

# **New York State Testing Program 2016: English Language Arts and Mathematics Grades 3–8**



**Technical Report**

**Questar Assessment, Inc.  
2016**

Developed and published under contract with the New York State Education Department by Questar Assessment, Inc., 5550 Upper 147<sup>th</sup> Street West, Apple Valley, MN 55124. Copyright © 2016 by the New York State Education Department.

Permission is hereby granted for New York State school administrators and educators to reproduce these materials, located online at <http://www.p12.nysed.gov/assessment/reports/>, in the quantities necessary for educational researchers' use, but not for sale, provided copyright notices are retained as they appear in these publications.

## 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.....	5
2.1. Test Descriptions .....	5
2.1.1. ELA Tests.....	5
2.1.2. Mathematics Tests.....	5
2.2. Test Configuration .....	6
2.2.1. Test Book Design .....	6
2.2.2. Embedded Field-Test Items.....	6
2.3. New York State Educators’ Involvement in Test Development.....	7
2.4. Test Blueprints .....	7
2.5. Passage Selection and Item Criteria Documents .....	8
2.5.1. Principles of Universal Design.....	10
2.6. Passage Finding .....	10
2.7. Item Development.....	11
2.8. Educator Item Review.....	11
2.9. Field-Testing.....	12
2.10. Rangefinding.....	13
2.11. Item Selection and Test Creation (Criteria and Process).....	13
2.12. Educator Form Construction.....	14
2.13. Test Form Production .....	15
2.14. Final Eyes Committees .....	15
2.15. Proficiency and Performance Standards .....	15
Section 3: Validity .....	17
3.1. Content Validity.....	17
3.2. Construct (Internal Structure) Validity .....	18
3.2.1. Internal Consistency .....	18
3.2.2. Unidimensionality .....	18
3.2.3. Detection of Bias.....	22
Section 4: Test Administration and Scoring.....	23

4.1. Test Administration .....	23
4.2. Scoring Procedures of Operational Tests.....	23
4.3. Scoring Models .....	23
4.4. Scoring of Constructed-Response Items.....	24
4.5. Scorer Qualifications and Training.....	24
4.6. Quality Control Process .....	25
Section 5: Operational Test Data Collection and Classical Analysis .....	26
5.1. Data Collection .....	26
5.2. Data Processing.....	26
5.2.1. Sampling Down for Representativeness.....	26
5.3. Classical Analysis and Calibration Sample Characteristics.....	31
5.4. Classical Data Analysis.....	40
5.4.1. Item Difficulty and Point Biserial Correlation Coefficients .....	40
5.4.2. Omit Rates.....	42
5.4.3. Differential Item Functioning (DIF).....	42
Section 6: IRT Calibration and Linking.....	45
6.1. IRT Models and Rationale for Use .....	45
6.2. Calibration Sample.....	46
6.2.1. Calibration Process.....	52
6.3. Item-Model Fit.....	53
6.4. Local Independence .....	55
6.5. Linking and Scaling.....	56
6.6. Anchor Set Evaluation .....	58
6.7. Test Characteristic Curves .....	60
6.8. Scoring Procedure.....	73
6.8.1. Raw Score-to-Scale Score and SEM Conversion Tables.....	73
Section 7: Reliability and Standard Error of Measurement .....	76
7.1. Test Reliability.....	76
7.1.1. Test Statistics and Reliability for Total Test .....	76
7.1.2. Reliability of MC Items.....	78
7.1.3. Reliability of CR Items.....	79
7.1.4. Test Reliability for Reporting Categories .....	79
7.2. Standard Error of Measurement (SEM).....	91
7.3. Performance Level Classification Consistency and Accuracy .....	92
7.3.1. Consistency .....	92
7.3.2. Accuracy.....	94
Section 8: Summary of Operational Test Results.....	96
8.1. Scale Score Distribution Summary.....	96
8.1.1. ELA Scale Score and Subscore Distributions .....	96
8.1.1.1. ELA Grade 3 .....	97
8.1.1.2. ELA Grade 4 .....	98
8.1.1.3. ELA Grade 5 .....	99

8.1.1.4. ELA Grade 6 .....	100
8.1.1.5. ELA Grade 7 .....	101
8.1.1.6. ELA Grade 8 .....	102
8.1.2. Mathematics Scale Score Distributions .....	103
8.1.2.1. Mathematics Grade 3.....	104
8.1.2.2. Mathematics Grade 4.....	106
8.1.2.3. Mathematics Grade 5.....	107
8.1.2.4. Mathematics Grade 6.....	108
8.1.2.5. Mathematics Grade 7.....	109
8.1.2.6. Mathematics Grade 8.....	111
8.2. Performance Level Distribution Summary .....	112
8.2.1. ELA Test Performance Level Distributions .....	112
8.2.1.1. ELA Grade 3 .....	113
8.2.1.2. ELA Grade 4 .....	114
8.2.1.3. ELA Grade 5 .....	115
8.2.1.4. ELA Grade 6 .....	116
8.2.1.5. ELA Grade 7 .....	117
8.2.1.6. ELA Grade 8 .....	118
8.2.2. Mathematics Test Performance Level Distributions .....	119
8.2.2.1. Mathematics Grade 3.....	120
8.2.2.2. Mathematics Grade 4.....	121
8.2.2.3. Mathematics Grade 5.....	122
8.2.2.4. Mathematics Grade 6.....	124
8.2.2.5. Mathematics Grade 7.....	125
8.2.2.6. Mathematics Grade 8.....	126
Section 9: References.....	128
Appendix A: ELA and Mathematics Test Configurations.....	131
Appendix B: ELA and Mathematics Test Blueprints .....	135
Appendix C: Passage Selection Guidelines for Assessing ELA.....	137
Appendix D: Universal Design Item Checklist .....	138
Appendix E: Criteria for Item Acceptability .....	141
Appendix F: Psychometric Guidelines for Operational Item Selection.....	143
Appendix G: Operational Item Maps.....	144
Appendix H: ELA Short-Response Rubric.....	160
Appendix I: ELA Extended-Response Rubric .....	161
Appendix J: Mathematics Short-Response Rubric .....	164
Appendix K: Mathematics Extended-Response Rubric .....	165
Appendix L: Factor Analysis Results for Select Subgroups.....	166
Appendix M: Classical Test Theory Statistics.....	183
Appendix N: Items Flagged for DIF .....	197
Appendix O: IRT Statistics.....	201
Appendix P: Derivation and Estimation of Classification Consistency and Accuracy .....	228
Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables.....	231

## 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 2015 Field Test Items .....	12
Table 3.1. ELA Tests Factor Analysis .....	20
Table 3.2. Mathematics Tests Factor Analysis .....	21
Table 5.1. ELA Grade 3 Data Cleaning .....	27
Table 5.2. ELA Grade 4 Data Cleaning .....	27
Table 5.3. ELA Grade 5 Data Cleaning .....	28
Table 5.4. ELA Grade 6 Data Cleaning .....	28
Table 5.5. ELA Grade 7 Data Cleaning .....	28
Table 5.6. ELA Grade 8 Data Cleaning .....	29
Table 5.7. Mathematics Grade 3 Data Cleaning .....	29
Table 5.8. Mathematics Grade 4 Data Cleaning .....	29
Table 5.9. Mathematics Grade 5 Data Cleaning .....	30
Table 5.10. Mathematics Grade 6 Data Cleaning .....	30
Table 5.11. Mathematics Grade 7 Data Cleaning .....	30
Table 5.12. Mathematics Grade 8 Data Cleaning .....	31
Table 5.13. ELA Grade 3 Sample Characteristics .....	31
Table 5.14. ELA Grade 4 Sample Characteristics .....	32
Table 5.15. ELA Grade 5 Sample Characteristics .....	33
Table 5.16. ELA Grade 6 Sample Characteristics .....	33
Table 5.17. ELA Grade 7 Sample Characteristics .....	34
Table 5.18. ELA Grade 8 Sample Characteristics .....	35
Table 5.19. Mathematics Grade 3 Sample Characteristics .....	36
Table 5.20. Mathematics Grade 4 Sample Characteristics .....	36
Table 5.21. Mathematics Grade 5 Sample Characteristics .....	37
Table 5.22. Mathematics Grade 6 Sample Characteristics .....	38
Table 5.23. Mathematics Grade 7 Sample Characteristics .....	39
Table 5.24. Mathematics Grade 8 Sample Characteristics .....	39
Table 5.25. ELA Classical DIF Sample N-Counts .....	43
Table 5.26. Mathematics Classical DIF Sample N-Counts .....	43
Table 5.27. ELA Items Flagged for DIF .....	43
Table 6.1. ELA Grades 3 and 4 Demographic Statistics .....	47
Table 6.2. ELA Grades 5 and 6 Demographic Statistics .....	48
Table 6.3. ELA Grades 7 and 8 Demographic Statistics .....	49
Table 6.4. Mathematics Grades 3 and 4 Demographic Statistics .....	50
Table 6.5. Mathematics Grades 5 and 6 Demographic Statistics .....	51
Table 6.6. Mathematics Grades 7 and 8 Demographic Statistics .....	52
Table 6.7. ELA Calibration Results .....	53

Table 6.8. Mathematics Calibration Results .....	53
Table 6.9. ELA Linking Coefficients .....	56
Table 6.10. Mathematics Linking Coefficients.....	57
Table 6.11. ELA Scaling Coefficients .....	58
Table 6.12. Mathematics Scaling Coefficients .....	58
Table 6.13. ELA Performance-Level Cut Scores .....	74
Table 6.14. Mathematics Performance-Level Cut Scores .....	75
Table 7.1. ELA Test Form Statistics.....	77
Table 7.2. ELA Test Reliability and Standard Error of Measurement .....	77
Table 7.3. Mathematics Test Form Statistics.....	77
Table 7.4. Mathematics Test Reliability and Standard Error of Measurement .....	78
Table 7.5. ELA MC Item Reliability and Standard Error of Measurement.....	78
Table 7.6. Mathematics MC Item Reliability and Standard Error of Measurement.....	78
Table 7.7. ELA CR Item Reliability and Standard Error of Measurement.....	79
Table 7.8. Mathematics CR Item Reliability and Standard Error of Measurement.....	79
Table 7.9. ELA Grade 3 Test Reliability by Subgroup.....	80
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.....	84
Table 7.14. ELA Grade 8 Test Reliability by Subgroup.....	85
Table 7.15. Mathematics Grade 3 Test Reliability by Subgroup.....	86
Table 7.16. Mathematics Grade 4 Test Reliability by Subgroup.....	87
Table 7.17. Mathematics Grade 5 Test Reliability by Subgroup.....	88
Table 7.18. Mathematics Grade 6 Test Reliability by Subgroup.....	89
Table 7.19. Mathematics Grade 7 Test Reliability by Subgroup.....	90
Table 7.20. Mathematics Grade 8 Test Reliability by Subgroup.....	91
Table 7.21. Decision Consistency (All Cuts).....	93
Table 7.22. Decision Consistency (Level III Cut) .....	94
Table 7.23. Decision Agreement (Accuracy) Estimates .....	95
Table 8.1. ELA Scale Score Distribution Summary .....	96
Table 8.2. ELA Subscore Summary .....	97
Table 8.3. ELA Grade 3 Scale Score Distribution by Subgroup .....	98
Table 8.4. ELA Grade 4 Scale Score Distribution by Subgroup .....	99
Table 8.5. ELA Grade 5 Scale Score Distribution by Subgroup .....	100
Table 8.6. ELA Grade 6 Scale Score Distribution by Subgroup .....	101
Table 8.7. ELA Grade 7 Scale Score Distribution by Subgroup .....	102
Table 8.8. ELA Grade 8 Scale Score Distribution by Subgroup .....	103
Table 8.9. Mathematics Scale Score Distribution Summary .....	104
Table 8.10. Mathematics Subscore Summary.....	104
Table 8.11. Mathematics Grade 3 Scale Score Distribution by Subgroup .....	105

Table 8.12. Mathematics Grade 4 Scale Score Distribution by Subgroup .....	106
Table 8.13. Mathematics Grade 5 Scale Score Distribution by Subgroup .....	107
Table 8.14. Mathematics Grade 6 Scale Score Distribution by Subgroup .....	109
Table 8.15. Mathematics Grade 7 Scale Score Distribution by Subgroup .....	110
Table 8.16. Mathematics Grade 8 Scale Score Distribution by Subgroup .....	111
Table 8.17. ELA Test Performance Level Distributions .....	113
Table 8.18. ELA Grade 3 Performance Level Distribution by Subgroup .....	113
Table 8.19. ELA Grade 4 Performance Level Distribution by Subgroup .....	114
Table 8.20. ELA Grade 5 Performance Level Distribution by Subgroup .....	115
Table 8.21. ELA Grade 6 Performance Level Distribution by Subgroup .....	116
Table 8.22. ELA Grade 7 Performance Level Distribution by Subgroup .....	118
Table 8.23. ELA Grade 8 Performance Level Distribution by Subgroup .....	119
Table 8.24. Mathematics Test Performance Level Distributions .....	120
Table 8.25. Mathematics Grade 3 Performance Level Distribution by Subgroup.....	120
Table 8.26. Mathematics Grade 4 Performance Level Distribution by Subgroup.....	122
Table 8.27. Mathematics Grade 5 Performance Level Distribution by Subgroup.....	123
Table 8.28. Mathematics Grade 6 Performance Level Distribution by Subgroup.....	124
Table 8.29. Mathematics Grade 7 Performance Level Distribution by Subgroup.....	125
Table 8.30. Mathematics Grade 8 Performance Level Distribution by Subgroup.....	127
Table A1. ELA Test Configuration .....	131
Table A2. Mathematics Test Configuration.....	132
Table A3. ELA Estimated Time on Task by Book.....	133
Table A4. Mathematics Estimated Time on Task by Book .....	133
Table B1. ELA Test Blueprint .....	135
Table B2. Mathematics Test Blueprint .....	136
Table G1. ELA Grade 3 Operational Item Map .....	144
Table G2. ELA Grade 4 Operational Item Map .....	145
Table G3. ELA Grade 5 Operational Item Map .....	146
Table G4. ELA Grade 6 Operational Item Map .....	147
Table G5. ELA Grade 7 Operational Item Map .....	148
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.....	155
Table G11. Mathematics Grade 7 Operational Item Map.....	156
Table G12. Mathematics Grade 8 Operational Item Map.....	158
Table L1. ELA Grade 3 Test Factor Analysis by Subgroup.....	166
Table L2. ELA Grade 4 Test Factor Analysis by Subgroup.....	167
Table L3. ELA Grade 5 Test Factor Analysis by Subgroup.....	168
Table L4. ELA Grade 6 Test Factor Analysis by Subgroup.....	170

Table L5. ELA Grade 7 Test Factor Analysis by Subgroup.....	172
Table L6. ELA Grade 8 Test Factor Analysis by Subgroup.....	173
Table L7. Mathematics Grade 3 Test Factor Analysis by Subgroup.....	174
Table L8. Mathematics Grade 4 Test Factor Analysis by Subgroup.....	175
Table L9. Mathematics Grade 5 Test Factor Analysis by Subgroup.....	176
Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup.....	177
Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup.....	179
Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup.....	180
Table M1. ELA Grade 3 Classical Item Analysis.....	183
Table M2. ELA Grade 4 Classical Item Analysis.....	184
Table M3. ELA Grade 5 Classical Item Analysis.....	184
Table M4. ELA Grade 6 Classical Item Analysis.....	186
Table M5. ELA Grade 7 Classical Item Analysis.....	187
Table M6. ELA Grade 8 Classical Item Analysis.....	188
Table M7. Mathematics Grade 3 Classical Item Analysis.....	189
Table M8. Mathematics Grade 4 Classical Item Analysis.....	190
Table M9. Mathematics Grade 5 Classical Item Analysis.....	191
Table M10. Mathematics Grade 6 Classical Item Analysis.....	192
Table M11. Mathematics Grade 7 Classical Item Analysis.....	193
Table M12. Mathematics Grade 8 Classical Item Analysis.....	194
Table N1. ELA MC Item Classical DIF Flags.....	197
Table N2. ELA CR Item Classical DIF Flags.....	199
Table N3. Mathematics MC Item Classical DIF Flags.....	199
Table N4. Mathematics CR Item Classical DIF Flags.....	200
Table O1. ELA Grade 3 Item Fit Statistics.....	201
Table O2. ELA Grade 4 Item Fit Statistics.....	202
Table O3. ELA Grade 5 Item Fit Statistics.....	202
Table O4. ELA Grade 6 Item Fit Statistics.....	204
Table O5. ELA Grade 7 Item Fit Statistics.....	205
Table O6. ELA Grade 8 Item Fit Statistics.....	206
Table O7. Mathematics Grade 3 Item Fit Statistics.....	207
Table O8. Mathematics Grade 4 Item Fit Statistics.....	208
Table O9. Mathematics Grade 5 Item Fit Statistics.....	209
Table O10. Mathematics Grade 6 Item Fit Statistics.....	210
Table O11. Mathematics Grade 7 Item Fit Statistics.....	211
Table O12. Mathematics Grade 8 Item Fit Statistics.....	213
Table O13. ELA Grade 3 OP Item Parameter Estimates.....	214
Table O14. ELA Grade 4 OP Item Parameter Estimates.....	215
Table O15. ELA Grade 5 OP Item Parameter Estimates.....	216
Table O16. ELA Grade 6 OP Item Parameter Estimates.....	217
Table O17. ELA Grade 7 OP Item Parameter Estimates.....	218

Table O18. ELA Grade 8 OP Item Parameter Estimates.....	219
Table O19. Mathematics Grade 3 OP Item Parameter Estimates.....	220
Table O20. Mathematics Grade 4 OP Item Parameter Estimates.....	221
Table O21. Mathematics Grade 5 OP Item Parameter Estimates.....	222
Table O22. Mathematics Grade 6 OP Item Parameter Estimates.....	223
Table O23. Mathematics Grade 7 OP Item Parameter Estimates.....	225
Table O24. Mathematics Grade 8 OP Item Parameter Estimates.....	226
Table Q1. ELA Grade 3 RSSS Table.....	231
Table Q2. ELA Grade 4 RSSS Table.....	231
Table Q3. ELA Grade 5 RSSS Table.....	232
Table Q4. ELA Grade 6 RSSS Table.....	233
Table Q6. ELA Grade 8 RSSS Table.....	234
Table Q7. Mathematics Grade 3 RSSS Table.....	235
Table Q8. Mathematics Grade 4 RSSS Table.....	236
Table Q9. Mathematics Grade 5 RSSS Table.....	237
Table Q10. Mathematics Grade 6 RSSS Table.....	238
Table Q11. Mathematics Grade 7 RSSS Table.....	239
Table Q12. Mathematics Grade 8 RSSS Table.....	240
Table Q13. ELA Grade 3 Scale Score Frequency Distribution.....	241
Table Q14. ELA Grade 4 Scale Score Frequency Distribution.....	242
Table Q15. ELA Grade 5 Scale Score Frequency Distribution.....	243
Table Q16. ELA Grade 6 Scale Score Frequency Distribution.....	245
Table Q17. ELA Grade 7 Scale Score Frequency Distribution.....	246
Table Q18. ELA Grade 8 Scale Score Frequency Distribution.....	248
Table Q19. Mathematics Grade 3 Scale Score Frequency Distribution.....	249
Table Q20. Mathematics Grade 4 Scale Score Frequency Distribution.....	251
Table Q21. Mathematics Grade 5 Scale Score Frequency Distribution.....	252
Table Q22. Mathematics Grade 6 Scale Score Frequency Distribution.....	254
Table Q23. Mathematics Grade 7 Scale Score Frequency Distribution.....	255
Table Q24. Mathematics Grade 8 Scale Score Frequency Distribution.....	257

## List of Figures

Figure 6.1. ELA Grade 3 TCC.....	61
Figure 6.2. ELA Grade 3 CSEM Curve.....	61
Figure 6.3. ELA Grade 4 TCC.....	62
Figure 6.4. ELA Grade 4 CSEM Curve.....	62
Figure 6.5. ELA Grade 5 TCC.....	63
Figure 6.6. ELA Grade 5 CSEM Curve.....	63
Figure 6.7. ELA Grade 6 TCC.....	64
Figure 6.8. ELA Grade 6 CSEM Curve.....	64
Figure 6.9. ELA Grade 7 TCC.....	65
Figure 6.10. ELA Grade 7 CSEM Curve.....	65
Figure 6.11. ELA Grade 8 TCC.....	66
Figure 6.12. ELA Grade 8 CSEM Curve.....	66
Figure 6.13. Mathematics Grade 3 TCC.....	67
Figure 6.14. Mathematics Grade 3 CSEM Curve.....	67
Figure 6.15. Mathematics Grade 4 TCC.....	68
Figure 6.16. Mathematics Grade 4 CSEM Curve.....	68
Figure 6.17. Mathematics Grade 5 TCC.....	69
Figure 6.18. Mathematics Grade 5 CSEM Curve.....	69
Figure 6.19. Mathematics Grade 6 TCC.....	70
Figure 6.20. Mathematics Grade 6 CSEM Curve.....	70
Figure 6.21. Mathematics Grade 7 TCC.....	71
Figure 6.22. Mathematics Grade 7 CSEM Curve.....	71
Figure 6.23. Mathematics Grade 8 TCC.....	72
Figure 6.24. Mathematics Grade 8 CSEM Curve.....	72

## **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 Common Core English Language Arts (ELA) and Mathematics 2016 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, linking, scaling, and student performance.

### **1.2. Test Purpose**

The 2016 Grades 3–8 Common Core ELA and Mathematics NYSTP has been designed to measure student knowledge and skills as defined by grade-level New York State Common Core Learning Standards (CCLS) 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 CCLS. 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 in the Grades 3–8 NYSTP. Non-public schools may participate in the testing program, but their participation is not mandatory. In 2016, some non-public 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 Common Core English Language Arts and Mathematics Tests School Administrator’s Manual (SAM)*, available online at <http://www.p12.nysed.gov/assessment/sam/ei/eisam16.pdf>.

### **1.4. Test Use and Decisions Based on Assessment**

The NYSTP Grades 3–8 Common Core ELA and Mathematics Tests are used to measure the extent to which individual students achieve the New York State CCLS in ELA and Mathematics, respectively, in order to determine whether or not 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 Common Core 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 and Linking. 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 were 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 were estimated based on the 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 Common Core 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 Common Core 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 Common Core 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 Common Core 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 Common Core 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 established during the process of standard setting in Summer 2013. The process is described in detail in Section 8 and Appendix P in the 2013 technical report (NYSED, 2013).

#### 1.4.4. *Subscores*

The Grades 3–8 Common Core 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 Common Core Mathematics tests have three subscores that are the domain-level scores for items measuring the *Major Clusters* in each grade. The CCLS 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 2016 tests. In 2016, 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**

Grade	Total Subscore Points	
	Reading	Writing to Sources
3	25	22
4	25	22
5	35	22
6	35	22
7	35	22
8	35	22

**Table 1.2. Mathematics Subscore Categories and Total Possible Score Points**

Grade	Reporting Subscores and Total Subscore Points		
	Subscore 1	Subscore 2	Subscore 3
3	Operations and Algebraic Thinking 25	Number and Operations—Fractions 11	Measurement and Data 11
4	Operations and Algebraic Thinking 11	Numbers and Operations in Base 10 16	Number and Operations—Fractions 17
5	Numbers and Operations in Base 10 16	Number and Operations—Fractions 23	Measurement and Data 7
6	Ratios and Proportional Relationships 17	The Number System 13	Expressions and Equations 23
7	Ratios and Proportional Relationships 20	The Number System 12	Expressions and Equations 21
8	Expressions and Equations 28	Functions 11	Geometry 12

### **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 [2016 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 Midland Information Resources, and the Braille editions are produced by gh, LLC. gh 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 Common Core 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 learners (ELLs) taking the Grades 3–8 Common Core 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 Common Core 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 ELLs taking the ELA Tests: separate testing location and bilingual glossaries.

## Section 2: Test Design and Development

---

### 2.1. Test Descriptions

The 2016 Grades 3–8 Common Core 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 CCLS. The tests were administered in New York State classrooms during a three-day period in April 2016. 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 Common Core English Language Arts and Mathematics Tests School Administrator’s Manual* (SAM), available at: <http://www.p12.nysed.gov/assessment/sam/ei/eisam16.pdf>.

#### 2.1.1. ELA Tests

The 2016 Grade 3–8 Common Core ELA Tests were designed to measure student literacy as defined by the CCLS. 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 Common Core 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 Common Core 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 or paired texts. 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 L provides the rubric for the extended-response items.

#### 2.1.2. Mathematics Tests

The 2016 Grade 3–8 Common Core Mathematics Tests were designed to measure student mathematic understanding as defined by the CCLS. 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 CCLS, 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, the application of multiple mathematics skills, and 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 Book Design*

The 2016 Grades 3–8 Common Core ELA Tests were composed of three books per grade and administered in three sessions over three days. Each day consisted of one book; Book 1 and Book 2 contained literary and informational reading passages and MC items based on the passages. Book 2 also contained reading passages with short-response items and an extended-response item based on those passages. Book 3 contained only reading passages with short-response items and an extended-response item based on those passages.

The 2016 Grades 3–8 Common Core Mathematics Tests were composed of three books per grade and administered in three sessions over three days. Each day consisted of one book: Book 1 and Book 2 contained MC items. Book 3 contained short- and extended-response items. The tables in Appendix A provide information on the numbers and types of items in each book for the Grades 3–8 Common Core 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 2016 administrations of the Common Core 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 2016, 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 Common Core ELA and Mathematics test development. New York State educators provide critical input throughout all stages of the test development process, which include standard setting, rangefinding, educator item review, operational forms construction, and “Final Eyes” meeting (a final review of the test books 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 Five Cities (i.e., Buffalo, New York City, 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 continuously 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-20 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 books), the representation and distribution of content were determined.

The CCLS for ELA are organized into four strands: Reading, Writing, Language, and Speaking/Listening. Due to administration constraints, Speaking/Listening was determined to best be assessed in the classroom, only; therefore, the Common Core 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 CCLS 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 (i.e., 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 Common Core 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 2016 operational tests.

## **2.5. Passage Selection and Item Criteria Documents**

The 2016 administration was the first year in which Questar delivered the New York State tests<sup>1</sup>. To guide test item development and to help ensure that New York State tests were measuring the CCLS 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 Common Core State Standards (CCSS) 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](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 or not it could be used to measure the CCSS 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:

[http://www.engageny.org/sites/default/files/resource/attachments/new\\_york\\_state\\_passage\\_review\\_criteria\\_protocol\\_document.doc](http://www.engageny.org/sites/default/files/resource/attachments/new_york_state_passage_review_criteria_protocol_document.doc).

---

<sup>1</sup> The items and passages selected for the operational test and field tested as embedded items were developed by the previous test delivery vendor. In general, the previous vendor completed the portion of the work prior to the construction of operational forms, while Questar worked with NYSED and educators to build the forms and performed all subsequent operational work.

*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 Common Core 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 or not 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 CCSS, 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 Common Core Mathematics Tests* document was developed to ensure that the tests measured the deep conceptual understanding that CCSS demand, rather than focusing on predictable Mathematics items that require only algorithmic strategies to be solved correctly. *Multiple Representations* are a broad set of specifications that describe, refer to, and symbolize 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 CCSS-aligned test items. To do so, in the 2013–2014 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, and it was 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 the reason for rejection.

In addition to the content reviews performed by NYSED staff and its vendors, the passages were also reviewed by executives in both organizations. 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 2016 test forms was conducted during the 2013–2014 development cycle. The goal of item development is to develop a sufficient number of high-quality, CCSS-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 replaced by content specialists. 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 the reason for rejection, 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, 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 2015 Field Test Items**

Grade	Unique ELA Items by Type*		Unique Mathematics Items by Type*	
	MC	CR	MC	CR
3	126	48	96	22
4	125	48	120	25
5	138	48	120	25
6	137	48	125	25
7	138	48	123	25
8	138	48	121	25

\* MC = multiple-choice. CR = constructed-response. All CR items were field-tested under stand-alone conditions, while MC items were administered under both embedded and stand-alone conditions.

Field test items were administered in Spring 2015 as embedded field test items within the 2015 operational test forms. The use of embedded field test items yields more reliable field-test data and reduces, but does not eliminate, the need for multiple-choice stand-alone field testing. One additional round of field testing was administered separately from the 2015 operational forms (i.e., as stand-alone tests) later in Spring 2015.

In order to better understand how the 2015 field test items may perform on future operational forms, a variety of analyses were conducted. 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, linking, 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**

Rangefinding for most items included on the 2016 test was conducted by Questar. 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. This process was overseen by NYSED content experts and Questar Scoring Directors. 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 2015 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 preapproved 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 Common Core ELA and Mathematics Tests were administered in April 2016. 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 2015.

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 with 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. Those items were then reviewed by the content specialists and their leads. 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 October 2015 to finalize item selection and test creation with NYSED staff (including content and assessment experts) and educators.

## **2.12. Educator Form Construction**

During an educator form construction meeting that took place from October 26 – November 2, 2015, in Albany, New York, educators from around the State worked with NYSED and Questar to review the content of the proposed 2016 operational ELA passages, and ELA and Mathematics individual test items, and how those items combine to entire operational forms, 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 a binder with item cards 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 for 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 for 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 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 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 nine 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. Appendix R contains the full Final Eyes meeting report.

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 April 2016 administration.

### **2.15. Proficiency and Performance Standards**

In Summer 2013, after the operational administration of the 2013 tests, a standard setting meeting occurred in Albany where 95 New York State educators went through a rigorous process, guided by the best practices indicated by this intensely studied process, to recommend performance standards for the new tests measuring the CCLS. These recommendations were presented to the Commissioner and the Board of Regents, who, in turn, adopted the

recommended standards set forth by the committees. For additional details, see Section 8 and Appendix P in the 2013 technical report (NYSED, 2013).

Each grade level has four performance levels. Three cut points demarcate the performance levels needed to demonstrate each ascending level of performance. Section 6.8.1. Raw Score-to-Scale Score and SEM Conversion Tables contains detailed information related to 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 Common Core Mathematics Learning Standards and in the area of ELA as defined by the New York State Common Core 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 CCLS. 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 .89 to .92. For all subgroups, the reliability coefficients were greater than or equal to .81. For the total population, the Mathematics reliability coefficients (Cronbach’s alpha) ranged from .93 to .95. For all subgroups, the reliability coefficients were greater than or equal to .80. Overall, high internal consistency of the NYSTP Grades 3–8 Common Core 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  $Q_I$  statistics (Yen, 1981), and the results are described in detail in Section 6: IRT Calibration and Linking. Most items demonstrated sound fit across grades and content areas, and only a few items were deemed to have deviate 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 non-public school students for whom data were available during the linking process. 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 Mathematics 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 further investigate the factor analysis results 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.”

Factor analyses related to the Grades 3–8 Common Core 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 Common Core ELA and Mathematics Tests exhibited first principal component accounting for more than 19% and 31% 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	<b>8.56</b>	<b>25.19</b>	<b>25.19</b>
	2	1.46	4.30	29.49
	3	1.26	3.72	33.21
4	1	<b>7.38</b>	<b>21.70</b>	<b>21.70</b>
	2	1.43	4.22	25.92
	3	1.03	3.04	28.95
5	1	<b>9.14</b>	<b>20.76</b>	<b>20.76</b>
	2	1.63	3.70	24.46
	3	1.29	2.94	27.41
	4	1.02	2.32	29.72
6	1	<b>8.33</b>	<b>18.93</b>	<b>18.93</b>
	2	1.61	3.67	22.60
	3	1.14	2.59	25.19
	4	1.09	2.47	27.66
	5	1.03	2.35	30.01
7	1	<b>9.32</b>	<b>21.18</b>	<b>21.18</b>
	2	1.59	3.61	24.79
	3	1.10	2.51	27.29
	4	1.04	2.35	29.65
8	1	<b>10.41</b>	<b>23.66</b>	<b>23.66</b>
	2	1.68	3.81	27.47
	3	1.31	2.97	30.44
	4	1.00	2.28	32.72

**Table 3.2. Mathematics Tests Factor Analysis**

Grade	Extracted Factor			
	#	Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
3	<b>1</b>	<b>11.42</b>	<b>25.39</b>	<b>25.39</b>
	2	1.58	3.51	28.90
	3	1.13	2.51	31.41
	4	1.10	2.45	33.86
4	<b>1</b>	<b>14.66</b>	<b>30.54</b>	<b>30.54</b>
	2	1.33	2.76	33.30
	3	1.22	2.54	35.84
	4	1.13	2.36	38.20
5	<b>1</b>	<b>12.70</b>	<b>27.02</b>	<b>27.02</b>
	2	1.84	3.92	30.95
	3	1.05	2.24	33.19
	4	1.02	2.16	35.35
	5	1.00	2.13	37.48
6	<b>1</b>	<b>12.79</b>	<b>24.13</b>	<b>24.13</b>
	2	1.74	3.28	27.41
	3	1.10	2.08	29.49
7	<b>1</b>	<b>14.34</b>	<b>26.56</b>	<b>26.56</b>
	2	1.53	2.83	29.39
	3	1.17	2.17	31.56
8	<b>1</b>	<b>12.16</b>	<b>22.52</b>	<b>22.52</b>
	2	1.49	2.77	25.29
	3	1.30	2.40	27.69
	4	1.00	1.86	29.55

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 (ELLs), students with disabilities (SWD), and students using test accommodations (SUA). 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, SWD, SUA, ELL/SUA, and SWD/SUA classifications. The ELL/SUA subgroup is defined as examinees who are ELLs and who use at least one ELL-related accommodation. The SWD/SUA subgroup includes examinees who are classified as having disabilities and who use at least one disability-related accommodation.

### 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 and 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). In addition, very few items were flagged by multiple methods. Items flagged for statistically significant DIF were carefully reviewed by multiple reviewers 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 (2016)* located here:

<http://www.p12.nysed.gov/assessment/sam/ei/scoringleaderhb16rev2.pdf>.

### **4.1. Test Administration**

The NYSTP Grades 3–8 Common Core ELA and Mathematics Tests were administered to students during April 2016. The testing window was Monday, April 4 – Thursday, April 7 for the Grades 3–8 Common Core ELA Tests and Wednesday, April 13 – Friday, April 15 for the Grades 3–8 Common Core Mathematics Tests. The makeup test administration window was Friday, April 8 – Tuesday, April 12 for the Grades 3–8 Common Core ELA Tests and Monday, April 18 – Wednesday, April 20 for the Grades 3–8 Common Core 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.

### **4.2. Scoring Procedures of Operational Tests**

The scoring of the NYSTP 2016 Grades 3–8 Common Core ELA and Mathematics Tests was performed at designated sites by qualified teachers and administrators. 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 books, 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. The trainers were assisted by facilitators or leaders, who also helped in monitoring the sessions and enforced scoring accuracy.

The titles for administrators, trainers, and facilitators vary by the scoring model that is selected. At the regional level, oversight was conducted by a site coordinator. 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.

### **4.3. Scoring Models**

For the 2015–2016 school year, schools and school districts were able to score Grades 3–8 Common Core 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 non-public schools in an affiliation group (non-public 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, non-public 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 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.
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 non-public school.
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 for both the training of scoring committee members and the scoring of CR items was the scoring guides. These documents 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. Additionally, scoring leader handbooks were also distributed to outline the responsibilities of the scoring roles.

Upon completion of the training of scoring committee members, scoring was conducted with pen-and-pencil scoring as opposed to electronic scoring, and each scoring-committee member evaluated actual student papers instead of electronically scanned papers. All scoring-committee members were trained by previously trained and approved trainers along with guidance from scoring guides. Each constructed-response test book was scored by three separate scoring committee members, who scored three distinct sections of the test book. After test books were completed, the table facilitator or subject (ELA or mathematics) leader conducted a “read behind” of approximately 12 sets of test books per hour to verify the accuracy of scoring. If an item arose that was not covered in the training materials, facilitators or trainers were to call the Questar Scoring Helpline for assistance with the ELA or mathematics scoring (see Section 4.6. Quality Control Process).

#### **4.5. Scorer Qualifications and Training**

The scoring of the 2016 Grades 3–8 Common Core ELA and Mathematics Tests was conducted by qualified administrators and teachers. 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 was 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/scoringleaderhb16rev2.pdf>.

#### **4.6. Quality Control Process**

Test books were randomly distributed throughout each scoring room so that books from each region, district, school, or class were evenly dispersed. Teams were divided into groups of three to ensure that a variety of scorers graded each book. If a scorer and a facilitator could not reach a decision on a paper after reviewing the scoring guides and audio files, they called the Questar Scoring Helpline. The call center was established to help teachers and administrators during scoring. The help-line staff consisted of trained Questar personnel, who answered items by phone or fax. 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 on each completed box of scored tests 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 2016. During Phase 2, “straggler files” were submitted to Questar in June 2016.

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 non-public 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, however, 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 book. 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 three test books 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 (i.e., all students testing in Spring 2015) in terms of key variables such as student gender, racial/ethnic identity, student disability status, English Language Learner (ELL) 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 2016 relative to 2015, but they were not used directly in the sampling down analyses. Comparison results between the

final 2016 sample and 2015 operational population are further described in Section 6, “IRT Calibration and Linking.” In Spring 2016, a sampling down approach was adopted to make the sample used for linking as similar as possible to the previous year’s testing population.

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 2015 and 2016 operational forms. There were some minor differences in the 2015 and 2016 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 – 12, respectively.

**Table 5.1. ELA Grade 3 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	175,071
Wrong Subject	0	175,071
No Grade	1	175,070
Wrong Grade	23	175,047
Language Mismatched Form	135	174,912
School Type	34	174,878
Missing Entire Book	1,169	173,709
Invalid Score	0	173,709
Out-of-Range CR Scores	0	173,709
Duplicated Record	14	173,695

**Table 5.2. ELA Grade 4 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	172,224
Wrong Subject	0	172,224
No Grade	2	172,222
Wrong Grade	13	172,209
Language Mismatched Form	132	172,077
School Type	0	172,077
Missing Entire Book	886	171,191
Invalid Score	0	171,191
Out-of-Range CR Scores	0	171,191
Duplicated Record	6	171,185

**Table 5.3. ELA Grade 5 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	162,075
Wrong Subject	0	162,075
No Grade	0	162,075
Wrong Grade	21	162,054
Language Mismatched Form	176	161,878
School Type	136	161,742
Missing Entire Book	920	160,822
Invalid Score	0	160,822
Out-of-Range CR Scores	0	160,822
Duplicated Record	14	160,808

**Table 5.4. ELA Grade 6 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	159,620
Wrong Subject	0	159,620
No Grade	0	159,620
Wrong Grade	21	159,599
Language Mismatched Form	220	159,379
School Type	111	159,268
Missing Entire Book	1,052	158,216
Invalid Score	0	158,216
Out-of-Range CR Scores	0	158,216
Duplicated Record	6	158,210

**Table 5.5. ELA Grade 7 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	150,384
Wrong Subject	0	150,384
No Grade	0	150,384
Wrong Grade	29	150,355
Language Mismatched Form	146	150,209
School Type	65	150,144
Missing Entire Book	1,283	148,861
Invalid Score	0	148,861
Out-of-Range CR Scores	0	148,861
Duplicated Record	4	148,857

**Table 5.6. ELA Grade 8 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	145,425
Wrong Subject	0	145,425
No Grade	0	145,425
Wrong Grade	37	145,388
Language Mismatched Form	147	145,241
School Type	66	145,175
Missing Entire Book	1,618	143,557
Invalid Score	0	143,557
Out-of-Range CR Scores	0	143,557
Duplicated Record	2	143,555

**Table 5.7. Mathematics Grade 3 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	179,827
Wrong Subject	0	179,827
No Grade	0	179,827
Wrong Grade	29	179,798
Language Mismatched Form	481	179,317
School Type	34	179,283
Missing Entire Book	397	178,886
Invalid Score	0	178,886
Out-of-Range CR Scores	0	178,886
Duplicated Record	16	178,870

**Table 5.8. Mathematics Grade 4 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	175,208
Wrong Subject	0	175,208
No Grade	0	175,208
Wrong Grade	13	175,195
Language Mismatched Form	535	174,660
School Type	0	174,660
Missing Entire Book	331	174,329
Invalid Score	0	174,329
Out-of-Range CR Scores	0	174,329
Duplicated Record	8	174,321

**Table 5.9. Mathematics Grade 5 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	163,890
Wrong Subject	0	163,890
No Grade	3	163,887
Wrong Grade	19	163,868
Language Mismatched Form	454	163,414
School Type	137	163,277
Missing Entire Book	271	163,006
Invalid Score	0	163,006
Out-of-Range CR Scores	0	163,006
Duplicated Record	14	162,992

**Table 5.10. Mathematics Grade 6 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	162,499
Wrong Subject	0	162,499
No Grade	1	162,498
Wrong Grade	27	162,471
Language Mismatched Form	735	161,736
School Type	103	161,633
Missing Entire Book	411	161,222
Invalid Score	0	161,222
Out-of-Range CR Scores	0	161,222
Duplicated Record	6	161,216

**Table 5.11. Mathematics Grade 7 Data Cleaning**

<b>Exclusion Rule</b>	<b># Deleted</b>	<b># Cases Remain</b>
Initial Number of Cases	n/a	148,630
Wrong Subject	0	148,630
No Grade	1	148,629
Wrong Grade	39	148,590
Language Mismatched Form	648	147,942
School Type	63	147,879
Missing Entire Book	623	147,256
Invalid Score	0	147,256
Out-of-Range CR Scores	0	147,256
Duplicated Record	4	147,252

**Table 5.12. Mathematics Grade 8 Data Cleaning**

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	116,810
Wrong Subject	0	116,810
No Grade	2	116,808
Wrong Grade	36	116,772
Language Mismatched Form	547	116,225
School Type	73	116,152
Missing Entire Book	960	115,192
Invalid Score	0	115,192
Out-of-Range CR Scores	0	115,192
Duplicated Record	2	115,190

**5.3. Classical Analysis and Calibration Sample Characteristics**

The cleaned and sampled-down data sets included more than 98% of New York State students and were used for classical analyses, calibration, and linking. 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	86,132	49.59
	Male	87,563	50.41
Ethnicity	Asian	17,910	10.31
	Black	31,562	18.17
	Hispanic	49,379	28.43
	American Indian	1,204	0.69
	Multiracial	4,343	2.50
	Pacific Islander	548	0.32
	White	68,749	39.58
NRC	New York	70,267	40.45
	Big 4 Cities	7,489	4.31
	Urban/Suburban	13,771	7.93
	Rural	9,539	5.49
	Average Needs	39,596	22.80
	Low Needs	17,480	10.06
	Charter School	9,645	5.55
	Non-Public	5,908	3.40

Demographic Category		N-Count	% of Total N-Count*
SWD	No	148,570	85.53
	Yes	25,125	14.47
SUA	No	149,680	86.17
	Yes	24,015	13.83
ELL	No	157,121	90.46
	Yes	16,574	9.54

\*The total n-count was 173,695.

**Table 5.14. ELA Grade 4 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	84,532	49.38
	Male	86,653	50.62
Ethnicity	Asian	17,504	10.23
	Black	31,862	18.61
	Hispanic	47,741	27.89
	American Indian	1,091	0.64
	Multiracial	3,689	2.15
	Pacific Islander	627	0.37
	White	68,671	40.12
NRC	New York	68,816	40.20
	Big 4 Cities	7,249	4.23
	Urban/Suburban	13,092	7.65
	Rural	9,061	5.29
	Average Needs	37,617	21.97
	Low Needs	16,928	9.89
	Charter School	8,189	4.78
	Non-Public	10,233	5.98
SWD	No	145,066	84.74
	Yes	26,119	15.26
SUA	No	144,297	84.29
	Yes	26,888	15.71
ELL	No	156,299	91.30
	Yes	14,886	8.70

\*The total n-count was 171,185.

**Table 5.15. ELA Grade 5 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	79,090	49.18
	Male	81,718	50.82
Ethnicity	Asian	16,724	10.40
	Black	30,617	19.04
	Hispanic	44,779	27.85
	American Indian	1,069	0.66
	Multiracial	2,948	1.83
	Pacific Islander	450	0.28
	White	64,221	39.94
NRC	New York	66,871	41.58
	Big 4 Cities	6,465	4.02
	Urban/Suburban	12,182	7.58
	Rural	8,489	5.28
	Average Needs	35,820	22.28
	Low Needs	16,833	10.47
	Charter School	8,373	5.21
	Non-Public	5,775	3.59
SWD	No	134,107	83.40
	Yes	26,701	16.60
SUA	No	133,429	82.97
	Yes	27,379	17.03
ELL	No	148,795	92.53
	Yes	12,013	7.47

\*The total n-count was 160,808.

**Table 5.16. ELA Grade 6 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	77,772	49.16
	Male	80,438	50.84
Ethnicity	Asian	17,183	10.86
	Black	30,271	19.13
	Hispanic	42,276	26.72
	American Indian	1,061	0.67
	Multiracial	2,513	1.59
	Pacific Islander	425	0.27
	White	64,481	40.76

Demographic Category		N-Count	% of Total N-Count*
NRC	New York	63,195	39.94
	Big 4 Cities	6,393	4.04
	Urban/Suburban	10,898	6.89
	Rural	8,184	5.17
	Average Needs	34,109	21.56
	Low Needs	17,046	10.77
	Charter School	9,189	5.81
	Non-Public	9,196	5.81
SWD	No	132,618	83.82
	Yes	25,592	16.18
SUA	No	132,198	83.56
	Yes	26,012	16.44
ELL	No	146,460	92.57
	Yes	11,750	7.43

\*The total n-count was 158,210.

**Table 5.17. ELA Grade 7 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	72,555	48.74
	Male	76,302	51.26
Ethnicity	Asian	16,249	10.92
	Black	29,565	19.86
	Hispanic	40,195	27.00
	American Indian	1,098	0.74
	Multiracial	2,036	1.37
	Pacific Islander	418	0.28
	White	59,296	39.83
NRC	New York	63,853	42.90
	Big 4 Cities	5,892	3.96
	Urban/Suburban	10,263	6.89
	Rural	7,777	5.22
	Average Needs	31,388	21.09
	Low Needs	16,503	11.09
	Charter School	8,180	5.50
	Non-Public	5,001	3.36
SWD	No	124,723	83.79
	Yes	24,134	16.21

Demographic Category		N-Count	% of Total N-Count*
SUA	No	124,861	83.88
	Yes	23,996	16.12
ELL	No	138,515	93.05
	Yes	10,342	6.95

\*The total n-count was 148,857.

**Table 5.18. ELA Grade 8 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	69,999	48.76
	Male	73,556	51.24
Ethnicity	Asian	16,027	11.16
	Black	30,083	20.96
	Hispanic	39,239	27.33
	American Indian	920	0.64
	Multiracial	1,599	1.11
	Pacific Islander	374	0.26
	White	55,313	38.53
NRC	New York	63,737	44.40
	Big 4 Cities	5,721	3.99
	Urban/Suburban	9,184	6.40
	Rural	7,307	5.09
	Average Needs	28,192	19.64
	Low Needs	14,983	10.44
	Charter School	6,816	4.75
	Non-Public	7,615	5.30
SWD	No	121,096	84.36
	Yes	22,459	15.64
SUA	No	120,996	84.29
	Yes	22,559	15.71
ELL	No	133,460	92.97
	Yes	10,095	7.03

\*The total n-count was 143,555.

**Table 5.19. Mathematics Grade 3 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	88,423	49.43
	Male	90,447	50.57
Ethnicity	Asian	18,673	10.44
	Black	32,281	18.05
	Hispanic	51,194	28.62
	American Indian	1,244	0.70
	Multiracial	4,341	2.43
	Pacific Islander	578	0.32
	White	70,559	39.45
NRC	New York	71,888	40.19
	Big 4 Cities	7,798	4.36
	Urban/Suburban	13,776	7.70
	Rural	9,429	5.27
	Average Needs	39,072	21.84
	Low Needs	17,440	9.75
	Charter School	9,565	5.35
	Non-Public	9,902	5.54
SWD	No	152,937	85.50
	Yes	25,933	14.50
SUA	No	154,205	86.21
	Yes	24,665	13.79
ELL	No	160,280	89.61
	Yes	18,590	10.39

\*The total n-count was 178,870.

**Table 5.20. Mathematics Grade 4 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	85,869	49.26
	Male	88,452	50.74
Ethnicity	Asian	18,124	10.40
	Black	32,575	18.69
	Hispanic	49,396	28.34
	American Indian	1,114	0.64
	Multiracial	3,693	2.12
	Pacific Islander	656	0.38
	White	68,763	39.45

Demographic Category		N-Count	% of Total N-Count*
NRC	New York	70,160	40.25
	Big 4 Cities	7,329	4.20
	Urban/Suburban	12,913	7.41
	Rural	8,920	5.12
	Average Needs	37,102	21.28
	Low Needs	17,038	9.77
	Charter School	8,453	4.85
	Non-Public	12,406	7.12
SWD	No	147,733	84.75
	Yes	26,588	15.25
SUA	No	147,276	84.49
	Yes	27,045	15.51
ELL	No	158,012	90.64
	Yes	16,309	9.36

\*The total n-count was 174,321.

**Table 5.21. Mathematics Grade 5 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	79,609	48.84
	Male	83,383	51.16
Ethnicity	Asian	17,389	10.67
	Black	31,457	19.30
	Hispanic	46,546	28.56
	American Indian	1,111	0.68
	Multiracial	3,027	1.86
	Pacific Islander	484	0.30
	White	62,978	38.64
NRC	New York	68,243	41.87
	Big 4 Cities	6,683	4.10
	Urban/Suburban	11,954	7.33
	Rural	8,188	5.02
	Average Needs	34,960	21.45
	Low Needs	16,695	10.24
	Charter School	9,051	5.55
	Non-Public	7,218	4.43
SWD	No	136,016	83.45
	Yes	26,976	16.55

Demographic Category		N-Count	% of Total N-Count*
SUA	No	135,559	83.17
	Yes	27,433	16.83
ELL	No	149,593	91.78
	Yes	13,399	8.22

\*The total n-count was 162,992.

**Table 5.22. Mathematics Grade 6 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	79,050	49.03
	Male	82,166	50.97
Ethnicity	Asian	17,833	11.06
	Black	31,008	19.23
	Hispanic	43,781	27.16
	American Indian	1,077	0.67
	Multiracial	2,513	1.56
	Pacific Islander	455	0.28
	White	64,549	40.04
NRC	New York	64,335	39.91
	Big 4 Cities	6,440	3.99
	Urban/Suburban	10,412	6.46
	Rural	7,757	4.81
	Average Needs	33,015	20.48
	Low Needs	16,735	10.38
	Charter School	9,825	6.09
	Non-Public	12,697	7.88
SWD	No	135,817	84.25
	Yes	25,399	15.75
SUA	No	135,817	84.25
	Yes	25,399	15.75
ELL	No	147,846	91.71
	Yes	13,370	8.29

\*The total n-count was 161,216.

**Table 5.23. Mathematics Grade 7 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	71,650	48.66
	Male	75,602	51.34
Ethnicity	Asian	16,614	11.28
	Black	29,690	20.16
	Hispanic	41,116	27.92
	American Indian	1,087	0.74
	Multiracial	1,942	1.32
	Pacific Islander	432	0.29
	White	56,371	38.28
NRC	New York	64,686	43.93
	Big 4 Cities	5,826	3.96
	Urban/Suburban	9,475	6.43
	Rural	7,140	4.85
	Average Needs	28,987	19.69
	Low Needs	15,649	10.63
	Charter School	8,474	5.75
	Non-Public	7,015	4.76
SWD	No	123,823	84.09
	Yes	23,429	15.91
SUA	No	124,359	84.45
	Yes	22,893	15.55
ELL	No	135,967	92.34
	Yes	11,285	7.66

\*The total n-count was 147,252.

**Table 5.24. Mathematics Grade 8 Sample Characteristics**

Demographic Category		N-Count	% of Total N-Count*
Gender	Female	55,286	48.00
	Male	59,904	52.00
Ethnicity	Asian	11,147	9.68
	Black	26,458	22.97
	Hispanic	35,547	30.86
	American Indian	761	0.66
	Multiracial	1,184	1.03
	Pacific Islander	315	0.27
	White	39,778	34.53

Demographic Category		N-Count	% of Total N-Count*
NRC	New York	53,996	46.88
	Big 4 Cities	5,128	4.45
	Urban/Suburban	7,474	6.49
	Rural	5,520	4.79
	Average Needs	18,111	15.72
	Low Needs	8,222	7.14
	Charter School	5,926	5.14
	Non-Public	10,813	9.39
SWD	No	94,527	82.06
	Yes	20,663	17.94
SUA	No	94,830	82.32
	Yes	20,360	17.68
ELL	No	103,743	90.06
	Yes	11,447	9.94

\*The total n-count was 115,190.

#### 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 (e.g., a printing error or two correct answers to one item), item analysis is the stage at which errors should be flagged and evaluated for rectification (suppression, credit, or other acceptable solution). 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) for the ELA tests ranged from 0.29 to 0.96. For Grade 3, the item p-values ranged from 0.30 to 0.90, with a mean of 0.57. For Grade 4, the item p-values ranged from 0.39 to 0.75, with a mean of 0.55. For Grade 5, the item p-values ranged from 0.36 to 0.87, with a mean of 0.62. For Grade 6, the item p-values ranged from 0.33 to 0.78, with a mean of 0.57. For Grade 7, the item p-values ranged from 0.29 to 0.79, with a mean of 0.57. For Grade 8, the item p-values ranged from 0.42 to 0.96, with a mean of 0.68. These p-value statistics are in Appendix M, Tables M1–M6, along with other classical test statistics of the keys.

Item difficulties (p-values) on the Mathematics tests ranged from 0.12 to 0.90. For Grade 3, the item p-values ranged from 0.24 to 0.90, with a mean of 0.63. For Grade 4, the item p-values ranged from 0.23 to 0.83, with a mean of 0.61. For Grade 5, the item p-values ranged from 0.20 to 0.86, with a mean of 0.56. For Grade 6, the item p-values ranged from 0.12 to 0.85, with a mean of 0.51. For Grade 7, the item p-values ranged from 0.28 to 0.80, with a mean of 0.49. For Grade 8, the item p-values ranged from 0.19 to 0.83, with a mean of 0.49. These statistics are provided in Appendix M, Tables M7–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 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 Common Core ELA and Mathematics Tests was that the point-biserial correlation for the key for MC items should be equal to or greater than .20, which would indicate that students who responded correctly to that item also tended to do well on the overall test. There were very few exceptions to this guideline, due to content considerations, which 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 0.09 to 0.72, as shown in Appendix M, Tables M1–M6. For Grade 3, the item pbis values ranged from 0.30 to 0.65, with a mean of 0.45. For Grade 4, the item pbis values ranged from 0.22 to 0.70, with a mean of 0.40. For Grade 5, the item pbis values ranged from 0.16 to 0.67, with a mean of 0.40. For Grade 6, the item pbis values ranged from 0.13 to 0.71, with a mean of 0.37. For Grade 7, the item pbis values ranged from 0.16 to 0.72, with a mean of 0.40. For Grade 8, the item pbis values ranged from 0.09 to 0.72, with a mean of 0.43.

Point biserials for correct answer options on the Mathematics tests ranged from 0.03 to 0.75, as shown in Appendix M, Tables M7–M12. For Grade 3, the item pbis values ranged from 0.23 to 0.69, with a mean of 0.46. For Grade 4, the item pbis values ranged from 0.28 to 0.73, with a mean of 0.52. For Grade 5, the item pbis values ranged from 0.03 to 0.69, with a mean of 0.48. For Grade 6, the item pbis values ranged from 0.21 to 0.70, with a mean of 0.45. For Grade 7,

the item pbis values ranged from 0.24 to 0.75, with a mean of 0.48. For Grade 8, the item pbis values ranged from 0.24 to 0.70, with a mean of 0.43.

#### 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 Common Core ELA and Mathematics Tests, respectively. The industry standard general rule of thumb is that omit rates for multiple-choice items should be less than 5.0%. Omit rates across multiple-choice and constructed-response items on the Grades 3–8 Common Core 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 booklets. 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 2015. Turning to the analysis of the 2015 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;  $\alpha = .01$ ) 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), and English language learners (focal group: English language learners; reference group: Non-English language learners). The DIF analyses were conducted using all cases from the clean data sets. Table 5.29 and Table 5.30 show the numbers of cases for the subgroups for ELA and Mathematics, respectively.

**Table 5.25. ELA Classical DIF Sample N-Counts**

Grade	Ethnicity				Gender		Needs/Resource Capacity		ELLs	
	Black	Hispanic	Asian	White	Female	Male	High	Low	ELL	Non-ELL
3	31,562	49,379	17,910	68,749	86,132	87,563	101,066	57,076	16,574	157,121
4	31,862	47,741	17,504	68,671	84,532	86,653	98,218	54,545	14,886	156,299
5	30,617	44,779	16,724	64,221	79,090	81,718	94,007	52,653	12,013	148,795
6	30,271	42,276	17,183	64,481	77,772	80,438	88,670	51,155	11,750	146,460
7	29,565	40,195	16,249	59,296	72,555	76,302	87,785	47,891	10,342	138,515
8	30,083	39,239	16,027	55,313	69,999	73,556	85,949	43,175	10,095	133,460

**Table 5.26. Mathematics Classical DIF Sample N-Counts**

Grade	Ethnicity				Gender		Needs/Resource Capacity		ELLs	
	Black	Hispanic	Asian	White	Female	Male	High	Low	ELL	Non-ELL
3	32,281	51,194	18,673	70,559	88,423	90,447	102,891	56,512	18,590	160,280
4	32,575	49,396	18,124	68,763	85,869	88,452	99,322	54,140	16,309	158,012
5	31,457	46,546	17,389	62,978	79,609	83,383	95,068	51,655	13,399	149,593
6	31,008	43,781	17,833	64,549	79,050	82,166	88,944	49,750	13,370	147,846
7	29,690	41,116	16,614	56,371	71,650	75,602	87,127	44,636	11,285	135,967
8	26,458	35,547	11,147	39,778	55,286	59,904	72,118	26,333	11,447	103,743

Table 5.31 (ELA) and Table 5.32 (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	2
4	4
5	10
6	6
7	9
8	8

Table 5.28. Mathematics Items Flagged for DIF

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

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 by the content experts again, and no sign of potential bias was found. In other words, based on combinations of statistical and content evaluations, none of the items on the 3–8 tests showed bias.

## Section 6: IRT Calibration and Linking

---

### 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  $\theta$  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  $\theta$  having a score  $(k - 1)$  at the  $k$ th level of the  $j$ th item is:

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

where

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

and

$k$  is the item response category ( $k = 1, 2, \dots, m_j$ ).

The  $m_j$  denotes the number of score levels for the  $j$ th 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

$\alpha_j$  and  $\gamma_{ji}$  are the free parameters to be estimated from the data.

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

## 6.2. Calibration Sample

The cleaned data were used for calibration and linking of the NYSTP 2016 Grades 3–8 Common Core ELA and Mathematics Tests. It should be noted that the sample sizes were adequate, as the calibration and linking were 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 2016 operational test samples were generally comparable to 2015 populations in terms of NRC, student race and ethnicity, proportions of ELLs, proportions of students with disabilities, and proportions of students using testing accommodations.

**Table 6.1. ELA Grades 3 and 4 Demographic Statistics**

Demographic Category		Grade 3		Grade 4	
		2015 Population	2016 Sample	2015 Population	2016 Sample
Gender	Female	49.08	49.59	49.23	49.38
	Male	50.92	50.41	50.77	50.62
Ethnicity	Asian	9.84	10.31	9.72	10.23
	Black	18.92	18.17	19.22	18.61
	Hispanic	28.22	28.43	27.39	27.89
	American Indian	0.66	0.69	0.62	0.64
	Multiracial	2.20	2.50	1.81	2.15
	Pacific Islander	0.35	0.32	0.29	0.37
	White	39.80	39.58	40.95	40.12
NRC	New York	39.58	40.45	39.02	40.20
	Big 4 Cities	4.24	4.31	3.99	4.23
	Urban/Suburban	7.88	7.93	7.36	7.65
	Rural	5.05	5.49	4.72	5.29
	Average Needs	22.18	22.80	21.60	21.97
	Low Needs	10.09	10.06	10.18	9.89
	Charter	5.20	5.55	4.49	4.78
	Non-Public	5.68	3.40	8.56	5.98
SWD	No	84.89	85.53	84.24	84.74
	Yes	15.11	14.47	15.76	15.26
SUA	No	88.28	86.17	88.40	84.29
	Yes	11.72	13.83	11.60	15.71
ELL	No	90.73	90.46	91.72	91.30
	Yes	9.27	9.54	8.28	8.70

**Table 6.2. ELA Grades 5 and 6 Demographic Statistics**

Demographic Category		Grade 5		Grade 6	
		2015 Population	2016 Sample	2015 Population	2016 Sample
Gender	Female	49.15	49.18	48.91	49.16
	Male	50.85	50.82	51.09	50.84
Ethnicity	Asian	10.24	10.40	9.95	10.86
	Black	19.36	19.04	19.71	19.13
	Hispanic	26.57	27.85	26.50	26.72
	American Indian	0.62	0.66	0.66	0.67
	Multiracial	1.50	1.83	1.39	1.59
	Pacific Islander	0.25	0.28	0.28	0.27
	White	41.46	39.94	41.50	40.76
NRC	New York	38.65	41.58	37.67	39.94
	Big 4 Cities	4.00	4.02	3.89	4.04
	Urban/Suburban	7.24	7.58	7.02	6.89
	Rural	4.78	5.28	4.73	5.17
	Average Needs	22.50	22.28	21.66	21.56
	Low Needs	11.27	10.47	10.82	10.77
	Charter	5.35	5.21	5.35	5.81
	Non-Public	6.12	3.59	8.76	5.81
SWD	No	83.31	83.40	83.93	83.82
	Yes	16.69	16.60	16.07	16.18
SUA	No	87.66	82.97	88.47	83.56
	Yes	12.34	17.03	11.53	16.44
ELL	No	92.19	92.53	93.03	92.57
	Yes	7.81	7.47	6.97	7.43

**Table 6.3. ELA Grades 7 and 8 Demographic Statistics**

Demographic Category		Grade 7		Grade 8	
		2015 Population	2016 Sample	2015 Population	2016 Sample
Gender	Female	48.78	48.74	48.49	48.76
	Male	51.22	51.26	51.51	51.24
Ethnicity	Asian	9.94	10.92	10.11	11.16
	Black	20.57	19.86	21.06	20.96
	Hispanic	26.49	27.00	26.34	27.33
	American Indian	0.61	0.74	0.59	0.64
	Multiracial	1.13	1.37	1.03	1.11
	Pacific Islander	0.25	0.28	0.25	0.26
	White	41.02	39.83	40.61	38.53
NRC	New York	39.69	42.90	40.42	44.40
	Big 4 Cities	3.92	3.96	3.93	3.99
	Urban/Suburban	7.03	6.89	6.91	6.40
	Rural	4.86	5.22	4.90	5.09
	Average Needs	21.25	21.09	20.44	19.64
	Low Needs	11.86	11.09	11.26	10.44
	Charter	4.89	5.50	3.71	4.75
	Non-Public	6.43	3.36	8.31	5.30
SWD	No	83.67	83.79	84.17	84.36
	Yes	16.33	16.21	15.83	15.64
SUA	No	88.91	83.88	89.28	84.29
	Yes	11.09	16.12	10.72	15.71
ELL	No	93.17	93.05	93.75	92.97
	Yes	6.83	6.95	6.25	7.03

**Table 6.4. Mathematics Grades 3 and 4 Demographic Statistics**

Demographic Category		Grade 3		Grade 4	
		2015 Population	2016 Sample	2015 Population	2016 Sample
Gender	Female	48.93	49.43	49.02	49.26
	Male	51.07	50.57	50.98	50.74
Ethnicity	Asian	10.17	10.44	10.09	10.40
	Black	18.91	18.05	19.18	18.69
	Hispanic	28.61	28.62	27.90	28.34
	American Indian	0.66	0.70	0.61	0.64
	Multiracial	2.14	2.43	1.73	2.12
	Pacific Islander	0.36	0.32	0.30	0.38
	White	39.15	39.45	40.18	39.45
NRC	New York	40.45	40.19	40.08	40.25
	Big 4 Cities	4.29	4.36	3.97	4.20
	Urban/Suburban	7.78	7.70	7.19	7.41
	Rural	4.88	5.27	4.51	5.12
	Average Needs	21.56	21.84	20.90	21.28
	Low Needs	9.92	9.75	10.08	9.77
	Charter	5.21	5.35	4.53	4.85
	Non-Public	5.81	5.54	8.65	7.12
SWD	No	85.02	85.50	84.34	84.75
	Yes	14.98	14.50	15.66	15.25
SUA	No	92.44	86.21	91.80	84.49
	Yes	7.56	13.79	8.20	15.51
ELL	No	88.13	89.61	88.67	90.64
	Yes	11.87	10.39	11.33	9.36

**Table 6.5. Mathematics Grades 5 and 6 Demographic Statistics**

Demographic Category		Grade 5		Grade 6	
		2015 Population	2016 Sample	2015 Population	2016 Sample
Gender	Female	48.96	48.84	48.80	49.03
	Male	51.04	51.16	51.20	50.97
Ethnicity	Asian	10.66	10.67	10.44	11.06
	Black	19.36	19.30	19.78	19.23
	Hispanic	27.19	28.56	27.20	27.16
	American Indian	0.59	0.68	0.65	0.67
	Multiracial	1.44	1.86	1.32	1.56
	Pacific Islander	0.26	0.30	0.29	0.28
	White	40.50	38.64	40.32	40.04
NRC	New York	40.01	41.87	39.48	39.91
	Big 4 Cities	4.01	4.10	3.85	3.99
	Urban/Suburban	7.05	7.33	6.72	6.46
	Rural	4.52	5.02	4.47	4.81
	Average Needs	21.63	21.45	20.45	20.48
	Low Needs	11.05	10.24	10.51	10.38
	Charter	5.45	5.55	5.49	6.09
	Non-Public	6.20	4.43	8.93	7.88
SWD	No	83.62	83.45	84.32	84.25
	Yes	16.38	16.55	15.68	15.75
SUA	No	88.15	83.17	88.46	84.25
	Yes	11.85	16.83	11.54	15.75
ELL	No	90.93	91.78	91.72	91.71
	Yes	9.07	8.22	8.28	8.29

**Table 6.6. Mathematics Grades 7 and 8 Demographic Statistics**

Demographic Category		Grade 7		Grade 8	
		2015 Population	2016 Sample	2015 Population	2016 Sample
Gender	Female	48.67	48.66	47.73	48.00
	Male	51.33	51.34	52.27	52.00
Ethnicity	Asian	10.49	11.28	8.93	9.68
	Black	20.63	20.16	23.67	22.97
	Hispanic	27.50	27.92	30.18	30.86
	American Indian	0.58	0.74	0.61	0.66
	Multiracial	1.05	1.32	0.95	1.03
	Pacific Islander	0.25	0.29	0.26	0.27
	White	39.50	38.28	35.40	34.53
NRC	New York	42.34	43.93	45.49	46.88
	Big 4 Cities	3.81	3.96	4.45	4.45
	Urban/Suburban	6.62	6.43	6.77	6.49
	Rural	4.43	4.85	4.67	4.79
	Average Needs	19.57	19.69	16.22	15.72
	Low Needs	11.19	10.63	7.69	7.14
	Charter	5.12	5.75	4.20	5.14
	Non-Public	6.84	4.76	10.41	9.39
SWD	No	84.12	84.09	81.80	82.06
	Yes	15.88	15.91	18.20	17.94
SUA	No	89.01	84.45	88.67	82.32
	Yes	10.99	15.55	11.33	17.68
ELL	No	91.53	92.34	90.58	90.06
	Yes	8.47	7.66	9.42	9.94

**6.2.1. Calibration Process**

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

The calibration of NYSTP 2016 Grades 3–8 Common Core 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 Common Core 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 in Section 6: IRT Calibration and Linking, above in Section 6.1. IRT Models and Rationale for Use,  $m_j$  denotes the number of score levels for the  $j$ th 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**

Grade	Item-level			Student-level		
	Largest a-Parameter	Range of b- / Step Parameters		N-Count	Theta Est.*	
					Mean	SD
3	1.304	-1.844	1.058	173,540	0.01	0.94
4	1.031	-1.120	1.320	171,061	0.00	0.94
5	1.304	-2.390	1.662	160,807	0.00	0.94
6	1.199	-1.323	2.746	158,161	0.00	0.94
7	1.362	-2.054	1.758	148,857	0.00	0.94
8	1.328	-2.447	1.005	143,555	-0.01	0.94

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

**Table 6.8. Mathematics Calibration Results**

Grade	Item-level			Student-level		
	Largest a-Parameter	Range of b- / Step Parameters		N-Count	Theta Est.*	
					Mean	SD
3	1.676	-2.820	1.363	178,870	0.01	0.93
4	1.725	-1.630	1.066	174,321	0.01	0.92
5	2.636	-4.310	1.354	162,795	0.01	0.93
6	2.053	-1.345	1.898	160,851	0.03	0.92
7	2.190	-1.494	1.240	146,870	0.04	0.91
8	1.867	-0.958	1.554	114,953	0.05	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  $Q_I$  procedure described by Yen (1981) was used to measure fit to the three-parameter

model. Students are rank-ordered on the basis of  $\hat{\theta}$  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:

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

with:

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

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

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

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,  $Q_{Ij}$  was transformed to  $Z_{Q_{Ij}}$  where:

$$Z_{Q_{Ij}} = (Q_{Ij} - df) / (2df)^{1/2}$$

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  $Z_{Q_{Ij}} \text{Crit}$  was used to flag items and was calculated using the expression

$$Z_{Q_{Ij}} \text{Crit} = \left( \frac{N}{1500} \right) * 4,$$

where  $N$  is the calibration sample size.

To compute the  $Q_{Ij}$  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  $Z_{Q_{Ij}}$  was greater than the value of  $Z_{Q_{Ij}}$  critical. If the obtained  $Z_{Q_{Ij}}$  was less than  $Z_{Q_{Ij}}$  critical, the items were rated as having acceptable fit. The fact that the majority of the items in the NYSTP 2016 Grades 3–8 Common Core ELA and Mathematics Tests demonstrated good model fit further supports the use of the chosen models. Item fit statistics are presented in Appendix O, in Tables O1–O12.

#### 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

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

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

$$Q_{3,ij} = r(d_j, d_{j'})$$

The generalization to items with multiple response categories uses

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

where

$$E_{ij} \equiv E(x|\hat{\theta}_i) = \sum_{k=1}^{m_j} kP_{jk}(\hat{\theta}_i)$$

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 .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.

The  $Q_3$  statistics were examined for all unique pairs of ELA and mathematics items. Items that were found to be significant in local dependency vary, depending on the subject and grade: one pair of items was found in ELA Grade 8. When reviewing the results for Mathematics, one pair of items each exceeded a correlation of .20 in Mathematics Grades 4, 7, and 8. The magnitudes of these statistics were not sufficient to warrant further concern or action (with the  $Q_3$  values being .27 for the ELA test and ranging from .23 to .28 for the Mathematics tests).

## 6.5. Linking and Scaling

With the new assessments being implemented in 2013, the scale was established after the data were collected. The purpose of linking was to place the 2016 item parameters and proficiency estimates on the same scale as those in 2015. The following steps constitute the linking process for each subject and grade:

1. Operational items as well as non-scored (i.e., external) anchor items were calibrated in IRTPRO.
2. The 2016 item parameter estimates for all anchor items—both scored and non-scored—enabled the establishment of the linking relationship via a test characteristic curve (TCC) method (Stocking and Lord, 1983; implemented in STUIRT, Kim, & Kolen, 2004) to the 2015 theta scale, using the established 2015 item parameter estimates for those same items. Tables 6.9 and 6.10 present the resulting linking coefficients. The following parameters were linked using the formulas below:

$$a_i^E = a_i^C / M_1^E,$$

$$b_i^E = M_1^E \cdot b_i^C + M_2^E, \text{ and}$$

$$d_{ij}^E = d_{ij}^C + \left[ \left( a_i^C / M_1^E \right) \right] \cdot M_2^E,$$

where

$M_1^E$  is defined as the multiplicative adjustment for linking and  $M_2^E$  is the additive adjustment for linking. The superscript “E” denotes linked item parameter estimates, while the superscript “C” denotes calibrated item parameter estimates.

**Table 6.9. ELA Linking Coefficients**

Grade	$M_1^E$	$M_2^E$
3	1.022	0.265
4	0.945	0.197
5	1.120	-0.082
6	1.015	-0.004
7	0.991	0.071
8	0.999	0.131

**Table 6.10. Mathematics Linking Coefficients**

Grade	$M_1^E$	$M_2^E$
3	1.141	0.197
4	1.175	0.156
5	1.148	0.202
6	1.179	0.170
7	1.175	0.169
8	1.188	-0.205

3. A raw-score-to-theta conversion chart was produced using the test characteristic curve (TCC) method (Stocking and Lord, 1983; see Section 6.8. Scoring Procedure for more details) and implemented in POLYEQUATE (Kolen & Cui, 2004). The theta estimates associated with the TCC method ( $\hat{\theta}_{TCC}$ ) must be linked back to the underlying theta scale established in the prior year (Spring 2015), and are computed as follows:

$$\theta^E = (M_1^E \cdot \hat{\theta}_{TCC}) + M_2^E$$

4. The TCC method does not produce theta estimates for raw scores below chance level or above the perfect score (highest obtainable raw score). In addition, for the scores at the low and high ends of the scale, some raw scores tended to have large theta estimates (for example, -7.999). Typically, the first obtainable theta value on a test corresponds to a very extreme theta value. The following adjustment/interpolation was conducted:

For any linked theta estimates ( $\theta^E$ ) that are outside of the range of -2.5 to 3, at the lower end of the scale, 0.25 was subtracted from the preceding theta value that is within the range; at the higher end of the scale, 0.25 was added to the previous theta value that is within the range, thus resulting in an adjusted theta estimate ( $\theta^A$ ) for those extremes. See the table below for an example at the lower end of the scale. Such an adjustment helps contain the theta scale within a reasonable range, and is standard practice in testing.

Raw Score	$\theta^E$	$\theta^A$
6	-5.30263	-3.37458
7	-3.66491	-3.12458
8	-3.03055	-2.87458
9	-2.76782	-2.62458
10	-2.37458	-2.37458

5. Once theta values were either estimated or interpolated for all raw scores, the raw-score-to-theta relationship was applied to each student, yielding a theta estimate corresponding to his or her raw score.
6. The adjusted theta estimates (presented in Tables 6.11 and 6.12) were then scaled using the established scaling coefficients from the prior year (Spring 2015) according to the following formula:

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

where

$M_1^S$  is defined as the multiplicative scaling coefficient, and  $M_2^S$  is the additive scaling coefficient.  $M_1^S$  and  $M_2^S$  are applied to a true score (i.e., the linked theta estimate) in order to obtain a scale score.

**Table 6.11. ELA Scaling Coefficients**

Grade	$M_1^S$	$M_2^S$
3	31.8145	301.4946
4	32.0356	300.7619
5	32.0160	300.9540
6	32.2585	300.6730
7	31.9257	300.8012
8	31.6273	300.9795

**Table 6.12. Mathematics Scaling Coefficients**

Grade	$M_1^S$	$M_2^S$
3	32.2491	299.8560
4	32.6982	300.1764
5	32.2199	300.6932
6	32.4213	300.3769
7	31.2289	301.1438
8	31.8685	301.1430

7. Scale scores range, approximately, from 100 to 400 across grades. The lowest and highest observed scale score (LOSS and HOSS, respectively) may vary by grade.
8. A series of anchor set stability checks were performed before finalizing the anchor set for each subject and grade; see Section 6.6. Anchor Set Evaluation, which follows this one.
9. For conditional standard error of measurement (CSEM), the scale scores (both estimated and interpolated) were used to compute the information function and CSEM.

Throughout this process, NYSED psychometricians have reviewed, and a senior scientist from HumRRO has independently verified, the results generated by Questar psychometricians.

## 6.6. Anchor Set Evaluation

In order to determine if each item from the anchor set performs similarly to when it was administered in the prior year, comparisons of individual item characteristic curves (ICCs) and item parameter estimates from the previous and current administrations were made. Initial

comparisons included a graphical inspection of the linearity of relationships between linked item parameter estimates from the 2015 and 2016 administrations. These revealed approximately linear relationships as well as similarities in item functions, and therefore provided support for the selected linking method used herein. Additional analyses of the correlations between linked item parameter estimates also provided evidence of strong linear relationships.

A formal process for validating the anchor set by using an objective criterion was used to determine if any items ought to be considered for removal from the anchor set. The linked item parameter estimates were used to calculate a weighted, squared deviation of the current ICC from the previous ICC, across the range of ability (i.e., theta, or  $\theta$ ) and under a hypothetical normal distribution for  $\theta$ . For a given item  $i$ , that quantity, called “d squared,” is given by

$$d_i^2 = \sum_k \{ [\text{Pr}_{i,16}(\theta_k) - \text{Pr}_{i,15}(\theta_k)]^2 \cdot g(\theta_k) \},$$

where  $i$  indexes anchor items;  $k$  indexes quadrature points for  $\theta$ ;  $\text{Pr}_{i,16}(\cdot)$  is the probability of a correct response to item  $i$  under the current calibration, while  $\text{Pr}_{i,15}(\cdot)$  is the same quantity under the previous calibration; and  $g(\theta_k)$  are weights for the quadrature points.

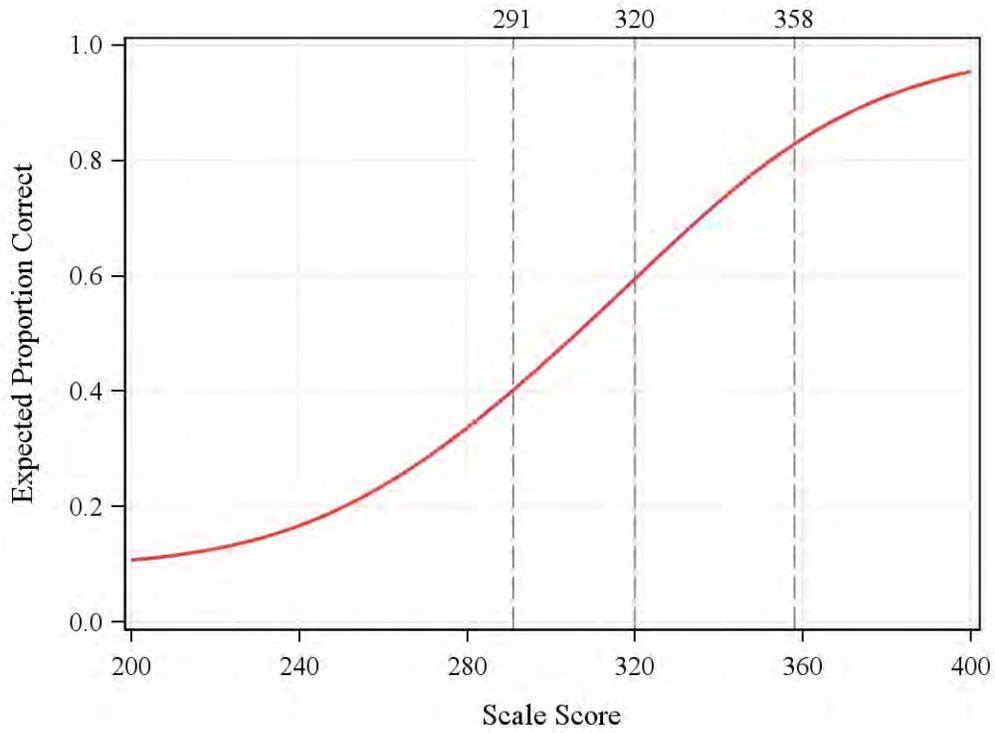
Historically, and as recently as the 2015 operational linking, a fixed criterion on this metric ( $d_i^2 \geq 0.05$ ) has been used for flagging items to be considered for removal from linking. The same approach and criterion were used for the linking of the 2016 operational forms to the 2015 scale score scale. This procedure minimizes the weighted squared differences between the two ICCs for each MC item: one based on 2015 item parameter estimates and the other on 2016 estimates. The differential item performance was evaluated by examining previous and current item parameters. The following steps were taken:

1. Before the iterative procedures start, the initial linking was performed, using all of the eligible anchor items as an anchor set, as described in Section 6.5: Linking and Scaling. The initial linking coefficients ( $M_1^E$  and  $M_2^E$ ) were obtained through the Stocking-Lord method.
2. The following process was repeated for at least five iterations or until the largest  $d_i^2 < 0.05$  is reached, whichever was greater:
  - a. For each anchor item,  $d_i^2$  was calculated as a weighted sum of the squared deviations between the ICCs based on old (2015) and new (2016) parameter estimates at each quadrature point and assuming a normal theta distribution.
  - b. The item having the largest  $d_i^2$  was identified and removed from the anchor set.
  - c. The linking procedures described in Section 6.5: Linking and Scaling were performed with the newly reduced anchor set.
  - d. New raw-score-to-scale-score tables were prepared as described in Section 6.8. Scoring Procedure.
3. Select the linking coefficients ( $M_1^E$  and  $M_2^E$ ) associated with the iteration selected in step 2 above.

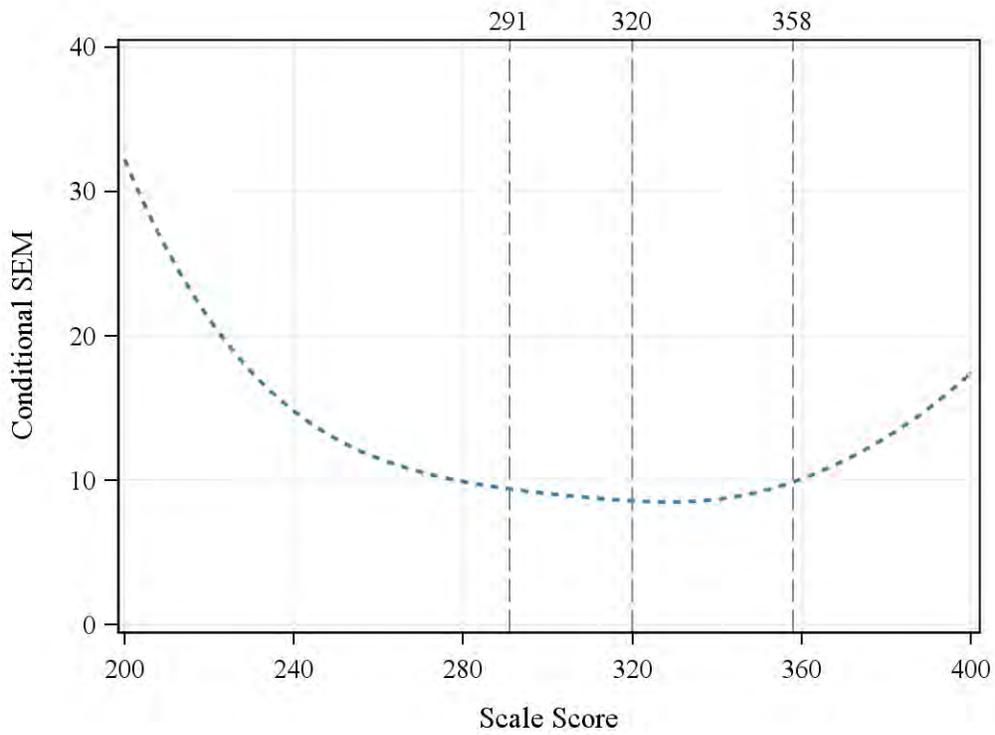
The items that are implicitly proposed for removal from the anchor set, based on the process described above, were summarized and evaluated. The only subject where items were proposed and ultimately approved for removal from the anchor set was mathematics, and one item each was removed from the anchor sets for Grades 5, 6, and 7.

### **6.7. Test Characteristic Curves**

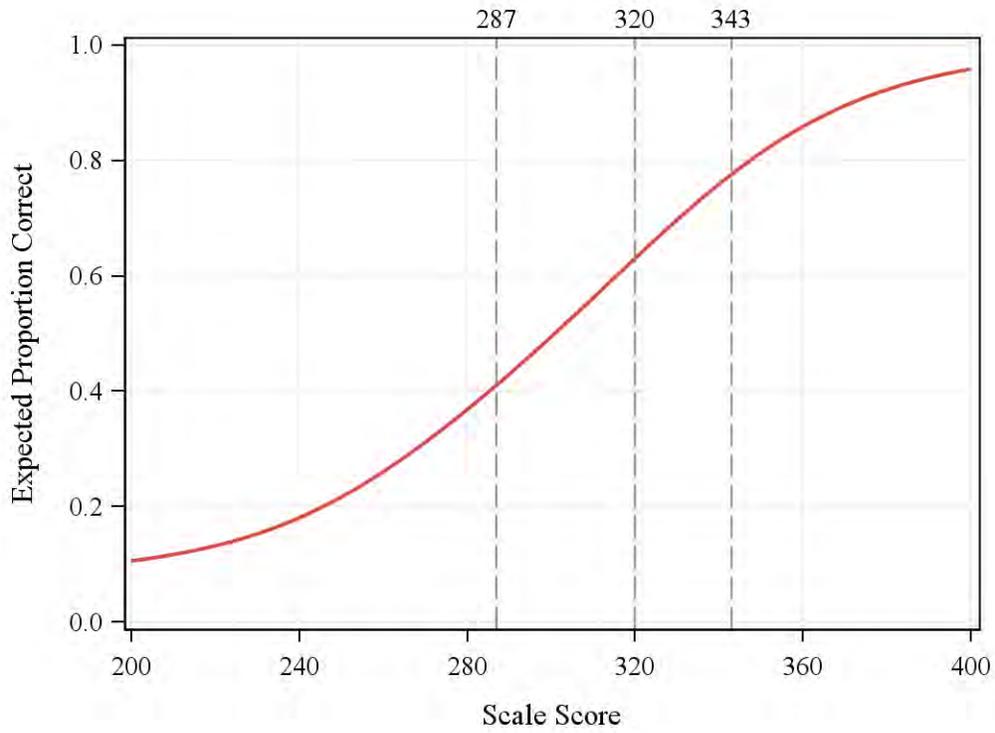
Test Characteristic Curves (TCCs) provide an overview of the tests in the IRT scale score metric. The 2016 TCCs were generated using final item parameters for all reporting test items administered in Spring 2016. TCCs are the summation of all the item characteristic curves (ICCs) for items that contribute 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.



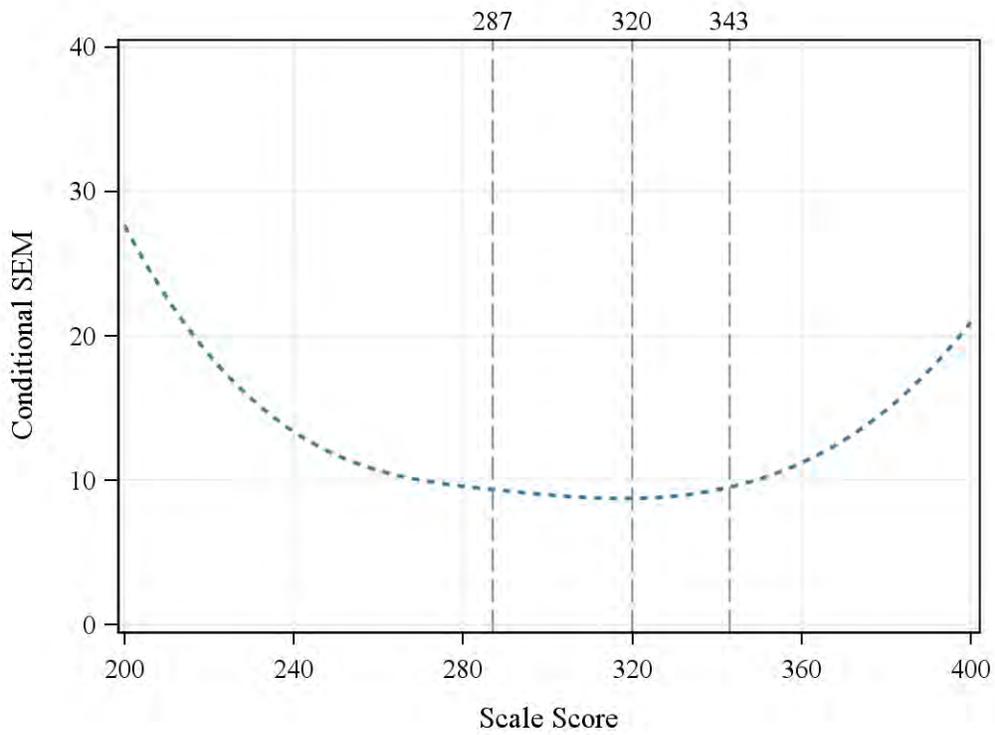
**Figure 6.1. ELA Grade 3 TCC**



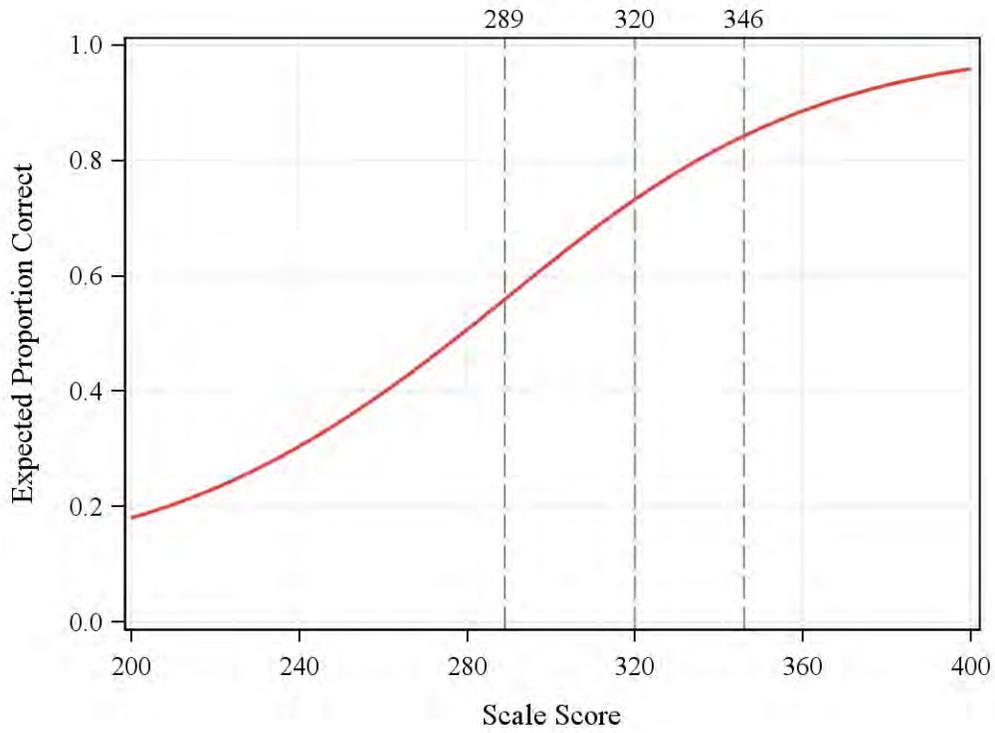
**Figure 6.2. ELA Grade 3 CSEM Curve**



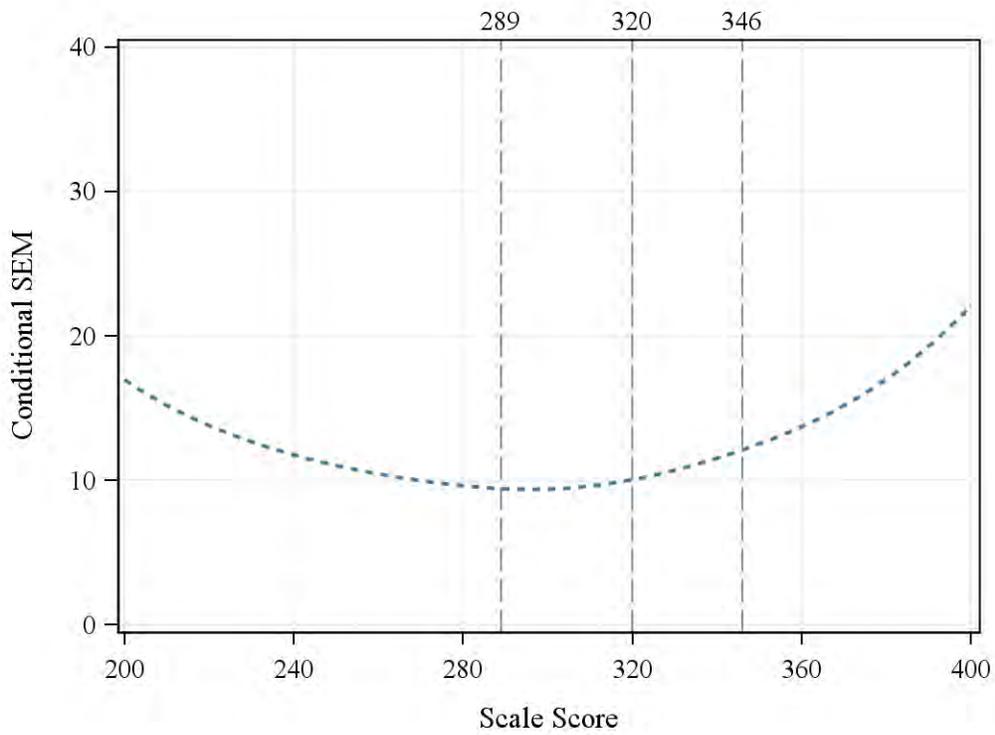
**Figure 6.3. ELA Grade 4 TCC**



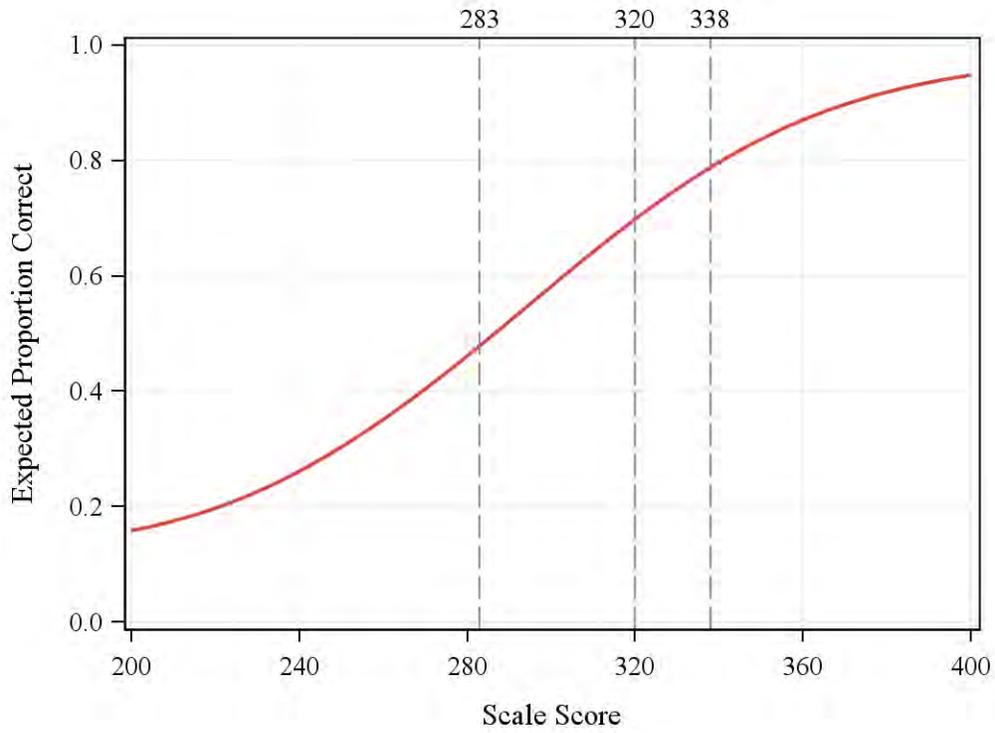
**Figure 6.4. ELA Grade 4 CSEM Curve**



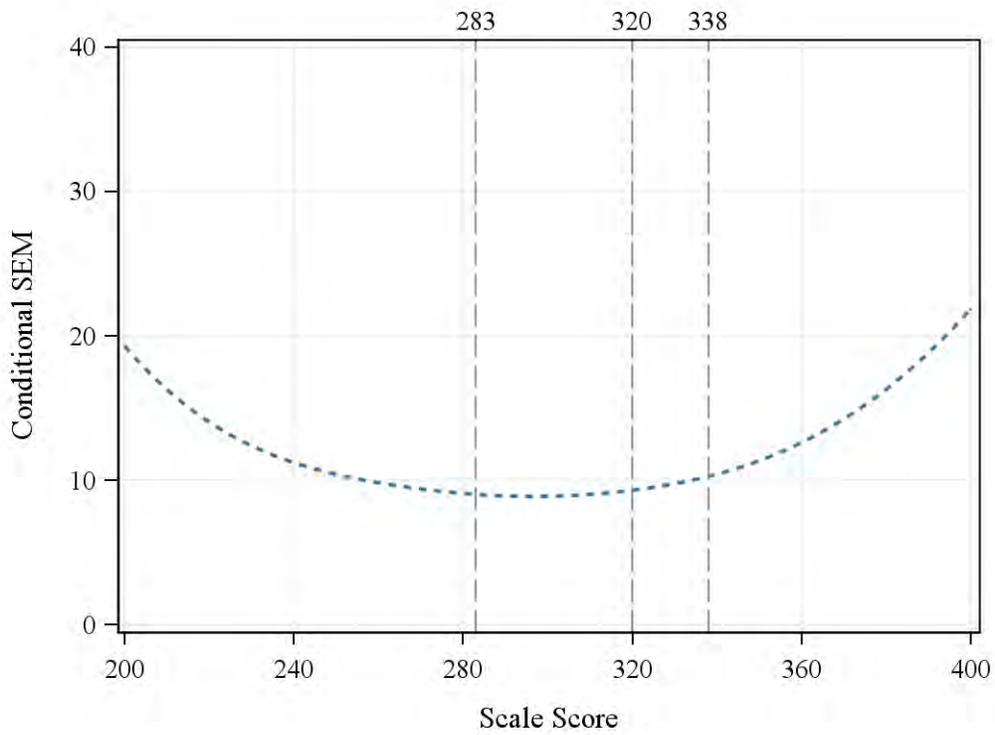
**Figure 6.5. ELA Grade 5 TCC**



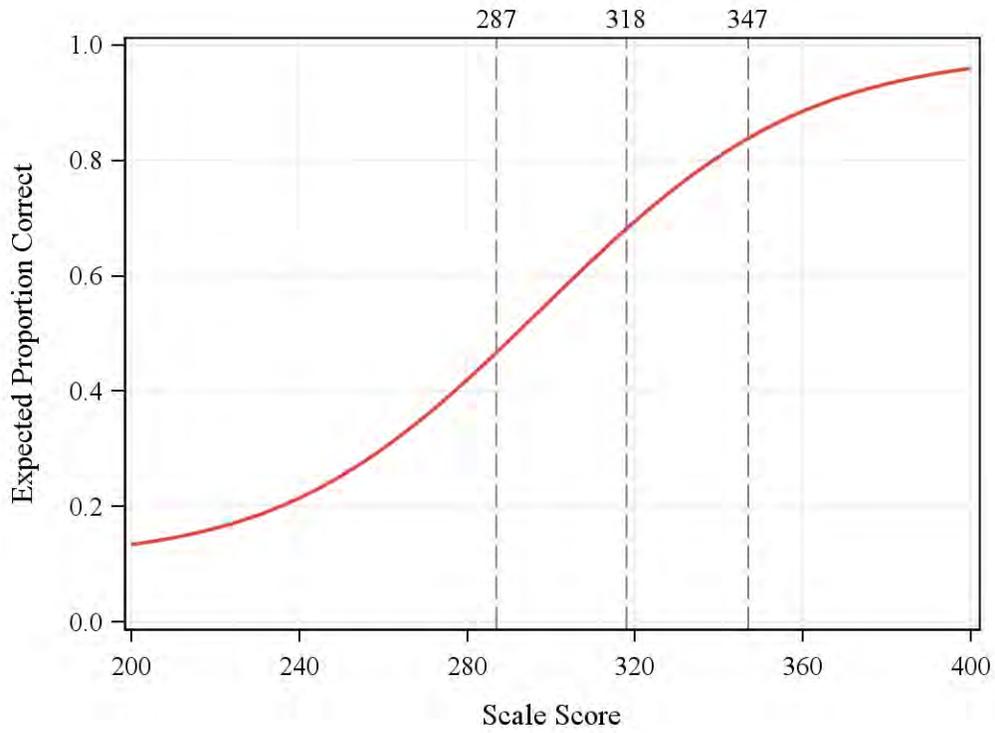
**Figure 6.6. ELA Grade 5 CSEM Curve**



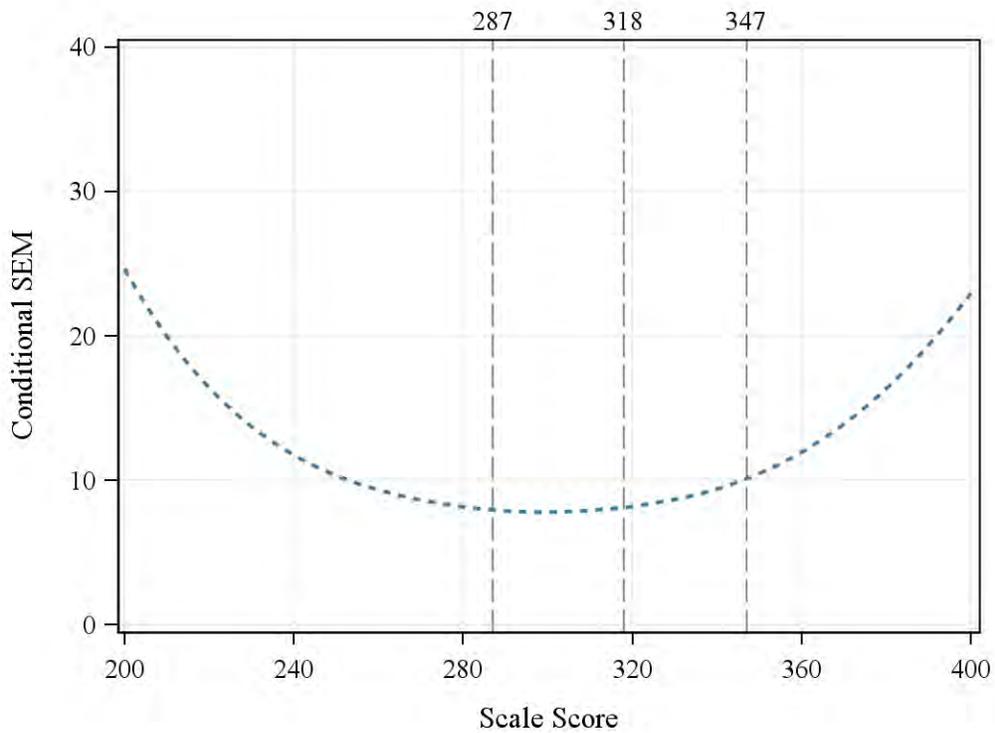
**Figure 6.7. ELA Grade 6 TCC**



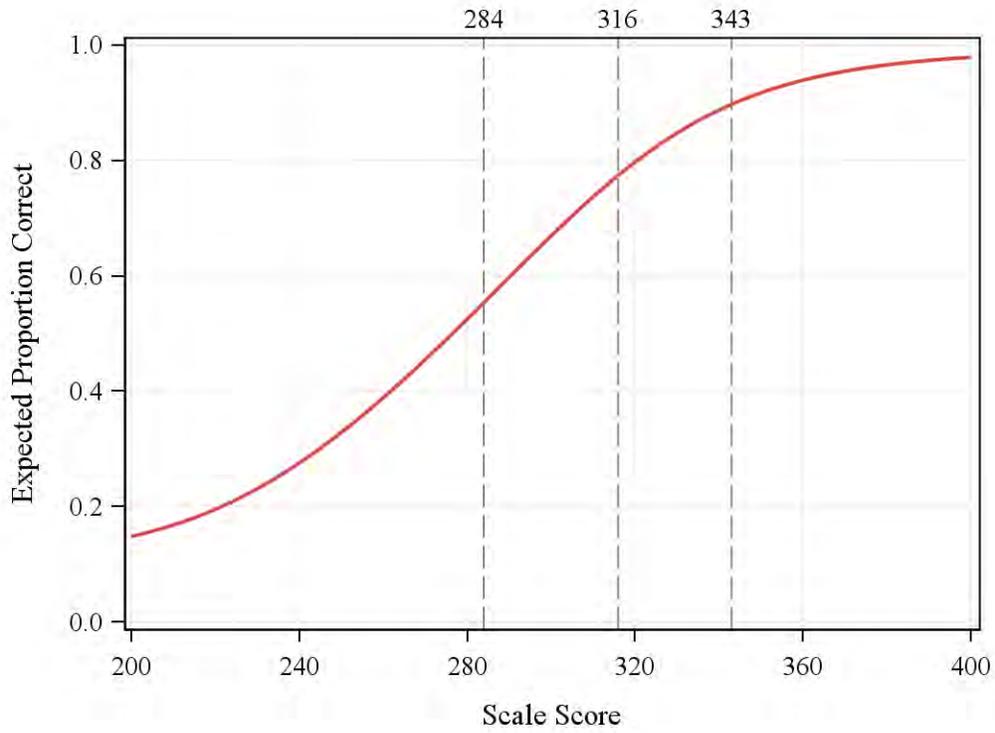
**Figure 6.8. ELA Grade 6 CSEM Curve**



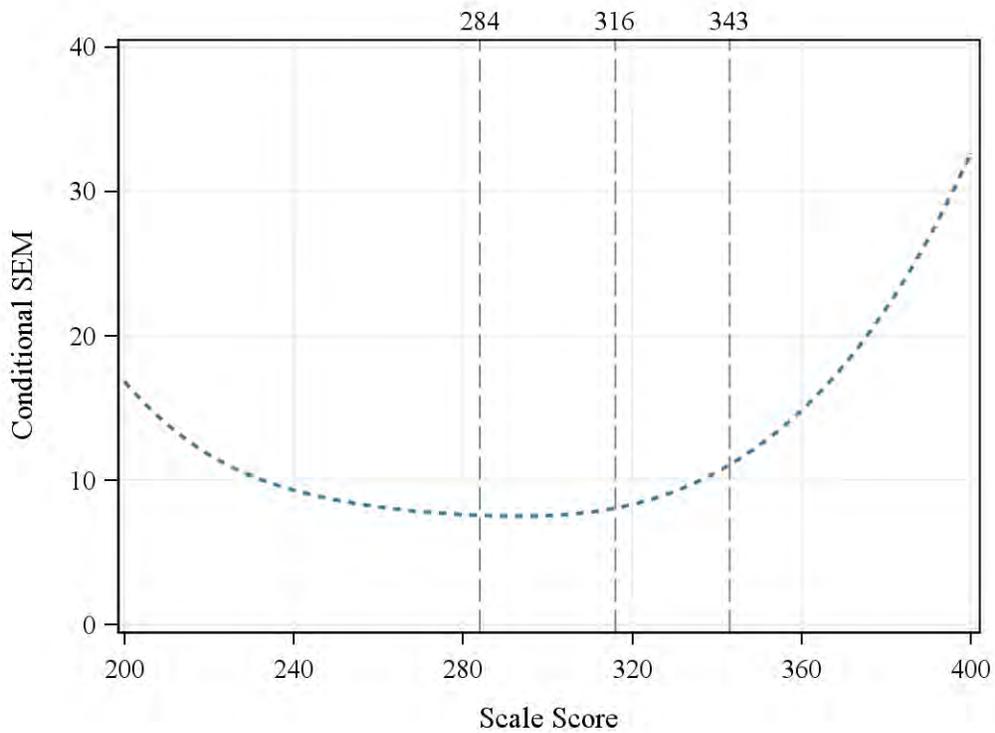
**Figure 6.9. ELA Grade 7 TCC**



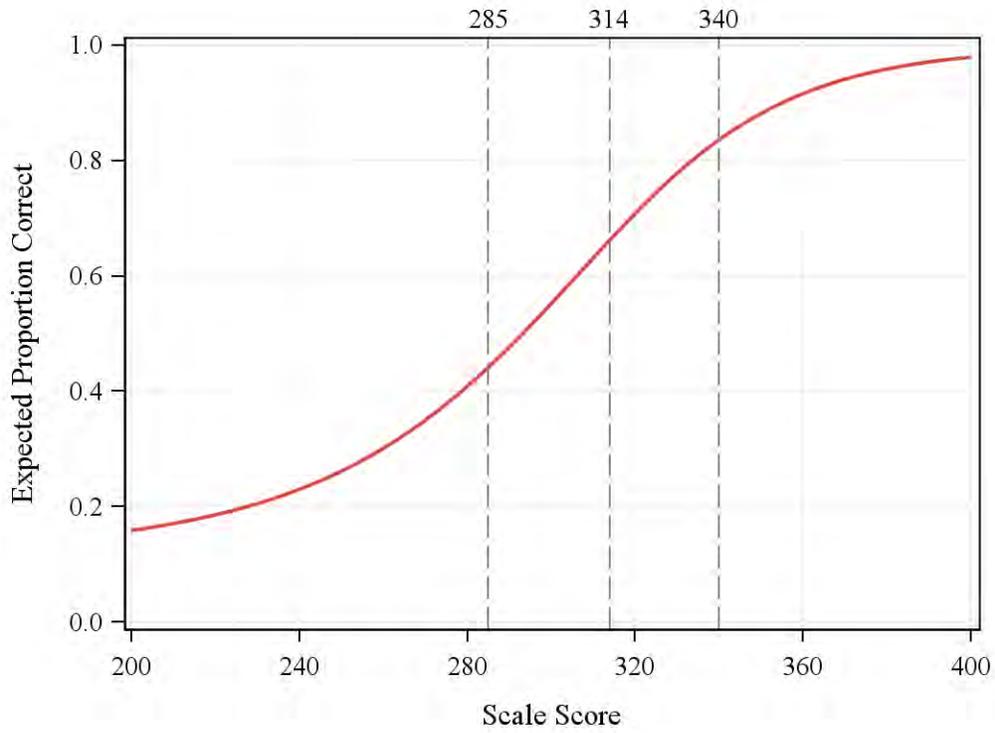
**Figure 6.10. ELA Grade 7 CSEM Curve**



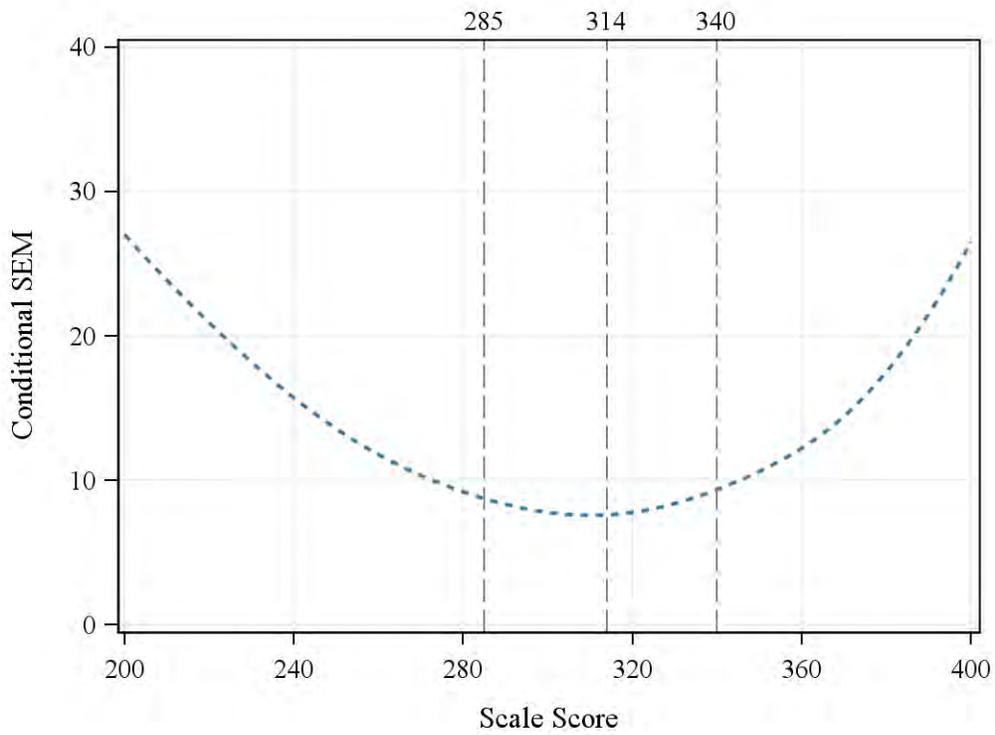
**Figure 6.11. ELA Grade 8 TCC**



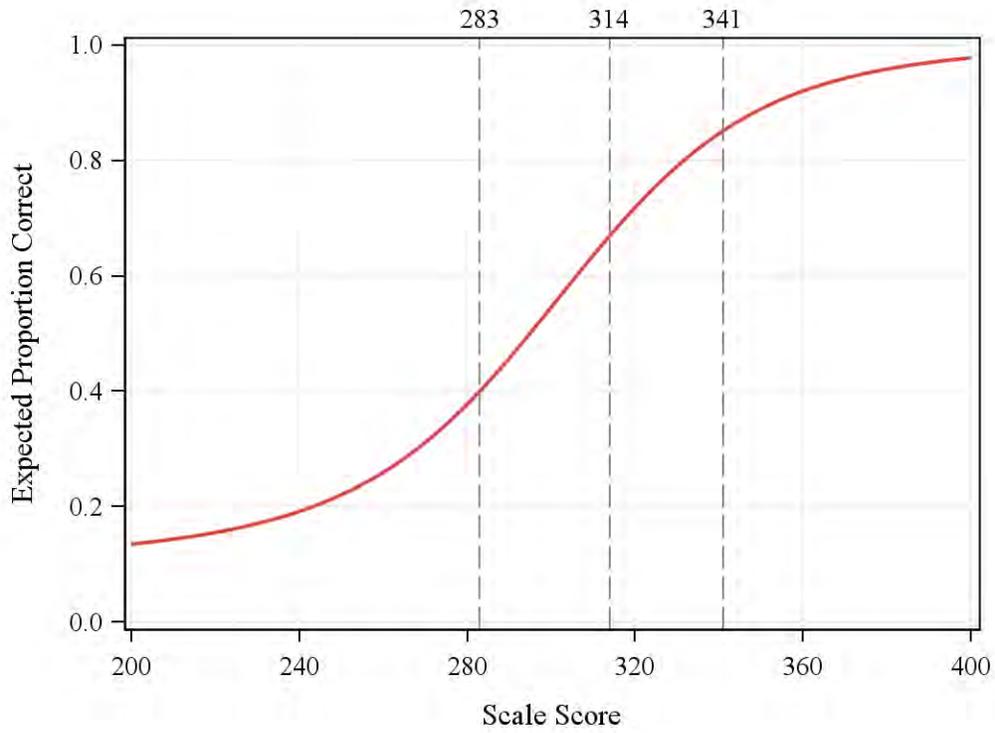
**Figure 6.12. ELA Grade 8 CSEM Curve**



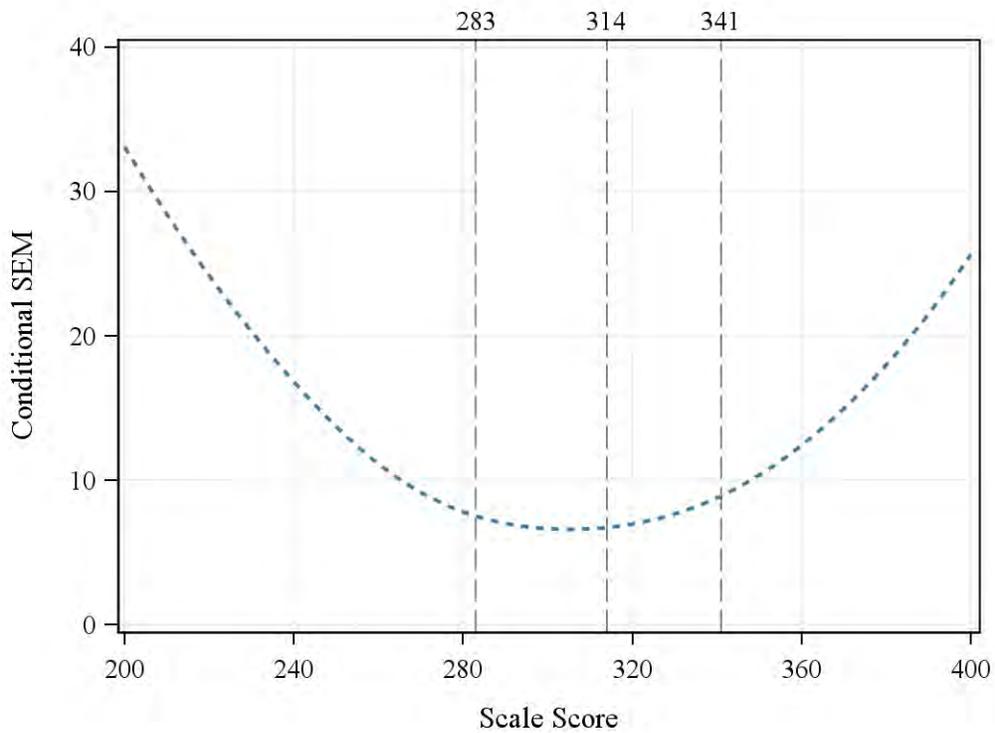
**Figure 6.13. Mathematics Grade 3 TCC**



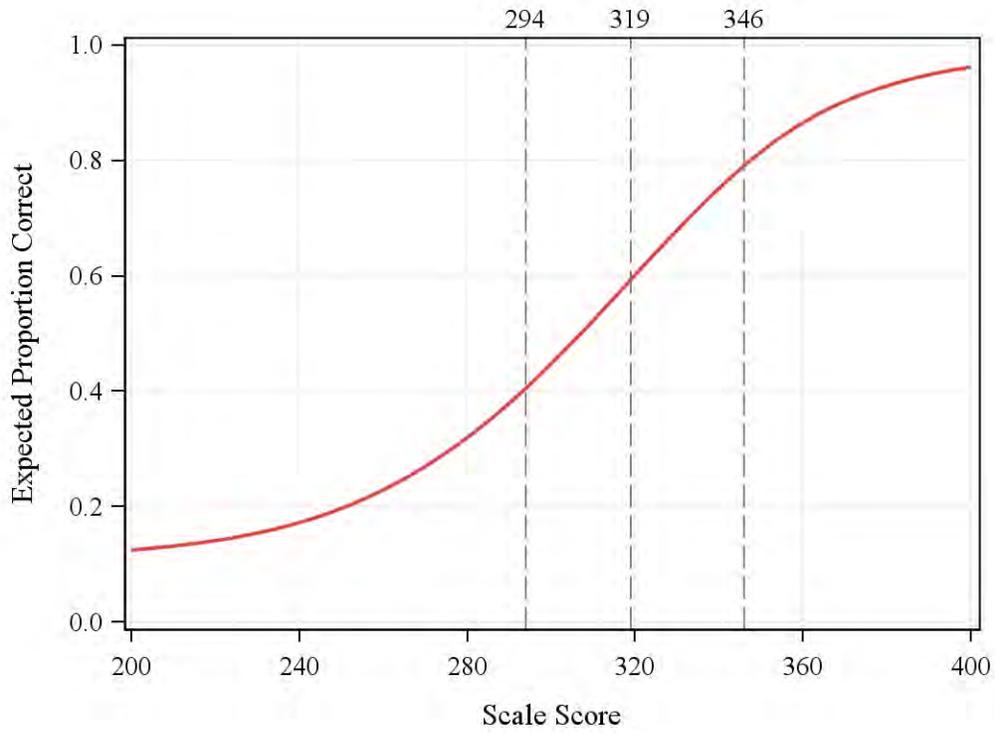
**Figure 6.14. Mathematics Grade 3 CSEM Curve**



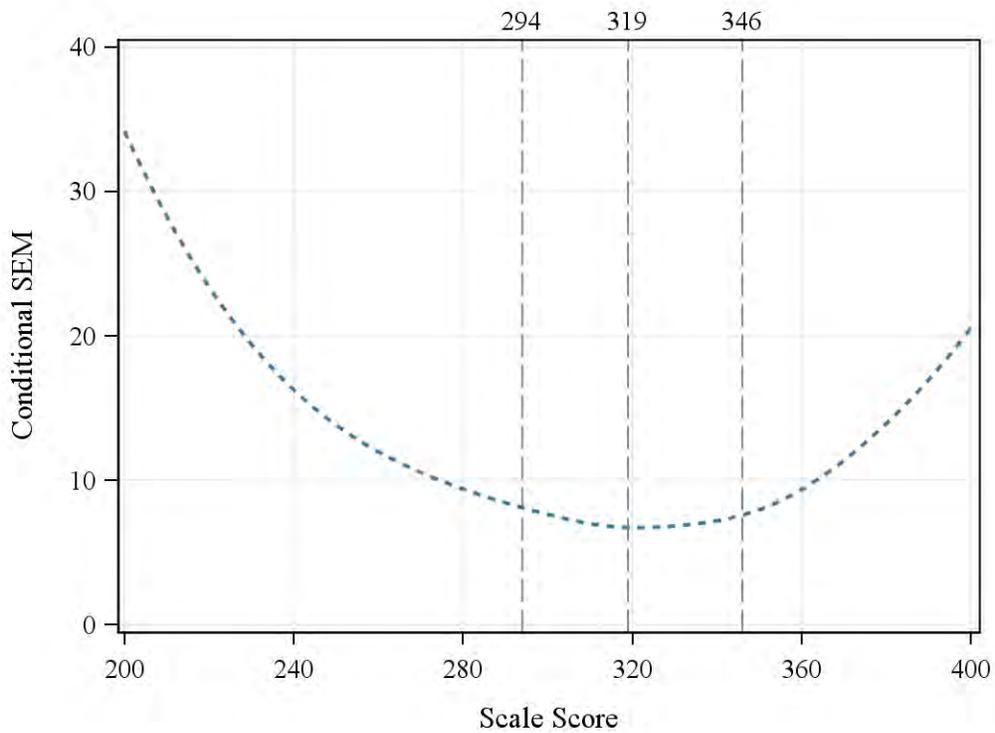
**Figure 6.15. Mathematics Grade 4 TCC**



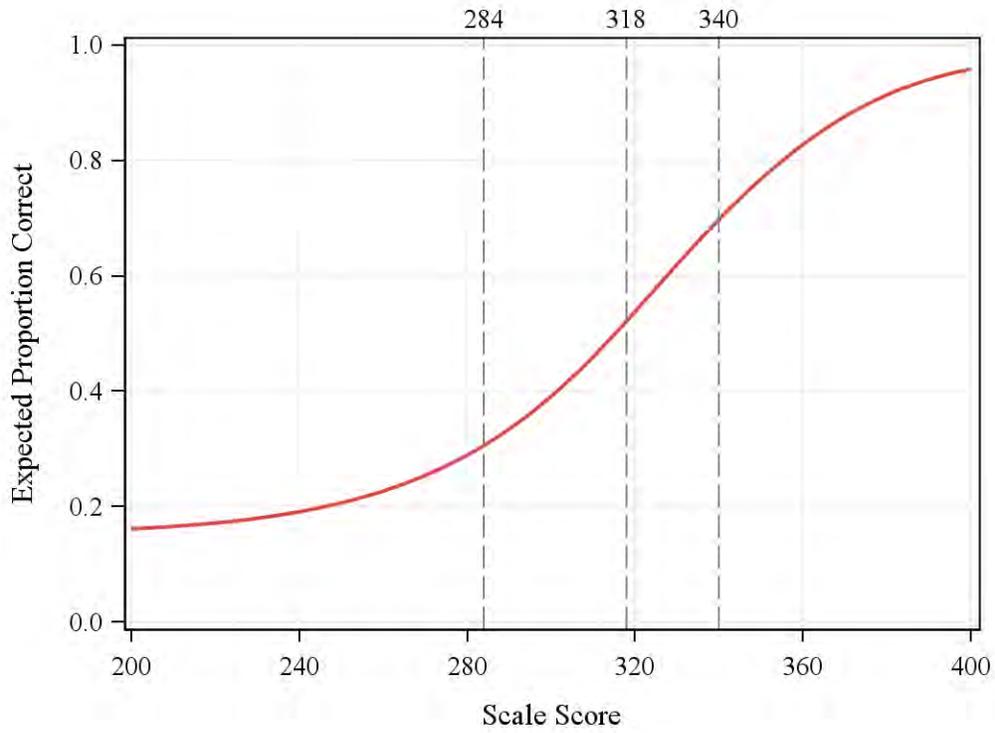
**Figure 6.16. Mathematics Grade 4 CSEM Curve**



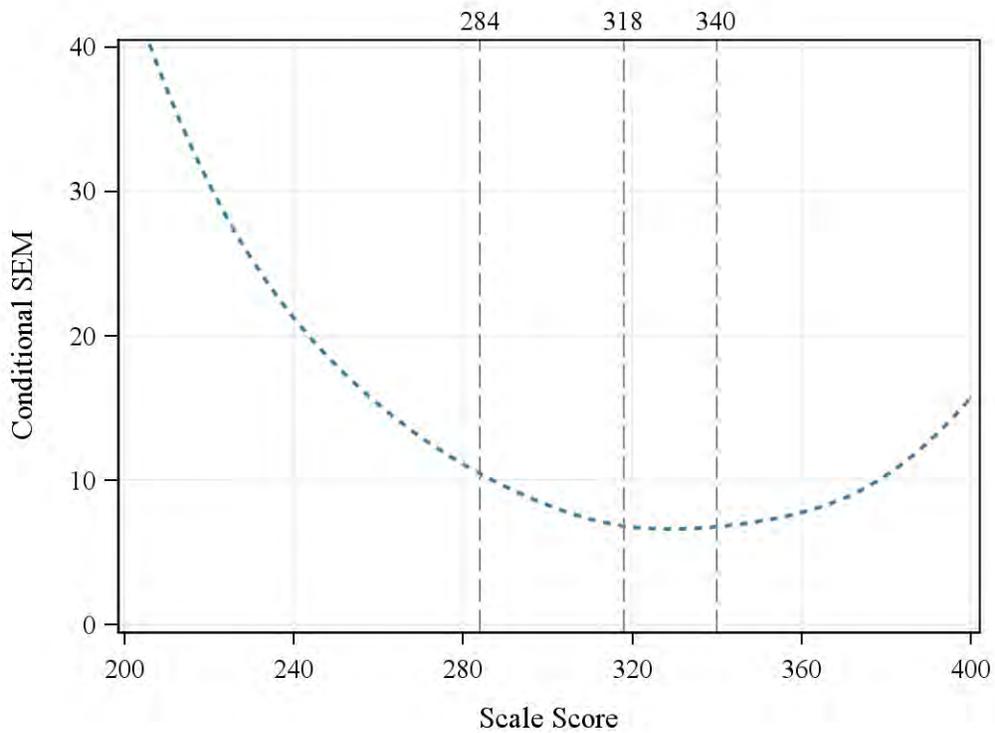
**Figure 6.17. Mathematics Grade 5 TCC**



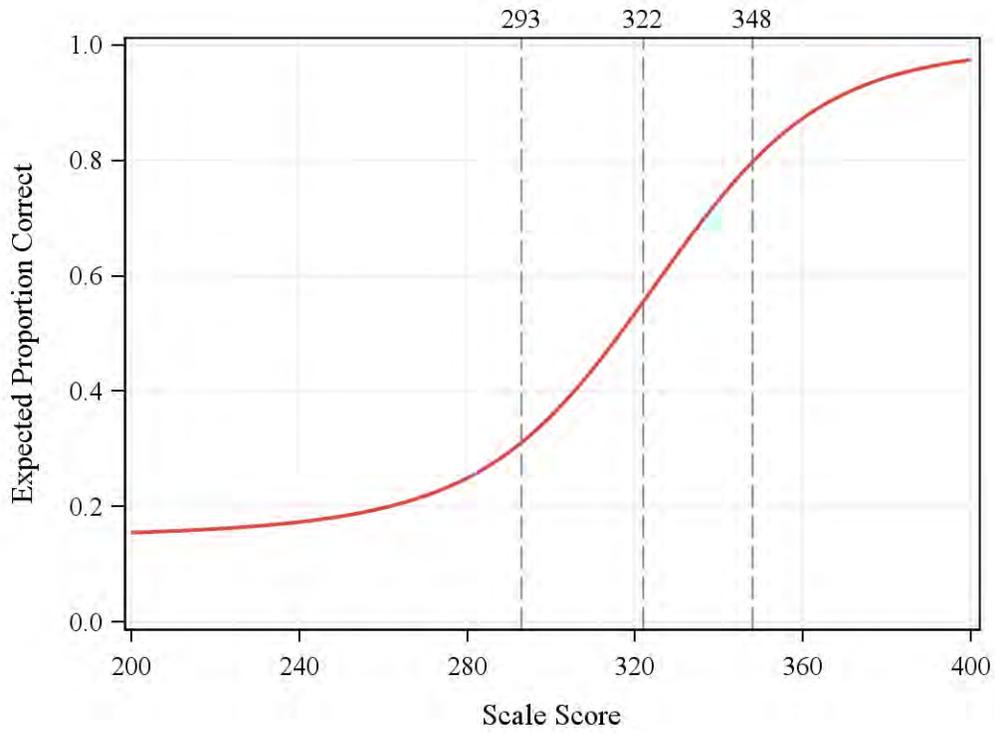
**Figure 6.18. Mathematics Grade 5 CSEM Curve**



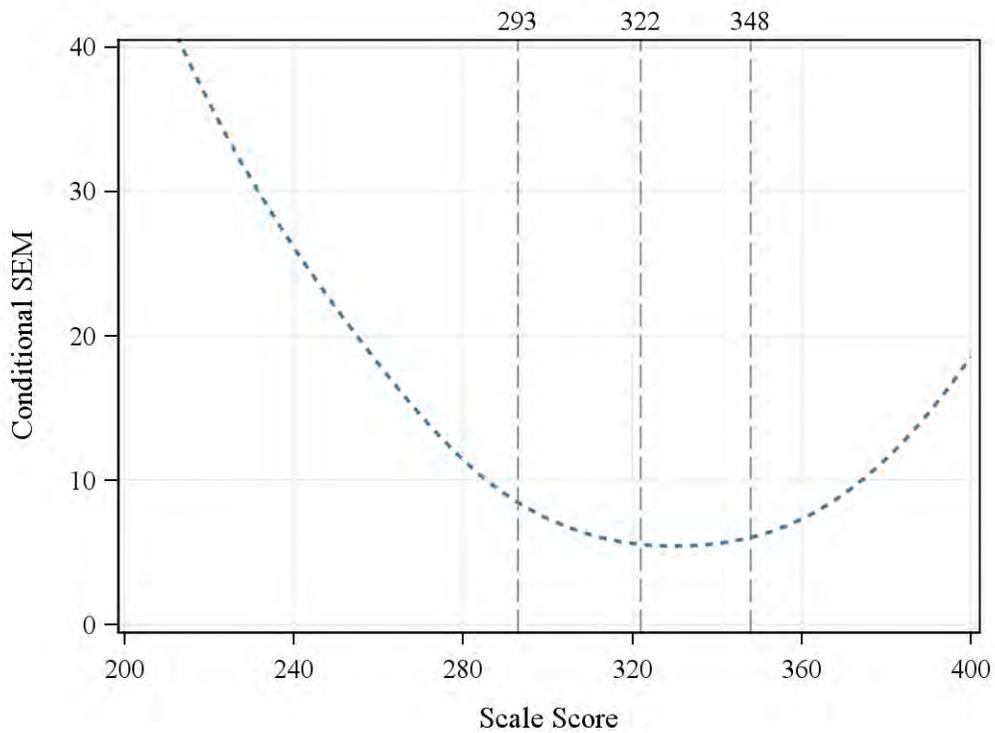
**Figure 6.19. Mathematics Grade 6 TCC**



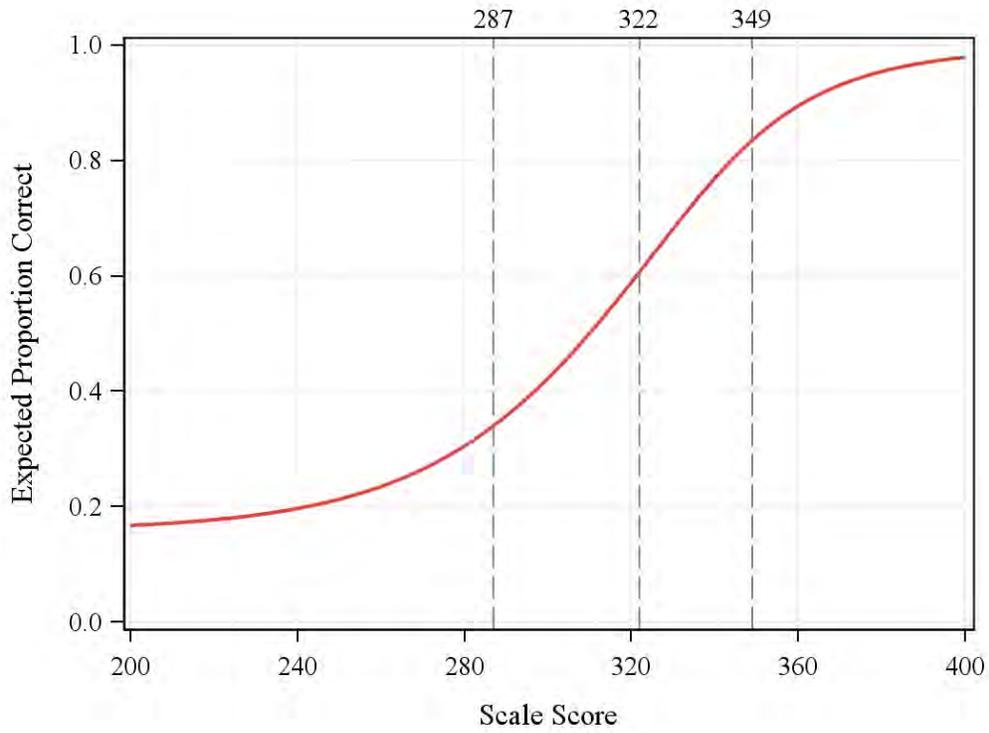
**Figure 6.20. Mathematics Grade 6 CSEM Curve**



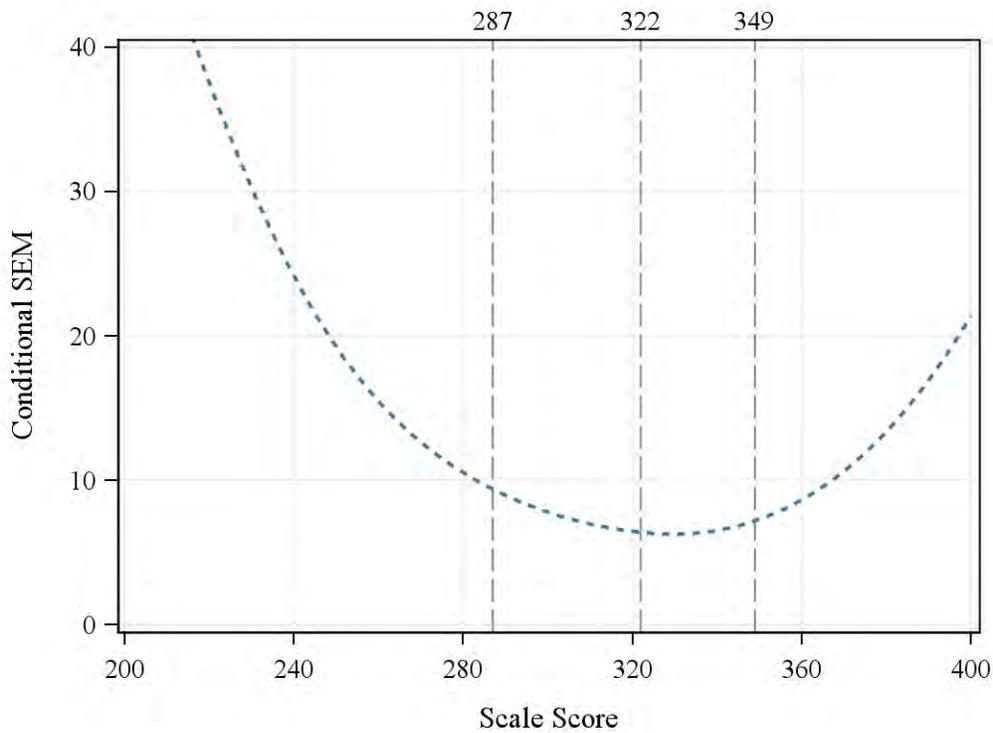
**Figure 6.21. Mathematics Grade 7 TCC**



**Figure 6.22. Mathematics Grade 7 CSEM Curve**



**Figure 6.23. Mathematics Grade 8 TCC**



**Figure 6.24. Mathematics Grade 8 CSEM Curve**

## 6.8. 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.

As described in Section 6.5: Linking and Scaling, 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 Tests.

An inverse TCC method was employed using POLYEQUATE (Kolen and 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 linking and 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.

### 6.8.1. 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).

$I(\theta)$  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.

The linking procedure described above does not guarantee that the same scale score scale points selected as performance-level cut scores will be observed. It was important to appropriately reflect the performance levels set by the standard setting panel and approved by the Commissioner in Summer 2013. To that end, if a given scale score cut was not observed in the 2016 RSSS table, the nearest, but lower, scale score value was rounded up to the established scale score cut. In this way, the approved scale score cuts set in 2013 were maintained for 2016.

Tables 6.13 and 6.14 for ELA and Mathematics, respectively, present the raw- and scale-score performance level cut scores.

**Table 6.13. ELA Performance-Level Cut Scores**

Performance Level	Raw Score Cut (Scale Score Cut)					
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level II	19 (291)	19 (287)	32 (289)	27 (283)	26 (287)	31 (284)
NYS Level III	28 (320)	29 (320)	41 (320)	39 (320)	39 (318)	44 (316)
NYS Level IV	39 (358)	36 (343)	48 (346)	45 (338)	47 (347)	51 (343)

**Table 6.14. Mathematics Performance-Level Cut Scores**

Performance Level	Raw Score Cut (Scale Score Cut)					
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
NYS Level II	24 (285)	25 (283)	24 (294)	20 (284)	21 (293)	23 (287)
NYS Level III	37 (314)	41 (314)	36 (319)	35 (318)	38 (322)	41 (322)
NYS Level IV	46 (340)	52 (341)	48 (346)	46 (340)	54 (348)	56 (349)

## 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 Common Core ELA and Mathematics Tests, we 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

Tables 7.1 and 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 Tables 7.2 and 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 0.89 to 0.93. Grades 3–8 Mathematics reliability estimates (Cronbach’s alpha and Feldt-Raju) ranged from 0.92 to 0.95. The reliabilities are similar across grades and slightly higher for the Mathematics tests than for the ELA tests. All reliabilities were at least .89 across all grades and both subjects, which is a good indication that the NYSTP Grades 3–8 Common Core ELA and Mathematics Tests are acceptably reliable.

**Table 7.1. ELA Test Form Statistics**

Grade	Item-level			Student-level			
	P-value			N-Count	Raw Score		
	Mean	Min.	Max.		Max.	Mean	SD
3	0.57	0.30	0.90	173,695	47	24.98	9.41
4	0.55	0.39	0.75	171,185	47	25.59	9.06
5	0.62	0.36	0.87	160,808	57	34.59	10.63
6	0.57	0.33	0.78	158,210	57	33.09	10.40
7	0.57	0.29	0.79	148,857	57	32.75	11.31
8	0.68	0.42	0.96	143,555	57	38.82	11.12

**Table 7.2. ELA Test Reliability and Standard Error of Measurement**

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	173,695	34	47	0.91	2.86	0.91	2.75
4	171,185	34	47	0.89	3.05	0.90	2.90
5	160,808	44	57	0.91	3.27	0.91	3.13
6	158,210	44	57	0.89	3.39	0.90	3.23
7	148,857	44	57	0.91	3.42	0.92	3.23
8	143,555	44	57	0.92	3.16	0.93	2.99

**Table 7.3. Mathematics Test Form Statistics**

Grade	Item-level			Student-level			
	P-value			N-Count	Raw Score		
	Mean	Min.	Max.		Max.	Mean	SD
3	0.63	0.24	0.90	178,870	56	33.51	12.63
4	0.61	0.23	0.83	174,321	62	36.26	15.41
5	0.56	0.20	0.86	162,992	61	31.72	13.82
6	0.51	0.12	0.85	161,216	67	31.75	14.70
7	0.49	0.28	0.80	147,252	68	31.83	16.39
8	0.49	0.19	0.83	115,190	68	30.14	15.04

**Table 7.4. Mathematics Test Reliability and Standard Error of Measurement**

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	178,870	45	56	0.92	3.51	0.93	3.28
4	174,321	48	62	0.95	3.60	0.95	3.38
5	162,992	47	61	0.93	3.54	0.94	3.38
6	161,216	53	67	0.94	3.74	0.94	3.53
7	147,252	54	68	0.94	3.86	0.95	3.63
8	115,190	54	68	0.93	3.91	0.94	3.70

**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. Tables 7.5 and 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		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	173,695	25	0.85	2.11	0.85	2.11
4	171,185	25	0.79	2.28	0.79	2.27
5	160,808	35	0.85	2.56	0.85	2.55
6	158,210	35	0.81	2.69	0.82	2.69
7	148,857	35	0.84	2.67	0.84	2.66
8	143,555	35	0.87	2.47	0.87	2.46

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

Grade	N-Count	Items	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	178,870	37	0.91	2.43	0.91	2.40
4	174,321	38	0.93	2.49	0.93	2.48
5	162,992	37	0.91	2.55	0.91	2.53
6	161,216	43	0.90	2.79	0.90	2.77
7	147,252	44	0.92	2.89	0.92	2.88
8	115,190	44	0.90	2.94	0.90	2.93

### 7.1.3. Reliability of CR Items

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

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

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	173,695	9	22	0.87	1.70	0.88	1.66
4	171,185	9	22	0.87	1.77	0.88	1.69
5	160,808	9	22	0.87	1.76	0.88	1.69
6	158,210	9	22	0.88	1.74	0.89	1.65
7	148,857	9	22	0.90	1.77	0.91	1.68
8	143,555	9	22	0.89	1.66	0.90	1.55

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**

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	178,870	8	19	0.81	2.31	0.82	2.20
4	174,321	10	24	0.87	2.38	0.88	2.28
5	162,992	10	24	0.85	2.28	0.86	2.23
6	161,216	10	24	0.87	2.23	0.88	2.17
7	147,252	10	24	0.89	2.23	0.90	2.17
8	115,190	10	24	0.88	2.26	0.88	2.22

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

### 7.1.4. Test Reliability for Reporting Categories

In this section, reliability coefficients that were estimated for the population and subgroups are presented. The reporting categories include the following: gender, ethnicity, NRC, ELL, all SWD, all SUA, students with disabilities using accommodations falling under 504 Plan (SWD/SUA), and English language learners using accommodations specific to their ELL status (ELL/SUA). Accommodations available to students under the 504 Plan 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, Third Reading of Listening Selection, 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 .79. Feldt-Raju reliability coefficients, which tend to be larger than the Cronbach's alpha estimates for the same group, were at least .80 each. 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		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	173,695	0.91	2.87	0.91	2.75
Gender	Female	86,132	0.90	2.87	0.91	2.75
	Male	87,563	0.91	2.86	0.92	2.74
Ethnicity	Asian	17,910	0.90	2.79	0.91	2.67
	Black	31,562	0.90	2.92	0.91	2.79
	Hispanic	49,379	0.89	2.89	0.90	2.79
	American Indian	1,204	0.89	2.90	0.90	2.77
	Multiracial	4,343	0.91	2.84	0.92	2.70
	Pacific Islander	548	0.89	2.86	0.90	2.76
	White	68,749	0.90	2.85	0.91	2.71
NRC	New York	70,267	0.91	2.87	0.91	2.75
	Big 4 Cities	7,489	0.90	2.87	0.91	2.76
	Urban/Suburban	13,771	0.89	2.88	0.90	2.78
	Rural	9,539	0.90	2.86	0.91	2.76
	Average Needs	39,596	0.90	2.84	0.91	2.73
	Low Needs	17,480	0.88	2.73	0.89	2.62
	Charter School	9,645	0.89	2.88	0.90	2.78
	Non-Public	5,908	0.91	3.01	0.92	2.82
SWD	All Codes	25,125	0.88	2.83	0.89	2.74
SUA	All Codes	24,015	0.88	2.83	0.89	2.75
ELL	ELL=Y	16,574	0.84	2.89	0.85	2.79
SWD/SUA	SUA=504 plan codes	21,150	0.87	2.82	0.88	2.74
ELL/SUA	SUA & ELL codes	3,703	0.80	2.79	0.81	2.72

**Table 7.10. ELA Grade 4 Test Reliability by Subgroup**

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	171,185	0.89	3.06	0.90	2.90
Gender	Female	84,532	0.88	3.04	0.89	2.89
	Male	86,653	0.89	3.06	0.90	2.90
Ethnicity	Asian	17,504	0.88	2.95	0.89	2.81
	Black	31,862	0.88	3.08	0.89	2.93
	Hispanic	47,741	0.87	3.05	0.88	2.91
	American Indian	1,091	0.88	3.08	0.89	2.91
	Multiracial	3,689	0.89	3.05	0.91	2.87
	Pacific Islander	627	0.88	3.04	0.89	2.87
	White	68,671	0.88	3.06	0.90	2.89
NRC	New York	68,816	0.89	3.02	0.90	2.86
	Big 4 Cities	7,249	0.88	3.05	0.89	2.89
	Urban/Suburban	13,092	0.87	3.07	0.88	2.93
	Rural	9,061	0.88	3.06	0.89	2.92
	Average Needs	37,617	0.88	3.05	0.89	2.90
	Low Needs	16,928	0.85	2.97	0.87	2.85
	Charter School	8,189	0.86	3.03	0.87	2.94
	Non-Public	10,233	0.88	3.23	0.90	2.99
SWD	All Codes	26,119	0.86	2.99	0.87	2.85
SUA	All Codes	26,888	0.86	2.99	0.87	2.87
ELL	ELL=Y	14,886	0.81	3.02	0.83	2.89
SWD/SUA	SUA=504 plan codes	22,933	0.85	2.97	0.86	2.85
ELL/SUA	SUA & ELL codes	3,724	0.77	2.90	0.79	2.79

**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	160,808	0.90	3.29	0.91	3.13
Gender	Female	79,090	0.90	3.24	0.91	3.09
	Male	81,718	0.91	3.30	0.92	3.16
Ethnicity	Asian	16,724	0.90	3.12	0.91	2.98
	Black	30,617	0.90	3.36	0.91	3.21
	Hispanic	44,779	0.89	3.33	0.90	3.19
	American Indian	1,069	0.90	3.34	0.91	3.17
	Multiracial	2,948	0.91	3.25	0.92	3.07
	Pacific Islander	450	0.89	3.23	0.90	3.10
	White	64,221	0.90	3.24	0.91	3.07
NRC	New York	66,871	0.90	3.26	0.91	3.12
	Big 4 Cities	6,465	0.91	3.37	0.92	3.22
	Urban/Suburban	12,182	0.90	3.35	0.90	3.22
	Rural	8,489	0.90	3.33	0.91	3.18
	Average Needs	35,820	0.90	3.24	0.91	3.10
	Low Needs	16,833	0.87	3.10	0.88	2.98
	Charter School	8,373	0.88	3.27	0.89	3.16
	Non-Public	5,775	0.91	3.46	0.93	3.20
SWD	All Codes	26,701	0.88	3.37	0.89	3.24
SUA	All Codes	27,379	0.89	3.36	0.90	3.24
ELL	ELL=Y	12,013	0.84	3.40	0.86	3.26
SWD/SUA	SUA=504 plan codes	23,570	0.88	3.37	0.89	3.24
ELL/SUA	SUA & ELL codes	3,388	0.81	3.32	0.82	3.21

**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	158,210	0.89	3.40	0.90	3.23
Gender	Female	77,772	0.88	3.34	0.89	3.20
	Male	80,438	0.90	3.43	0.91	3.24
Ethnicity	Asian	17,183	0.89	3.19	0.90	3.06
	Black	30,271	0.88	3.45	0.89	3.28
	Hispanic	42,276	0.88	3.44	0.89	3.28
	American Indian	1,061	0.88	3.43	0.89	3.27
	Multiracial	2,513	0.91	3.36	0.92	3.17
	Pacific Islander	425	0.88	3.31	0.89	3.18
	White	64,481	0.89	3.38	0.90	3.20
NRC	New York	63,195	0.90	3.35	0.91	3.19
	Big 4 Cities	6,393	0.89	3.53	0.90	3.32
	Urban/Suburban	10,898	0.89	3.49	0.90	3.30
	Rural	8,184	0.88	3.47	0.90	3.28
	Average Needs	34,109	0.89	3.39	0.90	3.23
	Low Needs	17,046	0.86	3.23	0.87	3.12
	Charter School	9,189	0.86	3.36	0.87	3.27
	Non-Public	9,196	0.89	3.58	0.91	3.29
SWD	All Codes	25,592	0.86	3.45	0.87	3.29
SUA	All Codes	26,012	0.87	3.46	0.88	3.29
ELL	ELL=Y	11,750	0.82	3.49	0.84	3.30
SWD/SUA	SUA=504 plan codes	22,171	0.85	3.45	0.86	3.29
ELL/SUA	SUA & ELL codes	3,359	0.76	3.39	0.78	3.25

**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	148,857	0.91	3.43	0.92	3.23
Gender	Female	72,555	0.90	3.36	0.91	3.19
	Male	76,302	0.91	3.44	0.92	3.24
Ethnicity	Asian	16,249	0.90	3.23	0.91	3.06
	Black	29,565	0.89	3.48	0.91	3.29
	Hispanic	40,195	0.90	3.45	0.91	3.27
	American Indian	1,098	0.90	3.43	0.91	3.25
	Multiracial	2,036	0.92	3.43	0.93	3.18
	Pacific Islander	418	0.91	3.36	0.92	3.18
	White	59,296	0.91	3.41	0.92	3.19
NRC	New York	63,853	0.91	3.36	0.92	3.17
	Big 4 Cities	5,892	0.90	3.49	0.91	3.29
	Urban/Suburban	10,263	0.90	3.51	0.91	3.31
	Rural	7,777	0.91	3.50	0.92	3.28
	Average Needs	31,388	0.91	3.44	0.92	3.23
	Low Needs	16,503	0.88	3.30	0.89	3.15
	Charter School	8,180	0.87	3.39	0.88	3.29
	Non-Public	5,001	0.92	3.60	0.93	3.28
SWD	All Codes	24,134	0.87	3.41	0.88	3.26
SUA	All Codes	23,996	0.88	3.42	0.89	3.27
ELL	ELL=Y	10,342	0.81	3.39	0.83	3.24
SWD/SUA	SUA=504 plan codes	20,811	0.86	3.41	0.88	3.26
ELL/SUA	SUA & ELL codes	2,750	0.76	3.30	0.77	3.19

**Table 7.14. ELA Grade 8 Test Reliability by Subgroup**

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	143,555	0.92	3.17	0.93	2.98
Gender	Female	69,999	0.91	3.07	0.92	2.91
	Male	73,556	0.92	3.23	0.93	3.03
Ethnicity	Asian	16,027	0.91	2.87	0.92	2.72
	Black	30,083	0.91	3.28	0.92	3.10
	Hispanic	39,239	0.91	3.24	0.92	3.07
	American Indian	920	0.91	3.25	0.92	3.07
	Multiracial	1,599	0.93	3.17	0.94	2.94
	Pacific Islander	374	0.90	3.08	0.92	2.89
	White	55,313	0.92	3.10	0.93	2.90
NRC	New York	63,737	0.91	3.13	0.92	2.96
	Big 4 Cities	5,721	0.92	3.42	0.93	3.21
	Urban/Suburban	9,184	0.92	3.33	0.92	3.14
	Rural	7,307	0.92	3.26	0.93	3.07
	Average Needs	28,192	0.92	3.16	0.93	2.97
	Low Needs	14,983	0.90	2.87	0.91	2.73
	Charter School	6,816	0.88	3.08	0.89	2.98
	Non-Public	7,615	0.92	3.34	0.94	3.03
SWD	All Codes	22,459	0.89	3.38	0.90	3.23
SUA	All Codes	22,559	0.90	3.37	0.91	3.22
ELL	ELL=Y	10,095	0.86	3.43	0.88	3.26
SWD/SUA	SUA=504 plan codes	19,319	0.89	3.38	0.90	3.23
ELL/SUA	SUA & ELL codes	2,554	0.83	3.36	0.84	3.24

**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	178,870	0.92	3.51	0.93	3.28
Gender	Female	88,423	0.92	3.51	0.93	3.28
	Male	90,447	0.93	3.52	0.94	3.28
Ethnicity	Asian	18,673	0.92	3.30	0.93	3.03
	Black	32,281	0.92	3.49	0.93	3.32
	Hispanic	51,194	0.91	3.52	0.92	3.34
	American Indian	1,244	0.92	3.52	0.92	3.33
	Multiracial	4,341	0.93	3.50	0.94	3.24
	Pacific Islander	578	0.92	3.42	0.93	3.19
	White	70,559	0.91	3.50	0.93	3.26
NRC	New York	71,888	0.92	3.49	0.93	3.26
	Big 4 Cities	7,798	0.92	3.45	0.93	3.30
	Urban/Suburban	13,776	0.91	3.53	0.92	3.36
	Rural	9,429	0.92	3.56	0.93	3.35
	Average Needs	39,072	0.91	3.53	0.92	3.29
	Low Needs	17,440	0.90	3.37	0.92	3.13
	Charter School	9,565	0.92	3.38	0.93	3.13
	Non-Public	9,902	0.91	3.61	0.92	3.40
SWD	All Codes	25,933	0.91	3.43	0.92	3.31
SUA	All Codes	24,665	0.91	3.44	0.91	3.32
ELL	ELL=Y	18,590	0.90	3.43	0.91	3.32
SWD/SUA	SUA=504 plan codes	21,837	0.90	3.42	0.91	3.31
ELL/SUA	SUA & ELL codes	3,805	0.89	3.32	0.89	3.25
ELL Test Language	English	174,967	0.92	3.51	0.93	3.28
	Chinese	671	0.90	3.38	0.91	3.14
	Haitian-Creole	62	0.89	3.37	0.90	3.23
	Korean	30	0.90	3.21	0.91	2.95
	Russian	86	0.92	3.44	0.93	3.27
	Spanish	3,054	0.90	3.38	0.90	3.28
	All Translations	3,903	0.92	3.44	0.93	3.28

**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	174,321	0.95	3.61	0.95	3.38
Gender	Female	85,869	0.94	3.62	0.95	3.40
	Male	88,452	0.95	3.59	0.95	3.36
Ethnicity	Asian	18,124	0.94	3.31	0.95	3.06
	Black	32,575	0.94	3.64	0.95	3.45
	Hispanic	49,396	0.94	3.65	0.94	3.46
	American Indian	1,114	0.94	3.64	0.95	3.41
	Multiracial	3,693	0.95	3.57	0.95	3.33
	Pacific Islander	656	0.94	3.56	0.95	3.33
	White	68,763	0.94	3.56	0.94	3.35
NRC	New York	70,160	0.95	3.60	0.95	3.36
	Big 4 Cities	7,329	0.94	3.57	0.95	3.39
	Urban/Suburban	12,913	0.94	3.63	0.94	3.45
	Rural	8,920	0.94	3.65	0.94	3.45
	Average Needs	37,102	0.94	3.60	0.94	3.39
	Low Needs	17,038	0.93	3.38	0.93	3.19
	Charter School	8,453	0.94	3.53	0.95	3.29
	Non-Public	12,406	0.93	3.72	0.94	3.52
SWD	All Codes	26,588	0.93	3.52	0.94	3.37
SUA	All Codes	27,045	0.93	3.55	0.94	3.40
ELL	ELL=Y	16,309	0.93	3.54	0.93	3.40
SWD/SUA	SUA=504 plan codes	23,246	0.93	3.51	0.93	3.37
ELL/SUA	SUA & ELL codes	3,782	0.90	3.39	0.91	3.31
ELL Test Language	English	170,566	0.94	3.61	0.95	3.38
	Chinese	596	0.93	3.51	0.94	3.27
	Haitian-Creole	70	0.90	3.33	0.90	3.27
	Korean	28	0.92	3.22	0.93	2.97
	Russian	107	0.93	3.71	0.94	3.52
	Spanish	2,954	0.92	3.47	0.93	3.36
	All Translations	3,755	0.94	3.54	0.95	3.36

**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	162,992	0.93	3.54	0.94	3.38
Gender	Female	79,609	0.93	3.54	0.94	3.39
	Male	83,383	0.94	3.55	0.94	3.37
Ethnicity	Asian	17,389	0.93	3.46	0.94	3.22
	Black	31,457	0.92	3.47	0.93	3.36
	Hispanic	46,546	0.92	3.47	0.92	3.39
	American Indian	1,111	0.93	3.51	0.94	3.36
	Multiracial	3,027	0.94	3.56	0.95	3.36
	Pacific Islander	484	0.93	3.55	0.94	3.37
	White	62,978	0.93	3.58	0.94	3.40
NRC	New York	68,243	0.94	3.51	0.94	3.35
	Big 4 Cities	6,683	0.93	3.39	0.93	3.27
	Urban/Suburban	11,954	0.92	3.46	0.93	3.37
	Rural	8,188	0.92	3.57	0.93	3.43
	Average Needs	34,960	0.93	3.58	0.93	3.42
	Low Needs	16,695	0.92	3.53	0.92	3.34
	Charter School	9,051	0.93	3.51	0.93	3.36
	Non-Public	7,218	0.92	3.62	0.93	3.46
SWD	All Codes	26,976	0.91	3.37	0.91	3.28
SUA	All Codes	27,433	0.91	3.39	0.92	3.29
ELL	ELL=Y	13,399	0.90	3.36	0.90	3.31
SWD/SUA	SUA=504 plan codes	23,802	0.90	3.36	0.91	3.27
ELL/SUA	SUA & ELL codes	3,408	0.86	3.23	0.86	3.18
ELL Test Language	English	159,330	0.93	3.55	0.94	3.38
	Chinese	542	0.92	3.61	0.93	3.38
	Haitian-Creole	58	0.81	3.21	0.82	3.12
	Korean	30	0.94	3.49	0.96	3.10
	Russian	76	0.92	3.62	0.93	3.37
	Spanish	2,956	0.87	3.29	0.88	3.20
	All Translations	3,662	0.92	3.23	0.92	3.34

**Table 7.18. Mathematics Grade 6 Test Reliability by Subgroup**

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	161,216	0.94	3.75	0.94	3.54
Gender	Female	79,050	0.93	3.76	0.94	3.55
	Male	82,166	0.94	3.73	0.94	3.52
Ethnicity	Asian	17,833	0.94	3.67	0.95	3.41
	Black	31,008	0.92	3.63	0.92	3.48
	Hispanic	43,781	0.91	3.68	0.92	3.53
	American Indian	1,077	0.92	3.70	0.93	3.54
	Multiracial	2,513	0.94	3.74	0.95	3.50
	Pacific Islander	455	0.93	3.76	0.94	3.55
	White	64,549	0.93	3.78	0.94	3.58
NRC	New York	64,335	0.94	3.73	0.95	3.49
	Big 4 Cities	6,440	0.91	3.48	0.92	3.36
	Urban/Suburban	10,412	0.91	3.60	0.92	3.47
	Rural	7,757	0.91	3.73	0.92	3.58
	Average Needs	33,015	0.93	3.77	0.93	3.59
	Low Needs	16,735	0.92	3.74	0.93	3.54
	Charter School	9,825	0.93	3.74	0.93	3.55
	Non-Public	12,697	0.92	3.76	0.93	3.60
SWD	All Codes	25,399	0.88	3.46	0.89	3.37
SUA	All Codes	25,399	0.89	3.49	0.90	3.40
ELL	ELL=Y	13,370	0.88	3.48	0.89	3.38
SWD/SUA	SUA=504 plan codes	21,808	0.87	3.44	0.88	3.37
ELL/SUA	SUA & ELL codes	3,163	0.77	3.32	0.78	3.27
ELL Test Language	English	156,840	0.93	3.75	0.94	3.54
	Chinese	836	0.92	3.81	0.93	3.58
	Haitian-Creole	59	0.87	3.42	0.88	3.33
	Korean	32	0.94	3.73	0.95	3.39
	Russian	122	0.94	3.68	0.94	3.45
	Spanish	3,327	0.81	3.37	0.82	3.32
	All Translations	4,376	0.92	3.55	0.93	3.40

**Table 7.19. Mathematics Grade 7 Test Reliability by Subgroup**

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	147,252	0.94	3.87	0.95	3.63
Gender	Female	71,650	0.94	3.88	0.95	3.64
	Male	75,602	0.95	3.84	0.95	3.61
Ethnicity	Asian	16,614	0.95	3.66	0.96	3.43
	Black	29,690	0.93	3.76	0.93	3.59
	Hispanic	41,116	0.93	3.83	0.93	3.64
	American Indian	1,087	0.93	3.83	0.94	3.64
	Multiracial	1,942	0.95	3.86	0.96	3.61
	Pacific Islander	432	0.95	3.86	0.95	3.62
	White	56,371	0.94	3.89	0.95	3.68
NRC	New York	64,686	0.95	3.82	0.96	3.57
	Big 4 Cities	5,826	0.91	3.63	0.92	3.48
	Urban/Suburban	9,475	0.91	3.76	0.92	3.60
	Rural	7,140	0.92	3.90	0.93	3.72
	Average Needs	28,987	0.93	3.93	0.94	3.72
	Low Needs	15,649	0.93	3.81	0.94	3.64
	Charter School	8,474	0.94	3.83	0.95	3.64
	Non-Public	7,015	0.93	3.92	0.94	3.72
SWD	All Codes	23,429	0.89	3.54	0.89	3.44
SUA	All Codes	22,893	0.90	3.58	0.91	3.47
ELL	ELL=Y	11,285	0.89	3.52	0.90	3.43
SWD/SUA	SUA=504 plan codes	19,956	0.88	3.52	0.88	3.43
ELL/SUA	SUA & ELL codes	2,520	0.77	3.32	0.77	3.29
ELL Test Language	English	143,169	0.94	3.87	0.95	3.64
	Chinese	814	0.94	3.79	0.95	3.59
	Haitian-Creole	55	0.64	3.13	0.65	3.12
	Korean	25	0.94	3.53	0.94	3.34
	Russian	88	0.89	3.86	0.90	3.73
	Spanish	3,101	0.83	3.43	0.83	3.38
	All Translations	4,083	0.94	3.60	0.94	3.44

**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	115,190	0.93	3.94	0.94	3.68
Gender	Female	55,286	0.93	3.97	0.94	3.70
	Male	59,904	0.93	3.91	0.94	3.66
Ethnicity	Asian	11,147	0.94	3.99	0.95	3.57
	Black	26,458	0.92	3.78	0.93	3.60
	Hispanic	35,547	0.92	3.85	0.93	3.65
	American Indian	761	0.92	3.83	0.93	3.62
	Multiracial	1,184	0.93	3.93	0.94	3.68
	Pacific Islander	315	0.94	4.02	0.95	3.64
	White	39,778	0.92	4.03	0.93	3.77
NRC	New York	53,996	0.94	3.92	0.95	3.63
	Big 4 Cities	5,128	0.91	3.56	0.92	3.42
	Urban/Suburban	7,474	0.89	3.69	0.89	3.57
	Rural	5,520	0.90	3.88	0.91	3.71
	Average Needs	18,111	0.90	3.99	0.91	3.79
	Low Needs	8,222	0.92	4.06	0.93	3.78
	Charter School	5,926	0.94	3.96	0.95	3.66
	Non-Public	10,813	0.93	4.03	0.94	3.76
SWD	All Codes	20,663	0.88	3.50	0.89	3.42
SUA	All Codes	20,360	0.89	3.54	0.90	3.45
ELL	ELL=Y	11,447	0.91	3.56	0.91	3.45
SWD/SUA	SUA=504 plan codes	17,652	0.88	3.48	0.88	3.40
ELL/SUA	SUA & ELL codes	2,449	0.82	3.32	0.83	3.27
ELL Test Language	English	111,234	0.93	3.95	0.94	3.69
	Chinese	743	0.93	3.99	0.94	3.60
	Haitian-Creole	48	0.75	3.57	0.75	3.52
	Korean	23	0.92	3.97	0.94	3.52
	Russian	122	0.93	3.86	0.94	3.65
	Spanish	3,020	0.87	3.47	0.87	3.43
	All Translations	3,956	0.94	3.70	0.94	3.49

**7.2. Standard Error of Measurement (SEM)**

Tables 7.2 and 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 2.75 to 3.91 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 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 2.62 to 4.06, which is acceptably close to those for the entire population. This narrow range indicates that across the Grades 3–8 Common Core 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 Common Core 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 most likely to be 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 results based solely on the Level III cut. Table 7.21 and 7.22 include case counts (n-count), classification consistency (Agreement), classification inconsistency (Inconsistency), and Cohen’s kappa (Kappa). Consistency indicates the rate 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, 69–75% of students were estimated to be classified consistently to one of the four performance categories with a hypothetical second administration. Kappa—which corrects for chance agreement—ranged from 0.56 to 0.63. These are between “moderate” and “substantial” agreement, as per Landis and Koch’s (1977) rules of thumb for kappa. For Mathematics, 74–79% of students were estimated to be classified consistently to one of the four performance categories, and kappa ranged from 0.64 to 0.70. 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
<b>ELA</b>				
3	173,695	75%	25%	0.63
4	171,185	71%	29%	0.56
5	160,808	70%	30%	0.58
6	158,210	69%	31%	0.56
7	148,857	73%	27%	0.61
8	143,555	73%	27%	0.61
<b>Mathematics</b>				
3	178,870	75%	25%	0.65
4	174,321	78%	22%	0.70
5	162,992	78%	22%	0.68
6	161,216	74%	26%	0.64
7	147,252	79%	21%	0.70
8	115,190	79%	21%	0.69

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, 92–98% of the classifications of individual students were estimated to remain stable with a second administration. Kappa coefficients for ELA classification consistency ranged from 0.64 to 0.71. These are considered “substantial” agreement, as per Landis and Koch’s (1977) rules of thumb for kappa. For Mathematics, 94–97% of the classifications were estimated consistently, and kappa coefficients ranged from 0.77 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
<b>ELA</b>				
3	173,695	98%	2%	0.66
4	171,185	96%	4%	0.71
5	160,808	93%	7%	0.64
6	158,210	92%	8%	0.67
7	148,857	94%	6%	0.67
8	143,555	92%	8%	0.64
<b>Mathematics</b>				
3	178,870	94%	6%	0.77
4	174,321	94%	6%	0.78
5	162,992	96%	4%	0.77
6	161,216	95%	5%	0.81
7	147,252	97%	3%	0.80
8	115,190	97%	3%	0.81

### 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 76% to 82% of the time across all performance levels and 94% to 99% 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 81% to 85% of the time across all performance levels and 96% to 98% 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
<b>ELA</b>			
3	173,695	82%	99%
4	171,185	78%	97%
5	160,808	78%	95%
6	158,210	76%	94%
7	148,857	80%	96%
8	143,555	80%	95%
<b>Mathematics</b>			
3	178,870	82%	96%
4	174,321	85%	96%
5	162,992	84%	97%
6	161,216	81%	96%
7	147,252	84%	98%
8	115,190	84%	98%

## Section 8: Summary of Operational Test Results

This section summarizes the distribution of scale score results on the NYSTP 2016 Grades 3–8 Common Core 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, 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/SUA subgroup is defined as English language learners 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 accommodations falling under the 504 Plan. 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.

### 8.1. 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.

#### 8.1.1. ELA Scale Score and Subscore Distributions

Table 8.1 shows some key statistics characterizing the distribution of ELA scale scores, while Table 8.2 summarizes the ELA subscores derived from the test in each grade. Tables 8.3 – 8.8 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 students, SWD, and/or SUA tended to under-perform the State population (All Students). This pattern of achievement was consistent across all grades.

**Table 8.1. ELA Scale Score Distribution Summary**

Grade	N-Count	Scale Score		Percentile Ranks				
		Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
3	180,303	309.01	34.97	264	288	311	333	350
4	177,092	306.38	33.28	263	287	309	331	345
5	167,409	297.38	39.51	247	274	301	325	346
6	166,040	299.71	36.09	253	279	303	324	342
7	156,248	302.18	34.69	256	280	305	327	347
8	150,849	304.09	34.80	257	284	307	329	343

**Table 8.2. ELA Subscore Summary**

Grade	Subscore	N-Count	Subscore		
			Max.	Mean	SD
3	Reading	180,303	25	15.19	5.40
	Writing	180,303	22	9.70	4.80
4	Reading	177,092	25	13.54	4.96
	Writing	177,092	22	11.96	4.97
5	Reading	167,409	35	21.55	6.58
	Writing	167,409	22	12.93	4.99
6	Reading	166,040	35	18.74	6.28
	Writing	166,040	22	14.22	5.14
7	Reading	156,248	35	19.30	6.69
	Writing	156,248	22	13.31	5.58
8	Reading	150,849	35	23.21	6.99
	Writing	150,849	22	15.35	5.18

#### 8.1.1.1. ELA Grade 3

Table 8.3 presents the scale score statistics and n-counts of demographic subgroups for Grade 3. The population scale score mean was 309.01 with a standard deviation of 34.97. Female students tended to outperform male students by around 9 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 (324.57). Across NRC categories, 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 (ELL) 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 49 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (311): Female (317), Asian (326), Multiracial (314), Pacific Islander (320), and White (317) students, those attending schools in Average (314) and Low (330) Needs districts and students attending Charter (320) and Non-Public (314) schools.

**Table 8.3. ELA Grade 3 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	180,303	309.01	34.97	264	288	311	333	350
Gender	Female	89,264	313.79	33.83	269	291	317	336	354
	Male	91,039	304.32	35.43	254	281	308	330	346
Ethnicity	Asian	18,237	324.57	32.81	281	305	326	346	363
	Black	33,101	300.63	34.77	254	277	301	326	343
	Hispanic	51,232	300.79	33.10	254	281	305	323	339
	American Indian	1,243	304.01	33.91	260	284	308	326	346
	Multiracial	4,476	311.65	36.03	264	288	314	336	354
	Pacific Islander	572	316.24	31.40	277	298	320	338	354
	White	71,442	314.68	34.36	269	295	317	339	354
NRC	New York	71,067	309.04	34.75	264	288	311	333	350
	Big 4 Cities	7,772	284.61	37.02	233	260	284	311	333
	Urban/Suburban	13,931	295.41	34.04	248	273	298	320	336
	Rural	9,662	299.44	34.35	254	281	301	323	339
	Average Needs	40,068	310.81	33.30	269	291	314	333	350
	Low Needs	17,567	326.76	29.19	291	311	330	346	358
	Charter	10,275	318.13	30.32	277	298	320	339	354
	Non-Public	9,927	308.06	36.15	260	288	314	333	350
SWD	All Codes	26,905	275.36	34.71	225	248	277	298	320
SUA	All Codes	12,231	271.13	35.22	225	248	273	295	317
ELL	ELL=Y	16,854	277.19	30.38	233	260	281	298	314
SWD/SUA	SUA=504 plan codes	9,998	265.85	33.83	225	241	264	291	311
ELL/SUA	SUA & ELL codes	1,122	260.31	29.16	225	241	260	277	298

**8.1.1.2. ELA Grade 4**

Table 8.4 contains Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 306.38 with a standard deviation of 33.28. Female students tended to outperform male students by around 9 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 (322.7). Across NRC categories, 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 subgroups scored, on average, about one standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 48 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (309): Female (312), Asian (324), Multiracial (312), Pacific Islander (315), and White (315) students, those from Average (312) and Low (324) Needs districts and those enrolled at Charter schools (315).

**Table 8.4. ELA Grade 4 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	177,092	306.38	33.28	263	287	309	331	345
Gender	Female	87,333	310.82	32.12	268	289	312	334	349
	Male	89,759	302.05	33.81	259	279	306	324	343
Ethnicity	Asian	17,770	322.70	31.06	283	306	324	345	358
	Black	33,190	298.31	32.35	254	275	299	321	338
	Hispanic	49,393	299.27	31.30	259	279	299	321	338
	American Indian	1,122	303.35	33.23	259	283	306	328	345
	Multiracial	3,809	308.75	34.85	263	287	312	334	349
	Pacific Islander	655	312.80	31.98	271	293	315	334	349
	White	71,153	310.86	33.21	268	293	315	334	349
NRC	New York	69,462	307.80	32.94	263	287	309	331	349
	Big 4 Cities	7,381	282.05	34.34	237	259	283	306	328
	Urban/Suburban	13,219	292.30	32.09	249	271	293	315	331
	Rural	9,168	295.88	32.89	254	275	299	320	334
	Average Needs	38,012	307.40	32.18	263	289	312	331	345
	Low Needs	16,999	322.27	28.14	287	306	324	343	353
	Charter	8,703	313.25	28.49	275	296	315	334	345
	Non-Public	14,148	305.96	33.33	259	287	309	328	345
SWD	All Codes	27,602	275.09	32.40	237	254	275	296	315
SUA	All Codes	13,680	272.13	33.53	228	249	271	296	315
ELL	ELL=Y	15,118	274.94	28.72	237	259	275	296	309
SWD/SUA	SUA=504 plan codes	10,555	265.47	32.01	220	243	263	287	309
ELL/SUA	SUA & ELL codes	1,148	258.66	26.73	220	243	259	275	293

**8.1.1.3. ELA Grade 5**

Table 8.5 provides the scale score summary statistics by key demographic subgroups for Grade 5 students. The population scale score mean was 297.38 with a standard deviation of 39.51. Female students tended to outperform male students by around 13 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 (315.52). Across NRC categories, 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 subgroups scored, on average, 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 62 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (301) Female (308), Asian (320), Pacific Islander (308), and White (308) students, those from Average (304) and Low (320) Needs districts and Charter schools (304).

**Table 8.5. ELA Grade 5 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	167,409	297.38	39.51	247	274	301	325	346
Gender	Female	82,133	304.10	37.12	258	283	308	328	346
	Male	85,276	290.91	40.66	239	268	295	320	337
Ethnicity	Asian	17,075	315.52	37.43	268	295	320	341	357
	Black	32,270	287.65	38.18	239	265	292	314	332
	Hispanic	46,573	288.87	37.08	243	268	292	314	332
	American Indian	1,118	291.99	39.16	243	268	295	320	341
	Multiracial	3,140	300.04	41.01	247	277	301	328	351
	Pacific Islander	475	305.53	36.38	258	286	308	328	346
	White	66,758	303.29	39.53	254	283	308	328	346
NRC	New York	67,570	299.04	38.96	251	277	301	325	346
	Big 4 Cities	6,751	268.91	43.52	208	243	271	298	321
	Urban/Suburban	12,302	280.63	38.70	229	258	283	308	325
	Rural	8,573	286.15	40.23	234	265	289	314	332
	Average Needs	36,269	299.58	37.98	251	277	304	325	346
	Low Needs	16,908	315.24	32.39	274	298	320	337	351
	Charter	9,349	300.91	33.11	258	280	304	325	341
	Non-Public	9,551	293.81	42.55	239	274	301	321	341
SWD	All Codes	28,145	259.98	39.32	208	234	265	286	308
SUA	All Codes	14,074	256.43	41.07	200	229	258	286	308
ELL	ELL=Y	12,300	252.86	35.91	200	234	258	277	295
SWD/SUA	SUA=504 plan codes	10,982	248.73	39.43	192	224	251	277	298
ELL/SUA	SUA & ELL codes	1,123	235.79	33.31	192	216	239	261	277

**8.1.1.4. ELA Grade 6**

Table 8.6 contains Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 299.71 with a standard deviation of 36.09. Female students tended to outperform male students by around 12 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 and Non-Public schools. Across ethnic groups, Asian students earned the highest mean score (318.64). Across NRC categories, 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 subgroups scored, on average, 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 54 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (303): Female (308), Asian (321), Multiracial (308), Pacific Islander (311), and White (308) students and those enrolled in Average (305) and Low (320) Needs districts and Non-Public schools (305).

**Table 8.6. ELA Grade 6 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	166,040	299.71	36.09	253	279	303	324	342
Gender	Female	81,474	305.73	33.58	263	285	308	327	347
	Male	84,566	293.92	37.45	245	270	297	321	338
Ethnicity	Asian	17,545	318.64	34.01	276	300	321	342	357
	Black	32,121	290.18	34.63	245	270	294	314	331
	Hispanic	44,634	291.21	33.72	245	270	294	314	331
	American Indian	1,137	293.00	35.05	249	273	297	320	335
	Multiracial	2,672	304.20	38.43	253	279	308	331	352
	Pacific Islander	450	309.17	33.37	267	291	311	331	347
	White	67,481	304.83	35.81	260	285	308	327	347
NRC	New York	63,916	300.70	35.71	253	279	303	324	342
	Big 4 Cities	6,567	273.69	38.28	225	249	276	300	321
	Urban/Suburban	11,045	283.92	36.43	236	260	288	308	327
	Rural	8,286	291.43	35.68	245	270	294	320	335
	Average Needs	35,060	301.19	35.17	257	279	305	324	342
	Low Needs	17,152	316.10	30.40	279	300	320	335	352
	Charter	10,479	301.07	29.84	263	283	303	321	338
	Non-Public	13,424	299.90	37.32	253	283	305	324	338
SWD	All Codes	27,171	265.36	34.18	217	245	267	288	308
SUA	All Codes	13,910	264.44	36.67	217	241	267	291	311
ELL	ELL=Y	12,212	259.03	32.46	217	241	260	283	297
SWD/SUA	SUA=504 plan codes	10,623	257.20	34.94	209	236	257	283	300
ELL/SUA	SUA & ELL codes	1,035	245.89	29.85	209	225	245	267	285

**8.1.1.5. ELA Grade 7**

Table 8.7 presents the Grade 7 scale score statistics and n-counts of demographic subgroups. The population scale score mean was 302.18 with a standard deviation of 34.69. Female students tended to outperform male students by around 14 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 and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (319.55). Across NRC categories, 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 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 51 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (305): Female (311), Asian (324), Multiracial (311), Pacific Islander (308), and White (311) students as well as those enrolled in Low Needs districts (321) and Non-Public schools (308).

**Table 8.7. ELA Grade 7 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	156,248	302.18	34.69	256	280	305	327	347
Gender	Female	76,119	309.28	32.76	266	288	311	333	348
	Male	80,129	295.44	35.12	248	272	298	321	337
Ethnicity	Asian	16,592	319.55	32.28	278	300	324	340	357
	Black	31,224	292.78	32.46	252	272	295	316	333
	Hispanic	42,218	294.61	32.18	252	275	295	316	333
	American Indian	1,139	297.25	34.09	256	278	298	321	340
	Multiracial	2,134	305.00	38.09	252	280	311	333	348
	Pacific Islander	438	305.94	32.84	263	287	308	330	347
	White	62,503	307.36	35.17	260	288	311	330	348
NRC	New York	64,587	304.18	33.25	263	283	305	327	347
	Big 4 Cities	6,230	277.32	35.38	233	252	278	303	324
	Urban/Suburban	10,436	284.25	34.94	239	260	287	311	327
	Rural	7,919	292.02	35.78	244	269	295	318	333
	Average Needs	31,962	302.77	35.02	256	280	305	327	347
	Low Needs	16,612	318.03	29.88	280	300	321	337	352
	Charter	8,901	304.18	27.42	269	287	305	324	337
	Non-Public	9,536	301.56	36.44	252	283	308	327	340
SWD	All Codes	25,573	270.02	31.53	226	248	272	291	308
SUA	All Codes	12,332	267.05	33.67	226	244	266	291	311
ELL	ELL=Y	10,645	261.31	28.32	226	244	263	280	295
SWD/SUA	SUA=504 plan codes	9,623	261.02	31.88	218	239	263	283	303
ELL/SUA	SUA & ELL codes	798	250.96	26.78	210	233	252	269	283

**8.1.1.6. ELA Grade 8**

Table 8.8 presents the Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 304.09 with a standard deviation of 34.80. Female students tended to outperform male students by around 13 scale score points. Asian, 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 (321.34). Across NRC categories, 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 subgroups scored, on average, one standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 51 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (307, Female (313), Asian (325), Pacific Islander (316), and White (313) students, as well as those enrolled in Low Needs districts (325) and Charter (310) and Non-Public (310) schools.

**Table 8.8. ELA Grade 8 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	150,849	304.09	34.80	257	284	307	329	343
Gender	Female	73,329	310.75	32.61	268	292	313	333	348
	Male	77,520	297.79	35.63	251	275	302	322	337
Ethnicity	Asian	16,338	321.34	32.52	280	302	325	343	355
	Black	31,832	295.31	32.05	254	275	297	319	333
	Hispanic	41,398	297.06	31.86	254	278	300	319	333
	American Indian	992	295.24	35.16	248	273	297	319	337
	Multiracial	1,731	304.06	37.43	251	280	307	329	348
	Pacific Islander	397	312.10	31.93	270	295	316	333	348
	White	58,161	309.14	36.08	262	290	313	333	348
NRC	New York	64,523	305.16	32.74	262	285	307	325	343
	Big 4 Cities	5,959	277.63	37.33	229	251	278	305	325
	Urban/Suburban	9,608	289.16	34.52	245	265	292	313	333
	Rural	7,445	295.35	35.66	248	273	300	319	337
	Average Needs	28,769	304.54	36.08	257	284	307	329	348
	Low Needs	15,112	320.93	31.06	280	305	325	343	355
	Charter	7,442	308.22	26.26	275	292	310	325	343
	Non-Public	11,925	303.98	36.28	260	288	310	325	343
SWD	All Codes	23,974	272.43	31.49	234	254	273	295	310
SUA	All Codes	11,509	270.34	34.04	229	248	270	292	313
ELL	ELL=Y	10,518	261.54	29.50	225	245	262	284	297
SWD/SUA	SUA=504 plan codes	8,921	264.31	32.31	225	245	265	288	305
ELL/SUA	SUA & ELL codes	672	252.76	26.40	225	237	254	270	285

**8.1.2. Mathematics Scale Score Distributions**

Table 8.9 shows some key statistics characterizing the distribution of mathematics scale scores, while Table 8.10 summarizes the mathematics subscores derived from the test in each grade. Tables 8.11 – 8.16 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 ELLs, SWDs, and/or SUAs achieved below the State mean in most percentile ranks. This pattern of achievement was fairly consistent across all grades.

**Table 8.9. Mathematics Scale Score Distribution Summary**

Grade	N-Count	Scale Score		Percentile Ranks				
		Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
3	180,824	305.89	39.50	257	280	307	331	353
4	177,147	304.60	40.95	252	279	308	333	354
5	166,838	306.51	39.29	256	282	308	334	354
6	163,927	304.67	41.29	252	279	306	333	354
7	151,897	304.56	39.80	244	280	309	333	352
8	117,643	292.72	41.22	236	270	296	320	341

**Table 8.10. Mathematics Subscore Summary**

Grade	Subscore	N-Count	Subscore		
			Max.	Mean	SD
3	Operations and Algebraic Thinking	180,824	25	13.56	6.14
	Number and Operations—Fractions	180,824	11	5.85	3.00
	Measurement and Data	180,824	11	7.43	2.57
4	Operations and Algebraic Thinking	177,147	11	5.88	3.12
	Number and Operations in Base Ten	177,147	16	9.89	4.28
	Number and Operations—Fractions	177,147	17	9.90	4.83
5	Number and Operations in Base Ten	166,838	16	9.42	4.06
	Number and Operations—Fractions	166,838	23	11.10	5.60
	Measurement and Data	166,838	7	3.10	1.78
6	Ratios and Proportional Relationships	163,927	17	7.86	4.09
	The Number System	163,927	13	6.57	3.06
	Expressions and Equations	163,927	23	11.23	5.27
7	Ratios and Proportional Relationships	151,897	20	7.91	5.16
	The Number System	151,897	12	5.88	3.49
	Expressions and Equations	151,897	21	10.71	5.07
8	Expressions and Equations	117,643	28	12.41	6.50
	Functions	117,643	11	5.00	2.76
	Geometry	117,643	12	5.16	3.24

**8.1.2.1. Mathematics Grade 3**

Table 8.11 presents the Grade 3 scale score statistics and n-counts of demographic subgroups. The population scale score mean was 305.89 with a standard deviation of 39.50. 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 Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (328.62). Across NRC categories, 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 subgroups scored, on average, 0.82 standard

deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 45 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (307): Asian (329), Multiracial (309), Pacific Islander (316), and White (316) students, as well as those enrolled at Average (312) and Low (326) Needs districts and Charter schools (321). In terms of the 50th-percentile ranks for students using translated forms, they ranged from 271 (Haitian-Creole, n = 86) to 323 (Chinese, n = 783).

**Table 8.11. Mathematics Grade 3 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	180,824	305.89	39.50	257	280	307	331	353
Gender	Female	89,256	306.38	38.08	257	285	307	331	353
	Male	91,568	305.42	40.82	252	280	307	331	358
Ethnicity	Asian	18,846	328.62	37.16	285	305	329	353	384
	Black	33,026	293.18	39.26	241	268	293	319	341
	Hispanic	51,784	294.78	36.85	247	271	296	319	341
	American Indian	1,256	299.19	38.04	252	278	300	323	344
	Multiracial	4,378	309.70	40.73	257	285	309	334	358
	Pacific Islander	585	314.89	37.10	265	293	316	340	358
	White	70,949	313.69	37.36	268	291	316	340	358
NRC	New York	72,428	304.26	39.51	257	280	305	329	353
	Big 4 Cities	7,883	278.72	40.46	226	252	278	305	331
	Urban/Suburban	13,862	290.92	37.55	241	268	293	316	340
	Rural	9,484	300.42	38.23	252	278	303	326	344
	Average Needs	39,280	309.82	36.99	265	288	312	334	353
	Low Needs	17,480	325.33	34.24	285	305	326	349	373
	Charter	10,295	320.84	37.44	275	296	321	344	373
	Non-Public	10,078	300.27	38.24	252	278	303	326	344
SWD	All Codes	26,877	274.90	39.49	218	247	275	300	323
SUA	All Codes	12,655	271.86	39.35	218	247	275	298	321
ELL	ELL=Y	18,934	277.03	37.04	226	252	278	300	323
SWD/SUA	SUA=504 plan codes	10,505	267.43	38.76	218	241	268	293	316
ELL/SUA	SUA & ELL codes	1,291	261.25	36.90	210	234	261	286	307
ELL Test Language	Chinese	783	324.86	33.86	285	303	323	344	373
	English	176,525	306.46	39.27	257	285	307	331	353
	Haitian-Creole	86	268.65	36.73	218	247	271	296	314
	Korean	46	321.72	43.04	261	314	329	349	365
	Russian	103	290.53	38.37	247	268	288	312	341
	Spanish	3,281	272.12	36.24	218	247	275	298	319
	All Translations	4,299	282.63	41.50	226	257	285	309	334

### 8.1.2.2. Mathematics Grade 4

Table 8.12 presents the Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 304.60 with a standard deviation of 40.95. 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 (330.43). Across NRC categories, 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 subgroups scored, on average, 0.84 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 47 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (308): Asian (333), Multiracial (311), Pacific Islander (314), and White (315) students, and those enrolled in Average (314) and Low (328) Needs districts and Charter schools (317). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 260 (Haitian-Creole, n = 88) to 323 (Chinese, n = 736).

**Table 8.12. Mathematics Grade 4 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	177,147	304.60	40.95	252	279	308	333	354
Gender	Female	87,170	304.92	39.92	252	279	306	330	354
	Male	89,977	304.28	41.93	247	277	308	333	354
Ethnicity	Asian	18,312	330.43	38.83	281	308	333	354	388
	Black	33,016	289.61	40.05	241	263	291	315	341
	Hispanic	49,917	292.87	38.65	241	269	295	319	341
	American Indian	1,124	300.34	40.34	252	275	300	327	354
	Multiracial	3,710	308.48	41.59	252	283	311	336	360
	Pacific Islander	667	312.70	40.34	260	288	314	341	367
	White	70,401	313.00	37.89	263	291	315	341	360
NRC	New York	70,714	303.08	42.17	247	275	304	330	360
	Big 4 Cities	7,428	274.12	41.55	216	247	275	304	328
	Urban/Suburban	12,988	286.87	39.05	234	260	289	314	336
	Rural	8,959	299.13	37.39	252	277	302	325	342
	Average Needs	37,253	309.64	37.21	260	289	314	333	354
	Low Needs	17,085	326.61	34.01	286	308	328	349	367
	Charter	8,731	316.40	38.16	269	291	317	342	367
	Non-Public	13,989	300.72	37.94	252	279	302	325	345
SWD	All Codes	27,416	270.93	39.32	216	247	269	297	321
SUA	All Codes	16,683	271.45	39.38	216	247	272	299	321
ELL	ELL=Y	17,115	272.32	37.91	225	247	272	297	319
SWD/SUA	SUA=504 plan codes	13,524	266.06	38.37	216	241	266	293	315

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
ELL/SUA	SUA & ELL codes	1,645	257.17	33.66	208	234	256	279	302
ELL Test Language	Chinese	736	323.17	36.90	281	302	323	345	367
	English	172,935	305.26	40.66	252	279	308	333	354
	Haitian-Creole	88	259.82	35.61	208	234	260	287	304
	Korean	67	315.91	42.41	256	283	319	349	360
	Russian	121	296.69	38.16	252	275	297	319	342
	Spanish	3,200	265.75	37.30	216	241	266	291	314
	All Translations	4,212	277.34	43.60	216	247	277	306	333

### 8.1.2.3. Mathematics Grade 5

Table 8.13 presents the Grade 5 demographic subgroup n-counts and scale score statistics. The population scale score mean was 306.51 with a standard deviation of 39.29. 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 Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (332.57). Across NRC categories, 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 subgroups scored, on average, about 0.85 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 45 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (308): Asian (334), Multiracial (312), Pacific Islander (312), and White (317) students, as well as those enrolled at Average (315) and Low (329) Needs districts and Charter schools (310). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 265 (Haitian-Creole, n = 71) to 327 (Korean, n = 57).

**Table 8.13. Mathematics Grade 5 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	166,838	306.51	39.29	256	282	308	334	354
Gender	Female	81,693	306.63	37.27	260	284	308	331	351
	Male	85,145	306.40	41.13	250	282	308	334	357
Ethnicity	Asian	17,581	332.57	37.66	287	310	334	357	382
	Black	31,935	290.42	37.00	244	268	294	315	336
	Hispanic	47,015	295.91	35.40	250	275	297	319	338
	American Indian	1,128	297.63	38.51	250	272	299	325	346
	Multiracial	3,045	309.26	41.33	256	282	312	338	361
	Pacific Islander	491	312.42	36.33	265	290	312	338	357
	White	65,643	314.93	37.46	268	294	317	340	357

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
NRC	New York	68,735	305.84	39.53	256	282	306	331	354
	Big 4 Cities	6,763	276.51	41.51	218	250	275	304	329
	Urban/Suburban	12,030	288.80	37.41	236	265	294	315	334
	Rural	8,240	299.39	36.81	250	279	302	325	343
	Average Needs	35,106	311.71	36.32	265	290	315	336	354
	Low Needs	16,744	328.92	33.47	287	308	329	351	370
	Charter	9,370	308.81	34.94	265	287	310	331	351
	Non-Public	9,712	300.19	37.50	250	279	302	325	346
SWD	All Codes	27,679	273.81	37.61	218	250	275	299	321
SUA	All Codes	16,295	274.49	38.72	218	250	275	302	323
ELL	ELL=Y	14,264	275.66	34.90	226	256	279	299	317
SWD/SUA	SUA=504 plan codes	13,203	269.23	37.59	218	244	268	295	317
ELL/SUA	SUA & ELL codes	1,577	261.49	32.05	218	236	265	284	302
ELL Test Language	Chinese	646	323.07	34.58	282	302	323	346	370
	English	162,834	306.98	39.28	256	284	308	334	354
	Haitian-Creole	71	259.30	35.81	210	226	265	287	299
	Korean	57	327.70	39.81	279	302	327	357	370
	Russian	88	289.50	38.03	236	263	294	318	343
	Spanish	3,142	279.81	28.14	244	260	279	299	315
	All Translations	4,004	287.32	34.32	244	265	284	308	331

#### 8.1.2.4. Mathematics Grade 6

Table 8.14 presents the Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 304.67 with a standard deviation of 41.29. Female students tended to outperform male students by around 4 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 Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (332.46). Across NRC categories, 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 subgroups scored, on average, 0.85 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 46 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (306): Female (308), Asian (335), Multiracial (312), Pacific Islander (312), and White (316) students, as well as those enrolled in Average (314) and Low (331) Needs districts and Charter schools (308). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 270 (Spanish, n = 3,850) to 335 (Korean, n = 102).

**Table 8.14. Mathematics Grade 6 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	163,927	304.67	41.29	252	279	306	333	354
Gender	Female	80,342	306.80	39.27	259	284	308	333	354
	Male	83,585	302.62	43.05	242	275	306	333	356
Ethnicity	Asian	18,008	332.46	39.25	284	308	335	359	379
	Black	31,597	287.96	39.10	230	265	289	314	337
	Hispanic	44,769	291.68	38.06	242	270	295	318	340
	American Indian	1,093	295.51	38.38	242	275	297	320	343
	Multiracial	2,539	311.22	42.86	259	286	312	343	365
	Pacific Islander	459	310.71	40.66	259	289	312	337	359
	White	65,462	313.83	38.21	265	292	316	340	359
NRC	New York	65,092	302.78	43.06	242	275	304	333	359
	Big 4 Cities	6,519	274.90	40.68	221	252	275	302	327
	Urban/Suburban	10,538	284.47	39.10	230	259	286	312	333
	Rural	7,807	299.04	36.98	252	279	302	324	343
	Average Needs	33,188	310.42	36.95	265	289	314	335	354
	Low Needs	16,783	329.17	34.03	286	310	331	351	368
	Charter	10,470	306.73	36.97	259	286	308	331	351
	Non-Public	13,427	300.81	38.66	252	279	304	325	345
SWD	All Codes	26,243	269.39	37.65	221	242	270	295	316
SUA	All Codes	16,464	273.24	38.99	221	252	275	300	322
ELL	ELL=Y	14,017	269.05	38.21	213	242	270	295	316
SWD/SUA	SUA=504 plan codes	13,327	268.13	37.65	213	242	270	292	314
ELL/SUA	SUA & ELL codes	1,668	258.60	33.36	213	230	259	284	300
ELL Test Language	Chinese	874	323.09	34.49	279	302	325	347	362
	English	158,869	305.56	40.96	252	284	308	333	356
	Haitian-Creole	89	269.02	35.87	213	242	270	297	316
	Korean	102	330.10	37.74	275	308	335	351	368
	Russian	143	292.36	44.82	230	259	292	320	345
	Spanish	3,850	264.47	34.21	213	242	270	289	306
	All Translations	5,058	276.79	41.89	221	252	275	304	331

**8.1.2.5. Mathematics Grade 7**

Table 8.15 presents the Grade 7 n-counts and scale score statistics for key demographic subgroups. The population scale score mean was 304.56 with a standard deviation of 39.80. Female students tended to outperform male students by around 4 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 (332.36). Across NRC

categories, 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 subgroups scored, on average, 0.87 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 47 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (309): Female (310), Asian (337), Multiracial (313), Pacific Islander (312), and White (318) students, those enrolled in Average (313) and Low (331) Needs districts and Charter schools (312). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 256 (Haitian-Creole, n = 83) to 336 (Korean, n = 89).

**Table 8.15. Mathematics Grade 7 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	151,897	304.56	39.80	244	280	309	333	352
Gender	Female	73,910	306.85	38.59	256	284	310	334	354
	Male	77,987	302.38	40.79	244	276	305	331	352
Ethnicity	Asian	16,761	332.36	37.28	284	312	337	356	373
	Black	30,239	287.87	37.85	236	265	290	315	336
	Hispanic	41,983	292.68	36.99	236	271	295	318	337
	American Indian	1,102	296.98	37.45	244	276	299	321	342
	Multiracial	1,964	309.76	40.99	256	284	313	339	359
	Pacific Islander	442	309.12	39.39	256	287	312	336	356
	White	59,406	313.54	36.42	265	295	318	339	356
NRC	New York	65,411	303.80	41.27	244	280	305	333	356
	Big 4 Cities	5,993	273.16	38.85	220	244	276	299	324
	Urban/Suburban	9,625	282.23	37.23	228	256	284	309	328
	Rural	7,230	296.17	35.53	244	276	301	319	337
	Average Needs	29,309	309.35	35.54	265	290	313	334	350
	Low Needs	15,736	327.76	31.50	290	312	331	348	362
	Charter	8,837	308.59	35.41	265	287	312	334	350
	Non-Public	9,693	301.61	36.95	244	280	305	327	344
SWD	All Codes	24,274	269.78	36.10	220	244	271	295	315
SUA	All Codes	13,498	272.94	37.49	220	244	276	299	321
ELL	ELL=Y	12,524	269.64	36.01	220	244	271	293	315
SWD/SUA	SUA=504 plan codes	10,944	267.88	35.86	220	236	271	293	313
ELL/SUA	SUA & ELL codes	1,030	257.44	32.16	213	236	256	280	297

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
ELL Test Language	Chinese	857	324.48	33.93	284	307	330	346	362
	English	147,216	305.41	39.45	244	280	309	333	354
	Haitian-Creole	83	257.60	34.06	213	228	256	280	305
	Korean	89	327.24	41.13	271	310	336	354	373
	Russian	112	301.36	30.72	271	284	306	321	336
	Spanish	3,540	264.87	32.98	220	236	271	287	305
	All Translations	4,681	277.71	41.23	220	244	276	305	334

#### 8.1.2.6. Mathematics Grade 8

Table 8.16 presents the Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 292.72 with a standard deviation of 41.22. Female students tended to outperform male students by around 6 scale score points. Asian, 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 and Non-Public schools. Across ethnic groups, Asian students earned the highest mean score (322.24). Across NRC categories, 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 subgroups scored, on average, about three-quarters of a standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed for English forms, scoring about 40 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (296): Female (299), Asian (325), Pacific Islander (306), and White (305) students, as well as those enrolled in Average (299) and Low (317) Needs districts and Charter (306) and Non-Public (303) schools. In terms of the 50th percentile ranks for students using translated forms, they ranged from: 266 (Spanish, n = 3,453) to 328 (Chinese, n = 777).

**Table 8.16. Mathematics Grade 8 Scale Score Distribution by Subgroup**

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
State	All Students	117,643	292.72	41.22	236	270	296	320	341
Gender	Female	56,305	295.66	39.80	236	274	299	322	343
	Male	61,338	290.01	42.30	228	266	294	318	341
Ethnicity	Asian	11,241	322.24	40.82	270	299	325	350	369
	Black	27,022	280.27	40.01	228	254	284	306	330
	Hispanic	36,370	284.93	38.85	228	260	287	310	331
	American Indian	786	282.50	40.15	228	260	284	310	330
	Multiracial	1,223	291.98	42.14	228	266	296	320	341
	Pacific Islander	315	305.38	40.32	254	278	306	333	355
	White	40,686	299.91	38.61	246	281	305	325	343

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
NRC	New York	54,791	293.40	42.22	236	266	294	322	349
	Big 4 Cities	5,353	262.01	41.50	212	228	260	292	317
	Urban/Suburban	7,668	271.72	37.66	220	246	278	299	317
	Rural	5,603	284.26	36.85	228	266	289	310	326
	Average Needs	18,369	293.53	35.60	246	274	299	318	333
	Low Needs	8,273	313.35	34.99	270	296	317	334	352
	Charter	6,077	305.70	38.20	254	281	306	331	352
	Non-Public	11,436	298.94	39.91	246	278	303	326	345
SWD	All Codes	21,514	261.71	37.79	212	236	266	289	310
SUA	All Codes	12,419	264.55	38.68	212	236	266	292	313
ELL	ELL=Y	12,050	265.50	39.40	212	236	266	292	315
SWD/SUA	SUA=504 plan codes	10,164	260.08	37.58	212	228	260	287	308
ELL/SUA	SUA & ELL codes	1,073	253.22	34.30	212	228	254	278	299
ELL Test Language	Chinese	777	325.60	34.79	284	306	328	350	364
	English	113,151	293.41	40.98	236	270	296	320	341
	Haitian-Creole	67	271.69	32.62	220	260	281	296	306
	Korean	55	319.62	34.58	274	303	323	343	357
	Russian	140	297.59	39.56	246	274	301	323	343
	Spanish	3,453	262.36	35.89	212	236	266	289	306
	All Translations	4,492	275.24	43.41	220	246	274	305	333

## 8.2. Performance Level Distribution Summary

Students are classified as NYS Level I, NYS Level II, NYS Level III, and NYS Level IV. The cut scores were established in 2013 during the standard-setting. Tables 6.13 and 6.14 show the ELA and Mathematics cut scores, respectively, used for classification of students into the four performance-level categories in 2016. It is inappropriate to compare scale scores across grades as they neither measure the same content, nor are they on the same scale. During the standard-setting process, while cut scores were set separately for different grades within a subject, additional care was taken to vertically articulate performance levels; see Section 8 and Appendix P in the 2013 technical report (NYSED, 2014) 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.

### 8.2.1. ELA Test Performance Level Distributions

Table 8.17 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 8.18 through 8.23, more Female students were classified in Level III and above categories than were Male students. Similarly, more Asian and White students were classified in Level III and above categories 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, SWD, and SUA subgroups were low, compared to the total population of examinees.

**Table 8.17. ELA Test Performance Level Distributions**

Grade	N-Count	Performance Levels				
		Level I	Level II	Level III	Level IV	Level III & IV
3	180,303	26.73	31.33	34.72	7.21	41.93
4	177,092	24.32	34.86	25.78	15.04	40.82
5	167,409	36.21	30.40	23.34	10.04	33.38
6	166,040	27.14	38.40	20.42	14.04	34.46
7	156,248	28.15	36.30	24.40	11.15	35.55
8	150,849	23.40	35.61	27.49	13.50	40.99

#### 8.2.1.1. ELA Grade 3

Table 8.18 presents the ELA Grade 3 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 41.93% of students achieved Level III and Level IV. About 47% of Female students were at Level III or above, as compared to 37% 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 (61%) students and students from Low Needs districts (66%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 18–32% of students in those same performance categories. Only about 9% of the SWD, SUA, and ELL 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 (42%), Female (47%), Asian (61%), Multiracial (46%), Pacific Islander (51%), White (50%) students, and those enrolled in Average (44%) and Low (66%) Needs districts and Charter (52%) schools.

**Table 8.18. 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	180,303	26.73	31.33	34.72	7.21	41.93
Gender	Female	89,264	22.31	30.90	37.51	9.27	46.78
	Male	91,039	31.07	31.75	31.98	5.20	37.18
Ethnicity	Asian	18,237	13.52	25.04	46.23	15.21	61.44
	Black	33,101	35.44	32.88	27.31	4.37	31.68
	Hispanic	51,232	34.35	35.10	27.16	3.39	30.56
	American Indian	1,243	31.13	34.19	29.53	5.15	34.67
	Multiracial	4,476	25.40	28.87	36.68	9.05	45.73
	Pacific Islander	572	18.36	30.59	43.71	7.34	51.05
	White	71,442	20.69	29.63	40.53	9.15	49.68

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	New York	71,067	27.30	31.81	33.18	7.71	40.89
	Big 4 Cities	7,772	54.66	26.99	15.86	2.48	18.35
	Urban/Suburban	13,931	40.33	34.29	23.03	2.35	25.38
	Rural	9,662	35.10	34.34	27.31	3.25	30.56
	Average Needs	40,068	23.50	32.74	37.13	6.63	43.76
	Low Needs	17,567	9.72	24.59	52.09	13.60	65.69
	Charter	10,275	17.27	30.83	42.49	9.41	51.90
	Non-Public	9,927	26.61	30.98	35.61	6.80	42.41
SWD	All Codes	26,905	65.45	23.37	10.34	0.84	11.18
SUA	All Codes	12,231	68.68	21.92	8.72	0.68	9.40
ELL	ELL=Y	16,854	64.32	28.05	7.38	0.25	7.63
SWD/SUA	SUA=504 plan codes	9,998	74.94	18.48	6.23	0.34	6.57
ELL/SUA	SUA & ELL codes	1,122	83.87	13.10	2.76	0.27	3.03

#### 8.2.1.2. ELA Grade 4

Table 8.19 presents the ELA Grade 4 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 40.82% of students achieved Level III and Level IV. About 46% of Female students were at Level III or above, as compared to 36% 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 (62%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 16–30% of students in those same performance categories. Only about 8% of the SWD, SUA, and ELL 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 (46%), Asian (62%), Multiracial (45%), Pacific Islander (50%), and White (47%) students as well as those enrolled in Average (42%) and Low (62%) Needs districts and Charter schools (49%).

**Table 8.19. 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	177,092	24.32	34.86	25.78	15.04	40.82
Gender	Female	87,333	20.18	34.03	27.71	18.08	45.79
	Male	89,759	28.35	35.67	23.90	12.08	35.98
Ethnicity	Asian	17,770	11.38	26.51	31.87	30.24	62.11
	Black	33,190	32.32	37.51	21.23	8.94	30.17
	Hispanic	49,393	30.58	39.00	21.78	8.63	30.42
	American Indian	1,122	27.81	35.56	24.33	12.30	36.63
	Multiracial	3,809	23.21	31.87	25.70	19.22	44.92
	Pacific Islander	655	17.25	33.13	28.40	21.22	49.62
	White	71,153	19.54	33.00	29.16	18.29	47.45

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	New York	69,462	23.55	35.02	24.95	16.47	41.43
	Big 4 Cities	7,381	53.61	30.00	11.95	4.44	16.39
	Urban/Suburban	13,219	38.66	38.14	17.85	5.35	23.19
	Rural	9,168	33.96	38.03	20.51	7.50	28.01
	Average Needs	38,012	21.98	35.86	27.73	14.43	42.16
	Low Needs	16,999	9.24	28.38	35.56	26.83	62.39
	Charter	8,703	15.49	35.96	31.71	16.83	48.55
	Non-Public	14,148	23.01	35.93	27.23	13.82	41.05
SWD	All Codes	27,602	61.77	28.33	7.71	2.18	9.90
SUA	All Codes	13,680	63.95	26.67	7.39	1.99	9.38
ELL	ELL=Y	15,118	61.85	31.85	5.60	0.71	6.30
SWD/SUA	SUA=504 plan codes	10,555	72.78	21.42	4.78	1.01	5.80
ELL/SUA	SUA & ELL codes	1,148	84.15	14.81	0.96	0.09	1.05

### 8.2.1.3. ELA Grade 5

Table 8.20 presents the ELA Grade 5 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 33.38% of students achieved Level III and Level IV. About 39% 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 (54%) students and students from Low Needs districts (53%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 14–23% of students in those same performance categories. Only about 5% of the SWD, SUA, and ELL 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 (33%): Female (39%), Asian (54%), Multiracial (37%), Pacific Islander (39%), and White (40%) students, as well as those enrolled in New York City (34%), Average (35%), and Low (53%) Needs districts and Charter schools (34%).

**Table 8.20. 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	167,409	36.21	30.40	23.34	10.04	33.38
Gender	Female	82,133	29.58	31.13	26.67	12.61	39.29
	Male	85,276	42.60	29.70	20.13	7.56	27.70
Ethnicity	Asian	17,075	19.56	26.87	32.17	21.40	53.57
	Black	32,270	46.30	30.62	18.00	5.08	23.08
	Hispanic	46,573	45.35	31.58	17.95	5.12	23.07
	American Indian	1,118	41.50	33.09	17.17	8.23	25.40
	Multiracial	3,140	34.27	28.69	23.73	13.31	37.04
	Pacific Islander	475	26.95	34.11	26.53	12.42	38.95
	White	66,758	29.29	30.39	27.49	12.83	40.32

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	New York	67,570	35.90	29.94	22.74	11.41	34.15
	Big 4 Cities	6,751	65.26	21.02	10.64	3.08	13.72
	Urban/Suburban	12,302	53.84	29.01	13.97	3.18	17.14
	Rural	8,573	47.09	29.98	17.52	5.41	22.93
	Average Needs	36,269	33.12	32.08	24.76	10.03	34.80
	Low Needs	16,908	16.76	30.71	35.68	16.84	52.53
	Charter	9,349	31.29	34.71	26.02	7.98	34.00
	Non-Public	9,551	36.59	31.32	23.83	8.26	32.09
SWD	All Codes	28,145	75.99	17.66	5.34	1.01	6.35
SUA	All Codes	14,074	77.65	16.10	5.28	0.97	6.25
ELL	ELL=Y	12,300	84.84	13.07	1.90	0.19	2.09
SWD/SUA	SUA=504 plan codes	10,982	84.58	12.10	2.90	0.42	3.31
ELL/SUA	SUA & ELL codes	1,123	96.17	3.29	0.53	--	0.53

#### 8.2.1.4. ELA Grade 6

Table 8.21 presents the ELA Grade 6 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 34.46% of students achieved Level III and Level IV. About 40% of Female students were at Level III or above, as compared to 29% 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 (54%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 13–25% of students in those same performance categories. Only about 5% of the SWD, SUA, and ELL 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 (34%): Female (40%), Asian (58%), Multiracial (42%), Pacific Islander (43%), and White (41%) students, as well as those from New York City (35%), Average (36%) and Low (54%) Needs districts and Non-Public schools (35%).

**Table 8.21. 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	166,040	27.14	38.40	20.42	14.04	34.46
Gender	Female	81,474	20.98	39.01	22.99	17.02	40.01
	Male	84,566	33.08	37.82	17.95	11.16	29.10

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
Ethnicity	Asian	17,545	12.38	30.08	26.70	30.84	57.54
	Black	32,121	36.40	40.42	15.72	7.45	23.18
	Hispanic	44,634	34.73	41.88	16.21	7.17	23.39
	American Indian	1,137	33.69	41.07	16.09	9.15	25.24
	Multiracial	2,672	25.37	32.71	22.19	19.72	41.92
	Pacific Islander	450	17.78	39.33	22.44	20.44	42.89
	White	67,481	21.57	37.48	23.79	17.15	40.95
NRC	New York	63,916	27.18	38.09	19.48	15.25	34.73
	Big 4 Cities	6,567	55.28	31.35	9.49	3.88	13.37
	Urban/Suburban	11,045	43.84	37.19	13.00	5.97	18.97
	Rural	8,286	34.79	39.85	16.75	8.60	25.36
	Average Needs	35,060	25.10	39.15	21.69	14.06	35.75
	Low Needs	17,152	11.33	34.96	29.31	24.40	53.71
	Charter	10,479	22.82	44.98	22.03	10.17	32.21
	Non-Public	13,424	23.76	40.81	22.59	12.84	35.43
SWD	All Codes	27,171	66.49	27.42	4.80	1.29	6.09
SUA	All Codes	13,910	66.11	26.50	5.61	1.78	7.39
ELL	ELL=Y	12,212	73.69	23.79	2.16	0.36	2.52
SWD/SUA	SUA=504 plan codes	10,623	74.40	21.54	3.32	0.73	4.06
ELL/SUA	SUA & ELL codes	1,035	88.12	11.79	0.10	--	0.10

#### 8.2.1.5. ELA Grade 7

Table 8.22 presents the ELA Grade 7 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 35.55% of students achieved Level III and Level IV. About 43% 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 (58%) students and students from Low Needs (56%) districts. The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 14–25% of students in those same performance categories. Only about 5% of the SWD, SUA, and ELL 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 (43%), Asian (58%), Multiracial (42%), Pacific Islander (38%), and White (43%) students, as well as those enrolled in Average (37%) and Low (56%) Needs districts and Non-Public schools (37%).

**Table 8.22. 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	156,248	28.15	36.30	24.40	11.15	35.55
Gender	Female	76,119	20.97	35.88	28.28	14.87	43.15
	Male	80,129	34.97	36.70	20.72	7.61	28.33
Ethnicity	Asian	16,592	13.21	28.99	33.80	24.00	57.80
	Black	31,224	37.23	39.71	18.18	4.88	23.06
	Hispanic	42,218	35.03	40.25	19.17	5.55	24.72
	American Indian	1,139	32.92	38.98	20.28	7.81	28.09
	Multiracial	2,134	27.04	31.40	26.05	15.51	41.57
	Pacific Islander	438	23.97	37.67	24.89	13.47	38.36
	White	62,503	22.91	33.98	28.56	14.54	43.10
NRC	New York	64,587	26.32	37.68	23.87	12.13	36.00
	Big 4 Cities	6,230	57.19	29.15	11.03	2.63	13.66
	Urban/Suburban	10,436	48.73	33.48	13.69	4.10	17.79
	Rural	7,919	38.11	36.61	19.16	6.12	25.28
	Average Needs	31,962	27.53	35.61	25.27	11.59	36.86
	Low Needs	16,612	12.15	31.83	36.35	19.67	56.02
	Charter	8,901	22.35	44.13	26.93	6.59	33.52
	Non-Public	9,536	26.17	37.24	26.71	9.88	36.59
SWD	All Codes	25,573	66.93	27.24	4.99	0.84	5.83
SUA	All Codes	12,332	68.85	24.85	5.32	0.99	6.31
ELL	ELL=Y	10,645	79.21	19.35	1.32	0.11	1.44
SWD/SUA	SUA=504 plan codes	9,623	76.49	20.00	3.15	0.35	3.50
ELL/SUA	SUA & ELL codes	798	90.85	8.65	0.50	--	0.50

**8.2.1.6. ELA Grade 8**

Table 8.23 presents the ELA Grade 8 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 40.99% of students achieved Level III and Level IV. About 48% of Female students were at Level III or above, as compared to 34% 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 (64%) students and students from Low Needs (64%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 16–31% of students in those same performance categories. Only about 6% of the SWD, SUA, and ELL 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 (48%), Asian (64%), Multiracial (42%), Pacific Islander (52%), and White (49%) students, as well as those attending Average (43%) and Low (64%) Needs districts and Charter (42%) and Non-Public (43%) schools.

**Table 8.23. 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	150,849	23.40	35.61	27.49	13.50	40.99
Gender	Female	73,329	17.34	34.57	30.79	17.30	48.09
	Male	77,520	29.13	36.60	24.37	9.90	34.27
Ethnicity	Asian	16,338	11.03	25.30	35.31	28.36	63.67
	Black	31,832	30.35	41.06	22.29	6.30	28.58
	Hispanic	41,398	28.46	41.05	23.45	7.04	30.49
	American Indian	992	31.96	38.61	20.56	8.87	29.44
	Multiracial	1,731	25.30	33.04	26.05	15.60	41.65
	Pacific Islander	397	16.37	31.74	33.50	18.39	51.89
	White	58,161	19.31	31.71	31.14	17.84	48.98
NRC	New York	64,523	22.11	37.38	27.04	13.47	40.51
	Big 4 Cities	5,959	53.35	30.17	12.52	3.96	16.48
	Urban/Suburban	9,608	38.38	37.90	18.18	5.54	23.72
	Rural	7,445	31.42	37.21	23.16	8.22	31.38
	Average Needs	28,769	23.59	33.77	28.02	14.62	42.64
	Low Needs	15,112	10.02	26.40	37.58	26.01	63.59
	Charter	7,442	14.79	43.47	31.55	10.19	41.74
	Non-Public	11,925	20.18	37.22	31.03	11.58	42.61
SWD	All Codes	23,974	59.74	32.42	6.82	1.02	7.85
SUA	All Codes	11,509	61.95	29.13	7.44	1.48	8.91
ELL	ELL=Y	10,518	74.48	23.19	2.22	0.10	2.33
SWD/SUA	SUA=504 plan codes	8,921	69.49	25.73	4.16	0.63	4.79
ELL/SUA	SUA & ELL codes	672	87.20	12.80	--	--	--

**8.2.2. Mathematics Test Performance Level Distributions**

Table 8.24 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 8.25 through Table 8.30, 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 Non-Public 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 8.24. Mathematics Test Performance Level Distributions**

Grade	N-Count	Performance Levels				
		Level I	Level II	Level III	Level IV	Level III & IV
3	180,824	25.41	30.88	22.02	21.69	43.71
4	177,147	27.63	28.12	23.44	20.80	44.25
5	166,838	32.29	28.03	23.86	15.81	39.67
6	163,927	25.88	34.43	18.56	21.14	39.70
7	151,897	33.76	30.72	21.94	13.57	35.51
8	117,643	39.09	36.58	16.21	8.12	24.33

**8.2.2.1. Mathematics Grade 3**

Table 8.25 presents the Mathematics Grade 3 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 43.71% 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 (68%) students and students from Low Needs (66%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 20–38% of students in those same performance categories. Only about 15% of the SWD, SUA, and ELL 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 (68%), Multiracial (48%), Pacific Islander (55%), and White (53%) students, as well as those enrolled at Average (48%) and Low (66%) Needs districts and Charter schools (59%). For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 10% (Haitian-Creole) to 76% (Korean).

**Table 8.25. Mathematics 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	180,824	25.41	30.88	22.02	21.69	43.71
Gender	Female	89,256	24.50	32.04	22.44	21.03	43.46
	Male	91,568	26.31	29.75	21.61	22.34	43.95
Ethnicity	Asian	18,846	9.83	22.21	25.47	42.49	67.97
	Black	33,026	37.30	32.75	16.80	13.15	29.95
	Hispanic	51,784	34.06	35.26	18.75	11.93	30.68
	American Indian	1,256	30.25	34.47	19.75	15.53	35.27
	Multiracial	4,378	23.00	29.21	22.89	24.90	47.78
	Pacific Islander	585	17.26	27.86	28.03	26.84	54.87
	White	70,949	17.83	29.18	25.85	27.14	52.99

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	New York	72,428	27.19	31.80	20.79	20.22	41.01
	Big 4 Cities	7,883	53.08	27.25	11.76	7.92	19.68
	Urban/Suburban	13,862	38.65	33.67	16.99	10.68	27.67
	Rural	9,484	28.17	33.58	21.55	16.69	38.24
	Average Needs	39,280	20.58	31.25	25.05	23.13	48.18
	Low Needs	17,480	9.36	24.69	28.32	37.63	65.95
	Charter	10,295	14.19	27.14	24.27	34.40	58.67
	Non-Public	10,078	28.38	33.85	21.17	16.60	37.78
SWD	All Codes	26,877	56.29	27.62	10.30	5.80	16.10
SUA	All Codes	12,655	58.94	26.73	9.58	4.75	14.33
ELL	ELL=Y	18,934	54.24	30.17	10.36	5.23	15.59
SWD_SUA	SUA=504 plan codes	10,505	63.87	24.47	8.14	3.51	11.65
ELL_SUA	SUA & ELL codes	1,291	71.73	20.06	5.65	2.56	8.21
ELL Test Language	Chinese	783	8.68	26.95	27.97	36.40	64.37
	English	176,525	24.83	30.96	22.23	21.98	44.21
	Haitian-Creole	86	62.79	26.74	6.98	3.49	10.47
	Korean	46	17.39	6.52	39.13	36.96	76.09
	Russian	103	41.75	33.98	11.65	12.62	24.27
	Spanish	3,281	59.22	27.98	9.48	3.32	12.80
	All Translations	4,299	49.22	27.68	13.17	9.93	23.10

#### 8.2.2.2. Mathematics Grade 4

Table 8.26 presents the Mathematics Grade 4 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 44.25% of students achieved Level III and Level IV. About 44% 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 (70%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 18–38% of students in those same performance categories. Only about 14% of the SWD, SUA, and ELL 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 (51%), and White (54%) students, as well as students enrolled in Average (50%) and Low (70%) Needs and Charter schools (55%). For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 5% (Haitian-Creole) to 64% (Chinese).

**Table 8.26. 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	177,147	27.63	28.12	23.44	20.80	44.25
Gender	Female	87,170	27.05	28.88	23.77	20.30	44.07
	Male	89,977	28.20	27.38	23.13	21.29	44.42
Ethnicity	Asian	18,312	10.71	18.76	25.85	44.68	70.53
	Black	33,016	41.90	29.92	17.03	11.14	28.17
	Hispanic	49,917	37.34	31.59	19.52	11.55	31.07
	American Indian	1,124	32.30	30.34	19.48	17.88	37.37
	Multiracial	3,710	24.91	26.31	24.42	24.37	48.79
	Pacific Islander	667	20.84	28.49	25.04	25.64	50.67
	White	70,401	18.59	27.30	28.61	25.50	54.10
NRC	New York	70,714	30.68	27.91	20.66	20.75	41.41
	Big 4 Cities	7,428	57.12	24.77	12.16	5.95	18.11
	Urban/Suburban	12,988	43.25	30.06	17.89	8.80	26.69
	Rural	8,959	29.46	32.25	24.76	13.54	38.30
	Average Needs	37,253	20.56	29.22	28.18	22.03	50.21
	Low Needs	17,085	8.94	21.36	31.98	37.73	69.70
	Charter	8,731	17.90	27.39	25.52	29.19	54.71
	Non-Public	13,989	28.64	32.28	23.52	15.56	39.08
SWD	All Codes	27,416	61.84	23.82	9.59	4.75	14.34
SUA	All Codes	16,683	60.34	24.47	10.75	4.43	15.18
ELL	ELL=Y	17,115	60.14	25.91	9.65	4.30	13.95
SWD_SUA	SUA=504 plan codes	13,524	66.31	22.17	8.39	3.14	11.52
ELL_SUA	SUA & ELL codes	1,645	77.20	17.93	4.07	0.79	4.86
ELL Test Language	Chinese	736	12.09	24.32	29.62	33.97	63.59
	English	172,935	26.96	28.23	23.72	21.09	44.81
	Haitian-Creole	88	71.59	23.86	4.55	.	4.55
	Korean	67	23.88	16.42	25.37	34.33	59.70
	Russian	121	31.40	37.19	19.01	12.40	31.40
	Spanish	3,200	66.59	22.97	7.84	2.59	10.44
	All Translations	4,212	55.48	23.53	12.18	8.81	20.99

**8.2.2.3. Mathematics Grade 5**

Table 8.27 presents the Mathematics Grade 5 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 39.67% of students achieved Level III and Level IV. About 39% of Female students were at Level III or above, as compared to 40% 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 (68%) students and students from Low Needs districts (65%). The Big 4

Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 16–32% of students in those same performance categories. Only about 11% of the SWD, SUA, and ELL 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%): Asian (68%), Multiracial (44%), Pacific Islander (45%), and White (50%) students, as well as those enrolled in Average (46%) and Low (65%) Needs districts and Charter schools (41%). For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 3% (Haitian-Creole) to 60% (Korean).

**Table 8.27. 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	166,838	32.29	28.03	23.86	15.81	39.67
Gender	Female	81,693	31.29	29.85	24.45	14.40	38.86
	Male	85,145	33.25	26.29	23.30	17.16	40.46
Ethnicity	Asian	17,581	12.14	20.12	29.96	37.77	67.73
	Black	31,935	47.89	29.41	16.78	5.91	22.70
	Hispanic	47,015	41.85	31.90	18.86	7.39	26.24
	American Indian	1,128	42.02	27.39	19.95	10.64	30.59
	Multiracial	3,045	31.66	24.70	23.65	20.00	43.65
	Pacific Islander	491	25.87	29.53	25.66	18.94	44.60
	White	65,643	23.16	26.87	29.33	20.65	49.97
NRC	New York	68,735	34.02	28.44	21.77	15.78	37.54
	Big 4 Cities	6,763	63.43	20.91	10.13	5.53	15.66
	Urban/Suburban	12,030	49.49	28.68	16.39	5.44	21.83
	Rural	8,240	37.49	30.45	22.57	9.49	32.06
	Average Needs	35,106	25.52	28.71	28.45	17.32	45.77
	Low Needs	16,744	11.34	23.20	33.83	31.62	65.46
	Charter	9,370	28.67	30.78	26.52	14.03	40.55
	Non-Public	9,712	37.06	30.59	22.15	10.20	32.35
SWD	All Codes	27,679	66.86	21.63	8.74	2.77	11.51
SUA	All Codes	16,295	66.13	20.90	9.59	3.38	12.97
ELL	ELL=Y	14,264	66.90	23.16	7.45	2.50	9.95
SWD_SUA	SUA=504 plan codes	13,203	71.59	18.80	7.39	2.22	9.61
ELL_SUA	SUA & ELL codes	1,577	82.75	14.27	2.54	0.44	2.98
ELL Test Language	Chinese	646	16.10	26.16	31.73	26.01	57.74
	English	162,834	31.68	28.10	24.16	16.05	40.22
	Haitian-Creole	71	81.69	15.49	1.41	1.41	2.82
	Korean	57	14.04	26.32	24.56	35.09	59.65
	Russian	88	48.86	26.14	15.91	9.09	25.00
	Spanish	3,142	65.98	25.21	7.57	1.24	8.82
	All Translations	4,004	57.09	25.22	11.79	5.89	17.68

#### 8.2.2.4. Mathematics Grade 6

Table 8.28 presents the Mathematics Grade 6 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 39.70% of students achieved Level III and Level IV. About 41% 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 (68%) students and students from Low Needs districts (68%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 15–32% of students in those same performance categories. Only about 10% of the SWD, SUA, and ELL 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 (41%), Asian (68%), Multiracial (46%), Pacific Islander (44%), and White (50%) students, as well as those enrolled in Average (46%) and Low (68%) Needs districts and Charter schools (41%). For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 8% (Haitian-Creole) to 72% (Korean).

**Table 8.28. 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	163,927	25.88	34.43	18.56	21.14	39.70
Gender	Female	80,342	23.29	35.80	19.70	21.22	40.92
	Male	83,585	28.37	33.11	17.45	21.07	38.52
Ethnicity	Asian	18,008	9.57	22.01	20.61	47.81	68.42
	Black	31,597	40.03	37.28	13.33	9.36	22.70
	Hispanic	44,769	35.63	39.27	14.82	10.27	25.10
	American Indian	1,093	30.92	40.99	16.01	12.08	28.09
	Multiracial	2,539	22.10	31.82	18.16	27.92	46.08
	Pacific Islander	459	17.21	38.56	19.61	24.62	44.23
	White	65,462	16.99	33.11	23.11	26.78	49.89
NRC	New York	65,092	29.03	34.11	15.77	21.09	36.87
	Big 4 Cities	6,519	54.38	30.80	9.04	5.78	14.82
	Urban/Suburban	10,538	43.79	35.82	12.74	7.64	20.38
	Rural	7,807	27.00	41.00	18.75	13.24	32.00
	Average Needs	33,188	18.59	35.81	22.86	22.74	45.60
	Low Needs	16,783	7.76	24.70	25.79	41.76	67.54
	Charter	10,470	22.18	37.33	20.63	19.87	40.50
	Non-Public	13,427	25.75	39.33	19.78	15.14	34.92
SWD	All Codes	26,243	61.16	29.40	6.15	3.30	9.45
SUA	All Codes	16,464	56.52	31.13	7.85	4.50	12.35
ELL	ELL=Y	14,017	61.07	29.46	5.86	3.61	9.47
SWD_SUA	SUA=504 plan codes	13,327	62.42	28.58	6.04	2.96	9.00
ELL_SUA	SUA & ELL codes	1,668	74.10	22.66	2.70	0.54	3.24

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
ELL Test Language	Chinese	874	11.78	28.60	23.91	35.70	59.61
	English	158,869	24.98	34.58	18.89	21.56	40.44
	Haitian-Creole	89	59.55	32.58	7.87	.	7.87
	Korean	102	12.75	15.69	27.45	44.12	71.57
	Russian	143	40.56	31.47	12.59	15.38	27.97
	Spanish	3,850	65.40	29.92	3.92	0.75	4.68
	All Translations	5,058	54.27	29.50	8.17	8.07	16.23

#### 8.2.2.5. Mathematics Grade 7

Table 8.29 presents the Mathematics Grade 7 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 35.51% of students achieved Level III and Level IV. About 37% of Female students were at Level III or above, as compared to 34% 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 (66%) students and students from Low Needs districts (64%). 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 7% of the SWD, SUA, and ELL 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 (37%), Asian (66%), Multiracial (42%), Pacific Islander (40%), and White (45%) students, as well as those enrolled in Average (40%) and Low (64%) Needs districts and Charter schools (39%). For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 2% (Haitian-Creole) to 63% (Korean).

**Table 8.29. 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	151,897	33.76	30.72	21.94	13.57	35.51
Gender	Female	73,910	31.09	31.90	22.96	14.05	37.01
	Male	77,987	36.30	29.61	20.97	13.12	34.10
Ethnicity	Asian	16,761	12.73	20.92	28.73	37.62	66.35
	Black	30,239	50.84	30.34	13.90	4.91	18.81
	Hispanic	41,983	45.18	33.03	15.94	5.85	21.79
	American Indian	1,102	41.83	33.30	17.06	7.80	24.86
	Multiracial	1,964	29.38	28.46	23.83	18.33	42.16
	Pacific Islander	442	27.83	31.90	24.66	15.61	40.27
	White	59,406	22.97	32.07	28.36	16.60	44.96

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	New York	65,411	36.57	29.39	18.90	15.14	34.04
	Big 4 Cities	5,993	67.15	21.94	8.31	2.60	10.91
	Urban/Suburban	9,625	56.82	28.88	11.36	2.94	14.30
	Rural	7,230	38.71	38.15	17.93	5.21	23.14
	Average Needs	29,309	25.79	34.55	27.30	12.36	39.66
	Low Needs	15,736	10.63	25.74	36.50	27.13	63.63
	Charter	8,837	28.38	33.03	25.86	12.73	38.59
	Non-Public	9,693	34.26	35.83	20.91	9.00	29.91
SWD	All Codes	24,274	71.97	21.06	5.41	1.56	6.97
SUA	All Codes	13,498	67.68	22.91	7.22	2.19	9.41
ELL	ELL=Y	12,524	72.91	19.80	5.56	1.73	7.29
SWD_SUA	SUA=504 plan codes	10,944	73.61	20.11	5.03	1.24	6.28
ELL_SUA	SUA & ELL codes	1,030	86.21	11.65	1.94	0.19	2.14
ELL Test Language	Chinese	857	14.35	26.02	35.59	24.04	59.63
	English	147,216	32.79	31.05	22.33	13.83	36.16
	Haitian-Creole	83	81.93	15.66	2.41	.	2.41
	Korean	89	15.73	21.35	29.21	33.71	62.92
	Russian	112	33.93	42.86	17.86	5.36	23.21
	Spanish	3,540	78.31	18.25	3.02	0.42	3.45
	All Translations	4,681	64.41	20.27	9.83	5.49	15.32

#### 8.2.2.6. Mathematics Grade 8

Table 8.30 presents the Mathematics Grade 8 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 24.33% of students achieved Level III and Level IV. About 26% of Female students were at Level III or above, as compared to 23% 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 (54%) students and students from Low Needs districts (44%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 8–17% of students in those same performance categories. Only about 6% of the SWD, SUA, and ELL 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 (24%): Female (26%), Asian (54%), Pacific Islander (37%), and White (30%) students, as well as those enrolled in New York City (25%) and Low Needs districts (44%) and Charter (35%) and Non-Public (30%) schools. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 1% (Haitian-Creole) to 58% (Korean).

**Table 8.30. 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	117,643	39.09	36.58	16.21	8.12	24.33
Gender	Female	56,305	36.00	38.35	17.04	8.62	25.65
	Male	61,338	41.93	34.95	15.45	7.67	23.12
Ethnicity	Asian	11,241	16.47	29.09	25.97	28.48	54.44
	Black	27,022	52.55	32.96	10.29	4.20	14.49
	Hispanic	36,370	47.16	36.26	12.12	4.47	16.59
	American Indian	786	50.64	33.46	11.83	4.07	15.90
	Multiracial	1,223	39.25	36.79	16.43	7.52	23.96
	Pacific Islander	315	31.11	32.06	21.90	14.92	36.83
	White	40,686	29.03	41.42	21.14	8.41	29.55
NRC	New York	54,791	40.57	34.40	14.96	10.07	25.03
	Big 4 Cities	5,353	70.76	20.90	5.88	2.45	8.33
	Urban/Suburban	7,668	60.33	32.04	6.47	1.16	7.63
	Rural	5,603	44.57	41.66	11.76	2.02	13.78
	Average Needs	18,369	33.65	45.15	17.82	3.38	21.20
	Low Needs	8,273	16.55	39.08	30.55	13.83	44.37
	Charter	6,077	27.71	37.06	21.70	13.53	35.23
	Non-Public	11,436	31.31	39.11	19.83	9.74	29.57
SWD	All Codes	21,514	72.46	22.26	4.28	1.01	5.29
SUA	All Codes	12,419	68.93	24.57	5.22	1.28	6.50
ELL	ELL=Y	12,050	68.74	23.43	5.65	2.18	7.83
SWD_SUA	SUA=504 plan codes	10,164	73.77	21.64	3.71	0.89	4.59
ELL_SUA	SUA & ELL codes	1,073	83.69	14.17	1.30	0.84	2.14
ELL Test Language	Chinese	777	10.04	32.18	29.73	28.06	57.79
	English	113,151	38.29	36.99	16.51	8.21	24.72
	Haitian-Creole	67	59.70	38.81	.	1.49	1.49
	Korean	55	12.73	29.09	38.18	20.00	58.18
	Russian	140	34.29	38.57	19.29	7.86	27.14
	Spanish	3,453	72.17	24.04	3.07	0.72	3.79
	All Translations	4,492	59.33	26.18	8.57	5.92	14.49

## Section 9: 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 Measurement* 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**

Grade	Day Book		Number of Items				
			Multiple-Choice		Constructed-Response		Total
			Operational	Embedded	Operational	Embedded	
3	1	1	18	6	0	0	24
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	<b>Total</b>		<b>25</b>	<b>6</b>	<b>9</b>	<b>0</b>	<b>40</b>
4	1	1	18	6	0	0	24
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	<b>Total</b>		<b>25</b>	<b>6</b>	<b>9</b>	<b>0</b>	<b>40</b>
5	1	1	28	7	0	0	35
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	<b>Total</b>		<b>35</b>	<b>7</b>	<b>9</b>	<b>0</b>	<b>51</b>
6	1	1	28	7	0	0	35
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	<b>Total</b>		<b>35</b>	<b>7</b>	<b>9</b>	<b>0</b>	<b>51</b>
7	1	1	28	7	0	0	35
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	<b>Total</b>		<b>35</b>	<b>7</b>	<b>9</b>	<b>0</b>	<b>51</b>
8	1	1	28	7	0	0	35
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	<b>Total</b>		<b>35</b>	<b>7</b>	<b>9</b>	<b>0</b>	<b>51</b>

Appendix A: ELA and Mathematics Test Configurations and Testing Times

**Table A2. Mathematics Test Configuration**

Grade	Day Book		Number of Items				Total
			Multiple-Choice		Constructed-Response		
			Operational	Embedded	Operational	Embedded	
3	1	1	18	4	0	0	22
	2	2	19	3	0	0	22
	3	3	0	0	8	0	8
	<b>Total</b>		<b>37</b>	<b>7</b>	<b>8</b>	<b>0</b>	<b>52</b>
4	1	1	18	4	0	0	22
	2	2	20	3	0	0	23
	3	3	0	0	10	0	10
	<b>Total</b>		<b>38</b>	<b>7</b>	<b>10</b>	<b>0</b>	<b>55</b>
5	1	1	18	4	0	0	22
	2	2	19	3	0	0	22
	3	3	0	0	10	0	10
	<b>Total</b>		<b>37*</b>	<b>7</b>	<b>10</b>	<b>0</b>	<b>54</b>
6	1	1	21	4	0	0	25
	2	2	22	3	0	0	25
	3	3	0	0	10	0	10
	<b>Total</b>		<b>43*</b>	<b>7</b>	<b>10</b>	<b>0</b>	<b>60</b>
7	1	1	22	4	0	0	26
	2	2	22	3	0	0	25
	3	3	0	0	10	0	10
	<b>Total</b>		<b>44</b>	<b>7</b>	<b>10</b>	<b>0</b>	<b>61</b>
8	1	1	22	4	0	0	26
	2	2	22	3	0	0	25
	3	3	0	0	10	0	10
	<b>Total</b>		<b>44</b>	<b>7</b>	<b>10</b>	<b>0</b>	<b>61</b>

\*One item each in Grades 5 and 6 were excluded from the analysis and scoring due to poor fit to the item response theory (IRT) model.

**Table A3. ELA Estimated Time on Task by Book**

Grades	Day	Book	Estimated Time on Task (min.)
3–4	1	1	60–70
	2	2	60–70
	3	3	60–70
	<b>Total</b>		<b>180–210</b>
5–8	1	1	80–90
	2	2	80–90
	3	3	80–90
	<b>Total</b>		<b>240–270</b>

Source: 2016 Common Core 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 Book**

Grade(s)	Day	Book	Estimated Time Needed (min.)
3	1	1	50–60
	2	2	50–60
	3	3	60–70
	<b>Total</b>		<b>160–190</b>
4	1	1	50–60
	2	2	50–60
	3	3	80–90
	<b>Total</b>		<b>180–210</b>
5–8	1	1	70–80
	2	2	70–80
	3	3	80–90
	<b>Total</b>		<b>220–250</b>

Source: 2016 Common Core 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

## Appendix A: ELA and Mathematics Test Configurations and Testing Times

details on security, scheduling, classroom organization and preparation, test materials, and administration can be found in the 2016 *Teacher's Directions* and the *School Administrator's Manual*, which are accessible online:

- *2016 Common Core ELA Teacher's Directions*
  - Grades 3–5: <http://www.p12.nysed.gov/assessment/sam/ei/td-35ela16.pdf>
  - Grades 6–8: <http://www.p12.nysed.gov/assessment/sam/ei/td-68ela16.pdf>
- *2016 Common Core Mathematics Teacher's Directions*
  - Grades 3–5: <http://www.p12.nysed.gov/assessment/sam/ei/td-35math16.pdf>
  - Grades 6–8: <http://www.p12.nysed.gov/assessment/sam/ei/td-68math16.pdf>
- *2016 Common Core ELA and Mathematics Tests School Administrator's Manual*
  - <http://www.p12.nysed.gov/assessment/sam/ei/eisam16.pdf>
- *2016 Common Core ELA and Mathematics Test Guides*
  - <https://www.engageny.org/resource/test-guides-for-english-language-arts-and-mathematics>

## Appendix B: ELA and Mathematics Test Blueprints

---

**Table B1. ELA Test Blueprint**

Grade	Total Points on OP Test	Standard	Point Range		% of Test	
			Target	Actual	Target	Actual
3	55	Literature	14–44	24	30%–94%	51%
		Information	14–44	22	30%–94%	47%
		Language	1–4	1	2%–9%	2%
4	55	Literature	14–44	20	30%–94%	43%
		Information	14–44	26	30%–94%	55%
		Language	1–4	1	2%–9%	2%
5	66	Literature	18–51	27	32%–89%	47%
		Information	18–51	28	32%–89%	49%
		Language	1–4	2	2%–7%	4%
6	65	Literature	11–44	25	19%–77%	44%
		Information	25–58	31	44%–102%	54%
		Language	1–4	1	2%–7%	2%
7	66	Literature	11–44	28	19%–77%	49%
		Information	25–58	28	44%–102%	49%
		Language	1–4	1	2%–7%	2%
8	66	Literature	11–44	26	19%–77%	46%
		Information	25–58	30	44%–102%	53%
		Language	1–4	1	2%–7%	2%

**Table B2. Mathematics Test Blueprint**

Grade	Total Points on OP Test	Standard	Point Range		% of Test	
			Target	Actual	Target	Actual
3	60	Operations and Algebraic Thinking	23–31	25	41%–55%	45%
		Number and Operations in Base Ten	3–5	4	5%–9%	7%
		Number and Operations – Fractions	10–14	11	18%–25%	20%
		Measurement and Data	12–18	14	21%–32%	25%
		Geometry*	1–3	2	2%–5%	4%
4	66	Operations and Algebraic Thinking	11–15	13	18%–24%	21%
		Number and Operations in Base Ten	14–20	16	23%–32%	26%
		Number and Operations – Fractions	15–21	17	24%–34%	27%
		Measurement and Data	9–15	10	15%–24%	16%
		Geometry	5–7	6	8%–11%	10%
5	66	Operations and Algebraic Thinking	3–5	4	5%–8%	7%
		Number and Operations in Base Ten	15–21	16	25%–34%	26%
		Number and Operations – Fractions	22–28	23	36%–46%	38%
		Measurement and Data	12–18	15	20%–30%	25%
		Geometry*	1–3	3	2%–5%	5%
6	72	Ratios and Proportional Relationships	16–20	17	24%–30%	25%
		The Number System	13–19	17	19%–28%	25%
		Expressions and Equations	23–33	23	34%–49%	34%
		Geometry	8–12	10	12%–18%	15%
7	72	Ratios and Proportional Relationships	18–22	20	26%–32%	29%
		The Number System	12–16	12	18%–24%	18%
		Expressions and Equations	19–25	21	28%–37%	31%
		Geometry	3–7	5	4%–10%	7%
		Statistics and Probability	8–14	10	12%–21%	15%
8	72	Expressions and Equations	26–34	28	38%–50%	41%
		Functions	16–22	19	24%–32%	28%
		Geometry	14–20	15	21%–29%	22%
		Statistics and Probability	5–7	6	7%–10%	9%

\*There is a slight difference between the “Target% of Test” shown in these tables and the tables presented in the Guides to the 2016 Common Core 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 Common Core mathematics standards.

## **Appendix C: Passage Selection Guidelines for Assessing ELA**

---

### **General Guidelines**

Along with instructional materials and teacher training, assessment development is essential to the successful implementation of the CCSS. While many of the expectations outlined in the CCSS align with previous versions of the New York State Learning Standards for ELA, the CCSS do represent some shifts in emphasis with direct implications for assessment development. In particular, the CCSS 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.

One of the major shifts of the CCSS is an emphasis on developing skills for comprehending and analyzing 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 this range of informational texts.

Another shift is an increased emphasis on the analysis across multiple texts, often of varied genres and media. Several standards, especially for reading literature, require intertextual and multi-media analysis. These expectations require special attention to the selection of related passages, chosen specifically to support the assessment of the full range of expectations. It will also require careful consideration of which standards are appropriate for large-scale assessment formats, and how these assessments might be modified to include passages of a variety of media.

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 CCSS 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	
<b>A.</b>	<b>Precisely Designed Constructs</b>
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>B.</b>	<b>Language Appropriateness</b>
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.
<b>C.</b>	<b>Gender Stereotypes</b>
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.
<b>D.</b>	<b>Ethnic Stereotypes</b>
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.
<b>E.</b>	<b>Cultural Familiarity</b>
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.

## Appendix D: Universal Design Item Checklist

Universal Design Item Checklist	
√	The item is free from unnecessary cultural references.
√	The item is free from religious references.
<b>F.</b>	<b>Geographic Bias</b>
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.
<b>G.</b>	<b>Disability Bias</b>
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).
<b>H.</b>	<b>Art Supports Text</b>
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).
<b>I.</b>	<b>Special Populations Considerations</b>
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.

## Appendix D: Universal Design Item Checklist

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 CCSS 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 2016 Common Core 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 2016 item selection included 2013, 2014, and 2015 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 2015 test)
  - Vertically 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 2015 operational test as a reference.

## Appendix G: Operational Item Maps

The following tables show the operational item maps for the 2016 NYSTP Grades 3–8 Common Core ELA and Mathematics Tests. External linking and field test items (i.e., those not contributing 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.RL.3.5
2	MC	1	CCSS.ELA-Literacy.RL.3.1
3	MC	1	CCSS.ELA-Literacy.RL.3.1
4	MC	1	CCSS.ELA-Literacy.RL.3.4
5	MC	1	CCSS.ELA-Literacy.RL.3.2
6	MC	1	CCSS.ELA-Literacy.RL.3.3
13	MC	1	CCSS.ELA-Literacy.RL.3.2
14	MC	1	CCSS.ELA-Literacy.RL.3.5
15	MC	1	CCSS.ELA-Literacy.RL.3.4
16	MC	1	CCSS.ELA-Literacy.RL.3.1
17	MC	1	CCSS.ELA-Literacy.RL.3.3
18	MC	1	CCSS.ELA-Literacy.RL.3.1
19	MC	1	CCSS.ELA-Literacy.RI.3.7
20	MC	1	CCSS.ELA-Literacy.RI.3.8
21	MC	1	CCSS.ELA-Literacy.RI.3.4
22	MC	1	CCSS.ELA-Literacy.RI.3.3
23	MC	1	CCSS.ELA-Literacy.RI.3.3
24	MC	1	CCSS.ELA-Literacy.RI.3.2
25	MC	1	CCSS.ELA-Literacy.L.3.4a
26	MC	1	CCSS.ELA-Literacy.RI.3.1
27	MC	1	CCSS.ELA-Literacy.RI.3.4
28	MC	1	CCSS.ELA-Literacy.RI.3.2
29	MC	1	CCSS.ELA-Literacy.RI.3.1
30	MC	1	CCSS.ELA-Literacy.RI.3.8
31	MC	1	CCSS.ELA-Literacy.RI.3.5
32	CR	2	CCSS.ELA-Literacy.RI.3.8
33	CR	2	CCSS.ELA-Literacy.RI.3.6
34	CR	4	CCSS.ELA-Literacy.W.3.2, CCSS.ELA-Literacy.RL.3.3
35	CR	2	CCSS.ELA-Literacy.RL.3.3
36	CR	2	CCSS.ELA-Literacy.RL.3.5
37	CR	2	CCSS.ELA-Literacy.RL.3.3

Item	Type	Points	Standard
38	CR	2	CCSS.ELA-Literacy.RL.3.3
39	CR	2	CCSS.ELA-Literacy.RI.3.1
40	CR	4	CCSS.ELA-Literacy.W.3.2, CCSS.ELA-Literacy.RI.3.3

**Table G2. ELA Grade 4 Operational Item Map**

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.4.1
2	MC	1	CCSS.ELA-Literacy.RL.4.1
3	MC	1	CCSS.ELA-Literacy.L.4.5a
4	MC	1	CCSS.ELA-Literacy.RL.4.2
5	MC	1	CCSS.ELA-Literacy.RL.4.1
6	MC	1	CCSS.ELA-Literacy.RL.4.3
13	MC	1	CCSS.ELA-Literacy.RI.4.8
14	MC	1	CCSS.ELA-Literacy.RI.4.8
15	MC	1	CCSS.ELA-Literacy.RI.4.1
16	MC	1	CCSS.ELA-Literacy.RI.4.3
17	MC	1	CCSS.ELA-Literacy.RI.4.3
18	MC	1	CCSS.ELA-Literacy.RI.4.5
19	MC	1	CCSS.ELA-Literacy.RI.4.8
20	MC	1	CCSS.ELA-Literacy.RI.4.4
21	MC	1	CCSS.ELA-Literacy.RI.4.3
22	MC	1	CCSS.ELA-Literacy.RI.4.2
23	MC	1	CCSS.ELA-Literacy.RI.4.2
24	MC	1	CCSS.ELA-Literacy.RI.4.5
25	MC	1	CCSS.ELA-Literacy.RL.4.5
26	MC	1	CCSS.ELA-Literacy.RL.4.4
27	MC	1	CCSS.ELA-Literacy.RL.4.1
28	MC	1	CCSS.ELA-Literacy.RL.4.1
29	MC	1	CCSS.ELA-Literacy.RL.4.5
30	MC	1	CCSS.ELA-Literacy.RL.4.3
31	MC	1	CCSS.ELA-Literacy.RL.4.2
32	CR	2	CCSS.ELA-Literacy.RI.4.7
33	CR	2	CCSS.ELA-Literacy.RI.4.6
34	CR	4	CCSS.ELA-Literacy.W.4.2, CCSS.ELA-Literacy.W.4.9, CCSS.ELA-Literacy.RL.4.3
35	CR	2	CCSS.ELA-Literacy.RL.4.2
36	CR	2	CCSS.ELA-Literacy.RL.4.3
37	CR	2	CCSS.ELA-Literacy.RI.4.1

Item	Type	Points	Standard
38	CR	2	CCSS.ELA-Literacy.RI.4.2
39	CR	2	CCSS.ELA-Literacy.RI.4.1
40	CR	4	CCSS.ELA-Literacy.W.4.2, CCSS.ELA-Literacy.W.4.9, CCSS.ELA-Literacy.RI.4.9

**Table G3. ELA Grade 5 Operational Item Map**

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.5.1
2	MC	1	CCSS.ELA-Literacy.RI.5.2
3	MC	1	CCSS.ELA-Literacy.RI.5.1
4	MC	1	CCSS.ELA-Literacy.RI.5.3
5	MC	1	CCSS.ELA-Literacy.L.5.5b
6	MC	1	CCSS.ELA-Literacy.RI.5.3
7	MC	1	CCSS.ELA-Literacy.RI.5.2
8	MC	1	CCSS.ELA-Literacy.RL.5.4
9	MC	1	CCSS.ELA-Literacy.RL.5.3
10	MC	1	CCSS.ELA-Literacy.RL.5.1
11	MC	1	CCSS.ELA-Literacy.RL.5.5
12	MC	1	CCSS.ELA-Literacy.RL.5.1
13	MC	1	CCSS.ELA-Literacy.RL.5.6
14	MC	1	CCSS.ELA-Literacy.RL.5.2
15	MC	1	CCSS.ELA-Literacy.RL.5.5
16	MC	1	CCSS.ELA-Literacy.L.5.4a
17	MC	1	CCSS.ELA-Literacy.RL.5.3
18	MC	1	CCSS.ELA-Literacy.RL.5.3
19	MC	1	CCSS.ELA-Literacy.RL.5.1
20	MC	1	CCSS.ELA-Literacy.RL.5.3
21	MC	1	CCSS.ELA-Literacy.RL.5.2
29	MC	1	CCSS.ELA-Literacy.RI.5.8
30	MC	1	CCSS.ELA-Literacy.RI.5.2
31	MC	1	CCSS.ELA-Literacy.RI.5.8
32	MC	1	CCSS.ELA-Literacy.RI.5.8
33	MC	1	CCSS.ELA-Literacy.RI.5.4
34	MC	1	CCSS.ELA-Literacy.RI.5.1
35	MC	1	CCSS.ELA-Literacy.RI.5.4
36	MC	1	CCSS.ELA-Literacy.RI.5.8
37	MC	1	CCSS.ELA-Literacy.RI.5.1
38	MC	1	CCSS.ELA-Literacy.RI.5.5
39	MC	1	CCSS.ELA-Literacy.RI.5.1

Item	Type	Points	Standard
40	MC	1	CCSS.ELA-Literacy.RI.5.3
41	MC	1	CCSS.ELA-Literacy.RI.5.2
42	MC	1	CCSS.ELA-Literacy.RI.5.2
43	CR	2	CCSS.ELA-Literacy.RI.5.2
44	CR	2	CCSS.ELA-Literacy.RI.5.2
45	CR	4	CCSS.ELA-Literacy.W.5.2, CCSS.ELA-Literacy.W.5.9, CCSS.ELA-Literacy.RI.5.8
46	CR	2	CCSS.ELA-Literacy.RL.5.3
47	CR	2	CCSS.ELA-Literacy.RL.5.5
48	CR	2	CCSS.ELA-Literacy.RL.5.3
49	CR	2	CCSS.ELA-Literacy.RL.5.4
50	CR	2	CCSS.ELA-Literacy.RL.5.2
51	CR	4	CCSS.ELA-Literacy.W.5.2, CCSS.ELA-Literacy.W.5.9, CCSS.ELA-Literacy.RL.5.3

**Table G4. ELA Grade 6 Operational Item Map**

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.6.5
2	MC	1	CCSS.ELA-Literacy.RL.6.4
3	MC	1	CCSS.ELA-Literacy.RL.6.3
4	MC	1	CCSS.ELA-Literacy.RL.6.2
5	MC	1	CCSS.ELA-Literacy.RL.6.1
6	MC	1	CCSS.ELA-Literacy.RL.6.3
7	MC	1	CCSS.ELA-Literacy.RL.6.6
8	MC	1	CCSS.ELA-Literacy.RI.6.4
9	MC	1	CCSS.ELA-Literacy.RI.6.3
10	MC	1	CCSS.ELA-Literacy.RI.6.8
11	MC	1	CCSS.ELA-Literacy.RI.6.6
12	MC	1	CCSS.ELA-Literacy.RI.6.2
13	MC	1	CCSS.ELA-Literacy.RI.6.5
14	MC	1	CCSS.ELA-Literacy.RI.6.2
22	MC	1	CCSS.ELA-Literacy.RL.6.2
23	MC	1	CCSS.ELA-Literacy.RL.6.4
24	MC	1	CCSS.ELA-Literacy.L.6.4c
25	MC	1	CCSS.ELA-Literacy.RL.6.1
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.6.1

Item	Type	Points	Standard
29	MC	1	CCSS.ELA-Literacy.RI.6.5
30	MC	1	CCSS.ELA-Literacy.RI.6.1
31	MC	1	CCSS.ELA-Literacy.RI.6.4
32	MC	1	CCSS.ELA-Literacy.RI.6.3
33	MC	1	CCSS.ELA-Literacy.RI.6.8
34	MC	1	CCSS.ELA-Literacy.RI.6.2
35	MC	1	CCSS.ELA-Literacy.RI.6.6
36	MC	1	CCSS.ELA-Literacy.RI.6.3
37	MC	1	CCSS.ELA-Literacy.RI.6.4
38	MC	1	CCSS.ELA-Literacy.RI.6.1
39	MC	1	CCSS.ELA-Literacy.RI.6.2
40	MC	1	CCSS.ELA-Literacy.RI.6.5
41	MC	1	CCSS.ELA-Literacy.RI.6.8
42	MC	1	CCSS.ELA-Literacy.RI.6.5
43	CR	2	CCSS.ELA-Literacy.RL.6.2
44	CR	2	CCSS.ELA-Literacy.RL.6.3
45	CR	4	CCSS.ELA-Literacy.W.6.2, CCSS.ELA-Literacy.W.6.9, CCSS.ELA-Literacy.RL.6.3
46	CR	2	CCSS.ELA-Literacy.RL.6.3
47	CR	2	CCSS.ELA-Literacy.RL.6.5
48	CR	2	CCSS.ELA-Literacy.RI.6.2
49	CR	2	CCSS.ELA-Literacy.RI.6.5
50	CR	2	CCSS.ELA-Literacy.RI.6.6
51	CR	4	CCSS.ELA-Literacy.W.6.2, CCSS.ELA-Literacy.W.6.9, CCSS.ELA-Literacy.RI.6.3

**Table G5. ELA Grade 7 Operational Item Map**

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.7.5
2	MC	1	CCSS.ELA-Literacy.RI.7.8
3	MC	1	CCSS.ELA-Literacy.RI.7.1
4	MC	1	CCSS.ELA-Literacy.RI.7.3
5	MC	1	CCSS.ELA-Literacy.RI.7.3
6	MC	1	CCSS.ELA-Literacy.RI.7.2
7	MC	1	CCSS.ELA-Literacy.RI.7.5
8	MC	1	CCSS.ELA-Literacy.RI.7.6
9	MC	1	CCSS.ELA-Literacy.RI.7.3
10	MC	1	CCSS.ELA-Literacy.L.7.4a

Appendix G: Operational Item Maps

Item	Type	Points	Standard
11	MC	1	CCSS.ELA-Literacy.RI.7.2
12	MC	1	CCSS.ELA-Literacy.RI.7.1
13	MC	1	CCSS.ELA-Literacy.RI.7.8
14	MC	1	CCSS.ELA-Literacy.RI.7.8
15	MC	1	CCSS.ELA-Literacy.RL.7.1
16	MC	1	CCSS.ELA-Literacy.RL.7.5
17	MC	1	CCSS.ELA-Literacy.RL.7.1
18	MC	1	CCSS.ELA-Literacy.RL.7.3
19	MC	1	CCSS.ELA-Literacy.RL.7.4
20	MC	1	CCSS.ELA-Literacy.RL.7.3
21	MC	1	CCSS.ELA-Literacy.RL.7.3
29	MC	1	CCSS.ELA-Literacy.RL.7.5
30	MC	1	CCSS.ELA-Literacy.RL.7.1
31	MC	1	CCSS.ELA-Literacy.RL.7.3
32	MC	1	CCSS.ELA-Literacy.RL.7.2
33	MC	1	CCSS.ELA-Literacy.RL.7.6
34	MC	1	CCSS.ELA-Literacy.RL.7.4
35	MC	1	CCSS.ELA-Literacy.RL.7.2
36	MC	1	CCSS.ELA-Literacy.RI.7.2
37	MC	1	CCSS.ELA-Literacy.RI.7.4
38	MC	1	CCSS.ELA-Literacy.RI.7.2
39	MC	1	CCSS.ELA-Literacy.RI.7.3
40	MC	1	CCSS.ELA-Literacy.RI.7.5
41	MC	1	CCSS.ELA-Literacy.RI.7.1
42	MC	1	CCSS.ELA-Literacy.RI.7.1
43	CR	2	CCSS.ELA-Literacy.RL.7.3
44	CR	2	CCSS.ELA-Literacy.RL.7.3
45	CR	4	CCSS.ELA-Literacy.W.7.2, CCSS.ELA-Literacy.W.7.9, CCSS.ELA-Literacy.RI.7.2
46	CR	2	CCSS.ELA-Literacy.RI.7.7
47	CR	2	CCSS.ELA-Literacy.RI.7.3
48	CR	2	CCSS.ELA-Literacy.RL.7.2
49	CR	2	CCSS.ELA-Literacy.RL.7.5
50	CR	2	CCSS.ELA-Literacy.RL.7.6
51	CR	4	CCSS.ELA-Literacy.W.7.2, CCSS.ELA-Literacy.W.7.9, CCSS.ELA-Literacy.RL.7.9

**Table G6. ELA Grade 8 Operational Item Map**

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.8.1
2	MC	1	CCSS.ELA-Literacy.RL.8.1
3	MC	1	CCSS.ELA-Literacy.RL.8.1
4	MC	1	CCSS.ELA-Literacy.RL.8.3
5	MC	1	CCSS.ELA-Literacy.RL.8.3
6	MC	1	CCSS.ELA-Literacy.RL.8.5
7	MC	1	CCSS.ELA-Literacy.RL.8.2
8	MC	1	CCSS.ELA-Literacy.RI.8.5
9	MC	1	CCSS.ELA-Literacy.L.8.4
10	MC	1	CCSS.ELA-Literacy.RI.8.4
11	MC	1	CCSS.ELA-Literacy.RI.8.3
12	MC	1	CCSS.ELA-Literacy.RI.8.6
13	MC	1	CCSS.ELA-Literacy.RI.8.8
14	MC	1	CCSS.ELA-Literacy.RI.8.3
22	MC	1	CCSS.ELA-Literacy.RL.8.3
23	MC	1	CCSS.ELA-Literacy.RL.8.4
24	MC	1	CCSS.ELA-Literacy.RL.8.1
25	MC	1	CCSS.ELA-Literacy.RL.8.3
26	MC	1	CCSS.ELA-Literacy.RL.8.6
27	MC	1	CCSS.ELA-Literacy.RL.8.6
28	MC	1	CCSS.ELA-Literacy.RL.8.2
29	MC	1	CCSS.ELA-Literacy.RI.8.4
30	MC	1	CCSS.ELA-Literacy.RI.8.1
31	MC	1	CCSS.ELA-Literacy.RI.8.3
32	MC	1	CCSS.ELA-Literacy.RI.8.3
33	MC	1	CCSS.ELA-Literacy.RI.8.8
34	MC	1	CCSS.ELA-Literacy.RI.8.5
35	MC	1	CCSS.ELA-Literacy.RI.8.2
36	MC	1	CCSS.ELA-Literacy.RI.8.3
37	MC	1	CCSS.ELA-Literacy.RI.8.5
38	MC	1	CCSS.ELA-Literacy.RI.8.4
39	MC	1	CCSS.ELA-Literacy.RI.8.1
40	MC	1	CCSS.ELA-Literacy.RI.8.7
41	MC	1	CCSS.ELA-Literacy.RI.8.2
42	MC	1	CCSS.ELA-Literacy.RI.8.2
43	CR	2	CCSS.ELA-Literacy.RL.8.3
44	CR	2	CCSS.ELA-Literacy.RL.8.3

Item	Type	Points	Standard
45	CR	4	CCSS.ELA-Literacy.W.8.2, CCSS.ELA-Literacy.W.8.9, CCSS.ELA-Literacy.RL.8.3
46	CR	2	CCSS.ELA-Literacy.RL.8.4
47	CR	2	CCSS.ELA-Literacy.RL.8.6
48	CR	2	CCSS.ELA-Literacy.RI.8.2
49	CR	2	CCSS.ELA-Literacy.RI.8.1
50	CR	2	CCSS.ELA-Literacy.RI.8.4
51	CR	4	CCSS.ELA-Literacy.W.8.2, CCSS.ELA-Literacy.W.8.9, CCSS.ELA-Literacy.RI.8.8

**Table G7. Mathematics Grade 3 Operational Item Map**

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.3.MD.A.1
2	MC	1	CCSS.Math.Content.3.NBT.A.1
3	MC	1	CCSS.Math.Content.3.NF.A.3c
4	MC	1	CCSS.Math.Content.3.G.A.2
6	MC	1	CCSS.Math.Content.3.OA.A.3
7	MC	1	CCSS.Math.Content.3.NBT.A.3
8	MC	1	CCSS.Math.Content.3.OA.A.4
9	MC	1	CCSS.Math.Content.3.MD.A.1
11	MC	1	CCSS.Math.Content.3.MD.C.6
12	MC	1	CCSS.Math.Content.3.OA.D.9
13	MC	1	CCSS.Math.Content.3.OA.B.6
14	MC	1	CCSS.Math.Content.3.MD.C.7d
16	MC	1	CCSS.Math.Content.3.MD.A.2
17	MC	1	CCSS.Math.Content.3.OA.D.8
19	MC	1	CCSS.Math.Content.3.OA.A.3
20	MC	1	CCSS.Math.Content.3.NF.A.1
21	MC	1	CCSS.Math.Content.3.OA.A.1
22	MC	1	CCSS.Math.Content.3.NF.A.3a
23	MC	1	CCSS.Math.Content.3.OA.A.4
24	MC	1	CCSS.Math.Content.3.NBT.A.3
25	MC	1	CCSS.Math.Content.3.OA.D.8
26	MC	1	CCSS.Math.Content.3.NF.A.1
27	MC	1	CCSS.Math.Content.3.OA.A.1
28	MC	1	CCSS.Math.Content.3.MD.C.5b
30	MC	1	CCSS.Math.Content.3.NF.A.2a
31	MC	1	CCSS.Math.Content.3.MD.C.6

Item	Type	Points	Standard
32	MC	1	CCSS.Math.Content.3.NBT.A.1
33	MC	1	CCSS.Math.Content.3.MD.A.2
34	MC	1	CCSS.Math.Content.3.G.A.2
35	MC	1	CCSS.Math.Content.3.OA.A.3
37	MC	1	CCSS.Math.Content.3.OA.B.6
38	MC	1	CCSS.Math.Content.3.MD.C.7a
39	MC	1	CCSS.Math.Content.3.OA.D.9
40	MC	1	CCSS.Math.Content.3.OA.A.3
41	MC	1	CCSS.Math.Content.3.NF.A.1
42	MC	1	CCSS.Math.Content.3.OA.D.8
43	MC	1	CCSS.Math.Content.3.MD.B.3
45	CR	2	CCSS.Math.Content.3.NF.A.2
46	CR	2	CCSS.Math.Content.3.OA.B.5
47	CR	2	CCSS.Math.Content.3.MD.B.3
48	CR	2	CCSS.Math.Content.3.OA.A.2
49	CR	2	CCSS.Math.Content.3.MD.C.7c
50	CR	3	CCSS.Math.Content.3.OA.A.3
51	CR	3	CCSS.Math.Content.3.NF.A.3b
52	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.NBT.A.2
2	MC	1	CCSS.Math.Content.4.OA.A.2
3	MC	1	CCSS.Math.Content.4.NF.A.1
4	MC	1	CCSS.Math.Content.4.NF.B.3c
5	MC	1	CCSS.Math.Content.4.NBT.A.1
6	MC	1	CCSS.Math.Content.4.OA.A.2
7	MC	1	CCSS.Math.Content.4.G.A.1
8	MC	1	CCSS.Math.Content.4.MD.C.5a
9	MC	1	CCSS.Math.Content.4.OA.A.3
10	MC	1	CCSS.Math.Content.4.NF.A.2
12	MC	1	CCSS.Math.Content.4.NBT.B.5
13	MC	1	CCSS.Math.Content.4.NF.B.4c
14	MC	1	CCSS.Math.Content.4.G.A.3
16	MC	1	CCSS.Math.Content.4.NBT.B.6
17	MC	1	CCSS.Math.Content.4.MD.C.6
18	MC	1	CCSS.Math.Content.4.NBT.A.1
19	MC	1	CCSS.Math.Content.3.MD.D.8

Item	Type	Points	Standard
20	MC	1	CCSS.Math.Content.4.G.A.1
23	MC	1	CCSS.Math.Content.4.NBT.B.5
24	MC	1	CCSS.Math.Content.4.G.A.1
25	MC	1	CCSS.Math.Content.4.NF.A.2
26	MC	1	CCSS.Math.Content.4.MD.C.5b
27	MC	1	CCSS.Math.Content.4.OA.C.5
28	MC	1	CCSS.Math.Content.4.MD.C.6
29	MC	1	CCSS.Math.Content.4.OA.A.1
30	MC	1	CCSS.Math.Content.4.NBT.B.6
31	MC	1	CCSS.Math.Content.4.NF.B.3a
32	MC	1	CCSS.Math.Content.4.NBT.B.5
33	MC	1	CCSS.Math.Content.4.MD.B.4
34	MC	1	CCSS.Math.Content.4.NF.B.4b
35	MC	1	CCSS.Math.Content.4.NBT.A.3
37	MC	1	CCSS.Math.Content.4.NF.A.1
38	MC	1	CCSS.Math.Content.4.OA.A.2
39	MC	1	CCSS.Math.Content.4.NBT.B.6
40	MC	1	CCSS.Math.Content.4.NBT.A.1
42	MC	1	CCSS.Math.Content.4.NF.B.4b
43	MC	1	CCSS.Math.Content.4.OA.B.4
45	MC	1	CCSS.Math.Content.4.NF.A.2
46	CR	2	CCSS.Math.Content.4.MD.A.3
47	CR	2	CCSS.Math.Content.4.NBT.A.2
48	CR	2	CCSS.Math.Content.4.NF.A.1
49	CR	2	CCSS.Math.Content.4.MD.C.7
50	CR	2	CCSS.Math.Content.4.NF.B.4c
51	CR	2	CCSS.Math.Content.4.G.A.2
52	CR	3	CCSS.Math.Content.4.OA.A.3
53	CR	3	CCSS.Math.Content.4.NF.B.3d
54	CR	3	CCSS.Math.Content.4.NBT.B.5
55	CR	3	CCSS.Math.Content.4.OA.A.2

**Table G9. Mathematics Grade 5 Operational Item Map**

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.5.NBT.B.7
2	MC	1	CCSS.Math.Content.5.NF.A.1
3	MC	1	CCSS.Math.Content.5.NBT.B.6
4	MC	1	CCSS.Math.Content.5.NF.A.2

Appendix G: Operational Item Maps

Item	Type	Points	Standard
5	MC	1	CCSS.Math.Content.5.G.B.4
6	MC	1	CCSS.Math.Content.4.MD.A.2
8	MC	1	CCSS.Math.Content.5.NBT.A.1
9	MC	1	CCSS.Math.Content.5.NF.B.7b
10	MC	1	CCSS.Math.Content.5.MD.C.3b
11	MC	1	CCSS.Math.Content.4.NF.C.5
13	MC	1	CCSS.Math.Content.5.NF.B.4a
14	MC	1	CCSS.Math.Content.5.MD.C.4
15	MC	1	CCSS.Math.Content.5.MD.B.2
16	MC	1	CCSS.Math.Content.5.MD.A.1
17	MC	1	CCSS.Math.Content.4.NF.C.7
18	MC	1	CCSS.Math.Content.5.NF.B.3
19	MC	1	CCSS.Math.Content.5.MD.A.1
20	MC	1	CCSS.Math.Content.5.NF.B.6
23	MC	1	CCSS.Math.Content.5.OA.A.1
24	MC	1	CCSS.Math.Content.5.G.B.4
25	MC	1	CCSS.Math.Content.4.NF.C.6
26	MC	1	CCSS.Math.Content.5.NBT.B.6
27	MC	1	CCSS.Math.Content.5.NF.B.4a
28	MC	1	CCSS.Math.Content.5.NBT.A.2
29	MC	1	CCSS.Math.Content.4.MD.A.1
31	MC	1	CCSS.Math.Content.5.NBT.B.6
33	MC	1	CCSS.Math.Content.5.MD.C.4
34	MC	1	CCSS.Math.Content.5.NF.B.5b
36	MC	1	CCSS.Math.Content.5.G.B.3
37	MC	1	CCSS.Math.Content.5.NF.B.3
39	MC	1	CCSS.Math.Content.5.NBT.A.4
40	MC	1	CCSS.Math.Content.5.NF.B.4b
41	MC	1	CCSS.Math.Content.5.MD.C.5b
42	MC	1	CCSS.Math.Content.5.NF.A.2
43	MC	1	CCSS.Math.Content.5.MD.B.2
44	MC	1	CCSS.Math.Content.5.NF.B.6
45	MC	1	CCSS.Math.Content.5.OA.A.1
46	CR	2	CCSS.Math.Content.5.NBT.A.3
47	CR	2	CCSS.Math.Content.5.NF.B.7c
48	CR	2	CCSS.Math.Content.5.NBT.B.6
49	CR	2	CCSS.Math.Content.5.NF.B.5b
50	CR	2	CCSS.Math.Content.5.MD.A.1

Item	Type	Points	Standard
51	CR	2	CCSS.Math.Content.5.OA.A.2
52	CR	3	CCSS.Math.Content.5.NF.A.2
53	CR	3	CCSS.Math.Content.5.NBT.B.7
54	CR	3	CCSS.Math.Content.5.NF.B.6
55	CR	3	CCSS.Math.Content.5.MD.C.5b

**Table G10. Mathematics Grade 6 Operational Item Map**

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.6.EE.B.6
2	MC	1	CCSS.Math.Content.5.G.A.1
4	MC	1	CCSS.Math.Content.6.RP.A.3b
5	MC	1	CCSS.Math.Content.6.NS.B.4
7	MC	1	CCSS.Math.Content.5.OA.B.3
8	MC	1	CCSS.Math.Content.6.G.A.4
9	MC	1	CCSS.Math.Content.6.G.A.2
11	MC	1	CCSS.Math.Content.6.EE.C.9
12	MC	1	CCSS.Math.Content.6.EE.A.4
13	MC	1	CCSS.Math.Content.6.NS.A.1
14	MC	1	CCSS.Math.Content.6.NS.C.6c
15	MC	1	CCSS.Math.Content.6.RP.A.3d
16	MC	1	CCSS.Math.Content.6.EE.B.8
17	MC	1	CCSS.Math.Content.6.NS.A.1
18	MC	1	CCSS.Math.Content.6.NS.C.6a
19	MC	1	CCSS.Math.Content.6.EE.C.9
20	MC	1	CCSS.Math.Content.6.RP.A.3a
21	MC	1	CCSS.Math.Content.6.EE.B.6
22	MC	1	CCSS.Math.Content.6.EE.A.2a
24	MC	1	CCSS.Math.Content.6.EE.A.2b
26	MC	1	CCSS.Math.Content.6.EE.A.3
27	MC	1	CCSS.Math.Content.6.RP.A.2
28	MC	1	CCSS.Math.Content.6.RP.A.3b
29	MC	1	CCSS.Math.Content.6.EE.B.7
30	MC	1	CCSS.Math.Content.6.G.A.1
31	MC	1	CCSS.Math.Content.6.EE.B.7
33	MC	1	CCSS.Math.Content.6.G.A.3
34	MC	1	CCSS.Math.Content.6.RP.A.3a
35	MC	1	CCSS.Math.Content.6.EE.A.4

Item	Type	Points	Standard
36	MC	1	CCSS.Math.Content.6.RP.A.1
37	MC	1	CCSS.Math.Content.6.NS.C.5
38	MC	1	CCSS.Math.Content.6.EE.C.9
39	MC	1	CCSS.Math.Content.6.RP.A.3d
40	MC	1	CCSS.Math.Content.6.NS.C.7a
41	MC	1	CCSS.Math.Content.6.EE.C.9
42	MC	1	CCSS.Math.Content.6.G.A.3
43	MC	1	CCSS.Math.Content.6.RP.A.3c
44	MC	1	CCSS.Math.Content.6.NS.A.1
45	MC	1	CCSS.Math.Content.6.G.A.4
46	MC	1	CCSS.Math.Content.6.EE.A.2a
47	MC	1	CCSS.Math.Content.6.RP.A.3a
48	MC	1	CCSS.Math.Content.6.EE.B.5
49	MC	1	CCSS.Math.Content.6.RP.A.3b
52	CR	2	CCSS.Math.Content.6.NS.C.8
53	CR	2	CCSS.Math.Content.6.NS.B.4
54	CR	2	CCSS.Math.Content.6.EE.A.1
55	CR	2	CCSS.Math.Content.6.G.A.1
56	CR	2	CCSS.Math.Content.6.NS.C.8
57	CR	2	CCSS.Math.Content.6.G.A.2
58	CR	3	CCSS.Math.Content.6.EE.A.3
59	CR	3	CCSS.Math.Content.6.EE.B.7
60	CR	3	CCSS.Math.Content.6.RP.A.2
61	CR	3	CCSS.Math.Content.6.RP.A.3c

**Table G11. Mathematics Grade 7 Operational Item Map**

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.7.G.A.1
2	MC	1	CCSS.Math.Content.7.NS.A.1d
4	MC	1	CCSS.Math.Content.7.RP.A.1
6	MC	1	CCSS.Math.Content.7.EE.A.1
7	MC	1	CCSS.Math.Content.7.RP.A.3
8	MC	1	CCSS.Math.Content.7.EE.B.4b
9	MC	1	CCSS.Math.Content.7.SP.B.3
10	MC	1	CCSS.Math.Content.7.NS.A.2c
11	MC	1	CCSS.Math.Content.7.EE.B.4a
12	MC	1	CCSS.Math.Content.7.SP.A.1
13	MC	1	CCSS.Math.Content.7.SP.C.8a

Appendix G: Operational Item Maps

Item	Type	Points	Standard
14	MC	1	CCSS.Math.Content.7.RP.A.2c
15	MC	1	CCSS.Math.Content.7.NS.A.3
16	MC	1	CCSS.Math.Content.7.SP.C.5
17	MC	1	CCSS.Math.Content.7.EE.A.2
18	MC	1	CCSS.Math.Content.7.RP.A.3
20	MC	1	CCSS.Math.Content.7.EE.B.4b
21	MC	1	CCSS.Math.Content.7.NS.A.2c
22	MC	1	CCSS.Math.Content.7.EE.B.3
23	MC	1	CCSS.Math.Content.7.RP.A.2a
24	MC	1	CCSS.Math.Content.7.EE.A.1
25	MC	1	CCSS.Math.Content.7.EE.B.4a
27	MC	1	CCSS.Math.Content.7.NS.A.1c
28	MC	1	CCSS.Math.Content.7.NS.A.2b
29	MC	1	CCSS.Math.Content.7.EE.A.1
30	MC	1	CCSS.Math.Content.7.RP.A.3
31	MC	1	CCSS.Math.Content.7.EE.A.1
33	MC	1	CCSS.Math.Content.7.EE.B.4a
34	MC	1	CCSS.Math.Content.7.EE.A.2
35	MC	1	CCSS.Math.Content.7.RP.A.3
36	MC	1	CCSS.Math.Content.7.SP.C.6
37	MC	1	CCSS.Math.Content.7.RP.A.1
38	MC	1	CCSS.Math.Content.7.NS.A.3
39	MC	1	CCSS.Math.Content.7.RP.A.2a
40	MC	1	CCSS.Math.Content.7.EE.A.1
41	MC	1	CCSS.Math.Content.7.RP.A.2b
42	MC	1	CCSS.Math.Content.7.EE.A.2
43	MC	1	CCSS.Math.Content.7.RP.A.1
44	MC	1	CCSS.Math.Content.7.EE.B.4a
45	MC	1	CCSS.Math.Content.7.RP.A.3
46	MC	1	CCSS.Math.Content.7.G.A.1
47	MC	1	CCSS.Math.Content.7.EE.B.3
48	MC	1	CCSS.Math.Content.7.SP.B.4
49	MC	1	CCSS.Math.Content.7.G.B.4
52	CR	2	CCSS.Math.Content.7.SP.C.6
53	CR	2	CCSS.Math.Content.7.RP.A.3
54	CR	2	CCSS.Math.Content.7.EE.B.4a
55	CR	2	CCSS.Math.Content.7.SP.A.2
56	CR	2	CCSS.Math.Content.7.G.B.4

Item	Type	Points	Standard
57	CR	2	CCSS.Math.Content.7.NS.A.3
58	CR	3	CCSS.Math.Content.7.RP.A.2a
59	CR	3	CCSS.Math.Content.7.EE.B.3
60	CR	3	CCSS.Math.Content.7.RP.A.3
61	CR	3	CCSS.Math.Content.7.NS.A.3

**Table G12. Mathematics Grade 8 Operational Item Map**

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.8.EE.C.8c
2	MC	1	CCSS.Math.Content.8.F.B.4
3	MC	1	CCSS.Math.Content.8.EE.A.3
4	MC	1	CCSS.Math.Content.8.G.A.2
5	MC	1	CCSS.Math.Content.8.EE.C.8b
6	MC	1	CCSS.Math.Content.8.G.C.9
7	MC	1	CCSS.Math.Content.8.F.A.3
8	MC	1	CCSS.Math.Content.8.SP.A.1
9	MC	1	CCSS.Math.Content.8.EE.B.5
10	MC	1	CCSS.Math.Content.8.F.A.3
11	MC	1	CCSS.Math.Content.8.EE.A.1
12	MC	1	CCSS.Math.Content.8.EE.C.7b
15	MC	1	CCSS.Math.Content.8.EE.B.6
16	MC	1	CCSS.Math.Content.8.F.A.2
17	MC	1	CCSS.Math.Content.8.SP.A.3
19	MC	1	CCSS.Math.Content.8.EE.A.3
20	MC	1	CCSS.Math.Content.8.G.A.4
21	MC	1	CCSS.Math.Content.8.F.A.2
22	MC	1	CCSS.Math.Content.8.G.A.1a
24	MC	1	CCSS.Math.Content.8.F.B.5
25	MC	1	CCSS.Math.Content.8.EE.A.4
26	MC	1	CCSS.Math.Content.8.F.A.1
27	MC	1	CCSS.Math.Content.8.EE.C.8b
28	MC	1	CCSS.Math.Content.8.G.A.3
29	MC	1	CCSS.Math.Content.8.EE.A.3
30	MC	1	CCSS.Math.Content.8.F.A.1
32	MC	1	CCSS.Math.Content.8.F.B.4
33	MC	1	CCSS.Math.Content.8.EE.B.6
34	MC	1	CCSS.Math.Content.8.SP.A.4
35	MC	1	CCSS.Math.Content.8.G.C.9

Appendix G: Operational Item Maps

Item	Type	Points	Standard
36	MC	1	CCSS.Math.Content.8.EE.B.5
37	MC	1	CCSS.Math.Content.8.F.A.3
38	MC	1	CCSS.Math.Content.8.EE.A.4
39	MC	1	CCSS.Math.Content.8.F.B.4
40	MC	1	CCSS.Math.Content.8.F.A.2
41	MC	1	CCSS.Math.Content.8.SP.A.2
42	MC	1	CCSS.Math.Content.8.EE.C.7b
44	MC	1	CCSS.Math.Content.8.G.C.9
45	MC	1	CCSS.Math.Content.8.F.B.5
46	MC	1	CCSS.Math.Content.8.EE.C.8a
47	MC	1	CCSS.Math.Content.8.G.A.5
48	MC	1	CCSS.Math.Content.8.EE.B.6
49	MC	1	CCSS.Math.Content.8.F.A.2
50	MC	1	CCSS.Math.Content.8.EE.C.8b
52	CR	2	CCSS.Math.Content.8.EE.A.1
53	CR	2	CCSS.Math.Content.8.G.A.2
54	CR	2	CCSS.Math.Content.8.F.A.3
55	CR	2	CCSS.Math.Content.8.EE.C.7a
56	CR	2	CCSS.Math.Content.8.SP.A.3
57	CR	2	CCSS.Math.Content.8.G.A.3
58	CR	3	CCSS.Math.Content.8.EE.B.5
59	CR	3	CCSS.Math.Content.8.F.B.4
60	CR	3	CCSS.Math.Content.8.G.A.4
61	CR	3	CCSS.Math.Content.8.EE.C.8c

## Appendix H: ELA Short-Response Rubric

---

### 2-Point Rubric–Short Response

Score	Response Features
<b>2 Point</b>	<p>The features of a 2-point response are</p> <ul style="list-style-type: none"> <li>• Valid inferences and/or claims from the text where required by the prompt</li> <li>• Evidence of analysis of the text where required by the prompt</li> <li>• Relevant facts, definitions, concrete details, and/or other information from the text to develop response according to the requirements of the prompt</li> <li>• Sufficient number of facts, definitions, concrete details, and/or other information from the text as required by the prompt</li> <li>• Complete sentences where errors do not impact readability</li> </ul>
<b>1 Point</b>	<p>The features of a 1-point response are</p> <ul style="list-style-type: none"> <li>• A mostly literal recounting of events or details from the text as required by the prompt</li> <li>• Some relevant facts, definitions, concrete details, and/or other information from the text to develop response according to the requirements of the prompt</li> <li>• Incomplete sentences or bullets</li> </ul>
<b>0 Point*</b>	<p>The features of a 0-point response are</p> <ul style="list-style-type: none"> <li>• A response that does not address any of the requirements of the prompt or is totally inaccurate</li> <li>• A response that is not written in English</li> <li>• A response that is unintelligible or indecipherable</li> </ul>

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

\* 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	<p>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.</p> <p>This response</p> <ul style="list-style-type: none"> <li>• indicates that the student has completed the task correctly, using mathematically sound procedures</li> <li>• contains sufficient work to demonstrate a thorough understanding of the mathematical concepts and/or procedures</li> <li>• may contain inconsequential errors that do not detract from the correct solution and the demonstration of a thorough understanding</li> </ul>
1 Point	<p>A one-point response demonstrates only a partial understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response</p> <ul style="list-style-type: none"> <li>• correctly addresses only some elements of the task</li> <li>• may contain an incorrect solution but applies a mathematically appropriate process</li> <li>• may contain the correct solution but required work is incomplete</li> </ul>
0 Points*	<p>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.</p>

\* 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	<p>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.</p> <p>This response</p> <ul style="list-style-type: none"> <li>• indicates that the student has completed the task correctly, using mathematically sound procedures</li> <li>• contains sufficient work to demonstrate a thorough understanding of the mathematical concepts and/or procedures</li> <li>• may contain inconsequential errors that do not detract from the correct solution(s) and the demonstration of a thorough understanding</li> </ul>
2 Points	<p>A two-point response demonstrates a partial understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response</p> <ul style="list-style-type: none"> <li>• appropriately addresses most, but not all, aspects of the task using mathematically sound procedures</li> <li>• may contain an incorrect solution but provides sound procedures, reasoning, and/or explanations</li> <li>• may reflect some minor misunderstanding of the underlying mathematical concepts and/or procedures</li> </ul>
1 Point	<p>A one-point response demonstrates only a limited understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response</p> <ul style="list-style-type: none"> <li>• may address some elements of the task correctly but reaches an inadequate solution and/or provides reasoning that is faulty or incomplete</li> <li>• exhibits multiple flaws related to misunderstanding of important aspects of the task, misuse of mathematical procedures, or faulty mathematical reasoning</li> <li>• reflects a lack of essential understanding of the underlying mathematical concepts</li> <li>• may contain the correct solution(s) but required work is limited</li> </ul>
0 Points*	<p>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.</p>

\* 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 Common Core ELA and Mathematics Tests data. The analyses were conducted for the total population of students and select subgroups: ELL, SWD, SUA, SWD students using disability accommodations (SWD & SUA), and ELL 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 Common Core ELA and Mathematics Tests, respectively.

**Table L1. ELA Grade 3 Test Factor Analysis by Subgroup**

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	<b>1</b>	<b>5.92</b>	<b>17.42</b>	<b>17.42</b>
		2	1.48	4.36	21.78
		3	1.23	3.62	25.40
		4	1.06	3.12	28.53
		5	1.04	3.05	31.58
		6	1.03	3.02	34.60
		7	1.01	2.98	37.58
		8	1.00	2.96	40.53
SWD	All Codes	<b>1</b>	<b>7.33</b>	<b>21.56</b>	<b>21.56</b>
		2	1.49	4.38	25.95
		3	1.21	3.54	29.49
		4	1.02	2.99	32.48
		5	1.01	2.96	35.44
SUA	All Codes	<b>1</b>	<b>7.19</b>	<b>21.14</b>	<b>21.14</b>
		2	1.50	4.42	25.57
		3	1.21	3.54	29.11
		4	1.03	3.03	32.14
		5	1.01	2.98	35.12
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>6.84</b>	<b>20.13</b>	<b>20.13</b>
		2	1.50	4.41	24.54
		3	1.21	3.55	28.09
		4	1.04	3.06	31.15
		5	1.03	3.03	34.18
		6	1.00	2.96	37.14

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
%	Cumulative %				
ELL/SUA	SUA & ELL Codes	<b>1</b>	<b>5.12</b>	<b>15.07</b>	<b>15.07</b>
		2	1.43	4.20	19.27
		3	1.21	3.57	22.84
		4	1.14	3.36	26.21
		5	1.12	3.30	29.50
		6	1.08	3.18	32.68
		7	1.07	3.14	35.82
		8	1.06	3.11	38.94
		9	1.02	2.99	41.92
		10	1.01	2.96	44.89

Table L2. ELA Grade 4 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
%	Cumulative %				
ELL	ELL=Y	<b>1</b>	<b>5.26</b>	<b>15.48</b>	<b>15.48</b>
		2	1.54	4.52	20.00
		3	1.13	3.31	23.31
		4	1.07	3.16	26.47
		5	1.06	3.13	29.60
		6	1.05	3.09	32.69
		7	1.04	3.05	35.73
		8	1.02	3.01	38.74
SWD	All Codes	<b>1</b>	<b>6.36</b>	<b>18.69</b>	<b>18.69</b>
		2	1.53	4.50	23.20
		3	1.09	3.20	26.39
		4	1.06	3.12	29.51
		5	1.04	3.07	32.58
		6	1.01	2.97	35.55
SUA	All Codes	<b>1</b>	<b>6.42</b>	<b>18.89</b>	<b>18.89</b>
		2	1.55	4.55	23.44
		3	1.08	3.17	26.61
		4	1.05	3.10	29.71
		5	1.04	3.06	32.77
		6	1.01	2.96	35.73

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>6.10</b>	<b>17.95</b>	<b>17.95</b>
		2	1.54	4.54	22.48
		3	1.09	3.22	25.70
		4	1.07	3.14	28.84
		5	1.05	3.09	31.94
		6	1.02	3.00	34.94
		7	1.01	2.97	37.91
ELL/SUA	SUA & ELL Codes	<b>1</b>	<b>4.71</b>	<b>13.87</b>	<b>13.87</b>
		2	1.48	4.35	18.22
		3	1.18	3.46	21.68
		4	1.15	3.39	25.07
		5	1.13	3.33	28.40
		6	1.10	3.22	31.63
		7	1.08	3.17	34.80
		8	1.06	3.11	37.91
		9	1.05	3.08	40.98
		10	1.02	3.00	43.98
		11	1.02	2.99	46.97

Table L3. ELA Grade 5 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	<b>1</b>	<b>6.44</b>	<b>14.64</b>	<b>14.64</b>
		2	1.69	3.83	18.48
		3	1.25	2.84	21.32
		4	1.13	2.56	23.88
		5	1.09	2.47	26.35
		6	1.08	2.45	28.80
		7	1.05	2.40	31.20
		8	1.04	2.36	33.56
		9	1.03	2.34	35.90
		10	1.02	2.33	38.23
		11	1.00	2.28	40.50

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD	All Codes	<b>1</b>	<b>7.78</b>	<b>17.69</b>	<b>17.69</b>
		2	1.73	3.93	21.61
		3	1.26	2.86	24.47
		4	1.10	2.49	26.96
		5	1.04	2.37	29.33
		6	1.02	2.32	31.65
		7	1.01	2.29	33.94
		8	1.00	2.28	36.22
SUA	All Codes	<b>1</b>	<b>7.98</b>	<b>18.14</b>	<b>18.14</b>
		2	1.73	3.93	22.06
		3	1.26	2.85	24.91
		4	1.09	2.48	27.40
		5	1.04	2.36	29.76
		6	1.02	2.31	32.07
		7	1.00	2.28	34.35
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>7.51</b>	<b>17.06</b>	<b>17.06</b>
		2	1.72	3.92	20.97
		3	1.25	2.84	23.82
		4	1.11	2.52	26.33
		5	1.05	2.39	28.73
		6	1.03	2.35	31.07
		7	1.01	2.31	33.38
		8	1.01	2.29	35.67
		9	1.00	2.28	37.95
ELL/SUA	SUA & ELL Codes	<b>1</b>	<b>5.62</b>	<b>12.78</b>	<b>12.78</b>
		2	1.57	3.58	16.35
		3	1.24	2.82	19.17
		4	1.19	2.71	21.89
		5	1.17	2.65	24.53
		6	1.14	2.58	27.12
		7	1.11	2.53	29.64
		8	1.10	2.50	32.14
		9	1.09	2.47	34.61
		10	1.07	2.43	37.03
		11	1.03	2.35	39.38
		12	1.03	2.34	41.73
		13	1.02	2.33	44.06
		14	1.02	2.31	46.37
		15	1.01	2.30	48.67

Appendix L: Factor Analysis Results for Select Subgroups

**Table L4. ELA Grade 6 Test Factor Analysis by Subgroup**

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	<b>1</b>	<b>5.93</b>	<b>13.48</b>	<b>13.48</b>
		2	1.58	3.60	17.07
		3	1.19	2.70	19.78
		4	1.13	2.57	22.35
		5	1.12	2.55	24.90
		6	1.10	2.51	27.41
		7	1.10	2.50	29.91
		8	1.09	2.48	32.39
		9	1.08	2.45	34.84
		10	1.07	2.43	37.26
		11	1.06	2.42	39.68
		12	1.03	2.35	42.03
		13	1.02	2.31	44.34
		14	1.00	2.28	46.62
SWD	All Codes	<b>1</b>	<b>6.73</b>	<b>15.29</b>	<b>15.29</b>
		2	1.66	3.77	19.06
		3	1.16	2.64	21.70
		4	1.15	2.62	24.31
		5	1.08	2.45	26.77
		6	1.07	2.43	29.20
		7	1.05	2.40	31.59
		8	1.04	2.37	33.97
		9	1.03	2.35	36.32
		10	1.02	2.32	38.64
		11	1.02	2.31	40.94
SUA	All Codes	<b>1</b>	<b>7.00</b>	<b>15.91</b>	<b>15.91</b>
		2	1.67	3.79	19.70
		3	1.16	2.63	22.33
		4	1.15	2.61	24.94
		5	1.07	2.44	27.38
		6	1.06	2.41	29.79
		7	1.05	2.40	32.19
		8	1.04	2.36	34.54
		9	1.03	2.34	36.88
		10	1.02	2.31	39.20
		11	1.01	2.30	41.50

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>6.54</b>	<b>14.87</b>	<b>14.87</b>
		2	1.66	3.76	18.63
		3	1.16	2.64	21.27
		4	1.15	2.62	23.89
		5	1.08	2.45	26.35
		6	1.07	2.44	28.78
		7	1.06	2.42	31.20
		8	1.05	2.40	33.60
		9	1.04	2.36	35.96
		10	1.03	2.34	38.30
		11	1.03	2.33	40.64
ELL/SUA	SUA & ELL Codes	<b>1</b>	<b>5.01</b>	<b>11.39</b>	<b>11.39</b>
		2	1.49	3.39	14.77
		3	1.24	2.82	17.59
		4	1.21	2.75	20.35
		5	1.18	2.69	23.03
		6	1.16	2.65	25.68
		7	1.15	2.62	28.30
		8	1.13	2.57	30.87
		9	1.11	2.52	33.39
		10	1.10	2.51	35.90
		11	1.08	2.46	38.35
		12	1.07	2.44	40.79
		13	1.05	2.39	43.19
		14	1.04	2.36	45.55
		15	1.02	2.33	47.88
		16	1.02	2.32	50.20
		17	1.01	2.31	52.50

Appendix L: Factor Analysis Results for Select Subgroups

**Table L5. ELA Grade 7 Test Factor Analysis by Subgroup**

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	<b>1</b>	<b>5.74</b>	<b>13.06</b>	<b>13.06</b>
		2	1.66	3.76	16.82
		3	1.17	2.66	19.48
		4	1.12	2.55	22.03
		5	1.12	2.53	24.56
		6	1.09	2.48	27.04
		7	1.08	2.46	29.50
		8	1.08	2.45	31.95
		9	1.06	2.41	34.35
		10	1.05	2.38	36.73
		11	1.03	2.34	39.07
		12	1.03	2.33	41.41
		13	1.01	2.29	43.70
		14	1.00	2.28	45.98
SWD	All Codes	<b>1</b>	<b>7.12</b>	<b>16.18</b>	<b>16.18</b>
		2	1.71	3.88	20.06
		3	1.14	2.59	22.65
		4	1.09	2.47	25.12
		5	1.06	2.40	27.52
		6	1.04	2.37	29.89
		7	1.03	2.34	32.23
		8	1.02	2.31	34.54
		9	1.00	2.28	36.82
SUA	All Codes	<b>1</b>	<b>7.45</b>	<b>16.94</b>	<b>16.94</b>
		2	1.71	3.89	20.83
		3	1.14	2.59	23.42
		4	1.07	2.44	25.86
		5	1.05	2.38	28.24
		6	1.03	2.35	30.59
		7	1.02	2.33	32.91
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>6.94</b>	<b>15.78</b>	<b>15.78</b>
		2	1.70	3.85	19.63
		3	1.14	2.59	22.22
		4	1.09	2.49	24.71
		5	1.07	2.43	27.13
		6	1.05	2.38	29.52
		7	1.04	2.36	31.87
		8	1.02	2.32	34.19
		9	1.01	2.29	36.48

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
%	Cumulative %				
ELL/SUA	SUA & ELL Codes	1	4.92	11.18	11.18
		2	1.47	3.34	14.52
		3	1.24	2.82	17.33
		4	1.19	2.72	20.05
		5	1.18	2.68	22.73
		6	1.16	2.64	25.37
		7	1.14	2.60	27.97
		8	1.13	2.58	30.55
		9	1.11	2.52	33.07
		10	1.08	2.46	35.54
		11	1.07	2.44	37.98
		12	1.06	2.41	40.39
		13	1.05	2.38	42.76
		14	1.04	2.35	45.12
		15	1.03	2.34	47.46
		16	1.02	2.32	49.78

Table L6. ELA Grade 8 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
%	Cumulative %				
ELL	ELL=Y	1	6.93	15.74	15.74
		2	1.80	4.09	19.84
		3	1.25	2.85	22.69
		4	1.16	2.64	25.33
		5	1.13	2.58	27.91
		6	1.08	2.46	30.37
		7	1.07	2.42	32.79
		8	1.04	2.36	35.15
		9	1.01	2.30	37.45
		10	1.01	2.30	39.75
SWD	All Codes	1	8.24	18.73	18.73
		2	1.79	4.07	22.80
		3	1.32	3.00	25.80
		4	1.10	2.50	28.29
		5	1.03	2.33	30.63
		6	1.02	2.32	32.94
		7	1.00	2.28	35.22

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SUA	All Codes	<b>1</b>	<b>8.63</b>	<b>19.61</b>	<b>19.61</b>
		2	1.80	4.08	23.69
		3	1.33	3.02	26.71
		4	1.08	2.46	29.17
		5	1.01	2.30	31.47
		6	1.01	2.29	33.76
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>8.02</b>	<b>18.22</b>	<b>18.22</b>
		2	1.79	4.07	22.30
		3	1.32	3.01	25.31
		4	1.10	2.49	27.80
		5	1.03	2.34	30.13
		6	1.03	2.33	32.47
		7	1.01	2.30	34.77
		8	1.01	2.29	37.05
ELL/SUA	SUA & ELL Codes	<b>1</b>	<b>5.83</b>	<b>13.25</b>	<b>13.25</b>
		2	1.69	3.83	17.08
		3	1.29	2.94	20.01
		4	1.21	2.75	22.77
		5	1.18	2.67	25.44
		6	1.15	2.61	28.05
		7	1.13	2.58	30.63
		8	1.11	2.53	33.16
		9	1.09	2.48	35.64
		10	1.08	2.45	38.08
		11	1.06	2.41	40.49
		12	1.04	2.37	42.86
		13	1.03	2.34	45.20
		14	1.02	2.32	47.52
		15	1.00	2.28	49.80

**Table L7. Mathematics Grade 3 Test Factor Analysis by Subgroup**

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	<b>1</b>	<b>9.23</b>	<b>20.51</b>	<b>20.51</b>
		2	1.78	3.95	24.46
		3	1.18	2.63	27.08
		4	1.09	2.42	29.51

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD	All Codes	1	<b>9.87</b>	<b>21.92</b>	<b>21.92</b>
		2	1.69	3.76	25.69
		3	1.18	2.63	28.31
		4	1.09	2.43	30.75
		5	1.01	2.23	32.98
SUA	All Codes	1	<b>9.48</b>	<b>21.06</b>	<b>21.06</b>
		2	1.67	3.71	24.77
		3	1.20	2.67	27.44
		4	1.10	2.43	29.87
		5	1.02	2.27	32.14
SWD/SUA	SUA=504 plan codes	1	<b>9.20</b>	<b>20.43</b>	<b>20.43</b>
		2	1.68	3.73	24.17
		3	1.21	2.69	26.86
		4	1.10	2.43	29.29
		5	1.03	2.29	31.58
		6	1.00	2.23	33.81
ELL/SUA	SUA & ELL Codes	1	<b>8.09</b>	<b>17.98</b>	<b>17.98</b>
		2	1.66	3.70	21.68
		3	1.23	2.73	24.40
		4	1.10	2.44	26.85
		5	1.08	2.39	29.24
		6	1.06	2.35	31.59
		7	1.02	2.28	33.86
		8	1.01	2.24	36.11

**Table L8. Mathematics Grade 4 Test Factor Analysis by Subgroup**

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	<b>11.51</b>	<b>23.97</b>	<b>23.97</b>
		2	1.49	3.11	27.08
		3	1.27	2.64	29.72
		4	1.19	2.48	32.20
		5	1.07	2.23	34.43
		6	1.01	2.11	36.53
SWD	All Codes	1	<b>12.15</b>	<b>25.31</b>	<b>25.31</b>
		2	1.38	2.87	28.18
		3	1.21	2.53	30.70
		4	1.18	2.46	33.16
		5	1.04	2.16	35.32

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SUA	All Codes	<b>1</b>	<b>12.16</b>	<b>25.33</b>	<b>25.33</b>
		2	1.37	2.86	28.19
		3	1.22	2.54	30.73
		4	1.18	2.46	33.19
		5	1.03	2.15	35.34
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>11.58</b>	<b>24.13</b>	<b>24.13</b>
		2	1.39	2.89	27.02
		3	1.23	2.56	29.59
		4	1.18	2.46	32.05
		5	1.06	2.20	34.25
ELL/SUA	SUA & ELL Codes	<b>1</b>	<b>9.18</b>	<b>19.13</b>	<b>19.13</b>
		2	1.51	3.15	22.27
		3	1.30	2.71	24.98
		4	1.22	2.53	27.51
		5	1.17	2.43	29.94
		6	1.09	2.27	32.21
		7	1.05	2.20	34.40
		8	1.02	2.12	36.53
		9	1.01	2.10	38.62

**Table L9. Mathematics Grade 5 Test Factor Analysis by Subgroup**

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	<b>1</b>	<b>8.89</b>	<b>18.91</b>	<b>18.91</b>
		2	1.96	4.17	23.08
		3	1.14	2.42	25.50
		4	1.11	2.36	27.86
		5	1.07	2.27	30.13
		6	1.01	2.15	32.28
SWD	All Codes	<b>1</b>	<b>9.64</b>	<b>20.51</b>	<b>20.51</b>
		2	1.89	4.02	24.53
		3	1.10	2.35	26.88
		4	1.06	2.25	29.13
		5	1.04	2.22	31.34
SUA	All Codes	<b>1</b>	<b>9.79</b>	<b>20.84</b>	<b>20.84</b>
		2	1.89	4.01	24.85
		3	1.10	2.35	27.20
		4	1.05	2.24	29.44
		5	1.04	2.20	31.64

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
%	Cumulative %				
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>9.15</b>	<b>19.46</b>	<b>19.46</b>
		2	1.86	3.97	23.43
		3	1.11	2.36	25.79
		4	1.06	2.27	28.06
		5	1.05	2.23	30.28
ELL/SUA	SUA & ELL Codes	<b>1</b>	<b>7.01</b>	<b>14.91</b>	<b>14.91</b>
		2	1.72	3.65	18.56
		3	1.21	2.57	21.13
		4	1.15	2.45	23.58
		5	1.13	2.40	25.98
		6	1.10	2.33	28.31
		7	1.06	2.26	30.57
		8	1.05	2.24	32.81
		9	1.05	2.22	35.03
		10	1.03	2.20	37.23
		11	1.01	2.15	39.38

Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
%	Cumulative %				
ELL	ELL=Y	<b>1</b>	<b>8.11</b>	<b>15.31</b>	<b>15.31</b>
		2	1.81	3.42	18.73
		3	1.13	2.13	20.86
		4	1.09	2.05	22.91
		5	1.07	2.02	24.92
		6	1.06	2.00	26.92
		7	1.03	1.95	28.87
		8	1.02	1.93	30.80
		9	1.02	1.92	32.72
SWD	All Codes	<b>1</b>	<b>7.95</b>	<b>15.00</b>	<b>15.00</b>
		2	1.63	3.08	18.08
		3	1.15	2.17	20.25
		4	1.09	2.05	22.30
		5	1.06	2.00	24.30
		6	1.04	1.96	26.26
		7	1.02	1.93	28.19
		8	1.02	1.92	30.12
		9	1.01	1.91	32.03
		10	1.00	1.89	33.91

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SUA	All Codes	<b>1</b>	<b>8.41</b>	<b>15.87</b>	<b>15.87</b>
		2	1.63	3.07	18.94
		3	1.16	2.18	21.12
		4	1.08	2.04	23.16
		5	1.06	1.99	25.15
		6	1.03	1.94	27.09
		7	1.02	1.92	29.01
		8	1.01	1.91	30.92
		9	1.00	1.90	32.81
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>7.45</b>	<b>14.05</b>	<b>14.05</b>
		2	1.61	3.03	17.09
		3	1.16	2.18	19.27
		4	1.10	2.07	21.34
		5	1.07	2.02	23.36
		6	1.05	1.97	25.33
		7	1.04	1.95	27.29
		8	1.03	1.95	29.24
		9	1.03	1.93	31.17
		10	1.01	1.91	33.08
		11	1.00	1.89	34.97
ELL/SUA	SUA & ELL Codes	<b>1</b>	<b>5.00</b>	<b>9.44</b>	<b>9.44</b>
		2	1.57	2.96	12.40
		3	1.24	2.34	14.73
		4	1.19	2.25	16.99
		5	1.19	2.24	19.23
		6	1.16	2.19	21.42
		7	1.15	2.17	23.59
		8	1.14	2.16	25.75
		9	1.11	2.10	27.85
		10	1.11	2.09	29.93
		11	1.10	2.07	32.00
		12	1.09	2.06	34.07
		13	1.07	2.01	36.08
		14	1.06	2.00	38.08
		15	1.04	1.97	40.05
		16	1.03	1.95	41.99
		17	1.03	1.93	43.93
		18	1.01	1.91	45.84
		19	1.00	1.89	47.73

Appendix L: Factor Analysis Results for Select Subgroups

**Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup**

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	<b>1</b>	<b>8.62</b>	<b>15.96</b>	<b>15.96</b>
		2	1.43	2.65	18.60
		3	1.20	2.22	20.83
		4	1.12	2.07	22.90
		5	1.07	1.98	24.88
		6	1.06	1.96	26.84
		7	1.05	1.94	28.79
		8	1.04	1.92	30.70
		9	1.02	1.90	32.60
		10	1.01	1.86	34.46
SWD	All Codes	<b>1</b>	<b>8.37</b>	<b>15.51</b>	<b>15.51</b>
		2	1.39	2.57	18.08
		3	1.26	2.34	20.41
		4	1.10	2.04	22.45
		5	1.07	1.97	24.43
		6	1.04	1.93	26.36
		7	1.03	1.91	28.27
		8	1.01	1.87	30.14
		9	1.01	1.86	32.00
SUA	All Codes	<b>1</b>	<b>8.99</b>	<b>16.66</b>	<b>16.66</b>
		2	1.41	2.61	19.26
		3	1.27	2.35	21.61
		4	1.09	2.02	23.63
		5	1.06	1.96	25.59
		6	1.04	1.92	27.50
		7	1.02	1.88	29.39
		8	1.00	1.86	31.24
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>7.84</b>	<b>14.51</b>	<b>14.51</b>
		2	1.38	2.56	17.07
		3	1.28	2.37	19.44
		4	1.11	2.06	21.50
		5	1.08	2.00	23.50
		6	1.05	1.95	25.45
		7	1.04	1.93	27.38
		8	1.02	1.88	29.26
		9	1.01	1.88	31.14
		10	1.01	1.87	33.01

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL/SUA	SUA & ELL Codes	1	5.03	9.31	9.31
		2	1.44	2.67	11.99
		3	1.25	2.32	14.31
		4	1.22	2.26	16.57
		5	1.22	2.26	18.83
		6	1.21	2.24	21.07
		7	1.18	2.19	23.26
		8	1.17	2.16	25.42
		9	1.16	2.15	27.56
		10	1.13	2.09	29.65
		11	1.12	2.07	31.72
		12	1.10	2.04	33.76
		13	1.09	2.01	35.77
		14	1.07	1.98	37.75
		15	1.06	1.97	39.72
		16	1.05	1.94	41.66
		17	1.04	1.92	43.58
		18	1.02	1.90	45.48
		19	1.02	1.89	47.36
		20	1.00	1.85	49.21

Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
ELL	ELL=Y	1	9.63	17.84	17.84
		2	1.48	2.75	20.58
		3	1.24	2.30	22.88
		4	1.15	2.14	25.02
		5	1.10	2.04	27.06
		6	1.06	1.97	29.03
		7	1.05	1.94	30.97
		8	1.03	1.91	32.88
		9	1.02	1.89	34.77
		10	1.01	1.87	36.64
		11	1.00	1.86	38.50

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
				%	Cumulative %
SWD	All Codes	<b>1</b>	<b>8.11</b>	<b>15.01</b>	<b>15.01</b>
		2	1.42	2.63	17.64
		3	1.30	2.41	20.05
		4	1.10	2.04	22.09
		5	1.08	2.00	24.10
		6	1.06	1.96	26.06
		7	1.05	1.94	28.00
		8	1.04	1.93	29.93
		9	1.03	1.92	31.84
		10	1.02	1.90	33.74
		11	1.01	1.87	35.61
SUA	All Codes	<b>1</b>	<b>8.52</b>	<b>15.78</b>	<b>15.78</b>
		2	1.43	2.64	18.42
		3	1.30	2.40	20.82
		4	1.09	2.03	22.85
		5	1.09	2.01	24.86
		6	1.05	1.95	26.81
		7	1.04	1.93	28.74
		8	1.03	1.92	30.65
		9	1.03	1.90	32.55
		10	1.01	1.87	34.42
		11	1.00	1.85	36.28
SWD/SUA	SUA=504 plan codes	<b>1</b>	<b>7.74</b>	<b>14.34</b>	<b>14.34</b>
		2	1.41	2.61	16.95
		3	1.30	2.41	19.36
		4	1.12	2.07	21.43
		5	1.10	2.04	23.47
		6	1.07	1.97	25.44
		7	1.06	1.95	27.39
		8	1.05	1.94	29.33
		9	1.04	1.93	31.26
		10	1.03	1.91	33.17
		11	1.02	1.88	35.05
		12	1.01	1.87	36.92

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor			
		#	Initial Eigenvalue	Variance Accounted for	
%	Cumulative %				
ELL/SUA	SUA & ELL Codes	1	6.03	11.17	11.17
		2	1.41	2.61	13.78
		3	1.28	2.38	16.16
		4	1.26	2.33	18.49
		5	1.22	2.27	20.75
		6	1.19	2.20	22.95
		7	1.18	2.18	25.14
		8	1.16	2.16	27.29
		9	1.14	2.12	29.41
		10	1.12	2.08	31.49
		11	1.11	2.06	33.55
		12	1.10	2.04	35.58
		13	1.09	2.02	37.60
		14	1.07	1.99	39.58
		15	1.05	1.95	41.53
		16	1.03	1.91	43.44
		17	1.03	1.90	45.34
		18	1.02	1.89	47.23
		19	1.01	1.87	49.10

## 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. External linking and field test items (i.e., those not contributing 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	173,557	0.76	0.04	0.44
2	MC	173,392	0.90	0.08	0.40
3	MC	173,388	0.57	0.10	0.43
4	MC	173,328	0.90	0.10	0.39
5	MC	173,343	0.77	0.12	0.38
6	MC	173,273	0.52	0.13	0.42
13	MC	173,243	0.65	0.16	0.41
14	MC	173,151	0.62	0.20	0.38
15	MC	173,197	0.45	0.19	0.33
16	MC	172,983	0.51	0.21	0.33
17	MC	173,304	0.54	0.17	0.36
18	MC	173,254	0.53	0.18	0.30
19	MC	173,175	0.72	0.20	0.38
20	MC	173,225	0.48	0.19	0.43
21	MC	173,214	0.52	0.21	0.42
22	MC	173,123	0.43	0.25	0.37
23	MC	173,073	0.50	0.31	0.30
24	MC	172,908	0.68	0.42	0.46
25	MC	173,577	0.73	0.03	0.46
26	MC	173,473	0.66	0.06	0.40
27	MC	173,365	0.34	0.09	0.39
28	MC	173,444	0.79	0.10	0.45
29	MC	173,512	0.65	0.07	0.33
30	MC	173,410	0.57	0.11	0.43
31	MC	173,304	0.47	0.20	0.40
32	CR2	172,801	0.61	0.51	0.56
33	CR2	172,136	0.48	0.90	0.57
34	CR4	171,975	0.39	0.99	0.65
35	CR2	173,397	0.53	0.17	0.62
36	CR2	172,872	0.54	0.47	0.57
37	CR2	172,402	0.50	0.74	0.58
38	CR2	171,801	0.47	1.09	0.57
39	CR2	171,520	0.42	1.25	0.63
40	CR4	170,874	0.30	1.62	0.64

**Table M2. ELA Grade 4 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	171,124	0.55	0.02	0.38
2	MC	171,104	0.49	0.03	0.38
3	MC	170,970	0.66	0.05	0.30
4	MC	170,974	0.54	0.06	0.24
5	MC	170,993	0.65	0.07	0.42
6	MC	170,988	0.63	0.06	0.22
13	MC	170,985	0.41	0.07	0.29
14	MC	170,940	0.44	0.10	0.33
15	MC	170,915	0.57	0.10	0.42
16	MC	170,922	0.55	0.11	0.40
17	MC	170,980	0.62	0.09	0.40
18	MC	170,952	0.54	0.11	0.39
19	MC	170,784	0.46	0.19	0.25
20	MC	170,892	0.54	0.12	0.28
21	MC	170,927	0.64	0.11	0.39
22	MC	170,867	0.43	0.14	0.25
23	MC	170,799	0.64	0.18	0.26
24	MC	170,743	0.43	0.25	0.31
25	MC	171,110	0.70	0.03	0.38
26	MC	171,055	0.39	0.04	0.28
27	MC	170,963	0.43	0.06	0.23
28	MC	171,020	0.39	0.07	0.30
29	MC	171,068	0.53	0.05	0.28
30	MC	171,016	0.66	0.08	0.36
31	MC	170,894	0.70	0.15	0.39
32	CR2	170,007	0.56	0.69	0.57
33	CR2	169,886	0.57	0.76	0.54
34	CR4	169,098	0.43	1.22	0.66
35	CR2	170,916	0.60	0.16	0.51
36	CR2	170,248	0.58	0.55	0.62
37	CR2	170,574	0.75	0.36	0.56
38	CR2	170,272	0.63	0.53	0.60
39	CR2	170,031	0.60	0.67	0.60
40	CR4	169,851	0.45	0.78	0.70

**Table M3. ELA Grade 5 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	160,780	0.87	0.01	0.33
2	MC	160,535	0.62	0.04	0.40
3	MC	160,681	0.58	0.04	0.48
4	MC	160,707	0.70	0.04	0.31
5	MC	160,708	0.51	0.03	0.26
6	MC	160,673	0.44	0.07	0.22

Appendix M: Classical Test Theory Statistics

Item	Type	N-Count	P-Value	% Omit	PBis Key
7	MC	160,689	0.85	0.04	0.36
8	MC	160,683	0.78	0.04	0.41
9	MC	160,595	0.74	0.11	0.36
10	MC	160,609	0.63	0.09	0.39
11	MC	160,630	0.42	0.09	0.19
12	MC	160,673	0.48	0.07	0.35
13	MC	160,669	0.82	0.07	0.48
14	MC	160,678	0.72	0.06	0.45
15	MC	160,624	0.52	0.09	0.22
16	MC	160,656	0.59	0.07	0.39
17	MC	160,576	0.61	0.11	0.43
18	MC	160,597	0.74	0.10	0.44
19	MC	160,596	0.50	0.10	0.39
20	MC	160,562	0.68	0.11	0.34
21	MC	160,583	0.52	0.12	0.34
29	MC	160,557	0.36	0.13	0.16
30	MC	160,564	0.51	0.11	0.18
31	MC	160,473	0.49	0.16	0.30
32	MC	160,508	0.65	0.15	0.48
33	MC	160,467	0.60	0.19	0.42
34	MC	160,538	0.56	0.15	0.36
35	MC	160,378	0.42	0.25	0.27
36	MC	160,744	0.37	0.03	0.26
37	MC	160,712	0.72	0.03	0.17
38	MC	160,604	0.57	0.06	0.38
39	MC	160,703	0.76	0.05	0.33
40	MC	160,644	0.66	0.04	0.46
41	MC	160,697	0.79	0.05	0.42
42	MC	160,667	0.82	0.08	0.36
43	CR2	160,462	0.75	0.22	0.52
44	CR2	159,941	0.64	0.54	0.58
45	CR4	159,895	0.48	0.57	0.63
46	CR2	160,633	0.77	0.11	0.58
47	CR2	160,224	0.69	0.36	0.55
48	CR2	160,298	0.63	0.32	0.59
49	CR2	159,963	0.58	0.53	0.57
50	CR2	159,801	0.66	0.63	0.65
51	CR4	159,454	0.42	0.84	0.67

**Table M4. ELA Grade 6 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	158,156	0.65	0.02	0.31
2	MC	158,141	0.70	0.03	0.37
3	MC	158,108	0.67	0.03	0.45
4	MC	158,052	0.60	0.07	0.31
5	MC	158,071	0.67	0.05	0.30
6	MC	158,115	0.73	0.03	0.46
7	MC	157,982	0.34	0.12	0.21
8	MC	158,061	0.34	0.07	0.17
9	MC	157,949	0.53	0.14	0.25
10	MC	158,070	0.64	0.07	0.36
11	MC	157,985	0.49	0.12	0.17
12	MC	158,069	0.72	0.06	0.41
13	MC	158,055	0.34	0.07	0.22
14	MC	158,023	0.59	0.10	0.41
22	MC	157,894	0.60	0.17	0.32
23	MC	158,025	0.46	0.09	0.27
24	MC	158,029	0.67	0.09	0.42
25	MC	157,901	0.43	0.15	0.30
26	MC	157,894	0.41	0.16	0.27
27	MC	157,942	0.48	0.12	0.21
28	MC	157,878	0.50	0.17	0.32
29	MC	157,938	0.66	0.15	0.34
30	MC	157,904	0.55	0.15	0.32
31	MC	157,906	0.33	0.14	0.14
32	MC	157,869	0.59	0.18	0.37
33	MC	157,796	0.46	0.23	0.23
34	MC	157,904	0.51	0.17	0.24
35	MC	157,849	0.59	0.21	0.40
36	MC	158,142	0.38	0.03	0.13
37	MC	158,143	0.76	0.03	0.40
38	MC	158,090	0.34	0.05	0.19
39	MC	158,125	0.53	0.04	0.38
40	MC	158,123	0.48	0.04	0.31
41	MC	158,053	0.53	0.08	0.43
42	MC	157,980	0.56	0.13	0.31
43	CR2	157,763	0.71	0.28	0.53
44	CR2	157,382	0.72	0.52	0.63
45	CR4	157,309	0.56	0.57	0.69
46	CR2	157,916	0.70	0.19	0.55
47	CR2	157,183	0.60	0.65	0.59
48	CR2	157,775	0.78	0.27	0.57
49	CR2	157,589	0.71	0.39	0.60
50	CR2	157,020	0.69	0.75	0.55
51	CR4	156,802	0.58	0.89	0.71

**Table M5. ELA Grade 7 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	148,805	0.49	0.03	0.44
2	MC	148,752	0.68	0.05	0.16
3	MC	148,717	0.66	0.05	0.45
4	MC	148,727	0.52	0.06	0.36
5	MC	148,686	0.53	0.10	0.27
6	MC	148,738	0.69	0.06	0.36
7	MC	148,725	0.63	0.07	0.46
8	MC	148,754	0.56	0.05	0.38
9	MC	148,660	0.54	0.11	0.41
10	MC	148,737	0.58	0.05	0.37
11	MC	148,689	0.75	0.09	0.40
12	MC	148,747	0.29	0.06	0.28
13	MC	148,715	0.63	0.08	0.38
14	MC	148,751	0.57	0.05	0.31
15	MC	148,756	0.53	0.05	0.26
16	MC	148,707	0.53	0.08	0.38
17	MC	148,709	0.51	0.08	0.32
18	MC	148,597	0.43	0.14	0.25
19	MC	148,703	0.74	0.07	0.53
20	MC	148,636	0.55	0.12	0.32
21	MC	148,663	0.51	0.11	0.22
29	MC	148,614	0.40	0.14	0.20
30	MC	148,625	0.44	0.12	0.30
31	MC	148,578	0.35	0.14	0.19
32	MC	148,516	0.51	0.19	0.39
33	MC	148,505	0.38	0.20	0.32
34	MC	148,571	0.53	0.17	0.32
35	MC	148,553	0.70	0.19	0.41
36	MC	148,793	0.66	0.03	0.38
37	MC	148,787	0.79	0.03	0.37
38	MC	148,750	0.53	0.04	0.20
39	MC	148,744	0.55	0.05	0.36
40	MC	148,779	0.41	0.04	0.22
41	MC	148,757	0.62	0.05	0.37
42	MC	148,607	0.58	0.16	0.37
43	CR2	147,974	0.65	0.59	0.61
44	CR2	147,369	0.71	1.00	0.65
45	CR4	147,424	0.54	0.96	0.69
46	CR2	148,527	0.76	0.22	0.58
47	CR2	147,888	0.70	0.65	0.64
48	CR2	147,737	0.64	0.75	0.61
49	CR2	147,388	0.65	0.99	0.63
50	CR2	146,152	0.61	1.82	0.65
51	CR4	145,945	0.49	1.96	0.72

**Table M6. ELA Grade 8 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	143,444	0.59	0.07	0.34
2	MC	143,480	0.58	0.04	0.25
3	MC	143,488	0.96	0.02	0.33
4	MC	143,472	0.89	0.03	0.38
5	MC	143,468	0.63	0.04	0.37
6	MC	143,424	0.73	0.07	0.37
7	MC	143,467	0.75	0.04	0.31
8	MC	143,453	0.86	0.05	0.46
9	MC	143,419	0.66	0.07	0.36
10	MC	143,440	0.66	0.06	0.40
11	MC	143,383	0.55	0.10	0.09
12	MC	143,447	0.87	0.07	0.51
13	MC	143,455	0.71	0.06	0.30
14	MC	143,423	0.42	0.08	0.28
22	MC	143,297	0.72	0.16	0.46
23	MC	143,403	0.69	0.09	0.49
24	MC	143,379	0.55	0.10	0.22
25	MC	143,370	0.52	0.10	0.46
26	MC	143,298	0.63	0.15	0.47
27	MC	143,338	0.73	0.12	0.48
28	MC	143,311	0.57	0.15	0.35
29	MC	143,329	0.61	0.14	0.40
30	MC	143,362	0.75	0.11	0.37
31	MC	143,289	0.65	0.15	0.44
32	MC	143,232	0.53	0.19	0.30
33	MC	143,240	0.67	0.19	0.48
34	MC	143,282	0.59	0.17	0.44
35	MC	143,236	0.60	0.20	0.40
36	MC	143,475	0.54	0.04	0.44
37	MC	143,464	0.57	0.05	0.36
38	MC	143,463	0.72	0.04	0.46
39	MC	143,458	0.72	0.05	0.45
40	MC	143,478	0.73	0.04	0.36
41	MC	143,433	0.54	0.07	0.37
42	MC	143,396	0.85	0.10	0.42
43	CR2	142,419	0.73	0.79	0.54
44	CR2	141,568	0.75	1.38	0.63
45	CR4	141,894	0.59	1.16	0.71
46	CR2	143,118	0.78	0.30	0.54
47	CR2	142,275	0.74	0.89	0.60
48	CR2	143,211	0.86	0.24	0.62
49	CR2	142,228	0.80	0.92	0.64
50	CR2	141,725	0.71	1.27	0.65
51	CR4	141,513	0.65	1.42	0.72

**Table M7. Mathematics Grade 3 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	178,772	0.79	0.03	0.41
2	MC	178,727	0.77	0.04	0.43
3	MC	177,536	0.31	0.21	0.24
4	MC	178,413	0.90	0.10	0.30
6	MC	178,401	0.69	0.13	0.44
7	MC	178,581	0.83	0.10	0.37
8	MC	178,359	0.58	0.14	0.47
9	MC	178,357	0.58	0.19	0.43
11	MC	178,631	0.89	0.09	0.23
12	MC	178,545	0.81	0.12	0.44
13	MC	178,482	0.55	0.11	0.43
14	MC	178,318	0.62	0.22	0.42
16	MC	178,325	0.66	0.18	0.36
17	MC	178,259	0.56	0.26	0.55
19	MC	178,487	0.65	0.17	0.57
20	MC	178,393	0.85	0.18	0.44
21	MC	178,255	0.73	0.31	0.47
22	MC	177,439	0.49	0.68	0.47
23	MC	178,781	0.84	0.03	0.43
24	MC	178,632	0.57	0.08	0.56
25	MC	178,397	0.53	0.17	0.58
26	MC	178,341	0.72	0.12	0.42
27	MC	178,492	0.64	0.11	0.41
28	MC	178,549	0.74	0.11	0.42
30	MC	178,499	0.48	0.12	0.34
31	MC	178,443	0.89	0.12	0.30
32	MC	178,566	0.67	0.11	0.52
33	MC	178,576	0.60	0.11	0.49
34	MC	178,630	0.89	0.10	0.31
35	MC	178,637	0.80	0.09	0.45
37	MC	178,365	0.54	0.17	0.41
38	MC	178,397	0.59	0.17	0.48
39	MC	178,394	0.41	0.19	0.41
40	MC	178,463	0.81	0.19	0.50
41	MC	178,636	0.58	0.10	0.53
42	MC	178,474	0.59	0.17	0.47
43	MC	178,404	0.64	0.23	0.59
45	CR2	178,271	0.43	0.33	0.61
46	CR2	178,652	0.63	0.12	0.33
47	CR2	178,474	0.69	0.22	0.58
48	CR2	178,262	0.24	0.34	0.56
49	CR2	178,379	0.55	0.27	0.63
50	CR3	178,166	0.37	0.39	0.56
51	CR3	178,156	0.53	0.40	0.58
52	CR3	177,942	0.34	0.52	0.69

**Table M8. Mathematics Grade 4 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	174,275	0.83	0.02	0.43
2	MC	174,195	0.77	0.04	0.47
3	MC	174,074	0.72	0.05	0.58
4	MC	174,051	0.49	0.10	0.58
5	MC	174,040	0.65	0.10	0.47
6	MC	174,057	0.72	0.10	0.53
7	MC	174,120	0.65	0.07	0.33
8	MC	173,988	0.71	0.13	0.46
9	MC	174,053	0.74	0.11	0.58
10	MC	174,039	0.58	0.10	0.64
12	MC	173,944	0.49	0.17	0.44
13	MC	174,031	0.43	0.11	0.53
14	MC	173,993	0.41	0.11	0.36
16	MC	173,810	0.72	0.25	0.57
17	MC	174,029	0.70	0.10	0.45
18	MC	173,954	0.61	0.18	0.52
19	MC	174,106	0.65	0.08	0.51
20	MC	174,039	0.75	0.11	0.43
23	MC	174,204	0.75	0.05	0.28
24	MC	174,203	0.66	0.04	0.41
25	MC	174,092	0.63	0.06	0.49
26	MC	174,090	0.61	0.08	0.51
27	MC	174,071	0.51	0.08	0.56
28	MC	174,128	0.75	0.07	0.43
29	MC	174,108	0.79	0.09	0.47
30	MC	173,999	0.68	0.15	0.54
31	MC	174,112	0.70	0.07	0.48
32	MC	174,020	0.69	0.12	0.55
33	MC	174,079	0.56	0.11	0.49
34	MC	174,093	0.46	0.08	0.45
35	MC	174,122	0.55	0.07	0.42
37	MC	174,127	0.66	0.08	0.61
38	MC	174,096	0.70	0.09	0.61
39	MC	173,913	0.60	0.18	0.51
40	MC	174,030	0.49	0.12	0.54
42	MC	174,042	0.69	0.12	0.39
43	MC	173,944	0.67	0.16	0.53
45	MC	173,615	0.59	0.39	0.63
46	CR2	173,886	0.47	0.25	0.60
47	CR2	173,878	0.68	0.25	0.52
48	CR2	173,891	0.65	0.25	0.59
49	CR2	173,788	0.40	0.31	0.63
50	CR2	173,670	0.49	0.37	0.66

Item	Type	N-Count	P-Value	% Omit	PBis Key
51	CR2	173,791	0.58	0.30	0.45
52	CR3	173,706	0.23	0.35	0.63
53	CR3	173,763	0.59	0.32	0.59
54	CR3	173,860	0.51	0.26	0.73
55	CR3	173,787	0.50	0.31	0.70

**Table M9. Mathematics Grade 5 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	162,832	0.57	0.08	0.53
2	MC	162,837	0.64	0.07	0.56
3	MC	162,835	0.86	0.07	0.46
4	MC	162,610	0.64	0.19	0.03
5	MC	162,775	0.68	0.08	0.44
6	MC	162,752	0.47	0.11	0.41
8	MC	162,763	0.53	0.10	0.46
9	MC	162,687	0.45	0.14	0.43
10	MC	162,735	0.61	0.09	0.51
11	MC	162,876	0.83	0.04	0.44
13	MC	162,832	0.71	0.07	0.30
14	MC	162,756	0.64	0.10	0.48
15	MC	162,550	0.38	0.22	0.41
16	MC	162,660	0.49	0.14	0.52
17	MC	162,788	0.62	0.08	0.58
18	MC	162,701	0.56	0.14	0.48
19	MC	162,610	0.30	0.21	0.39
20	MC	162,654	0.31	0.17	0.47
23	MC	162,846	0.75	0.08	0.47
24	MC	162,869	0.57	0.05	0.26
25	MC	162,880	0.74	0.05	0.35
26	MC	162,805	0.78	0.09	0.43
27	MC	162,821	0.46	0.07	0.49
28	MC	162,770	0.62	0.10	0.42
29	MC	162,809	0.63	0.08	0.51
31	MC	162,768	0.76	0.11	0.53
33	MC	162,821	0.50	0.07	0.50
34	MC	162,778	0.59	0.10	0.39
36	MC	162,742	0.51	0.11	0.29
37	MC	162,830	0.47	0.07	0.43
39	MC	162,730	0.67	0.11	0.50
40	MC	162,789	0.60	0.09	0.37
41	MC	162,781	0.72	0.10	0.49
42	MC	162,725	0.50	0.12	0.62
43	MC	162,747	0.52	0.13	0.61
44	MC	162,784	0.37	0.11	0.51

Appendix M: Classical Test Theory Statistics

Item	Type	N-Count	P-Value	% Omit	PBis Key
45	MC	162,400	0.74	0.35	0.44
46	CR2	162,894	0.59	0.06	0.62
47	CR2	162,576	0.51	0.26	0.55
48	CR2	162,739	0.71	0.16	0.60
49	CR2	161,483	0.45	0.93	0.60
50	CR2	162,306	0.40	0.42	0.59
51	CR2	161,883	0.57	0.68	0.43
52	CR3	162,228	0.51	0.47	0.69
53	CR3	162,276	0.24	0.44	0.66
54	CR3	162,216	0.20	0.48	0.62
55	CR3	159,463	0.20	2.17	0.48

**Table M10. Mathematics Grade 6 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	161,157	0.83	0.03	0.27
2	MC	161,114	0.71	0.05	0.37
4	MC	160,880	0.68	0.18	0.42
5	MC	161,048	0.54	0.06	0.39
7	MC	161,021	0.63	0.09	0.44
8	MC	161,059	0.70	0.07	0.27
9	MC	160,869	0.14	0.19	0.30
11	MC	161,023	0.62	0.08	0.38
12	MC	160,856	0.54	0.10	0.49
13	MC	160,846	0.47	0.19	0.50
14	MC	161,012	0.77	0.10	0.45
15	MC	161,067	0.35	0.07	0.56
16	MC	160,936	0.38	0.14	0.37
17	MC	160,769	0.48	0.24	0.34
18	MC	161,021	0.64	0.09	0.48
19	MC	160,991	0.47	0.10	0.50
20	MC	160,988	0.62	0.11	0.45
21	MC	160,994	0.54	0.11	0.52
22	MC	161,021	0.59	0.09	0.35
25	MC	160,843	0.63	0.19	0.26
26	MC	160,495	0.29	0.42	0.32
27	MC	161,138	0.82	0.04	0.24
28	MC	161,020	0.71	0.11	0.50
29	MC	161,039	0.72	0.08	0.48
30	MC	160,957	0.38	0.12	0.49
31	MC	161,033	0.71	0.07	0.47
33	MC	161,005	0.46	0.11	0.51
34	MC	161,016	0.62	0.11	0.40
35	MC	161,011	0.54	0.08	0.29
36	MC	160,990	0.78	0.09	0.47

Appendix M: Classical Test Theory Statistics

Item	Type	N-Count	P-Value	% Omit	PBis Key
37	MC	161,064	0.29	0.07	0.21
38	MC	161,003	0.46	0.10	0.53
39	MC	160,991	0.41	0.10	0.43
40	MC	160,998	0.46	0.10	0.42
41	MC	161,073	0.59	0.06	0.49
42	MC	160,943	0.68	0.12	0.43
43	MC	160,966	0.34	0.12	0.42
44	MC	160,993	0.26	0.11	0.21
45	MC	160,947	0.40	0.12	0.36
46	MC	161,030	0.48	0.09	0.45
47	MC	160,954	0.42	0.14	0.30
48	MC	161,069	0.85	0.07	0.38
49	MC	160,926	0.54	0.15	0.56
52	CR2	161,022	0.55	0.12	0.62
53	CR2	160,475	0.41	0.46	0.60
54	CR2	160,790	0.55	0.26	0.56
55	CR2	160,693	0.35	0.32	0.66
56	CR2	160,328	0.39	0.55	0.62
57	CR2	160,292	0.28	0.57	0.66
58	CR3	158,952	0.20	1.40	0.54
59	CR3	160,047	0.34	0.73	0.68
60	CR3	160,217	0.12	0.62	0.55
61	CR3	160,462	0.41	0.47	0.70

**Table M11. Mathematics Grade 7 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	147,029	0.70	0.14	0.46
2	MC	146,822	0.40	0.26	0.37
4	MC	146,749	0.44	0.31	0.37
6	MC	147,094	0.80	0.08	0.33
7	MC	146,879	0.44	0.22	0.33
8	MC	147,078	0.54	0.08	0.43
9	MC	147,026	0.48	0.11	0.44
10	MC	146,871	0.47	0.24	0.51
11	MC	147,060	0.69	0.10	0.49
12	MC	147,090	0.57	0.08	0.39
13	MC	147,032	0.51	0.10	0.45
14	MC	147,010	0.33	0.13	0.47
15	MC	146,883	0.47	0.22	0.49
16	MC	147,072	0.63	0.08	0.37
17	MC	146,935	0.57	0.17	0.37
18	MC	146,881	0.31	0.21	0.38
20	MC	146,840	0.60	0.25	0.54
21	MC	146,829	0.44	0.26	0.42

Item	Type	N-Count	P-Value	% Omit	PBis Key
22	MC	146,907	0.38	0.20	0.41
23	MC	146,875	0.47	0.21	0.37
24	MC	146,762	0.34	0.29	0.34
25	MC	146,851	0.70	0.24	0.54
27	MC	147,224	0.67	0.01	0.47
28	MC	147,091	0.51	0.10	0.50
29	MC	147,071	0.48	0.10	0.54
30	MC	146,822	0.51	0.27	0.50
31	MC	147,072	0.51	0.10	0.53
33	MC	147,109	0.67	0.07	0.54
34	MC	146,988	0.39	0.16	0.33
35	MC	146,992	0.48	0.14	0.54
36	MC	146,995	0.51	0.15	0.53
37	MC	147,068	0.42	0.10	0.38
38	MC	146,889	0.45	0.22	0.42
39	MC	147,020	0.35	0.12	0.32
40	MC	147,070	0.57	0.10	0.52
41	MC	147,052	0.39	0.11	0.43
42	MC	147,059	0.50	0.11	0.24
43	MC	146,890	0.64	0.22	0.51
44	MC	147,049	0.48	0.11	0.31
45	MC	147,058	0.39	0.10	0.46
46	MC	146,988	0.60	0.16	0.42
47	MC	147,071	0.54	0.10	0.56
48	MC	147,119	0.55	0.07	0.40
49	MC	147,006	0.49	0.14	0.47
52	CR2	145,763	0.30	1.01	0.63
53	CR2	146,648	0.44	0.41	0.75
54	CR2	146,700	0.58	0.37	0.62
55	CR2	146,377	0.46	0.59	0.59
56	CR2	145,143	0.28	1.43	0.68
57	CR2	144,673	0.56	1.75	0.60
58	CR3	145,117	0.32	1.45	0.61
59	CR3	145,491	0.31	1.20	0.60
60	CR3	145,619	0.34	1.11	0.74
61	CR3	146,269	0.48	0.67	0.73

**Table M12. Mathematics Grade 8 Classical Item Analysis**

Item	Type	N-Count	P-Value	% Omit	PBis Key
1	MC	115,097	0.83	0.07	0.34
2	MC	115,110	0.51	0.05	0.49
3	MC	115,035	0.46	0.11	0.38
4	MC	115,093	0.61	0.05	0.40
5	MC	114,926	0.58	0.20	0.41

Appendix M: Classical Test Theory Statistics

Item	Type	N-Count	P-Value	% Omit	PBis Key
6	MC	114,932	0.51	0.18	0.34
7	MC	114,976	0.44	0.16	0.46
8	MC	115,070	0.49	0.07	0.34
9	MC	114,979	0.39	0.15	0.37
10	MC	115,055	0.55	0.09	0.28
11	MC	115,030	0.57	0.11	0.43
12	MC	114,959	0.51	0.18	0.40
15	MC	115,003	0.27	0.14	0.43
16	MC	114,978	0.36	0.15	0.29
17	MC	114,983	0.57	0.15	0.48
19	MC	115,028	0.55	0.10	0.49
20	MC	115,026	0.72	0.11	0.46
21	MC	115,043	0.31	0.10	0.24
22	MC	115,050	0.76	0.10	0.39
24	MC	114,920	0.64	0.20	0.29
25	MC	114,961	0.65	0.16	0.30
26	MC	114,934	0.53	0.20	0.43
27	MC	115,087	0.66	0.08	0.35
28	MC	114,966	0.52	0.16	0.53
29	MC	115,040	0.58	0.10	0.47
30	MC	115,041	0.53	0.09	0.42
32	MC	114,982	0.33	0.14	0.43
33	MC	114,997	0.54	0.15	0.26
34	MC	115,007	0.50	0.12	0.49
35	MC	114,965	0.60	0.16	0.40
36	MC	114,806	0.48	0.32	0.42
37	MC	115,053	0.67	0.09	0.39
38	MC	115,040	0.54	0.10	0.44
39	MC	115,065	0.41	0.07	0.27
40	MC	114,870	0.48	0.26	0.42
41	MC	115,088	0.74	0.06	0.38
42	MC	115,051	0.65	0.10	0.45
44	MC	115,039	0.47	0.09	0.39
45	MC	115,046	0.49	0.09	0.40
46	MC	115,076	0.48	0.07	0.37
47	MC	115,061	0.42	0.09	0.32
48	MC	115,030	0.43	0.11	0.44
49	MC	114,989	0.45	0.15	0.33
50	MC	114,993	0.33	0.14	0.30
52	CR2	113,885	0.40	1.13	0.49
53	CR2	114,032	0.37	1.01	0.54
54	CR2	110,790	0.38	3.82	0.58
55	CR2	112,705	0.45	2.16	0.65
56	CR2	112,551	0.26	2.29	0.64

Appendix M: Classical Test Theory Statistics

<b>Item</b>	<b>Type</b>	<b>N-Count</b>	<b>P-Value</b>	<b>% Omit</b>	<b>PBis Key</b>
57	CR2	110,792	0.38	3.82	0.59
58	CR3	111,958	0.27	2.81	0.64
59	CR3	111,214	0.23	3.45	0.68
60	CR3	111,121	0.25	3.53	0.70
61	CR3	110,384	0.19	4.17	0.67

## 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–N3 show items flagged by the SMD, or Mantel-Haenszel methods. No mathematics constructed-response items were flagged for DIF, so that table has been omitted. Positive values of SMD and Delta in Tables N1–N3 indicate DIF in favor of a focal group, and negative values of SMD and Delta indicate DIF against a focal group. External linking and field test items (i.e., those not contributing to students’ scores) have been omitted.

**Table N1. ELA MC Item Classical DIF Flags**

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
3	21	Black	Against	1.55	827.60	-1.03
3	21	Hispanic	Against	1.89	2332.50	-1.49
3	21	Asian	Against	1.61	624.80	-1.11
3	21	High Needs	Against	1.65	1789.70	-1.18
3	21	ELL	Against	1.70	720.40	-1.24
3	25	Female	Against	1.66	1620.60	-1.19
3	25	Hispanic	Against	1.67	1141.40	-1.20
3	25	ELL	Against	1.77	999.10	-1.34
4	6	Asian	Against	1.67	787.00	-1.21
4	6	High Needs	Against	1.63	1705.70	-1.15
4	15	ELL	Against	1.58	533.80	-1.07
4	16	Hispanic	Against	1.56	1161.50	-1.05
5	1	ELL	Against	1.60	415.70	-1.10
5	3	Black	Against	2.06	2027.90	-1.70
5	3	Hispanic	Against	2.11	2666.00	-1.75
5	3	Asian	Against	2.40	1683.60	-2.06
5	3	High Needs	Against	2.03	2849.70	-1.66
5	3	ELL	Against	1.95	801.60	-1.57
5	8	Black	Against	1.60	653.30	-1.11
5	8	Hispanic	Against	1.73	1078.50	-1.29
5	8	Asian	Against	1.78	522.90	-1.36
5	8	High Needs	Against	1.78	1337.40	-1.36
5	8	ELL	Against	1.87	883.40	-1.47
5	16	Black	Against	1.58	883.70	-1.07
5	16	Hispanic	Against	1.69	1471.20	-1.23
5	16	ELL	Against	1.60	447.00	-1.11
5	18	Black	Against	1.55	638.30	-1.03
5	32	Black	Against	1.65	954.00	-1.18
5	32	Hispanic	Against	1.65	1176.70	-1.18
5	32	High Needs	Against	1.58	1140.80	-1.08
5	32	ELL	Against	1.54	381.30	-1.01
5	33	ELL	Against	1.70	578.20	-1.24
5	40	Asian	Against	1.54	380.20	-1.01
6	1	ELL	Against	1.85	826.10	-1.44

Appendix N: Items Flagged for DIF

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
6	2	Hispanic	Against	1.65	1191.70	-1.18
6	2	Asian	Against	1.59	449.20	-1.10
6	2	High Needs	Against	1.56	1061.80	-1.04
6	2	ELL	Against	1.90	936.40	-1.51
6	31	Female	Against	1.68	2227.70	-1.22
6	37	Hispanic	Against	1.69	1090.60	-1.24
6	37	Asian	Against	1.68	431.40	-1.22
6	37	High Needs	Against	1.56	849.60	-1.04
6	37	ELL	Against	2.02	1132.00	-1.65
6	41	Female	Against	1.65	1877.10	-1.17
7	1	Female	Against	1.56	1391.20	-1.05
7	1	Black	Against	1.63	932.90	-1.15
7	1	Hispanic	Against	1.65	1180.80	-1.17
7	1	High Needs	Against	1.65	1529.50	-1.17
7	1	ELL	Against	1.59	280.20	-1.09
7	3	Female	Against	1.68	1681.50	-1.22
7	3	Asian	Against	1.53	342.40	-1.00
7	10	Female	Against	1.63	1743.60	-1.15
7	10	Hispanic	Against	1.55	959.00	-1.04
7	10	Asian	Against	1.73	754.70	-1.29
7	10	ELL	Against	2.24	1056.10	-1.89
7	12	Hispanic	Against	1.62	952.50	-1.13
7	12	High Needs	Against	1.59	1281.90	-1.10
7	17	Asian	Against	1.88	1083.40	-1.48
7	17	ELL	Against	1.77	526.60	-1.34
7	19	Hispanic	Against	1.73	985.80	-1.28
7	19	Asian	Against	1.66	313.60	-1.20
7	19	High Needs	Against	1.56	729.30	-1.05
7	19	ELL	Against	1.60	398.50	-1.10
8	2	ELL	Against	1.61	415.70	-1.12
8	3	Black	In Favor	0.60	169.10	1.21
8	4	ELL	Against	1.64	361.90	-1.16
8	8	ELL	Against	1.91	697.20	-1.53
8	10	Asian	Against	1.96	992.20	-1.58
8	36	Black	Against	1.98	1797.50	-1.61
8	36	Hispanic	Against	1.95	1988.80	-1.56
8	36	Asian	Against	1.56	428.30	-1.04
8	36	High Needs	Against	1.76	1703.30	-1.33

**Table N2. ELA CR Item Classical DIF Flags**

Grade	Item	Subgroup	DIF	SMD	Effect
4	33	High Needs	In Favor	0.12	0.18
5	43	Black	In Favor	0.12	0.20
5	43	Hispanic	In Favor	0.12	0.20
5	43	Asian	In Favor	0.12	0.20
5	43	High Needs	In Favor	0.12	0.20
5	45	Asian	In Favor	0.21	0.20
6	45	Female	In Favor	0.18	0.18
7	48	High Needs	In Favor	0.13	0.18
7	49	High Needs	In Favor	0.14	0.19
7	51	Female	In Favor	0.22	0.18
8	45	Female	In Favor	0.21	0.19
8	46	Black	In Favor	0.10	0.17
8	46	Hispanic	In Favor	0.10	0.18
8	46	High Needs	In Favor	0.12	0.21

**Table N3. Mathematics MC Item Classical DIF Flags**

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
3	24	Asian	In Favor	0.52	803.70	1.55
3	33	Black	Against	1.62	955.10	-1.14
4	4	Female	Against	1.66	1781.70	-1.20
4	6	Black	Against	1.55	616.20	-1.04
4	6	Asian	Against	1.61	346.80	-1.12
4	29	Asian	Against	1.54	227.40	-1.02
4	43	Black	In Favor	0.62	743.10	1.13
4	43	Asian	In Favor	0.65	294.90	1.02
5	5	ELL	Against	1.62	605.20	-1.13
5	10	ELL	Against	1.54	429.30	-1.02
5	26	Asian	In Favor	0.65	195.50	1.01
6	5	High Needs	In Favor	0.64	1171.70	1.04
6	15	Black	Against	1.55	522.00	-1.02
7	9	Black	Against	1.56	715.00	-1.04
7	12	Hispanic	Against	1.66	1214.40	-1.19
7	12	Asian	Against	1.80	777.20	-1.38
7	12	High Needs	Against	1.78	1844.60	-1.36
7	12	ELL	Against	1.74	672.30	-1.31
7	13	High Needs	Against	1.54	1031.00	-1.01
8	29	Female	Against	1.57	1052.40	-1.05

**Table N4. Mathematics CR Item Classical DIF Flags**

<b>Grade</b>	<b>Item</b>	<b>Subgroup</b>	<b>DIF</b>	<b>SMD</b>	<b>Effect</b>
5	55	ELL	In Favor	0.14	0.17
6	54	Female	In Favor	0.14	0.18
6	56	Black	Against	-0.13	-0.18
8	58	Black	Against	-0.23	-0.20
8	58	Hispanic	Against	-0.22	-0.19

## Appendix O: IRT Statistics

---

External linking and field test items (i.e., those not contributing 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	388.95	8	95.24	462.41	Y
2	3PL	379.36	8	92.84	461.97	Y
3	3PL	353.25	8	86.31	461.95	Y
4	3PL	757.27	8	187.32	461.79	Y
5	3PL	317.82	8	77.45	461.83	Y
6	3PL	345.19	8	84.30	461.65	Y
13	3PL	720.58	8	178.15	461.57	Y
14	3PL	255.26	8	61.82	461.32	Y
15	3PL	382.35	8	93.59	461.45	Y
16	3PL	423.22	8	103.80	460.88	Y
17	3PL	240.68	8	58.17	461.73	Y
18	3PL	153.22	8	36.31	461.60	Y
19	3PL	251.29	8	60.82	461.39	Y
20	3PL	624.37	8	154.09	461.52	Y
21	3PL	681.02	8	168.26	461.49	Y
22	3PL	737.73	8	182.43	461.25	Y
23	3PL	243.48	8	58.87	461.11	Y
24	3PL	399.43	8	97.86	460.67	Y
25	3PL	414.45	8	101.61	462.46	Y
26	3PL	245.63	8	59.41	462.18	Y
27	3PL	1492.10	8	371.03	461.89	Y
28	3PL	344.44	8	84.11	462.11	Y
29	3PL	347.46	8	84.86	462.29	Y
30	3PL	325.40	8	79.35	462.01	Y
31	3PL	581.90	8	143.48	461.73	Y
32	2PPC	586.47	17	97.66	460.39	Y
33	2PPC	469.91	17	77.67	458.62	Y
34	2PPC	613.98	35	69.20	458.19	Y
35	2PPC	430.10	17	70.85	461.98	Y
36	2PPC	749.24	17	125.58	460.58	Y
37	2PPC	834.61	17	140.22	459.33	Y
38	2PPC	1385.40	17	234.67	457.72	Y
39	2PPC	1101.80	17	186.03	456.97	Y
40	2PPC	324.55	35	34.61	455.25	Y

**Table O2. ELA Grade 4 Item Fit Statistics**

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	349.81	8	85.45	456.00	Y
2	3PL	346.60	8	84.65	455.95	Y
3	3PL	193.35	8	46.34	455.59	Y
4	3PL	399.37	8	97.84	455.60	Y
5	3PL	353.58	8	86.40	455.65	Y
6	3PL	170.65	8	40.66	455.64	Y
13	3PL	324.25	8	79.06	455.63	Y
14	3PL	418.17	8	102.54	455.51	Y
15	3PL	343.74	8	83.93	455.44	Y
16	3PL	393.09	8	96.27	455.46	Y
17	3PL	597.78	8	147.44	455.62	Y
18	3PL	397.64	8	97.41	455.54	Y
19	3PL	188.16	8	45.04	455.10	Y
20	3PL	356.33	8	87.08	455.38	Y
21	3PL	343.03	8	83.76	455.47	Y
22	3PL	224.29	8	54.07	455.31	Y
23	3PL	167.57	8	39.89	455.13	Y
24	3PL	321.56	8	78.39	454.98	Y
25	3PL	298.11	8	72.53	455.96	Y
26	3PL	238.50	8	57.62	455.82	Y
27	3PL	388.20	8	95.05	455.57	Y
28	3PL	586.61	8	144.65	455.72	Y
29	3PL	792.27	8	196.07	455.85	Y
30	3PL	279.26	8	67.82	455.71	Y
31	3PL	314.12	8	76.53	455.39	Y
32	2PPC	641.36	17	107.08	453.02	Y
33	2PPC	737.22	17	123.52	452.70	Y
34	2PPC	653.24	35	73.89	450.60	Y
35	2PPC	699.45	17	117.04	455.45	Y
36	2PPC	778.61	17	130.61	453.66	Y
37	2PPC	637.94	17	106.49	454.53	Y
38	2PPC	980.11	17	165.17	453.73	Y
39	2PPC	566.52	17	94.24	453.09	Y
40	2PPC	1043.40	35	120.52	452.61	Y

**Table O3. ELA Grade 5 Item Fit Statistics**

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	227.08	8	54.77	428.74	Y
2	3PL	236.39	8	57.10	428.09	Y
3	3PL	441.66	8	108.42	428.48	Y
4	3PL	160.71	8	38.18	428.55	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
5	3PL	203.69	8	48.92	428.55	Y
6	3PL	221.22	8	53.31	428.46	Y
7	3PL	164.36	8	39.09	428.50	Y
8	3PL	233.71	8	56.43	428.49	Y
9	3PL	188.11	8	45.03	428.25	Y
10	3PL	352.11	8	86.03	428.29	Y
11	3PL	178.36	8	42.59	428.34	Y
12	3PL	414.33	8	101.58	428.46	Y
13	3PL	247.61	8	59.90	428.45	Y
14	3PL	749.88	8	185.47	428.47	Y
15	3PL	136.35	8	32.09	428.33	Y
16	3PL	391.94	8	95.98	428.41	Y
17	3PL	332.77	8	81.19	428.20	Y
18	3PL	228.17	8	55.04	428.26	Y
19	3PL	393.46	8	96.37	428.25	Y
20	3PL	203.04	8	48.76	428.16	Y
21	3PL	236.68	8	57.17	428.22	Y
29	3PL	447.97	8	109.99	428.15	Y
30	3PL	103.86	8	23.96	428.17	Y
31	3PL	245.66	8	59.41	427.93	Y
32	3PL	481.93	8	118.48	428.02	Y
33	3PL	338.73	8	82.68	427.91	Y
34	3PL	314.01	8	76.50	428.10	Y
35	3PL	260.76	8	63.19	427.67	Y
36	3PL	408.71	8	100.18	428.65	Y
37	3PL	1692.60	8	421.14	428.56	Y
38	3PL	265.44	8	64.36	428.27	Y
39	3PL	660.86	8	163.21	428.54	Y
40	3PL	358.73	8	87.68	428.38	Y
41	3PL	283.52	8	68.88	428.52	Y
42	3PL	1044.70	8	259.18	428.44	Y
43	2PPC	247.25	17	39.49	427.90	Y
44	2PPC	1549.10	17	262.75	426.51	Y
45	2PPC	492.84	35	54.72	426.38	Y
46	2PPC	246.32	17	39.33	428.35	Y
47	2PPC	364.42	17	59.58	427.26	Y
48	2PPC	406.78	17	66.85	427.46	Y
49	2PPC	400.93	17	65.84	426.57	Y
50	2PPC	1045.10	17	176.31	426.13	Y
51	2PPC	523.93	35	58.44	425.21	Y

**Table O4. ELA Grade 6 Item Fit Statistics**

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	301.94	8	73.48	421.62	Y
2	3PL	365.33	8	89.33	421.58	Y
3	3PL	308.86	8	75.22	421.49	Y
4	3PL	207.70	8	49.92	421.34	Y
5	3PL	179.69	8	42.92	421.39	Y
6	3PL	271.06	8	65.77	421.51	Y
7	3PL	257.44	8	62.36	421.15	Y
8	3PL	197.67	8	47.42	421.37	Y
9	3PL	129.12	8	30.28	421.07	Y
10	3PL	303.25	8	73.81	421.39	Y
11	3PL	1199.80	8	297.95	421.16	Y
12	3PL	305.04	8	74.26	421.39	Y
13	3PL	446.20	8	109.55	421.35	Y
14	3PL	303.01	8	73.75	421.26	Y
22	3PL	181.47	8	43.37	420.92	Y
23	3PL	218.96	8	52.74	421.27	Y
24	3PL	275.28	8	66.82	421.28	Y
25	3PL	619.42	8	152.85	420.94	Y
26	3PL	339.11	8	82.78	420.92	Y
27	3PL	324.97	8	79.24	421.05	Y
28	3PL	251.32	8	60.83	420.88	Y
29	3PL	176.26	8	42.06	421.04	Y
30	3PL	394.90	8	96.72	420.95	Y
31	3PL	50.59	8	10.65	420.95	Y
32	3PL	246.78	8	59.69	420.85	Y
33	3PL	204.84	8	49.21	420.66	Y
34	3PL	124.87	8	29.22	420.95	Y
35	3PL	257.92	8	62.48	420.80	Y
36	3PL	233.82	8	56.46	421.58	Y
37	3PL	245.02	8	59.25	421.58	Y
38	3PL	171.61	8	40.90	421.44	Y
39	3PL	312.62	8	76.15	421.54	Y
40	3PL	338.40	8	82.60	421.53	Y
41	3PL	357.83	8	87.46	421.34	Y
42	3PL	274.76	8	66.69	421.15	Y
43	2PPC	416.19	17	68.46	420.57	Y
44	2PPC	632.56	17	105.57	419.55	Y
45	2PPC	654.86	35	74.09	419.36	Y
46	2PPC	373.44	17	61.13	420.98	Y
47	2PPC	446.41	17	73.64	419.02	Y
48	2PPC	317.72	17	51.57	420.60	Y
49	2PPC	517.21	17	85.79	420.11	Y
50	2PPC	1307.50	17	221.31	418.59	Y
51	2PPC	800.87	35	91.54	418.01	Y

**Table O5. ELA Grade 7 Item Fit Statistics**

<b>Item</b>	<b>Model</b>	<b>Chi Square</b>	<b>DF</b>	<b>Z-observed</b>	<b>Z-critical</b>	<b>Fit OK?</b>
1	3PL	250.76	8	60.69	396.81	Y
2	3PL	512.92	8	126.23	396.67	Y
3	3PL	240.66	8	58.17	396.58	Y
4	3PL	156.09	8	37.02	396.61	Y
5	3PL	104.75	8	24.19	396.50	Y
6	3PL	224.91	8	54.23	396.63	Y
7	3PL	214.12	8	51.53	396.60	Y
8	3PL	207.43	8	49.86	396.68	Y
9	3PL	257.55	8	62.39	396.43	Y
10	3PL	166.44	8	39.61	396.63	Y
11	3PL	143.88	8	33.97	396.50	Y
12	3PL	668.36	8	165.09	396.66	Y
13	3PL	156.22	8	37.06	396.57	Y
14	3PL	114.29	8	26.57	396.67	Y
15	3PL	350.14	8	85.54	396.68	Y
16	3PL	211.34	8	50.83	396.55	Y
17	3PL	156.44	8	37.11	396.56	Y
18	3PL	179.32	8	42.83	396.26	Y
19	3PL	272.30	8	66.07	396.54	Y
20	3PL	947.32	8	234.83	396.36	Y
21	3PL	132.81	8	31.20	396.43	Y
29	3PL	119.60	8	27.90	396.30	Y
30	3PL	151.73	8	35.93	396.33	Y
31	3PL	229.29	8	55.32	396.21	Y
32	3PL	359.12	8	87.78	396.04	Y
33	3PL	515.63	8	126.91	396.01	Y
34	3PL	175.09	8	41.77	396.19	Y
35	3PL	358.38	8	87.60	396.14	Y
36	3PL	196.23	8	47.06	396.78	Y
37	3PL	202.44	8	48.61	396.77	Y
38	3PL	86.28	8	19.57	396.67	Y
39	3PL	161.21	8	38.30	396.65	Y
40	3PL	156.72	8	37.18	396.74	Y
41	3PL	185.94	8	44.49	396.69	Y
42	3PL	232.91	8	56.23	396.29	Y
43	2PPC	891.90	17	150.04	394.60	Y
44	2PPC	542.53	17	90.13	392.98	Y
45	2PPC	611.24	35	68.87	393.13	Y
46	2PPC	288.95	17	46.64	396.07	Y
47	2PPC	581.55	17	96.82	394.37	Y
48	2PPC	557.00	17	92.61	393.97	Y
49	2PPC	786.40	17	131.95	393.03	Y
50	2PPC	283.03	17	45.62	389.74	Y
51	2PPC	810.93	35	92.74	389.19	Y

**Table O6. ELA Grade 8 Item Fit Statistics**

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	458.16	8	112.54	382.52	Y
2	3PL	124.52	8	29.13	382.61	Y
3	3PL	156.00	8	37.00	382.63	Y
4	3PL	271.25	8	65.81	382.59	Y
5	3PL	286.97	8	69.74	382.58	Y
6	3PL	154.48	8	36.62	382.46	Y
7	3PL	335.79	8	81.95	382.58	Y
8	3PL	147.40	8	34.85	382.54	Y
9	3PL	201.84	8	48.46	382.45	Y
10	3PL	210.10	8	50.52	382.51	Y
11	3PL	1172.00	8	290.99	382.35	Y
12	3PL	162.67	8	38.67	382.53	Y
13	3PL	295.91	8	71.98	382.55	Y
14	3PL	422.45	8	103.61	382.46	Y
22	3PL	213.61	8	51.40	382.13	Y
23	3PL	200.69	8	48.17	382.41	Y
24	3PL	125.65	8	29.41	382.34	Y
25	3PL	618.49	8	152.62	382.32	Y
26	3PL	497.05	8	122.26	382.13	Y
27	3PL	217.21	8	52.30	382.23	Y
28	3PL	1212.70	8	301.18	382.16	Y
29	3PL	311.53	8	75.88	382.21	Y
30	3PL	601.95	8	148.49	382.30	Y
31	3PL	460.82	8	113.21	382.10	Y
32	3PL	172.34	8	41.08	381.95	Y
33	3PL	256.50	8	62.13	381.97	Y
34	3PL	349.74	8	85.43	382.09	Y
35	3PL	666.98	8	164.74	381.96	Y
36	3PL	520.55	8	128.14	382.60	Y
37	3PL	253.02	8	61.26	382.57	Y
38	3PL	166.95	8	39.74	382.57	Y
39	3PL	166.13	8	39.53	382.55	Y
40	3PL	199.47	8	47.87	382.61	Y
41	3PL	299.47	8	72.87	382.49	Y
42	3PL	118.22	8	27.56	382.39	Y
43	2PPC	366.44	17	59.93	379.78	Y
44	2PPC	599.67	17	99.93	377.51	Y
45	2PPC	738.02	35	84.03	378.38	Y
46	2PPC	299.67	17	48.48	381.65	Y
47	2PPC	575.24	17	95.74	379.40	Y
48	2PPC	115.81	17	16.95	381.90	Y
49	2PPC	748.76	17	125.50	379.27	Y
50	2PPC	316.48	17	51.36	377.93	Y
51	2PPC	891.14	35	102.33	377.37	Y

**Table O7. Mathematics Grade 3 Item Fit Statistics**

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	303.61	8	73.90	476.73	Y
2	3PL	268.51	8	65.13	476.61	Y
3	3PL	893.51	8	221.38	473.43	Y
4	3PL	369.87	8	90.47	475.77	Y
6	3PL	250.90	8	60.73	475.74	Y
7	3PL	209.34	8	50.33	476.22	Y
8	3PL	273.17	8	66.29	475.62	Y
9	3PL	205.74	8	49.44	475.62	Y
11	3PL	307.03	8	74.76	476.35	Y
12	3PL	463.98	8	114.00	476.12	Y
13	3PL	216.34	8	52.09	475.95	Y
14	3PL	272.26	8	66.07	475.51	Y
16	3PL	172.75	8	41.19	475.53	Y
17	3PL	451.93	8	110.98	475.36	Y
19	3PL	424.76	8	104.19	475.97	Y
20	3PL	279.46	8	67.87	475.71	Y
21	3PL	282.93	8	68.73	475.35	Y
22	3PL	418.08	8	102.52	473.17	Y
23	3PL	541.44	8	133.36	476.75	Y
24	3PL	386.86	8	94.72	476.35	Y
25	3PL	463.31	8	113.83	475.73	Y
26	3PL	222.93	8	53.73	475.58	Y
27	3PL	310.19	8	75.55	475.98	Y
28	3PL	264.47	8	64.12	476.13	Y
30	3PL	449.57	8	110.39	476.00	Y
31	3PL	496.80	8	122.20	475.85	Y
32	3PL	321.08	8	78.27	476.18	Y
33	3PL	296.78	8	72.19	476.20	Y
34	3PL	458.79	8	112.70	476.35	Y
35	3PL	1086.50	8	269.62	476.37	Y
37	3PL	419.52	8	102.88	475.64	Y
38	3PL	367.11	8	89.78	475.73	Y
39	3PL	338.27	8	82.57	475.72	Y
40	3PL	718.47	8	177.62	475.90	Y
41	3PL	328.49	8	80.12	476.36	Y
42	3PL	265.40	8	64.35	475.93	Y
43	3PL	569.80	8	140.45	475.74	Y
45	2PPC	528.54	17	87.73	475.39	Y
46	2PPC	4028.90	17	688.03	476.41	N
47	2PPC	95.94	17	13.54	475.93	Y
48	2PPC	375.59	17	61.50	475.37	Y
49	2PPC	2004.30	17	340.82	475.68	Y
50	2PPC	139.24	26	15.70	475.11	Y
51	2PPC	1469.00	26	200.10	475.08	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
52	2PPC	358.62	26	46.13	474.51	Y

**Table O8. Mathematics Grade 4 Item Fit Statistics**

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	506.74	8	124.68	464.73	Y
2	3PL	128.53	8	30.13	464.52	Y
3	3PL	321.94	8	78.49	464.20	Y
4	3PL	666.65	8	164.66	464.14	Y
5	3PL	151.33	8	35.83	464.11	Y
6	3PL	924.75	8	229.19	464.15	Y
7	3PL	248.16	8	60.04	464.32	Y
8	3PL	252.83	8	61.21	463.97	Y
9	3PL	320.71	8	78.18	464.14	Y
10	3PL	251.39	8	60.85	464.10	Y
12	3PL	277.40	8	67.35	463.85	Y
13	3PL	476.52	8	117.13	464.08	Y
14	3PL	166.70	8	39.67	463.98	Y
16	3PL	216.84	8	52.21	463.49	Y
17	3PL	215.15	8	51.79	464.08	Y
18	3PL	414.45	8	101.61	463.88	Y
19	3PL	191.32	8	45.83	464.28	Y
20	3PL	236.51	8	57.13	464.10	Y
23	3PL	2375.50	8	591.88	464.54	N
24	3PL	252.99	8	61.25	464.54	Y
25	3PL	196.71	8	47.18	464.25	Y
26	3PL	244.78	8	59.19	464.24	Y
27	3PL	270.45	8	65.61	464.19	Y
28	3PL	138.99	8	32.75	464.34	Y
29	3PL	175.84	8	41.96	464.29	Y
30	3PL	246.55	8	59.64	464.00	Y
31	3PL	400.25	8	98.06	464.30	Y
32	3PL	423.09	8	103.77	464.05	Y
33	3PL	209.60	8	50.40	464.21	Y
34	3PL	435.32	8	106.83	464.25	Y
35	3PL	198.74	8	47.69	464.33	Y
37	3PL	394.93	8	96.73	464.34	Y
38	3PL	296.27	8	72.07	464.26	Y
39	3PL	439.00	8	107.75	463.77	Y
40	3PL	351.41	8	85.85	464.08	Y
42	3PL	140.91	8	33.23	464.11	Y
43	3PL	357.99	8	87.50	463.85	Y
45	3PL	281.27	8	68.32	462.97	Y
46	2PPC	2927.60	17	499.16	463.70	N

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
47	2PPC	212.74	17	33.57	463.67	Y
48	2PPC	2738.70	17	466.76	463.71	N
49	2PPC	533.90	17	88.65	463.43	Y
50	2PPC	1441.40	17	244.29	463.12	Y
51	2PPC	854.80	17	143.68	463.44	Y
52	2PPC	173.38	26	20.44	463.22	Y
53	2PPC	379.89	26	49.08	463.37	Y
54	2PPC	375.34	26	48.45	463.63	Y
55	2PPC	240.42	26	29.73	463.43	Y

**Table O9. Mathematics Grade 5 Item Fit Statistics**

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	586.68	8	144.67	433.69	Y
2	3PL	234.80	8	56.70	433.71	Y
3	3PL	1412.70	8	351.18	433.70	Y
4	3PL	1292.30	8	321.07	433.10	Y
5	3PL	152.67	8	36.17	433.54	Y
6	3PL	384.68	8	94.17	433.48	Y
8	3PL	129.49	8	30.37	433.51	Y
9	3PL	371.83	8	90.96	433.31	Y
10	3PL	187.29	8	44.82	433.43	Y
11	3PL	224.43	8	54.11	433.81	Y
13	3PL	1277.10	8	317.27	433.69	Y
14	3PL	152.22	8	36.06	433.49	Y
15	3PL	422.35	8	103.59	432.94	Y
16	3PL	755.05	8	186.76	433.23	Y
17	3PL	251.92	8	60.98	433.58	Y
18	3PL	158.53	8	37.63	433.35	Y
19	3PL	777.05	8	192.26	433.10	Y
20	3PL	541.44	8	133.36	433.22	Y
23	3PL	538.78	8	132.69	433.73	Y
24	3PL	124.01	8	29.00	433.79	Y
25	3PL	165.52	8	39.38	433.82	Y
26	3PL	1622.40	8	403.60	433.62	Y
27	3PL	173.63	8	41.41	433.66	Y
28	3PL	168.79	8	40.20	433.53	Y
29	3PL	150.96	8	35.74	433.63	Y
31	3PL	460.49	8	113.12	433.52	Y
33	3PL	201.50	8	48.37	433.66	Y
34	3PL	202.97	8	48.74	433.55	Y
36	3PL	98.08	8	22.52	433.45	Y
37	3PL	446.87	8	109.72	433.69	Y
39	3PL	186.00	8	44.50	433.42	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
40	3PL	194.04	8	46.51	433.58	Y
41	3PL	484.29	8	119.07	433.56	Y
42	3PL	1034.10	8	256.53	433.41	Y
43	3PL	517.23	8	127.31	433.47	Y
44	3PL	679.06	8	167.76	433.57	Y
45	3PL	284.58	8	69.15	432.54	Y
46	2PPC	946.06	17	159.33	433.86	Y
47	2PPC	1988.50	17	338.12	433.01	Y
48	2PPC	604.59	17	100.77	433.45	Y
49	2PPC	992.21	17	167.25	430.10	Y
50	2PPC	866.78	17	145.74	432.29	Y
51	2PPC	358.81	17	58.62	431.16	Y
52	2PPC	1556.50	26	212.25	432.08	Y
53	2PPC	302.91	26	38.40	432.21	Y
54	2PPC	210.07	26	25.53	432.05	Y
55	2PPC	593.98	26	78.76	424.71	Y

Table O10. Mathematics Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	307.00	8	74.75	428.78	Y
2	3PL	437.88	8	107.47	428.66	Y
4	3PL	1038.20	8	257.55	428.04	Y
5	3PL	160.67	8	38.17	428.49	Y
7	3PL	196.20	8	47.05	428.42	Y
8	3PL	98.31	8	22.58	428.52	Y
9	3PL	1066.60	8	264.66	428.01	Y
11	3PL	172.04	8	41.01	428.42	Y
12	3PL	235.81	8	56.95	427.98	Y
13	3PL	230.58	8	55.65	427.95	Y
14	3PL	682.48	8	168.62	428.39	Y
15	3PL	572.52	8	141.13	428.54	Y
16	3PL	160.49	8	38.12	428.19	Y
17	3PL	279.89	8	67.97	427.74	Y
18	3PL	332.02	8	81.01	428.42	Y
19	3PL	394.84	8	96.71	428.34	Y
20	3PL	248.16	8	60.04	428.33	Y
21	3PL	188.49	8	45.12	428.34	Y
22	3PL	321.81	8	78.45	428.42	Y
25	3PL	215.52	8	51.88	427.94	Y
26	3PL	740.14	8	183.03	427.01	Y
27	3PL	181.52	8	43.38	428.73	Y
28	3PL	325.99	8	79.50	428.41	Y
29	3PL	397.32	8	97.33	428.46	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
30	3PL	206.81	8	49.70	428.25	Y
31	3PL	211.87	8	50.97	428.45	Y
33	3PL	177.36	8	42.34	428.37	Y
34	3PL	239.04	8	57.76	428.40	Y
35	3PL	582.97	8	143.74	428.39	Y
36	3PL	509.71	8	125.43	428.33	Y
37	3PL	57.85	8	12.46	428.53	Y
38	3PL	148.47	8	35.12	428.37	Y
39	3PL	634.36	8	156.59	428.34	Y
40	3PL	177.79	8	42.45	428.35	Y
41	3PL	175.28	8	41.82	428.55	Y
42	3PL	724.70	8	179.18	428.21	Y
43	3PL	335.08	8	81.77	428.27	Y
44	3PL	125.58	8	29.40	428.34	Y
45	3PL	244.67	8	59.17	428.22	Y
46	3PL	178.88	8	42.72	428.44	Y
47	3PL	216.86	8	52.22	428.24	Y
48	3PL	1369.60	8	340.39	428.54	Y
49	3PL	321.78	8	78.45	428.16	Y
52	2PPC	2278.00	17	387.77	428.42	Y
53	2PPC	55.02	17	6.52	426.96	Y
54	2PPC	521.02	17	86.44	427.80	Y
55	2PPC	80.61	17	10.91	427.54	Y
56	2PPC	444.72	17	73.35	426.57	Y
57	2PPC	467.13	17	77.20	426.47	Y
58	2PPC	55.92	26	4.15	422.90	Y
59	2PPC	301.12	26	38.15	425.82	Y
60	2PPC	80.19	26	7.51	426.27	Y
61	2PPC	159.71	26	18.54	426.93	Y

Table O11. Mathematics Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	259.06	8	62.77	391.06	Y
2	3PL	149.69	8	35.42	390.51	Y
4	3PL	208.94	8	50.24	390.32	Y
6	3PL	2010.20	8	500.55	391.23	N
7	3PL	122.78	8	28.70	390.66	Y
8	3PL	87.49	8	19.87	391.19	Y
9	3PL	80.77	8	18.19	391.05	Y
10	3PL	87.96	8	19.99	390.64	Y
11	3PL	169.06	8	40.27	391.14	Y
12	3PL	129.30	8	30.33	391.22	Y
13	3PL	170.45	8	40.61	391.07	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
14	3PL	221.38	8	53.35	391.01	Y
15	3PL	248.71	8	60.18	390.67	Y
16	3PL	614.98	8	151.74	391.17	Y
17	3PL	416.81	8	102.20	390.81	Y
18	3PL	468.69	8	115.17	390.66	Y
20	3PL	141.55	8	33.39	390.55	Y
21	3PL	118.09	8	27.52	390.53	Y
22	3PL	135.50	8	31.88	390.73	Y
23	3PL	319.88	8	77.97	390.65	Y
24	3PL	269.43	8	65.36	390.35	Y
25	3PL	503.25	8	123.81	390.58	Y
27	3PL	640.40	8	158.10	391.58	Y
28	3PL	124.51	8	29.13	391.22	Y
29	3PL	142.94	8	33.74	391.17	Y
30	3PL	100.96	8	23.24	390.51	Y
31	3PL	178.05	8	42.51	391.17	Y
33	3PL	249.15	8	60.29	391.27	Y
34	3PL	292.44	8	71.11	390.95	Y
35	3PL	146.58	8	34.64	390.96	Y
36	3PL	111.89	8	25.97	390.97	Y
37	3PL	155.90	8	36.97	391.16	Y
38	3PL	102.44	8	23.61	390.69	Y
39	3PL	120.69	8	28.17	391.04	Y
40	3PL	375.62	8	91.91	391.17	Y
41	3PL	84.38	8	19.10	391.12	Y
42	3PL	114.98	8	26.75	391.14	Y
43	3PL	140.90	8	33.22	390.69	Y
44	3PL	207.37	8	49.84	391.11	Y
45	3PL	509.66	8	125.42	391.14	Y
46	3PL	935.81	8	231.95	390.95	Y
47	3PL	428.14	8	105.04	391.17	Y
48	3PL	76.37	8	17.09	391.30	Y
49	3PL	98.89	8	22.72	391.00	Y
52	2PPC	211.22	17	33.31	387.69	Y
53	2PPC	157.82	17	24.15	390.04	Y
54	2PPC	343.84	17	56.05	390.18	Y
55	2PPC	885.78	17	148.99	389.33	Y
56	2PPC	408.19	17	67.09	386.03	Y
57	2PPC	278.32	17	44.82	384.78	Y
58	2PPC	791.74	26	106.19	385.96	Y
59	2PPC	94.40	26	9.49	386.96	Y
60	2PPC	318.05	26	40.50	387.30	Y
61	2PPC	132.71	26	14.80	389.03	Y

**Table O12. Mathematics Grade 8 Item Fit Statistics**

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	846.02	8	209.50	306.29	Y
2	3PL	176.35	8	42.09	306.33	Y
3	3PL	117.28	8	27.32	306.13	Y
4	3PL	413.44	8	101.36	306.28	Y
5	3PL	90.68	8	20.67	305.84	Y
6	3PL	112.17	8	26.04	305.85	Y
7	3PL	168.83	8	40.21	305.97	Y
8	3PL	256.10	8	62.02	306.22	Y
9	3PL	82.82	8	18.70	305.98	Y
10	3PL	205.67	8	49.42	306.18	Y
11	3PL	186.62	8	44.65	306.11	Y
12	3PL	146.52	8	34.63	305.93	Y
15	3PL	251.49	8	60.87	306.04	Y
16	3PL	124.88	8	29.22	305.98	Y
17	3PL	222.34	8	53.58	305.99	Y
19	3PL	444.22	8	109.06	306.11	Y
20	3PL	766.14	8	189.53	306.10	Y
21	3PL	79.63	8	17.91	306.15	Y
22	3PL	394.96	8	96.74	306.17	Y
24	3PL	1596.70	8	397.18	305.82	N
25	3PL	90.91	8	20.73	305.93	Y
26	3PL	187.97	8	44.99	305.86	Y
27	3PL	326.28	8	79.57	306.27	Y
28	3PL	365.56	8	89.39	305.94	Y
29	3PL	491.96	8	120.99	306.14	Y
30	3PL	252.44	8	61.11	306.14	Y
32	3PL	211.52	8	50.88	305.99	Y
33	3PL	294.86	8	71.71	306.03	Y
34	3PL	124.57	8	29.14	306.05	Y
35	3PL	166.68	8	39.67	305.94	Y
36	3PL	159.22	8	37.80	305.52	Y
37	3PL	268.65	8	65.16	306.18	Y
38	3PL	116.21	8	27.05	306.14	Y
39	3PL	150.37	8	35.59	306.21	Y
40	3PL	180.61	8	43.15	305.69	Y
41	3PL	490.43	8	120.61	306.27	Y
42	3PL	388.05	8	95.01	306.17	Y
44	3PL	98.72	8	22.68	306.14	Y
45	3PL	192.30	8	46.07	306.16	Y
46	3PL	92.57	8	21.14	306.24	Y
47	3PL	57.56	8	12.39	306.20	Y
48	3PL	80.29	8	18.07	306.11	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
49	3PL	75.11	8	16.78	306.01	Y
50	3PL	75.68	8	16.92	306.02	Y
52	2PPC	530.22	17	88.02	303.06	Y
53	2PPC	97.91	17	13.88	303.45	Y
54	2PPC	53.56	17	6.27	294.81	Y
55	2PPC	101.00	17	14.41	299.91	Y
56	2PPC	72.44	17	9.51	299.51	Y
57	2PPC	58.72	17	7.16	294.82	Y
58	2PPC	93.32	26	9.34	297.93	Y
59	2PPC	44.28	26	2.53	295.94	Y
60	2PPC	113.97	26	12.20	295.70	Y
61	2PPC	88.64	26	8.69	293.73	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	1.039	-0.270	0.294		
2	1	1.095	-1.413	0.148		
3	1	0.886	0.384	0.180		
4	1	0.979	-1.619	0.005		
5	1	0.639	-1.086	0.010		
6	1	0.790	0.519	0.134		
13	1	0.641	-0.369	0.041		
14	1	0.713	0.158	0.191		
15	1	1.004	1.167	0.243		
16	1	0.638	0.769	0.189		
17	1	0.796	0.706	0.230		
18	1	0.596	0.814	0.229		
19	1	0.690	-0.417	0.183		
20	1	1.194	0.803	0.194		
21	1	0.828	0.522	0.145		
22	1	1.000	1.032	0.191		
23	1	0.727	1.031	0.254		
24	1	1.027	0.005	0.226		
25	1	0.943	-0.377	0.169		
26	1	0.743	-0.047	0.184		
27	1	1.275	1.220	0.143		
28	1	0.970	-0.691	0.177		
29	1	0.507	-0.369	0.087		
30	1	0.924	0.408	0.192		
31	1	1.111	0.909	0.213		
32	2	1.394	-2.012	1.194		
33	2	1.362	-0.853	1.847		
34	4	1.383	-1.624	0.388	2.278	4.122

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
35	2	1.678	-1.441	1.885		
36	2	1.463	-1.647	1.825		
37	2	1.416	-1.123	1.965		
38	2	1.431	-0.942	2.265		
39	2	1.686	-0.402	2.507		
40	4	1.348	-0.544	1.160	2.626	4.019

**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	0.845	0.510	0.213		
2	1	0.711	0.524	0.105		
3	1	0.443	-0.661	0.039		
4	1	0.325	-0.076	0.004		
5	1	0.758	-0.197	0.125		
6	1	0.313	-0.787	0.034		
13	1	0.527	1.001	0.095		
14	1	0.893	0.985	0.208		
15	1	0.823	0.275	0.159		
16	1	0.925	0.470	0.214		
17	1	0.645	-0.229	0.055		
18	1	0.771	0.399	0.149		
19	1	0.608	1.207	0.235		
20	1	0.389	-0.031	0.007		
21	1	0.683	-0.181	0.130		
22	1	0.728	1.346	0.244		
23	1	0.402	-0.393	0.127		
24	1	0.719	1.053	0.182		
25	1	0.651	-0.541	0.097		
26	1	0.522	1.236	0.109		
27	1	1.001	1.444	0.305		
28	1	0.674	1.211	0.150		
29	1	0.383	0.041	0.007		
30	1	0.612	-0.270	0.144		
31	1	0.825	-0.132	0.260		
32	2	1.535	-1.570	1.442		
33	2	1.424	-1.679	1.282		
34	4	1.449	-1.644	0.084	1.574	3.048
35	2	1.424	-2.391	1.364		
36	2	1.764	-1.614	1.262		
37	2	1.579	-2.570	-0.150		
38	2	1.800	-2.231	0.963		
39	2	1.854	-2.247	1.410		
40	4	1.774	-1.999	-0.104	1.701	3.242

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	0.587	-2.174	0.168		
2	1	0.670	-0.236	0.181		
3	1	1.020	-0.013	0.180		
4	1	0.430	-1.070	0.128		
5	1	0.415	0.559	0.170		
6	1	0.621	1.348	0.274		
7	1	0.584	-2.205	0.016		
8	1	0.640	-1.456	0.059		
9	1	0.500	-1.451	0.049		
10	1	0.551	-0.585	0.079		
11	1	0.569	1.629	0.269		
12	1	0.831	0.629	0.213		
13	1	0.972	-1.231	0.191		
14	1	0.677	-1.051	0.050		
15	1	0.410	0.929	0.261		
16	1	0.545	-0.328	0.081		
17	1	0.862	-0.073	0.212		
18	1	0.737	-0.945	0.158		
19	1	0.834	0.449	0.191		
20	1	0.458	-0.973	0.086		
21	1	0.631	0.423	0.195		
29	1	0.934	1.779	0.282		
30	1	0.199	-0.067	0.023		
31	1	0.675	0.754	0.240		
32	1	1.164	-0.128	0.264		
33	1	0.831	-0.033	0.214		
34	1	0.784	0.339	0.257		
35	1	0.647	1.123	0.218		
36	1	0.552	1.343	0.153		
37	1	0.214	-2.759	0.005		
38	1	0.657	0.027	0.164		
39	1	0.447	-1.778	0.003		
40	1	0.911	-0.324	0.205		
41	1	0.631	-1.578	0.023		
42	1	0.548	-2.069	0.003		
43	2	1.187	-3.478	-0.371		
44	2	1.235	-2.217	0.245		
45	4	1.109	-2.480	-0.786	0.995	2.437
46	2	1.455	-3.856	-0.652		
47	2	1.170	-2.565	-0.086		
48	2	1.312	-2.232	0.345		
49	2	1.184	-1.733	0.598		
50	2	1.490	-2.057	-0.126		

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
51	4	1.216	-1.679	-0.335	1.157	2.699

**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	0.419	-0.971	0.005		
2	1	0.750	-0.252	0.301		
3	1	1.035	-0.170	0.263		
4	1	0.472	-0.208	0.139		
5	1	0.406	-1.059	0.028		
6	1	0.948	-0.604	0.203		
7	1	0.755	1.640	0.208		
8	1	0.563	2.005	0.212		
9	1	0.480	0.635	0.226		
10	1	0.538	-0.570	0.078		
11	1	0.208	0.118	0.004		
12	1	0.664	-0.915	0.062		
13	1	0.797	1.520	0.200		
14	1	0.708	-0.073	0.133		
22	1	0.592	0.075	0.228		
23	1	0.662	0.967	0.230		
24	1	0.842	-0.220	0.239		
25	1	1.181	0.994	0.260		
26	1	0.786	1.144	0.224		
27	1	0.264	0.274	0.008		
28	1	0.673	0.571	0.192		
29	1	0.567	-0.394	0.190		
30	1	0.544	0.222	0.162		
31	1	0.203	2.783	0.070		
32	1	0.724	0.142	0.217		
33	1	0.302	0.538	0.034		
34	1	0.384	0.510	0.144		
35	1	0.723	-0.030	0.161		
36	1	0.825	1.928	0.308		
37	1	0.821	-0.679	0.274		
38	1	0.493	1.960	0.184		
39	1	0.844	0.426	0.216		
40	1	0.915	0.847	0.269		
41	1	0.860	0.266	0.161		
42	1	0.738	0.558	0.300		
43	2	1.217	-2.298	-0.174		
44	2	1.709	-2.554	-0.466		
45	4	1.637	-3.336	-1.498	0.517	2.542
46	2	1.319	-2.380	-0.110		
47	2	1.481	-2.039	0.675		
48	2	1.659	-3.786	-0.683		

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
49	2	1.673	-3.134	-0.081		
50	2	1.324	-2.364	-0.003		
51	4	1.612	-2.899	-1.505	0.239	1.924

**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	0.909	0.426	0.138		
2	1	0.221	-1.965	0.008		
3	1	0.999	-0.046	0.259		
4	1	0.608	0.335	0.124		
5	1	0.518	0.634	0.220		
6	1	0.593	-0.576	0.148		
7	1	0.900	-0.117	0.165		
8	1	0.588	-0.027	0.082		
9	1	0.942	0.426	0.215		
10	1	0.643	0.096	0.163		
11	1	0.740	-0.752	0.188		
12	1	1.006	1.444	0.153		
13	1	0.760	0.004	0.232		
14	1	0.508	0.170	0.159		
15	1	0.340	-0.127	0.006		
16	1	0.673	0.300	0.135		
17	1	0.596	0.545	0.180		
18	1	0.521	1.171	0.186		
19	1	1.373	-0.384	0.249		
20	1	0.414	-0.224	0.004		
21	1	0.279	0.090	0.018		
29	1	0.332	1.490	0.119		
30	1	0.487	0.745	0.092		
31	1	0.770	1.813	0.236		
32	1	1.185	0.627	0.254		
33	1	1.215	1.097	0.202		
34	1	0.748	0.665	0.257		
35	1	0.722	-0.558	0.157		
36	1	0.602	-0.580	0.074		
37	1	0.643	-1.339	0.066		
38	1	0.346	0.828	0.215		
39	1	0.632	0.219	0.149		
40	1	0.461	1.444	0.176		
41	1	0.643	-0.068	0.174		
42	1	0.961	0.473	0.302		
43	2	1.483	-1.685	0.197		
44	2	1.828	-2.308	-0.305		
45	4	1.582	-2.994	-1.157	0.827	2.437
46	2	1.612	-3.167	-0.427		

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
47	2	1.829	-2.690	0.015		
48	2	1.634	-2.186	0.508		
49	2	1.694	-1.889	0.237		
50	2	1.755	-1.777	0.649		
51	4	1.677	-1.967	-0.557	1.081	2.481

**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	1.090	0.598	0.341		
2	1	0.315	-0.442	0.013		
3	1	1.088	-2.314	0.075		
4	1	0.766	-1.941	0.015		
5	1	0.810	0.211	0.272		
6	1	0.580	-0.934	0.064		
7	1	0.452	-1.436	0.005		
8	1	0.983	-1.233	0.156		
9	1	0.601	-0.354	0.137		
10	1	0.797	-0.065	0.226		
11	1	0.108	-0.792	0.012		
12	1	1.324	-1.115	0.201		
13	1	0.427	-1.228	0.012		
14	1	0.879	1.135	0.216		
22	1	0.990	-0.337	0.240		
23	1	0.936	-0.365	0.131		
24	1	0.275	-0.254	0.007		
25	1	1.329	0.459	0.186		
26	1	1.199	0.129	0.251		
27	1	1.033	-0.371	0.219		
28	1	0.456	-0.257	0.004		
29	1	0.848	0.155	0.225		
30	1	0.550	-1.205	0.005		
31	1	1.182	0.125	0.288		
32	1	0.539	0.470	0.157		
33	1	1.110	-0.087	0.229		
34	1	0.980	0.212	0.193		
35	1	0.762	0.067	0.171		
36	1	1.005	0.359	0.172		
37	1	0.707	0.300	0.197		
38	1	0.880	-0.520	0.162		
39	1	0.845	-0.489	0.179		
40	1	0.519	-1.138	0.006		
41	1	0.815	0.450	0.202		
42	1	0.903	-1.190	0.223		
43	2	1.191	-2.071	-0.273		
44	2	1.675	-2.556	-0.390		

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3	step4
45	4	1.594	-3.054	-1.451	0.490	2.081
46	2	1.440	-3.558	-0.327		
47	2	1.565	-2.673	-0.262		
48	2	2.114	-4.680	-1.502		
49	2	1.943	-3.541	-0.780		
50	2	1.839	-2.652	0.032		
51	4	1.655	-3.216	-1.923	-0.042	1.795

Table O19. Mathematics Grade 3 OP Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	0.794	-0.840	0.219	
2	1	0.807	-0.729	0.201	
3	1	1.470	1.752	0.211	
4	1	0.623	-2.321	0.017	
6	1	1.074	0.068	0.327	
7	1	0.798	-0.859	0.357	
8	1	0.848	0.190	0.151	
9	1	0.816	0.329	0.210	
11	1	0.413	-3.020	0.012	
12	1	0.847	-1.096	0.135	
13	1	0.719	0.322	0.134	
14	1	0.854	0.253	0.257	
16	1	0.537	-0.294	0.160	
17	1	1.304	0.312	0.165	
19	1	1.313	-0.087	0.150	
20	1	1.185	-0.866	0.335	
21	1	1.108	-0.219	0.305	
22	1	1.028	0.644	0.171	
23	1	0.872	-1.393	0.044	
24	1	1.098	0.123	0.093	
25	1	1.238	0.291	0.095	
26	1	0.843	-0.257	0.280	
27	1	0.840	0.184	0.274	
28	1	0.702	-0.687	0.163	
30	1	1.230	1.135	0.301	
31	1	0.572	-2.342	0.008	
32	1	0.975	-0.252	0.134	
33	1	1.040	0.223	0.206	
34	1	0.603	-2.282	0.012	
35	1	0.819	-1.234	0.004	
37	1	0.586	0.150	0.049	
38	1	1.117	0.336	0.233	
39	1	0.769	0.966	0.113	
40	1	1.100	-0.994	0.101	
41	1	1.285	0.285	0.197	

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
42	1	1.174	0.385	0.253	
43	1	1.161	-0.198	0.063	
45	2	1.159	0.004	1.243	
46	2	0.567	-1.659	0.472	
47	2	0.973	0.445	-1.522	
48	2	1.263	1.164	2.380	
49	2	1.202	-0.676	0.625	
50	3	0.576	2.288	0.759	-1.787
51	3	0.596	0.875	0.255	-1.004
52	3	1.215	1.660	0.229	1.185

**Table O20. Mathematics Grade 4 OP Item Parameter Estimates**

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	0.790	-1.419	0.088	
2	1	1.159	-0.346	0.355	
3	1	1.271	-0.405	0.165	
4	1	1.397	0.475	0.120	
5	1	1.006	0.105	0.266	
6	1	0.870	-0.730	0.043	
7	1	0.502	-0.133	0.226	
8	1	0.944	-0.206	0.290	
9	1	1.261	-0.572	0.143	
10	1	1.305	0.012	0.063	
12	1	1.033	0.749	0.205	
13	1	1.468	0.755	0.136	
14	1	0.516	0.924	0.067	
16	1	1.226	-0.410	0.166	
17	1	1.004	-0.010	0.328	
18	1	0.996	0.041	0.170	
19	1	0.907	-0.179	0.144	
20	1	0.903	-0.313	0.348	
23	1	0.384	-1.760	0.002	
24	1	0.702	-0.100	0.226	
25	1	0.937	0.028	0.198	
26	1	0.874	-0.010	0.137	
27	1	0.971	0.274	0.063	
28	1	0.864	-0.390	0.321	
29	1	1.024	-0.618	0.294	
30	1	1.034	-0.312	0.155	
31	1	0.732	-0.664	0.068	
32	1	0.927	-0.547	0.056	
33	1	1.112	0.406	0.211	
34	1	0.907	0.726	0.146	

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
35	1	0.646	0.272	0.130	
37	1	1.169	-0.332	0.063	
38	1	1.326	-0.423	0.118	
39	1	1.224	0.292	0.239	
40	1	1.052	0.469	0.108	
42	1	1.173	0.367	0.438	
43	1	0.896	-0.384	0.094	
45	1	1.379	0.046	0.103	
46	2	1.001	0.138	0.503	
47	2	0.911	-1.681	0.055	
48	2	0.939	-0.556	-0.467	
49	2	1.231	0.134	1.428	
50	2	1.327	-0.405	1.016	
51	2	0.689	-1.130	0.558	
52	3	1.292	1.488	1.071	2.900
53	3	0.652	-0.928	1.749	-1.656
54	3	1.113	0.319	0.094	0.083
55	3	0.939	0.404	0.146	-0.041

**Table O21. Mathematics Grade 5 OP Item Parameter Estimates**

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	1.164	0.288	0.163	
2	1	1.187	-0.055	0.141	
3	1	1.195	-1.371	0.006	
4	1	0.056	-4.746	0.050	
5	1	0.882	-0.066	0.259	
6	1	0.925	0.923	0.205	
8	1	1.118	0.633	0.230	
9	1	0.612	0.596	0.039	
10	1	1.047	0.126	0.177	
11	1	1.177	-0.782	0.318	
13	1	0.403	-1.288	0.003	
14	1	1.047	0.138	0.250	
15	1	1.313	1.187	0.186	
16	1	0.942	0.443	0.088	
17	1	1.340	0.008	0.137	
18	1	1.087	0.443	0.217	
19	1	1.761	1.407	0.154	
20	1	1.189	1.214	0.091	
23	1	0.828	-0.900	0.021	
24	1	0.389	0.581	0.224	
25	1	0.641	-0.415	0.297	
26	1	0.755	-1.165	0.002	

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
27	1	1.166	0.780	0.166	
28	1	0.741	0.132	0.201	
29	1	1.190	0.197	0.239	
31	1	1.220	-0.674	0.108	
33	1	0.903	0.461	0.109	
34	1	0.675	0.301	0.206	
36	1	0.679	1.162	0.300	
37	1	0.584	0.428	0.027	
39	1	1.310	0.080	0.285	
40	1	0.839	0.577	0.313	
41	1	0.890	-0.604	0.091	
42	1	2.297	0.500	0.155	
43	1	1.622	0.397	0.130	
44	1	1.869	1.033	0.142	
45	1	0.869	-0.431	0.237	
46	2	1.166	-0.844	0.419	
47	2	1.063	-1.115	1.423	
48	2	1.321	-1.358	-0.454	
49	2	1.076	-0.141	1.101	
50	2	1.098	0.035	1.492	
51	2	0.496	2.356	-2.513	
52	3	1.052	0.012	0.159	0.421
53	3	1.316	1.222	1.648	2.192
54	3	1.173	2.390	1.341	1.102
55	3	0.792	0.588	3.239	0.345

Table O22. Mathematics Grade 6 OP Item Parameter Estimates

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	0.664	-0.414	0.539	
2	1	0.598	-0.656	0.152	
4	1	0.626	-0.733	0.006	
5	1	0.813	0.728	0.261	
7	1	0.961	0.218	0.279	
8	1	1.276	1.009	0.577	
9	1	1.580	2.090	0.070	
11	1	0.692	0.191	0.249	
12	1	1.040	0.515	0.205	
13	1	1.324	0.800	0.201	
14	1	0.968	-0.760	0.170	
15	1	1.162	0.945	0.062	
16	1	1.414	1.388	0.229	
17	1	1.230	1.284	0.327	
18	1	0.881	-0.105	0.163	

Appendix O: IRT Statistics

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
19	1	1.072	0.701	0.156	
20	1	0.969	0.259	0.266	
21	1	1.448	0.555	0.228	
22	1	0.698	0.583	0.301	
25	1	0.386	0.039	0.227	
26	1	1.681	1.672	0.190	
27	1	1.041	0.541	0.697	
28	1	1.313	-0.173	0.265	
29	1	1.044	-0.369	0.213	
30	1	1.345	1.087	0.154	
31	1	1.311	-0.018	0.330	
33	1	1.072	0.698	0.151	
34	1	1.139	0.597	0.379	
35	1	1.366	1.323	0.416	
36	1	1.249	-0.605	0.264	
37	1	0.668	2.383	0.183	
38	1	1.538	0.784	0.195	
39	1	1.116	1.112	0.189	
40	1	0.947	0.916	0.208	
41	1	1.404	0.462	0.289	
42	1	0.670	-0.628	0.057	
43	1	0.968	1.305	0.131	
44	1	0.756	2.408	0.169	
45	1	1.129	1.367	0.241	
46	1	0.808	0.680	0.154	
47	1	0.623	1.436	0.218	
48	1	0.924	-1.415	0.116	
49	1	1.742	0.478	0.214	
52	2	1.328	-1.518	1.291	
53	2	0.892	1.867	-0.864	
54	2	0.977	-0.601	0.434	
55	2	1.156	1.678	0.100	
56	2	1.235	-0.037	1.751	
57	2	1.409	1.249	1.881	
58	3	0.712	1.993	0.991	0.536
59	3	1.103	1.226	-0.455	3.134
60	3	1.222	1.950	3.495	1.677
61	3	0.948	0.638	1.559	-0.792

**Table O23. Mathematics Grade 7 OP Item Parameter Estimates**

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	1.192	0.038	0.346	
2	1	1.379	1.343	0.243	
4	1	1.749	1.254	0.292	
6	1	0.569	-1.587	0.004	
7	1	0.819	1.324	0.256	
8	1	0.902	0.637	0.239	
9	1	0.965	0.835	0.215	
10	1	1.119	0.720	0.171	
11	1	1.342	0.029	0.322	
12	1	0.719	0.462	0.236	
13	1	1.090	0.780	0.241	
14	1	1.272	1.242	0.124	
15	1	1.501	0.862	0.225	
16	1	0.638	0.166	0.245	
17	1	0.717	0.587	0.271	
18	1	1.521	1.488	0.177	
20	1	1.346	0.274	0.221	
21	1	1.621	1.129	0.265	
22	1	1.275	1.268	0.199	
23	1	1.287	1.212	0.303	
24	1	1.565	1.543	0.214	
25	1	1.303	-0.261	0.189	
27	1	1.069	0.054	0.294	
28	1	1.306	0.699	0.227	
29	1	1.276	0.668	0.171	
30	1	1.531	0.733	0.248	
31	1	1.331	0.625	0.206	
33	1	1.349	-0.042	0.228	
34	1	1.322	1.469	0.262	
35	1	1.633	0.754	0.202	
36	1	1.676	0.690	0.238	
37	1	0.730	1.118	0.174	
38	1	1.576	1.109	0.269	
39	1	1.076	1.627	0.218	
40	1	1.206	0.413	0.222	
41	1	1.159	1.190	0.194	
42	1	0.636	1.531	0.344	
43	1	1.270	0.151	0.257	
44	1	0.881	1.305	0.303	
45	1	1.257	1.116	0.172	
46	1	0.587	-0.243	0.050	
47	1	1.863	0.526	0.228	

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
48	1	0.949	0.744	0.289	
49	1	1.287	0.822	0.237	
52	2	1.120	1.851	0.371	
53	2	1.862	0.502	1.072	
54	2	1.270	-0.914	0.502	
55	2	1.022	-0.244	0.992	
56	2	1.616	0.978	2.778	
57	2	1.042	-0.247	0.124	
58	3	0.746	1.222	0.390	0.819
59	3	0.829	0.629	0.737	1.661
60	3	1.447	0.719	0.573	2.795
61	3	1.207	0.709	-0.038	0.459

**Table O24. Mathematics Grade 8 OP Item Parameter Estimates**

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
1	1	0.757	-1.342	0.328	
2	1	1.289	0.311	0.222	
3	1	0.836	0.684	0.229	
4	1	0.647	-0.356	0.163	
5	1	1.102	0.265	0.323	
6	1	1.572	0.864	0.380	
7	1	0.987	0.520	0.173	
8	1	0.486	0.211	0.111	
9	1	0.721	0.894	0.160	
10	1	0.865	0.912	0.393	
11	1	0.938	0.120	0.256	
12	1	1.144	0.572	0.301	
15	1	1.249	1.156	0.112	
16	1	1.230	1.356	0.251	
17	1	0.986	-0.078	0.193	
19	1	0.853	-0.206	0.097	
20	1	1.014	-0.799	0.194	
21	1	1.350	1.641	0.234	
22	1	0.904	-0.725	0.353	
24	1	0.384	-1.178	0.003	
25	1	1.110	0.577	0.499	
26	1	0.827	0.164	0.198	
27	1	0.690	-0.241	0.304	
28	1	1.001	-0.090	0.093	
29	1	0.858	-0.220	0.148	
30	1	0.751	0.133	0.182	
32	1	1.366	1.001	0.163	
33	1	1.118	1.112	0.435	

Appendix O: IRT Statistics

Item	Max Pts	a-par / alpha	b-par / step1	c-par / step2	step3
34	1	1.315	0.341	0.226	
35	1	0.928	0.159	0.316	
36	1	1.242	0.628	0.275	
37	1	0.859	-0.215	0.321	
38	1	0.992	0.240	0.242	
39	1	1.248	1.348	0.308	
40	1	1.044	0.533	0.243	
41	1	0.797	-0.702	0.311	
42	1	0.935	-0.358	0.233	
44	1	0.776	0.567	0.200	
45	1	0.684	0.335	0.169	
46	1	0.776	0.608	0.225	
47	1	0.790	1.077	0.242	
48	1	1.111	0.652	0.208	
49	1	1.264	1.003	0.311	
50	1	1.291	1.380	0.231	
52	2	0.729	-0.105	0.708	
53	2	0.877	0.281	0.529	
54	2	0.875	1.275	-0.647	
55	2	1.156	0.745	-0.642	
56	2	1.313	1.018	1.211	
57	2	0.914	1.201	-0.546	
58	3	0.872	1.414	0.455	0.164
59	3	1.127	1.482	0.865	0.892
60	3	1.312	0.618	1.415	1.457
61	3	1.286	1.652	1.966	0.472

## Appendix P: Derivation and Estimation of Classification Consistency and Accuracy

---

### Classification Consistency

Assume that  $\theta$  is a single latent trait measured by a test and denote  $\Phi$  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

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

where

$g(\theta)$  is the density of  $\theta$ .

In this report, the marginal distribution  $P(X = x)$  is denoted as  $f(x)$ , and the conditional error distribution  $P(X = x | \Phi = \theta)$  is denoted as  $f(x | \theta)$ . It is assumed that examinees are classified into one of  $H$  mutually exclusive categories on the basis of predetermined  $H - 1$  observed score cutoffs,  $C_1, C_2, \dots, C_{H-1}$ . Let  $L_h$  represent the  $h$ th category into which examinees with  $C_{h-1} \leq X < C_h$  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:

$$P(X \in L_h | \theta) = \sum_{x=C_{h-1}}^{C_h-1} f(x | \theta), \quad h = 1, 2, \dots, H$$

$$P(X \in L_h) = \int \sum_{x=C_{h-1}}^{C_h-1} f(x | \theta)g(\theta)d\theta, \quad h = 1, 2, \dots, H$$

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  $\theta$ ,  $X_1$  and  $X_2$  are independent and identically distributed. Consequently, the conditional bivariate distribution of  $X_1$  and  $X_2$  is

$$f(x_1, x_2 | \theta) = f(x_1 | \theta)f(x_2 | \theta)$$

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

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

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

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

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

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

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.

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

Then, Kappa (Cohen, 1960) is

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

### Classification Accuracy

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

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

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,  $\theta$ .

### **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 3 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 raw-to-scale score conversion tables, while Tables Q13–Q24 show the scale score distributions, by frequency (n-count), percent, cumulative frequency, and cumulative percent. The data in the tables include all students with valid scores.

**Table Q1. ELA Grade 3 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	177	54	24	308	9
1	185	45	25	311	9
2	193	38	26	314	9
3	201	32	27	317	9
4	209	27	28	320	9
5	217	22	29	323	9
6	225	19	30	326	9
7	233	17	31	330	8
8	241	15	32	333	9
9	248	13	33	336	9
10	254	12	34	339	9
11	260	12	35	343	9
12	264	11	36	346	9
13	269	11	37	350	9
14	273	10	38	354	10
15	277	10	39	358	10
16	281	10	40	363	10
17	284	10	41	368	11
18	288	10	42	374	12
19	291	9	43	381	13
20	295	9	44	390	15
21	298	9	45	398	17
22	301	9	46	406	19
23	305	9	47	414	22

**Table Q2. ELA Grade 4 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	172	48	24	303	9
1	180	41	25	306	9
2	188	35	26	309	9
3	196	30	27	312	9
4	204	26	28	315	9
5	212	22	29	320	9

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

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
6	220	19	30	321	9
7	228	16	31	324	9
8	237	14	32	328	9
9	243	13	33	331	9
10	249	12	34	334	9
11	254	11	35	338	9
12	259	11	36	343	10
13	263	10	37	345	10
14	268	10	38	349	10
15	271	10	39	353	10
16	275	10	40	358	11
17	279	10	41	364	12
18	283	9	42	370	13
19	287	9	43	377	14
20	289	9	44	386	16
21	293	9	45	394	19
22	296	9	46	402	22
23	299	9	47	410	25

**Table Q3. ELA Grade 5 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	112	66	29	280	10
1	120	58	30	283	10
2	128	51	31	286	9
3	136	44	32	289	9
4	144	39	33	292	9
5	152	34	34	295	9
6	160	30	35	298	9
7	168	26	36	301	9
8	176	23	37	304	9
9	184	21	38	308	10
10	192	19	39	311	10
11	200	17	40	314	10
12	208	16	41	320	10
13	216	14	42	321	10
14	224	13	43	325	10
15	229	13	44	328	11
16	234	12	45	332	11
17	239	12	46	337	11
18	243	12	47	341	12

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

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
19	247	11	48	346	12
20	251	11	49	351	13
21	254	11	50	357	13
22	258	11	51	363	14
23	261	10	52	371	15
24	265	10	53	380	17
25	268	10	54	391	20
26	271	10	55	399	22
27	274	10	56	407	24
28	277	10	57	415	27

Table Q4. ELA Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	128	76	29	288	9
1	136	66	30	291	9
2	144	57	31	294	9
3	152	49	32	297	9
4	161	41	33	300	9
5	169	35	34	303	9
6	177	30	35	305	9
7	185	26	36	308	9
8	193	22	37	311	9
9	201	19	38	314	9
10	209	17	39	320	9
11	217	15	40	321	9
12	225	13	41	324	9
13	231	12	42	327	10
14	236	12	43	331	10
15	241	11	44	335	10
16	245	11	45	338	10
17	249	10	46	342	11
18	253	10	47	347	11
19	257	10	48	352	12
20	260	10	49	357	12
21	263	10	50	362	13
22	267	10	51	369	14
23	270	9	52	377	16
24	273	9	53	387	18
25	276	9	54	395	20
26	279	9	55	403	23

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

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
27	283	9	56	411	26
28	285	9	57	419	29

**Table Q5. ELA Grade 7 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	147	74	29	293	8
1	154	65	30	295	8
2	162	55	31	298	8
3	170	47	32	300	8
4	178	40	33	303	8
5	186	33	34	305	8
6	194	28	35	308	8
7	202	24	36	311	8
8	210	20	37	313	8
9	218	17	38	316	8
10	226	15	39	318	8
11	233	13	40	321	8
12	239	12	41	324	8
13	244	11	42	327	8
14	248	11	43	330	9
15	252	10	44	333	9
16	256	10	45	337	9
17	260	9	46	340	9
18	263	9	47	347	10
19	266	9	48	348	10
20	269	9	49	352	11
21	272	9	50	357	11
22	275	8	51	363	12
23	278	8	52	370	14
24	280	8	53	378	16
25	283	8	54	389	19
26	287	8	55	397	22
27	288	8	56	405	25
28	291	8	57	413	28

**Table Q6. ELA Grade 8 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	130	69	29	278	8
1	138	59	30	280	8
2	146	51	31	284	8

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

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
3	154	44	32	285	8
4	161	38	33	288	8
5	169	32	34	290	8
6	177	27	35	292	8
7	185	23	36	295	8
8	193	19	37	297	8
9	201	16	38	300	8
10	209	14	39	302	8
11	217	12	40	305	8
12	225	11	41	307	8
13	229	10	42	310	8
14	234	10	43	313	8
15	237	10	44	316	8
16	241	9	45	319	8
17	245	9	46	322	8
18	248	9	47	325	9
19	251	9	48	329	9
20	254	8	49	333	10
21	257	8	50	337	10
22	260	8	51	343	11
23	262	8	52	348	12
24	265	8	53	355	14
25	268	8	54	365	16
26	270	8	55	379	21
27	273	8	56	387	25
28	275	8	57	395	30

**Table Q7. Mathematics Grade 3 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	137	58	29	296	8
1	145	52	30	298	8
2	153	47	31	300	8
3	161	43	32	303	8
4	170	39	33	305	8
5	178	35	34	307	8
6	186	32	35	309	8
7	194	29	36	312	8
8	202	26	37	314	8
9	210	24	38	316	8
10	218	21	39	319	8

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

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
11	226	19	40	321	8
12	234	17	41	323	8
13	241	15	42	326	8
14	247	14	43	329	8
15	252	13	44	331	8
16	257	12	45	334	9
17	261	12	46	340	9
18	265	11	47	341	9
19	268	11	48	344	10
20	271	10	49	349	10
21	275	10	50	353	11
22	278	9	51	358	12
23	280	9	52	365	13
24	285	9	53	373	15
25	286	9	54	384	19
26	288	8	55	392	22
27	291	8	56	401	27
28	293	8			

**Table Q8. Mathematics Grade 4 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	143	68	32	297	7
1	151	62	33	299	7
2	159	57	34	300	7
3	167	51	35	302	7
4	176	46	36	304	7
5	184	41	37	306	7
6	192	37	38	308	7
7	200	33	39	309	7
8	208	29	40	311	7
9	216	26	41	314	7
10	225	22	42	315	7
11	234	19	43	317	7
12	241	16	44	319	7
13	247	15	45	321	7
14	252	13	46	323	7
15	256	12	47	325	7
16	260	11	48	328	7
17	263	10	49	330	8
18	266	10	50	333	8

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

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
19	269	9	51	336	8
20	272	9	52	341	9
21	275	8	53	342	9
22	277	8	54	345	10
23	279	8	55	349	10
24	281	8	56	354	11
25	283	8	57	360	12
26	286	7	58	367	14
27	288	7	59	375	16
28	289	7	60	388	21
29	291	7	61	396	24
30	293	7	62	405	28
31	295	7			

**Table Q9. Mathematics Grade 5 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	153	78	31	308	7
1	161	68	32	310	7
2	169	60	33	312	7
3	177	52	34	315	7
4	185	45	35	317	7
5	193	39	36	319	7
6	201	34	37	321	7
7	210	28	38	323	7
8	218	24	39	325	7
9	226	21	40	327	7
10	236	17	41	329	7
11	244	15	42	331	7
12	250	14	43	334	7
13	256	13	44	336	7
14	260	12	45	338	7
15	265	11	46	340	7
16	268	11	47	343	7
17	272	10	48	346	8
18	275	10	49	348	8
19	279	9	50	351	8
20	282	9	51	354	8
21	284	9	52	357	9
22	287	9	53	361	10
23	290	8	54	365	10

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

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
24	294	8	55	370	11
25	295	8	56	375	13
26	297	8	57	382	14
27	299	8	58	392	18
28	302	7	59	400	21
29	304	7	60	408	24
30	306	7	61	416	28

**Table Q10. Mathematics Grade 6 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	132	165	34	316	7
1	140	142	35	318	7
2	148	123	36	320	7
3	157	104	37	322	7
4	165	89	38	324	7
5	173	77	39	325	7
6	181	66	40	327	7
7	189	56	41	329	7
8	197	48	42	331	7
9	205	41	43	333	7
10	213	35	44	335	7
11	221	30	45	337	7
12	230	25	46	340	7
13	242	21	47	341	7
14	252	17	48	343	7
15	259	16	49	345	7
16	265	14	50	347	7
17	270	13	51	349	7
18	275	12	52	351	7
19	279	11	53	354	7
20	284	10	54	356	7
21	286	10	55	359	8
22	289	10	56	362	8
23	292	9	57	365	8
24	295	9	58	368	9
25	297	9	59	371	9
26	300	8	60	375	9
27	302	8	61	379	10
28	304	8	62	384	11
29	306	8	63	390	13
30	308	7	64	398	15

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

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
31	310	7	65	406	18
32	312	7	66	414	21
33	314	7	67	423	25

**Table Q11. Mathematics Grade 7 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	150	112	35	318	6
1	158	98	36	319	6
2	166	86	37	321	6
3	174	75	38	322	6
4	181	67	39	324	6
5	189	59	40	325	5
6	197	52	41	327	5
7	205	46	42	328	5
8	213	40	43	330	5
9	220	36	44	331	5
10	228	32	45	333	5
11	236	28	46	334	5
12	244	24	47	336	6
13	256	20	48	337	6
14	265	16	49	339	6
15	271	14	50	340	6
16	276	13	51	342	6
17	280	11	52	344	6
18	284	10	53	346	6
19	287	10	54	348	6
20	290	9	55	350	6
21	293	8	56	352	6
22	295	8	57	354	7
23	297	8	58	356	7
24	299	7	59	359	7
25	301	7	60	362	8
26	303	7	61	365	8
27	305	7	62	369	9
28	307	7	63	373	10
29	309	6	64	379	11
30	310	6	65	386	13
31	312	6	66	394	16
32	313	6	67	402	19
33	315	6	68	409	23
34	316	6			

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

**Table Q12. Mathematics Grade 8 RSSS Table**

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	132	139	35	312	7
1	140	126	36	313	7
2	148	114	37	315	7
3	156	103	38	317	7
4	164	93	39	318	7
5	172	84	40	320	6
6	180	75	41	322	6
7	188	67	42	323	6
8	196	59	43	325	6
9	204	51	44	326	6
10	212	44	45	328	6
11	220	38	46	330	6
12	228	32	47	331	6
13	236	26	48	333	6
14	246	21	49	334	6
15	254	18	50	336	6
16	260	15	51	338	6
17	266	14	52	340	7
18	270	13	53	341	7
19	274	12	54	343	7
20	278	11	55	345	7
21	281	10	56	349	7
22	284	10	57	350	7
23	287	9	58	352	8
24	289	9	59	355	8
25	292	9	60	357	8
26	294	8	61	361	9
27	296	8	62	364	9
28	299	8	63	369	10
29	301	8	64	374	12
30	303	8	65	381	14
31	305	7	66	391	17
32	306	7	67	399	21
33	308	7	68	407	25
34	310	7			

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

**Table Q13. ELA Grade 3 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
177	31	0.02%	31	0.02%
185	56	0.03%	87	0.05%
193	152	0.08%	239	0.13%
201	318	0.18%	557	0.31%
209	727	0.40%	1,284	0.71%
217	1,154	0.64%	2,438	1.35%
225	1,702	0.94%	4,140	2.30%
233	2,152	1.19%	6,292	3.49%
241	2,524	1.40%	8,816	4.89%
248	2,830	1.57%	11,646	6.46%
254	2,955	1.64%	14,601	8.10%
260	3,117	1.73%	17,718	9.83%
264	3,476	1.93%	21,194	11.8%
269	3,694	2.05%	24,888	13.8%
273	3,988	2.21%	28,876	16.0%
277	4,360	2.42%	33,236	18.4%
281	4,616	2.56%	37,852	21.0%
284	4,951	2.75%	42,803	23.7%
288	5,401	3.00%	48,204	26.7%
291	5,505	3.05%	53,709	29.8%
295	5,889	3.27%	59,598	33.1%
298	5,892	3.27%	65,490	36.3%
301	6,245	3.46%	71,735	39.8%
305	6,492	3.60%	78,227	43.4%
308	6,510	3.61%	84,737	47.0%
311	6,770	3.75%	91,507	50.8%
314	6,597	3.66%	98,104	54.4%
317	6,589	3.65%	104,693	58.1%
320	6,684	3.71%	111,377	61.8%
323	6,602	3.66%	117,979	65.4%
326	6,589	3.65%	124,568	69.1%
330	6,193	3.43%	130,761	72.5%
333	6,209	3.44%	136,970	76.0%
336	6,156	3.41%	143,126	79.4%
339	5,822	3.23%	148,948	82.6%
343	5,195	2.88%	154,143	85.5%
346	4,827	2.68%	158,970	88.2%
350	4,440	2.46%	163,410	90.6%
354	3,886	2.16%	167,296	92.8%
358	3,360	1.86%	170,656	94.6%
363	2,920	1.62%	173,576	96.3%
368	2,316	1.28%	175,892	97.6%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
374	1,807	1.00%	177,699	98.6%
381	1,250	0.69%	178,949	99.2%
390	766	0.42%	179,715	99.7%
398	383	0.21%	180,098	99.9%
406	165	0.09%	180,263	100%
414	40	0.02%	180,303	100%

**Table Q14. ELA Grade 4 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
172	15	0.01%	15	0.01%
180	31	0.02%	46	0.03%
188	108	0.06%	154	0.09%
196	230	0.13%	384	0.22%
204	463	0.26%	847	0.48%
212	756	0.43%	1,603	0.91%
220	1,127	0.64%	2,730	1.54%
228	1,488	0.84%	4,218	2.38%
237	1,757	0.99%	5,975	3.37%
243	2,275	1.28%	8,250	4.66%
249	2,504	1.41%	10,754	6.07%
254	2,849	1.61%	13,603	7.68%
259	3,269	1.85%	16,872	9.53%
263	3,567	2.01%	20,439	11.5%
268	3,989	2.25%	24,428	13.8%
271	4,293	2.42%	28,721	16.2%
275	4,506	2.54%	33,227	18.8%
279	4,796	2.71%	38,023	21.5%
283	5,048	2.85%	43,071	24.3%
287	5,193	2.93%	48,264	27.3%
289	5,477	3.09%	53,741	30.3%
293	5,784	3.27%	59,525	33.6%
296	5,943	3.36%	65,468	37.0%
299	6,156	3.48%	71,624	40.4%
303	6,390	3.61%	78,014	44.1%
306	6,450	3.64%	84,464	47.7%
309	6,567	3.71%	91,031	51.4%
312	6,835	3.86%	97,866	55.3%
315	6,941	3.92%	104,807	59.2%
320	6,809	3.84%	111,616	63.0%
321	6,911	3.90%	118,527	66.9%
324	6,879	3.88%	125,406	70.8%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
328	6,723	3.80%	132,129	74.6%
331	6,635	3.75%	138,764	78.4%
334	6,046	3.41%	144,810	81.8%
338	5,652	3.19%	150,462	85.0%
343	5,305	3.00%	155,767	88.0%
345	4,965	2.80%	160,732	90.8%
349	4,171	2.36%	164,903	93.1%
353	3,533	2.00%	168,436	95.1%
358	2,800	1.58%	171,236	96.7%
364	2,210	1.25%	173,446	97.9%
370	1,594	0.90%	175,040	98.8%
377	1,034	0.58%	176,074	99.4%
386	620	0.35%	176,694	99.8%
394	275	0.16%	176,969	99.9%
402	104	0.06%	177,073	100%
410	19	0.01%	177,092	100%

**Table Q15. ELA Grade 5 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
112	9	0.01%	9	0.01%
120	14	0.01%	23	0.01%
128	11	0.01%	34	0.02%
136	32	0.02%	66	0.04%
144	53	0.03%	119	0.07%
152	141	0.08%	260	0.16%
160	208	0.12%	468	0.28%
168	389	0.23%	857	0.51%
176	515	0.31%	1,372	0.82%
184	737	0.44%	2,109	1.26%
192	961	0.57%	3,070	1.83%
200	1,137	0.68%	4,207	2.51%
208	1,253	0.75%	5,460	3.26%
216	1,407	0.84%	6,867	4.10%
224	1,554	0.93%	8,421	5.03%
229	1,668	1.00%	10,089	6.03%
234	1,782	1.06%	11,871	7.09%
239	1,910	1.14%	13,781	8.23%
243	2,057	1.23%	15,838	9.46%
247	2,231	1.33%	18,069	10.8%
251	2,428	1.45%	20,497	12.2%
254	2,555	1.53%	23,052	13.8%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
258	2,827	1.69%	25,879	15.5%
261	2,844	1.70%	28,723	17.2%
265	3,147	1.88%	31,870	19.0%
268	3,280	1.96%	35,150	21.0%
271	3,680	2.20%	38,830	23.2%
274	3,848	2.30%	42,678	25.5%
277	4,043	2.42%	46,721	27.9%
280	4,409	2.63%	51,130	30.5%
283	4,647	2.78%	55,777	33.3%
286	4,846	2.89%	60,623	36.2%
289	4,973	2.97%	65,596	39.2%
292	5,129	3.06%	70,725	42.2%
295	5,371	3.21%	76,096	45.5%
298	5,626	3.36%	81,722	48.8%
301	5,738	3.43%	87,460	52.2%
304	5,846	3.49%	93,306	55.7%
308	5,960	3.56%	99,266	59.3%
311	6,094	3.64%	105,360	62.9%
314	6,161	3.68%	111,521	66.6%
320	6,161	3.68%	117,682	70.3%
321	6,116	3.65%	123,798	73.9%
325	6,002	3.59%	129,800	77.5%
328	5,751	3.44%	135,551	81.0%
332	5,367	3.21%	140,918	84.2%
337	5,103	3.05%	146,021	87.2%
341	4,576	2.73%	150,597	90.0%
346	4,118	2.46%	154,715	92.4%
351	3,528	2.11%	158,243	94.5%
357	2,950	1.76%	161,193	96.3%
363	2,308	1.38%	163,501	97.7%
371	1,650	0.99%	165,151	98.7%
380	1,129	0.67%	166,280	99.3%
391	687	0.41%	166,967	99.7%
399	321	0.19%	167,288	99.9%
407	99	0.06%	167,387	100%
415	22	0.01%	167,409	100%

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

**Table Q16. ELA Grade 6 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
128	5	0.00%	5	0.00%
136	19	0.01%	24	0.01%
144	23	0.01%	47	0.03%
152	30	0.02%	77	0.05%
161	56	0.03%	133	0.08%
169	144	0.09%	277	0.17%
177	262	0.16%	539	0.32%
185	377	0.23%	916	0.55%
193	624	0.38%	1,540	0.93%
201	801	0.48%	2,341	1.41%
209	1,005	0.61%	3,346	2.02%
217	1,257	0.76%	4,603	2.77%
225	1,369	0.82%	5,972	3.60%
231	1,620	0.98%	7,592	4.57%
236	1,823	1.10%	9,415	5.67%
241	1,981	1.19%	11,396	6.86%
245	2,198	1.32%	13,594	8.19%
249	2,253	1.36%	15,847	9.54%
253	2,441	1.47%	18,288	11.0%
257	2,653	1.60%	20,941	12.6%
260	2,752	1.66%	23,693	14.3%
263	3,170	1.91%	26,863	16.2%
267	3,288	1.98%	30,151	18.2%
270	3,408	2.05%	33,559	20.2%
273	3,657	2.20%	37,216	22.4%
276	3,764	2.27%	40,980	24.7%
279	4,086	2.46%	45,066	27.1%
283	4,239	2.55%	49,305	29.7%
285	4,502	2.71%	53,807	32.4%
288	4,653	2.80%	58,460	35.2%
291	5,018	3.02%	63,478	38.2%
294	5,130	3.09%	68,608	41.3%
297	5,299	3.19%	73,907	44.5%
300	5,537	3.33%	79,444	47.8%
303	5,669	3.41%	85,113	51.3%
305	5,811	3.50%	90,924	54.8%
308	5,873	3.54%	96,797	58.3%
311	5,975	3.60%	102,772	61.9%
314	6,057	3.65%	108,829	65.5%
320	5,999	3.61%	114,828	69.2%
321	6,032	3.63%	120,860	72.8%
324	5,760	3.47%	126,620	76.3%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
327	5,668	3.41%	132,288	79.7%
331	5,372	3.24%	137,660	82.9%
335	5,076	3.06%	142,736	86.0%
338	4,727	2.85%	147,463	88.8%
342	4,185	2.52%	151,648	91.3%
347	3,757	2.26%	155,405	93.6%
352	3,073	1.85%	158,478	95.4%
357	2,524	1.52%	161,002	97.0%
362	2,012	1.21%	163,014	98.2%
369	1,320	0.79%	164,334	99.0%
377	824	0.50%	165,158	99.5%
387	511	0.31%	165,669	99.8%
395	250	0.15%	165,919	99.9%
403	90	0.05%	166,009	100%
411	29	0.02%	166,038	100%
419	2	0.00%	166,040	100%

**Table Q17. ELA Grade 7 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
147	11	0.01%	11	0.01%
154	13	0.01%	24	0.02%
162	33	0.02%	57	0.04%
170	41	0.03%	98	0.06%
178	98	0.06%	196	0.13%
186	200	0.13%	396	0.25%
194	377	0.24%	773	0.49%
202	582	0.37%	1,355	0.87%
210	821	0.53%	2,176	1.39%
218	1,094	0.70%	3,270	2.09%
226	1,365	0.87%	4,635	2.97%
233	1,524	0.98%	6,159	3.94%
239	1,744	1.12%	7,903	5.06%
244	1,958	1.25%	9,861	6.31%
248	2,127	1.36%	11,988	7.67%
252	2,220	1.42%	14,208	9.09%
256	2,412	1.54%	16,620	10.6%
260	2,462	1.58%	19,082	12.2%
263	2,702	1.73%	21,784	13.9%
266	2,796	1.79%	24,580	15.7%
269	2,790	1.79%	27,370	17.5%
272	2,986	1.91%	30,356	19.4%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
275	3,172	2.03%	33,528	21.5%
278	3,400	2.18%	36,928	23.6%
280	3,475	2.22%	40,403	25.9%
283	3,580	2.29%	43,983	28.1%
287	3,646	2.33%	47,629	30.5%
288	3,906	2.50%	51,535	33.0%
291	3,809	2.44%	55,344	35.4%
293	4,138	2.65%	59,482	38.1%
295	4,111	2.63%	63,593	40.7%
298	4,263	2.73%	67,856	43.4%
300	4,390	2.81%	72,246	46.2%
303	4,631	2.96%	76,877	49.2%
305	4,629	2.96%	81,506	52.2%
308	4,716	3.02%	86,222	55.2%
311	4,753	3.04%	90,975	58.2%
313	4,878	3.12%	95,853	61.3%
316	4,851	3.10%	100,704	64.5%
318	5,029	3.22%	105,733	67.7%
321	4,954	3.17%	110,687	70.8%
324	5,057	3.24%	115,744	74.1%
327	4,862	3.11%	120,606	77.2%
330	4,755	3.04%	125,361	80.2%
333	4,657	2.98%	130,018	83.2%
337	4,464	2.86%	134,482	86.1%
340	4,351	2.78%	138,833	88.9%
347	3,915	2.51%	142,748	91.4%
348	3,496	2.24%	146,244	93.6%
352	3,004	1.92%	149,248	95.5%
357	2,401	1.54%	151,649	97.1%
363	1,813	1.16%	153,462	98.2%
370	1,323	0.85%	154,785	99.1%
378	765	0.49%	155,550	99.6%
389	448	0.29%	155,998	99.8%
397	185	0.12%	156,183	100%
405	55	0.04%	156,238	100%
413	10	0.01%	156,248	100%

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

**Table Q18. ELA Grade 8 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
130	16	0.01%	16	0.01%
138	14	0.01%	30	0.02%
146	24	0.02%	54	0.04%
154	24	0.02%	78	0.05%
161	41	0.03%	119	0.08%
169	85	0.06%	204	0.14%
177	151	0.10%	355	0.24%
185	241	0.16%	596	0.40%
193	328	0.22%	924	0.61%
201	454	0.30%	1,378	0.91%
209	532	0.35%	1,910	1.27%
217	701	0.46%	2,611	1.73%
225	752	0.50%	3,363	2.23%
229	934	0.62%	4,297	2.85%
234	967	0.64%	5,264	3.49%
237	1,129	0.75%	6,393	4.24%
241	1,272	0.84%	7,665	5.08%
245	1,319	0.87%	8,984	5.96%
248	1,463	0.97%	10,447	6.93%
251	1,517	1.01%	11,964	7.93%
254	1,624	1.08%	13,588	9.01%
257	1,675	1.11%	15,263	10.1%
260	1,804	1.20%	17,067	11.3%
262	1,856	1.23%	18,923	12.5%
265	1,970	1.31%	20,893	13.9%
268	2,055	1.36%	22,948	15.2%
270	2,221	1.47%	25,169	16.7%
273	2,320	1.54%	27,489	18.2%
275	2,444	1.62%	29,933	19.8%
278	2,622	1.74%	32,555	21.6%
280	2,738	1.82%	35,293	23.4%
284	2,880	1.91%	38,173	25.3%
285	3,219	2.13%	41,392	27.4%
288	3,317	2.20%	44,709	29.6%
290	3,576	2.37%	48,285	32.0%
292	3,680	2.44%	51,965	34.4%
295	3,906	2.59%	55,871	37.0%
297	4,101	2.72%	59,972	39.8%
300	4,326	2.87%	64,298	42.6%
302	4,576	3.03%	68,874	45.7%
305	4,743	3.14%	73,617	48.8%
307	4,981	3.30%	78,598	52.1%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
310	5,077	3.37%	83,675	55.5%
313	5,340	3.54%	89,015	59.0%
316	5,593	3.71%	94,608	62.7%
319	5,736	3.80%	100,344	66.5%
322	5,937	3.94%	106,281	70.5%
325	6,050	4.01%	112,331	74.5%
329	6,050	4.01%	118,381	78.5%
333	6,135	4.07%	124,516	82.5%
337	5,973	3.96%	130,489	86.5%
343	5,596	3.71%	136,085	90.2%
348	4,842	3.21%	140,927	93.4%
355	4,158	2.76%	145,085	96.2%
365	2,940	1.95%	148,025	98.1%
379	1,849	1.23%	149,874	99.4%
387	767	0.51%	150,641	99.9%
395	208	0.14%	150,849	100%

**Table Q19. Mathematics Grade 3 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
137	6	0.00%	6	0.00%
145	11	0.01%	17	0.01%
153	21	0.01%	38	0.02%
161	29	0.02%	67	0.04%
170	82	0.05%	149	0.08%
178	171	0.09%	320	0.18%
186	322	0.18%	642	0.36%
194	564	0.31%	1,206	0.67%
202	856	0.47%	2,062	1.14%
210	1,250	0.69%	3,312	1.83%
218	1,576	0.87%	4,888	2.70%
226	1,944	1.08%	6,832	3.78%
234	2,251	1.24%	9,083	5.02%
241	2,455	1.36%	11,538	6.38%
247	2,690	1.49%	14,228	7.87%
252	2,995	1.66%	17,223	9.52%
257	3,120	1.73%	20,343	11.3%
261	3,321	1.84%	23,664	13.1%
265	3,361	1.86%	27,025	14.9%
268	3,469	1.92%	30,494	16.9%
271	3,715	2.05%	34,209	18.9%
275	3,854	2.13%	38,063	21.0%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
278	3,913	2.16%	41,976	23.2%
280	3,976	2.20%	45,952	25.4%
285	4,125	2.28%	50,077	27.7%
286	4,159	2.30%	54,236	30.0%
288	4,232	2.34%	58,468	32.3%
291	4,224	2.34%	62,692	34.7%
293	4,283	2.37%	66,975	37.0%
296	4,451	2.46%	71,426	39.5%
298	4,276	2.36%	75,702	41.9%
300	4,334	2.40%	80,036	44.3%
303	4,271	2.36%	84,307	46.6%
305	4,394	2.43%	88,701	49.1%
307	4,374	2.42%	93,075	51.5%
309	4,367	2.42%	97,442	53.9%
312	4,345	2.40%	101,787	56.3%
314	4,353	2.41%	106,140	58.7%
316	4,270	2.36%	110,410	61.1%
319	4,450	2.46%	114,860	63.5%
321	4,399	2.43%	119,259	66.0%
323	4,475	2.47%	123,734	68.4%
326	4,505	2.49%	128,239	70.9%
329	4,451	2.46%	132,690	73.4%
331	4,450	2.46%	137,140	75.8%
334	4,462	2.47%	141,602	78.3%
340	4,598	2.54%	146,200	80.9%
341	4,486	2.48%	150,686	83.3%
344	4,370	2.42%	155,056	85.7%
349	4,167	2.30%	159,223	88.1%
353	4,074	2.25%	163,297	90.3%
358	4,000	2.21%	167,297	92.5%
365	3,766	2.08%	171,063	94.6%
373	3,424	1.89%	174,487	96.5%
384	2,855	1.58%	177,342	98.1%
392	2,276	1.26%	179,618	99.3%
401	1,206	0.67%	180,824	100%

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

**Table Q20. Mathematics Grade 4 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
143	3	0.00%	3	0.00%
151	10	0.01%	13	0.01%
159	11	0.01%	24	0.01%
167	39	0.02%	63	0.04%
176	160	0.09%	223	0.13%
184	340	0.19%	563	0.32%
192	580	0.33%	1,143	0.65%
200	1,011	0.57%	2,154	1.22%
208	1,453	0.82%	3,607	2.04%
216	2,020	1.14%	5,627	3.18%
225	2,455	1.39%	8,082	4.56%
234	2,752	1.55%	10,834	6.12%
241	2,927	1.65%	13,761	7.77%
247	3,011	1.70%	16,772	9.47%
252	3,018	1.70%	19,790	11.2%
256	2,995	1.69%	22,785	12.9%
260	2,945	1.66%	25,730	14.5%
263	2,978	1.68%	28,708	16.2%
266	2,922	1.65%	31,630	17.9%
269	2,954	1.67%	34,584	19.5%
272	2,918	1.65%	37,502	21.2%
275	2,877	1.62%	40,379	22.8%
277	2,841	1.60%	43,220	24.4%
279	2,871	1.62%	46,091	26.0%
281	2,861	1.62%	48,952	27.6%
283	2,922	1.65%	51,874	29.3%
286	2,883	1.63%	54,757	30.9%
288	2,939	1.66%	57,696	32.6%
289	2,848	1.61%	60,544	34.2%
291	3,002	1.69%	63,546	35.9%
293	3,018	1.70%	66,564	37.6%
295	2,983	1.68%	69,547	39.3%
297	3,086	1.74%	72,633	41.0%
299	3,153	1.78%	75,786	42.8%
300	3,130	1.77%	78,916	44.5%
302	3,106	1.75%	82,022	46.3%
304	3,267	1.84%	85,289	48.1%
306	3,246	1.83%	88,535	50.0%
308	3,265	1.84%	91,800	51.8%
309	3,371	1.90%	95,171	53.7%
311	3,594	2.03%	98,765	55.8%
314	3,384	1.91%	102,149	57.7%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
315	3,580	2.02%	105,729	59.7%
317	3,600	2.03%	109,329	61.7%
319	3,625	2.05%	112,954	63.8%
321	3,638	2.05%	116,592	65.8%
323	3,701	2.09%	120,293	67.9%
325	3,869	2.18%	124,162	70.1%
328	3,977	2.25%	128,139	72.3%
330	4,043	2.28%	132,182	74.6%
333	4,096	2.31%	136,278	76.9%
336	4,018	2.27%	140,296	79.2%
341	4,105	2.32%	144,401	81.5%
342	4,134	2.33%	148,535	83.8%
345	4,181	2.36%	152,716	86.2%
349	4,211	2.38%	156,927	88.6%
354	4,037	2.28%	160,964	90.9%
360	4,006	2.26%	164,970	93.1%
367	3,682	2.08%	168,652	95.2%
375	3,315	1.87%	171,967	97.1%
388	2,718	1.53%	174,685	98.6%
396	1,777	1.00%	176,462	99.6%
405	685	0.39%	177,147	100%

**Table Q21. Mathematics Grade 5 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
153	6	0.00%	6	0.00%
161	19	0.01%	25	0.01%
169	28	0.02%	53	0.03%
177	77	0.05%	130	0.08%
185	199	0.12%	329	0.20%
193	479	0.29%	808	0.48%
201	803	0.48%	1,611	0.97%
210	1,301	0.78%	2,912	1.75%
218	1,783	1.07%	4,695	2.81%
226	2,177	1.30%	6,872	4.12%
236	2,508	1.50%	9,380	5.62%
244	2,739	1.64%	12,119	7.26%
250	2,995	1.80%	15,114	9.06%
256	3,053	1.83%	18,167	10.9%
260	3,155	1.89%	21,322	12.8%
265	3,234	1.94%	24,556	14.7%
268	3,360	2.01%	27,916	16.7%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
272	3,471	2.08%	31,387	18.8%
275	3,435	2.06%	34,822	20.9%
279	3,726	2.23%	38,548	23.1%
282	3,784	2.27%	42,332	25.4%
284	3,777	2.26%	46,109	27.6%
287	3,830	2.30%	49,939	29.9%
290	3,936	2.36%	53,875	32.3%
294	3,928	2.35%	57,803	34.6%
295	3,975	2.38%	61,778	37.0%
297	4,097	2.46%	65,875	39.5%
299	4,017	2.41%	69,892	41.9%
302	4,004	2.40%	73,896	44.3%
304	3,997	2.40%	77,893	46.7%
306	3,966	2.38%	81,859	49.1%
308	3,850	2.31%	85,709	51.4%
310	3,853	2.31%	89,562	53.7%
312	3,743	2.24%	93,305	55.9%
315	3,674	2.20%	96,979	58.1%
317	3,667	2.20%	100,646	60.3%
319	3,606	2.16%	104,252	62.5%
321	3,553	2.13%	107,805	64.6%
323	3,546	2.13%	111,351	66.7%
325	3,434	2.06%	114,785	68.8%
327	3,379	2.03%	118,164	70.8%
329	3,381	2.03%	121,545	72.9%
331	3,295	1.97%	124,840	74.8%
334	3,194	1.91%	128,034	76.7%
336	3,137	1.88%	131,171	78.6%
338	3,205	1.92%	134,376	80.5%
340	3,079	1.85%	137,455	82.4%
343	3,005	1.80%	140,460	84.2%
346	2,798	1.68%	143,258	85.9%
348	2,804	1.68%	146,062	87.5%
351	2,679	1.61%	148,741	89.2%
354	2,610	1.56%	151,351	90.7%
357	2,461	1.48%	153,812	92.2%
361	2,406	1.44%	156,218	93.6%
365	2,092	1.25%	158,310	94.9%
370	2,008	1.20%	160,318	96.1%
375	1,786	1.07%	162,104	97.2%
382	1,465	0.88%	163,569	98.0%
392	1,227	0.74%	164,796	98.8%
400	970	0.58%	165,766	99.4%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
408	696	0.42%	166,462	99.8%
416	376	0.23%	166,838	100%

**Table Q22. Mathematics Grade 6 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
132	8	0.00%	8	0.00%
140	11	0.01%	19	0.01%
148	7	0.00%	26	0.02%
157	20	0.01%	46	0.03%
165	48	0.03%	94	0.06%
173	117	0.07%	211	0.13%
181	217	0.13%	428	0.26%
189	382	0.23%	810	0.49%
197	815	0.50%	1,625	0.99%
205	1,300	0.79%	2,925	1.78%
213	1,997	1.22%	4,922	3.00%
221	2,725	1.66%	7,647	4.66%
230	3,440	2.10%	11,087	6.76%
242	3,929	2.40%	15,016	9.16%
252	4,256	2.60%	19,272	11.8%
259	4,611	2.81%	23,883	14.6%
265	4,702	2.87%	28,585	17.4%
270	4,590	2.80%	33,175	20.2%
275	4,668	2.85%	37,843	23.1%
279	4,581	2.79%	42,424	25.9%
284	4,370	2.67%	46,794	28.5%
286	4,334	2.64%	51,128	31.2%
289	4,345	2.65%	55,473	33.8%
292	4,311	2.63%	59,784	36.5%
295	4,000	2.44%	63,784	38.9%
297	3,983	2.43%	67,767	41.3%
300	3,813	2.33%	71,580	43.7%
302	3,802	2.32%	75,382	46.0%
304	3,544	2.16%	78,926	48.1%
306	3,533	2.16%	82,459	50.3%
308	3,410	2.08%	85,869	52.4%
310	3,337	2.04%	89,206	54.4%
312	3,326	2.03%	92,532	56.4%
314	3,221	1.96%	95,753	58.4%
316	3,103	1.89%	98,856	60.3%
318	3,069	1.87%	101,925	62.2%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
320	2,980	1.82%	104,905	64.0%
322	2,961	1.81%	107,866	65.8%
324	2,832	1.73%	110,698	67.5%
325	2,797	1.71%	113,495	69.2%
327	2,766	1.69%	116,261	70.9%
329	2,680	1.63%	118,941	72.6%
331	2,579	1.57%	121,520	74.1%
333	2,635	1.61%	124,155	75.7%
335	2,620	1.60%	126,775	77.3%
337	2,498	1.52%	129,273	78.9%
340	2,573	1.57%	131,846	80.4%
341	2,399	1.46%	134,245	81.9%
343	2,333	1.42%	136,578	83.3%
345	2,342	1.43%	138,920	84.7%
347	2,179	1.33%	141,099	86.1%
349	2,227	1.36%	143,326	87.4%
351	2,112	1.29%	145,438	88.7%
354	2,108	1.29%	147,546	90.0%
356	2,005	1.22%	149,551	91.2%
359	1,842	1.12%	151,393	92.4%
362	1,827	1.11%	153,220	93.5%
365	1,700	1.04%	154,920	94.5%
368	1,579	0.96%	156,499	95.5%
371	1,439	0.88%	157,938	96.3%
375	1,328	0.81%	159,266	97.2%
379	1,140	0.70%	160,406	97.9%
384	1,024	0.62%	161,430	98.5%
390	833	0.51%	162,263	99.0%
398	701	0.43%	162,964	99.4%
406	500	0.31%	163,464	99.7%
414	324	0.20%	163,788	99.9%
423	139	0.08%	163,927	100%

**Table Q23. Mathematics Grade 7 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
150	13	0.01%	13	0.01%
158	13	0.01%	26	0.02%
166	14	0.01%	40	0.03%
174	55	0.04%	95	0.06%
181	108	0.07%	203	0.13%
189	236	0.16%	439	0.29%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
197	528	0.35%	967	0.64%
205	869	0.57%	1,836	1.21%
213	1,463	0.96%	3,299	2.17%
220	2,156	1.42%	5,455	3.59%
228	2,904	1.91%	8,359	5.50%
236	3,661	2.41%	12,020	7.91%
244	4,248	2.80%	16,268	10.7%
256	4,638	3.05%	20,906	13.8%
265	4,849	3.19%	25,755	17.0%
271	4,633	3.05%	30,388	20.0%
276	4,624	3.04%	35,012	23.0%
280	4,402	2.90%	39,414	25.9%
284	4,140	2.73%	43,554	28.7%
287	3,949	2.60%	47,503	31.3%
290	3,783	2.49%	51,286	33.8%
293	3,563	2.35%	54,849	36.1%
295	3,446	2.27%	58,295	38.4%
297	3,198	2.11%	61,493	40.5%
299	3,142	2.07%	64,635	42.6%
301	2,896	1.91%	67,531	44.5%
303	2,871	1.89%	70,402	46.3%
305	2,830	1.86%	73,232	48.2%
307	2,654	1.75%	75,886	50.0%
309	2,701	1.78%	78,587	51.7%
310	2,538	1.67%	81,125	53.4%
312	2,567	1.69%	83,692	55.1%
313	2,563	1.69%	86,255	56.8%
315	2,485	1.64%	88,740	58.4%
316	2,333	1.54%	91,073	60.0%
318	2,382	1.57%	93,455	61.5%
319	2,291	1.51%	95,746	63.0%
321	2,205	1.45%	97,951	64.5%
322	2,252	1.48%	100,203	66.0%
324	2,159	1.42%	102,362	67.4%
325	2,140	1.41%	104,502	68.8%
327	2,205	1.45%	106,707	70.2%
328	2,141	1.41%	108,848	71.7%
330	2,186	1.44%	111,034	73.1%
331	2,108	1.39%	113,142	74.5%
333	2,111	1.39%	115,253	75.9%
334	2,049	1.35%	117,302	77.2%
336	2,035	1.34%	119,337	78.6%
337	2,098	1.38%	121,435	79.9%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
339	1,936	1.27%	123,371	81.2%
340	1,984	1.31%	125,355	82.5%
342	1,961	1.29%	127,316	83.8%
344	1,969	1.30%	129,285	85.1%
346	1,992	1.31%	131,277	86.4%
348	1,960	1.29%	133,237	87.7%
350	1,912	1.26%	135,149	89.0%
352	1,821	1.20%	136,970	90.2%
354	1,793	1.18%	138,763	91.4%
356	1,769	1.16%	140,532	92.5%
359	1,699	1.12%	142,231	93.6%
362	1,627	1.07%	143,858	94.7%
365	1,679	1.11%	145,537	95.8%
369	1,465	0.96%	147,002	96.8%
373	1,351	0.89%	148,353	97.7%
379	1,173	0.77%	149,526	98.4%
386	1,038	0.68%	150,564	99.1%
394	754	0.50%	151,318	99.6%
402	433	0.29%	151,751	99.9%
409	146	0.10%	151,897	100%

**Table Q24. Mathematics Grade 8 Scale Score Frequency Distribution**

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
132	12	0.01%	12	0.01%
140	10	0.01%	22	0.02%
148	20	0.02%	42	0.04%
156	27	0.02%	69	0.06%
164	71	0.06%	140	0.12%
172	137	0.12%	277	0.24%
180	281	0.24%	558	0.47%
188	519	0.44%	1,077	0.92%
196	943	0.80%	2,020	1.72%
204	1,410	1.20%	3,430	2.92%
212	2,038	1.73%	5,468	4.65%
220	2,592	2.20%	8,060	6.85%
228	3,112	2.65%	11,172	9.50%
236	3,395	2.89%	14,567	12.4%
246	3,668	3.12%	18,235	15.5%
254	3,639	3.09%	21,874	18.6%
260	3,684	3.13%	25,558	21.7%
266	3,591	3.05%	29,149	24.8%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
270	3,588	3.05%	32,737	27.8%
274	3,421	2.91%	36,158	30.7%
278	3,355	2.85%	39,513	33.6%
281	3,329	2.83%	42,842	36.4%
284	3,145	2.67%	45,987	39.1%
287	3,098	2.63%	49,085	41.7%
289	3,020	2.57%	52,105	44.3%
292	2,917	2.48%	55,022	46.8%
294	2,788	2.37%	57,810	49.1%
296	2,797	2.38%	60,607	51.5%
299	2,600	2.21%	63,207	53.7%
301	2,637	2.24%	65,844	56.0%
303	2,481	2.11%	68,325	58.1%
305	2,423	2.06%	70,748	60.1%
306	2,424	2.06%	73,172	62.2%
308	2,339	1.99%	75,511	64.2%
310	2,246	1.91%	77,757	66.1%
312	2,096	1.78%	79,853	67.9%
313	1,951	1.66%	81,804	69.5%
315	1,916	1.63%	83,720	71.2%
317	1,782	1.51%	85,502	72.7%
318	1,811	1.54%	87,313	74.2%
320	1,704	1.45%	89,017	75.7%
322	1,636	1.39%	90,653	77.1%
323	1,569	1.33%	92,222	78.4%
325	1,461	1.24%	93,683	79.6%
326	1,441	1.22%	95,124	80.9%
328	1,430	1.22%	96,554	82.1%
330	1,313	1.12%	97,867	83.2%
331	1,383	1.18%	99,250	84.4%
333	1,235	1.05%	100,485	85.4%
334	1,175	1.00%	101,660	86.4%
336	1,194	1.01%	102,854	87.4%
338	1,113	0.95%	103,967	88.4%
340	1,055	0.90%	105,022	89.3%
341	1,028	0.87%	106,050	90.1%
343	1,033	0.88%	107,083	91.0%
345	1,005	0.85%	108,088	91.9%
349	951	0.81%	109,039	92.7%
350	903	0.77%	109,942	93.5%
352	950	0.81%	110,892	94.3%
355	913	0.78%	111,805	95.0%
357	839	0.71%	112,644	95.8%

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

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
361	828	0.70%	113,472	96.5%
364	835	0.71%	114,307	97.2%
369	790	0.67%	115,097	97.8%
374	684	0.58%	115,781	98.4%
381	653	0.56%	116,434	99.0%
391	571	0.49%	117,005	99.5%
399	436	0.37%	117,441	99.8%
407	202	0.17%	117,643	100%