993) was used to calibrate all CR items.

Once the item parameters were identified, the response probability for each item was calculated using a probability of 0.67. Using these calculated values, the OIBs were created, with each item appearing on one page in the OIB developed by Questar. For the CR items, each item appeared multiple times, once for each score on each trait. As with the item calibrations, the values for the response probabilities were calculated by two independent parties, and the OIBs were only calculated once both parties had verified the match on values and the correct order for the items in the OIB.

After the calibration of items was completed, the cut points were equated by identifying the theta value on the 2018 test that was comparable to each of the three cut points from the 2017 test. Once these equated cut point values were identified, the page with the closest associated theta for each of the three cut points were identified. By doing so, the page that was equivalent to the current cut point recommendation could be determined, for all three cut points, for all test forms. This information was provided to all panelists after the training on the Bookmark method. Panelists were told that in addition to identifying the specific page that represented the equated cut point, there was also an expected range of variability (plus or minus two pages around each cut point). Panelists were provided a handout for each exam that listed all items on the exam with the one page identified as comparable to the cut point (dark green) and a range of two pages below/above identified to represent the expected variability (light green). Additional field-test items were added to the OIB list if a gap was observed around the equated cut points. Panelists were told that if they wished to provide a recommendation outside of the green zone, they would also be asked to provide a content-based rationale for their rating. An example of the item map handout that was provided is included in Appendix D of this report.

After the review of the PLDs and the item maps, panelists were provided further training on the Bookmark procedure and how they would review the test and complete their ratings. After the training on the process was completed, panelists were then provided an opportunity for a practice round with the Bookmark method with a set of six items from the spring test forms. They were asked to determine a Bookmark rating for the first threshold level - Level 2. As panelists completed their practice ratings, the facilitator ensured that panelists were comfortable with the process and understood how to record their ratings. After the practice ratings were completed, the panelists discussed how they made their judgments and reviewed any questions or concerns they had.

After completing the practice round, panelists were provided the OIB for the complete test to be used and began their Round 1 rating. Panelists completed their ratings on a hard copy rating form, and once they had completed all three ratings for Round 1, provided the rating form to the facilitator. The facilitator entered all ratings into an EXCEL worksheet to create the feedback for each round. Once all panelists had completed their ratings and the facilitator had compiled all ratings, panelists were provided multiple pieces of feedback. Examples of all feedback provided are included in Appendix E of this report. The feedback included:

- The minimum, maximum and median recommendation received from panelists
- The distribution of recommendations received across all panelists
- The difficulty level of items

Figure 1: Example of figure used to demonstrate variability of cut point recommendations

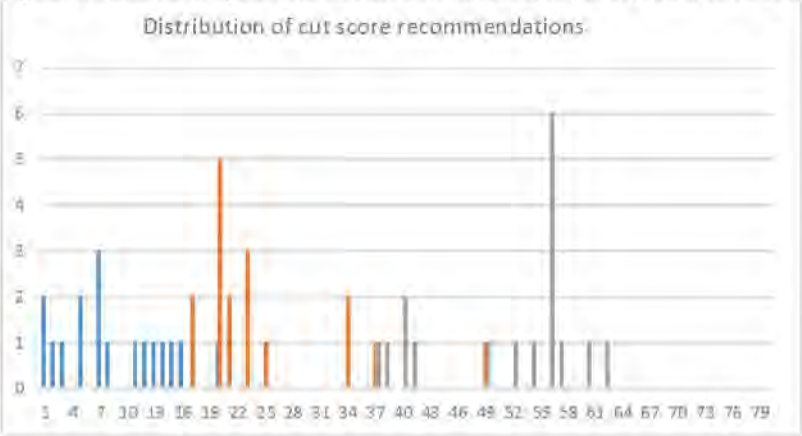


Figure 1 shows an example of how the distribution of cut point recommendations were illustrated for the panelists. After presenting the feedback associated with the recommendations received, the facilitator instructed the panelists to discuss the results and their ratings at the table level. Panelists were encouraged to share their rationale for selecting the page they had selected, and to discuss how they felt the items could be linked to the threshold PLDs. Panelists discussed Figure 1 and individual panelists who had provided some of the highest and lowest recommendations at each level were asked to provide their rationale for their rating. Panelists also discussed why they may have observed significant variability for some levels, while other levels showed little to no variability.

After these discussions were completed, estimated impact data was also shared with the panelists. The impact data was based upon the recommendation of the entire group and included the estimated percentage of students classified into each of the four performance levels. Table 4 below provides an example of the feedback provided to panelists. This table includes not only the cut point recommendation, but also the minimum and maximum recommendation received. Lastly, it also includes the estimated percentage of students classified into each of the four performance categories based upon the current cut point recommendation.

Table 4: Example of cut point feedback, including impact data

	Minimum	Maximum	Median	% students
Level 1				63%
Level 2	1	20	7	24%
Level 3	17	49	21	12%
Level 4	37	62	56	1%

After providing the impact data, the moderator facilitated a discussion of the impact data with all panelists. The moderators asked panelists to discuss whether they felt that the impact data was consistent with their expectations or if any particular components of the impact data was surprising to them. After the discussion of the impact data was completed, the panelists were instructed to complete their Round 2 rating.

After completing this discussion, panelists were asked to make their round 2 (final) ratings. After completing these ratings, the results were compiled, and panelists were provided a brief summary of their final recommendations and the resulting impact. After this review, panelists completed a survey indicating their comfort level with the overall process and the all materials for the first test were collected by the panel facilitators.

Each panel proceeded forward to complete the standards review process for the remaining tests. Panelists completed their work on their second assigned assessment closely following the process described above. The one difference was that panelists did not complete a practice round for the second assessment. Panelists still completed two rounds and were provided feedback in the same manner as was done with the first test.

In addition to determining the recommended cut points, a vertical articulation process was also included in the workshop. For both English and math, the vertical articulation process was completed on Wednesday afternoon and included representatives from each respective panel. During the vertical articulation, the panelists first discussed the knowledge and skills necessary to complete each of the assessments. They also discussed the expected challenges and new materials that were introduced each year and whether they would expect to observe consistent performance across years and content areas. After this discussion, each table reviewed the recommended cut points and impact data across all grade levels and content areas.

Panelists were shown a figure that indicated the estimated percentage of students in each of the four performance categories for all tests within the content area. Panelists could also compare these percentages using a different cut point recommendation. An example of the estimates that were originally reviewed by both vertical articulation panels are included in Appendix F. Vertical articulation panelists recommended changes to the cut points based upon this review and reached a consensus on the most appropriate cut points for each test.

Results

Round 2

The recommended cut points recorded after Round 2 are provided in Table 5 and 6 below. Tables 5 and 6 include, for each level:

- The recommended cut point, or page number that the committee recommended
- The corresponding theta value associated with the cut point recommendation
- The corresponding raw score cut for each level
- The estimated percentage of students who would be classified into each of the levels.

Table 5: Recommended cut points after Round 2 for the NYSED ELA assessments

Round 2 recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	17.4%	19.1%	32.8%	27.1%	28.3%	18.2%
Level 2	Page #	5	4	15	11	18	14
	Theta	-0.9061	-0.9521	-0.3710	-0.5556	-0.5325	-0.8798
	Raw score cut	12	13	24	23	24	23
	% students	31.7%	33.1%	30.5%	23.7%	31.5%	33.5%
Level 3	Page #	13	19	30	28	36	33
	Theta cut point	0.1046	0.2332	0.4600	0.1718	0.3917	0.1646
	Raw score cut	19	21	31	30	33	33
	% students	43.7%	29.8%	22.5%	22.0%	28.1%	27.5%
Level 4	Page #	36	30	42	40	48	45
	Theta cut point	1.5069	1.0535	1.2514	0.7967	1.2212	0.9229
	Raw score cut	29	27	36	35	40	39
	% students	7.3%	18.1%	14.3%	27.2%	12.1%	20.9%

Table 6: Recommended cut points after Round 2 for the NYSED Math assessments

Round 2 recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	23.7%	25.0%	31.5%	30.1%	31.7%	36.1%
Level 2	Page #	9	6	9	7	6	4
	Theta	-0.6954	-0.6109	-0.4231	-0.4525	-0.4241	-0.2439
	Raw score cut	18	19	19	16	19	18
	% students	21.9%	26.3%	24.1%	27.7%	25.9%	32.4%
Level 3	Page #	24	24	21	22	21	23
	Theta cut point	-0.0020	0.0959	0.2137	0.0062	0.2938	0.5379
	Raw score cut	26	30	28	26	31	31
	% students	30.9%	23.2%	26.1%	20.8%	23.6%	18.7%
Level 4	Page #	38	42	41	37	48	45
	Theta cut point	0.7955	0.7825	1.0200	0.8366	0.9632	1.1324
	Raw score cut	35	38	38	35	42	41
	% students	23.5%	25.4%	18.4%	21.4%	18.8%	12.8%

Vertical Articulation

During the vertical articulation, the participants first discussed the knowledge and skills expected at each performance level for each assessment. They also discussed the expected challenges and new materials that were introduced each year and whether they would expect to observe consistent performance across years. After this discussion, the panel reviewed the recommended cut points and impact data across all grades. The panelists recommended slight changes to the cut points based upon this review and reached a consensus on the most appropriate cut point recommendations for each test. The results from each panel are recorded in Table 7 below. As can be seen in Table 7, the ELA panel recommended no changes to the cut point recommendations, while the Math panel recommended modifications to cut point recommendations for 2 of the 6 grade levels.

Table 7: Changes recommended during the Vertical Articulation process

OIB Page # recommendation			
	Level 2	**Level 3**	**Level 4**
English Language Arts			
Grade 3	No change		
Grade 4	No change		
Grade5	No change		
Grade 6	No change		
Grade 7	No change		
Grade 8	No change		

OIB Page # recommendation			
	Level 2	**Level 3**	**Level 4**
Mathematics			
Grade 3	No change		
Grade 4	No change		
Grade5	No change	No change	From 41 to 39
Grade 6	No change	From 22 to 21	No change
Grade 7	No change		
Grade 8	No change		

In all instances, panels had reviewed the OIB and determined that the recommended changes were more consistent with the threshold PLDs and the expected performance of students at that level.

Final recommendations

As the workshop concluded, all ratings received by panelists were compiled and the final recommended cut points were determined along with the resulting impact. Tables 8 and 9 include, for each level:

- The recommended cut point, or page number that the committee recommended
- The theta corresponding to the page number recommendation from the committee
- The corresponding raw score cut for each level
- The estimated percentage of students who would be classified into each of the levels.

In Appendix G, both sets of results (post-Round 2, post-vertical articulation) are presented to document any changes that were introduced during the vertical articulation phase.

Table 8: Recommended cut points for the NYSED English Language Arts assessments

		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	17.4%	19.1%	32.8%	27.1%	28.3%	18.2%
Level 2	Page #	5	4	15	11	18	14
	Theta	-0.9061	-0.9521	-0.3710	-0.5556	-0.5325	-0.8798
	Raw score cut	12	13	24	23	24	23
	% students	31.7%	33.1%	30.5%	23.7%	31.5%	33.5%
Level 3	Page #	13	19	30	28	36	33
	Theta cut point	0.1046	0.2332	0.4600	0.1718	0.3917	0.1646
	Raw score cut	19	21	31	30	33	33
	% students	43.7%	29.8%	22.5%	22.0%	28.1%	27.5%
Level 4	Page #	36	30	42	40	48	45
	Theta cut point	1.5069	1.0535	1.2514	0.7967	1.2212	0.9229
	Raw score cut	29	27	36	35	40	39
	% students	7.3%	18.1%	14.3%	27.2%	12.1%	20.9%

Table 9: Recommended cut points for the NYSED Mathematics assessments

		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	23.7%	25.0%	31.5%	30.1%	31.7%	36.1%
Level 2	Page #	9	6	9	7	6	4
	Theta	-0.6954	-0.6109	-0.4231	-0.4525	-0.4241	-0.2439
	Raw score cut	18	19	19	16	19	18
	% students	21.9%	26.3%	24.1%	25.1%	25.9%	32.4%
Level 3	Page #	24	24	21	21	21	23
	Theta cut point	-0.0020	0.0959	0.2137	0.2201	0.2938	0.5379
	Raw score cut	26	30	28	25	31	31
	% students	30.9%	23.2%	23.4%	23.4%	23.6%	18.7%
Level 4	Page #	38	42	39	37	48	45
	Theta cut point	0.7955	0.7825	0.8416	0.8366	0.9632	1.1324
	Raw score cut	35	38	37	35	42	41
	% students	23.5%	25.4%	21.0%	21.4%	18.8%	12.8%

Results across rounds

As was described earlier, all panelists completed two sets of ratings for each test during the standards review workshop. For each round, the recommended cut points for each panel were determined, with the results and resulting impact data shared with the panelists. In addition to the recommended cut point, the standard error of the median was also determined. The standard error was determined to evaluate the consistency of the panelists ratings and to help evaluate how consistent the panelist recommendations were. The standard error of the mean was calculated using the panelists' individual ratings. In addition, the standard error of the mean was then multiplied by 1.2538 to estimate the standard error of the median as per McCann and Stanley (2004).

The recommended cut points during each round, along with the standard error of the median for each cut point, are provided in Appendix H. As can be observed in the tables, the recommendations for each cut point generally did not shift dramatically from round 1 to round 2. However, the standard error of the cut points did generally decrease from Round 1 to Round 2.

For each recommended cut point, the standard error of the median was used to estimate error bands for the cut point recommendation. Once these error estimates were calculated and associated with a page number recommendation, these page numbers were converted to the theta value for each page, and then to the raw score for each theta value. The raw scores were used to estimate the impact at that score to be determined. In Appendix I, error bands are provided for each assessment, and for each cut point recommendation.

Conditional Standard Error of Measurement

In order to evaluate how measurement error could impact the interpretation of the cut scores, the conditional standard error of measurement (CSEM) of the scaled score was also estimated. This estimate is done independent of the standards review activities and are based on the test administration data from the spring of 2018. The CSEM reflects the reliability of the test scores as well as the variability of scores throughout the scale. The CSEM provides further information regarding the expected variability of student performance around each of the recommended cut score. In Table 10 below, for each assessment, the recommended scale score cut points are provided along with the CSEM for that scale score.

Table 10: Scale score cut scores and associated conditional standard error of measurement (CSEM)

	Level 2		Level 3		Level 4	
	Scale score cut	CSEM	Scale score Cut	CSEM	Scale score cut	CSEM
English Grade 3	583	7	602	6	629	7
English Grade 4	584	6	603	6	619	7
English Grade 5	594	6	609	6	622	7
English Grade 6	590	5	602	5	614	6
English Grade 7	591	5	607	5	623	6
English Grade 8	584	5	603	5	617	6
Math Grade 3	587	5	600	4	615	5
Math Grade 4	588	4	602	4	614	5
Math Grade 5	592	4	604	4	616	4
Math Grade 6	592	5	604	4	616	4
Math Grade 7	593	4	606	3	618	4
Math Grade 8	596	5	610	3	622	4

Comparison to equated cut scores

Once the final cut point recommendations were identified, the recommendations were compared to the equated recommendations. The first review was completed by comparing the equated page number recommendation with the final page number recommendation received by each panel. Table 11 provides the equated page number and the expected range of variability provided to panelists along with the final recommendation that panelists provided. As can be seen in the table, 22% of the recommendations in English Language Arts matched the equated page number, while an additional 61% of recommendations received were within the expected range of variability. Approximately 17% of the English language Arts page number recommendations were outside of the expected range. The recommendations in Math followed a similar pattern, with 28% of recommendations matching the equated page number, and an additional 50% of recommendations falling within the expected range of variability.

Table 11: Equated page number, expected range of variability and final cut page number recommendations

	Level 2			Level 3			Level 4		
	Equated Page #	Range Page #s	Final Page #	Equated Page #	Range Page #s	Final Page #	Equated Page #	Range Page #s	Final Page #
ELA									
Grade 3	6	4 - 8	5	17	15 - 19	13	37	35 - 39	36
Grade 4	5	3 - 7	4	21	19 - 23	19	30	28 - 32	30
Grade 5	16	14 - 18	15	32	30 - 34	30	41	39 - 43	42
Grade 6	11	9 - 13	11	34	32 - 36	28	43	41 - 45	40
Grade 7	19	17 - 21	18	37	35 - 39	36	49	47 - 51	48
Grade 8	14	12 - 16	14	33	31 - 35	33	46	44 - 48	45
Math									
Grade 3	10	8 - 12	9	27	25 - 29	24	40	38 - 42	38
Grade 4	7	5 - 9	6	26	24 - 28	24	44	42 - 46	42
Grade 5	9	7 - 11	9	22	20 - 24	21	42	40 - 44	37
Grade 6	6	4 - 8	7	23	21 - 25	21	37	35 - 39	37
Grade 7	6	4 - 8	6	21	19 - 23	21	47	45 - 49	48
Grade 8	4	2 - 6	4	30	28 - 32	23	51	49 - 53	45

In addition to comparing the page recommended, the cut point recommendations were also compared using the theta values associated with each recommendation. Each recommended cut point was translated to the theta value associated with it, and the equated theta cut point was then subtracted from it. In other words, in the event that the recommended cut point lowered or decreased the expected cut point, we would expect to see a negative value. Figure 2 below shows the difference values for all theta cut points recommendations received in English Language Arts while Figure 3 shows the same information for the Mathematics assessments. As can be seen in Figure 2, the theta cut point recommendations generally remained fairly close to the equated theta. It is also interesting to note that when differences were observed, the recommendations were consistently slightly below the equated thetas. Lastly, in English, it does appear that the Level 3 cut point is the level that had the most disagreement between the equated theta and the recommended theta values. In Mathematics, it appears that the Level 4 cut point is the recommendations where the largest discrepancy exists.

Figure 2: Difference in theta values for English Language Arts cut point recommendations

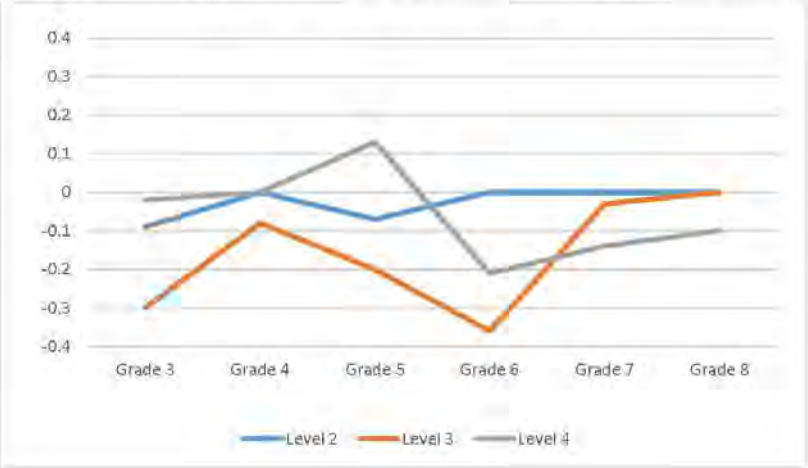
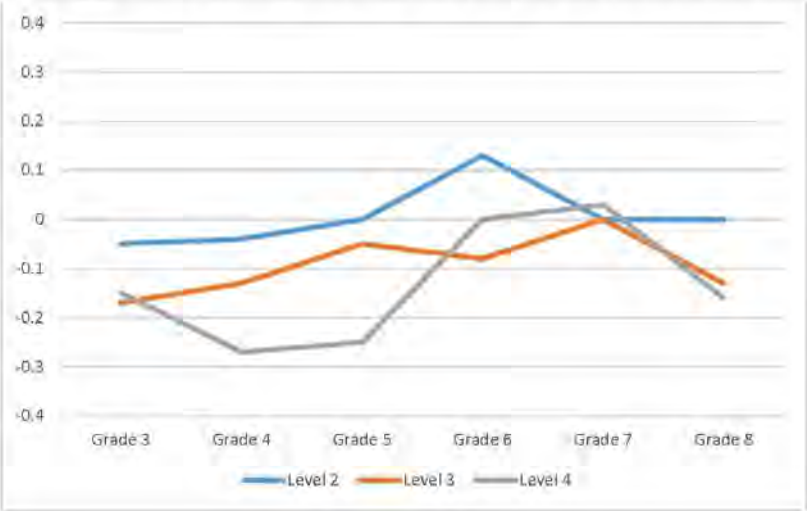


Figure 3: Difference in theta values for Mathematics cut point recommendations



The data presented in Figures 2 and 3 are focused on comparing the difference in thetas values across the different levels of cut points. Reviewing these figures, we can see that the cut point recommendations received from panelists for the Level 2 cut points were, in general, more consistent with the equated cut points. We can also see that the cut points recommendation in Math were generally more consistent with the equated cut points.

Another way to review these results is to focus on specific assessments. Figures 4 and 5 below provides this information for each of the English Language Arts and Mathematics assessments respectively. In English Language Arts, the recommendations received for grade 4 and grade 8 appear to closely approximate the equated values. On the other hand, the grade 6 recommendations diverge a bit more than the other grade levels. In Mathematics, the grades 3 and 4 appear to have more significant differences between the recommended and equated theta values, whereas the grade 6 and grade 7 recommendations appear to closely align with the equated theta values.

Figure 4: Difference in theta cut values for English Language Arts assessments

Figure 5: Difference in theta cut values for Mathematics assessments

Validity of the standards review/Panelist Surveys

Throughout the entire duration of the project, and as the results were finalized, Kane's (2001) framework for validating standards review activities was applied. Kane (1994) suggested three sources of evidence should be considered in the validation process: procedural, internal, and external. Evidence within each of these areas that was observed in this study is discussed here.

Procedural evidence can be viewed through the surveys and feedback provided throughout the entire standards review workshop. In addition to providing the recommended cut scores, panelists also completed surveys designed to evaluate how well prepared they felt to provide their ratings and whether they felt the ratings were appropriate. In the first survey completed, all panelists were asked about the effectiveness of the general session by indicating their agreement with the statement "The general session clarified what procedures were going to be followed and how the work would progress." They provided ratings ranging from Strong Disagree (value = 1) to Strongly Agree (value = 4). Across all panelists, the average rating received was 3.40, indicating that the general session provided a good overview of the overall process.

In the second and third evaluation surveys, panelists were asked how well comfortable they were with the final cut score recommendation. This survey question was asked after they completed their Round 2 recommendations, but *prior to* the vertical articulation panels. The ratings for two of the survey items are summarized in Table 12 below. Appendix J contains the survey questions for each survey along with the average response to each question on all surveys completed.

Table 12: Mean response for the questions below (1 = strongly disagree, 4 = strongly agree).

	I am confident my recorded ratings reflect my best judgment on where to place the Bookmark.	Based upon the feedback, I understood my recommended cut point and how I compared to other panelists.
Panel 1 (ELA grade 3/4)	3.9	3.5
Panel 2 (ELA grade 5/6)	3.9	4.0
Panel 3 (ELA grade 7/8)	3.6	3.9
Panel 4 (Math grade 3/4)	3.8	4.0
Panel 5 (Math grade 5/6)	3.9	4.0
Panel 6 (Math grade 7/8)	3.9	3.2

The primary source of internal validity evidence can be observed when looking at the variability of the cut point recommendations. The standard error of the median was calculated for every round of ratings, for each of the cut point recommendations. Table 13 below provides the mean standard error value across all assessments, for each cut point recommendation, and for each round. For both ELA and Math, the variability of the standard error did decline as panelists moved from the first to the second round, which is indicative of an increased degree of agreement across panelists. The standard error for each recommended cut point, within each grade, content area, and round, is provided in Appendix H.

Table 13: Median standard error of the ratings by round

		Level 2	Level 3	Level 4
ELA	Round 1	0.67	0.80	0.65
	Round 2	0.36	0.58	0.56
Math	Round 1	0.52	1.45	0.64
	Round 2	0.27	0.40	0.50
All	Round 1	0.59	1.13	0.65
	Round 2	0.32	0.49	0.53

Conclusion

At the conclusion of all workshop activities, the cut score recommendations were provided to the New York State Department of Education for their review. Based upon the evidence collected and the review of the performance of panelists, it does appear that the cut point review provided support for the current cut score recommendations. The results of this study were used by NYSEDNYSED to make final decisions on the appropriate cut scores. Moving forward, it would also be appropriate to closely monitor the item and test performance across all students as the New York State assessment program continues to be administered across New York.

References

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

Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. Cizek (ed.), Setting performance standards: Concepts, methods, and perspectives (pp. 53-88). Mahwah, NJ: Lawrence Erlbaum Associates.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The Bookmark standard setting procedure. In G. J. Cizek (Ed.) *Setting Performance Standards: Foundations, Methods, and Innovations* (2nd Ed.). New York, NY: Routledge.

Lord, F.M. (1980). *Applications of item response theory to practical problems*. Hillsdale, NJ: Erlbaum.

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Appendix A: Demographic information for standards review panelists

Table A.1 Demographic characteristics of standards review panelists

Standards review panelists	Sex		Race			
	Female	Male	Black	White	Asian	Other
ELA Grades 3 and 4	100%	--		100%		
ELA Grades 5 and 6	90%	10%	10%	90%		
ELA Grades 7 and 8	80%	20%	10%	80%		10%
Math Grades 3 and 4	89%	11%	22%	78%		
Math Grades 5 and 6	90%	10%	10%	80%	10%	
Math Grades 7 and 8	67%	33%		89%	11%	

Table A.2: Current professional roles of standards review panelists

Standards review panelists	Teacher	Curriculum	Administrator	Other
ELA Grades 3 and 4	75%			25%
ELA Grades 5 and 6	90%			10%
ELA Grades 7 and 8	60%			40%
Math Grades 3 and 4	67%		11%	22%
Math Grades 5 and 6	100%			
Math Grades 7 and 8	89%		11%	

Table A.3: Regional area of standards review panels

	English language Arts			Mathematics		
Standards review panelists	Grade 3	Grade 5	Grade 7	Grade 3	Grade 5	Grade 7
North County	38%	10%	10%		10%	11%
Long Island	25%		20%	11%		
New York City		10%	10%	33%	20%	11%
Big 4 City		20%	10%	11%	10%	22%
Lower to mid-Hudson						33%
Capital Region	13%	20%	20%	11%	10%	
Central NY	13%	30%	20%	22%	30%	11%
Western NY	13%	10%	10%	11%	20%	11%

Appendix B: - Agenda

Monday, July 9

Time		Activity	Room
8:00-8:30	AM	Registration and Breakfast	Governor A
8:30-10:00	AM	Welcome and Introductions General Session	Governor A
10:00 -10:15	AM	Break – Coffee Service *10:15 – 10:30* *10:30 – 11:00* *11:00 – 12:00*	Governor A *Get settled, signatures, introductions* *Review assessments* *Table level drafts of PLDs*
10:15 - 12:00	PM	Move to Breakout Rooms	Breakout Rooms
12:00-1:00	PM	Lunch	Governor A
1:00-2:30	PM	Work in Breakout Rooms *Finish PLDs by this time*	Breakout Rooms
2:30-2:45	PM	Break – Snack *PLDs will be printed*	Governor A
2:45-5:00	PM	Work in Breakout Rooms *Training and practice by 3:45* *Ratins begin by 3:45*	Breakout Rooms

Tuesday, July 10

Time		Activity	Room
8:00-8:30	AM	Registration and Breakfast	Governor A
8:30-10:00	AM	Work in Breakout Rooms *Round 1 discussion*	Breakout Rooms
10:00-10:15	AM	Break – Coffee Service	Governor A
10:15-12:00	PM	Work in Breakout Rooms *10:15-11:15* *11:30-12:00*	Breakout Rooms *Round 2 rating* *Round 2 feedback and wrap up*
12:00-1:00	PM	Lunch	Governor A
1:00-2:30	PM	Work in Breakout Rooms *1:00 – 1:30* *1:30 – 2:30*	Breakout Rooms *Review assessment* *Table level PLDs*
2:30-2:45	PM	Break – Snack	Governor A
2:45-5:00	PM	Work in Breakout Rooms *Finish PLDs*	Breakout Rooms

Wednesday, July 11

Time		Activity	Room
8:00-8:30	AM	Registration and Breakfast	Governor A
8:30-10:00	AM	Round 1 ratings	Breakout Rooms
10:00-10:15	AM	Break – Coffee Service	Governor A
10:15-12:00	PM	Work in Breakout Rooms	Breakout Rooms
		10:15-11:15	*Round 1 feedback & discussion*
		11:15-12:00	*Round 2 ratings*
		12:15-12:30	*Round 2 wrap up*
12:30-1:30	PM	Lunch	Governor A
1:30-4:30	PM	Vertical Articulation	

Appendix C – Threshold Performance Level Descriptors for all panels

English Language Arts **Grade 3**

Threshold Level 2

basic/partial/limited

- literal test questions/right-there questions (e.g., events, characters, etc.)
- Do not need to understand the entire text
- Grab a detail, but not use it the right way or thoroughly
- what happens first/last

-imprecise

- who/what/where/why – but not able to explain and with some errors
- able to identify literal meaning, but not identify correctly the non-literal

Students show partial, basic comprehension of the text (some inaccuracies may present) and use limited text evidence to support their thinking.

- emerging use of commands of standard English grammar
    - writing makes partial sense/understandable even if it has errors
    - may have several errors that hinders comprehension
    - proper tense
    - complete simple sentences
- inconsistent use of punctuation (attempts to use punctuation, some long-run on sentences), capitalization, spelling (most words are understandable)
- determine meaning of unknown words at times, but inconsistently and perhaps not generalizing it as it is used in the text
    - can identify meaning of words in isolation, but unable to pick up on nuance of the word meaning in text
    - can identify meaning of words in simple text
- can understand how words connect to the meaning, but not in-depth
- using words that do not connect to the purpose/task; not relevant to the task
- have limited understanding of grade 3 vocabulary – may understand some domain specific vocabulary
    - can answer more easily if distractors are farther apart
- addresses a topic with insufficient clarity in writing – limited in how they convey their ideas
    - limited details/incorrect details (what do you mean by that?)

Threshold Level 3

- complete and accurate understanding

  some parts may be vague

-referring to the text to answer accurately

-full recount of the text (summarize, relevant supporting details, main idea, etc.)

-competent use of information (heading, graphic details, text features)

-make distinctions between point of view, who is the narrator

Students show complete and accurate comprehension of the text and use text evidence to support their thinking.

- using conventions accurately with occasional errors
  - writing makes sense/understandable with errors that do not affect comprehension
  - proper tense,
  - complete complex sentences
  - may begin sentences in a similar way
- can determine the meaning of words using a range of strategies
- use words that connect to the purpose/task
- accurate understanding of grade 3 vocabulary
- addresses a topic clearly in writing

Threshold Level 4

- in-depth understanding (more abstract, multi-step, theme, connect paragraphs, supporting details)
- explicitly referring to the text – (e.g., what showed why Johnny was kind to Sally; character's actions, what does it reveal about the character)
- asking insightful questions
- detailed, nuanced and accurate explanations, abstract connections to the text
- understanding subtlety and complexity
- clear and precise distinctions
- in-depth textual analysis aptly and insightfully
- confident command of conventions with few minor errors
- skillful and precise use of language (e.g., multiple meaning, transition words, punctuation)
- can determine the meaning of words precisely using a range of strategies
- precisely understand word relationships and how words and phrases connect
- advanced understanding of grade 3 vocabulary
- addresses a topic clearly and precisely in writing
- engages their reader in their writing (voice)

English Language Arts **Grade 4**

Threshold Level 2

-basic or partial understanding by inconsistently or incorrectly referring to details when drawing inferences

-basic or inaccurate ability to make connections across a text (details, characters, events, paragraphs)

-partial understanding of main idea (appropriate detail vs the main idea; may give main idea of only part of asked text)

-basic or limited summary (e.g., omit important details, include irrelevant details, limited sequence)

-partial understanding of theme without correct supporting details

-partial understanding of words and phrases within a literary text

-basic/partial ability to explain major differences across genres

-basic/partial to describe the overall structure

-basic/partial ability to make observations and/or comparisons about point of view (identify but not compare)

-basic/partial ability to description of differences between accounts

-limited connections between texts and interpretations of information presented visually (identify but not make the connections; limited interpretation)

- partially supports an argument with the claim

-basic ability to compare and contrast

- emerging use of commands
    - writing makes partial sense/understandable with some errors
    - may have several errors that hinders comprehension
    - proper tense
    - complete simple sentences
- inconsistent use of punctuation, capitalization, spelling, grammar and usage
- can hinder comprehension
- determine meaning of unknown words inconsistently and perhaps not generalizing it as it is used in the text including figurative language
- can understand how words connect but not in-depth
- using words that do not connect to the purpose/task
- have limited understanding of grade 4 vocabulary
- addresses a topic with insufficient clarity in writing
- inconsistently use domain-specific words that are essential to a particular topic (ex. wildlife, conservation when discussing animal preservation)
- inconsistently draw evidence from texts to support analysis

Threshold Level 3

-Consistently and correctly referring to details when drawing inferences

-Thorough ability to make connections across a text (details, characters, events, paragraphs)

-use relevant and appropriate details to support

-complete and accurate understanding of main idea

-full and accurate ability to summarize

-thorough describe of characters, setting or events

-draws on specific details

-determine the meaning of words and phrases within a literary text

-determine academic and domain-specific words or phrases

- thorough understanding of structures of a text and explain the differences throughout

-correctly identify theme and supporting details

-thorough analysis of point of view across different texts/genres

-complete and thorough understanding of oral and visual presentation of a text

-thorough explanation of how an author uses reasons and evidence to support their claim

-thorough textual analysis by comparing and contrasting the treatment of similar themes and topics

- using conventions accurately with occasional errors
- can determine the meaning of words using a range of strategies including figurative language
- use words that connect to the purpose/task
- accurate understanding of grade 4 vocabulary
- addresses a topic clearly in writing
- carefully and accurately use domain-specific words that are essential to a particular topic (ex. wildlife, conservation when discussing animal preservation)

carefully draw evidence from texts to support analysis

Threshold Level 4

-in depth and illuminating details when drawing inferences

-in depth understanding of a text noting subtle connections and providing a detailed nuanced summary of the text

-in depth understanding of a text formulating a sophisticated statement of main idea

-in depth ability to make connections across a text (details, characters, events, paragraphs)

-detailed and nuanced explanation of events

-thorough describe of characters, setting or events-draws on specific details

-determine with precision and detail the meaning of words and phrases within a literary text

-determine with precision and detail academic and domain-specific words or phrases

- thorough understanding of structures of a text and explain the differences throughout

-insightful analysis of comparing and contrasting

-in-depth, understanding of theme and nuanced supporting details

-in-depth analysis of point of view across different texts

- confident command of conventions with few errors
- skillful and precise use of language
- can determine the meaning of words precisely using a range of strategies including figurative language
- precisely understand word relationships and how words and phrases connect
- advanced understanding of grade 4 vocabulary
- addresses a topic clearly and precisely in writing
- engages their reader in their writing
- precise and nuanced use domain-specific words that are essential to a particular topic (ex. wildlife, conservation when discussing animal preservation)
- skillfully and purposefully draw evidence from texts to support analysis
-

English Language Arts **Grade 5** **Reading**

Threshold Level 2

- partially use of explicit and implicit information to support emerging inferences
- partially analyze relationships among literary elements in texts of varying complexity and genre
- Summarize central ideas and **events** using some insufficient or irrelevant details
- Student may just retell all aspects of the text presented
- partially determine meanings of academic and domain specific words/phrases and words with multiple meanings based on context-word relationships and differentiating vocabulary meanings
- partially demonstrates a general understanding of author's purpose and point-of-view

Threshold Level 3

- adequate use explicit and implicit information to justify inferences
- adequately analyze of relationships among literary elements in texts of varying complexity and genre
- Summarize central ideas and key events using **sufficient** details
- adequately determine meanings of academic and domain specific words/phrases and words with multiple meanings based on context-word relationships and differentiating vocabulary meanings
- adequately demonstrate a general understanding of author's purpose and point-of-view

Threshold Level 4

- Captures insightful aspects of reading passages that go beyond literal interpretations of text
- consistently use of explicit and implicit information to justify inferences
- consistently analyze of relationships among literary elements in texts of varying complexity and genre
- consistently summarize central ideas and key events using insightful details
- consistently determine meanings of academic and domain specific words/phrases and words with multiple meanings based on context-word relationships and differentiating vocabulary meanings
- consistently demonstrates a insightful understanding of author's purpose and point-of-view

Grade 5 Writing and Language

- At level 1 it's basically unreadable with many errors. Key terms limited and inaccuracy throughout making it unreadable.

Threshold Level 2

- Inconsistent basic structure, partially logical when using transitional words. Basic transitional terms: next, then
- Partially addresses the topic,
- Students can demonstrate minimal connections with attempts to use details
- Literal interpretation of a story; retelling of a story with a lack of inferential thinking
- Consistent errors in conventions that impede readability

Threshold Level 3

- Demonstrates an appropriate structure and organization; simple structures
- Accurate and appropriate vocabulary
- Competent is a good word to think of for this level.
- Occasional errors that impede readability
- Adequately answers the question
- Makes an inference

Threshold Level 4

- Demonstrates an appropriate structure and organization throughout; more comprehensive structures; engaging
- Engaging language and vocabulary.
- Clear structure
- Analytical; know that the student understands at an advanced level
- Precise and clever writing.
- Not only answers correctly, but is precise and insightful.
- Infers and explains this thinking

English Language Arts **Grade 6** **Reading**

Threshold Level 2

- Basic inference with inconsistent citing of a text.
- Determines a theme or central idea without connecting how they are conveyed or having sufficient evidence
- Provides key events or ideas using **limited** understanding of the story's plot or character development.
- Limited understanding of figurative language and how words describe tone and meaning.
- Limited understanding of how one section relates to the whole.
- Has a general understanding of point of view, limited ability to explain.
- Limited ability to connect media to text.
- Inconsistently distinguishes evidence based claims with claims that have no evidence.
- Limited ability to compare and contrast text.

Threshold Level 3

- Thoroughly cites textual evidence as well as drawing inferences.
- Show a thorough understanding of the theme or central idea without inserting opinions or judgements.
- Provides key events or ideas using **thorough** understanding of the story's plot or character development.
- Determines the meaning of figurative and connotative language.
- Thorough understanding of how pieces of text contribute to the overall structure of the text.
- Thorough understanding of point of view and explains how it is conveyed.
- Analyzes different media and how it relates to text.
- Consistently distinguishes evidence based claims from claims based on opinion.
- Show a thorough analysis of texts by comparing and contrasting.

Threshold Level 4

- In depth understanding.
- Precisely determines meanings.
- Answers questions in clever and insightful ways.

Grade 6 Writing and Language

Threshold Level 2

- Begin to address the topic but is lacking in organization, relevant details, clarity, and coherence
- Consistent errors in conventions that may impeded readability
- Basic understanding and use of figurative language
- Inconsistent basic structure
- Simple language

Threshold Level 3

- Addresses the topic, includes relevant details, shows general analysis of the task
- Occasional errors that do not impede readability
- Makes inferences
- Adequate understanding of figurative language
- Demonstrates appropriate organization and structure
- Appropriate language

Threshold Level 4

- Insightfully address the topic, uses purposeful relevant details, shows more in depth analysis of the task
- Few errors that do not impede readability
- Make and explain inferences with details
- Clear understanding of figurative language
- Demonstrates appropriate grade level organization and structure
- Engaging language

English Language Arts Grade 7

Threshold Level 2

Reading

- Students can recognize similarities and/or differences, although there may be some inaccuracies in their understanding of the text.
- Students can recognize a central idea, although they may not connect it to sufficient details or evidence of its development.
- Students can demonstrate basic understanding of structure with limited analysis of its contribution to the meaning of the text as a whole.

Writing and Language

- Students may make many errors and some of the errors impede comprehension of their writing
- Students use basic language to show limited understanding (of the text) when expressing ideas in writing
- Students address the task using some evidence from the text

Threshold Level 3

- Students can compare and contrast accurately and consistently, recognizing similarities and differences.
- Students can provide (recognize) an accurate summary with limited analysis of the central idea.
- Students can demonstrate an understanding of structure with accurate and sufficient analysis of its contribution to the meaning of the text as a whole.

Writing and Language

- Students have some errors, but errors do not impede overall comprehension
- Students show a competent use of language to demonstrate a clear understanding (of the text) when expressing ideas in writing
- Student use relevant evidence to identify and address the purpose of the task

Threshold Level 4

Reading

- Students can compare and contrast using insightful connections between details to show a deeper understanding of the text.
- Students can provide (recognize) a detailed summary with thorough analysis of the central idea.
- Students can demonstrate in-depth understanding of structure with detailed analysis of its contribution to the meaning of the text as a whole.

Language and Writing

- Students may make few errors and the errors do not impede comprehension of their writing
- Students show a precise and purposeful use of language to demonstrate a clear understanding (of the text) when expressing ideas in writing
- Students analyze and evaluate evidence to address the purpose of the task
- Students include details and/or descriptions to support their analysis

English Language Arts Grade 8

Threshold Level 2

Reading

- Students use text-based evidence that provides literal support for the central idea/theme without analyzing its development
- Students inconsistently determine the meaning of words and phrases in text with limited analysis of the impact that word choice has on meaning or tone
- Students inconsistently or insufficiently relate text structure to the bigger picture

Writing/Language

- Students make some errors in conventions and usage that may hinder comprehension
- Students demonstrate inconsistent ability to determine the meaning of some unfamiliar words and/or figurative language
- Students demonstrate some level of understanding of text, with an inconsistent use of relevant evidence
- Students demonstrate a limited level of organization in written response

Threshold Level 3

Reading

- Students use and analyze text-based evidence to determine the central idea/theme and the impact of author's choices on the development of the central idea/theme
- Students can determine the meaning of most words and phrases in text and/or determine the impact that word choice has on meaning or tone
- Students can sufficiently relate text structure to the bigger picture

Writing/Language

- Students make some errors in conventions and usage but most do not hinder comprehension
- Students demonstrate an ability to determine the meaning of most unfamiliar words and figurative language
- Students demonstrate logical and coherent interpretation of the text, with adequate use of relevant evidence
- Students demonstrate a sufficient level of organization in written response

Threshold Level 4

Reading

- Students provide precise details to support their use and analysis of text-based evidence to determine the central idea/theme and the impact of author's choices on the development of the central idea/theme
- Students can determine the meaning of words and phrases in text with precision and analyze the impact that word choice has on meaning or tone

- Students can analyze text structure and how it relates to the bigger picture

Writing/Language

- Students make few errors in conventions and usage that do not hinder comprehension
- Students demonstrate ability to determine the meaning of unfamiliar words and figurative language
- Students demonstrate consistently logical and coherent interpretation of the text, with purposeful use of relevant evidence
- Students demonstrate a clear level of organization in written response

Mathematics Grade 3

Threshold Level 2

3.OA

- Understand the relationship between multiplication and division (inverse operations) with factors less than or equal to 10.
- Identify equivalent expressions that illustrate the commutative property beyond 10
- Identify an arithmetic pattern
- Represent one-step word problems with or without a letter representing the unknown quantity

3.NF

- Use 2, 4 as denominators to represent unit fractions
- Given a model, identify equivalent fractions
- Compare two fractions that have the same denominator
- Express whole numbers as fractions

3.MD

- Tell time to the nearest minute or hours
- Given visual, solve addition word problems in minutes or hours
- Add (without a visual) or subtract (with a visual) to solve one-step word problems involving masses or volumes that are given in the same units
- Use a visual model to demonstrate area and understand that area it is measured in square units
- Given the visual model and the formula, find the area of a rectangular figure

Threshold Level 3

3.OA

- Demonstrate relationship between multiplication and division and becoming more competent in determining products with factors greater than 5 but less than 10.
- Identify the unknown whole number in division equations.
- Identifying the relationship of multiplication and division when the product is unknown (inverse operation)
- Apply associative and/or distributive properties along with the commutative property as strategies to multiply
- Identify arithmetic patterns using addition and/or subtraction
- Represent one-step word problems with a variable and represent two-step word problems with or without a variable

3.NF

- Generate equivalent fractions using 2, 3, 4, 6, 8 as denominators
- Identify equivalent fractions using 2, 3, 4, 6, 8, as denominators
- Identify and represent equivalent fractions on a number line
- Represent fractions using numerators other than 1 with the denominators 2, 3, 4, 6, 8
- Compare two fractions that have the same numerator or the same denominator
- Identify fractions that are equivalent to whole numbers

3.MD

- Tell time and write time to the nearest minute and measure intervals of time
- Solve two-step word problems involving addition and subtraction in minutes and/or hours
- Measure and estimate liquid volumes and mass of objects using standard units of grams (g), kilograms (kg), and liters (l).
- Add, subtract, multiply and/or divide to solve one-step word problems involving masses or volumes that are given in the same units.
- Demonstrate area of rectangles and understand that area it is measured in square units
- Use real-world context to solve mathematical problems involving rectangular areas by multiplying the side lengths

Threshold Level 4

3.OA

- Interpret products and quotients of whole numbers and beginning to articulate and justify reasoning.
- Application of the properties of operations as strategies to multiply and beginning to justify the use of the properties.
- Apply arithmetic patterns when solving word problems and beginning to justify thinking.
- Represent two-step word problems with a letter standing in for the unknown quantity and approaching the use of visual and/or text to justify the reasonableness of the answer

3.NF

- Explain/justify the comparison of two fractions using visual evidence or written evidence
- Physically plot the location of fractions on a number line
- Identify and label a number line given a real-world context
- Evaluate and/or generate equivalent fractions including fractions of whole numbers

3.MD

- Identify the context and solve two-step word problems involving addition and/or subtraction of time intervals
- Create a visual model showing unit squares to demonstrate area of a rectangle
- Solve real-world problems involving the composition of rectilinear figures
- Using the area model to show the distributive property

Mathematics Grade 4

Threshold Level 2

4.OA

- Given a visual model and/or manipulatives, use multiplication to solve some problems involving multiplicative comparisons as multiplicative equations
- Solve one-step word problems using the four operations with one- and two-digit whole numbers.

4.NBT

- Round three-digit whole numbers to the largest place or specific place value
- Identify the place value in any two- or three-digit number is larger as you move to the left
- Compare two- or three-digit whole numbers using base-ten numerals, number names, and expanded form, or inequality symbols.
- Multiply a two-digit by a one-digit number based on place value and the properties of operations (without a visual)
- Divide whole numbers up to two-digit dividends and one-digit divisors 5 or less based on place value, the properties of operations, and/or the relationship between multiplication and division

4.NF

- Identify some equivalent fractions using visual models with denominators 2, 3, 4, 6, 8
- Comparing fractions with words, pictures, or symbols (>, <, =).
- Given a visual model and/or manipulatives solve some mathematical problems involving the addition or subtraction of fractions with like denominators.
- Decompose a fraction into a sum of fractions with the same denominator in at least one way using a visual model.
- Given a visual model solve mathematical problems by recognizing a fraction is a multiple of a unit fraction times a whole number.

Threshold Level 3

4.OA

- Solve two-step word problems using the four operations with whole numbers (multiplying a 2-digit number by 1-digit number or dividing 2-digit dividends by 1-digit divisors with remainders)
- Solve one-step word problems using the four operations with whole numbers (multiplying a 2-digit number by a 2-digit number.
- Represent two-step word problems using equations with a letter standing in for the unknown quantity

4.NBT

- In any two or three digit whole number and some multi-digit whole numbers determine that a digit in one place represents ten times as much as it represents in the place to its right
- Read, write and/or compare multi-digit whole numbers using base-ten numerals, number names in expanded form, and inequality (>, <, =) symbols.
- Round multi-digit whole numbers to the largest place or specific place value
- Divide whole numbers up to 3-digit dividends and one-digit divisors based on place value, the properties of operations, and/or the relationship between multiplication and division

4.NF

- Generate some equivalent fractions using visual models with denominators 2, 3, 4, 5, 6, 8, 10, 12, 100 (e.g., a student can do one denominator but not another)
- Compare two fractions, with like or unlike numerators or denominators, by comparing to a benchmark fraction.
- Solve mathematical problems involving the addition and subtraction of fractions and mixed numbers.
    - Not necessarily word problems yet
- Decompose a fraction into a sum of fractions with the same denominator in at least one way.
- Solve mathematical problems by recognizing a fraction is a multiple of a unit fraction multiplied by a whole number.

Threshold Level 4

4.OA

- Use multiplication or division to solve two-step word problems involving multiplicative comparisons.
- Solve two-step word problems using the four operations with whole numbers (Multiplying a 4-digit number by a 1-digit number, multiplying a 2-digit number by a 2-digit number or dividing up to a 4-digit dividend by a 1-digit divisor)

4.NBT

- Explain that a digit one place value to the left represents ten times as much as it represents as the place to its right
- Evaluate whether a rounded number fits the context
- Illustrate or explain the product by using equations, rectangular arrays, and/or area models
- Illustrate or explain the quotient using equations, rectangular arrays, and/or area models.

4.NF

- Generate equivalent fractions with some denominators 2, 3, 4, 5, 6, 8, 10, 12, 100
- Create or solve mathematical word problems involving the addition and subtraction of fractions and mixed numbers with like denominators by joining and separating parts.
- Recognize a fraction is a multiple of a unit fraction multiplied by a whole number.
- Create or solve mathematical word problems by recognizing a fraction is a multiple of a unit fraction multiplied by a whole number.

Mathematics Grade 5

Threshold 2

- Extend their understanding of the place value system from whole numbers to the tenths (Read, Write, Compare, Round).
- Add, subtract, multiply and divide using strategies based on place values.
- Divide by a 2-digit divisor.
- Add fractions and mixed numbers with **unlike** denominators
- Interpret multiplication as scaling using visual models
- Either visually or numerically, be able to divide whole/unit or unit/whole
- understand a prism maintains the same volume regardless of how it is oriented

Threshold 3

- Extend their understanding of the place value system from whole numbers to the hundredths (Read, Write, Compare, Round).
- Represent powers of 10 using exponents.
- Add, subtract decimals to the hundredths; multiply decimals tenths by tenths or tenths by hundredths; and divide decimals involving tenths and/or hundredths using written strategies based on place values.
- Justify reasonableness of answers.
- solve real-world problems involving the addition and/or subtraction of fractions with unlike denominators
- Interpret product / quotient
- Multiply fractions and mixed numbers using standard algorithm
- Division of unit fraction by whole or whole by unit fraction without the need of visual models

Threshold 4

- Extend their understanding of the place value system to determine that a digit is 1/10 of what it represents in the place to its left.
- Apply understanding exponents with the powers of ten to read, write and compare decimals to the thousandths.
- Add, subtract, multiply and divide decimals and apply to real world context.
- +/- Two **or more** fractions
- Apply multiplication and division for rectangular area problems
- create real-world word problems involving fraction operations

Mathematics Grade 6

Threshold Level 2

- Use ratio reasoning and language to describe and solve
- Computing quotients of fractions without models
- Use visual models to solve basic fraction word problems
- Use negative whole numbers on a number line.
- Plot ordered pairs on all 4 quadrants
- Evaluate basic numerical expressions WITH exponents
- Identify SOME parts of mathematical expressions
- Write numerical expression involving all four operations
- Solve single step equations with whole numbers
- Write an equation or inequality
- Identify solutions to inequalities
- Find the distance between two points on a grid by counting
- Solve percent problems where you have to find the part or percent without having to simply a fraction

Threshold Level 3

- Use rate reasoning and language to describe and solve
- Solve real-world problems with equivalent ratio, rate and percent of a number and simple unit conversion.
- Interpret quotient of fractions without variables
- Order and compare rational numbers
- Use negative numbers without a number line
- Read, Write, AND evaluate numerical and algebraic expressions, including exponents
- Identify parts of algebraic AND numerical expressions
- Identify equivalent expressions
- Solve single step equations including fractions or decimals (fractions only as constants)
- Write an equation or inequality using a real-world problem
- Identify and GRAPH solutions to inequalities
- Identify the relationship between independent and dependent variable
- Identify the reflection of a point
- Solve percent problems where you have to find the part, percent, or whole

Threshold Level 4

- Making a connection between the different representations and strategies.
- Create and solve word problems involving division of fraction by fraction
- Create visual models to solve word problems
- Recognize that when two ordered pairs differ only by signs, the locations of the points are related by multiple reflections across one or both axes.
- Patterns and characteristics of positive and negative numbers.

- Generate equivalent expressions
- Solve AND explain single-step equations and inequalities
- Write an equation using a real-world problem AND Analyze solution
- Recognize that there are an infinite number of solutions for an inequality
- Solve multi-step percent problems

Mathematics Grade 7

7.RP.1-3

Threshold Level 2

- Apply given information to compute unit rate and proportions from a word problem. Identify the concept of a unit rate. Know whether a given scenario represents a proportional relationship.

Threshold Level 3

- More complex rational numbers used to determine if a unit rate exists through a verbal/written description. Real world multi-step application using multiple representations (graph, table, point, etc.) through explicit description. Use multiple representations to determine the unit rate and put it into equation form.

Threshold Level 4

- Explain if multiple points given multiple representations represent a proportional relationship in the context of the problem. Understand limitations of a unit rate by putting reasonings into their own words. Apply unit rates in different scenarios using multiple accurate reasons that describe the given problem. Interpret the point (x,y) in the context of the problem.

7.NS.1-3

Threshold Level 2 (Some)

- Use properties of operations with integers to produce an accurate solution. Identify and understand scenarios where the situation results in a quantity of zero. Solving some two step real world problems accurately. Using a number line to show addition and subtraction of integers.

Threshold Level 3 (All)

- Use properties of operations with rational numbers to produce an accurate solution. Solving all two step real world problems accurately. Using a number line to show addition and subtraction of rational numbers as well as showing absolute value. Convert rational numbers to a decimal using accurate long division.

Threshold Level 4

- Describe through justification and show the steps involved in solving multi-step word problems using the four operations. Use the additive inverse as the complexity of the problem increases with real world situations. Understand the vocabulary of a real world situation and use that understanding to solve those multi-step problems.

Threshold Level 2

7.EE.1,2

- Students can add AND subtract linear expressions.
- Students can rewrite expressions with BOTH positive AND negative integer coefficients.
- Students can recognize equivalent expressions.

7.EE.3,4

- Students must now solve TWO-STEP mathematical problems.
- Students must be able to SOLVE TWO-STEP linear equations AND inequalities.

Threshold Level 3

7.EE.1,2

- Students must be able to use the distributive property to FACTOR and EXPAND linear expressions. (multiply both terms)
- Students must be able rewrite expressions with RATIONAL coefficients.

7.EE.3,4

- Students must solve MULTI-STEP mathematical situations using algebra.
- Students must DERIVE and/or SOLVE linear equations/inequalities in linear or factored form.

Threshold Level 4

7.EE.1,2

- Students must understand and EXPLAIN why the algebra is applicable in the context of the situation.

7.EE.3,4

- EXPLAIN the process of solving linear equations using an algebraic AND an arithmetic solution.
- Students must GRAPH and INTERPRET solution sets of linear inequalities.
- Students must KNOW/EXPLAIN if their answer makes SENSE in context of the problem.

Mathematics Grade 8

8.EE.1,3,4

Threshold Level 2

- Multiply and Divide numbers in scientific notation.
- Use scientific notation to estimate very large quantities and interpret its meaning.
- EVALUATE simple numerical expressions (monomials) using properties of INTEGER (any) exponents

Threshold Level 3

- GENERATE and/or EVALUATE equivalent expressions.
- Apply ONE property of integer exponents within a real-world context.
- ADD, SUBTRACT, Multiply and Divide numbers in scientific notation.
- Use scientific notation to estimate very large and/or very small quantities, interpret its meaning, compare the quantities

Threshold Level 4

- Apply TWO properties of integer exponents within a real-world context.
- Add, Subtract, Multiply and Divide numbers in scientific notation AND decimal notation
-
- **8.EE.5,6**

Threshold Level 2

- ACCURATELY graph a proportional relationship given a table.
- KNOW that the unit rate IS the slope.
- KNOW that when presented with the form y=mx, m=slope and the line passes through the origin.

Threshold Level 3

- ACCURATELY graph a proportional relationship given a table and/or an equation (y=mx).
- KNOW that the unit rate IS the slope and be able to interpret its meaning in context.
- Compare two different proportional relationships represented in different ways.
- Use similar triangles to determine the slope.
- KNOW that when presented with the form y=mx+b, m=slope and b=y-intercept.

Threshold Level 4

- ACCURATELY WRITE and graph a proportional relationship given a table and/or an equation (y=mx and y=mx+b).
- KNOW that the unit rate IS the slope and be able to interpret its meaning in context using real-world examples.

8.EE.7,8

Threshold Level 2

- Solve linear equations with rational numbers.
- Solve linear equations by combining like terms and/or distributive property.
- Graph pairs of linear equations to find point of intersection.
- Attempt to solve pairs of equations without a graph.

Threshold Level 3

- Solve linear equations with different types of solutions (one, infinite or none).
- Recognize what the point of intersection means on the graph.
- Understand the relation between a graphic solution and algebraic solution.

Threshold Level 4

- Justify method used to solve.
- Solve real world system of linear equations to identify and understand the solution.
- Explain relationships between solving graphically versus algebraically.

8.F.1-3

Threshold Level 2

- Graph a function from ordered pairs or equation in $y = mx + b$ form.
- Compare functions in the same form.
- Identify some linear functions by examining a table or graph.

Threshold Level 3

- Determine if a set of ordered pairs represents a function.
- Compare functions in different forms.
- Identify functions that are nonlinear.
- Interpret the equation $y = mx + b$ as defining a linear function.

Threshold Level 4

- Create functions in various forms to represent real-world situations.
- Compare properties of functions.
- Create examples that represent linear and nonlinear functions.
- Construct linear functions given specific properties.

8.G.1-5

Threshold Level 2

- Describe the change made to a two dimensional shape under a specific transformation (excluding dilation).
- identify congruence and similarity between transformations (excluding dilation).

Threshold Level 3

- Describe the change from transformations using coordinates.
- Justify congruence or similarity of two figures after one transformation.

Threshold Level 4

- Describe the change from more than one transformation.
- Justify congruence or similarity of two figures after more than one transformation.

Appendix D: Example of standards review item map

Sequence	Item ID	Item Type	CR Score Point	Level 2 Cut	Level 3 Cut	Level 4 Cut
1	NY_1440	MC		1	1	1
2	NYM160	MC		2	2	2
3	NYM1601	MC		3	3	3
4	NYM1602	MC		4	4	4
5	NY_164	CR2	1	5	5	5
6	NY_1340	MC		6	6	6
7	NY154	MC		7	7	7
8	NY105	MC		8	8	8
9	NYM1603	MC		9	9	9
10	NY_16403	CR2	2	10	10	10
11	NY_164030	MC		11	11	11
12	NY_164031	MC		12	12	12
13	NYM16037	MC		13	13	13
14	NYM11144	MC		14	14	14
15	NY_150203	CR2	1	15	15	15
16	NY_1646	MC		16	16	16
17	NYM11836	MC		17	17	17
18	NYM322163	CR2	1	18	18	18
19	NYM16824	MC		19	19	19
20	NYM160035	MC		20	20	20
21	NYM160193	CR2	1	21	21	21
22	NYM1602423	CR2	1	22	22	22
23	NYM160151	CR2	1	23	23	23
24	NYM13287	MC		24	24	24
25	NYM3611	MC		25	25	25
26	NY1154	MC		26	26	26
27	NY_130304	CR3	1	27	27	27
28	NY_15403	CR2	2	28	28	28
29	NY_16403	MC		29	29	29
30	NY_16403	MC		30	30	30
31	NYM1378	MC		31	31	31
32	NYM16211	MC		32	32	32
33	NYM16023	CR2	2	33	33	33
34	NYM16033	MC		34	34	34
35	NYM16033	CR2	2	35	35	35
36	NYM1151	CR2	2	36	36	36
37	NYM13210	MC		37	37	37
38	NY_16404	CR3	2	38	38	38
39	NY_16403	MC		39	39	39

40	NYM16174	MC		40	40	40
41	NY_16303	CR3	3	41	41	41
42	NY_15303	MC		42	42	42
43	NYM16021	CR2	2	43	43	43
44	NYM1686	MC		44	44	44
45	NYM1658	MC		45	45	45

Appendix E: Sample of feedback provided after Round 1 and Round 2

Table E.1 Example of summary table provided after Round 1 and Round 2

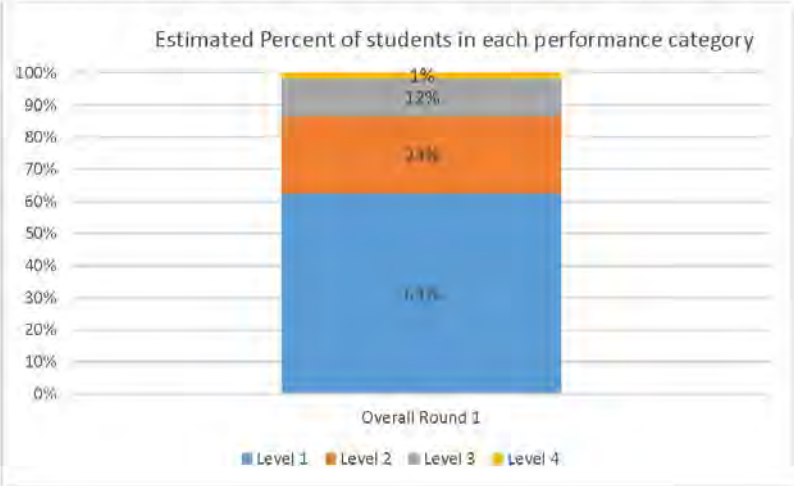
	Minimum	Maximum	Median
Level 1			
Level 2	1	20	7
Level 3	17	49	21
Level 4	37	62	56

Table E.2: Example of summary and impact table provided after Round 1 and Round 2

	Minimum	Maximum	Median	% students
Level 1				63%
Level 2	1	20	7	24%
Level 3	17	49	21	12%
Level 4	37	62	56	1%

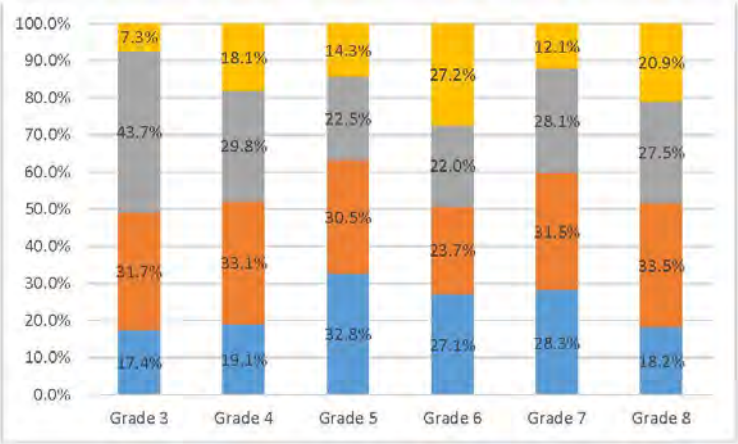
Figure E.1: Sample of figure used to demonstrate rater variability after Round 1 and Round 2

Figure E.2: Sample of figure used to demonstrate impact data after Round 1 and Round 2



Appendix F: Example of vertical articulation Impact data

Figure F.1 Example impact data used during vertical articulation



Appendix G: Round 2 and Vertical Articulation Cut Point Recommendations

Table G1: Recommended cut points after vertical articulation for the NYSED ELA assessments

Final recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	17.4%	19.1%	32.8%	27.1%	28.3%	18.2%
Level 2	Page #	5	4	15	11	18	14
	Theta	-0.9061	-0.9521	-0.3710	-0.5556	-0.5325	-0.8798
	Raw score cut	12	13	24	23	24	23
	% students	31.7%	33.1%	30.5%	23.7%	31.5%	33.5%
Level 3	Page #	13	19	30	28	36	33
	Theta cut point	0.1046	0.2332	0.4600	0.1718	0.3917	0.1646
	Raw score cut	19	21	31	30	33	33
	% students	43.7%	29.8%	22.5%	22.0%	28.1%	27.5%
Level 4	Page #	36	30	42	40	48	45
	Theta cut point	1.5069	1.0535	1.2514	0.7967	1.2212	0.9229
	Raw score cut	29	27	36	35	40	39
	% students	7.3%	18.1%	14.3%	27.2%	12.1%	20.9%

Table G2: Recommended cut points after Round 2 for the NYSED ELA assessments

Round 2 recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	17.4%	19.1%	32.8%	27.1%	28.3%	18.2%
Level 2	Page #	5	4	15	11	18	14
	Theta	-0.9061	-0.9521	-0.3710	-0.5556	-0.5325	-0.8798
	Raw score cut	12	13	24	23	24	23
	% students	31.7%	33.1%	30.5%	23.7%	31.5%	33.5%
Level 3	Page #	13	19	30	28	36	33
	Theta cut point	0.1046	0.2332	0.4600	0.1718	0.3917	0.1646
	Raw score cut	19	21	31	30	33	33
	% students	43.7%	29.8%	22.5%	22.0%	28.1%	27.5%
Level 4	Page #	36	30	42	40	48	45
	Theta cut point	1.5069	1.0535	1.2514	0.7967	1.2212	0.9229
	Raw score cut	29	27	36	35	40	39
	% students	7.3%	18.1%	14.3%	27.2%	12.1%	20.9%

Table G3: Recommended cut points after vertical articulation for the NYSED Math assessments

Final recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	23.7%	25.0%	31.5%	30.1%	31.7%	36.1%
Level 2	Page #	9	6	9	7	6	4
	Theta	-0.6954	-0.6109	-0.4231	-0.4525	-0.4241	-0.2439
	Raw score cut	18	19	19	16	19	18
	% students	21.9%	26.3%	24.1%	25.1%	25.9%	32.4%
Level 3	Page #	24	24	21	21	21	23
	Theta cut point	-0.0020	0.0959	0.2137	0.2201	0.2938	0.5379
	Raw score cut	26	30	28	25	31	31
	% students	30.9%	23.2%	23.4%	23.4%	23.6%	18.7%
Level 4	Page #	38	42	39	37	48	45
	Theta cut point	0.7955	0.7825	0.8416	0.8366	0.9632	1.1324
	Raw score cut	35	38	37	35	42	41
	% students	23.5%	25.4%	21.0%	21.4%	18.8%	12.8%

Table G4: Recommended cut points after Round 2 for the NYSED Math assessments

Round 2 recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	23.7%	25.0%	31.5%	30.1%	31.7%	36.1%
Level 2	Page #	9	6	9	7	6	4
	Theta	-0.6954	-0.6109	-0.4231	-0.4525	-0.4241	-0.2439
	Raw score cut	18	19	19	16	19	18
	% students	21.9%	26.3%	24.1%	27.7%	25.9%	32.4%
Level 3	Page #	24	24	21	22	21	23
	Theta cut point	-0.0020	0.0959	0.2137	0.0062	0.2938	0.5379
	Raw score cut	26	30	28	26	31	31
	% students	30.9%	23.2%	26.1%	20.8%	23.6%	18.7%
Level 4	Page #	38	42	41	37	48	45
	Theta cut point	0.7955	0.7825	1.0200	0.8366	0.9632	1.1324
	Raw score cut	35	38	38	35	42	41
	% students	23.5%	25.4%	18.4%	21.4%	18.8%	12.8%

Figure G1: Impact data for English Language Arts after vertical articulation

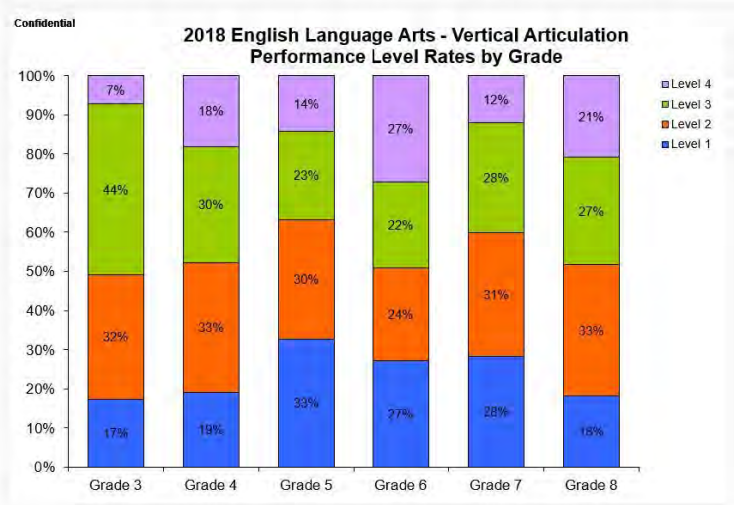


Figure G2: Impact data for English Language Arts after Round

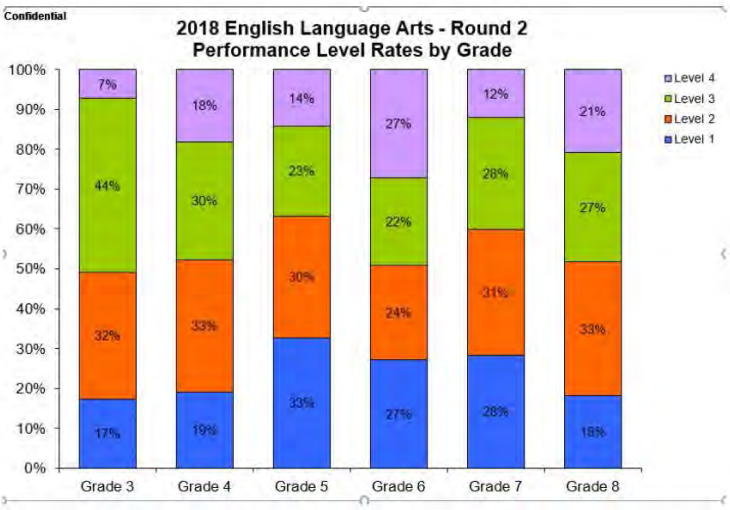


Figure G3: Impact data for Mathematics Vertical Articulation

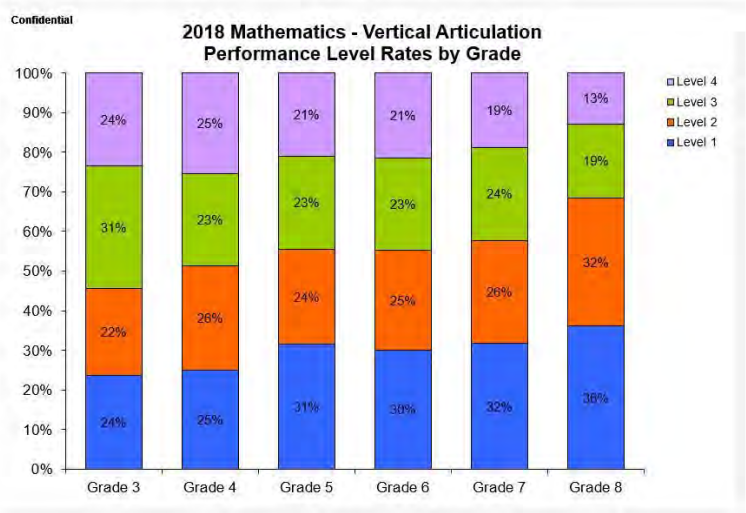
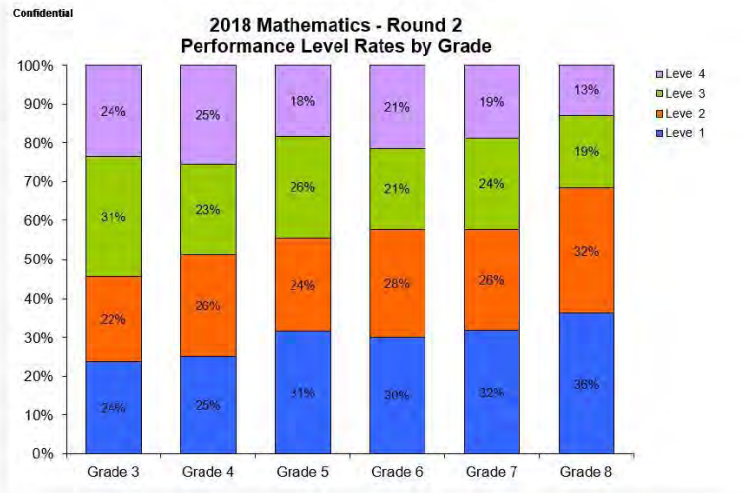


Figure G4: Impact data for Mathematics after round 2



Appendix H: Total group recommendations and standard error of the median by round

Table H1: Cut point recommendations by round, English Grade 3

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	3.5	0.58	-1.31	15	0.67	0.37	36.5	0.88	1.53
Round 2	5	0.41	-0.91	13	0.65	0.10	36	0.33	1.51

Table H2: Cut point recommendations by round, English Grade 4

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	4	0.55	-0.95	19	1.02	0.23	30	0.74	1.05
Round 2	4	0.28	0.95	19	0.41	0.23	30	0.53	1.05

Table H3: Cut point recommendations by round, English Grade 5

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	15	1.54	-0.37	32	0.84	0.66	42	1.01	1.25
Round 2	15	0.38	-0.37	30	0.59	0.46	42	1.33	1.25

Table H4: Cut point recommendations by round, English Grade 6

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	11	0.38	-0.56	32	0.96	0.36	40.5	0.28	0.94
Round 2	11	0.25	-0.56	28	1.06	0.17	40	0.17	0.80

Table H5: Cut point recommendations by round, English Grade 7

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	19	0.47	-0.53	36.5	0.50	0.42	49	0.51	1.36
Round 2	18	0.49	-0.53	36	0.44	0.39	48	0.46	1.22

Table H6: Cut point recommendations by round, English Grade 8

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	13	0.51	-1.07	32.5	0.82	0.16	45	0.47	0.92
Round 2	14	0.36	-0.88	33	0.33	0.16	45	0.52	0.92

Table H7: Cut point recommendations by round, Math Grade 3

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	9	0.42	-0.70	25	1.02	0.11	38	0.65	0.80
Round 2	9	0.21	-0.70	24	0.49	-0.00	38	0.60	0.80

Table H8: Cut point recommendations by round, Math Grade 4

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	6	0.28	-0.61	24	1.71	0.09	42	0.21	0.78
Round 2	6	0.33	-0.61	24	0.14	0.09	42	0.21	0.78

Table H9: Cut point recommendations by round, Math Grade 5

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	9	0.54	-0.42	21	1.12	0.21	42	0.67	1.06
Round 2	9	0.33	-0.42	21	0.49	0.21	41	0.44	1.02

Table H10: Cut point recommendations by round, Math Grade 6

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	7	0.68	-0.45	21.5	1.11	0.23	36	0.47	0.80
Round 2	7	0.26	-0.45	22	0.50	0.23	37	0.41	0.84

Table H11: Cut point recommendations by round, Math Grade 7

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	6	0.56	-0.42	21	1.28	0.29	48	1.08	0.96
Round 2	6	0.14	-0.42	21	0.21	0.29	48	0.78	0.96

Table H12: Cut point recommendations by round, Math Grade 8

	Level 2			Level 3			Level 4		
	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	4	0.61	-0.24	29	2.48	0.61	49	0.77	1.24
Round 2	4	0.35	-0.24	23	0.58	0.54	45	0.54	1.13

Appendix I: Standard error bands for ratings

Table I.1: English standard error estimates using panelists level ratings

	Level 2			Level 3			Level 4		
	Page #	Theta	% of students at or above	Page #	Theta	% of students at or above	Page #	Theta	% of students at or above
GRADE 3									
- 2 SEM	4	-1.31	89.4%	12	0.01	55.9%	35	1.49	7.3%
-1 SEM	5	-0.91	82.6%	12	0.01	55.9%	36	1.51	7.3%
Recommendation	5	-0.91	82.6%	13	0.10	50.9%	36	1.51	7.3%
+1 SEM	5	-0.91	82.6%	14	0.16	46.1%	36	1.51	7.3%
+ 2 SEM	6	-0.81	78.7%	14	0.16	46.1%	37	1.53	7.3%
GRADE 4									
- 2 SEM	3	-1.07	84.2%	18	0.18	47.8%	29	0.95	22.8%
-1 SEM	4	-0.95	80.9%	19	0.23	47.8%	29	0.95	22.8%
Recommendation	4	-0.95	80.9%	19	0.23	47.8%	30	1.05	18.1%
+1 SEM	4	-0.95	80.9%	19	0.23	47.8%	31	1.08	18.1%
+ 2 SEM	5	-0.95	80.9%	20	0.31	42.9%	31	1.08	18.1%
GRADE 5									
- 2 SEM	14	-0.48	70.8%	29	0.40	41.5%	39	1.00	18.2%
-1 SEM	15	-0.37	67.2%	29	0.40	41.5%	41	1.12	14.3%
Recommendation	15	-0.37	67.2%	30	0.46	36.8%	42	1.25	14.3%
+1 SEM	15	-0.37	67.2%	31	0.63	32.0%	43	1.27	14.3%
+ 2 SEM	16	-0.30	67.2%	31	0.63	32.0%	45	1.28	10.6%
GRADE 6									
- 2 SEM	10	-0.74	78.1%	26	0.07	53.1%	40	0.80	27.2%
-1 SEM	11	-0.56	72.9%	27	0.09	49.2%	40	0.80	27.2%
Recommendation	11	-0.56	72.9%	28	0.17	49.2%	40	0.80	27.2%
+1 SEM	11	-0.56	72.9%	29	0.21	45.1%	40	0.80	27.2%
+ 2 SEM	12	-0.42	70.0%	30	0.25	45.1%	40	0.80	27.2%

GRADE 7									
- 2 SEM	17	-0.61	74.5%	35	0.37	40.2%	47	1.19	12.1%
-1 SEM	18	-0.53	71.7%	36	0.39	40.2%	48	1.22	12.1%
Recommendation	18	-0.53	71.7%	36	0.39	40.2%	48	1.22	12.1%
+1 SEM	18	-0.53	71.7%	36	0.39	40.2%	48	1.22	12.1%
+ 2 SEM	19	-0.53	71.7%	37	0.42	40.2%	49	1.36	12.1%
GRADE 8									
- 2 SEM	13	-1.07	85.9%	32	0.05	52.6%	44	0.77	25.5%
-1 SEM	14	-0.88	81.8%	33	0.16	48.3%	44	0.77	25.5%
Recommendation	14	-0.88	81.8%	33	0.16	48.3%	45	0.92	20.9%
+1 SEM	14	-0.88	81.8%	33	0.16	48.3%	46	1.02	16.4%
+ 2 SEM	15	-0.81	79.4%	34	0.20	43.9%	46	1.02	16.4%

Table I.2: Mathematics standard error estimates using panelists level ratings

	Level 2			Level 3			Level 4		
	Page #	Theta	% of students at or above	Page #	Theta	% of students at or above	Page #	Theta	% of students at or above
GRADE 3									
- 2 SEM	9	-0.70	76.3%	23	-0.00	54.4%	37	0.77	23.5%
-1 SEM	9	-0.70	76.3%	24	-0.00	54.4%	37	0.77	23.5%
Recommendation	9	-0.70	76.3%	24	-0.00	54.4%	38	0.80	23.5%
+1 SEM	9	-0.70	76.3%	24	-0.00	54.4%	39	0.83	23.5%
+ 2 SEM	9	-0.70	76.3%	25	0.11	47.9%	39	0.83	23.5%
GRADE 4									
- 2 SEM	5	-0.79	79.4%	24	0.10	48.7%	42	0.78	25.4%
-1 SEM	6	-0.61	75.0%	24	0.10	48.7%	42	0.78	25.4%
Recommendation	6	-0.61	75.0%	24	0.10	48.7%	42	0.78	25.4%
+1 SEM	6	-0.61	75.0%	24	0.10	48.7%	42	0.78	25.4%
+ 2 SEM	7	-0.57	72.8%	24	0.10	48.7%	42	0.78	25.4%
GRADE 5									
- 2 SEM	8	-0.48	68.5%	20	0.16	44.5%	38	0.83	23.7%
-1 SEM	9	-0.42	68.5%	21	0.21	44.5%	39	0.84	21.0%
Recommendation	9	-0.42	68.5%	21	0.21	44.5%	39	0.84	21.0%
+1 SEM	9	-0.42	68.5%	21	0.21	44.5%	39	0.84	21.0%
+ 2 SEM	10	-0.39	65.8%	22	0.26	41.9%	40	0.94	18.4%
GRADE 6									
- 2 SEM	6	-0.58	72.9%	20	0.11	47.4%	36	0.80	23.5%
-1 SEM	7	-0.45	69.9%	20	0.11	47.4%	37	0.84	21.4%
Recommendation	7	-0.45	69.9%	21	0.22	44.8%	37	0.84	21.4%
+1 SEM	7	-0.45	69.9%	22	0.23	42.2%	37	0.84	21.4%
+ 2 SEM	8	-0.38	67.0%	22	0.23	42.2%	38	0.85	21.4%

GRADE 7									
- 2 SEM	6	-0.42	68.3%	21	0.29	42.4%	45	0.89	21.0%
-1 SEM	6	-0.42	68.3%	21	0.29	42.4%	47	0.93	18.8%
Recommendation	6	-0.42	68.3%	21	0.29	42.4%	48	0.96	18.8%
+1 SEM	6	-0.42	68.3%	21	0.29	42.4%	49	1.05	16.5%
+ 2 SEM	6	-0.42	68.3%	21	0.29	42.4%	51	1.07	16.5%
GRADE 8									
- 2 SEM	3	-0.34	66.8%	22	0.49	33.7%	44	1.11	14.4%
-1 SEM	4	-0.24	63.9%	22	0.49	33.7%	44	1.11	14.4%
Recommendation	4	-0.24	63.9%	23	0.54	31.5%	45	1.13	12.8%
+1 SEM	4	-0.24	63.9%	24	0.56	31.5%	46	1.17	12.8%
+ 2 SEM	5	-0.18	61.0%	24	0.56	31.5%	46	1.17	12.8%

Appendix J: Survey results for all panels

Survey #1 ELA Grade 3 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

	Strongly Disagree	Disagree	Agree	Strongly Agree
General Session				
The general session helped me understand the purpose of the standard setting workshop.	-	-	8	1
The general session helped me understand my role in determining the cut scores for New York state assessments	-	-	8	1
The general session clarified what procedures were going to be followed and how the work would progress.	-	-	8	1
Performance Level Descriptions (PLDs)				
The review of the current performance level descriptions provided me sufficient information to allow me to understand the role and purpose of the PLDs.	-	-	8	1
The current PLDs had clear distinctions across the performance levels.	-	1	6	2
I understand how and when the threshold PLDs will be used during the standards review process.	-	1	6	2
I was provided sufficient opportunity to discuss the current PLDs and the threshold PLDs that we developed.	-	-	7	2
The threshold PLDs that we developed clearly identify knowledge and skills near the different cut points.	-	-	7	2

Survey #1 ELA Grade 5 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

	Strongly Disagree	Disagree	Agree	Strongly Agree
General Session				
The general session helped me understand the purpose of the standard setting workshop.	-	-	7	3
The general session helped me understand my role in determining the cut scores for New York state assessments	-	-	6	4
The general session clarified what procedures were going to be followed and how the work would progress.	-	-	5	5
Performance Level Descriptions (PLDs)				
The review of the current performance level descriptions provided me sufficient information to allow me to understand the role and purpose of the PLDs.	-	-	6	4
The current PLDs had clear distinctions across the performance levels.	-	-	7	3
I understand how and when the threshold PLDs will be used during the standards review process.	-	-	3	7
I was provided sufficient opportunity to discuss the current PLDs and the threshold PLDs that we developed.	-	-	2	8
The threshold PLDs that we developed clearly identify knowledge and skills near the different cut points.	-	-	4	6

Survey #1 ELA Grade 7 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

	Strongly Disagree	Disagree	Agree	Strongly Agree
General Session				
The general session helped me understand the purpose of the standard setting workshop.	-	-	6	4
The general session helped me understand my role in determining the cut scores for New York state assessments	-	-	8	2
The general session clarified what procedures were going to be followed and how the work would progress.	-	-	4	6
Performance Level Descriptions (PLDs)				
The review of the current performance level descriptions provided me sufficient information to allow me to understand the role and purpose of the PLDs.	-	-	6	4
The current PLDs had clear distinctions across the performance levels.	-	1	5	4
I understand how and when the threshold PLDs will be used during the standards review process.	-	2	3	5
I was provided sufficient opportunity to discuss the current PLDs and the threshold PLDs that we developed.	-	1	3	6
The threshold PLDs that we developed clearly identify knowledge and skills near the different cut points.	-	1	8	1

COMMENTS

ELA Grades 7 and 8

I think the PLDs are partially done at this point

Survey #1 Math Grade 3 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

	Strongly Disagree	Disagree	Agree	Strongly Agree
General Session				
The general session helped me understand the purpose of the standard setting workshop.	-	-	4	5
The general session helped me understand my role in determining the cut scores for New York state assessments	-	-	5	4
The general session clarified what procedures were going to be followed and how the work would progress.	-	1	4	4
Performance Level Descriptions (PLDs)				
The review of the current performance level descriptions provided me sufficient information to allow me to understand the role and purpose of the PLDs.	-	-	6	3
The current PLDs had clear distinctions across the performance levels.	-	1	6	2
I understand how and when the threshold PLDs will be used during the standards review process.	-	-	7	2
I was provided sufficient opportunity to discuss the current PLDs and the threshold PLDs that we developed.	-	-	5	4
The threshold PLDs that we developed clearly identify knowledge and skills near the different cut points.	-	-	5	4

Survey #1 Math Grade 5 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

	Strongly Disagree	Disagree	Agree	Strongly Agree
General Session				
The general session helped me understand the purpose of the standard setting workshop.	-	-	3	7
The general session helped me understand my role in determining the cut scores for New York state assessments	-	-	4	6
The general session clarified what procedures were going to be followed and how the work would progress.	-	-	5	5
Performance Level Descriptions (PLDs)				
The review of the current performance level descriptions provided me sufficient information to allow me to understand the role and purpose of the PLDs.	-	-	3	7
The current PLDs had clear distinctions across the performance levels.	-	1	5	4
I understand how and when the threshold PLDs will be used during the standards review process.	-	-	3	7
I was provided sufficient opportunity to discuss the current PLDs and the threshold PLDs that we developed.	-	-	-	10
The threshold PLDs that we developed clearly identify knowledge and skills near the different cut points.	-	-	5	5

Survey #1 Math Grade 7 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

	Strongly Disagree	Disagree	Agree	Strongly Agree
General Session				
The general session helped me understand the purpose of the standard setting workshop.	-	1	5	4
The general session helped me understand my role in determining the cut scores for New York state assessments	-	-	7	3
The general session clarified what procedures were going to be followed and how the work would progress.	-	-	4	6
Performance Level Descriptions (PLDs)				
The review of the current performance level descriptions provided me sufficient information to allow me to understand the role and purpose of the PLDs.	-	2	4	4
The current PLDs had clear distinctions across the performance levels.	-	1	8	1
I understand how and when the threshold PLDs will be used during the standards review process.	-	2	6	2
I was provided sufficient opportunity to discuss the current PLDs and the threshold PLDs that we developed.	-	-	3	7
The threshold PLDs that we developed clearly identify knowledge and skills near the different cut points.	-	1	7	2

Survey #1 Survey results ELA All panels – Mean responses

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

	English Language Arts			Mathematics		
	Panel 3/4	Panel 5/6	Panel 7/8	Panel 3/4	Panel 5/6	Panel 7/8
General Session						
The general session helped me understand the purpose of the standard setting workshop.	3.1	3.3	3.4	3.6	3.7	3.3
The general session helped me understand my role in determining the cut scores for New York state assessments	3.1	3.4	3.2	3.4	3.6	3.3
The general session clarified what procedures were going to be followed and how the work would progress.	3.1	3.5	3.6	3.3	3.5	3.6
Performance Level Descriptions (PLDs)						
The review of the current performance level descriptions provided me sufficient information to allow me to understand the role and purpose of the PLDs.	3.1	3.4	3.4	3.3	3.7	3.2
The current PLDs had clear distinctions across the performance levels.	3.1	3.3	3.3	3.1	3.3	3.0
I understand how and when the threshold PLDs will be used during the standards review process.	3.1	3.7	3.3	3.2	3.7	3.0
I was provided sufficient opportunity to discuss the current PLDs and the threshold PLDs that we developed.	3.2	3.8	3.5	3.4	4.0	3.7
The threshold PLDs that we developed clearly identify knowledge and skills near the different cut points.	3.2	3.6	3.0	3.4	3.5	3.1

Survey #2 ELA Grade 3 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Training

Training	Strongly Disagree	Disagree	Agree	Strongly Agree
The training on how to complete the Bookmark procedure allowed me to understand how to complete the work.	-	-	4	4
The facilitator of our session walked through an example of completing a Bookmark rating and explained how it was to be completed.	-	-	5	3
The facilitator of our session allowed panelists the opportunity to ask questions and took steps to ensure that all panelists understand how to complete the Bookmark ratings.	-	-	3	5
The materials provided as training tools were clear, understandable and useful during the workshop.	-	-	4	4
I am confident my recorded ratings reflect my best judgment on where to place the Bookmark.	-	-	1	7

Survey #2 ELA Grade 5 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Training

Training	Strongly Disagree	Disagree	Agree	Strongly Agree
The training on how to complete the Bookmark procedure allowed me to understand how to complete the work.	-	-	2	8
The facilitator of our session walked through an example of completing a Bookmark rating and explained how it was to be completed.	-	-	2	8
The facilitator of our session allowed panelists the opportunity to ask questions and took steps to ensure that all panelists understand how to complete the Bookmark ratings.	-	-	1	9
The materials provided as training tools were clear, understandable and useful during the workshop.	-	-	4	6
I am confident my recorded ratings reflect my best judgment on where to place the Bookmark.	-	-	1	9

Survey #2 ELA Grade 7 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Training

Training	Strongly Disagree	Disagree	Agree	Strongly Agree
The training on how to complete the Bookmark procedure allowed me to understand how to complete the work.	-	-	4	6
The facilitator of our session walked through an example of completing a Bookmark rating and explained how it was to be completed.	-	-	5	5
The facilitator of our session allowed panelists the opportunity to ask questions and took steps to ensure that all panelists understand how to complete the Bookmark ratings.	-	-	2	8
The materials provided as training tools were clear, understandable and useful during the workshop.	-	-	4	6
I am confident my recorded ratings reflect my best judgment on where to place the Bookmark.	-	-	4	6

COMMENTS

ELA Grades 7 and 8

- Lori is very knowledgeable and helpful in clarifying confusion, and the process run smooth
- Need more time to read passages
- suggest more time to review passages for content during standard review. However, send PLDs to members prior to attending
- provide more time to read PLDs & passages in order to be more knowledgeable of test

Survey #2 Math Grade 3 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Training

Training	Strongly Disagree	Disagree	Agree	Strongly Agree
The training on how to complete the Bookmark procedure allowed me to understand how to complete the work.	-	-	2	7
The facilitator of our session walked through an example of completing a Bookmark rating and explained how it was to be completed.	-	-	2	7
The facilitator of our session allowed panelists the opportunity to ask questions and took steps to ensure that all panelists understand how to complete the Bookmark ratings.	-	-	-	9
The materials provided as training tools were clear, understandable and useful during the workshop.	-	-	1	8
I am confident my recorded ratings reflect my best judgment on where to place the Bookmark.	-	-	2	7

COMMENTS

Math Grades 3 and 4

- When we review the items, it would be better if we did not see the green boxes

Survey #2 Math Grade 5 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Training

Training	Strongly Disagree	Disagree	Agree	Strongly Agree
The training on how to complete the Bookmark procedure allowed me to understand how to complete the work.	-	-	1	9
The facilitator of our session walked through an example of completing a Bookmark rating and explained how it was to be completed.	-	-	-	10
The facilitator of our session allowed panelists the opportunity to ask questions and took steps to ensure that all panelists understand how to complete the Bookmark ratings.	-	-	-	10
The materials provided as training tools were clear, understandable and useful during the workshop.	-	-	3	7
I am confident my recorded ratings reflect my best judgment on where to place the Bookmark.	-	-	1	9

COMMENTS

Math Grades 5 and 6

- I now have a better understanding of the process and feel confident the "cut-off" is fair and reflect student progress
- Providing "equated" bookmarks should not be provided before panelists do their work to prevent bias

Survey #2 Math Grade 7 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Training

Training	Strongly Disagree	Disagree	Agree	Strongly Agree
The training on how to complete the Bookmark procedure allowed me to understand how to complete the work.	-	-	6	2
The facilitator of our session walked through an example of completing a Bookmark rating and explained how it was to be	-	1	1	6
The facilitator of our session allowed panelists the opportunity to ask questions and took steps to ensure that all panelists understand how to complete the Bookmark ratings.	-	-	1	7
The materials provided as training tools were clear, understandable and useful during the workshop.	-	-	5	3
I am confident my recorded ratings reflect my best judgment on where to place the Bookmark.	-	-	1	7

Survey #2 Survey results All panels – Mean responses

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

	English Language Arts			Mathematics		
	Panel 3/4	Panel 5/6	Panel 7/8	Panel 3/4	Panel 5/6	Panel 7/8
The training on how to complete the Bookmark procedure allowed me to understand how to complete the work.	3.5	3.8	3.6	3.8	3.9	3.3
The facilitator of our session walked through an example of completing a Bookmark rating and explained how it was to be completed.	3.4	3.8	3.5	3.8	4.0	3.6
The facilitator of our session allowed panelists the opportunity to ask questions and took steps to ensure that all panelists understand how to complete the Bookmark ratings.	3.6	3.9	3.8	4.0	4.0	3.9
The materials provided as training tools were clear, understandable and useful during the workshop.	3.5	3.6	3.6	3.8	3.7	3.4
I am confident my recorded ratings reflect my best judgment on where to place the Bookmark.	3.9	3.9	3.6	3.8	3.9	3.9

Survey #3 ELA Grade 3 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Feedback – Comprehension	Strongly Disagree	Disagree	Agree	Strongly Agree
Based upon the feedback, I understood my recommended cut score and how I compared to other panelists.	-	-	4	4
Based upon the feedback, I understood the group's recommended cut score and the variability of the ratings across the panelists.	-	-	4	4
Based upon the feedback, I understood the estimated difficulty level of the items in the ordered item booklet (OIB) and how difficulty relates to our cut scores.	-	-	4	4
Based upon the feedback, I understood the estimated percentage of students that would be classified into each of the four performance categories.	-	-	4	4

Survey #3 ELA Grade 5 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Feedback – Comprehension	Strongly Disagree	Disagree	Agree	Strongly Agree
Based upon the feedback, I understood my recommended cut score and how I compared to other panelists.	-	-	-	10
Based upon the feedback, I understood the group's recommended cut score and the variability of the ratings across the panelists.	-	-	2	8
Based upon the feedback, I understood the estimated difficulty level of the items in the ordered item booklet (OIB) and how difficulty relates to our cut scores.	-	-	2	8
Based upon the feedback, I understood the estimated percentage of students that would be classified into each of the four performance categories.	-	-	1	9

Survey #3 ELA Grade 7 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Feedback – Comprehension	Strongly Disagree	Disagree	Agree	Strongly Agree
Based upon the feedback, I understood my recommended cut score and how I compared to other panelists.	-	-	1	9
Based upon the feedback, I understood the group's recommended cut score and the variability of the ratings across the panelists.	-	-	2	8
Based upon the feedback, I understood the estimated difficulty level of the items in the ordered item booklet (OIB) and how difficulty relates to our cut scores.	-	-	3	7
Based upon the feedback, I understood the estimated percentage of students that would be classified into each of the four performance categories.	-	-	2	8

Survey #3 Math Grade 3 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Feedback – Comprehension	Strongly Disagree	Disagree	Agree	Strongly Agree
Based upon the feedback, I understood my recommended cut score and how I compared to other panelists.	-	-	-	9
Based upon the feedback, I understood the group's recommended cut score and the variability of the ratings across the panelists.	-	-	-	9
Based upon the feedback, I understood the estimated difficulty level of the items in the ordered item booklet (OIB) and how difficulty relates to our cut scores.	-	-	1	8
Based upon the feedback, I understood the estimated percentage of students that would be classified into each of the four performance categories.	-	-	-	9

COMMENTS

Math Grades 3 and 4

- I appreciate the process as it allows for a deeper understanding of how the state arrives at the cut scores

Survey #3 Math Grade 5 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Feedback – Comprehension	Strongly Disagree	Disagree	Agree	Strongly Agree
Based upon the feedback, I understood my recommended cut score and how I compared to other panelists.	-	-	-	10
Based upon the feedback, I understood the group's recommended cut score and the variability of the ratings across the panelists.	-	-	-	10
Based upon the feedback, I understood the estimated difficulty level of the items in the ordered item booklet (OIB) and how difficulty relates to our cut scores.	-	-	1	9
Based upon the feedback, I understood the estimated percentage of students that would be classified into each of the four performance categories.	-	-	-	10

COMMENTS

Math Grades 5 and 6

- Susan was a wonderful facilitator! Thank you
- As part of the general session - include a short description of the different panelist opportunities (e.g. standards review, range finding, etc.
- Very enjoyable experience; I was very interested in data
- point was brought up that on 6th grade assessment level 1 and level 1 performing students would not be able to achieve any points on constructed response questions based upon 2018 data

Survey #3 Math Grade 7 panel frequency distribution

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

Feedback – Comprehension	Strongly Disagree	Disagree	Agree	Strongly Agree
Based upon the feedback, I understood my recommended cut score and how I compared to other panelists.	-	-	7	2
Based upon the feedback, I understood the group's recommended cut score and the variability of the ratings across the panelists.	-	-	5	4
Based upon the feedback, I understood the estimated difficulty level of the items in the ordered item booklet (OIB) and how difficulty relates to our cut scores.	-	2	6	1
Based upon the feedback, I understood the estimated percentage of students that would be classified into each of the four performance categories.	-	-	5	4

COMMENTS

Math Grades 7 and 8

- while some undisclosed metric created the sequence, it did not seem to reflect our experience with students in the classroom or the P-values. Not having the questions in an appropriate order made this much more difficult
- I would be willing to participate in other conferences if needed
- I would be interested in participation again; appreciated the opportunity
- The grade 7 reference sheet should include simple interest formula
- grade 7 reference sheet should include simple interest rate formula
- I am interested in other review processes
- Thank you for the opportunity. I would like to participate in the future
- I believe the questions should be ordered based on the actual difficulty of the question for the OIB
- All questions requiring a formula (such as simple interest) should provide the formula (not done on the 7th grade test)

Survey #3 **Survey results** **All panels – Mean responses**

Thank you for participating in the New York State standards review workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by indicating if you strongly disagree, disagree, agree, or strongly agree with it. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____ Grade Level: _____

	English Language Arts			Mathematics		
	Panel 3/4	Panel 5/6	Panel 7/8	Panel 3/4	Panel 5/6	Panel 7/8
Based upon the feedback, I understood my recommended cut score and how I compared to other panelists.	3.5	4.0	3.9	4.0	4.0	3.2
Based upon the feedback, I understood the group's recommended cut score and the variability of the ratings across the panelists.	3.5	3.8	3.8	4.0	4.0	3.4
Based upon the feedback, I understood the estimated difficulty level of the items in the ordered item booklet (OIB) and how difficulty relates to our cut scores.	3.5	3.8	3.7	3.9	3.9	2.9
Based upon the feedback, I understood the estimated percentage of students that would be classified into each of the four performance categories.	3.5	3.9	3.8	4.0	4.0	3.4

Survey #4 Vertical Articulation ELA Frequency distribution

Thank you for participating in the New York State standard setting workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by checking the box that corresponds with your feeling on each statement. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____

Vertical articulation	**Strongly disagree**	**Disagree**	**Agree**	**Strongly Agree**
The orientation to the vertical articulation process was comprehensive and allowed me to understand the purpose of the vertical articulation procedure.	-	-	2	7
I understood the data and information provided as part of the feedback across all grade levels.	-	-	2	7
I believe the Vertical Articulation impact data reflects my expectations as far as the percent of students within each performance category.	-	-	3	6
Feedback – Value **How important were the following factors when completing the Vertical Articulation process?**	**Not important**	**Somewhat important**	**Important**	**Very important**
Round 3 impact data	-	-	2	7
Discussion with other panelists	-	-	1	8
Your expectations and experience with New York students	-	-	1	8

COMMENTS:
ELA:

- When we are deliberating over the cut points for constructed response items, it would be so helpful to look at what is allowed in scoring

Survey #4 Vertical Articulation Math frequency distribution

Thank you for participating in the New York State standard setting workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by checking the box that corresponds with your feeling on each statement. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____

Vertical articulation	**Strongly disagree**	**Disagree**	**Agree**	**Strongly Agree**
The orientation to the vertical articulation process was comprehensive and allowed me to understand the purpose of the vertical articulation procedure.	-	-	2	7
I understood the data and information provided as part of the feedback across all grade levels.	-	-	3	6
I believe the Vertical Articulation impact data reflects my expectations as far as the percent of students within each performance category.	-	-	3	6
Feedback – Value **How important were the following factors when completing the Vertical Articulation process?**	**Not important**	**Somewhat important**	**Important**	**Very important**
Round 3 impact data	-	-	3	6
Discussion with other panelists	-	-	2	7
Your expectations and experience with New York students	-	-	3	6

Survey #4 Vertical Articulation Survey results – Mean responses

Thank you for participating in the New York State standard setting workshop. Throughout the workshop, we will be asking you to complete brief surveys so that we can continually evaluate how well panelists understand their tasks and if the training on the tasks has been sufficient. Please respond to each statement below by checking the box that corresponds with your feeling on each statement. Upon completing the survey, please provide the survey to your facilitator.

Content Area: _____

	ELA	Math
Vertical articulation		
The orientation to the vertical articulation process was comprehensive and allowed me to understand the purpose of the vertical articulation procedure.	3.8	3.8
I understood the data and information provided as part of the feedback across all grade levels.	3.8	3.7
I believe the Vertical Articulation impact data reflects my expectations as far as the percent of students within each performance category.	3.7	3.7
Feedback – Value		
How important were the following factors when completing the Vertical Articulation process?		
Round 3 impact data	3.8	3.7
Discussion with other panelists	3.9	3.8
Your expectations and experience with New York students	3.9	3.7