New York State Testing Program 2023: English Language Arts and Mathematics Grades 3–8

Technical Report
Prepared for the New York State Education Department by NWEA

2023

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. Performance Level Cut Scores and Classification 2
1.4.3. Subscores 2
1.5. Testing Accommodations 3
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 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 Review Criteria Documents 8
2.5.1. Principles of Universal Design 9
2.6. Passage Finding 9
2.7. Item Development 10
2.8. Educator Item Review 11
2.9. Field Testing 11
2.10. Rangefinding 12
2.11. Item Selection and Test Creation (Criteria and Process) 12
2.12. Educator Form Construction 13
2.13. Test Form Production 14
2.14. Final Eyes Committees 14
2.15. Standard Setting 15
Section 3: Validity 16
3.1. Content Validity 16
3.2. Construct (Internal Structure) Validity 17
3.2.1. Internal Consistency 17
3.2.2. Unidimensionality 17
3.2.3. Detection of Bias 20
Section 4: Test Administration and Scoring 22
4.1. Test Administration 22

4.2. Scoring Models 22
4.3. Scoring Procedures of Operational Tests 23
4.4. Scoring of Constructed-Response Items 23
4.5. Scorer Qualifications and Training 24
4.6. Quality Control Process 24
Section 5: Operational Test Data Collection and Classical Analysis 25
5.1. Data Collection 25
5.2. Data Processing 25
5.3. Classical Analysis and Calibration Sample Characteristics 30
5.4. Classical Data Analysis 41
5.4.1. Item Difficulty and Point-Biserial Correlation Coefficients 41
5.4.2. Omit Rates 43
5.4.3. Differential Item Functioning (DIF) 43
Section 6: IRT Calibration 44
6.1. IRT Models and Rationale for Use 44
6.2. Calibration Sample 45
6.2.1. Calibration Process 47
6.2.2. Item-Model Fit 48
6.3. Scaling and Scoring Procedure 48
6.3.1. Raw-Score-to-Theta-Score Conversion Tables 49
6.3.2. Theta Adjustments 49
6.3.3. Mean and Standard Deviation of Adjusted Theta Scores 50
6.3.4. Scaling Coefficients 51
6.3.5. Raw-Score-to-Scale Score Conversion Tables, CSEMs, and Performance Levels 52
6.4. Test Characteristic Curves and CSEMs 53
Section 7: Reliability and Standard Error of Measurement 66
7.1. Test Reliability 66
7.1.1. Test Statistics and Reliability for Total Test 66
7.1.2. Reliability of MC Items 67
7.1.3. Reliability of CR Items 68
7.1.4. Test Reliability for Subgroups 69
7.2. Standard Error of Measurement (SEM) 78
7.3. Performance Level Classification Consistency and Accuracy 79
7.3.1. Consistency 79
7.3.2. Accuracy 81
Section 8: Standard Setting 83
8.1. Goals of Standard Setting 83
8.2. Participants 83
8.3. Methodology 83
8.4. Standard-Setting Process 83
8.5. Results 84
Section 9: Summary of Operational Test Results 85

9.1. Scale Score Distribution Summary 85
9.1.1. ELA Scale Score and Subscore Distributions 85
9.1.1.1. ELA Grade 3 86
9.1.1.2. ELA Grade 4 87
9.1.1.3. ELA Grade 5 88
9.1.1.4. ELA Grade 6 89
9.1.1.5. ELA Grade 7 90
9.1.1.6. ELA Grade 8 91
9.1.2. Mathematics Scale Score and Subscore Distributions 92
9.1.2.1. Mathematics Grade 3 93
9.1.2.2. Mathematics Grade 4 95
9.1.2.3. Mathematics Grade 5 96
9.1.2.4. Mathematics Grade 6 97
9.1.2.5. Mathematics Grade 7 98
9.1.2.6. Mathematics Grade 8 100
9.2. Performance Level Distribution Summary 101
9.2.1. ELA Test Performance Level Distributions 101
9.2.1.1. ELA Grade 3 102
9.2.1.2. ELA Grade 4 103
9.2.1.3. ELA Grade 5 104
9.2.1.4. ELA Grade 6 105
9.2.1.5. ELA Grade 7 106
9.2.1.6. ELA Grade 8 107
9.2.2. Mathematics Test Performance Level Distributions 108
9.2.2.1. Mathematics Grade 3 109
9.2.2.2. Mathematics Grade 4 110
9.2.2.3. Mathematics Grade 5 112
9.2.2.4. Mathematics Grade 6 113
9.2.2.5. Mathematics Grade 7 115
9.2.2.6. Mathematics Grade 8 116
References 118
Appendix A: ELA and Mathematics Test Configurations and Testing Times 121
Appendix B: ELA and Mathematics Test Blueprints 124
Appendix C: Passage Selection Guidelines for Assessing ELA 126
Appendix D: Universal Design Item Checklist 127
Appendix E: Criteria for Item Acceptability 130
Appendix F: Psychometric Guidelines for Operational Item Selection 132
Appendix G: Operational Item Maps 133
Appendix H: ELA Short-Response Rubric 157
Appendix I: ELA Extended-Response Rubrics 158
Appendix J: Mathematics Short-Response Rubrics 162
Appendix K: Mathematics Extended-Response Rubric 163
Appendix L: Factor Analysis Results for Selected Subgroups 164
Appendix M: Classical Test Theory Statistics 171

Appendix N: IRT Statistics 182
Appendix O: Derivation and Estimation of Classification Consistency and Accuracy 193
Appendix P: Raw-Score-to-Scale Score and Scale Score Frequency Tables 196
Appendix Q: Standard Setting Technical Report 224

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 3.1. ELA Tests Factor Analysis 19
Table 3.2. Mathematics Tests Factor Analysis 19
Table 5.1. ELA Grade 3 Data Cleaning 25
Table 5.2. ELA Grade 4 Data Cleaning 26
Table 5.3. ELA Grade 5 Data Cleaning 26
Table 5.4. ELA Grade 6 Data Cleaning 27
Table 5.5. ELA Grade 7 Data Cleaning 27
Table 5.6. ELA Grade 8 Data Cleaning 27
Table 5.7. Mathematics Grade 3 Data Cleaning 28
Table 5.8. Mathematics Grade 4 Data Cleaning 28
Table 5.9. Mathematics Grade 5 Data Cleaning 29
Table 5.10. Mathematics Grade 6 Data Cleaning 29
Table 5.11. Mathematics Grade 7 Data Cleaning 29
Table 5.12. Mathematics Grade 8 Data Cleaning 30
Table 5.13. ELA Grade 3 Sample Characteristics 30
Table 5.14. ELA Grade 4 Sample Characteristics 31
Table 5.15. ELA Grade 5 Sample Characteristics 32
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 40
Table 5.25. Item Analysis Flagging Criteria 41
Table 5.26. Number of Flagged Items 41
Table 5.27. Item Difficulty Distribution 42
Table 5.28. Item Discrimination Distribution 42
Table 6.1. ELA Demographic Statistics 45
Table 6.2. Mathematics Demographic Statistics 46
Table 6.3. ELA Calibration Results 47
Table 6.4. Mathematics Calibration Results 47
Table 6.5. Smoothing Rules for the 2023 Administration 50
Table 6.6. Example of Smoothing at the End of the Scale for ELA and Mathematics 50
Table 6.7. Mean and Standard Deviation of Adjusted Theta Scores 51

Table 6.8. Level 3 Cut Score and Standard Deviation of Scale Scores 51
Table 6.9. Scaling Coefficients 52
Table 6.10. ELA Scale Score Ranges Associated with Each Performance Level 53
Table 6.11. Mathematics Scale Score Ranges Associated with Each Performance Level 53
Table 7.1. ELA Test Form Statistics 66
Table 7.2. ELA Test Reliability and Standard Error of Measurement 67
Table 7.3. Mathematics Test Form Statistics 67
Table 7.4. Mathematics Test Reliability and Standard Error of Measurement 67
Table 7.5. ELA MC Item Reliability and Standard Error of Measurement 68
Table 7.6. Mathematics MC Item Reliability and Standard Error of Measurement 68
Table 7.7. ELA CR Item Reliability and Standard Error of Measurement 68
Table 7.8. Mathematics CR Item Reliability and Standard Error of Measurement 69
Table 7.9. ELA Grade 3 Test Reliability by Subgroup 69
Table 7.10. ELA Grade 4 Test Reliability by Subgroup 70
Table 7.11. ELA Grade 5 Test Reliability by Subgroup 71
Table 7.12. ELA Grade 6 Test Reliability by Subgroup 72
Table 7.13. ELA Grade 7 Test Reliability by Subgroup 72
Table 7.14. ELA Grade 8 Test Reliability by Subgroup 73
Table 7.15. Mathematics Grade 3 Test Reliability by Subgroup 74
Table 7.16. Mathematics Grade 4 Test Reliability by Subgroup 75
Table 7.17. Mathematics Grade 5 Test Reliability by Subgroup 75
Table 7.18. Mathematics Grade 6 Test Reliability by Subgroup 76
Table 7.19. Mathematics Grade 7 Test Reliability by Subgroup 77
Table 7.20. Mathematics Grade 8 Test Reliability by Subgroup 78
Table 7.21. Decision Consistency (All Cuts) 80
Table 7.22. Decision Consistency (Level 3 Cut) 81
Table 7.23. Decision Agreement (Accuracy) Estimates 81
Table 8.1. ELA Performance-Level Cut Scores 84
Table 8.2. Mathematics Performance-Level Cut Scores 84
Table 9.1. ELA Scale Score Distribution Summary 86
Table 9.2. ELA Subscore Summary 86
Table 9.3. ELA Grade 3 Scale Score Distribution by Subgroup 87
Table 9.4. ELA Grade 4 Scale Score Distribution by Subgroup 88
Table 9.5. ELA Grade 5 Scale Score Distribution by Subgroup 89
Table 9.6. ELA Grade 6 Scale Score Distribution by Subgroup 90
Table 9.7. ELA Grade 7 Scale Score Distribution by Subgroup 91
Table 9.8. ELA Grade 8 Scale Score Distribution by Subgroup 92
Table 9.9. Mathematics Scale Score Distribution Summary 93
Table 9.10. Mathematics Subscore Summary 93
Table 9.11. Mathematics Grade 3 Scale Score Distribution by Subgroup 94
Table 9.12. Mathematics Grade 4 Scale Score Distribution by Subgroup 95

Table 9.13. Mathematics Grade 5 Scale Score Distribution by Subgroup 96
Table 9.14. Mathematics Grade 6 Scale Score Distribution by Subgroup 98
Table 9.15. Mathematics Grade 7 Scale Score Distribution by Subgroup 99
Table 9.16. Mathematics Grade 8 Scale Score Distribution by Subgroup 100
Table 9.17. ELA Test Performance Level Distributions 102
Table 9.18. ELA Grade 3 Performance Level Distribution by Subgroup 102
Table 9.19. ELA Grade 4 Performance Level Distribution by Subgroup 103
Table 9.20. ELA Grade 5 Performance Level Distribution by Subgroup 104
Table 9.21. ELA Grade 6 Performance Level Distribution by Subgroup 106
Table 9.22. ELA Grade 7 Performance Level Distribution by Subgroup 107
Table 9.23. ELA Grade 8 Performance Level Distribution by Subgroup 108
Table 9.24. Mathematics Test Performance Level Distributions 109
Table 9.25. Mathematics Grade 3 Performance Level Distribution by Subgroup..... 109
Table 9.26. Mathematics Grade 4 Performance Level Distribution by Subgroup..... 111
Table 9.27. Mathematics Grade 5 Performance Level Distribution by Subgroup..... 112
Table 9.28. Mathematics Grade 6 Performance Level Distribution by Subgroup..... 114
Table 9.29. Mathematics Grade 7 Performance Level Distribution by Subgroup..... 115
Table 9.30. Mathematics Grade 8 Performance Level Distribution by Subgroup..... 116
Table A1. ELA Test Configuration 121
Table A2. Mathematics Test Configuration..... 121
Table A3. ELA Estimated Time on Task by Session 122
Table A4. Mathematics Estimated Time on Task by Session 122
Table B1. ELA Test Blueprint 124
Table B2. Mathematics Test Blueprint 124
Table G1. ELA Grade 3 Operational Item Map 133
Table G2. ELA Grade 4 Operational Item Map 134
Table G3. ELA Grade 5 Operational Item Map 135
Table G4. ELA Grade 6 Operational Item Map 137
Table G5. ELA Grade 7 Operational Item Map 138
Table G6. ELA Grade 8 Operational Item Map 140
Table G7. Mathematics Grade 3 Operational Item Map..... 142
Table G8. Mathematics Grade 4 Operational Item Map..... 144
Table G9. Mathematics Grade 5 Operational Item Map..... 146
Table G10. Mathematics Grade 6 Operational Item Map..... 149
Table G11. Mathematics Grade 7 Operational Item Map..... 151
Table G12. Mathematics Grade 8 Operational Item Map..... 154
Table L1. ELA Grade 3 Test Factor Analysis by Subgroup..... 164
Table L2. ELA Grade 4 Test Factor Analysis by Subgroup..... 164
Table L3. ELA Grade 5 Test Factor Analysis by Subgroup..... 165
Table L4. ELA Grade 6 Test Factor Analysis by Subgroup..... 165
Table L5. ELA Grade 7 Test Factor Analysis by Subgroup..... 166

Table L6. ELA Grade 8 Test Factor Analysis by Subgroup 167
Table L7. Mathematics Grade 3 Test Factor Analysis by Subgroup 167
Table L8. Mathematics Grade 4 Test Factor Analysis by Subgroup 168
Table L9. Mathematics Grade 5 Test Factor Analysis by Subgroup 168
Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup 169
Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup 169
Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup 170
Table M1. ELA Grade 3 Classical Item Analysis 171
Table M2. ELA Grade 4 Classical Item Analysis 171
Table M3. ELA Grade 5 Classical Item Analysis 172
Table M4. ELA Grade 6 Classical Item Analysis 173
Table M5. ELA Grade 7 Classical Item Analysis 174
Table M6. ELA Grade 8 Classical Item Analysis 175
Table M7. Mathematics Grade 3 Classical Item Analysis 176
Table M8. Mathematics Grade 4 Classical Item Analysis 177
Table M9. Mathematics Grade 5 Classical Item Analysis 178
Table M10. Mathematics Grade 6 Classical Item Analysis 179
Table M11. Mathematics Grade 7 Classical Item Analysis 180
Table M12. Mathematics Grade 8 Classical Item Analysis 181
Table N1. ELA Grade 3 OP Item Parameter Estimates 182
Table N2. ELA Grade 4 OP Item Parameter Estimates 182
Table N3. ELA Grade 5 OP Item Parameter Estimates 183
Table N4. ELA Grade 6 OP Item Parameter Estimates 184
Table N5. ELA Grade 7 OP Item Parameter Estimates 185
Table N6. ELA Grade 8 OP Item Parameter Estimates 186
Table N7. Mathematics Grade 3 OP Item Parameter Estimates 187
Table N8. Mathematics Grade 4 OP Item Parameter Estimates 188
Table N9. Mathematics Grade 5 OP Item Parameter Estimates 189
Table N10. Mathematics Grade 6 OP Item Parameter Estimates 190
Table N11. Mathematics Grade 7 OP Item Parameter Estimates 191
Table N12. Mathematics Grade 8 OP Item Parameter Estimates 192
Table P1. ELA Grade 3 RSSS Table 196
Table P2. ELA Grade 4 RSSS Table 197
Table P3. ELA Grade 5 RSSS Table 198
Table P4. ELA Grade 6 RSSS Table 199
Table P5. ELA Grade 7 RSSS Table 200
Table P6. ELA Grade 8 RSSS Table 201
Table P7. Mathematics Grade 3 RSSS Table 202
Table P8. Mathematics Grade 4 RSSS Table 203
Table P9. Mathematics Grade 5 RSSS Table 204
Table P10. Mathematics Grade 6 RSSS Table 205

Table P11. Mathematics Grade 7 RSSS Table 207
Table P12. Mathematics Grade 8 RSSS Table 208
Table P13. ELA Grade 3 Scale Score Frequency Distribution..... 209
Table P14. ELA Grade 4 Scale Score Frequency Distribution..... 210
Table P15. ELA Grade 5 Scale Score Frequency Distribution..... 211
Table P16. ELA Grade 6 Scale Score Frequency Distribution..... 212
Table P17. ELA Grade 7 Scale Score Frequency Distribution..... 213
Table P18. ELA Grade 8 Scale Score Frequency Distribution..... 215
Table P19. Mathematics Grade 3 Scale Score Frequency Distribution..... 216
Table P20. Mathematics Grade 4 Scale Score Frequency Distribution..... 217
Table P21. Mathematics Grade 5 Scale Score Frequency Distribution..... 218
Table P22. Mathematics Grade 6 Scale Score Frequency Distribution..... 219
Table P23. Mathematics Grade 7 Scale Score Frequency Distribution..... 221
Table P24. Mathematics Grade 8 Scale Score Frequency Distribution..... 222

List of Figures

Figure 6.1. Example Item Fit Plot 48
Figure 6.2. ELA Grade 3 TCC 54
Figure 6.3. ELA Grade 3 CSEM Curve 54
Figure 6.4. ELA Grade 4 TCC 55
Figure 6.5. ELA Grade 4 CSEM Curve 55
Figure 6.6. ELA Grade 5 TCC 56
Figure 6.7. ELA Grade 5 CSEM Curve 56
Figure 6.8. ELA Grade 6 TCC 57
Figure 6.9. ELA Grade 6 CSEM Curve 57
Figure 6.10. ELA Grade 7 TCC 58
Figure 6.11. ELA Grade 7 CSEM Curve 58
Figure 6.12. ELA Grade 8 TCC 59
Figure 6.13. ELA Grade 8 CSEM Curve 59
Figure 6.14. Mathematics Grade 3 TCC 60
Figure 6.15. Mathematics Grade 3 CSEM Curve 60
Figure 6.16. Mathematics Grade 4 TCC 61
Figure 6.17. Mathematics Grade 4 CSEM Curve 61
Figure 6.18. Mathematics Grade 5 TCC 62
Figure 6.19. Mathematics Grade 5 CSEM Curve 62
Figure 6.20. Mathematics Grade 6 TCC 63
Figure 6.21. Mathematics Grade 6 CSEM Curve 63
Figure 6.22. Mathematics Grade 7 TCC 64
Figure 6.23. Mathematics Grade 7 CSEM Curve 64
Figure 6.24. Mathematics Grade 8 TCC 65
Figure 6.25. Mathematics Grade 8 CSEM Curve 65

Section 1: Introduction and Overview

1.1. Introduction

This technical report provides detailed information regarding the technical, statistical, and measurement attributes of the New York State Testing Program (NYSTP) for the Grades 3–8 English Language Arts (ELA) and Mathematics 2023 Operational Tests. This report includes information about test content and test development, item (i.e., individual test question) and test statistics, validity and reliability, test administration, standard setting, scoring, scaling, and student performance.

1.2. Test Purpose

The 2023 Grades 3–8 ELA and Mathematics NYSTP has been designed to measure student knowledge and skills as defined by grade in the New York State Next Generation Learning Standards for ELA and mathematics. The 2023 tests were the first administration measuring these new standards. The tests are designed to allow the classification of student proficiency into four performance levels (Level 1, Level 2, Level 3, and Level 4). Likewise, the test provides opportunities for students at each of these performance levels to demonstrate their knowledge and skills in the Next Generation Learning Standards. Details about the content standards for ELA and mathematics are described in Test Blueprints.

1.3. Expected Participants

Students in New York State public school Grades 3, 4, 5, 6, 7, and 8 (and ungraded students of equivalent chronological ages) are the expected participants for the Grades 3–8 NYSTP. Religious and independent schools may participate in the testing program, but their participation is not mandatory. In 2023, some religious and independent schools participated in the testing program across all grades. These schools were included in the data analyses. Public school and charter school students were required to take all State assessments administered at their grade, 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 2023 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (*SAM*), available online at https://www.nysed.gov/sites/default/files/programs/state-assessment/38-sam-2023.pdf.

1.4. Test Use and Decisions Based on Assessment

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

1.4.1. *Scale Scores*

The scale scores are a quantification of the proficiency measured by the Grades 3–8 ELA and Mathematics Tests at each grade. Scale scores are comparable only within a given subject and grade. Scale scores are not comparable across grades nor 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, including the range of scale scores for each subject and grade, is provided in Section 6: IRT Calibration. The Grades 3–8 ELA and Mathematics Tests' scale scores are the basis for placing students into performance levels, which are used to determine student progress within schools and districts, support registration of schools and districts, determine eligibility of students for additional educational services, and provide teachers with indicators of a student's need, or lack of need, for remediation in specific content-area knowledge.

1.4.2. *Performance Level Cut Scores and Classification*

Student performance is classified as Level 1, Level 2, Level 3, or Level 4 for the Grades 3–8 ELA and Mathematics Tests. The definition of each performance level is as follows:

- **NYS Level 1:** Students performing at this level are below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices, as embodied by the Next Generation Learning Standards, that are considered insufficient for the expectations at this grade.

- **NYS Level 2:** Students performing at this level are partially proficient in standards for their grade. They demonstrate knowledge, skills, and practices, as embodied by the Next Generation Learning Standards, that are considered partial but insufficient for the expectations at this grade. Students performing at Level 2 are considered on track to meet current New York State high school graduation requirements but are not yet proficient in the Next Generation Learning Standards at this grade.

- **NYS Level 3:** Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices, as embodied by the Next Generation Learning Standards, that are considered sufficient for the expectations at this grade.

- **NYS Level 4:** Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices, as embodied by the Next Generation Learning Standards, that are considered more than sufficient for the expectations at this grade.

The performance level cut scores used to distinguish between Levels 1, 2, 3, and 4 were established during the standard-setting process in summer 2023. This process is described in detail in Section 8: and Appendix Q: Standard Setting Technical Report.

1.4.3. *Subscores*

The Grades 3–8 ELA Tests have two subscores: reading (which includes all multiple-choice items assessing both reading and language standards) and writing to sources (which includes all constructed-response items assessing reading, writing, and language standards). The Grades 3–8 Mathematics Tests have three subscores that are the domain-level scores for items measuring the major clusters in each grade. The New York State Next Generation Learning Standards are divided into *Major*, *Supporting*, and *Additional* clusters. Standards within major clusters are the intended focus of instruction and assessment; these standards account for the majority of the mathematics test items. The *Supporting* and *Additional* clusters are mathematics standards that

both introduce and reinforce the major clusters. Tables 1.1 and 1.2 present the reporting subscore categories and the point values that correspond to each on the 2023 tests.

Table 1.1. ELA Subscore Categories and Total Possible Score Points

Grade	Total Subscore Points	
	Reading	Writing to Sources
3	23	10
4	23	14
5	26	14
6	26	14
7	33	14
8	33	14

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 12	Number and Operations—Fractions 8	Measurement and Data 11
4	Operations and Algebraic Thinking 9	Numbers and Operations in Base 10 10	Number and Operations—Fractions 11
5	Numbers and Operations in Base 10 13	Number and Operations—Fractions 16	Measurement and Data 13
6	Ratios and Proportional Relationships 12	The Number System 9	Expressions and Equations 18
7	Ratios and Proportional Relationships 14	The Number System 10	Expressions and Equations 15
8	Expressions and Equations 15	Functions 12	Geometry 17

1.5. Testing Accommodations

In accordance with federal law under the Individuals with Disabilities Education Act (IDEA, 2004) and the "Fairness in Testing" section of the *Standards for Educational and Psychological Testing* (AERA et al., 2014, pp. 49–72), accommodations that do not alter the measurement of any construct being tested are allowed for test takers. This 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 2023 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (*SAM*), available online at https://www.nysed.gov/sites/default/files/programs/state-assessment/38-sam-2023.pdf.

1.6. Test Transcriptions

For visually impaired students, large type and braille editions of the test books are provided. In most cases, students dictate and/or record their responses, and teachers transcribe student responses to multiple-choice items onto scannable answer sheets and transcribe responses to 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 and printed by NWEA. SeeWriteHear, LLC, produces the braille editions. SeeWriteHear employs certified Library of Congress braille transcribers and delivers braille in accordance with the Braille Authority of North America (BANA) standards. Camera-ready versions of the regular test books are provided to the braille vendor, which then produces the braille editions.

1.7. Test Translations

The NYSTP Grades 3–8 Mathematics Tests are translated into eight languages: Arabic, Bengali, Chinese (Simplified), Chinese (Traditional), Haitian-Creole, Korean, Russian, and Spanish. These tests are translated in order to provide students with 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: https://www.nysedregents.org/ei/translations.html.

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

The NYSTP Grades 3–8 ELA Tests are not translated into any other language because they are assessments of proficiency in English language arts. The following testing accommodations are made available to ELLs taking the ELA Tests: separate testing location and bilingual glossaries.

Section 2: Test Design and Development

2.1. Test Descriptions

The 2023 Grades 3–8 ELA and Mathematics Tests are criterion-referenced tests composed of multiple-choice (MC) and constructed-response (CR) items based on the New York State Next Generation Learning Standards. The tests were administered in New York State classrooms during a three-day period for paper-based tests and a six-day period for computer-based tests from April to May 2023. Details on the administration and scoring of these tests can be found in Test Administration and Scoring. Additional information can be found in the 2023 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (*SAM*), available online at https://www.nysed.gov/sites/default/files/programs/state-assessment/38-sam-2023.pdf.

2.1.1. *ELA Tests*

The 2023 Grade 3–8 ELA Tests were designed to measure student literacy, as defined by the Next Generation Learning Standards. The tests assessed reading, writing, and language standards by using multiple-choice, short-response, and extended-response items. All items were based on close readings of informational, literary, or paired texts. All texts were drawn from authentic, grade-appropriate works.

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

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

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

2.1.2. *Mathematics Tests*

The 2023 Grade 3–8 Mathematics Tests were designed to measure student understanding of mathematics, as defined by the Next Generation Learning Standards. The tests required that students understand mathematics conceptually, use prerequisite skills with grade-level mathematical facts, decide which formulas and tools (e.g., protractors and rulers) to use, and

solve mathematics problems rooted in the real world. The tests contained multiple-choice, short-response (1-point and 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 Next Generation Learning Standards, some in real-world applications. Many multiple-choice items required students to complete multiple steps. Likewise, many of these items were linked to more than one standard, drawing on the simultaneous application of multiple skills and concepts.

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

2.2. Test Configuration

2.2.1. *Test Design*

The 2023 Grades 3–8 ELA Tests were composed of two sessions per grade and administered over two days during the testing administration window. Each day consisted of one session. Session 1 contained literary and informational reading passages, multiple-choice (MC) items, and two 2-point constructed-response (CR) items based on the passages. For Grade 3, Session 2 contained reading passages, MC items, and three 2-point CR items based on those passages. For Grades 4 through 8, Session 2 contained reading passages, MC items, three 2-point CR items, and one 4-point CR item based on those passages.

The 2023 Grades 3–8 Mathematics Tests were composed of two sessions per grade and administered over two days during the testing administration window. Each day consisted of one session. Session 1 contained MC items, and Session 2 contained MC items as well as 1-point CR items, 2-point CR items, and one 3-point CR item.

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

2.2.2. *Embedded Field Test Items*

In 2010, NYSED announced its commitment to embed multiple-choice items for field testing within the Spring 2012 Grades 3–8 ELA and Mathematics Operational Tests. This commitment continued for the Spring 2023 administrations of the tests. Embedding field test items allows for a better representation of student responses and provides more reliable field test data on which to build future operational tests. In other words, since the specific locations of the embedded field test items are not disclosed and they look the same as operational test items, students are 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 standalone field test forms during Spring 2023, although it did not eliminate the need for them.

2.3. New York State Educators' Involvement in Test Development

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

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

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

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

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

2.4. Test Blueprints

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

The New York State Next Generation Learning Standards for ELA are organized into four strands: reading, writing, language, and speaking/listening. Due to administration constraints, speaking/listening was determined to be best assessed only in the classroom; therefore, the ELA tests assess three of the four strands: reading, writing, and language. Content experts reviewed the reading, writing, and language standards and recommended content coverage by standard and item type, based on the depth and breadth of each standard.

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

Tables B1 and B2 in Appendix B: ELA and Mathematics Test Blueprints show the test blueprints and actual number of score points in the Grades 3–8 ELA and Mathematics Tests, respectively. The tables include the ranges of allowable points for each ELA strand, mathematics domain, and the actual number of points on the 2023 operational tests. Tables A3 and A4 in Appendix A: ELA and Mathematics Test Configurations and Testing Times display the anticipated testing times by grade for ELA and mathematics, respectively.

2.5. Passage Selection and Item Review Criteria Documents

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

Passage review criteria documents were created based on the passage-selection guidelines and were used to evaluate each potential passage and determine whether it could be used to measure the New York State Next Generation Learning Standards for ELA. The criteria documents were used to determine whether each passage suggested for testing use was grade appropriate, fair, and possessed the necessary characteristics to assess each standard. Specifically, passages were evaluated for the presence and quality of key ideas and details, craft and structure, and integration of knowledge and ideas.

Item review criteria for the Grades 3–8 ELA Tests were used to help ensure that each item was clear and fair, measured a specific standard or standards with fidelity, and conformed to the specifications for each item type. Each section of the criteria includes pertinent questions used to determine whether an item is of sufficient quality to move forward in the development process. The first two 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 being 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 grade appropriate 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 item needs to address. Finally, the criteria establish key requirements for each item type (e.g., requiring that each 2-point constructed-response item asks students to make a clear statement that can be supported with two independent, text-based pieces of evidence).

Item review criteria for the Grades 3–8 Mathematics Tests were used to ensure clarity, language and graphical appropriateness, fairness, freedom from bias, fidelity of measurement to the New York State Next Generation Learning Standards, and conformity to the expectations for specific item types and formats for each test item. Each section of the criteria includes pertinent questions that determine whether an item is of sufficient quality. The first two criteria (clarity of text 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 being 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 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 (e.g., making sure that each 3-point constructed-response item involves a multi-step process and requires students to show work).

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 et al., 2002, p. 5). 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: Universal Design Item Checklist 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 Next Generation Learning Standards–aligned test items. To do so, independent passage finders were recruited and trained, using passage-selection resources such as the passage-selection criteria. Passage finders were given assignments based on the test-blueprint requirements. Passage finders submitted passages, along with completed criteria documents and source information, to ELA content specialists, who reviewed the passages against the agreed-upon criteria. Passages that did not meet the criteria were rejected, and passages that did meet the criteria were moved forward in the process, where the text from scanned copies of the original sources was entered into templates. Once in the templates, readability metrics were determined for each text. Passages were then proofread by copyeditors, fact checked by research librarians, reviewed for content issues by science and social studies content specialists, when necessary, 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 NYSED staff provided an explanation for the rejection.

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

2.7. Item Development

Item development for the 2023 test forms was conducted during recent annual development cycles. The goal of item development is to develop a sufficient number of high-quality, Next Generation Learning Standards-aligned items to populate the test forms. Using the criteria documents for both content areas and the multiple-perspective document for mathematics, content leads trained item writers. The item writers had teaching or assessment experience in the content area for which they were writing items; experience in writing for large-scale, high-stakes assessments; and, at a 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 blueprints. For ELA, the item writers were also provided with the completed passage-criteria documents.

Item writers provided items to content specialists for review. At least two content specialists reviewed each item. After the content specialists were satisfied that all 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, when necessary. The ELA and mathematics content specialists evaluated the feedback from the different internal groups and edited the items accordingly. The items were then posted for NYSED's review and approval for moving forward in the process.

NYSED content experts retrieved and reviewed the items. If NYSED staff determined that an item did not meet the criteria, NYSED staff provided an explanation for rejection or revision. 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 during meetings at which content staff and NYSED staff reviewed and edited all the items to ensure that they met the criteria. All passages and items accepted at that meeting were moved forward for 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- and content-specific groups where they reviewed the items. The reviews were facilitated by NWEA content specialists and were attended by NYSED staff. For ELA, reviewers first read and then discussed the passages before reviewing items. For ELA and mathematics, the educators used the following checklists to review each item.

ELA Item Checklist:

- Is the passage reading level appropriate for the grade?
- Is the passage appropriate and fair for the grade?
- Are the passage graphics accurate and appropriate?
- Is the item aligned to the intended standard?
- Is there one and only one key?
- Are the distractors plausible?
- Is the item clearly worded and free of errors (e.g., spelling, punctuation, grammar)?
- Is the item free of bias and sensitivity concerns?

Math Item Checklist:

- Is the item aligned to the intended standard?
- Are the mathematics correct?
- Is there one and only one key?
- Are the distractors plausible?
- Is the item clearly worded?
- Is the item free of bias and sensitivity concerns?
- Are the item graphics accurate and appropriate?

As the educators reviewed the items, they discussed their judgments about them. If the educators felt that an item did not align to the standards, did not meet quality standards, or was not fair, they made recommendations for editing the item. NYSED staff and NWEA 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 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. More items are field tested than are needed for the operational forms because that enables tests to be constructed with items that include the best possible characteristics from both a content and psychometric perspective.

There were two types of approaches used for field testing in Spring 2023: embedded field testing and standalone field testing. With embedded field testing, field test items were administered within the 2023 operational test forms for most multiple-choice items. With standalone field testing, field test items were administered separately from the 2023 operational forms later in Spring 2023, which included CR items and a small number of MC items.

A variety of analyses were conducted to better understand how the items field tested in 2023 may perform on future operational forms. All 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 standalone field testing, it was necessary to ensure that the standalone field test samples were representative of the entire New York State population in terms of student achievement on prior years' tests, student gender, student ethnicity, and school Needs/Resource Capacity (NRC) category. Finally, a variety of psychometric analyses were conducted, including classical item analysis, inter-rater reliability for constructed-response items, differential item functioning (DIF), item response theory (IRT), item calibration, scaling, and fit evaluation. Many of these analyses are described at length in the *2023 Field Test Technical Report*.

2.10. Rangefinding

NWEA conducts rangefinding after constructed-response items have been field tested. The purpose of rangefinding is to have New York State educators review student-constructed responses and arrive at consensus scores based on the standards established by NYSED and the scoring rubrics. The consensus scores become the basis for operational rating guides and scoring ancillaries. To arrive at consensus, committees of New York State educators review, discuss, and rate student responses to the constructed-response field test items. NYSED content experts and NWEA Scoring Directors oversaw this process.

The first step in the rangefinding process was to have the educator committees review rubrics and a NYSED-approved grounding guide set, derived from operational scoring training materials, to familiarize teachers with the application of NYSED standards and rubrics. A grounding guide set contains student responses that illustrate the full range of scores on the rubric. This set is composed of student responses that had previously gone through the rangefinding process and been approved by NYSED and are used to guide the scoring of field test and operational student responses. Referencing the previously approved guide-set papers during the rangefinding sessions ensures consistency in the application of NYSED standards and rubrics from year-to-year.

After the committee reviewed the pre-approved grounding guide set, groups of committee members familiarized themselves with each item type, scoring a small number of responses representative of each of the different score points. After a 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 NWEA scorers, who scored the field test responses to constructed-response items.

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

The NYSTP Grades 3–8 ELA and Mathematics Tests were administered from April to May 2023. The test items were selected from the pools of available ELA and mathematics items. These items were field tested either by embedded field testing or standalone field testing in 2019 or 2022.

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

NWEA content specialists were provided the test designs, blueprints, and psychometric guidelines for item selection. The psychometric guidelines are based on the classical and IRT statistics associated with the test items. Appendix F: Psychometric Guidelines for Operational Item Selection provides general psychometric guidelines for operational item selection. For example, one of the guidelines for building the NYSTP Grades 3–8 ELA and Mathematics Tests was that the point-biserial correlation for MC items should be equal to or greater than 0.20, which would indicate that students who responded correctly to that item also tended to do well on the overall test. The few exceptions to this guideline were due to content considerations that required the inclusion of particular items. Decisions to use such items were made very carefully, and no item with a negative point-biserial correlation was allowed on the test.

Using the pool of field tested items, NWEA 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 NWEA psychometricians. If items with undesirable statistics were selected, the psychometricians proposed items with more desirable statistics. The content specialists and their leads then reviewed those items. Once the NWEA 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. NWEA content specialists and psychometricians traveled to Albany, New York, in October 2022 to finalize item selection and test creation with NYSED staff (including content and assessment experts) and New York State educators.

2.12. Educator Form Construction

During an educator form construction meeting that took place from October 24–25, 2022, in Albany, New York, educators from around the State worked with NYSED and NWEA to review the content of the proposed operational ELA passages and individual ELA and mathematics items. They looked at how those items combine to create entire operational forms and for quality and appropriateness, using their subject-matter expertise. The goal was to ensure that all test

items and forms are defensible from content and psychometric perspectives. The outcome was test forms that meet psychometric parameters and contain items that meet content criteria.

On October 24, 2022, educators reviewed Grades 4, 6, and 8. Different educators reviewed Grades 3, 5, and 7 the following day. Each grade and subject group had 5–6 educators. Since different groups of educators participated in the review of each subject and grade's test form, each morning began with a general session and then 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 their own checklist and had access to NWEA's Content Management System, which displayed the items corresponding to the order of items in the test.

- For ELA, the educators initially reviewed the first passage and a single item from the passage. Once they got used to the process, the educators reviewed the passages and the corresponding items. During this review, educators confirmed that there was only one correct answer for each multiple-choice item and that the item was aligned to the standard that it purported to address.
- 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 a 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.

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

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

Given other constraints, it was not always possible to make every change that educators recommended, but they were given the opportunity to voice any and all concerns that they had; 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 NWEA and NYSED proposed for any necessary replacements. Following each session with educators, NYSED and NWEA 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 sessions.

2.13. Test Form Production

Once the selection of items for the operational and embedded field test positions was completed, NWEA created test forms. The test forms were reviewed by NWEA content specialists and were posted for NYSED to review. NYSED and NWEA 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 NWEA reviewed copies of the test forms, the test forms were reviewed by the Final Eyes committees. For each content area, a committee consisted of thirty New York State educators from around the State. During that review, the educators were charged with taking the test to make sure that each multiple-choice item had a single correct answer and to look for errors in spelling, capitalization, punctuation, grammar, and formatting.

After the Final Eyes review and after NYSED approved edits made as a result of the review, the tests were then considered final and produced for the 2023 administration.

2.15. Standard Setting

The 2023 Grades 3–8 ELA and Mathematics Tests were the first administration based on the New York State Next Generation Learning Standards. In August 2023, after the operational administration of the 2023 tests, a standard setting meeting occurred in Albany, where approximately 65 New York State educators went through a rigorous process (guided by the best practices indicated by this intensely studied process) to recommend updated performance standards for the Next Generation Learning Standards. These recommendations were presented to the Commissioner, who, in turn, adopted the recommended standards set forth by the committees. For additional details, see Section 8: and Appendix Q: Standard Setting Technical Report.

Each grade has four performance levels. Three cut points demarcate the performance levels needed to demonstrate each ascending level of performance. 6.3.5. contains the raw-to-scale score conversion tables, SEMs, and detailed information related to the performance standards.

Section 3: Validity

The *Standards for Educational and Psychological Testing* refers to validity as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (AERA et al., 2014, p. 11). Test validation is an ongoing process of gathering evidence from many sources to evaluate the soundness of the desired score interpretations or uses. This evidence is acquired from studies of the content of the test as well as studies involving scores produced by the test. Additionally, reliability has to be taken into account 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* addresses 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 & Lehmann, 1991). Tests are now also used for the purposes of accountability. Specific to student-level outcomes, the NYSTP documents student performance in ELA as defined by the New York State ELA Next Generation Learning Standards and in mathematics as defined by the New York State Mathematics Next Generation Learning Standards.

For test-score interpretations to be appropriate for this purpose, the content of the test must be carefully matched to the specified standards. The *Standards for Educational and Psychological Testing* states that content-related evidence of validity is a central concern during test development (AERA et al., 2014). 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 a 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 standards (see Tables B1 and B2 in Appendix B: ELA and Mathematics Test Blueprints). The NYSTP test-development process requires specific attention to content representation and balance within each test form. New York State educators were involved in test construction at various development stages. For example, during the item-review process, they reviewed field test items for alignment with the Next Generation Learning Standards. Educators also participated in a process of establishing scoring rubrics for constructed-response items during rangefinding. Test Design and Development contains more information specific to the item-review process.

As a means of collecting further content validity evidence, a third-party alignment study was conducted by ACS Ventures, LLC in November 2023 to evaluate the degree to which the tests

measure the content standards they are supposed to measure. See the *Evaluation of Alignment of New York State Assessment Program to the New York Next Generation Learning Standards* for the full details of this alignment study.

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 is 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 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 Test Reliability. For the total population, the ELA reliability coefficients (Cronbach's alpha) ranged from 0.87 to 0.90. For all subgroups, the reliability coefficients were greater than or equal to 0.76, except for the non-binary gender group. For the total population, the mathematics reliability coefficients (Cronbach's alpha) ranged from 0.92 to 0.93. For all subgroups, the reliability coefficients were greater than or equal to 0.76, except for the non-binary gender group. Overall, high internal consistency of the NYSTP Grades 3–8 ELA and Mathematics Tests provided sound evidence of construct validity.

3.2.2. *Unidimensionality*

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

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

Additional evidence for the efficacy of the model involves demonstrating that the items on the New York State tests are related to one another within their respective content areas. This relationship of the items within the ELA or mathematics tests shows 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. A large main factor found using this analysis would suggest a primary construct that may be related to what the items were designed to have in common (i.e., ELA proficiency or mathematics proficiency).

To demonstrate the common factor underlying student responses to the ELA and mathematics items, principal component factor analyses were conducted on a correlation matrix of individual items for the ELA and mathematics tests. The study was conducted on New York State public, charter, and religious or independent school students for whom data were available. A large first principal component was evident in each analysis, demonstrating essential unidimensionality of the trait (i.e., proficiency) measured by each test. In other words, statistical evidence indicates that the ELA items are measuring one underlying construct, ELA proficiency, and that the mathematic items are measuring one underlying construct, mathematics proficiency.

The factor analyses conducted with the ELA and mathematics data will show almost as many underlying constructs, or factors, as there are items on the test. Therefore, it is necessary to investigate the factor analysis results further to determine the number of "meaningful" factors. Specifically, more than one factor with an eigenvalue greater than 1.0 present in each dataset would suggest the presence of small additional factors (Kaiser, 1960). The magnitude of the ratio of the variance accounted for by the first factor compared with the remaining factors also provides evidence as to the number of meaningful factors (Cattell, 1966). In addition, the total amount of variance accounted for by the main factor was evaluated.

Factor analyses related to the Grades 3–8 ELA and Mathematics Tests indicate 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 show that the first eigenvalues were at least 5 times and 6.5 times as large as the second eigenvalues for all grades.

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

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

Table 3.1. ELA Tests Factor Analysis

Grade	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
3	1	**7.89**	**28.17**	**28.17**
	2	1.25	4.48	32.65
	3	1.14	4.08	36.73
4	1	**7.24**	**24.97**	**24.97**
	2	1.22	4.22	29.19
	3	1.02	3.52	32.72
5	1	**7.04**	**21.99**	**21.99**
	2	1.41	4.4	26.39
	3	1.18	3.7	30.09
6	1	**7.09**	**22.17**	**22.17**
	2	1.4	4.38	26.55
	3	1.05	3.29	29.84
	4	1.01	3.17	33.01
7	1	**8.16**	**20.93**	**20.93**
	2	1.58	4.06	24.99
	3	1.06	2.72	27.71
	4	1.01	2.58	30.3
8	1	**8.4**	**21.54**	**21.54**
	2	1.47	3.76	25.3
	3	1.1	2.83	28.13
	4	1.08	2.77	30.9
	5	1.03	2.63	33.54

Table 3.2. Mathematics Tests Factor Analysis

Grade	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
3	1	**9.51**	**29.72**	**29.72**
	2	1.37	4.29	34.01
	3	1.08	3.37	37.37
	4	1.01	3.17	40.54
4	1	**10.3**	**27.83**	**27.83**
	2	1.36	3.67	31.5
	3	1.09	2.94	34.44
5	1	**11.31**	**30.57**	**30.57**
	2	1.73	4.68	35.25
6	1	**11.66**	**29.89**	**29.89**

Grade	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
7	2	1.34	3.44	33.32
	3	1.08	2.76	36.08
	1	**11.98**	**29.22**	**29.22**
	2	1.46	3.57	32.79
	3	1.18	2.88	35.67
8	1	**10.33**	**25.19**	**25.19**
	2	1.21	2.95	28.14
	3	1.05	2.56	30.7
	4	1.01	2.46	33.17

As additional evidence for construct validity, the same factor-analysis procedure was employed to assess the dimensionality of the ELA and mathematics construct for selected subgroups of students in each grade: ELLs, students with disabilities (SWD), and students using test accommodations (SUA). Appendix L: Factor Analysis Results for Selected Subgroups provides factor analysis results for these subgroup classifications. The results were comparable to those obtained from the total population data. Evaluation of the magnitude of the eigenvalue and proportion of variance explained by the main factor provide evidence of essential unidimensionality of the construct measured by the tests for these subgroups.

3.2.3. *Detection of Bias*

Minimizing item bias means 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 a 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 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 NWEA'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 (Jensen, 1980; Sandoval & Mille, 1980). Thus, empirical studies were conducted.

Statistical 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 (MH) methods. In each grade, for both ELA and mathematics, few items were flagged for DIF. Multiple reviewers carefully reviewed items flagged for statistically significant DIF during the operational test item selection. All such items were deemed by the reviewers to be free of bias (i.e., judged not to adversely affect any demographic subgroup) and remained in the tests. Please refer to the *2023 Field Test Technical Report* for details about the DIF analysis.

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 2023 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (*SAM*) located at https://www.nysed.gov/sites/default/files/programs/state-assessment/38-sam-2023.pdf and the 2023 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests Scoring Leader Handbook* located at https://www.nysed.gov/sites/default/files/programs/state-assessment/scoring-leader-handbook-2023.pdf.

4.1. Test Administration

The NYSTP Grades 3–8 ELA and Mathematics Tests were administered to students in either a paper-based (PBT) or computer-based (CBT) testing mode in 2023. The PBT testing window for the Grades 3–8 ELA Tests was Wednesday, April 19–Friday, April 21, and the CBT testing window for Grades 3–8 ELA Tests was Wednesday, April 19–Wednesday, April 26. The PBT testing window for the Grades 3–8 Mathematics Tests was Tuesday, May 2–Thursday, May 4, and the CBT testing window for Grades 3–8 Mathematics Tests was Tuesday, May 2–Tuesday, May 9.

Makeup-test administration windows allowed students who were ill or otherwise unable to test during their original assigned window to still take the tests. The makeup-test administration window for both PBT and CBT was Monday, April 24–Friday, April 28 for the Grades 3–8 ELA Tests and Friday, May 5–Thursday, May 11 for the Grades 3–8 Mathematics Tests.

4.2. Scoring Models

For the 2022–2023 school year, schools and school districts were able to score Grades 3–8 ELA and/or Mathematics Tests regionally, multi-districtwide, districtwide, or schoolwide, 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 include either staff from three or more school districts or staff from two or more religious or independent schools in an affiliation group (religious, independent, or charter schools may participate in regional scoring with public school districts, and each religious, independent, or charter school may be counted as one district).
2. Schools from two districts—The scorers for the school's test papers include either staff from two school districts, two religious or independent schools, two charter schools, or a combination of two from those.
3. Three or more schools within a district—The scorers for the school's test papers include staff from at least three schools within a district.
4. Two schools within a district—The scorers for the school's test papers include staff from two schools within a district (not available for CBT schools).
5. One school only (local scoring)—The scorers for the school's test papers include three or more staff for each grade, all from the same school (not available for CBT schools).
6. Private contractor—Scoring is conducted by a private contractor that does not belong to the Boards of Cooperative Educational Services (BOCES).

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

4.3. Scoring Procedures of Operational Tests

Operational tests contain multiple-choice (MC) and constructed-response (CR) items. All operational MC items are machine scored. This section describes the scoring of the operational CR items.

Qualified teachers and administrators performed the scoring of the NYSTP 2023 Grades 3–8 ELA and Mathematics Tests at designated sites. The number of personnel at a given site varies, as districts have the option of regional, districtwide, or schoolwide scoring (please refer to 4.2. for more details). Administrators are responsible for the oversight of scoring operations, including preparation of the test site, security of the test materials, and supervision of the scoring process. At each site, designated trainers teach scoring committee members the basic criteria for scoring each item and monitor the scoring sessions in the room. Facilitators or leaders, who also help monitor the sessions and enforce scoring accuracy, assist the trainers.

The titles for administrators, trainers, and facilitators vary by the scoring model that is selected. At the regional level, a site coordinator conducts oversight. A scoring leader trains the scoring committee members and monitors the sessions, and a table facilitator assists in monitoring the sessions. For each subject, oversight is structured in the same way for district- and schoolwide models. At the districtwide level, a school district administrator oversees scoring. A district subject leader trains the scoring committee members and monitors the sessions, and a school subject leader assists in monitoring the sessions. For schoolwide scoring, oversight is provided by the principal; otherwise, titles for the schoolwide model are the same as those for the districtwide model. The general title "scoring committee members" includes scorers at every site. Details on titles and responsibilities for each scoring model can be found on page 16 of the 2023 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (*SAM*), available online at https://www.nysed.gov/sites/default/files/programs/state-assessment/38-sam-2023.pdf.

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

- Scoring Model 4 (two schools within a district) and Scoring Model 5 (one school only) are not available for CBT. Please refer to page 15 of the 2023 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* for descriptions of all the scoring models.
- For CBT, scorers use the ScorePoint system to score responses.

4.4. Scoring of Constructed-Response Items

The key resources used to train scoring committee members on how to score student responses for constructed-response (CR) items are scoring guides. These guides were created by NWEA from sets of actual field test student responses that were consensus scored by NYSED and New York State teachers during rangefinding sessions. Trainers use these materials to train scoring

committee members on the criteria for scoring CR items and rubric application. Additionally, *Scoring Leader Handbooks* are distributed to provide guidelines, information, and procedures for both the scorers and Scoring Site Coordinators to facilitate scoring.

The constructed-response items are divided into three groups for scoring, and three separate scoring committee members score each constructed-response item in the group they are assigned. After scoring is completed, the table facilitator or subject (ELA or mathematics) leader conducts read behinds for the scorers and items assigned to their scoring group.

4.5. Scorer Qualifications and Training

Qualified administrators and teachers conducted the scoring of the 2023 Grades 3–8 ELA and Mathematics Tests. Trainers use scoring guides to train scoring committee members on the criteria for scoring constructed-response items. Part of the training process is the administration of a consistency assurance set (CAS) that provides the State's scoring sites with information regarding strengths and weaknesses of their scorers. This tool allows trainers to retrain their scorers, if necessary. The CAS also acknowledges those scorers who grasp all aspects of the content area being scored and are 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 New York 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 "Assigning Scorer Numbers and Questions to PBT Scoring Committee Members" section (page 25) of the 2023 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests Scoring Leader Handbook* located at https://www.nysed.gov/sites/default/files/programs/state-assessment/scoring-leader-handbook-2023.pdf.

4.6. Quality Control Process

Test books and electronic responses are randomly distributed throughout each scoring room so that completed tests from each region, district, school, or class are evenly dispersed. Scoring teams are divided into groups of three to ensure that a variety of scorers grade each test. If a scorer and a facilitator cannot reach a decision after reviewing the scoring guides, they call the NWEA Scoring Helpline, a call center established to help teachers and administrators during scoring. The helpline staff consists of trained NWEA personnel who answer questions by phone. When a member of the staff is unable to resolve an issue, it is referred to NYSED for a scoring decision. A quality check is also performed to certify that all the items are scored and that the scoring committee members darkened each score on the answer document appropriately. The log of calls received by the scoring helpline is delivered to NYSED daily during the scoring window. To affirm that scoring guidelines and policies are followed, approximately 5% of student results for both ELA and mathematics 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 more than 95% of the student test records was received from the data warehouse and delivered to NWEA at the end of May 2023. During Phase 2, "straggler files" were submitted to NWEA in June 2023.

The "straggler files" contained fewer than 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 Summary of Operational Test Results were based on the data collected from both Phase 1 and Phase 2. Data collected from public, charter, and religious or independent schools were included in all data analyses.

5.2. Data Processing

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

Data processing 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. NWEA's psychometric team performed data cleaning on 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 does not mean that the student record was invalidated. According to 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 9:).

The major groups of cases excluded from the data set (used for analyses in Sections 5, 6, and 7) were students with missing school types and those with at least one entirely missing test session. Other deleted cases included students with incorrect or incomplete grade information, duplicate record cases, no-response record cases, and/or mismatched form codes. For mathematics, the "Form Code Mismatch" category includes students tested with math translation forms. Therefore, the number of deleted cases for this category is generally higher for mathematics than for ELA.

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

Table 5.1. ELA Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	187,584
Wrong Subject	0	187,584
No Grade	0	187,584
Wrong Grade	43	187,541

Exclusion Rule	# Deleted	# Cases Remaining
Form Code Mismatch	492	187,049
School Type	2,903	184,146
Missing Entire Session	13,077	171,069
Invalid Score	0	171,069
Not Tested Reason	7,985	163,084
Out-of-Range CR Scores	0	163,084
Duplicated Record	8	163,076
Test Mode Discrepancy	0	163,076

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

Table 5.2. ELA Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	191,187
Wrong Subject	0	191,187
No Grade	0	191,187
Wrong Grade	47	191,140
Form Code Mismatch	367	190,773
School Type	3,248	187,525
Missing Entire Session	21,360	166,165
Invalid Score	0	166,165
Not Tested Reason	3,298	162,867
Out-of-Range CR Scores	0	162,867
Duplicated Record	15	162,852
Test Mode Discrepancy	0	162,852

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

Table 5.3. ELA Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	191,099
Wrong Subject	0	191,099
No Grade	0	191,099
Wrong Grade	40	191,059
Form Code Mismatch	437	190,622
School Type	3,085	187,537
Missing Entire Session	13,172	174,365
Invalid Score	0	174,365
Not Tested Reason	13,089	161,276
Out-of-Range CR Scores	0	161,276
Duplicated Record	20	161,256
Test Mode Discrepancy	0	161,256

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

Table 5.4. ELA Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	193,375
Wrong Subject	0	193,375
No Grade	0	193,375
Wrong Grade	69	193,306
Form Code Mismatch	519	192,787
School Type	3,685	189,102
Missing Entire Session	14,861	174,241
Invalid Score	0	174,241
Not Tested Reason	14,674	159,567
Out-of-Range CR Scores	0	159,567
Duplicated Record	26	159,541
Test Mode Discrepancy	0	159,541

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

Table 5.5. ELA Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	195,620
Wrong Subject	0	195,620
No Grade	0	195,620
Wrong Grade	39	195,581
Form Code Mismatch	451	195,130
School Type	4,044	191,086
Missing Entire Session	15,311	175,775
Invalid Score	0	175,775
Not Tested Reason	20,300	155,475
Out-of-Range CR Scores	0	155,475
Duplicated Record	25	155,450
Test Mode Discrepancy	0	155,450

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

Table 5.6. ELA Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	195,980
Wrong Subject	0	195,980
No Grade	0	195,980
Wrong Grade	101	195,879
Form Code Mismatch	330	195,549

Exclusion Rule	# Deleted	# Cases Remaining
School Type	4,516	191,033
Missing Entire Session	21,956	169,077
Invalid Score	0	169,077
Not Tested Reason	22,507	146,570
Out-of-Range CR Scores	0	146,570
Duplicated Record	38	146,532
Test Mode Discrepancy	0	146,532

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

Table 5.7. Mathematics Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	190,854
Wrong Subject	0	190,854
No Grade	0	190,854
Wrong Grade	45	190,809
Form Code Mismatch	7,421	183,388
School Type	3,131	180,257
Missing Entire Session	13,633	166,624
Invalid Score	0	166,624
Not Tested Reason	7,179	159,445
Out-of-Range CR Scores	0	159,445
Duplicated Record	6	159,439
Test Mode Discrepancy	0	159,439

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

Table 5.8. Mathematics Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	193,958
Wrong Subject	0	193,958
No Grade	0	193,958
Wrong Grade	54	193,904
Form Code Mismatch	7,689	186,215
School Type	3,089	183,126
Missing Entire Session	13,337	169,789
Invalid Score	0	169,789
Not Tested Reason	10,579	159,210
Out-of-Range CR Scores	0	159,210
Duplicated Record	12	159,198
Test Mode Discrepancy	0	159,198

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

Table 5.9. Mathematics Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	195,159
Wrong Subject	0	195,159
No Grade	0	195,159
Wrong Grade	41	195,118
Form Code Mismatch	7,574	187,544
School Type	3,363	184,181
Missing Entire Session	24,603	159,578
Invalid Score	0	159,578
Not Tested Reason	2,531	157,047
Out-of-Range CR Scores	0	157,047
Duplicated Record	18	157,029
Test Mode Discrepancy	0	157,029

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

Table 5.10. Mathematics Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	197,648
Wrong Subject	0	197,648
No Grade	0	197,648
Wrong Grade	52	197,596
Form Code Mismatch	8,787	188,809
School Type	4,135	184,674
Missing Entire Session	17,470	167,204
Invalid Score	0	167,204
Not Tested Reason	15,040	152,164
Out-of-Range CR Scores	0	152,164
Duplicated Record	28	152,136
Test Mode Discrepancy	0	152,136

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

Table 5.11. Mathematics Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	198,587
Wrong Subject	0	198,587
No Grade	0	198,587
Wrong Grade	43	198,544
Form Code Mismatch	9,079	189,465
School Type	4,223	185,242

Exclusion Rule	# Deleted	# Cases Remaining
Missing Entire Session	19,949	165,293
Invalid Score	0	165,293
Not Tested Reason	19,359	145,934
Out-of-Range CR Scores	0	145,934
Duplicated Record	27	145,907
Test Mode Discrepancy	0	145,907

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

Table 5.12. Mathematics Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remaining
Initial Number of Cases	n/a	154,286
Wrong Subject	0	154,286
No Grade	0	154,286
Wrong Grade	115	154,171
Form Code Mismatch	6,177	147,994
School Type	4,163	143,831
Missing Entire Session	49,890	93,941
Invalid Score	0	93,941
Not Tested Reason	674	93,267
Out-of-Range CR Scores	0	93,267
Duplicated Record	32	93,235
Test Mode Discrepancy	0	93,235

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

5.3. Classical Analysis and Calibration Sample Characteristics

The cleaned data were used for classical analyses and calibration. The demographic characteristics of students in these data sets are presented in Tables 5.13–5.18 and Tables 5.19–5.24 for ELA and mathematics, respectively, including gender, ethnicity, Needs/Resource Capacity (NRC) category, English Language Learner (ELL) status, students with disabilities (SWDs), students using test accommodations (SUAs), SWD/SUA (includes examinees who are classified as having a disability and who use at least one disability-related accommodation), and ELLs using accommodations specific to their ELL status (ELL/SUA). The NRC category 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	81,468	49.96
	Male	81,602	50.04
	Non-Binary	6	0.00
Ethnicity	Asian	17,409	10.69

Demographic Category		N-Count	% of Total N-Count
	African American	22,569	13.86
	Hispanic	44,064	27.06
	American Indian	1160	0.71
	Multiracial	6,002	3.69
	Pacific Islander	310	0.19
	White	71,350	43.81
NRC	New York	47,579	29.18
	Big 4 Cities	6,106	3.74
	Urban/Suburban	12,853	7.88
	Rural	9,411	5.77
	Average Needs	44,170	27.09
	Low Needs	20,694	12.69
	Charter School	11,754	7.21
	Religious or Independent	10,509	6.44
SWD	No	140,355	86.07
	Yes	22,721	13.93
SUA	No	143,129	87.77
	Yes	19,947	12.23
ELL	No	141,608	86.84
	Yes	21,468	13.16
SWD/ SUA	No	145,931	89.49
	Yes	17,145	10.51
ELL/ SUA	No	160,316	98.31
	Yes	2,760	1.69

Note. The total n-count was 163,076.

Table 5.14. ELA Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	80,811	49.62
	Male	82,035	50.37
	Non-Binary	6	0.00
Ethnicity	Asian	17,097	10.51
	African American	23,473	14.43
	Hispanic	44,484	27.35
	American Indian	1183	0.73
	Multiracial	5,681	3.49
	Pacific Islander	361	0.22

Demographic Category		N-Count	% of Total N-Count
NRC	White	70,342	43.26
	New York	48,769	29.95
	Big 4 Cities	6,292	3.86
	Urban/Suburban	12,220	7.50
	Rural	9,232	5.67
	Average Needs	42,381	26.02
	Low Needs	19,572	12.02
	Charter School	11,219	6.89
	Religious or Independent	13,167	8.09
SWD	No	139,376	85.58
	Yes	23,476	14.42
SUA	No	140,629	86.35
	Yes	22,223	13.65
ELL	No	144,148	88.51
	Yes	18,704	11.49
SWD/ SUA	No	144,072	88.47
	Yes	18,780	11.53
ELL/ SUA	No	160,014	98.26
	Yes	2,838	1.74

Note. The total n-count was 162,852.

Table 5.15. ELA Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	79,887	49.54
	Male	81,360	50.45
	Non-Binary	9	0.01
Ethnicity	Asian	18,159	11.27
	African American	24,016	14.91
	Hispanic	43,998	27.32
	American Indian	1180	0.73
	Multiracial	5,430	3.37
	Pacific Islander	340	0.21
	White	67,944	42.18
NRC	New York	50,758	31.48
	Big 4 Cities	6,121	3.80
	Urban/Suburban	12,496	7.75
	Rural	9,325	5.78
	Average Needs	41,812	25.93

Demographic Category		N-Count	% of Total N-Count
	Low Needs	19,625	12.17
	Charter School	10,958	6.80
	Religious or Independent	10,161	6.30
SWD	No	137,430	85.22
	Yes	23,826	14.78
SUA	No	138,032	85.60
	Yes	23,224	14.40
ELL	No	143,636	89.07
	Yes	17,620	10.93
SWD/ SUA	No	141,534	87.77
	Yes	19,722	12.23
ELL/ SUA	No	158,224	98.12
	Yes	3,032	1.88

Note. The total n-count was 161,256.

Table 5.16. ELA Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	78,699	49.33
	Male	80,814	50.65
	Non-Binary	28	0.02
Ethnicity	Asian	17,139	10.76
	African American	24,214	15.20
	Hispanic	44,113	27.70
	American Indian	1138	0.71
	Multiracial	5,317	3.34
	Pacific Islander	367	0.23
	White	66,990	42.06
NRC	New York	49,124	30.79
	Big 4 Cities	6,008	3.77
	Urban/Suburban	12,027	7.54
	Rural	9,240	5.79
	Average Needs	40,128	25.15
	Low Needs	18,974	11.89
	Charter School	11,552	7.24
	Religious or Independent	12,488	7.83
SWD	No	136,998	85.87
	Yes	22,543	14.13

Demographic Category		N-Count	% of Total N-Count
SUA	No	137,089	85.93
	Yes	22,452	14.07
ELL	No	143,355	89.85
	Yes	16,186	10.15
SWD/ SUA	No	141,326	88.58
	Yes	18,215	11.42
ELL/ SUA	No	156,600	98.16
	Yes	2,941	1.84

Note. The total n-count was 159,541.

Table 5.17. ELA Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	76,280	49.07
	Male	79,122	50.90
	Non-Binary	48	0.03
Ethnicity	Asian	16,780	10.81
	African American	24,632	15.87
	Hispanic	44,657	28.78
	American Indian	1113	0.72
	Multiracial	4,671	3.01
	Pacific Islander	317	0.20
	White	63,016	40.61
NRC	New York	50,462	32.46
	Big 4 Cities	6,057	3.90
	Urban/Suburban	11,679	7.51
	Rural	8,923	5.74
	Average Needs	36,758	23.65
	Low Needs	18,505	11.90
	Charter School	12,130	7.80
	Religious or Independent	10,936	7.04
SWD	No	133,333	85.77
	Yes	22,117	14.23
SUA	No	133,642	85.97
	Yes	21,808	14.03
ELL	No	141,954	91.32
	Yes	13,496	8.68
SWD/ SUA	No	137,449	88.42
	Yes	18,001	11.58

Demographic Category		N-Count	% of Total N-Count
ELL/ SUA	No	152,932	98.38
	Yes	2,518	1.62

Note. The total n-count was 155,450.

Table 5.18. ELA Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	71,382	48.71
	Male	75,099	51.25
	Non-Binary	51	0.03
Ethnicity	Asian	16,560	11.32
	African American	24,767	16.93
	Hispanic	42,324	28.93
	American Indian	1033	0.71
	Multiracial	4,110	2.81
	Pacific Islander	329	0.22
	White	57,182	39.08
NRC	New York	50,761	34.64
	Big 4 Cities	6,192	4.23
	Urban/Suburban	11,509	7.85
	Rural	8,651	5.90
	Average Needs	33,634	22.95
	Low Needs	16,544	11.29
	Charter School	11,280	7.70
	Religious or Independent	7,961	5.43
SWD	No	125,517	85.66
	Yes	21,015	14.34
SUA	No	126,240	86.15
	Yes	20,292	13.85
ELL	No	134,285	91.64
	Yes	12,247	8.36
SWD/ SUA	No	129,604	88.45
	Yes	16,928	11.55
ELL/ SUA	No	144,366	98.52
	Yes	2,166	1.48

Note. The total n-count was 146,532.

Table 5.19. Mathematics Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	79,626	49.94
	Male	79,808	50.06
	Non-Binary	5	0.00
Ethnicity	Asian	17,296	10.86
	African American	22,416	14.08
	Hispanic	41,102	25.81
	American Indian	1141	0.72
	Multiracial	5,934	3.73
	Pacific Islander	311	0.20
	White	71,030	44.61
NRC	New York	45,820	28.74
	Big 4 Cities	5,917	3.71
	Urban/Suburban	11,874	7.45
	Rural	9,441	5.92
	Average Needs	43,662	27.38
	Low Needs	20,606	12.92
	Charter School	11,616	7.29
	Religious or Independent	10,503	6.59
SWD	No	138,085	86.61
	Yes	21,354	13.39
SUA	No	140,663	88.22
	Yes	18,776	11.78
ELL	No	140,180	87.92
	Yes	19,259	12.08
SWD/ SUA	No	143,402	89.94
	Yes	16,037	10.06
ELL/ SUA	No	156,907	98.41
	Yes	2,532	1.59

Note. The total n-count was 159,439.

Table 5.20. Mathematics Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	78,908	49.57
	Male	80,282	50.43
	Non-Binary	8	0.01
Ethnicity	Asian	16,973	10.68
	African American	23,242	14.62

Demographic Category		N-Count	% of Total N-Count
	Hispanic	41,417	26.05
	American Indian	1152	0.72
	Multiracial	5,626	3.54
	Pacific Islander	347	0.22
	White	70,228	44.17
NRC	New York	46,699	29.33
	Big 4 Cities	6,167	3.87
	Urban/Suburban	11,445	7.19
	Rural	9,194	5.78
	Average Needs	41,904	26.32
	Low Needs	19,696	12.37
	Charter School	11,293	7.09
	Religious or Independent	12,800	8.04
SWD	No	137,223	86.20
	Yes	21,975	13.80
SUA	No	138,284	86.86
	Yes	20,914	13.14
ELL	No	142,573	89.56
	Yes	16,625	10.44
SWD/ SUA	No	141,743	89.04
	Yes	17,455	10.96
ELL/ SUA	No	156,496	98.30
	Yes	2,702	1.70

Note. The total n-count was 159,198.

Table 5.21. Mathematics Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	77,567	49.40
	Male	79,454	50.60
	Non-Binary	8	0.01
Ethnicity	Asian	17,957	11.45
	African American	23,593	15.04
	Hispanic	41,075	26.19
	American Indian	1155	0.74
	Multiracial	5,325	3.40
	Pacific Islander	329	0.21
	White	67,413	42.98
NRC	New York	48,108	30.64

Demographic Category		N-Count	% of Total N-Count
	Big 4 Cities	5,960	3.80
	Urban/Suburban	11,640	7.41
	Rural	9,114	5.80
	Average Needs	41,062	26.15
	Low Needs	19,508	12.42
	Charter School	11,357	7.23
	Religious or Independent	10,280	6.55
SWD	No	135,183	86.09
	Yes	21,846	13.91
SUA	No	135,724	86.43
	Yes	21,305	13.57
ELL	No	141,355	90.02
	Yes	15,674	9.98
SWD/ SUA	No	139,275	88.69
	Yes	17,754	11.31
ELL/ SUA	No	154,267	98.24
	Yes	2,762	1.76

Note. The total n-count was 157,029.

Table 5.22. Mathematics Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	74,892	49.23
	Male	77,217	50.76
	Non-Binary	27	0.02
Ethnicity	Asian	16,446	10.83
	African American	23,975	15.78
	Hispanic	39,637	26.10
	American Indian	1098	0.72
	Multiracial	5,067	3.34
	Pacific Islander	363	0.24
	White	65,308	43.00
NRC	New York	44,788	29.44
	Big 4 Cities	5,878	3.86
	Urban/Suburban	10,873	7.15
	Rural	8,909	5.86
	Average Needs	38,209	25.12
	Low Needs	18,621	12.24
	Charter School	12,387	8.14

Demographic Category		N-Count	% of Total N-Count
	Religious or Independent	12,471	8.20
SWD	No	132,027	86.78
	Yes	20,109	13.22
SUA	No	132,452	87.06
	Yes	19,684	12.94
ELL	No	139,038	91.39
	Yes	13,098	8.61
SWD/ SUA	No	136,227	89.54
	Yes	15,909	10.46
ELL/ SUA	No	149,877	98.52
	Yes	2,259	1.48

Note. The total n-count was 152,136.

Table 5.23. Mathematics Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	71,558	49.04
	Male	74,303	50.92
	Non-Binary	46	0.03
Ethnicity	Asian	15,832	10.87
	African American	23,626	16.22
	Hispanic	39,232	26.93
	American Indian	1056	0.72
	Multiracial	4,426	3.04
	Pacific Islander	310	0.21
	White	61,201	42.01
NRC	New York	45,143	30.94
	Big 4 Cities	5,880	4.03
	Urban/Suburban	10,167	6.97
	Rural	8,487	5.82
	Average Needs	35,069	24.04
	Low Needs	17,933	12.29
	Charter School	12,024	8.24
	Religious or Independent	11,204	7.68
SWD	No	126,949	87.01
	Yes	18,958	12.99
SUA	No	127,561	87.43
	Yes	18,346	12.57

Demographic Category		N-Count	% of Total N-Count
ELL	No	135,534	92.89
	Yes	10,373	7.11
SWD/ SUA	No	131,044	89.81
	Yes	14,863	10.19
ELL/ SUA	No	144,150	98.80
	Yes	1,757	1.20

Note. The total n-count was 145,907.

Table 5.24. Mathematics Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	45,151	48.43
	Male	48,042	51.53
	Non-Binary	42	0.05
Ethnicity	Asian	7,404	7.96
	African American	17,042	18.31
	Hispanic	25,672	27.59
	American Indian	640	0.69
	Multiracial	2,608	2.80
	Pacific Islander	203	0.22
	White	39,496	42.44
NRC	New York	27,221	29.20
	Big 4 Cities	4,907	5.26
	Urban/Suburban	7,349	7.88
	Rural	6,968	7.47
	Average Needs	22,948	24.61
	Low Needs	8,295	8.90
	Charter School	7,820	8.39
	Religious or Independent	7,727	8.29
SWD	No	78,271	83.95
	Yes	14,964	16.05
SUA	No	79,051	84.79
	Yes	14,184	15.21
ELL	No	85,009	91.18
	Yes	8,226	8.82
SWD/ SUA	No	81,450	87.36
	Yes	11,785	12.64
ELL/ SUA	No	91,886	98.55
	Yes	1,349	1.45

Note. The total n-count was 93,235.

5.4. Classical Data Analysis

Classical data analysis of the NYSTP Grades 3–8 ELA and Mathematics Tests consists of several important elements. One element is the analysis of item-level statistical information about student performance. It is important to verify that the items and test forms function as intended. If any serious error were to occur with an item, errors should be flagged and evaluated for rectification (suppression, credit, or other acceptable solution) during item analysis. Analyses of test-level data comprise the second element of classical data analysis. These include examination of the raw score (RS) statistics (mean and standard deviation, or "SD") and test reliability measures Cronbach's alpha (Cronbach, 1951) and the 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 Validity and 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 dichotomous item or the average proportion of the maximum score that students earned on each polytomous item. It is important to have a good range of *p* values in order to increase test reliability and 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. Point-biserial statistics are used to examine item-test correlations, or item discrimination. Items are flagged for review by a subject-matter expert according to the criteria listed in Table 5.25.

Table 5.25. Item Analysis Flagging Criteria

Item Type	*P* Value	Point-Biserial
Dichotomous	< 0.25 or > 0.90	< 0.20
Polytomous	< 0.30 or > 0.85	< 0.40

The number of 2023 OP items flagged for each content area and grade are given in Table 5.26.

Table 5.26. Number of Flagged Items

			# Flagged Items	
Subject	**Grade**	**#Items**	***P* Value**	**Point-Biserial**
ELA	3	28	–	–
	4	29	–	–
	5	32	–	–
	6	32	–	2
	7	39	–	–
	8	39	–	1
Mathematics	3	32	1	–
	4	37	2	–

			# Flagged Items	
Subject	**Grade**	**#Items**	***P* Value**	**Point-Biserial**
	5	37	1	–
	6	39	1	–
	7	41	–	–
	8	41	–	1

If an MC item is flagged, a subject-matter expert reviews the item and intended key to verify that the item was scored correctly. Choices are checked to verify that one and only one correct answer exists. If a CR item is flagged, a subject-matter expert reviews the item to ensure that all components are present (e.g., art was not omitted) and to ensure the item is clearly worded.

If no defects are found in a flagged item, a subject-matter expert may suggest a reason for the statistical flag, if apparent.

The summary statistics of the item difficulty (*p* values) and item discrimination (point-biserial correlations) for the operational tests are shown in Table 5.27 and Table 5.28, respectively. The data show there was a reasonably wide range of item difficulties for each test. The mean item difficulties ranged from 0.59 to 0.65 for ELA and 0.51 to 0.65 for mathematics.

Point-biserial correlations ranged from 0.14 to 0.66 for the ELA tests and 0.09 to 0.74 for the mathematics tests. The mean point-biserial correlations ranged from 0.39 to 0.47 for ELA and 0.45 to 0.51 for mathematics.

Table 5.27. Item Difficulty Distribution

Subject	Grade	N-Count	Mean	SD	Min	Max
ELA	3	163,076	0.62	0.12	0.41	0.84
	4	162,852	0.59	0.10	0.36	0.74
	5	161,256	0.61	0.14	0.33	0.88
	6	159,541	0.63	0.13	0.37	0.86
	7	155,450	0.62	0.10	0.42	0.79
	8	146,532	0.65	0.13	0.39	0.86
Mathematics	3	159,439	0.65	0.15	0.38	0.92
	4	159,198	0.60	0.15	0.22	0.92
	5	157,029	0.59	0.16	0.26	0.87
	6	152,136	0.54	0.12	0.13	0.75
	7	145,907	0.62	0.14	0.36	0.89
	8	93,235	0.51	0.12	0.27	0.73

Table 5.28. Item Discrimination Distribution

Subject	Grade	N-Count	Mean	SD	Min	Max
ELA	3	163,076	0.47	0.09	0.29	0.65
	4	162,852	0.43	0.11	0.23	0.66

Subject	Grade	N-Count	Mean	SD	Min	Max
	5	161,256	0.40	0.12	0.22	0.64
	6	159,541	0.39	0.14	0.14	0.66
	7	155,450	0.39	0.12	0.20	0.64
	8	146,532	0.40	0.13	0.18	0.66
Mathematics	3	159,439	0.50	0.08	0.33	0.65
	4	159,198	0.48	0.11	0.30	0.67
	5	157,029	0.51	0.10	0.32	0.74
	6	152,136	0.51	0.09	0.36	0.71
	7	145,907	0.50	0.12	0.29	0.74
	8	93,235	0.45	0.13	0.09	0.69

In Appendix M: Classical Test Theory Statistics, Tables M1–M12 illustrate classical test statistics for all items at each grade.

5.4.2. *Omit Rates*

Omit rates (i.e., the percentage of students not answering a given item) are routinely checked, based on test data, after each administration. Tables M1–M12 in Appendix M: Classical Test Theory Statistics show the omit rates for items on the Grades 3–8 ELA and Mathematics Tests. The industry standard general rule is that omit rates for multiple-choice items should be less than 5%; omit rates for items on the Grades 3–8 ELA and Mathematics Tests were less than 1%.

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. The psychometric phenomenon of DIF has been extensively investigated, and experts' judgments of bias was collected when items were field tested, which reduced the likelihood of including any differentially functioning items on the operational forms. DIF was evaluated for operational items using two methods: the Mantel-Haenszel Delta method (Dorans, & Holland, 1992) for dichotomous items and the standardized mean difference method (Dorans et al., 1992) for polytomous items. Please refer to the *2023 Field-test Technical Report* for details about these DIF methods and item-flagging criteria. 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) to identify potential systematic issues that could be addressed in future item writing.

Section 6: IRT Calibration

6.1. IRT Models and Rationale for Use

Item response theory (IRT) allows for comparisons between item-level statistics, 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).

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, two parameters are estimated: the discrimination parameter and the difficulty parameter(s). 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. A difficulty parameter is an index of how easy or difficult an item is; the higher the difficulty parameter is, the harder the item is.

Because the characteristics of dichotomous and polytomous items are different, two IRT models were used in item calibration. The two-parameter logistic (2PL) model (Lord, 1980; Lord & Novick, 1968) was used in the analysis of dichotomous items. In this model, the probability that a student with proficiency θ responds correctly to item i is:

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

where a_i is the item discrimination and b_i is the item difficulty.

For analysis of the polytomous items, the generalized partial-credit (GPC) model (Muraki, 1992) was used. The probability of a student with proficiency θ obtaining score k on item i is:

$$P_{ik}(\theta) = P(x_i = k) = \frac{e^{\sum_{h=0}^{k} Da_i(\theta - b_i + d_{ih})}}{\sum_{c=0}^{m_i} e^{\sum_{h=0}^{c} Da_i(\theta - b_i + d_{ih})}},$$

where:

- m_i is the maximum number of score points of item i,
- a_i is the discrimination parameter of item i,
- b_i is the location parameter of item i,
- d_{ih} is the category parameter of item i on score h,
- $d_{i0} = 0$,
- $\sum_{k=1}^{m_i} d_{ik} = 0$, and
- D is a scaling constant of 1.7.

Each item has one discrimination parameter, one location parameter, and m_i -1 independent category parameters.

6.2. Calibration Sample

The cleaned data were used for calibration of the NYSTP 2023 Grades 3–8 ELA and Mathematics Tests. It should be noted that the sample sizes were adequate, as the calibration was performed using nearly all the New York State public and non-public school student population data in each tested grade. Tables 6.1 and 6.2 show the percentages of the 2023 operational test samples by demographic group for ELA and mathematics, respectively. The subgroups include gender, ethnicity, Needs/Resource Capacity (NRC) category, English Language Learner (ELL) status, students with disabilities (SWDs), students using test accommodations (SUAs), SWD/SUA (includes examinees who are classified as having a disability and who use at least one disability-related accommodation), and ELLs using accommodations specific to their ELL status (ELL/SUA).

Table 6.1. ELA Demographic Statistics

Demographic Category		Grade					
		3	4	5	6	7	8
Gender	Female	49.99	49.67	49.55	49.39	49.13	48.77
	Male	50.01	50.33	50.44	50.59	50.84	51.20
	Non-Binary	0.00	0.00	0.01	0.02	0.03	0.04
Ethnicity	Asian	10.65	10.53	11.29	10.78	10.84	11.36
	African American	13.79	14.39	14.82	15.05	15.74	16.72
	Hispanic	26.93	27.28	27.23	27.61	28.63	28.84
	American Indian	0.71	0.73	0.73	0.71	0.71	0.71
	Multiracial	3.70	3.50	3.38	3.34	3.02	2.81
	Pacific Islander	0.19	0.22	0.21	0.23	0.20	0.23
	White	44.04	43.35	42.33	42.29	40.85	39.34
NRC	New York	28.81	29.81	31.28	30.62	32.19	34.54
	Big 4 Cities	3.78	3.88	3.80	3.79	3.92	4.26
	Urban/Suburban	7.93	7.51	7.77	7.35	7.57	7.61
	Rural	5.78	5.69	5.81	5.84	5.79	5.96
	Average Needs	27.28	26.12	26.05	25.34	23.81	23.16
	Low Needs	12.81	12.09	12.24	11.99	11.98	11.39
	Charter School	7.17	6.86	6.76	7.18	7.70	7.63
	Religious or Independent	6.46	8.05	6.28	7.88	7.04	5.45
SWD	No	86.19	85.68	85.33	86.02	85.89	85.77
	Yes	13.81	14.32	14.67	13.98	14.11	14.23
SUA	No	87.87	86.45	85.68	86.01	86.06	86.19
	Yes	12.13	13.55	14.32	13.99	13.94	13.81
ELL	No	86.91	88.55	89.12	89.90	91.37	91.72
	Yes	13.09	11.45	10.88	10.10	8.63	8.28
SWD/	No	89.59	88.56	87.85	88.67	88.52	88.49

Demographic Category		Grade					
		3	4	5	6	7	8
SUA	Yes	10.41	11.44	12.15	11.33	11.48	11.51
ELL/	No	98.34	98.29	98.13	98.17	98.40	98.53
SUA	Yes	1.66	1.71	1.87	1.83	1.60	1.47

Table 6.2. Mathematics Demographic Statistics

Demographic Category		Grade					
		3	4	5	6	7	8
Gender	Female	50.22	49.68	49.47	49.49	49.01	47.96
	Male	49.78	50.31	50.52	50.49	50.96	52.00
	Non-Binary	0.00	0.00	0.00	0.02	0.03	0.05
Ethnicity	Asian	12.02	11.76	12.38	11.51	11.47	8.72
	African American	16.56	16.51	16.61	17.45	17.61	20.48
	Hispanic	27.88	28.12	27.74	27.78	29.01	29.69
	American Indian	0.97	0.95	0.94	0.91	0.89	0.88
	Multiracial	3.55	3.36	3.26	3.32	2.95	2.64
	Pacific Islander	0.23	0.25	0.24	0.27	0.24	0.25
	White	38.79	39.07	38.83	38.76	37.83	37.35
NRC	New York	42.92	41.05	42.21	40.27	40.71	39.22
	Big 4 Cities	5.57	5.42	5.24	5.31	5.35	7.10
	Urban/Suburban	3.57	4.39	3.38	3.25	4.25	3.53
	Rural	5.96	5.81	6.04	5.92	5.86	7.33
	Average Needs	25.44	24.86	25.04	24.65	24.36	24.78
	Low Needs	5.42	7.41	7.61	8.09	7.78	5.69
	Charter School	9.77	8.93	9.00	10.39	10.22	10.55
	Religious or Independent	1.34	2.13	1.48	2.12	1.47	1.80
SWD	No	85.41	84.94	84.90	85.71	85.86	82.79
	Yes	14.59	15.06	15.10	14.29	14.14	17.21
SUA	No	87.36	85.99	85.52	86.22	86.69	83.84
	Yes	12.64	14.01	14.48	13.78	13.31	16.16
ELL	No	88.62	90.81	91.64	93.09	94.10	93.59
	Yes	11.38	9.19	8.36	6.91	5.90	6.41
SWD/	No	89.03	88.05	87.69	88.64	88.84	86.35
SUA	Yes	10.97	11.95	12.31	11.36	11.16	13.65
ELL/	No	98.06	98.02	98.02	98.32	98.61	98.42
SUA	Yes	1.94	1.98	1.98	1.68	1.39	1.58

6.2.1. *Calibration Process*

Item parameters were estimated using Scientific Software International (SSI) Inc.'s IRTPRO Version 5.2 (Cai et al., 2011) package. Dichotomous and polytomous items were calibrated simultaneously using marginal maximum likelihood procedures.

The calibration of NYSTP 2023 Grades 3–8 ELA and Mathematics Tests did not exhibit any test-level issues. The estimated parameters were on the standard normal scale, and all the items were well within the prescribed parameter ranges (i.e., non-negative a and $|b| < 4$). For both the Grades 3–8 ELA and Mathematics Tests, all calibration estimation results were reasonable. Tables 6.3 and 6.4 present the summaries of the calibration results for ELA and mathematics, respectively. Additional details, including individual item-parameter estimates, may be found in Appendix N: IRT Statistics in Tables N1–N12. The parameter estimates are expressed on the theta metric and are defined as:

- Dichotomous items:
  - a is a discrimination parameter.
  - b is a difficulty parameter.

- Polytomous items:
  - a is an item discrimination parameter.
  - b is an item-location parameter.
  - d_k is the item-category parameter for category k.

Table 6.3. ELA Calibration Results

Grade	N-Count	Range of *a*-Parameters		Range of *b*-Parameters	
3	163,076	0.41	1.40	-1.33	0.51
4	162,852	0.33	1.01	-1.09	1.09
5	161,256	0.29	1.01	-1.81	1.26
6	159,541	0.19	1.12	-1.67	1.62
7	155,450	0.26	1.18	-1.29	0.53
8	146,532	0.23	1.11	-1.84	0.89

Table 6.4. Mathematics Calibration Results

Grade	N-Count	Range of *a*-Parameters		Range of *b*-Parameters	
3	159,439	0.48	1.29	-1.93	0.65
4	159,198	0.41	1.32	-1.84	1.39
5	157,029	0.48	1.54	-1.89	0.91
6	152,136	0.51	1.47	-0.91	1.63
7	145,907	0.44	1.55	-2.07	0.71
8	93,235	0.14	1.36	-1.15	2.87

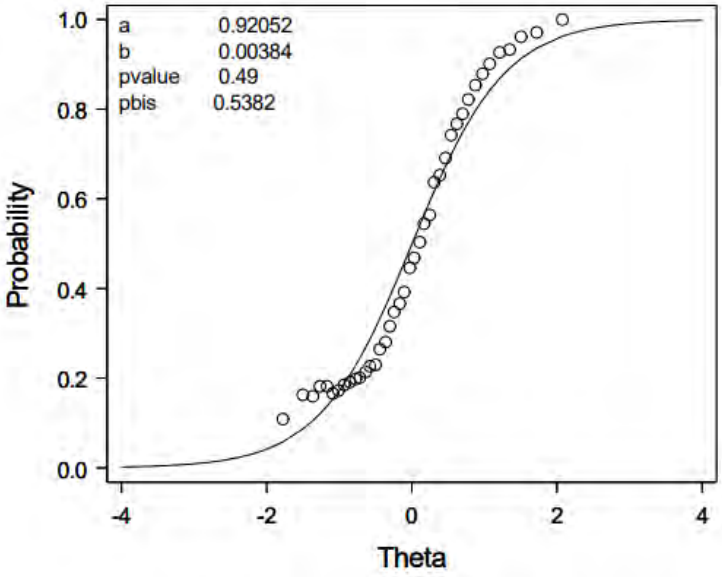
Since 2023 was the first administration of the NYSTP 2023 Grades 3–8 ELA and Mathematics Tests that measure the New York State Next Generation Learning Standards, the IRT calibration

results of 2023 will be used as the base year to anchor the IRT parameters of subsequent years. For Year 2 and beyond, the IRT item parameters will be transformed onto the scale of 2023 so that item statistics (such as item difficulty and item discrimination) and test-level statistics (such as scale scores) are comparable across years.

6.2.2. *Item-Model Fit*

The *Standards for Educational and Psychological Testing* suggests documenting evidence of model fit when model-based methods such as IRT are used to estimate item parameters in test development. The standard process of assessing the fit of an item under unidimensional IRT models involves steps such as (a) defining a number of examinee groups ("buckets") and then (b) making an informed judgment by comparing the observed and model-predicted proportion-correct scores for the item by the examinees in different "buckets" (Sinharay, 2006). To make this judgment on each item, Hambleton and Swaminathan (1985) recommend the use of graphical plots comparing the estimated/predicted item-response function to the empirical student-response data for an item. To visually examine the model-data fit for each item, fit plots were produced and closely examined for all operational items. An example item fit plot is shown in Figure 6.1.

Figure 6.1. Example Item Fit Plot



All items showed adequate model-data fit. The fact that the items in the NYSTP 2023 Grades 3–8 ELA and Mathematics Tests demonstrated good model fit further supports the use of the chosen models.

6.3. Scaling and Scoring Procedure

The 2023 Grades 3–8 ELA and Mathematics Tests are new assessments developed based on the New York State Next Generation Learning Standards, which are different from previous content standards. Even though there is overlap between the old and new standards, there are significant content shifts and depth of learning changes. The 2023 Grades 3–8 ELA and Mathematics Tests

also have new item formats that led to substantial changes in test specifications. The *Standards for Education and Psychological Testing* states that "When substantial changes in test specifications occur, scores should be reported on a new scale, or a clear statement should be provided to alert users that the scores are not directly comparable with those on earlier versions of the test" (AERA et al., 2014, p. 107). Being the first administration of the NYSTP tests to measure the Next Generation Learning Standards, a new reporting scale was established following the standard setting meeting in summer 2023. The reporting scale was developed to quantify the information captured by the assessment about what students know and can do. The reporting scale was developed to interpret changes, make comparisons, facilitate inferences, and inform educational decisions.

New York State student examinations were scored using the number correct (NC) scoring method. This method considers how many score points a student obtained on a test in determining their reported score, also called a 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.

6.3.1. *Raw-Score-to-Theta-Score Conversion Tables*

To create a raw-score-to-scale-score table, each raw score is first converted to a theta score that represents the student's proficiency under the IRT model. An inversed test characteristic curve (TCC) procedure is used to obtain the theta estimates. 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 dichotomous 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 points. For tests with a mixture of dichotomous and polytomous items, the MLE and TCC estimates are even more similar (Yen, 1984).

The inverse TCC method relies on the following equation:

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

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.

6.3.2. *Theta Adjustments*

With the adoption of the 2PL model, the θ scores can be obtained for all raw-score points, except the zero, and perfect scores using the inverse TCC method. However, the θ scores at the two

ends of the scale are much less reliable, as indicated by the large conditional standard errors of measurement (CSEMs). Therefore, an adjustment and interpolation were conducted to derive the adjusted theta scores following the rules outlined in Table 6.5.

Table 6.5. Smoothing Rules for the 2023 Administration

		Smoothing	
Subject	**Grade**	**Starting Point**	**Step Size**
ELA	3–8	CSEM > 0.56	0.16
Mathematics	3–8	CSEM > 0.44	0.12

At both ends of the scale, for any theta estimates with CSEMs greater than 0.56 for ELA and 0.44 for mathematics, 0.16 and 0.12 were subtracted (at the low end) or added (at the high end) from the preceding theta value for ELA and mathematics, respectively. Table 6.6 shows an example of smoothing at the lower end of the scale for Grade 5 ELA and mathematics.

Table 6.6. Example of Smoothing at the End of the Scale for ELA and Mathematics

ELA				Mathematics			
Raw Score	**Theta**	**CSEM**	**Adjusted Theta**	**Raw Score**	**Theta**	**CSEM**	**Adjusted Theta**
0	-99	–	-3.4369	0	-99	–	-2.5979
1	-4.4605	1.173	-3.2769	1	-3.1313	0.763	-2.4779
2	-3.6007	0.75	-3.1169	2	-2.5668	0.52	-2.3579
3	-3.1266	0.586	-2.9569	3	-2.2379	0.418	-2.2379
4	-2.7969	0.497	-2.7969	–	–	–	–

6.3.3. *Mean and Standard Deviation of Adjusted Theta Scores*

The mean and standard deviation (SD) of the theta scores were computed from the 2023 Grades 3–8 ELA and Mathematics calibration sample and are summarized in Table 6.7.

Table 6.7. Mean and Standard Deviation of Adjusted Theta Scores

Subject	Grade	Mean	SD
ELA	3	-0.00218	1.03234
	4	0.00183	1.05483
	5	0.00574	1.06079
	6	-0.00306	1.04001
	7	0.00413	1.0431
	8	-0.00263	1.03319
Mathematics	3	-0.0322	0.96453
	4	-0.01906	0.99152
	5	-0.01344	0.99162
	6	-0.00113	1.01244
	7	-0.01288	0.98599
	8	-0.00376	1.01906

6.3.4. *Scaling Coefficients*

The adjusted θ scores were converted to scale scores using a linear transformation by fixing two desired properties: the Level 3 cut score and the SD of scale scores (as shown in Table 6.8). The scale score of 450 was chosen as the desired Level 3 cut score so that the scale-score ranges of the new 2023 scale would not overlap with previous Grades 3–8 tests or other NYSTP tests. The desired SD of scale scores was set as 23 for ELA and 27 for mathematics.

Table 6.8. Level 3 Cut Score and Standard Deviation of Scale Scores

		Scaling	
Subject	Grade	Level 3 Cut	Standard Deviation
ELA	3–8	450	23
Mathematics	3–8	450	27

The scaling slope and intercept are computed as follows:

$$Slope = \frac{\sigma(ScaleScore)}{\sigma(\theta)},$$

$$Intercept = cut(ScaleScore) - \frac{\sigma(ScaleScore)}{\sigma(\theta)} cut(\theta)$$

where $\sigma(ScaleScore)$ is the desired standard deviation of scale scores (23 for ELA and 27 for mathematics); $\sigma(\theta)$ is the standard deviation of the adjusted theta scores based on the calibration sample (as shown in Table 6.7); $cut(ScaleScore)$ is 450 for both ELA and mathematics; and

$cut(\theta)$ is the theta score in the raw-to-theta conversion table that corresponds to the Level 3 cut score obtained from standard setting. Table 6.9 shows the resulting scaling coefficients for ELA and mathematics.

Table 6.9. Scaling Coefficients

Grade	Slope (M_1)	Intercept (M_2)
ELA		
3	22.27948	444.7527
4	21.80446	447.6488
5	21.68195	445.6578
6	22.11517	445.3819
7	22.04966	447.8788
8	22.26115	451.1351
Mathematics		
3	27.99291	453.3983
4	27.23092	453.381
5	27.22817	451.2373
6	26.66825	451.1035
7	27.38364	454.3584
8	26.49501	446.1055

6.3.5. *Raw-Score-to-Scale Score Conversion Tables, CSEMs, and Performance Levels*
The scale scores are calculated by applying the scaling coefficients to the theta scores as follows:

$$ScaleScore = M_1\theta + M_2$$

The raw-score-to-scale score (RSSS) conversion tables are presented in Appendix P: Raw-Score-to-Scale Score and Scale Score Frequency Tables, Tables P1–P12.

The conditional standard error of measurement (CSEM) of a scale score is calculated as follows:

$$M_1 \frac{1}{\sqrt{I(\hat{\theta})}},$$

where $\hat{\theta}$ is the theta estimate corresponding to the scale score, $I(\hat{\theta})$ is the value of the test information function (TIF) at $\hat{\theta}$, and M_1 is the scaling slope coefficient defined in Table 6.9.
The final element of the RSSS tables is the application of the performance level cut scores. Scale score cuts were obtained based on the raw score cut set in summer 2023 through the standard setting procedure (see Section 8: and Appendix Q: Standard Setting Technical Report for more information on the standard setting process).

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

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

Grade	NYS Level 1	NYS Level 2	NYS Level 3	NYS Level 4
3	380–431	432–449	450–473	474–491
4	380–430	431–449	450–470	471–500
5	371–431	432–449	450–473	474–502
6	366–430	431–449	450–469	470–497
7	366–432	433–449	450–471	472–503
8	366–429	430–449	450–471	472–503

Table 6.11. Mathematics Scale Score Ranges Associated with Each Performance Level

Grade	NYS Level 1	NYS Level 2	NYS Level 3	NYS Level 4
3	377–423	424–449	450–486	487–501
4	378–430	431–449	450–485	486–507
5	381–431	432–449	450–482	483–506
6	388–430	431–449	450–484	485–513
7	379–429	430–449	450–476	477–509
8	379–435	436–449	450–481	482–514

6.4. Test Characteristic Curves and CSEMs

Test characteristic curves (TCCs) provide an overview of the tests in the IRT scale score metric. The 2023 TCCs were generated using final item parameters for all reporting test items administered in Spring 2023. 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 6.2–6.25.

Figure 6.2. ELA Grade 3 TCC

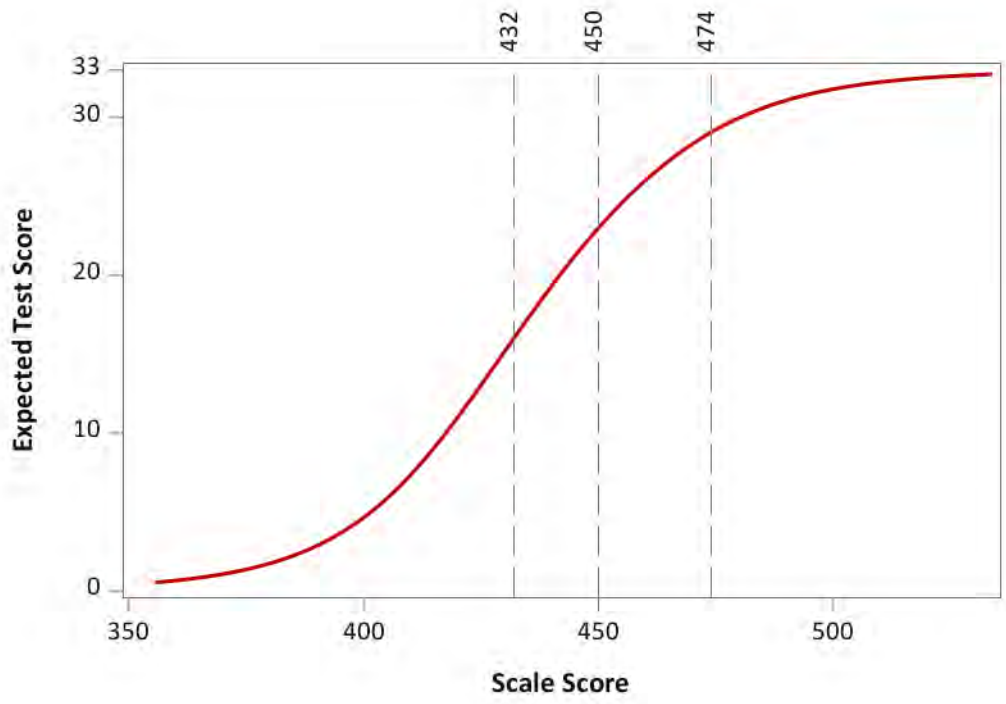


Figure 6.3. ELA Grade 3 CSEM Curve

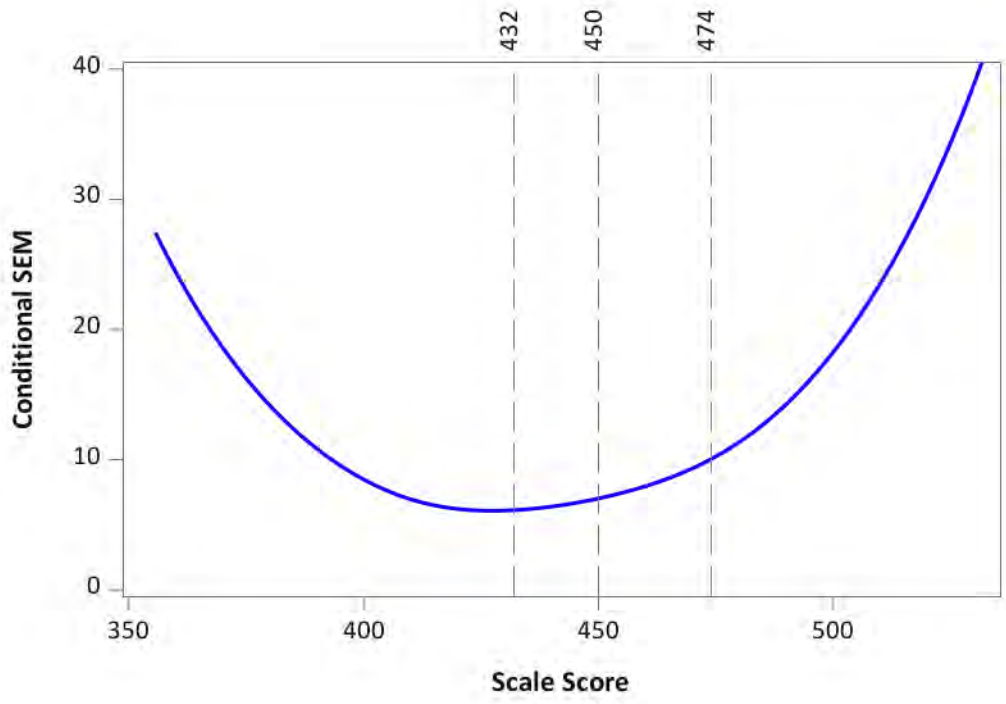


Figure 6.4. ELA Grade 4 TCC

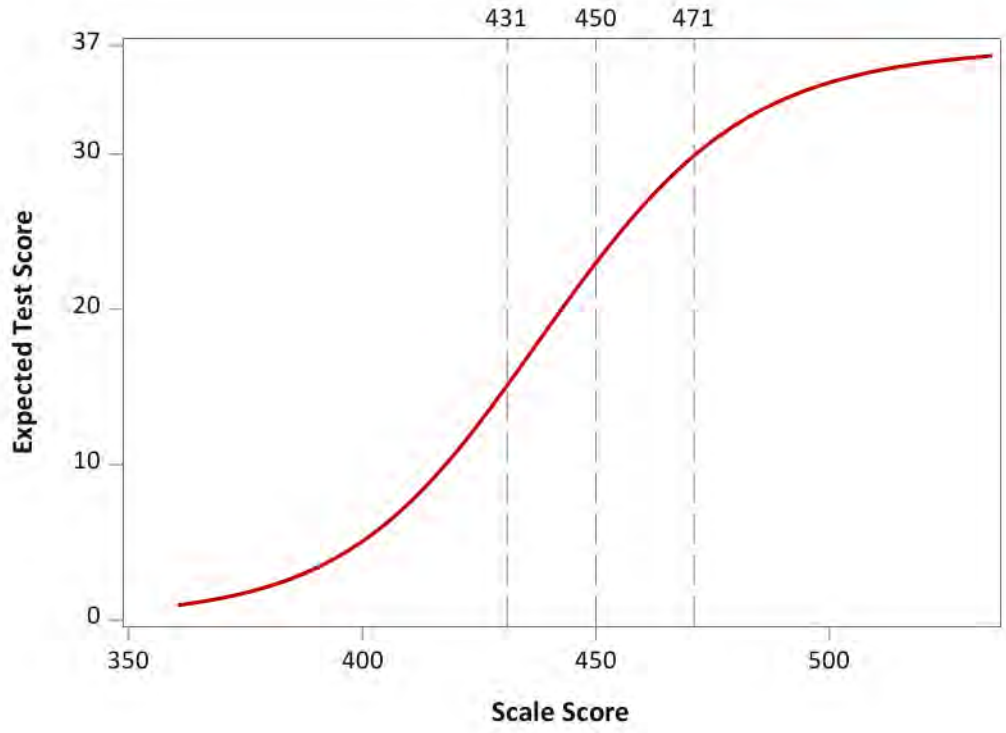


Figure 6.5. ELA Grade 4 CSEM Curve

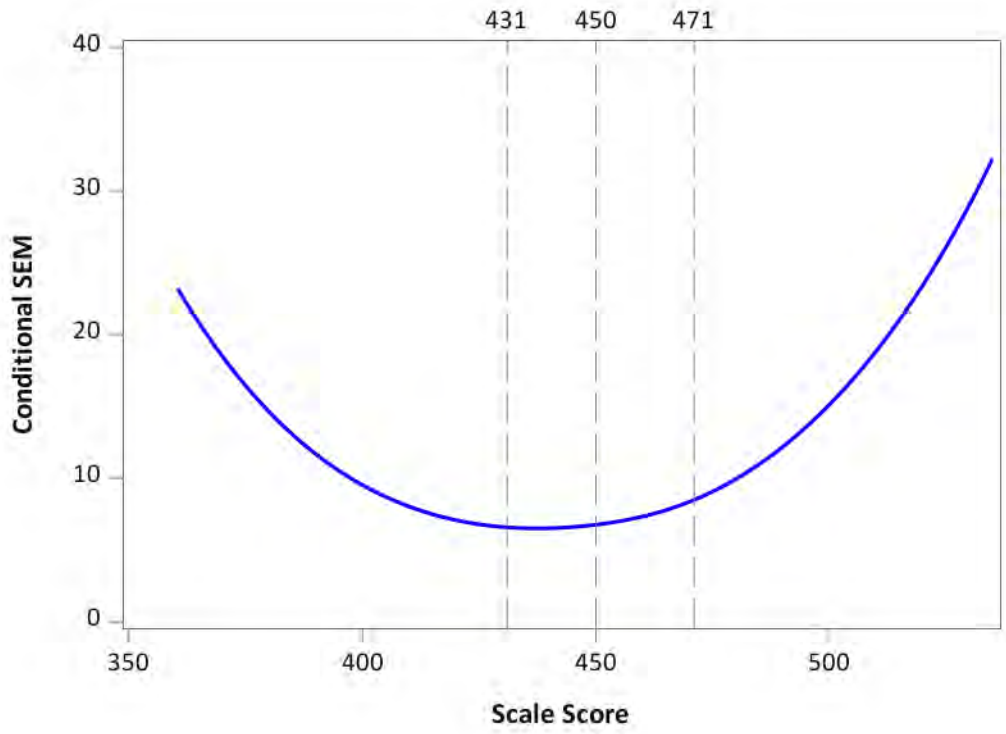


Figure 6.6. ELA Grade 5 TCC

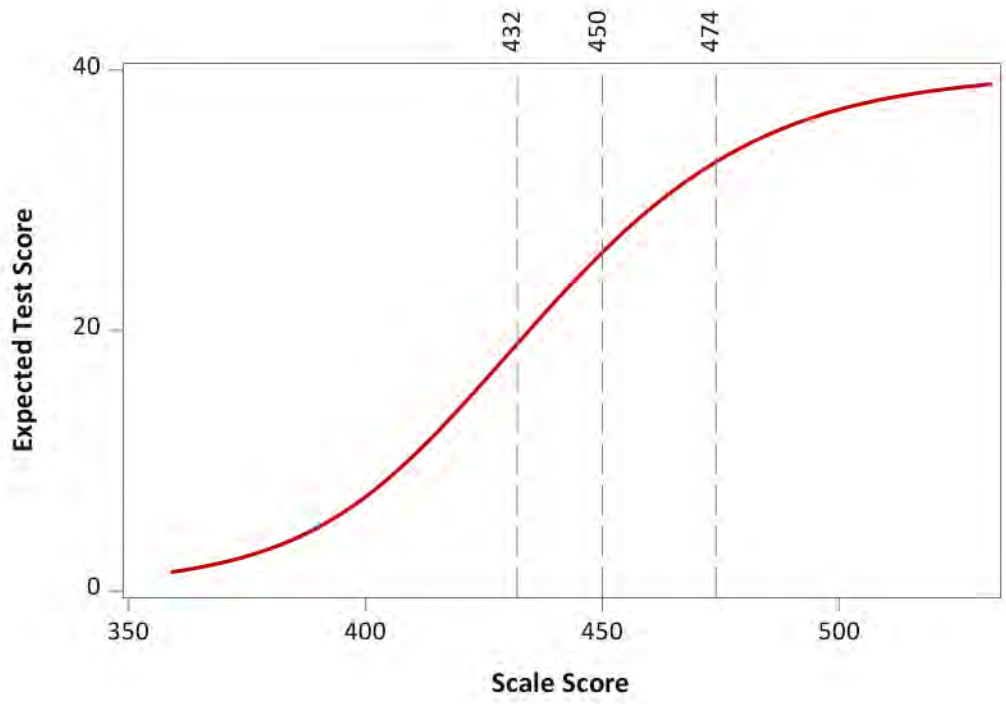


Figure 6.7. ELA Grade 5 CSEM Curve

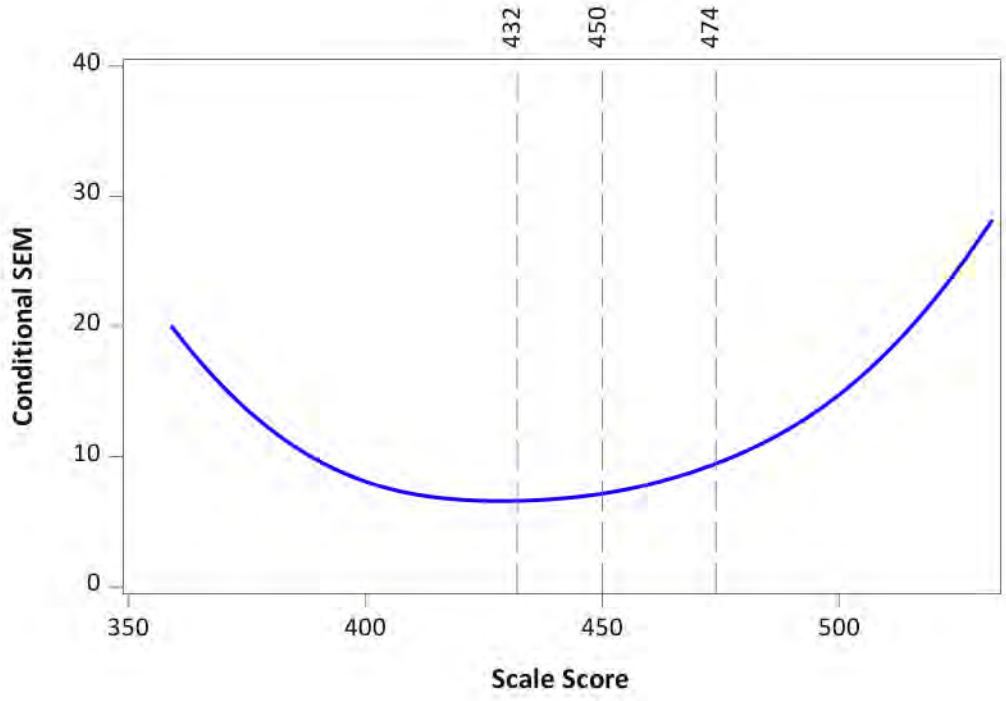


Figure 6.8. ELA Grade 6 TCC

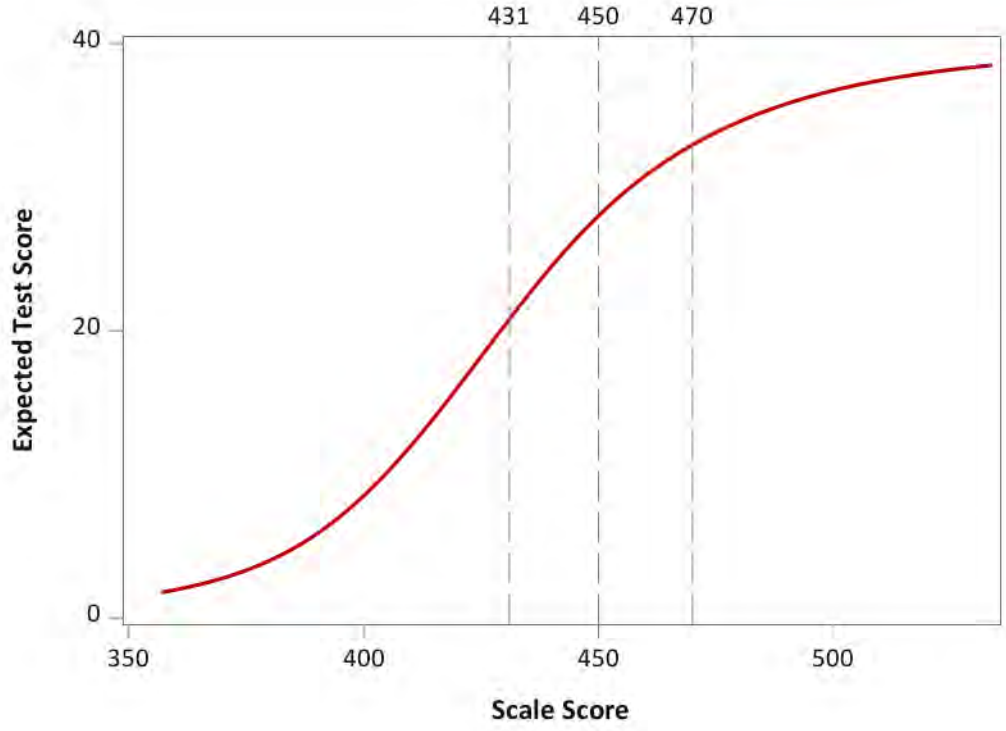


Figure 6.9. ELA Grade 6 CSEM Curve

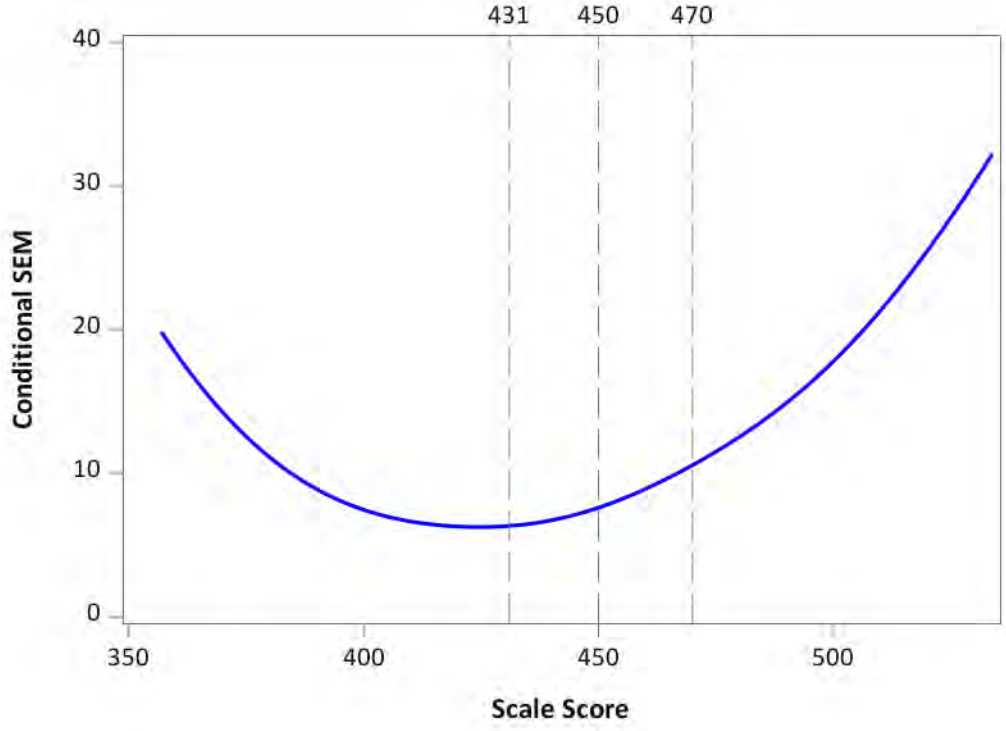


Figure 6.10. ELA Grade 7 TCC

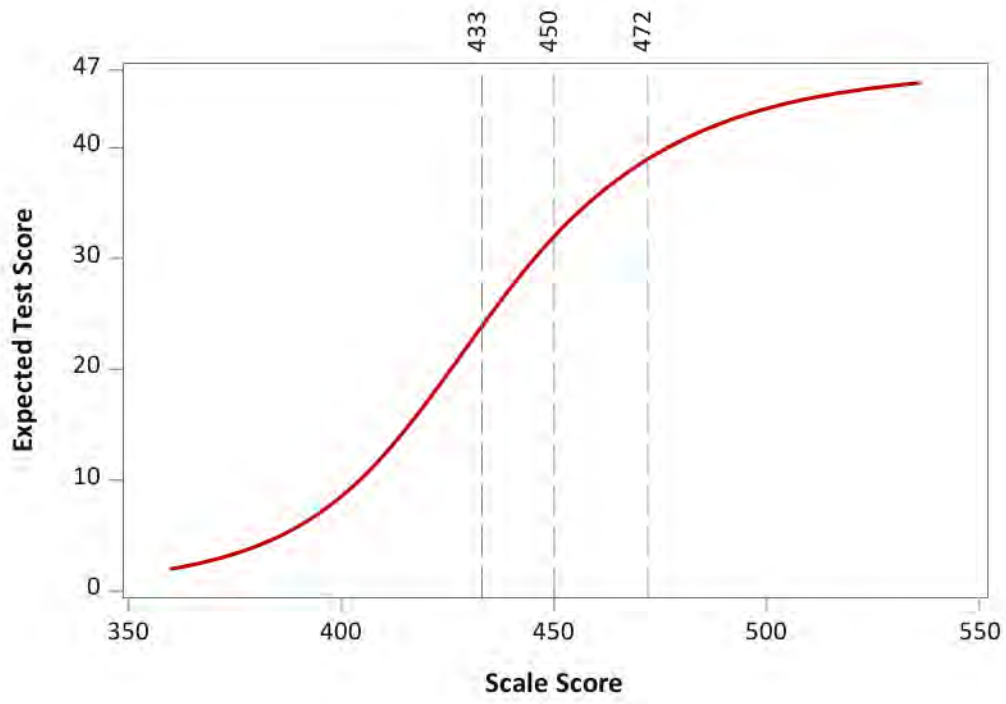


Figure 6.11. ELA Grade 7 CSEM Curve

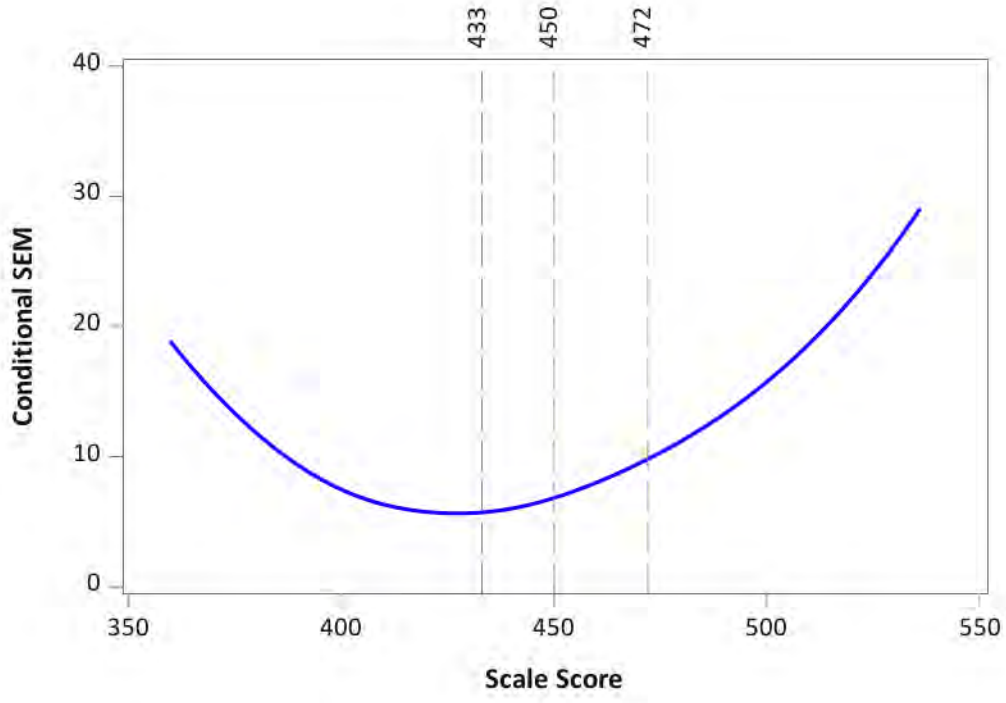


Figure 6.12. ELA Grade 8 TCC

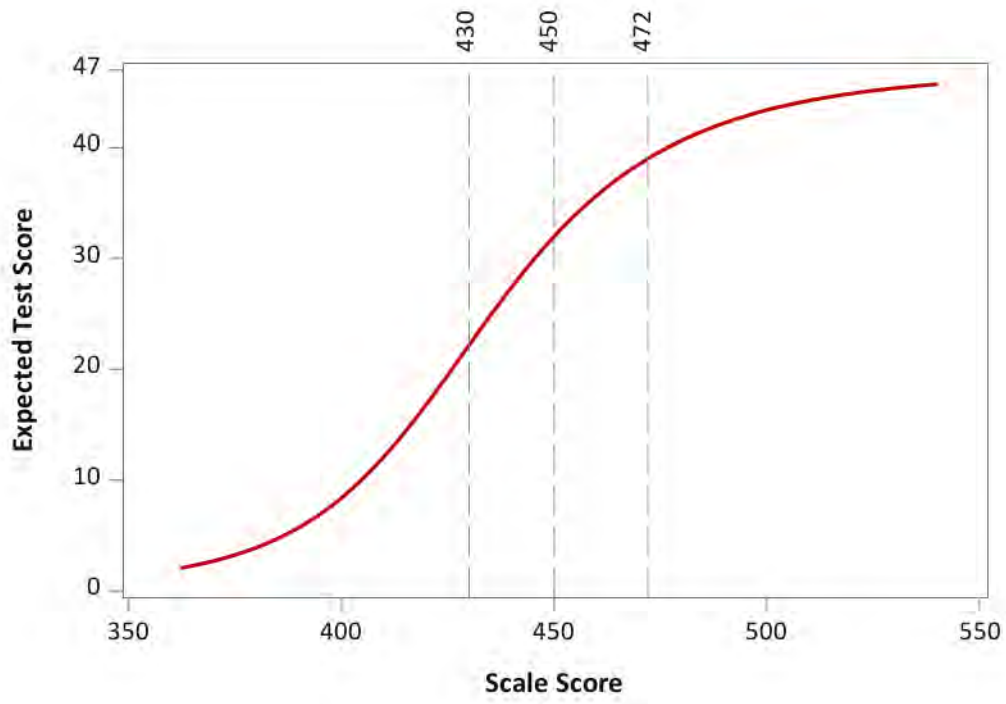


Figure 6.13. ELA Grade 8 CSEM Curve

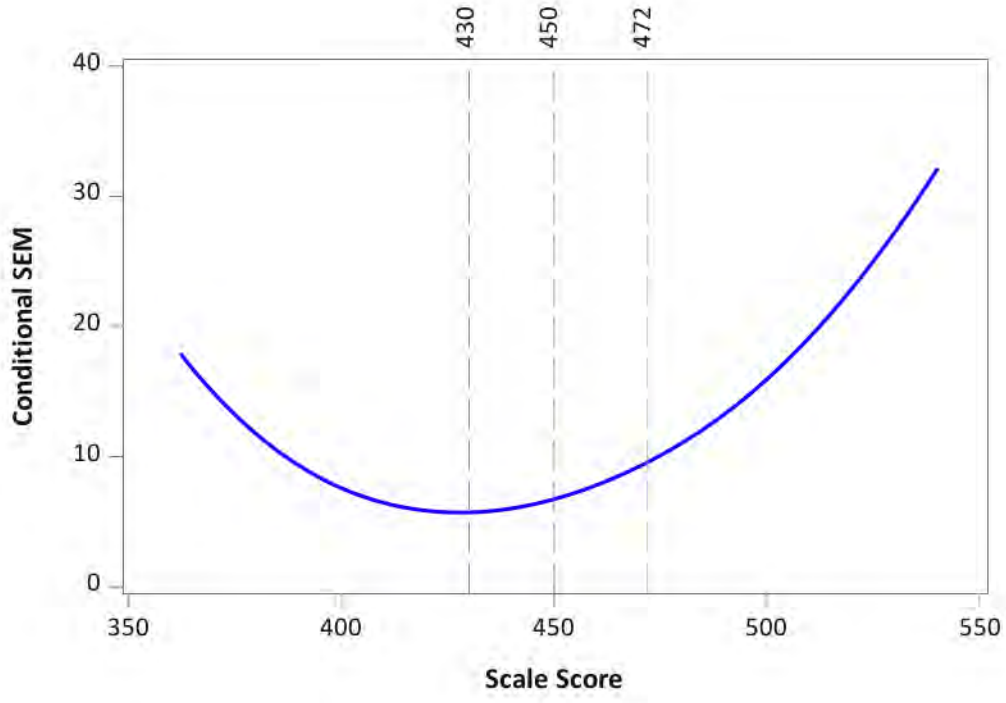


Figure 6.14. Mathematics Grade 3 TCC

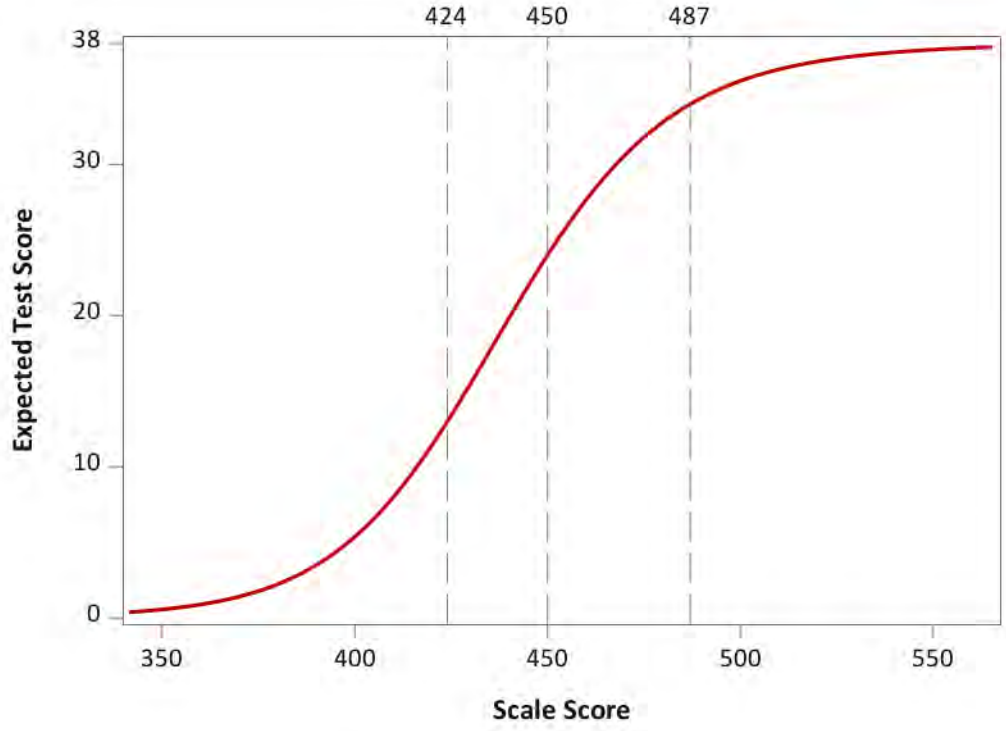


Figure 6.15. Mathematics Grade 3 CSEM Curve

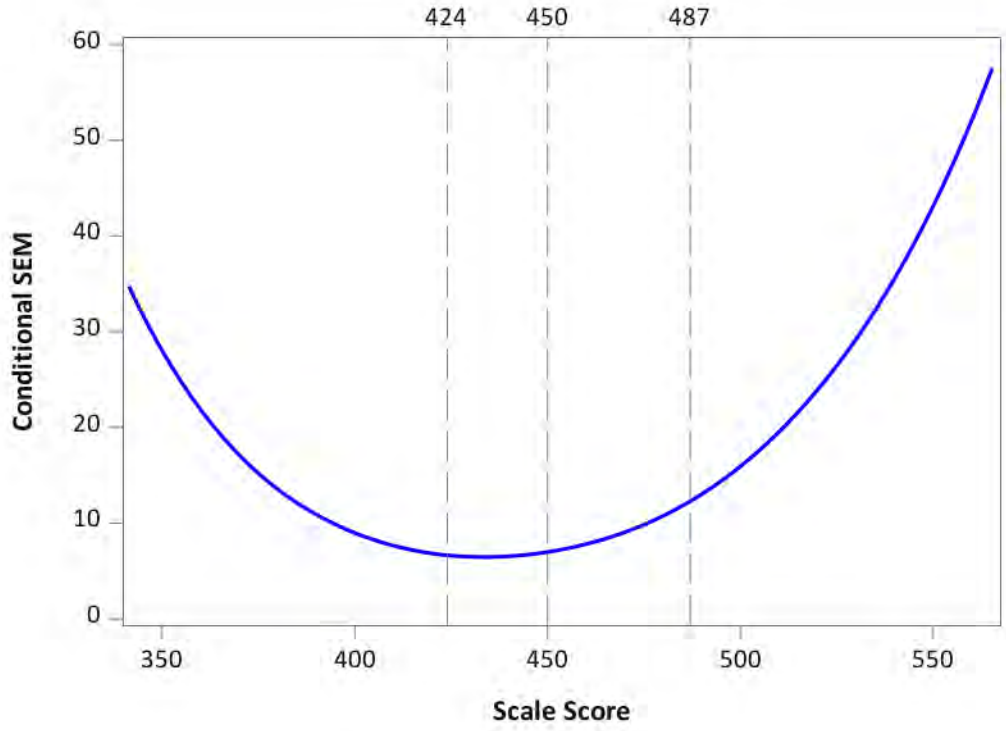


Figure 6.16. Mathematics Grade 4 TCC

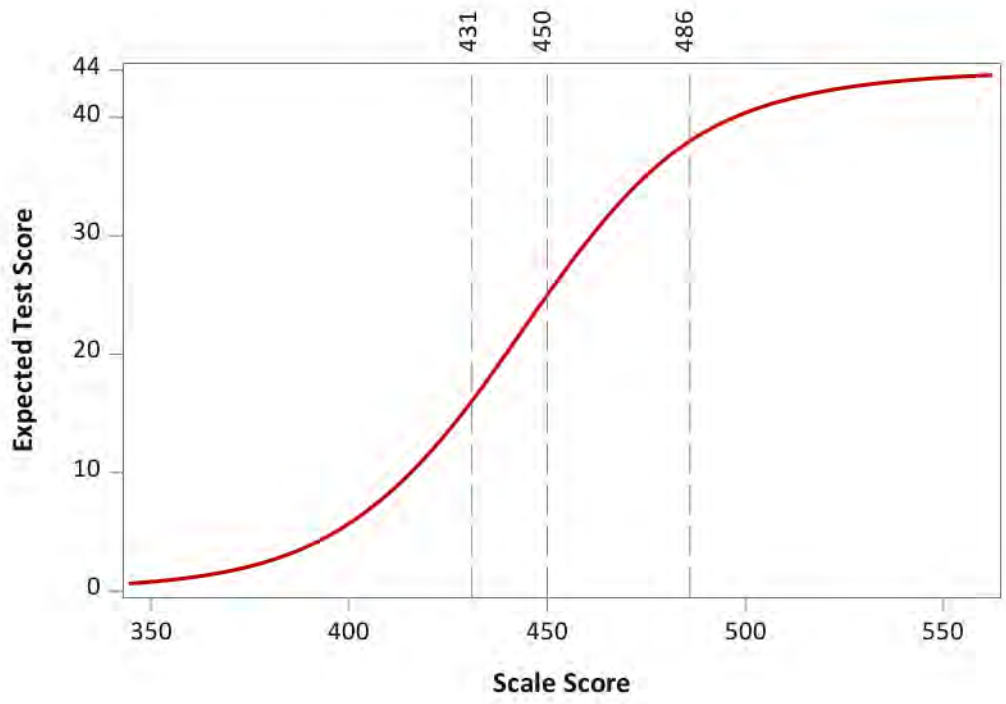


Figure 6.17. Mathematics Grade 4 CSEM Curve

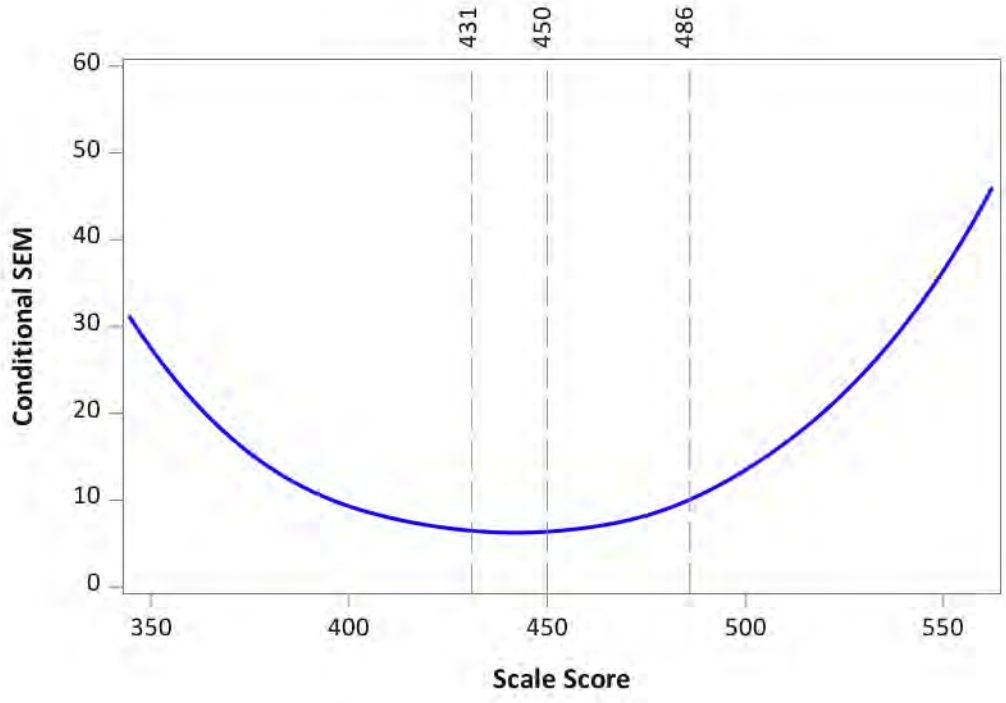


Figure 6.18. Mathematics Grade 5 TCC

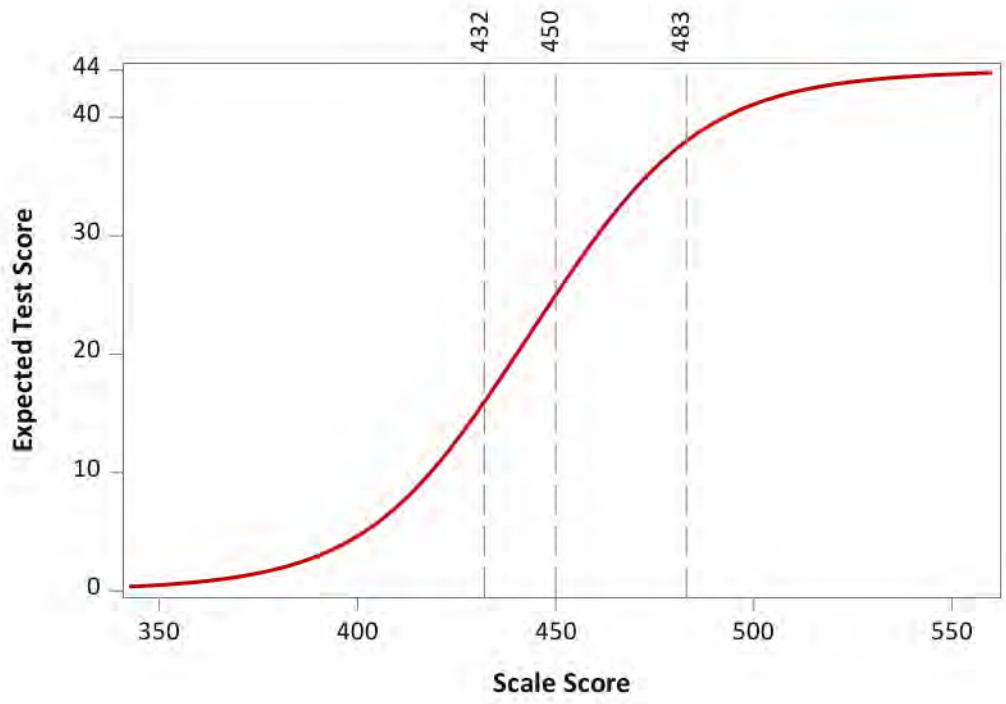


Figure 6.19. Mathematics Grade 5 CSEM Curve

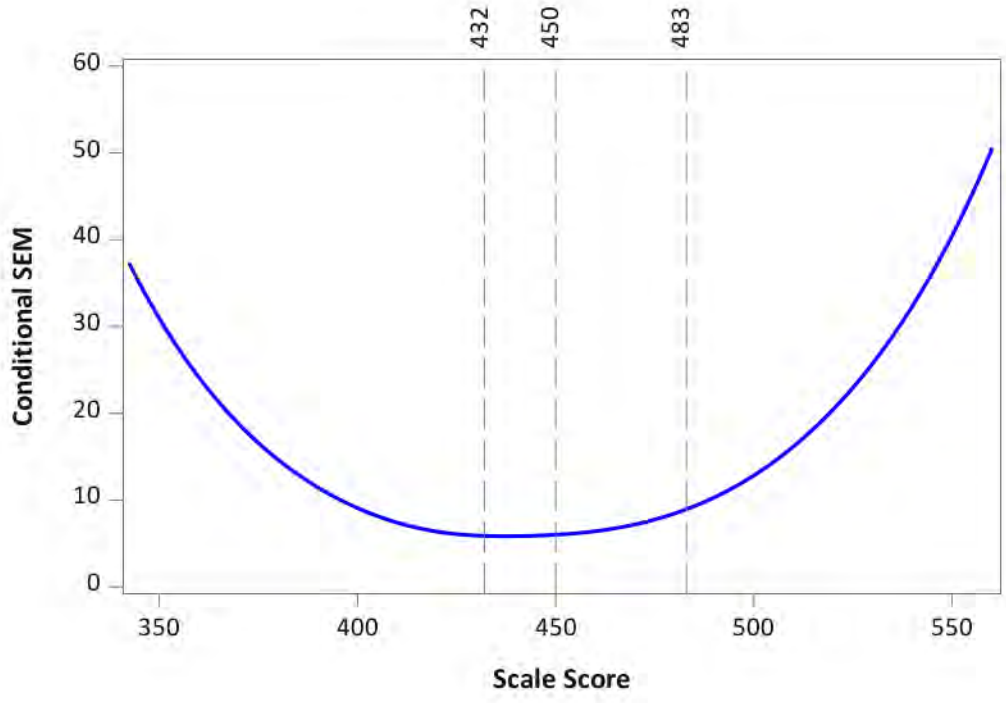


Figure 6.20. Mathematics Grade 6 TCC

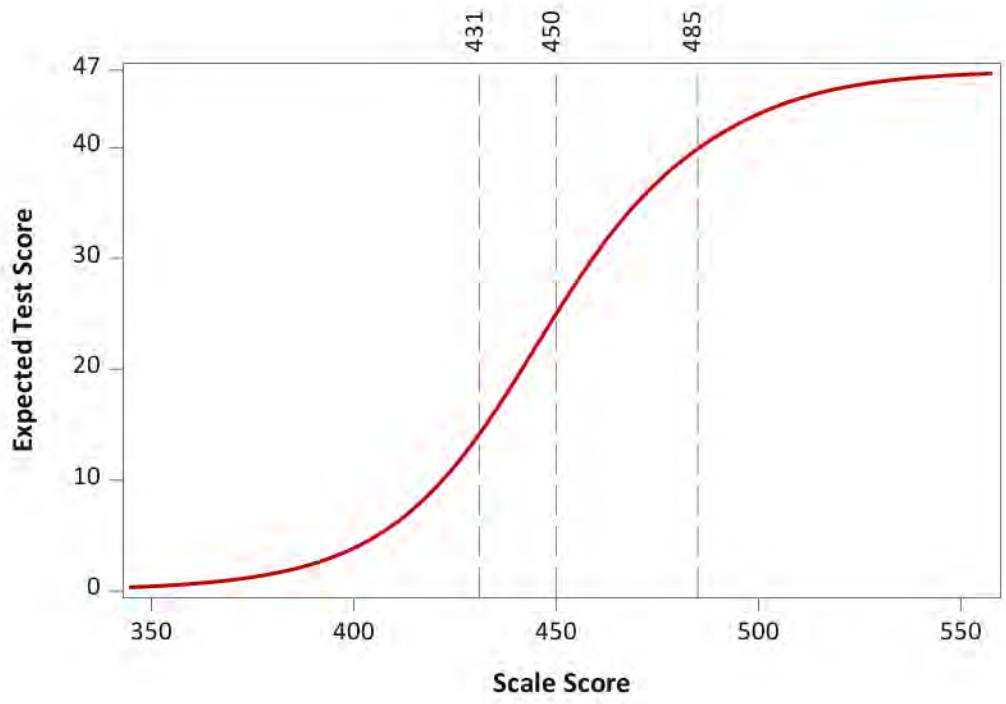


Figure 6.21. Mathematics Grade 6 CSEM Curve

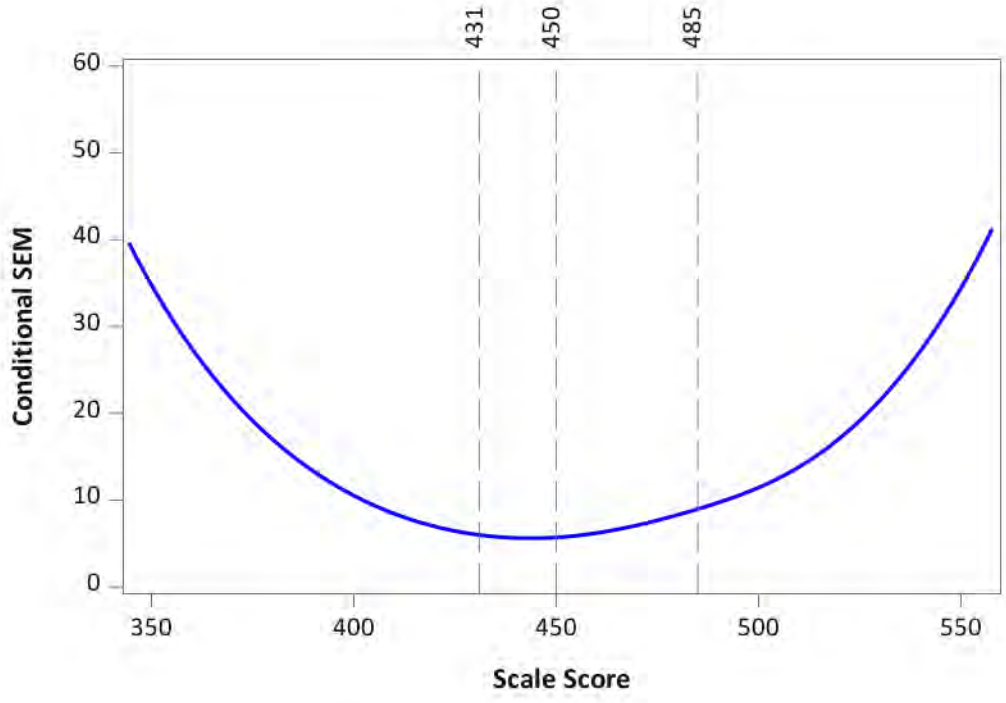


Figure 6.22. Mathematics Grade 7 TCC

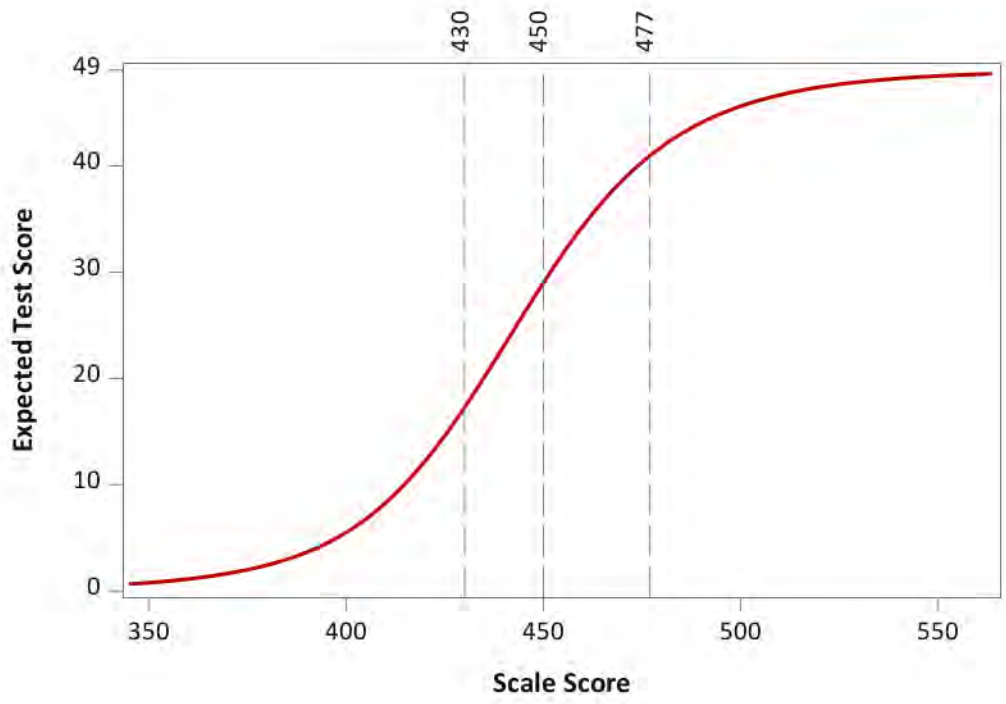


Figure 6.23. Mathematics Grade 7 CSEM Curve

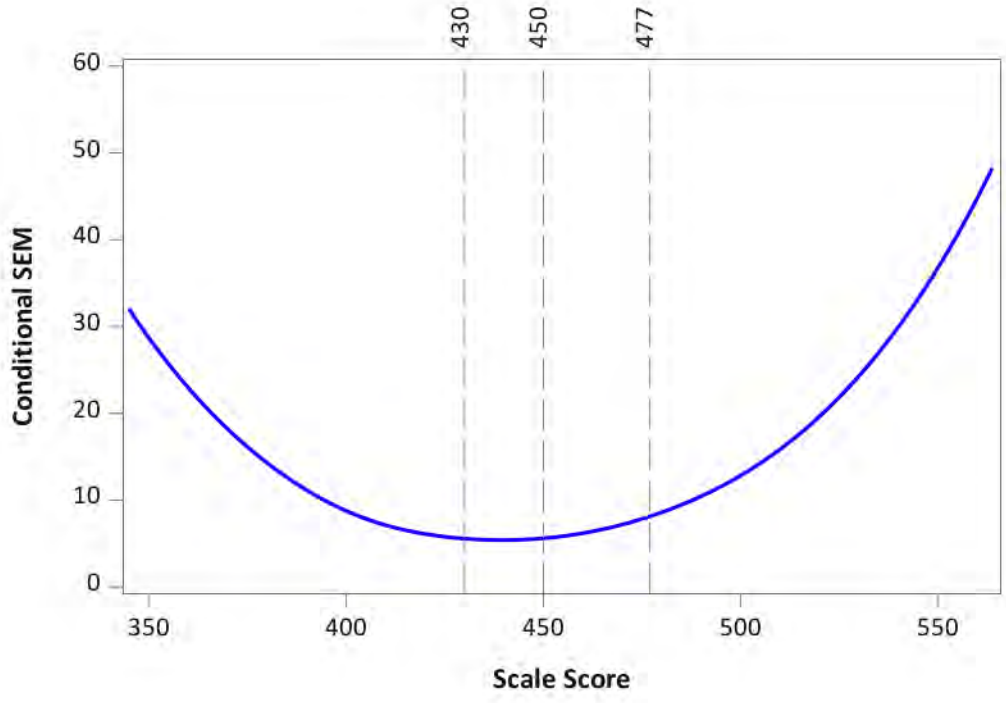


Figure 6.24. Mathematics Grade 8 TCC

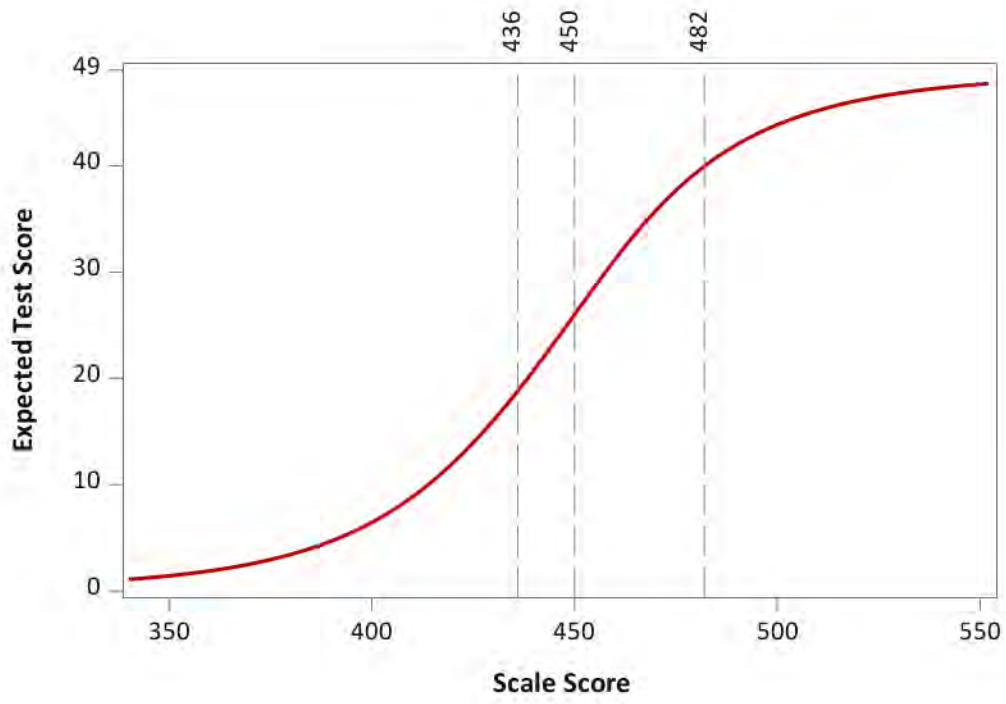
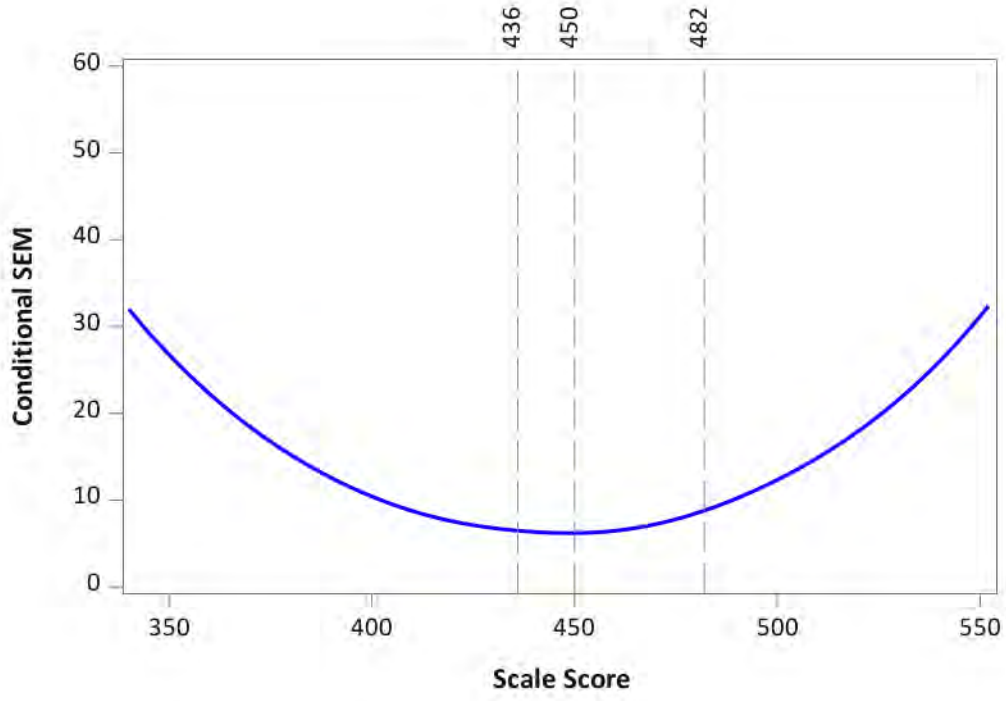


Figure 6.25. Mathematics Grade 8 CSEM Curve



Section 7: Reliability and Standard Error of Measurement

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

7.1. Test Reliability

Test reliability is directly related to score stability and standard error and, as such, is an essential element of fairness and validity. Test reliability can be directly measured with an alpha statistic, or the alpha statistic can be used to derive the SEM. For the Grades 3–8 ELA and Mathematics Tests, NWEA calculated two types of reliability statistics: Cronbach's alpha (Cronbach, 1951) and the 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 reflecting 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 the Feldt-Raju coefficient measures are appropriate for tests consisting of multiple item formats (MC and CR items).

7.1.1. *Test Statistics and Reliability for Total Test*

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

Grades 3–8 ELA reliability estimates (Cronbach's alpha and Feldt-Raju) ranged from 0.87 to 0.91. Grades 3–8 mathematics reliability estimates (Cronbach's alpha and Feldt-Raju) ranged from 0.92 to 0.94. The reliabilities were similar across grades and slightly higher for the mathematics tests than for the ELA tests. All reliabilities were at least 0.87 across all grades and both subjects, which is a good indication that the NYSTP Grades 3–8 ELA and Mathematics Tests are acceptably reliable.

Table 7.1. ELA Test Form Statistics

	Item-Level			Student-Level			
	P Value				Raw Score		
Grade	Mean	Min.	Max.	N-Count	Max.	Mean	SD
3	0.62	0.41	0.84	163,076	33	20.02	7.70
4	0.59	0.36	0.74	162,852	37	21.32	8.07
5	0.61	0.33	0.88	161,256	40	23.57	7.88
6	0.63	0.37	0.86	159,541	40	25.13	7.76

Grade	Item-Level			Student-Level			
	P Value				Raw Score		
	Mean	Min.	Max.	N-Count	Max.	Mean	SD
7	0.62	0.42	0.79	155,450	47	29.44	9.35
8	0.65	0.39	0.86	146,532	47	30.64	9.26

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	163,076	28	33	0.90	2.41	0.91	2.35
4	162,852	29	37	0.89	2.71	0.90	2.58
5	161,256	32	40	0.88	2.75	0.89	2.63
6	159,541	32	40	0.87	2.76	0.89	2.63
7	155,450	39	47	0.89	3.06	0.90	2.94
8	146,532	39	47	0.89	3.00	0.90	2.86

Table 7.3. Mathematics Test Form Statistics

Grade	Item-Level			Student-Level			
	P Value				Raw Score		
	Mean	Min.	Max.	N-Count	Max.	Mean	SD
3	0.65	0.38	0.92	159,439	38	23.37	9.30
4	0.60	0.22	0.92	159,198	44	25.09	10.71
5	0.59	0.26	0.87	157,029	44	24.13	11.29
6	0.54	0.13	0.75	152,136	47	23.88	12.22
7	0.62	0.36	0.89	145,907	49	28.59	12.28
8	0.51	0.27	0.73	93,235	49	23.31	11.48

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	159,439	32	38	0.92	2.70	0.92	2.56
4	159,198	37	44	0.92	3.00	0.93	2.82
5	157,029	37	44	0.93	2.95	0.94	2.77
6	152,136	39	47	0.93	3.13	0.94	2.98
7	145,907	41	49	0.93	3.16	0.94	2.94
8	93,235	41	49	0.92	3.24	0.93	3.09

7.1.2. *Reliability of MC Items*

In addition to overall test reliability, Cronbach's alpha and the Feldt-Raju coefficient were computed separately for MC and CR item sets. It is important to recognize that reliability is directly affected by test length; therefore, reliability estimates for tests by item type will always

be lower than reliability estimates for the overall test form. Table 7.5 and Table 7.6 present reliabilities for the subsets of MC items.

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

Grade	N-Count	#Items	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	163,076	23	0.87	1.97	0.87	1.96
4	162,852	23	0.83	2.08	0.83	2.08
5	161,256	26	0.82	2.19	0.82	2.19
6	159,541	26	0.81	2.22	0.81	2.21
7	155,450	33	0.85	2.56	0.85	2.55
8	146,532	33	0.85	2.44	0.85	2.44

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	159,439	24	0.89	1.90	0.89	1.88
4	159,198	28	0.88	2.17	0.89	2.15
5	157,029	28	0.90	2.11	0.90	2.10
6	152,136	29	0.90	2.27	0.91	2.27
7	145,907	31	0.89	2.26	0.90	2.25
8	93,235	31	0.86	2.47	0.87	2.46

7.1.3. *Reliability of CR Items*

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

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

Grade	N-Count	#Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	163,076	5	10	0.83	1.17	0.83	1.17
4	162,852	6	14	0.84	1.47	0.85	1.41
5	161,256	6	14	0.83	1.38	0.85	1.33
6	159,541	6	14	0.84	1.33	0.86	1.25
7	155,450	6	14	0.85	1.33	0.87	1.25
8	146,532	6	14	0.86	1.36	0.87	1.30

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

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	159,439	8	14	0.81	1.75	0.83	1.67
4	159,198	9	16	0.83	1.88	0.85	1.77
5	157,029	9	16	0.86	1.86	0.87	1.76
6	152,136	10	18	0.85	1.99	0.87	1.89
7	145,907	10	18	0.88	1.93	0.89	1.82
8	93,235	10	18	0.86	1.91	0.88	1.83

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

7.1.4. *Test Reliability for Subgroups*

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

As shown in Tables 7.9–7.14 and Tables 7.15–7.20 for ELA and mathematics, respectively, the estimated reliabilities for subgroups were close in magnitude to the test reliability estimates of the population. Except for the non-binary group, Cronbach's alpha reliability coefficients were all at least 0.76, and the Feldt-Raju reliability coefficients, which tend to be larger than the Cronbach's alpha estimates for the same group, were at least 0.78. These indicate a very good internal test 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	163,076	0.90	2.41	0.91	2.35
Gender	Female	81,468	0.90	2.38	0.91	2.32
	Male	81,602	0.90	2.43	0.91	2.37
	Non-Binary	6	0.61	2.32	0.73	1.93
Ethnicity	Asian	17,409	0.89	2.24	0.90	2.19
	African American	22,569	0.90	2.46	0.91	2.40
	Hispanic	44,064	0.90	2.47	0.90	2.41
	American Indian	1,160	0.90	2.44	0.90	2.38
	Multiracial	6,002	0.91	2.36	0.91	2.29
	Pacific Islander	310	0.90	2.35	0.91	2.29

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
NRC	White	71,350	0.90	2.38	0.90	2.32
	New York	47,579	0.91	2.36	0.91	2.30
	Big 4 Cities	6,106	0.90	2.53	0.91	2.45
	Urban/Suburban	12,853	0.89	2.49	0.89	2.45
	Rural	9,411	0.88	2.48	0.89	2.43
	Average Needs	44,170	0.89	2.40	0.89	2.36
	Low Needs	20,694	0.87	2.25	0.87	2.21
	Charter School	11,754	0.90	2.41	0.90	2.35
	Religious or Independent	10,509	0.91	2.45	0.92	2.38
SWD	All Codes	22,721	0.87	2.52	0.88	2.48
SUA	All Codes	19,947	0.86	2.52	0.87	2.48
ELL	ELL=Y	21,468	0.86	2.56	0.87	2.50
SWD/SUA	SWD & SUA codes	17,145	0.86	2.53	0.86	2.49
ELL/SUA	SUA & ELL codes	2,760	0.81	2.55	0.82	2.51

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	162,852	0.89	2.69	0.90	2.58
Gender	Female	80,811	0.88	2.68	0.89	2.57
	Male	82,035	0.89	2.69	0.90	2.58
	Non-Binary	6	0.93	2.31	0.95	2.03
Ethnicity	Asian	17,097	0.88	2.50	0.89	2.42
	African American	23,473	0.88	2.74	0.89	2.63
	Hispanic	44,484	0.88	2.73	0.89	2.63
	American Indian	1,183	0.89	2.71	0.90	2.59
	Multiracial	5,681	0.90	2.66	0.91	2.53
	Pacific Islander	361	0.89	2.64	0.90	2.54
	White	70,342	0.89	2.68	0.90	2.56
NRC	New York	48,769	0.90	2.63	0.91	2.51
	Big 4 Cities	6,292	0.88	2.80	0.90	2.64
	Urban/Suburban	12,220	0.87	2.76	0.87	2.66
	Rural	9,232	0.86	2.74	0.87	2.65
	Average Needs	42,381	0.87	2.69	0.88	2.60
	Low Needs	19,572	0.85	2.56	0.86	2.48
	Charter School	11,219	0.87	2.69	0.88	2.61
	Religious or Independent	13,167	0.90	2.77	0.91	2.59

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
SWD	All Codes	23,476	0.86	2.75	0.87	2.65
SUA	All Codes	22,223	0.85	2.74	0.86	2.65
ELL	ELL=Y	18,704	0.82	2.78	0.83	2.67
SWD/SUA	SWD & SUA codes	18,780	0.85	2.74	0.86	2.65
ELL/SUA	SUA & ELL codes	2,838	0.78	2.73	0.79	2.66

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	161,256	0.88	2.73	0.89	2.63
Gender	Female	79,887	0.87	2.71	0.88	2.61
	Male	81,360	0.88	2.74	0.89	2.64
	Non-Binary	9	0.24	2.82	0.27	2.76
Ethnicity	Asian	18,159	0.87	2.56	0.88	2.47
	African American	24,016	0.87	2.78	0.87	2.68
	Hispanic	43,998	0.86	2.77	0.87	2.67
	American Indian	1,180	0.87	2.75	0.88	2.64
	Multiracial	5,430	0.89	2.70	0.89	2.59
	Pacific Islander	340	0.89	2.73	0.89	2.63
	White	67,944	0.88	2.71	0.89	2.61
NRC	New York	50,758	0.88	2.70	0.89	2.59
	Big 4 Cities	6,121	0.87	2.84	0.88	2.72
	Urban/Suburban	12,496	0.86	2.76	0.87	2.68
	Rural	9,325	0.85	2.74	0.86	2.67
	Average Needs	41,812	0.86	2.70	0.87	2.63
	Low Needs	19,625	0.85	2.59	0.86	2.53
	Charter School	10,958	0.85	2.75	0.86	2.67
	Religious or Independent	10,161	0.91	2.80	0.92	2.64
SWD	All Codes	23,826	0.85	2.80	0.86	2.72
SUA	All Codes	23,224	0.85	2.79	0.85	2.72
ELL	ELL=Y	17,620	0.83	2.84	0.84	2.74
SWD/SUA	SWD & SUA codes	19,722	0.84	2.80	0.85	2.72
ELL/SUA	SUA & ELL codes	3,032	0.79	2.80	0.80	2.73

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	159,541	0.87	2.74	0.88	2.62
Gender	Female	78,699	0.86	2.69	0.87	2.59
	Male	80,814	0.88	2.76	0.89	2.64
	Non-Binary	28	0.71	2.59	0.74	2.49
Ethnicity	Asian	17,139	0.85	2.50	0.86	2.42
	African American	24,214	0.86	2.80	0.87	2.69
	Hispanic	44,113	0.86	2.80	0.87	2.69
	American Indian	1,138	0.86	2.81	0.88	2.68
	Multiracial	5,317	0.88	2.69	0.89	2.58
	Pacific Islander	367	0.85	2.67	0.86	2.58
	White	66,990	0.88	2.71	0.89	2.59
NRC	New York	49,124	0.87	2.71	0.88	2.59
	Big 4 Cities	6,008	0.87	2.88	0.89	2.73
	Urban/Suburban	12,027	0.87	2.79	0.88	2.71
	Rural	9,240	0.86	2.76	0.87	2.68
	Average Needs	40,128	0.86	2.71	0.87	2.62
	Low Needs	18,974	0.82	2.53	0.83	2.47
	Charter School	11,552	0.83	2.70	0.84	2.64
	Religious or Independent	12,488	0.91	2.86	0.92	2.64
SWD	All Codes	22,543	0.84	2.87	0.85	2.77
SUA	All Codes	22,452	0.85	2.87	0.86	2.76
ELL	ELL=Y	16,186	0.82	2.93	0.84	2.79
SWD/SUA	SWD & SUA codes	18,215	0.83	2.88	0.84	2.78
ELL/SUA	SUA & ELL codes	2,941	0.76	2.88	0.78	2.79

Table 7.13. ELA Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	155,450	0.89	3.03	0.90	2.93
Gender	Female	76,280	0.89	2.96	0.89	2.88
	Male	79,122	0.89	3.08	0.90	2.97
	Non-Binary	48	0.88	2.90	0.89	2.77
Ethnicity	Asian	16,780	0.88	2.75	0.88	2.69
	African American	24,632	0.88	3.08	0.89	2.98
	Hispanic	44,657	0.89	3.07	0.89	2.98

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
NRC	American Indian	1,113	0.89	3.06	0.90	2.95
	Multiracial	4,671	0.90	3.00	0.91	2.90
	Pacific Islander	317	0.89	2.99	0.89	2.88
	White	63,016	0.89	3.02	0.90	2.91
	New York	50,462	0.90	2.94	0.90	2.85
	Big 4 Cities	6,057	0.89	3.18	0.90	3.04
	Urban/Suburban	11,679	0.88	3.15	0.89	3.05
	Rural	8,923	0.88	3.09	0.88	3.01
	Average Needs	36,758	0.88	3.04	0.89	2.96
	Low Needs	18,505	0.86	2.87	0.86	2.81
	Charter School	12,130	0.87	2.93	0.87	2.88
	Religious or Independent	10,936	0.92	3.12	0.93	2.91
SWD	All Codes	22,117	0.86	3.17	0.87	3.07
SUA	All Codes	21,808	0.87	3.17	0.87	3.07
ELL	ELL=Y	13,496	0.81	3.21	0.82	3.09
SWD/SUA	SWD & SUA codes	18,001	0.85	3.17	0.86	3.08
ELL/SUA	SUA & ELL codes	2,518	0.76	3.16	0.78	3.07

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	146,532	0.89	2.97	0.90	2.85
Gender	Female	71,382	0.88	2.91	0.89	2.81
	Male	75,099	0.90	3.00	0.91	2.87
	Non-Binary	51	0.84	2.80	0.85	2.67
Ethnicity	Asian	16,560	0.88	2.66	0.88	2.58
	African American	24,767	0.89	3.02	0.89	2.91
	Hispanic	42,324	0.89	3.02	0.90	2.90
	American Indian	1,033	0.88	3.00	0.89	2.88
	Multiracial	4,110	0.90	2.94	0.91	2.83
	Pacific Islander	329	0.88	2.80	0.88	2.73
	White	57,182	0.89	2.96	0.90	2.84
NRC	New York	50,761	0.89	2.88	0.90	2.77
	Big 4 Cities	6,192	0.90	3.14	0.91	2.98
	Urban/Suburban	11,509	0.91	3.05	0.91	2.96
	Rural	8,651	0.88	3.05	0.89	2.95
	Average Needs	33,634	0.89	2.99	0.89	2.89

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
	Low Needs	16,544	0.86	2.80	0.87	2.71
	Charter School	11,280	0.85	2.86	0.86	2.81
	Religious or Independent	7,961	0.92	3.10	0.93	2.88
SWD	All Codes	21,015	0.87	3.16	0.88	3.05
SUA	All Codes	20,292	0.88	3.15	0.89	3.04
ELL	ELL=Y	12,247	0.85	3.18	0.86	3.06
SWD/SUA	SWD & SUA codes	16,928	0.87	3.16	0.88	3.05
ELL/SUA	SUA & ELL codes	2,166	0.81	3.15	0.82	3.07

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	159,439	0.92	2.66	0.92	2.54
Gender	Female	79,626	0.91	2.69	0.92	2.57
	Male	79,808	0.92	2.63	0.93	2.51
	Non-Binary	5	0.77	2.11	0.81	1.92
Ethnicity	Asian	17,296	0.91	2.39	0.92	2.26
	African American	22,416	0.92	2.71	0.92	2.60
	Hispanic	41,102	0.91	2.72	0.92	2.61
	American Indian	1,141	0.92	2.67	0.93	2.54
	Multiracial	5,934	0.92	2.63	0.93	2.51
	Pacific Islander	311	0.92	2.65	0.92	2.52
	White	71,030	0.91	2.65	0.92	2.55
NRC	New York	45,820	0.92	2.62	0.93	2.48
	Big 4 Cities	5,917	0.92	2.68	0.92	2.58
	Urban/Suburban	11,874	0.91	2.72	0.92	2.65
	Rural	9,441	0.90	2.71	0.91	2.62
	Average Needs	43,662	0.90	2.68	0.91	2.58
	Low Needs	20,606	0.89	2.48	0.90	2.38
	Charter School	11,616	0.91	2.63	0.92	2.51
	Religious or Independent	10,503	0.91	2.78	0.92	2.68
SWD	All Codes	21,354	0.91	2.71	0.91	2.63
SUA	All Codes	18,776	0.90	2.72	0.90	2.64
ELL	ELL=Y	19,259	0.90	2.75	0.90	2.68
SWD/SUA	SWD & SUA codes	16,037	0.89	2.72	0.90	2.64
ELL/SUA	SUA & ELL codes	2,532	0.88	2.70	0.88	2.64

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	159,198	0.92	2.95	0.93	2.80
Gender	Female	78,908	0.92	2.96	0.93	2.82
	Male	80,282	0.93	2.93	0.93	2.77
	Non-Binary	8	0.84	2.92	0.86	2.81
Ethnicity	Asian	16,973	0.92	2.74	0.93	2.56
	African American	23,242	0.92	2.96	0.93	2.83
	Hispanic	41,417	0.91	2.98	0.92	2.84
	American Indian	1,152	0.92	2.95	0.93	2.81
	Multiracial	5,626	0.93	2.93	0.94	2.75
	Pacific Islander	347	0.92	2.91	0.93	2.76
	White	70,228	0.92	2.95	0.92	2.80
NRC	New York	46,699	0.93	2.92	0.94	2.74
	Big 4 Cities	6,167	0.92	2.89	0.93	2.77
	Urban/Suburban	11,445	0.91	2.95	0.92	2.83
	Rural	9,194	0.91	2.95	0.91	2.84
	Average Needs	41,904	0.91	2.95	0.92	2.82
	Low Needs	19,696	0.90	2.84	0.91	2.69
	Charter School	11,293	0.91	2.95	0.92	2.80
	Religious or Independent	12,800	0.92	3.02	0.92	2.91
SWD	All Codes	21,975	0.91	2.89	0.92	2.79
SUA	All Codes	20,914	0.90	2.89	0.91	2.80
ELL	ELL=Y	16,625	0.89	2.93	0.89	2.86
SWD/SUA	SWD & SUA codes	17,455	0.90	2.86	0.90	2.78
ELL/SUA	SUA & ELL codes	2,702	0.88	2.83	0.88	2.77

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	157,029	0.93	2.92	0.94	2.75
Gender	Female	77,567	0.93	2.94	0.94	2.78
	Male	79,454	0.93	2.90	0.94	2.73
	Non-Binary	8	0.92	2.82	0.94	2.55
Ethnicity	Asian	17,957	0.93	2.67	0.94	2.47
	African American	23,593	0.92	2.91	0.93	2.79
	Hispanic	41,075	0.92	2.94	0.93	2.81

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
NRC	American Indian	1,155	0.93	2.93	0.93	2.77
	Multiracial	5,325	0.94	2.89	0.95	2.70
	Pacific Islander	329	0.93	2.91	0.94	2.73
	White	67,413	0.93	2.92	0.93	2.77
	New York	48,108	0.94	2.89	0.94	2.69
	Big 4 Cities	5,960	0.92	2.80	0.93	2.71
	Urban/Suburban	11,640	0.92	2.89	0.93	2.78
	Rural	9,114	0.92	2.92	0.92	2.80
	Average Needs	41,062	0.92	2.94	0.93	2.79
	Low Needs	19,508	0.91	2.81	0.92	2.64
	Charter School	11,357	0.92	2.93	0.93	2.78
	Religious or Independent	10,280	0.92	2.95	0.93	2.84
SWD	All Codes	21,846	0.91	2.84	0.92	2.75
SUA	All Codes	21,305	0.91	2.84	0.92	2.75
ELL	ELL=Y	15,674	0.89	2.86	0.90	2.79
SWD/SUA	SWD & SUA codes	17,754	0.90	2.82	0.91	2.74
ELL/SUA	SUA & ELL codes	2,762	0.88	2.79	0.89	2.72

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	152,136	0.93	3.10	0.94	2.97
Gender	Female	74,892	0.93	3.12	0.94	2.99
	Male	77,217	0.94	3.08	0.94	2.94
	Non-Binary	27	0.90	3.16	0.91	3.00
Ethnicity	Asian	16,446	0.94	2.80	0.94	2.68
	African American	23,975	0.93	3.11	0.93	2.98
	Hispanic	39,637	0.92	3.15	0.93	3.02
	American Indian	1,098	0.93	3.11	0.94	2.96
	Multiracial	5,067	0.94	3.08	0.95	2.94
	Pacific Islander	363	0.94	3.06	0.95	2.90
	White	65,308	0.93	3.11	0.93	2.99
NRC	New York	44,788	0.94	3.07	0.95	2.92
	Big 4 Cities	5,878	0.93	3.02	0.93	2.89
	Urban/Suburban	10,873	0.92	3.09	0.93	2.99
	Rural	8,909	0.92	3.14	0.92	3.03
	Average Needs	38,209	0.92	3.14	0.93	3.02

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
	Low Needs	18,621	0.92	2.98	0.92	2.88
	Charter School	12,387	0.93	3.08	0.94	2.95
	Religious or Independent	12,471	0.93	3.15	0.93	3.02
SWD	All Codes	20,109	0.90	2.99	0.91	2.90
SUA	All Codes	19,684	0.91	3.01	0.91	2.91
ELL	ELL=Y	13,098	0.88	3.02	0.89	2.94
SWD/SUA	SWD & SUA codes	15,909	0.89	2.96	0.89	2.88
ELL/SUA	SUA & ELL codes	2,259	0.85	2.93	0.86	2.86

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	145,907	0.93	3.11	0.94	2.91
Gender	Female	71,558	0.93	3.12	0.94	2.92
	Male	74,303	0.94	3.09	0.94	2.90
	Non-Binary	46	0.93	3.08	0.94	2.86
Ethnicity	Asian	15,832	0.93	2.69	0.94	2.51
	African American	23,626	0.93	3.17	0.93	3.00
	Hispanic	39,232	0.92	3.18	0.93	3.01
	American Indian	1,056	0.93	3.15	0.94	2.95
	Multiracial	4,426	0.94	3.07	0.95	2.85
	Pacific Islander	310	0.93	3.07	0.94	2.89
	White	61,201	0.93	3.07	0.94	2.90
NRC	New York	45,143	0.94	3.06	0.95	2.84
	Big 4 Cities	5,880	0.92	3.14	0.93	2.97
	Urban/Suburban	10,167	0.92	3.16	0.93	3.02
	Rural	8,487	0.92	3.18	0.93	3.02
	Average Needs	35,069	0.92	3.13	0.93	2.95
	Low Needs	17,933	0.92	2.89	0.93	2.74
	Charter School	12,024	0.93	3.09	0.93	2.91
	Religious or Independent	11,204	0.93	3.16	0.94	2.99
SWD	All Codes	18,958	0.91	3.12	0.92	2.98
SUA	All Codes	18,346	0.91	3.13	0.92	2.99
ELL	ELL=Y	10,373	0.89	3.13	0.90	3.04
SWD/SUA	SWD & SUA codes	14,863	0.90	3.10	0.91	2.98
ELL/SUA	SUA & ELL codes	1,757	0.85	3.01	0.85	2.94

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	93,235	0.92	3.23	0.93	3.08
Gender	Female	45,151	0.92	3.25	0.92	3.10
	Male	48,042	0.92	3.20	0.93	3.06
	Non-Binary	42	0.88	3.38	0.89	3.24
Ethnicity	Asian	7,404	0.93	3.06	0.94	2.89
	African American	17,042	0.92	3.17	0.93	3.04
	Hispanic	25,672	0.91	3.22	0.92	3.09
	American Indian	640	0.92	3.23	0.92	3.09
	Multiracial	2,608	0.92	3.22	0.92	3.08
	Pacific Islander	203	0.94	3.19	0.94	2.99
	White	39,496	0.91	3.25	0.92	3.12
NRC	New York	27,221	0.93	3.21	0.94	3.04
	Big 4 Cities	4,907	0.91	2.99	0.91	2.89
	Urban/Suburban	7,349	0.89	3.12	0.90	3.05
	Rural	6,968	0.89	3.23	0.90	3.12
	Average Needs	22,948	0.89	3.27	0.90	3.16
	Low Needs	8,295	0.90	3.21	0.91	3.08
	Charter School	7,820	0.93	3.17	0.94	3.00
	Religious or Independent	7,727	0.93	3.23	0.94	3.07
SWD	All Codes	14,964	0.89	3.07	0.89	2.99
SUA	All Codes	14,184	0.89	3.09	0.89	3.01
ELL	ELL=Y	8,226	0.89	3.10	0.90	3.01
SWD/SUA	SWD & SUA codes	11,785	0.88	3.05	0.88	2.98
ELL/SUA	SUA & ELL codes	1,349	0.81	2.93	0.82	2.89

7.2. Standard Error of Measurement (SEM)

Table 7.2 and Table 7.4 presented the SEMs, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, for ELA and mathematics, respectively. The SEMs ranged from 2.35 to 3.24 across subjects, grades, and the two methods of estimation, which is reasonable and small. The SEMs are directly related to reliability: the higher the reliability, the lower the standard error. As discussed, the reliability of these tests is relatively high, so it was expected that the SEMs would be very low.

The SEMs for the subpopulations, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, are presented in Tables 7.9–7.14 and Tables 7.15–7.20. The SEMs associated with all reliability estimates for all subjects, grades, methods of estimation, and subpopulations, except for the non-binary group, ranged from 2.38 to 3.27, which is acceptably

close to those for the entire population. This narrow range indicates that across the Grades 3–8 ELA and Mathematics Tests, all students' test scores are reasonably reliable with minimal error.

7.3. Performance Level Classification Consistency and Accuracy

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 & Lewis, 1995).

In conjunction with measures of internal consistency, classification consistency is an important type of reliability and is particularly relevant to high-stakes 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 performance is near the proficiency cut score. For example, consider the cut score delineating Levels 2 and 3, or simply the "Level 3 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. Students whose true scores are close to the cut score are a more serious concern. These students' true scores will likely lie within the SEM of the cut score. For this reason, the measurement error at the cut scores should be considered when evaluating the classification consistency of a test. Furthermore, the number of students near the cut scores should also be considered when evaluating classification consistency, as these numbers show the number of students who are at risk of being misclassified.

Scoring tables with SEMs are located in IRT Calibration, and student scale score frequency distributions are located in Appendix P: Raw-Score-to-Scale Score and Scale Score Frequency Tables. Classification consistency and accuracy were estimated using the IRT procedure suggested by Lee et al. (2002) and Wang et al. (2000). Appendix O: Derivation and Estimation of Classification Consistency and Accuracy includes a description of the calculations and procedure based on the paper by Lee et al. (2002).

7.3.1. *Consistency*

The results for classifying students into four performance levels are separated from the results based solely on the Level 3 cut. Table 7.21 and Table 7.22 include case counts ("N-Count"), classification consistency ("Agreement"), classification inconsistency ("Inconsistency"), and Cohen's kappa ("Kappa"). Consistency indicates the rate at which 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, 63–67% of students were estimated to be classified consistently into one of the four performance categories following a hypothetical second administration. Kappa—which corrects for chance agreement—ranged from 0.50 to 0.55. These are between "moderate" and "substantial" agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For mathematics, 73–76% of students were estimated to be classified consistently into one of the four performance categories, and kappa ranged from 0.61 to 0.66. 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. For example, by random chance, students testing twice may be classified first as a Level 3 and second as a Level 4. This is expected to occur more often for students scoring around the selected cut score and less often for students closer to the middle of the performance level (i.e., close to the mid-point of two adjacent cut scores).

Table 7.21. Decision Consistency (All Cuts)

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	163,076	67%	33%	0.55
4	162,852	67%	33%	0.55
5	161,256	66%	34%	0.53
6	159,541	63%	37%	0.5
7	155,450	65%	35%	0.53
8	146,532	66%	34%	0.54
Mathematics				
3	159,439	73%	27%	0.61
4	159,198	74%	26%	0.63
5	157,029	75%	25%	0.65
6	152,136	76%	24%	0.66
7	145,907	75%	25%	0.66
8	93,235	73%	27%	0.62

Table 7.22 depicts the ELA and mathematics consistency study results based on two performance levels (NYS Level 2 and NYS Level 3) as defined by the Level 3 cut. For ELA, 85–87% of the classifications of individual students were estimated to remain stable with a second administration. Kappa coefficients for ELA classification consistency ranged from 0.69 to 0.73. These are considered "substantial" agreement per Landis and Koch's (1977) rules of thumb for kappa. For mathematics, 89–91% of the classifications were estimated consistently, and kappa coefficients ranged from 0.77 to 0.81. These statistics indicate at least "substantial" agreement (where kappa > 0.60) for all tests and "almost perfect" agreement (where kappa > 0.80) for a few tests per Landis and Koch's (1977) rules of thumb for kappa.

Table 7.22. Decision Consistency (Level 3 Cut)

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	163,076	87%	13%	0.73
4	162,852	86%	14%	0.72
5	161,256	85%	15%	0.7
6	159,541	85%	15%	0.69
7	155,450	86%	14%	0.72
8	146,532	86%	14%	0.72
Mathematics				
3	159,439	89%	11%	0.77
4	159,198	90%	10%	0.79
5	157,029	90%	10%	0.8
6	152,136	90%	10%	0.81
7	145,907	91%	9%	0.81
8	93,235	89%	11%	0.78

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 3 cut score. By definition, accuracy associated with the Level 3 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 their underlying proficiency 71% to 75% of the time across all performance levels and 88% to 90% of the time in regard to the Level 3 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 their true proficiency 80% to 83% of the time across all performance levels and 92% to 93% of the time in regard to the Level 3 cut score.

Table 7.23. Decision Agreement (Accuracy) Estimates

		Accuracy	
Grade	N-Count	All Cuts	Level 3 Cut
ELA			
3	163,076	75%	90%
4	162,852	75%	90%
5	161,256	74%	88%
6	159,541	71%	88%
7	155,450	75%	90%
8	146,532	75%	90%
Mathematics			
3	159,439	80%	92%

		Accuracy	
Grade	N-Count	All Cuts	Level 3 Cut
4	159,198	81%	93%
5	157,029	82%	93%
6	152,136	83%	93%
7	145,907	81%	93%
8	93,235	80%	93%

Section 8: Standard Setting

Standard setting is the formal process by which panels of educators and subject-matter experts recommend performance standards. These performance standards include cut points that divide the test scale into performance levels (i.e., Level 1, Level 2, Level 3, and Level 4). Students are placed into one of these performance levels based on their test results.

The adoption of the Next Generation Learning Standards in 2017 included the creation of new performance level descriptions for each standard in both content areas and all grades. These new guiding documents informed the subsequent implementation for the Spring 2023 operational assessments. These changes compelled the establishment of new cut points for each of the assessments in ELA and mathematics Grades 3 through 8.

Standard setting was conducted in summer 2023 to set performance standards for the new assessments. This section summarizes the background, methodology, and process of standard setting.

8.1. Goals of Standard Setting

The goals of standard setting were as follows:

- Provide vertically articulated performance standards for the assessments in ELA and mathematics and indicate the degree to which students have met the standards for their grades
- Recommend rigorous and attainable performance standards
- Incorporate existing and future policy considerations relevant to New York State's educational system into the established performance standards

8.2. Participants

The standard-setting panelists were comprised of 65 qualified New York State educators who had knowledge of the current NYSED standards and were from diverse backgrounds regarding demographic characteristics and geographic locations within the State.

8.3. Methodology

The Bookmark method was used in the standard-setting process for setting the cut scores. This method requires panelists to work through a test booklet in which the items have been re-ordered from easiest to hardest based on student performance data. Panelists are asked to place a bookmark in the ordered booklet to demarcate each performance standard. The bookmarks are placed with the assumption that borderline students will perform successfully at a given performance level with a probability of at least 67%.

The cut scores are derived by taking the median of the corresponding bookmarks placed for the various performance standards across panelists.

8.4. Standard-Setting Process

The following steps were used as the standard-setting process:

1. Standards review committees are convened.
2. Panelists review the current performance level descriptors (PLDs) and develop threshold PLDs.
3. Panelists review and recommend cut score points following the Bookmark standard-setting methodology (3 rounds of judgements).
4. Vertical articulation is conducted.

8.5. Results

The cut-score recommendations from Round 3 were affirmed during vertical articulation and then approved by the Commissioner of Education. The final raw score cuts are shown in Tables 8.1 and 8.2 for ELA and mathematics, respectively, along with the corresponding scale score cuts.

Table 8.1. ELA Performance-Level Cut Scores

Grade	Raw Score Cuts			Scale Score Cuts		
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
3	16	23	29	432	450	474
4	15	23	30	431	450	471
5	19	26	33	432	450	474
6	21	28	33	431	450	470
7	24	32	39	433	450	472
8	22	32	39	430	450	472

Table 8.2. Mathematics Performance-Level Cut Scores

Grade	Raw Score Cuts			Scale Score Cuts		
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
3	13	24	34	424	450	487
4	16	25	38	431	450	486
5	16	25	38	432	450	483
6	14	25	40	431	450	485
7	17	29	41	430	450	477
8	19	26	40	436	450	482

Appendix Q: Standard Setting Technical Report presents the full 2023 standard-setting technical report that describes the general process, the composition of the committees, ratings from the various rounds, evaluation forms, and other materials.

Section 9: Summary of Operational Test Results

This section summarizes the distribution of scale-score results on the NYSTP 2023 Grades 3–8 ELA and Mathematics Tests. These include the scale score means, standard deviations, and performance level distributions for each grade's population and specific subgroups. Gender, ethnicity, Needs/Resource Capacity (NRC) category, English Language Learner (ELL) status, students with disabilities (SWDs), and students using test accommodations (SUAs) 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 ELLs who use one or more ELL-related accommodation, and the SWD/SUA subgroup is defined as SWDs who use one or more disability-related accommodation. For the mathematics analyses, the test translation language is also indicated (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 P: Raw-Score-to-Scale Score and Scale Score Frequency Tables.

9.1. Scale Score Distribution Summary

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. Caution is recommended when interpreting the statistics for subgroups with small number counts that are included in the scale-score summaries.

9.1.1. *ELA Scale Score and Subscore Distributions*

Table 9.1 shows some key statistics characterizing the distribution of ELA scale scores, while Table 9.2 summarizes the ELA subscores derived from the test for each grade. Tables 9.3–9.8 break down the scale scores by selected subgroups. Some general observations from these tables include:

- Females outperformed Males.
- Asian and White students outperformed their peers from other reported ethnic groups.
- Students from Low Needs (as identified by NRC) districts outperformed students from other districts (New York City, Big 4 Cities, Urban/Suburban, Rural, Average Needs, and Charter).
- ELLs, SWD, SUA, and SWD/SUA students tended to under-perform against the State population (All Students).

This pattern of achievement was consistent across all grades.

Table 9.1. ELA Scale Score Distribution Summary

Grade	N-Count	Scale Score	
		Mean	SD
3	166,155	444.41	23.06
4	166,173	447.39	23.06
5	165,259	445.53	23.07
6	165,051	444.95	22.93
7	160,467	447.56	23.05
8	152,212	450.78	23.00

Table 9.2. ELA Subscore Summary

Grade	Subscore Category	N-Count	Subscore		
			Max	Mean	SD
3	Reading	166,155	23	14.70	5.46
	Writing	166,155	10	5.26	2.83
4	Reading	166,173	23	13.83	5.08
	Writing	166,173	14	7.45	3.63
5	Reading	165,259	26	15.80	5.18
	Writing	165,259	14	7.74	3.40
6	Reading	165,051	26	15.86	5.02
	Writing	165,051	14	9.27	3.41
7	Reading	160,467	33	19.65	6.59
	Writing	160,467	14	9.78	3.50
8	Reading	152,212	33	21.28	6.25
	Writing	152,212	14	9.43	3.63

9.1.1.1. ELA Grade 3

Table 9.3 presents the Grade 3 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 444.41 with a standard deviation of 23.06. Female students tended to outperform Male students by around five scale-score points. Asian, Multiracial, Pacific Islander, and White students' scale-score means exceeded the State mean scale score, as did those of students from New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (455.48). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score of 432.51—about 12 scale-score points below the population mean. The students with disabilities (SWD), students tested under accommodations (SUA), and ELL subgroups scored about 15–17 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed, scoring more than 21 scale-score points below the State mean.

Table 9.3. ELA Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	166,155	444.41	23.06
Gender	Female	82,592	446.99	22.98
	Male	83,557	441.86	22.85
	Non-Binary	6	459.67	12.88
Ethnicity	Asian	17,602	455.48	22.20
	African American	23,448	440.16	22.94
	Hispanic	45,285	439.05	22.40
	American Indian	1,178	442.75	22.43
	Multiracial	6,102	446.85	23.49
	Pacific Islander	316	447.23	23.10
	White	72,002	446.31	22.40
NRC	New York	47,840	447.43	23.97
	Big 4 Cities	6,128	432.51	23.48
	Urban/Suburban	12,893	435.02	21.30
	Rural	9,488	437.20	20.56
	Average Needs	44,171	443.91	21.01
	Low Needs	20,742	454.92	19.88
	Charter	12,241	446.75	22.84
	Religious or Independent	10,619	437.91	24.23
SWD	All Codes	25,265	429.51	19.92
SUA	All Codes	14,450	428.47	18.70
ELL	ELL=Y	21,857	426.71	19.05
ELL/SUA	SUA & ELL codes	1,353	422.82	16.17
SWD/SUA	SWD & SUA codes	11,822	426.61	17.77

9.1.1.2. ELA Grade 4

Table 9.4 contains Grade 4 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 447.39 with a standard deviation of 23.06. Female students tended to outperform Male students by around four scale-score points. Asian, Multiracial, Pacific Islander, and White students' scale-score means exceeded the State mean scale score, as did those of students from New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (460.14). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—about 13 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored about 16–21 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed, scoring about 22 scale-score points below the State mean.

Table 9.4. ELA Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	166,173	447.39	23.06
Gender	Female	81,980	449.64	22.80
	Male	84,187	445.20	23.11
	Non-Binary	6	451.67	24.39
Ethnicity	Asian	17,287	460.14	22.34
	African American	24,414	442.62	22.26
	Hispanic	45,849	442.59	21.83
	American Indian	1,207	445.44	23.46
	Multiracial	5,773	449.94	24.07
	Pacific Islander	366	450.16	22.99
	White	71,034	448.88	22.76
NRC	New York	49,017	451.16	23.99
	Big 4 Cities	6,288	434.38	23.22
	Urban/Suburban	12,292	438.23	20.65
	Rural	9,274	440.18	20.21
	Average Needs	42,317	446.97	20.84
	Low Needs	19,584	457.71	19.93
	Charter	11,706	449.31	21.30
	Religious or Independent	13,308	441.16	24.95
SWD	All Codes	26,093	431.46	20.30
SUA	All Codes	15,880	430.37	19.23
ELL	ELL=Y	19,170	426.38	17.87
ELL/SUA	SUA & ELL codes	1,417	424.63	16.12
SWD/SUA	SWD & SUA codes	12,584	427.93	18.40

9.1.1.3. ELA Grade 5

Table 9.5 provides the Grade 5 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 445.53 with a standard deviation of 23.07. Female students tended to outperform Male students by around five scale-score points. Asian, Multiracial, Pacific Islander, and White students' scale-score means exceeded the State mean scale score, as did those of students enrolled in New York City, Low Needs districts, and Charter schools. Across all ethnic groups, Asian students earned the highest mean score (459.07). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—about 12 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored about 16–21 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed, scoring about 23 scale-score points below the State mean.

Table 9.5. ELA Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	165,259	445.53	23.07
Gender	Female	81,369	448.25	22.56
	Male	83,881	442.89	23.27
	Non-Binary	9	448.67	8.92
Ethnicity	Asian	18,360	459.07	22.38
	African American	25,173	440.20	21.60
	Hispanic	45,712	441.43	21.51
	American Indian	1,206	443.89	22.02
	Multiracial	5,528	447.41	23.88
	Pacific Islander	351	447.24	24.02
	White	68,729	446.50	23.17
NRC	New York	50,926	449.95	23.53
	Big 4 Cities	6,132	433.14	22.29
	Urban/Suburban	12,491	436.81	20.60
	Rural	9,301	438.25	19.93
	Average Needs	41,848	444.78	20.64
	Low Needs	19,638	456.08	20.58
	Charter	11,875	446.31	21.01
	Religious or Independent	10,310	435.59	26.72
SWD	All Codes	26,863	429.01	19.86
SUA	All Codes	16,090	427.69	19.20
ELL	ELL=Y	18,156	424.55	18.52
ELL/SUA	SUA & ELL codes	1,504	422.19	15.72
SWD/SUA	SWD & SUA codes	12,819	425.00	18.00

9.1.1.4. ELA Grade 6

Table 9.6 contains Grade 6 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 444.95 with a standard deviation of 22.93. Female students tended to outperform Male students by around six scale-score points. Asian, Multiracial, Pacific Islander, and White students' scale-score means exceeded the State mean scale score, as did those of students enrolled in New York City, Average Needs and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (458.15). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—about 12 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored about 17–24 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed, scoring about 26 scale-score points below the State mean.

Table 9.6. ELA Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	165,051	444.95	22.93
Gender	Female	80,890	448.08	22.16
	Male	84,133	441.94	23.26
	Non-Binary	28	454.32	16.35
Ethnicity	Asian	17,433	458.15	21.27
	African American	26,068	440.15	21.55
	Hispanic	46,408	440.08	21.55
	American Indian	1,184	440.96	22.25
	Multiracial	5,369	446.82	23.43
	Pacific Islander	385	448.73	21.40
	White	67,927	446.70	23.14
NRC	New York	49,364	447.81	23.08
	Big 4 Cities	6,064	432.63	22.67
	Urban/Suburban	12,027	436.57	21.11
	Rural	9,200	439.60	20.76
	Average Needs	40,015	445.22	21.36
	Low Needs	18,975	455.55	19.42
	Charter	13,254	446.16	19.96
	Religious or Independent	12,815	438.52	27.22
SWD	All Codes	26,146	426.81	19.69
SUA	All Codes	16,486	427.06	20.06
ELL	ELL=Y	16,906	420.47	18.34
ELL/SUA	SUA & ELL codes	1,596	418.89	15.20
SWD/SUA	SWD & SUA codes	12,560	423.26	18.26

9.1.1.5. ELA Grade 7

Table 9.7 presents the Grade 7 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 447.56 with a standard deviation of 23.05. Female students tended to outperform Male students by around seven scale-score points. Asian, Multiracial, Pacific Islander, and White students' scale-score means exceeded the State mean scale score, as did those of students from New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (461.15). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—about 11 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored about 17–26 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed, scoring about 28 scale-score points below the State mean.

Table 9.7. ELA Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	160,467	447.56	23.05
Gender	Female	78,213	451.43	22.49
	Male	82,205	443.87	22.97
	Non-Binary	49	457.59	22.36
Ethnicity	Asian	17,027	461.15	21.96
	African American	26,124	443.27	21.73
	Hispanic	46,560	442.96	21.85
	American Indian	1,141	445.25	22.36
	Multiracial	4,748	449.38	23.85
	Pacific Islander	329	450.35	22.48
	White	64,246	449.00	23.03
NRC	New York	50,600	451.15	23.36
	Big 4 Cities	6,086	435.97	22.27
	Urban/Suburban	11,668	437.28	21.21
	Rural	8,871	440.55	20.37
	Average Needs	36,918	446.35	21.26
	Low Needs	18,594	456.96	20.49
	Charter	12,886	451.08	20.35
	Religious or Independent	11,234	443.97	26.67
SWD	All Codes	25,660	430.30	19.08
SUA	All Codes	14,625	429.25	19.22
ELL	ELL=Y	14,042	421.25	15.88
ELL/SUA	SUA & ELL codes	1,075	419.03	12.67
SWD/SUA	SWD & SUA codes	11,090	425.63	17.47

9.1.1.6. ELA Grade 8

Table 9.8 presents the Grade 8 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 450.78 with a standard deviation of 23.00. Female students tended to outperform Male students by six scale-score points. Asian, Multiracial, Pacific Islander, and White students' scale-score means exceeded the State mean scale score, as did those of students enrolled in New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (464.91). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—about 11 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored about 17–26 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed, scoring about 28 scale-score points below the State mean.

Table 9.8. ELA Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	152,212	450.78	23.00
Gender	Female	73,586	454.13	21.88
	Male	78,571	447.64	23.58
	Non-Binary	55	461.95	19.05
Ethnicity	Asian	16,824	464.91	21.30
	African American	26,346	446.76	21.61
	Hispanic	44,283	446.84	22.27
	American Indian	1,067	449.39	22.03
	Multiracial	4,275	452.11	23.32
	Pacific Islander	344	456.97	21.52
	White	58,807	451.48	22.93
NRC	New York	50,885	455.14	22.79
	Big 4 Cities	6,221	439.05	22.99
	Urban/Suburban	11,528	441.55	21.78
	Rural	8,594	443.76	20.90
	Average Needs	34,288	448.72	21.66
	Low Needs	16,627	459.96	20.34
	Charter	12,137	454.49	19.72
	Religious or Independent	8,129	445.66	26.47
SWD	All Codes	24,687	433.50	19.91
SUA	All Codes	14,449	432.20	20.31
ELL	ELL=Y	12,771	424.21	17.73
ELL/SUA	SUA & ELL codes	1,112	422.17	15.16
SWD/SUA	SWD & SUA codes	11,372	429.29	19.07

9.1.2. *Mathematics Scale Score and Subscore Distributions*

Table 9.9 shows some key statistics characterizing the distribution of mathematics scale scores, while Table 9.10 summarizes the mathematics subscores derived from the test for each grade. Tables 9.11–9.16 break down the scale scores by selected subgroups. Some general observations from the mathematics data include:

- Female and Male students performed fairly consistently.
- Asian students scored considerably higher than other reported ethnic groups.
- Students from Low Needs districts (as identified by the NRC code) outperformed students from High Needs (New York City, Big 4 Cities, Urban/Suburban, and Rural) and Average Needs districts.
- ELLs, SWDs, and/or SUAs tended to under-perform against the State population (All Students).

- Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

This pattern of achievement was fairly consistent across all grades.

Table 9.9. Mathematics Scale Score Distribution Summary

Grade	N-Count	Scale Score	
		Mean	SD
3	169,444	451.61	27.06
4	169,293	451.69	27.39
5	167,238	449.7	27.1
6	164,792	449.8	27.02
7	158,339	452.57	27.17
8	102,560	444.63	26.97

Table 9.10. Mathematics Subscore Summary

Grade	Subscore Category	N-Count	Subscore		
			Max	Mean	SD
3	Operations and Algebraic Thinking	169,444	12	7.76	3.16
	Number and Operations - Fractions	169,444	8	4.72	2.27
	Measurement and Data	169,444	11	7.13	2.75
4	Operations and Algebraic Thinking	169,293	9	5.26	2.48
	Number and Operations in Base Ten	169,293	10	6.34	2.74
	Number and Operations - Fractions	169,293	11	6.05	3.33
5	Number and Operations in Base Ten	167,238	13	7.54	3.79
	Number and Operations - Fractions	167,238	16	7.67	4.34
	Measurement and Data	167,238	13	7.64	3.55
6	Ratios and Proportional Relationships	164,792	12	6.83	3.67
	The Number System	164,792	9	4.99	2.57
	Expressions and Equations	164,792	18	8.30	4.75
7	Ratios and Proportional Relationships	158,339	14	8.02	3.90
	The Number System	158,339	10	7.05	2.64
	Expressions and Equations	158,339	15	8.23	4.01
8	Expressions and Equations	102,560	15	7.29	3.84
	Functions	102,560	12	5.45	3.31
	Geometry	102,560	17	7.76	4.33

9.1.2.1. Mathematics Grade 3

Table 9.11 presents the Grade 3 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 451.61 with a standard deviation of 27.06. Male students outperformed Female students. Asian, Multiracial, Pacific Islander, and White

students' scale-score means exceeded the State mean scale score, as did those of students enrolled in New York City, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (468.30). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—17 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored 16–20 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 23 scale-score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 9.11. Mathematics Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	169,444	451.61	27.06
Gender	Female	83,909	450.82	26.52
	Male	85,529	452.39	27.56
	Non-Binary	6	480.17	16.92
Ethnicity	Asian	18,176	468.30	25.29
	African American	23,497	444.60	26.74
	Hispanic	47,307	444.77	25.76
	American Indian	1174	450.16	26.91
	Multiracial	6,078	453.61	28.01
	Pacific Islander	324	454.67	26.88
	White	72,639	454.06	26.05
NRC	New York	50,220	454.58	27.81
	Big 4 Cities	6,222	434.20	25.97
	Urban/Suburban	13,160	440.46	24.70
	Rural	9,478	443.74	24.75
	Average Needs	44,307	451.24	25.09
	Low Needs	20,892	465.51	23.47
	Charter	12,252	455.38	25.83
	Religious or Independent	10,844	443.07	26.49
SWD	All Codes	25,496	434.55	25.22
SUA	All Codes	14,685	431.07	23.52
ELL	ELL=Y	25,239	435.51	23.53
ELL/SUA	SUA & ELL codes	1643	427.72	21.18
SWD/SUA	SWD & SUA codes	12,001	428.77	22.66
ELL Test Language	Arabic	144	436.12	24.64
	Bengali	84	436.45	23.41
	Chinese (Traditional)	35	458.43	24.59
	Chinese (Simplified)	347	464.86	23.22
	Haitian-Creole	50	432.00	21.85

Demographic Category		N-Count	Scale Score	
			Mean	SD
	Korean	43	463.77	25.55
	Russian	388	445.81	23.48
	Spanish	4,744	430.56	21.10
	All Translations	5,835	434.26	23.48

9.1.2.2. Mathematics Grade 4

Table 9.12 presents the Grade 4 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 451.69 with a standard deviation of 27.39. Male students outperformed Female students. Asian, Multiracial, Pacific Islander, and White students' scale-score means exceeded the State mean scale score, as did those of students enrolled in New York City, Average Needs and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (469.83). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—18 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored 18–20 scale-score points below the mean scale score for the population. Students tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 23 scale-score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 9.12. Mathematics Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	169,293	451.69	27.39
Gender	Female	83,150	450.24	26.64
	Male	86,135	453.08	28.03
	Non-Binary	8	448.38	16.59
Ethnicity	Asian	17,845	469.83	26.11
	African American	24,366	443.39	26.20
	Hispanic	47,625	443.95	25.65
	American Indian	1217	448.80	27.02
	Multiracial	5,756	454.51	28.40
	Pacific Islander	367	453.87	26.89
	White	71,865	455.01	26.23
NRC	New York	51,179	453.49	28.74
	Big 4 Cities	6,406	433.68	26.38
	Urban/Suburban	12,558	440.79	25.01
	Rural	9,249	445.81	23.98
	Average Needs	42,581	452.92	24.71
	Low Needs	19,921	466.86	23.72
	Charter	11,776	454.63	25.74
	Religious or Independent	13,236	442.97	26.51

Demographic Category		N-Count	Scale Score	
			Mean	SD
SWD	All Codes	26,275	433.24	24.67
SUA	All Codes	16,437	431.88	23.70
ELL	ELL=Y	22,429	431.10	21.74
ELL/SUA	SUA & ELL codes	1806	428.17	20.49
SWD/SUA	SWD & SUA codes	13,039	428.67	22.57
ELL Test Language	Arabic	137	433.82	22.38
	Bengali	96	441.68	26.02
	Chinese (Traditional)	42	462.21	24.77
	Chinese (Simplified)	329	465.71	26.69
	Haitian-Creole	45	424.31	18.51
	Korean	28	467.54	25.87
	Russian	364	447.40	23.45
	Spanish	4,551	425.87	19.81
	All Translations	5,592	430.55	23.60

9.1.2.3. Mathematics Grade 5

Table 9.13 presents the Grade 5 n-counts and scale-score statistics for key demographic subgroups. The population scale-score mean was 449.70 with a standard deviation of 27.10. Male students outperformed Female students. 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 (469.67). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—19 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored 18–20 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 26 scale-score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 9.13. Mathematics Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	167,238	449.70	27.10
Gender	Female	81,848	448.62	26.33
	Male	85,382	450.74	27.78
	Non-Binary	8	451.38	22.89
Ethnicity	Asian	18,865	469.67	26.36
	African American	24,882	439.98	24.41
	Hispanic	47,230	442.15	24.54
	American Indian	1197	447.29	25.62

Demographic Category		N-Count	Scale Score	
			Mean	SD
NRC	Multiracial	5,457	452.37	28.60
	Pacific Islander	354	452.10	27.14
	White	69,022	452.82	26.21
	New York	52,824	452.81	28.21
	Big 4 Cities	6,195	430.70	23.32
	Urban/Suburban	12,595	437.62	23.58
	Rural	9,167	442.24	23.42
	Average Needs	41,565	450.86	24.83
	Low Needs	19,716	465.06	24.28
	Charter	11,862	449.48	25.22
	Religious or Independent	10,617	438.96	25.29
SWD	All Codes	26,487	431.47	23.10
SUA	All Codes	15,848	429.69	22.06
ELL	ELL=Y	21,338	429.42	20.41
ELL/SUA	SUA & ELL codes	1606	423.60	17.66
SWD/SUA	SWD & SUA codes	12,457	426.31	20.34
ELL Test Language	Arabic	144	429.79	21.51
	Bengali	78	436.60	19.37
	Chinese (Traditional)	25	455.24	26.57
	Chinese (Simplified)	330	464.11	25.59
	Haitian-Creole	63	428.13	20.13
	Korean	23	462.91	29.55
	Russian	369	443.11	23.43
	Spanish	4,402	425.59	18.44
	All Translations	5,434	429.71	22.07

9.1.2.4. Mathematics Grade 6

Table 9.14 presents the Grade 6 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 449.80 with a standard deviation of 27.02. Female and Male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale-score means exceeded the State mean scale score, as did those of students enrolled in New York City, Average Needs and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (470.81). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—16 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored 18–20 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 24 scale-score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 9.14. Mathematics Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	164,792	449.80	27.02
Gender	Female	80,250	449.66	26.55
	Male	84,515	449.94	27.46
	Non-Binary	27	455.81	21.68
Ethnicity	Asian	17,715	470.81	27.97
	African American	25,550	440.33	24.30
	Hispanic	47,262	441.75	24.12
	American Indian	1168	443.15	25.71
	Multiracial	5,216	451.78	28.27
	Pacific Islander	381	453.41	28.50
	White	67,214	453.56	25.63
NRC	New York	50,765	450.90	28.88
	Big 4 Cities	6,082	433.83	23.96
	Urban/Suburban	11,953	438.15	22.70
	Rural	8,973	443.28	22.86
	Average Needs	38,741	450.83	24.45
	Low Needs	18,836	466.08	24.76
	Charter	13,264	451.17	25.94
	Religious or Independent	12,930	445.04	25.14
SWD	All Codes	25,411	430.73	20.93
SUA	All Codes	15,741	431.63	21.66
ELL	ELL=Y	19,926	429.46	18.75
ELL/SUA	SUA & ELL codes	1635	425.22	16.47
SWD/SUA	SWD & SUA codes	11,996	427.57	18.76
ELL Test Language	Arabic	219	430.53	20.00
	Bengali	151	437.88	23.96
	Chinese (Traditional)	27	460.85	26.00
	Chinese (Simplified)	681	463.77	25.25
	Haitian-Creole	93	427.08	16.58
	Korean	28	462.89	29.86
	Russian	410	441.32	21.41
	Spanish	5,617	426.90	16.59
	All Translations	7,226	431.80	21.48

9.1.2.5. Mathematics Grade 7

Table 9.15 presents the Grade 7 n-counts and scale-score statistics for key demographic subgroups. The population scale-score mean was 452.57 with a standard deviation of 27.17. Female and Male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and

White students' scale-score means exceeded the State mean scale score, as did those of students from New York City, Average Needs and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (473.32). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—17 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored 20–22 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 29 scale-score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 9.15. Mathematics Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	158,339	452.57	27.17
Gender	Female	76,627	452.77	26.75
	Male	81,665	452.39	27.55
	Non-Binary	47	453.57	24.74
Ethnicity	Asian	17,038	473.32	26.69
	African American	25,340	442.94	24.34
	Hispanic	46,691	444.92	24.34
	American Indian	1120	448.63	26.19
	Multiracial	4,548	454.62	28.64
	Pacific Islander	335	455.83	27.00
	White	62,978	456.52	26.32
NRC	New York	51,096	455.10	28.30
	Big 4 Cities	6,085	434.89	23.53
	Urban/Suburban	11,245	438.95	22.58
	Rural	8,541	444.47	23.34
	Average Needs	35,513	453.27	24.90
	Low Needs	18,095	467.57	24.78
	Charter	12,888	454.55	25.18
	Religious or Independent	11,449	448.52	26.88
SWD	All Codes	24,650	432.42	21.61
SUA	All Codes	14,453	432.36	21.76
ELL	ELL=Y	16,667	430.23	19.46
ELL/SUA	SUA & ELL codes	1210	423.40	15.49
SWD/SUA	SWD & SUA codes	11,026	428.47	19.60
ELL Test Language	Arabic	213	433.15	23.12
	Bengali	123	439.42	23.68
	Chinese (Traditional)	41	456.27	27.64
	Chinese (Simplified)	611	466.15	25.84
	Haitian-Creole	91	425.92	16.36

Demographic Category		N-Count	Scale Score	
			Mean	SD
	Korean	26	464.00	28.36
	Russian	408	449.44	23.70
	Spanish	5,301	428.14	17.07
	All Translations	6,814	433.46	22.34

9.1.2.6. Mathematics Grade 8

Table 9.16 presents the Grade 8 scale-score statistics and n-counts for key demographic subgroups. The population scale-score mean was 444.63 with a standard deviation of 26.97. Female students tended to outperform Male students by around four 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, Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (464.52). Across NRC subgroups, students from Big 4 Cities districts earned the lowest mean score—18 scale-score points below the population mean. The SWD, SUA, and ELL subgroups scored 16 scale-score points below the mean scale score for the population. ELLs tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 24 scale-score points below the State mean. Students taking the Chinese and Korean translations tended to outperform the other translation subgroups.

Table 9.16. Mathematics Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score	
			Mean	SD
State	All Students	102,560	444.63	26.97
Gender	Female	48,939	446.70	26.43
	Male	53,576	442.74	27.31
	Non-Binary	45	452.49	20.57
Ethnicity	Asian	8,211	464.52	29.23
	African American	18,328	438.26	26.42
	Hispanic	31,030	439.63	25.50
	American Indian	682	441.17	25.67
	Multiracial	2,733	442.42	26.26
	Pacific Islander	228	449.50	30.66
	White	41,111	447.58	25.49
NRC	New York	31,431	447.77	28.65
	Big 4 Cities	5,135	426.64	23.55
	Urban/Suburban	8,160	431.60	21.16
	Rural	7,026	440.15	22.66
	Average Needs	23,348	444.37	22.70
	Low Needs	8,405	457.70	24.49
	Charter	8,222	453.59	28.74

Demographic Category		N-Count	Scale Score	
			Mean	SD
	Religious or Independent	7,910	446.08	28.82
SWD	All Codes	19,196	428.32	21.71
SUA	All Codes	11,843	428.39	21.80
ELL	ELL=Y	12,883	428.54	21.89
ELL/SUA	SUA & ELL codes	1021	420.53	16.20
SWD/SUA	SWD & SUA codes	9,483	425.68	20.57
ELL Test Language	Arabic	179	428.21	22.97
	Bengali	58	433.02	22.98
	Chinese (Traditional)	22	454.50	26.11
	Chinese (Simplified)	444	463.45	28.54
	Haitian-Creole	57	424.60	20.02
	Korean	8	469.25	32.29
	Russian	368	444.35	24.82
	Spanish	3,905	425.28	18.64
	All Translations	5,041	430.42	23.57

9.2. Performance Level Distribution Summary

Students under the New York State Testing Program are classified into performance levels as Level 1, Level 2, Level 3, or Level 4. The cut scores for these performance levels were established in summer 2023 during the standards review (see Section 8: and Appendix Q: Standard Setting Technical Report for details on the standards review). While vertical articulation helps apply 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. It is also inappropriate to compare scale scores across grades because they neither measure the same content nor are on the same scale.

9.2.1. *ELA Test Performance Level Distributions*

Table 9.17 shows the performance level distributions for all examinees from public, charter, and non-public schools with valid ELA scores. Performance level data for selected subgroups[1] of students were also examined. In general, these distributions reflect the same achievement trends as in the scale-score summary discussion. Across Tables 9.18 through 9.23, more Female students were classified as Level 3 and above than Male students were. Similarly, more Asian students were classified as Level 3 and above than their peers from other reported ethnic groups were. Consistent with the pattern shown in scale-score distributions across the subgroups, students from Low Needs districts outperformed students from High Needs districts (New York City, Big 4 Cities, Urban/Suburban, and Rural). The Level 3 and above rates for students in the ELL, SWD, and SUA subgroups were low compared with the total population of examinees.

[1] In 2021–2022, a new gender category was introduced, "Non-Binary." Since processes for data collection were still in development during the 2021–2022 reporting year, school district access to this code was significantly limited and, thus, the 2021–2022 technical report does not disaggregate data by non-binary students.

Table 9.17. ELA Test Performance Level Distributions

Grade	N-Count	Performance Levels				
		Level 1	Level 2	Level 3	Level 4	Level 3 & 4
3	166,155	30.42	25.65	29.24	14.69	43.93
4	166,173	23.16	29.12	29.59	18.13	47.72
5	165,259	27.29	28.57	30.40	13.74	44.14
6	165,051	26.76	28.09	27.17	17.98	45.15
7	160,467	26.77	25.47	29.76	18.00	47.76
8	152,212	17.71	27.46	33.17	21.66	54.83

9.2.1.1. ELA Grade 3

Table 9.18 presents the ELA Grade 3 performance level distributions and n-counts for key demographic subgroups. Statewide, a combined 44% of students achieved Level 3 and Level 4. About 48% of Female students were at Level 3 or above, as compared with 40% of Male students. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (65%) students and students from Low Needs districts (64%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 25–36% of students in those same performance categories. Only about 14–17% of the SWD, SUA, and ELL subgroups earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (44%): Female (48%), Asian (65%), Multiracial (48%), Pacific Islander (49%), White (47%) students, as well as those enrolled in New York City (49%), Low Needs districts (64%), and Charter schools (48%).

Table 9.18. ELA Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	166,155	30.42	25.65	29.24	14.69	43.93
Gender	Female	82,592	26.73	24.90	30.90	17.47	48.37
	Male	83,557	34.07	26.39	27.59	11.94	39.53
	Non-Binary	6	–	16.67	66.67	16.67	83.34
Ethnicity	Asian	17,602	15.55	19.88	36.29	28.29	64.58
	African American	23,448	38.23	25.67	24.86	11.24	36.10
	Hispanic	45,285	39.13	26.95	24.22	9.71	33.93
	American Indian	1,178	33.28	25.81	28.18	12.73	40.91
	Multiracial	6,102	27.19	24.48	30.24	18.09	48.33
	Pacific Islander	316	26.90	23.73	31.33	18.04	49.37
	White	72,002	26.22	26.34	32.06	15.38	47.44
NRC	New York	47,840	27.66	23.31	29.12	19.91	49.03
	Big 4 Cities	6,128	53.28	21.44	17.22	8.06	25.28

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
	Urban/Suburban	12,893	46.56	26.61	20.55	6.27	26.82
	Rural	9,488	41.07	29.26	23.61	6.06	29.67
	Average Needs	44,171	28.43	29.62	30.83	11.11	41.94
	Low Needs	20,742	12.59	23.44	40.56	23.41	63.97
	Charter	12,241	26.74	25.12	31.08	17.07	48.15
	Religious or Independent	10,619	41.32	23.69	24.74	10.25	34.99
SWD	All Codes	25,265	57.80	25.04	13.28	3.88	17.16
SUA	All Codes	14,450	59.13	26.45	11.88	2.54	14.42
ELL	ELL=Y	21,857	62.71	23.40	11.61	2.28	13.89
ELL/SUA	SUA & ELL	1,353	71.32	22.54	5.17	0.96	6.13
SWD/SUA	SWD & SUA codes	11,822	62.98	25.64	9.61	1.77	11.38

9.2.1.2. ELA Grade 4

Table 9.19 presents the ELA Grade 4 performance level distributions and n-counts for key demographic subgroups. Statewide, a combined 48% of students achieved Level 3 and Level 4. About 51% of Female students were at Level 3 or above, as compared with 44% of Male students. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (71%) students and students from Low Needs districts (68%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 26–39% of students in those same performance categories. Only about 11–19% of the SWD, SUA, and ELL subgroups on average earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (48%): Female (51%), Asian (71%), Multiracial (51%), Pacific Islander (53%), and White (51%) students, as well as those enrolled in New York City (54%), Low Needs districts (68%), and Charter schools (51%).

Table 9.19. ELA Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	166,173	23.16	29.12	29.59	18.13	47.72
Gender	Female	81,980	19.96	28.60	30.85	20.60	51.45
	Male	84,187	26.27	29.63	28.37	15.73	44.10
	Non-Binary	6	16.67	33.33	–	50.00	50.00
Ethnicity	Asian	17,287	9.90	19.54	33.79	36.77	70.56
	African American	24,414	29.57	31.82	26.20	12.41	38.61
	Hispanic	45,849	28.78	32.70	26.73	11.79	38.52
	American Indian	1,207	25.35	30.65	27.01	16.98	43.99

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
NRC	Multiracial	5,773	21.38	28.03	27.68	22.92	50.60
	Pacific Islander	366	19.67	27.05	33.06	20.22	53.28
	White	71,034	20.60	28.28	31.80	19.31	51.11
	New York	49,017	19.74	26.53	29.46	24.26	53.72
	Big 4 Cities	6,288	46.20	27.94	17.22	8.64	25.86
	Urban/Suburban	12,292	35.71	33.98	22.78	7.53	30.31
	Rural	9,274	31.96	35.40	24.16	8.48	32.64
	Average Needs	42,317	20.93	32.59	31.76	14.72	46.48
	Low Needs	19,584	8.40	23.66	38.98	28.96	67.94
	Charter	11,706	18.95	30.27	32.46	18.32	50.78
	Religious or Independent	13,308	33.39	27.06	25.65	13.90	39.55
SWD	All Codes	26,093	50.06	30.56	14.96	4.42	19.38
SUA	All Codes	15,880	51.75	31.38	13.59	3.28	16.87
ELL	ELL=Y	19,170	57.45	31.55	9.98	1.02	11.00
ELL/SUA	SUA & ELL	1,417	63.23	29.64	6.63	0.49	7.12
SWD/SUA	SWD & SUA codes	12,584	56.90	29.86	11.05	2.19	13.24

9.2.1.3. ELA Grade 5

Table 9.20 presents the ELA Grade 5 performance level distributions and n-counts for key demographic subgroups. Statewide, a combined 44% of students achieved Level 3 and Level 4. About 48% of Female students were at Level 3 or above, as compared with 40% of Male students. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (68%) students and students from Low Needs districts (64%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 24–36% of students in those same performance categories. Only about 10–16% of the SWD, SUA, and ELL subgroups on average earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (44%): Female (48%), Asian (68%), Multiracial (47%), Pacific Islander (46%), and White (47%) students, as well as those enrolled in New York City (51%), Low Needs districts (64%), and Charter schools (45%).

Table 9.20. ELA Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	165,259	27.29	28.57	30.40	13.74	44.14
Gender	Female	81,369	22.94	28.77	32.26	16.02	48.28

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
Ethnicity	Male	83,881	31.51	28.37	28.59	11.53	40.12
	Non-Binary	9	–	44.44	55.56	–	55.56
	Asian	18,360	11.41	20.16	37.53	30.90	68.43
	African American	25,173	34.87	30.92	26.28	7.93	34.21
	Hispanic	45,712	32.33	31.67	27.37	8.63	36.00
	American Indian	1,206	28.28	30.85	29.77	11.11	40.88
	Multiracial	5,528	26.48	26.92	29.43	17.17	46.60
	Pacific Islander	351	26.50	27.07	28.77	17.66	46.43
	White	68,729	25.37	28.02	32.15	14.46	46.61
NRC	New York	50,926	21.95	26.71	32.17	19.17	51.34
	Big 4 Cities	6,132	49.25	26.47	19.08	5.20	24.28
	Urban/Suburban	12,491	40.41	32.14	22.03	5.43	27.46
	Rural	9,301	37.96	32.74	23.54	5.76	29.30
	Average Needs	41,848	26.04	32.24	31.45	10.26	41.71
	Low Needs	19,638	11.71	24.51	40.10	23.68	63.78
	Charter	11,875	24.57	30.07	33.41	11.95	45.36
	Religious or Independent	10,310	43.39	23.60	23.51	9.50	33.01
SWD	All Codes	26,863	57.06	27.05	13.31	2.58	15.89
SUA	All Codes	16,090	60.08	26.13	11.74	2.05	13.79
ELL	ELL=Y	18,156	63.65	26.66	8.98	0.71	9.69
ELL/SUA	SUA & ELL	1,504	71.81	23.40	4.52	0.27	4.79
SWD/SUA	SWD & SUA codes	12,819	65.76	24.30	8.90	1.04	9.94

9.2.1.4. ELA Grade 6

Table 9.21 presents the ELA Grade 6 performance level distributions and n-counts for key demographic subgroups. Statewide, a combined 45% of students achieved Level 3 and Level 4. About 50% of Female students were at Level 3 or above, as compared with 40% of Male students. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (70%) students and students from Low Needs districts (66%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 25–35% of students in those same performance categories. Only about 6–15% of the SWD, SUA, and ELL subgroups on average earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (45%): Female (50%), Asian (70%), Multiracial (48%), Pacific Islander (51%), and White (49%) students, as well as those from New York City (49%), Low Needs districts (66%), and Charter schools (46%).

Table 9.21. ELA Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	165,051	26.76	28.09	27.17	17.98	45.15
Gender	Female	80,890	22.01	27.85	28.99	21.15	50.14
	Male	84,133	31.33	28.33	25.41	14.93	40.34
	Non-Binary	28	10.71	28.57	39.29	21.43	60.72
Ethnicity	Asian	17,433	11.05	19.11	32.16	37.68	69.84
	African American	26,068	33.39	31.25	24.12	11.24	35.36
	Hispanic	46,408	33.33	31.51	24.05	11.12	35.17
	American Indian	1,184	33.36	29.05	25.08	12.50	37.58
	Multiracial	5,369	25.57	26.04	26.86	21.53	48.39
	Pacific Islander	385	22.60	25.97	30.13	21.30	51.43
	White	67,927	23.65	27.03	29.28	20.04	49.32
NRC	New York	49,364	23.78	26.75	27.11	22.35	49.46
	Big 4 Cities	6,064	49.32	25.96	17.02	7.70	24.72
	Urban/Suburban	12,027	39.83	31.54	20.35	8.28	28.63
	Rural	9,200	34.15	31.70	23.93	10.22	34.15
	Average Needs	40,015	24.93	30.43	28.33	16.30	44.63
	Low Needs	18,975	10.70	23.69	36.23	29.38	65.61
	Charter	13,254	21.62	32.60	30.59	15.20	45.79
	Religious or Independent	12,815	36.10	24.11	24.43	15.36	39.79
SWD	All Codes	26,146	60.59	25.44	10.70	3.27	13.97
SUA	All Codes	16,486	60.01	25.17	11.08	3.74	14.82
ELL	ELL=Y	16,906	70.15	23.41	5.67	0.77	6.44
ELL/SUA	SUA & ELL	1,596	79.20	17.17	3.45	0.19	3.64
SWD/SUA	SWD & SUA codes	12,560	68.01	22.64	7.64	1.72	9.36

9.2.1.5. ELA Grade 7

Table 9.22 presents the ELA Grade 7 performance level distributions and n-counts for key demographic subgroups. Statewide, a combined 48% of students achieved Level 3 and Level 4. About 54% of Female students were at Level 3 or above, as compared with 41% of Male students. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (72%) students and students from Low Needs districts (66%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 27–39% of students in those same performance categories. Only about 5–16% of the SWD, SUA, and ELL subgroups on average earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (48%): Female (54%), Asian (72%), Multiracial (51%), Pacific Islander

(53%), and White (51%) students, as well as those enrolled in New York City (54%), Low Needs districts (66%), and Charter schools (54%).

Table 9.22. ELA Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	160,467	26.77	25.47	29.76	18.00	47.76
Gender	Female	78,213	21.03	24.54	32.26	22.18	54.44
	Male	82,205	32.25	26.36	27.37	14.02	41.39
	Non-Binary	49	10.20	26.53	26.53	36.73	63.26
Ethnicity	Asian	17,027	10.82	16.99	34.53	37.65	72.18
	African American	26,124	32.55	27.99	27.46	12.01	39.47
	Hispanic	46,560	33.14	27.69	27.27	11.89	39.16
	American Indian	1,141	31.29	26.21	27.17	15.34	42.51
	Multiracial	4,748	25.95	23.46	28.85	21.74	50.59
	Pacific Islander	329	21.28	26.14	31.31	21.28	52.59
	White	64,246	23.93	25.20	31.39	19.47	50.86
NRC	New York	50,600	22.52	23.95	30.54	22.99	53.53
	Big 4 Cities	6,086	47.85	24.66	19.14	8.35	27.49
	Urban/Suburban	11,668	43.56	27.76	21.02	7.66	28.68
	Rural	8,871	36.88	30.38	23.79	8.95	32.74
	Average Needs	36,918	26.75	28.31	30.56	14.38	44.94
	Low Needs	18,594	11.96	22.39	37.79	27.86	65.65
	Charter	12,886	18.76	27.11	35.57	18.56	54.13
	Religious or Independent	11,234	32.66	20.93	28.68	17.73	46.41
SWD	All Codes	25,660	58.06	25.68	12.83	3.43	16.26
SUA	All Codes	14,625	60.50	24.05	12.25	3.19	15.44
ELL	ELL=Y	14,042	76.62	18.69	4.28	0.41	4.69
ELL/SUA	SUA & ELL	1,075	85.95	12.09	1.86	0.09	1.95
SWD/SUA	SWD & SUA codes	11,090	68.18	21.80	8.39	1.63	10.02

9.2.1.6. ELA Grade 8

Table 9.23 presents the ELA Grade 8 performance level distributions and n-counts for key demographic subgroups. Statewide, a combined 55% of students achieved Level 3 and Level 4. About 61% of Female students were at Level 3 or above, as compared with 49% of Male students. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (79%) students and students from Low Needs districts (73%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 34–48% of students in those same performance categories. Only about 8–22% of the SWD,

SUA, and ELL subgroups on average earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (55%): Female (61%), Asian (79%), Multiracial (57%), Pacific Islander (67%), and White (56%) students, as well as those enrolled in New York City (62%) and Low Needs districts (73%), and Charter schools (62%).

Table 9.23. ELA Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	152,212	17.71	27.46	33.17	21.66	54.83
Gender	Female	73,586	13.17	26.06	35.66	25.11	60.77
	Male	78,571	21.96	28.77	30.84	18.42	49.26
	Non-Binary	55	7.27	16.36	34.55	41.82	76.37
Ethnicity	Asian	16,824	6.47	14.54	33.96	45.03	78.99
	African American	26,346	21.12	31.61	32.41	14.86	47.27
	Hispanic	44,283	21.48	30.67	31.99	15.86	47.85
	American Indian	1,067	17.90	29.62	33.93	18.56	52.49
	Multiracial	4,275	17.01	25.71	33.71	23.58	57.29
	Pacific Islander	344	10.47	22.97	40.12	26.45	66.57
	White	58,807	16.53	26.98	34.14	22.35	56.49
NRC	New York	50,885	13.48	24.66	33.89	27.97	61.86
	Big 4 Cities	6,221	34.74	31.51	23.71	10.05	33.76
	Urban/Suburban	11,528	28.89	33.52	27.21	10.38	37.59
	Rural	8,594	24.47	34.37	30.13	11.03	41.16
	Average Needs	34,288	18.36	30.54	33.92	17.19	51.11
	Low Needs	16,627	7.30	20.18	39.40	33.13	72.53
	Charter	12,137	10.19	27.59	39.10	23.11	62.21
	Religious or Independent	8,129	25.58	25.62	29.68	19.12	48.80
SWD	All Codes	24,687	42.42	35.93	17.37	4.28	21.65
SUA	All Codes	14,449	45.91	33.78	16.04	4.26	20.30
ELL	ELL=Y	12,771	60.08	31.45	7.75	0.72	8.47
ELL/SUA	SUA & ELL	1,112	66.82	29.05	3.96	0.18	4.14
SWD/SUA	SWD & SUA codes	11,372	51.36	33.33	12.81	2.50	15.31

9.2.2. *Mathematics Test Performance Level Distributions*

Table 9.24 shows the performance level distributions for all examinees from public, charter, and non-public schools with valid scores and presents mathematics performance level data for total populations of students in Grades 3–8. Performance level data for selected subgroups of students were also examined. In general, these summaries reflect the same achievement trends as in the scale-score summary discussion. Across Table 9.25 through Table 9.30, Male students

outperformed Female students in terms of Level 3 and above classifications, except for Grade 8. More White, Pacific Islander, and Asian students were classified in Level 3 and above, as compared with 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, Urban/Suburban, and Rural), and Religious or Independent schools. The Level 3 and above rates for SWD, SUA, and ELL subgroups were low compared with the total population of examinees. The subgroups that used the Korean or Chinese translations outperformed other test translation subgroups. The n-counts for some translation subgroups were low, and the results might have been heavily influenced by very high- and/or very low-achieving individual students.

Table 9.24. Mathematics Test Performance Level Distributions

Grade	N-Count	Performance Levels				
		Level 1	Level 2	Level 3	Level 4	Level 3 & 4
3	169,444	16.48	31.28	35.94	16.30	52.24
4	169,293	23.94	23.80	37.45	14.81	52.26
5	167,238	28.10	23.65	32.69	15.56	48.25
6	164,792	27.17	25.75	33.79	13.29	47.08
7	158,339	21.70	26.90	29.71	21.69	51.40
8	102,560	40.62	18.20	30.10	11.08	41.18

9.2.2.1. Mathematics Grade 3

Table 9.25 presents the mathematics Grade 3 performance level summaries and n-counts for key demographic subgroups. Statewide, a combined 52% of students achieved Level 3 and Level 4. About 51% of Female and 54% of Male students were at Level 3 or above. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (76%) students and students from Low Needs districts (75%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 26–41% of students in those same performance categories. Only about 21–26% of the SWD, SUA, and ELL subgroups earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (52%): Male (54%), Asian (76%), Multiracial (55%), Pacific Islander (56%), and White (57%) students, as well as those enrolled at New York City (56%), Low Needs districts (75%) and Charter schools (58%). For ELLs who used translated test forms, the percentages of students earning at least a Level 3 ranged from 16% (Haitian-Creole) to 79% (Korean).

Table 9.25. Mathematics Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	169,444	16.48	31.28	35.94	16.30	52.24
Gender	Female	83,909	16.21	33.24	35.39	15.16	50.55
	Male	85,529	16.75	29.36	36.48	17.42	53.90
	Non-Binary	6	–	–	50.00	50.00	100.0

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
Ethnicity	Asian	18,176	5.77	17.79	39.67	36.77	76.44
	African American	23,497	23.88	34.80	30.41	10.91	41.32
	Hispanic	47,307	21.97	36.96	31.15	9.93	41.08
	American Indian	1,174	17.38	33.22	34.41	14.99	49.40
	Multiracial	6,078	16.16	29.27	34.70	19.87	54.57
	Pacific Islander	324	12.96	31.17	38.27	17.59	55.86
	White	72,639	13.14	29.94	40.09	16.84	56.93
NRC	New York	50,220	14.95	29.11	35.26	20.68	55.94
	Big 4 Cities	6,222	40.32	33.49	20.09	6.09	26.18
	Urban/Suburban	13,160	27.01	38.69	27.36	6.94	34.30
	Rural	9,478	22.11	37.89	31.91	8.09	40.00
	Average Needs	44,307	13.85	34.18	38.77	13.20	51.97
	Low Needs	20,892	4.51	20.61	46.37	28.52	74.89
	Charter	12,252	11.55	30.05	40.16	18.24	58.40
	Religious or Independent	10,844	25.37	35.71	29.39	9.54	38.93
SWD	All Codes	25,496	37.55	36.16	20.69	5.60	26.29
SUA	All Codes	14,685	42.23	36.57	17.90	3.31	21.21
ELL	ELL=Y	25,239	33.67	40.11	21.65	4.58	26.23
ELL/SUA	SUA & ELL codes	1643	46.80	38.10	12.96	2.13	15.09
SWD/SUA	SWD & SUA codes	12,001	46.01	35.92	15.51	2.56	18.07
ELL Test Language	Arabic	144	32.64	40.97	20.14	6.25	26.39
	Bengali	84	29.76	42.86	22.62	4.76	27.38
	Chinese (Traditional)	35	5.71	34.29	42.86	17.14	60.00
	Chinese (Simplified)	347	4.61	20.17	48.99	26.22	75.21
	Haitian-Creole	50	36.00	48.00	14.00	2.00	16.00
	Korean	43	9.30	11.63	58.14	20.93	79.07
	Russian	388	16.49	40.98	33.51	9.02	42.53
	Spanish	4,744	40.94	41.25	15.43	2.38	17.81
	All Translations	5,835	36.30	39.79	19.31	4.59	23.90

9.2.2.2. Mathematics Grade 4

Table 9.26 presents the mathematics Grade 4 performance level summaries and n-counts for key demographic subgroups. Statewide, a combined 52% of students achieved Level 3 and Level 4. About 50% of Female students and 55% of Male students were at Level 3 or above. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were

Asian (77%) students and students from Low Needs districts (76%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 25–40% of students in those same performance categories. Only about 19–25% of the SWD, SUA, and ELL subgroups earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (52%): Asian (77%), Multiracial (56%), Pacific Islander (56%), and White (59%) students, as well as students enrolled in New York City (54%), Average Needs (55%) and Low Needs (76%) districts, and Charter schools (57%). For ELLs who used translated test forms, the percentages of students earning at least a Level 3 ranged from 11% (Haitian-Creole) to 79% (Korean).

Table 9.26. Mathematics Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	169,293	23.94	23.80	37.45	14.81	52.26
Gender	Female	83,150	24.73	25.51	36.81	12.95	49.76
	Male	86,135	23.17	22.15	38.07	16.61	54.68
	Non-Binary	8	25.00	–	75.00	–	75.00
Ethnicity	Asian	17,845	8.50	14.44	40.69	36.37	77.06
	African American	24,366	34.64	26.14	30.91	8.31	39.22
	Hispanic	47,625	32.59	27.49	31.78	8.14	39.92
	American Indian	1,217	28.51	24.57	34.92	12.00	46.92
	Multiracial	5,756	21.89	22.55	36.81	18.75	55.56
	Pacific Islander	367	22.34	21.53	38.96	17.17	56.13
	White	71,865	18.39	22.99	42.78	15.85	58.63
NRC	New York	51,179	24.09	22.40	34.87	18.65	53.52
	Big 4 Cities	6,406	52.53	22.09	19.43	5.95	25.38
	Urban/Suburban	12,558	37.23	27.48	29.20	6.09	35.29
	Rural	9,249	27.18	29.62	36.52	6.67	43.19
	Average Needs	42,581	18.63	26.20	42.81	12.36	55.17
	Low Needs	19,921	6.75	16.83	48.63	27.79	76.42
	Charter	11,776	19.24	23.86	41.65	15.24	56.89
	Religious or Independent	13,236	34.69	26.00	31.06	8.26	39.32
SWD	All Codes	26,275	51.05	24.41	20.26	4.27	24.53
SUA	All Codes	16,437	52.45	24.91	19.41	3.23	22.64
ELL	ELL=Y	22,429	52.81	27.77	17.31	2.12	19.43
ELL/SUA	SUA & ELL codes	1806	58.53	25.25	15.34	0.89	16.23
SWD/SUA	SWD & SUA codes	13,039	58.39	23.50	15.80	2.31	18.11
ELL Test Language	Arabic	137	49.64	28.47	19.71	2.19	21.90
	Bengali	96	41.67	15.63	35.42	7.29	42.71
	Chinese (Traditional)	42	14.29	14.29	57.14	14.29	71.43

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
	Chinese (Simplified)	329	10.03	20.06	39.21	30.70	69.91
	Haitian-Creole	45	66.67	22.22	11.11	–	11.11
	Korean	28	14.29	7.14	50.00	28.57	78.57
	Russian	364	23.35	33.52	34.62	8.52	43.14
	Spanish	4,551	63.48	23.84	11.78	0.90	12.68
	All Translations	5,592	56.42	24.05	16.01	3.52	19.53

9.2.2.3. Mathematics Grade 5

Table 9.27 presents the mathematics Grade 5 performance level summaries and n-counts for key demographic subgroups. Statewide, a combined 48% of students achieved Level 3 and Level 4. About 46% of Female students and 50% of Male students were at Level 3 or above. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (76%) students and students from Low Needs districts (73%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 21–36% of students in those same performance categories. Only about 16–21% of the SWD, SUA, and ELL subgroups earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (48%): Male (50%), Asian (76%), Multiracial (52%), Pacific Islander (51%), and White (54%) students, as well as those enrolled in New York City (52%), and Average Needs (51%) and Low Needs (73%) districts. For ELLs who used translated test forms, the percentages of students earning at least a Level 3 ranged from 11% (Spanish) to 74% (Korean).

Table 9.27. Mathematics Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	167,238	28.10	23.65	32.69	15.56	48.25
Gender	Female	81,848	28.56	25.20	32.31	13.93	46.24
	Male	85,382	27.65	22.17	33.06	17.12	50.18
	Non-Binary	8	25.00	12.50	50.00	12.50	62.50
Ethnicity	Asian	18,865	9.35	14.38	36.01	40.26	76.27
	African American	24,882	40.79	26.06	26.17	6.98	33.15
	Hispanic	47,230	36.82	27.07	27.96	8.15	36.11
	American Indian	1,197	29.74	25.81	32.50	11.95	44.45
	Multiracial	5,457	26.88	20.89	32.42	19.81	52.23
	Pacific Islander	354	25.71	23.45	32.20	18.64	50.84
	White	69,022	22.64	23.18	37.46	16.73	54.19
NRC	New York	52,824	25.94	22.27	31.64	20.15	51.79
	Big 4 Cities	6,195	58.76	20.52	16.88	3.84	20.72

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
	Urban/Suburban	12,595	44.52	26.08	23.88	5.52	29.40
	Rural	9,167	34.94	28.58	29.81	6.67	36.48
	Average Needs	41,565	23.11	26.18	37.28	13.43	50.71
	Low Needs	19,716	9.06	17.83	43.71	29.40	73.11
	Charter	11,862	26.05	26.17	34.63	13.15	47.78
	Religious or Independent	10,617	43.15	24.88	24.55	7.42	31.97
SWD	All Codes	26,487	56.91	22.53	16.43	4.13	20.56
SUA	All Codes	15,848	59.38	22.54	15.17	2.92	18.09
ELL	ELL=Y	21,338	59.13	24.66	14.04	2.18	16.22
ELL/SUA	SUA & ELL codes	1606	70.11	20.67	8.66	0.56	9.22
SWD/SUA	SWD & SUA codes	12,457	65.79	20.82	11.59	1.80	13.39
ELL Test Language	Arabic	144	57.64	25.00	14.58	2.78	17.36
	Bengali	78	42.31	35.90	20.51	1.28	21.79
	Chinese (Traditional)	25	20.00	24.00	36.00	20.00	56.00
	Chinese (Simplified)	330	9.70	20.00	37.58	32.73	70.31
	Haitian-Creole	63	57.14	20.63	22.22	–	22.22
	Korean	23	17.39	8.70	47.83	26.09	73.92
	Russian	369	34.15	32.25	25.75	7.86	33.61
	Spanish	4,402	67.86	20.90	10.22	1.02	11.24
	All Translations	5,434	60.84	21.90	13.62	3.64	17.26

9.2.2.4. Mathematics Grade 6

Table 9.28 presents the mathematics Grade 6 performance level summaries and n-counts for key demographic subgroups. Statewide, a combined 47% of students achieved Level 3 and Level 4. About 46% of Female students and 48% of Male students were at Level 3 or above. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (76%) students and students from Low Needs districts (74%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 23–34% of students in those same performance categories. Only about 14–18% of the SWD, SUA, and ELL subgroups earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (47%): Male (48%), Asian (76%), Multiracial (49%), Pacific Islander (50%), and White (54%) students, as well as those enrolled in Average Needs (50%) and Low Needs (74%) districts and Charter schools (50%). For ELLs who used translated test forms, the percentages of students earning at least a Level 3 ranged from 10% (Spanish) to 69% (Simplified Chinese).

Table 9.28. Mathematics Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	164,792	27.17	25.75	33.79	13.29	47.08
Gender	Female	80,250	26.57	27.14	33.46	12.84	46.30
	Male	84,515	27.75	24.43	34.10	13.72	47.82
	Non-Binary	27	11.11	25.93	51.85	11.11	62.96
Ethnicity	Asian	17,715	9.34	14.50	38.57	37.59	76.16
	African American	25,550	40.03	27.84	25.79	6.35	32.14
	Hispanic	47,262	36.98	28.86	27.59	6.57	34.16
	American Indian	1,168	37.07	26.97	27.23	8.73	35.96
	Multiracial	5,216	26.28	24.54	32.61	16.56	49.17
	Pacific Islander	381	23.62	25.98	34.38	16.01	50.39
	White	67,214	19.86	25.84	40.20	14.10	54.30
NRC	New York	50,765	28.52	24.56	30.40	16.52	46.92
	Big 4 Cities	6,082	53.96	23.10	17.84	5.10	22.94
	Urban/Suburban	11,953	42.56	28.95	24.04	4.46	28.50
	Rural	8,973	32.22	31.29	30.60	5.88	36.48
	Average Needs	38,741	21.66	28.12	39.34	10.88	50.22
	Low Needs	18,836	7.64	18.16	48.00	26.19	74.19
	Charter	13,264	24.14	25.80	37.08	12.98	50.06
	Religious or Independent	12,930	31.08	30.02	29.91	9.00	38.91
SWD	All Codes	25,411	58.85	24.05	14.45	2.65	17.10
SUA	All Codes	15,741	57.05	24.46	15.14	3.35	18.49
ELL	ELL=Y	19,926	59.09	27.31	11.97	1.62	13.59
ELL/SUA	SUA & ELL codes	1635	69.30	22.32	7.71	0.67	8.38
SWD/SUA	SWD & SUA codes	11,996	64.85	22.49	11.13	1.53	12.66
ELL Test Language	Arabic	219	56.16	25.11	16.89	1.83	18.72
	Bengali	151	46.36	22.52	25.17	5.96	31.13
	Chinese (Traditional)	27	14.81	22.22	40.74	22.22	62.96
	Chinese (Simplified)	681	10.72	19.82	44.64	24.82	69.46
	Haitian-Creole	93	61.29	27.96	9.68	1.08	10.76
	Korean	28	14.29	25.00	28.57	32.14	60.71
	Russian	410	32.68	35.61	27.56	4.15	31.71
	Spanish	5,617	65.25	24.53	9.56	0.66	10.22
	All Translations	7,226	57.15	24.73	14.63	3.49	18.12

9.2.2.5. Mathematics Grade 7

Table 9.29 presents the mathematics Grade 7 performance level summaries and n-counts for key demographic subgroups. Statewide, 51% of students achieved Level 3 and Level 4, and Male and Female students performed similarly. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (79%) students and students from Low Needs districts (76%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 24–39% of students in those same performance categories. Only about 15–20% of the SWD, SUA, and ELL subgroups earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (51%): Asian (79%), Multiracial (55%), Pacific Islander (57%), and White (59%) students, as well as those enrolled in New York City (53%), Average Needs (55%) and Low Needs (76%) districts and Charter schools (55%). For ELLs who used translated test forms, the percentages of students earning at least a Level 3 ranged from 11% (Spanish) to 69% (Korean).

Table 9.29. Mathematics Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	158,339	21.70	26.90	29.71	21.69	51.40
Gender	Female	76,627	20.75	27.94	29.84	21.48	51.32
	Male	81,665	22.59	25.92	29.59	21.90	51.49
	Non-Binary	47	21.28	21.28	38.30	19.15	57.45
Ethnicity	Asian	17,038	6.64	14.29	28.00	51.06	79.06
	African American	25,340	32.19	31.74	25.08	10.99	36.07
	Hispanic	46,691	28.82	32.29	26.66	12.23	38.89
	American Indian	1,120	26.61	27.95	28.39	17.05	45.44
	Multiracial	4,548	21.28	24.16	28.96	25.59	54.55
	Pacific Islander	335	19.70	22.99	31.64	25.67	57.31
	White	62,978	16.08	24.55	34.45	24.93	59.38
NRC	New York	51,096	20.10	26.95	27.40	25.55	52.95
	Big 4 Cities	6,085	48.09	28.28	16.53	7.10	23.63
	Urban/Suburban	11,245	37.77	32.55	22.40	7.28	29.68
	Rural	8,541	28.09	32.30	29.07	10.54	39.61
	Average Needs	35,513	17.82	27.52	34.93	19.73	54.66
	Low Needs	18,095	7.06	16.95	36.15	39.83	75.98
	Charter	12,888	16.64	28.30	33.37	21.69	55.06
	Religious or Independent	11,449	25.22	28.46	29.44	16.88	46.32
SWD	All Codes	24,650	51.59	29.09	14.43	4.90	19.33
SUA	All Codes	14,453	51.73	27.94	15.64	4.68	20.32
ELL	ELL=Y	16,667	54.08	31.15	11.86	2.90	14.76

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
ELL/SUA	SUA & ELL codes	1210	71.32	22.89	5.04	0.74	5.78
SWD/SUA	SWD & SUA codes	11,026	59.10	26.70	11.40	2.80	14.20
ELL Test Language	Arabic	213	45.07	32.86	16.90	5.16	22.06
	Bengali	123	38.21	30.89	23.58	7.32	30.90
	Chinese (Traditional)	41	17.07	21.95	36.59	24.39	60.98
	Chinese (Simplified)	611	9.82	17.68	33.55	38.95	72.50
	Haitian-Creole	91	59.34	31.87	7.69	1.10	8.79
	Korean	26	11.54	19.23	34.62	34.62	69.24
	Russian	408	19.36	34.31	33.09	13.24	46.33
	Spanish	5,301	58.40	30.60	9.51	1.49	11.00
	All Translations	6,814	50.51	29.66	13.80	6.03	19.83

9.2.2.6. Mathematics Grade 8

Table 9.30 presents the mathematics Grade 8 performance level summaries and n-counts for key demographic subgroups. Statewide, a combined 41% of students achieved Level 3 and Level 4. About 44% of Female students were at Level 3 or above, as compared with 39% of Male students. The percentage of students in Levels 3 and 4 varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level 3 and above were Asian (69%) students and students from Low Needs districts (64%). The Big 4 Cities, High Needs/Urban/Suburban, African American, and Hispanic students had a range of 16–32% of students in those same performance categories. Only about 16–17% of the SWD, SUA, and ELL subgroups earned at least a Level 3. Except for the Non-Binary gender group, each of the following subgroups had a higher percentage of students in Levels 3 and 4 than statewide (41%): Female (44%), Asian (69%), Pacific Islander (48%), and White (47%) students, as well as those enrolled in New York City (44%), Average Needs (42%) and Low Needs (64%) districts, Charter schools (53%), and Religious or Independent (45%) schools. For ELLs who used translated test forms, the percentages of students earning at least a Level 3 ranged from 10% (Spanish) to 75% (Korean).

Table 9.30. Mathematics Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
State	All Students	102,560	40.62	18.20	30.10	11.08	41.18
Gender	Female	48,939	36.79	19.19	32.32	11.70	44.02
	Male	53,576	44.13	17.30	28.06	10.50	38.56
	Non-Binary	45	24.44	17.78	46.67	11.11	57.78
Ethnicity	Asian	8,211	18.60	12.58	36.13	32.69	68.82
	African American	18,328	52.18	16.53	23.20	8.08	31.28

Demographic Category		N-Count	Performance Levels				
			Level 1	Level 2	Level 3	Level 4	Level 3 & 4
NRC	Hispanic	31,030	49.05	18.50	24.51	7.94	32.45
	American Indian	682	47.07	18.04	26.39	8.50	34.89
	Multiracial	2,733	42.59	19.14	29.38	8.89	38.27
	Pacific Islander	228	39.04	13.16	28.51	19.30	47.81
	White	41,111	33.04	19.86	36.44	10.66	47.10
	New York	31,431	39.01	16.85	28.69	15.45	44.14
	Big 4 Cities	5,135	72.62	11.61	11.78	3.99	15.77
	Urban/Suburban	8,160	61.32	19.24	16.96	2.48	19.44
	Rural	7,026	43.75	22.39	29.19	4.67	33.86
	Average Needs	23,348	35.74	22.25	36.04	5.97	42.01
	Low Needs	8,405	18.39	17.26	46.19	18.16	64.35
	Charter	8,222	30.09	16.74	32.67	20.51	53.18
	Religious or Independent	7,910	38.96	16.42	30.75	13.87	44.62
SWD	All Codes	19,196	68.09	15.13	14.35	2.43	16.78
SUA	All Codes	11,843	67.56	15.22	14.87	2.36	17.23
ELL	ELL=Y	12,883	68.59	14.97	13.62	2.82	16.44
ELL/SUA	SUA & ELL codes	1021	84.43	9.89	5.39	0.29	5.68
SWD/SUA	SWD & SUA codes	9,483	72.69	14.06	11.58	1.68	13.26
ELL Test Language	Arabic	179	70.39	14.53	10.61	4.47	15.08
	Bengali	58	62.07	13.79	20.69	3.45	24.14
	Chinese (Traditional)	22	27.27	18.18	40.91	13.64	54.55
	Chinese (Simplified)	444	20.50	11.04	38.06	30.41	68.47
	Haitian-Creole	57	73.68	10.53	15.79	–	15.79
	Korean	8	25.00	–	37.50	37.50	75.00
	Russian	368	42.66	19.84	27.72	9.78	37.50
	Spanish	3,905	75.01	14.62	9.35	1.02	10.37
	All Translations	5,041	67.23	14.62	13.65	4.50	18.15

References

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

Cai, L., Thissen, D. J. & du Toit, S. (2011). *IRTPRO* (Version 5.2) [Computer software]. Vector Psychometric Group (VPG). Available from https://store.vpgcentral.com/

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Dorans, N. J. & Holland, P. W. (1992). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 33–66). Lawrence Erlbaum Associates, Inc. https://doi.org/10.1002/j.2333-8504.1992.tb01440.x

Dorans, N. J., Schmitt, A. P. & Bleistein, C. A. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement, 29*(4), 309–319. http://www.jstor.org/stable/1435087

Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Springer. https://doi.org/10.1007/978-94-017-1988-9

Individuals with Disabilities Education Act (IDEA). Pub L. No. 108–446 (2004). https://www.congress.gov/108/plaws/publ446/PLAW-108publ446.pdf

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

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kim S. & Kolen M. J. (2004). *STUIRT: A computer program for scale transformation under unidimensional item response theory models* (Version 1.0). [Computer software]. Iowa Testing Programs, University of Iowa.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Lee, W.-C. (2010). Classification consistency and accuracy for complex assessments using item response theory. *Journal of Educational Measurement*, *47*(1), 1–17. https://doi.org/10.1111/j.1745-3984.2009.00096.x

Lee, W.-C., Hanson, B. A. & Brennan, R. L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement*, *26*(4), 412–432. https://doi.org/10.1177/014662102237797

Lee, W.-C. & Kolen, M. J. (2006, Revised 2008). *IRT-CLASS: A computer program for item response theory classification consistency and accuracy* (Version 2.0) [Computer software]. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

Livingston, S. A. & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*(2), 179–197. http://www.jstor.org/stable/1435147

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge. https://doi.org/10.4324/9780203056615

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Addison-Wesley.

Mehrens, W. A. & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (3rd Ed.). Holt, Rinehart, and Winston.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2), 159–176. https://doi.org/10.1177/014662169201600206

Qualls, A. L. (1995). Estimating the reliability of a test containing multiple item formats. *Applied Measurement in Education, 8*(2), 111–120. https://doi.org/10.1207/s15324818ame0802_1

Sandoval, J. & Mille, M. W. (1980) Accuracy of judgments of WISC-R item difficulty for minority groups. *Journal of Consulting and Clinical Psychology*, *48*(2), 249–253. https://doi.org/10.1037/0022-006X.48.2.249

Sinharay, S. (2006) Bayesian item fit analysis for unidimensional item response theory models. *British Journal of Mathematical and Statistical Psychology*, *59*(2), 429–449. https://doi.org/10.1348/000711005X66888

Thompson, S. J., Johnstone C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. https://nceo.umn.edu/docs/onlinepubs/synth44.pdf

Wang, T., Kolen, M. J., & Harris, D. J. (2000). Psychometric properties of scale scores and performance levels for performance assessment using polytomous IRT. *Journal of Educational Measurement*, *37*(2), 141–162. https://psycnet.apa.org/doi/10.1111/j.1745-3984.2000.tb01080.x

Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. *Journal of Educational Measurement, 21*(2), 93–111. https://www.jstor.org/stable/1434536

Appendix A: ELA and Mathematics Test Configurations and Testing Times

Table A1. ELA Test Configuration

			Number of Items				
			Multiple-Choice		Constructed-Response		
Grade	Day	Session	Operational	Embedded	Operational	Embedded	Total
3	1	1	17	6	2	0	25
	2	2	6	0	3	0	9
		Total	23	6	5	0	34
4	1	1	17	6	2	0	25
	2	2	6	0	4	0	10
		Total	23	6	6	0	35
5	1	1	19	7	2	0	28
	2	2	7	0	4	0	11
		Total	26	7	6	0	39
6	1	1	19	7	2	0	28
	2	2	7	0	4	0	11
		Total	26	7	6	0	39
7	1	1	19	7	2	0	28
	2	2	14	0	4	0	18
		Total	33	7	6	0	46
8	1	1	19	7	2	0	28
	2	2	14	0	4	0	18
		Total	33	7	6	0	46

Table A2. Mathematics Test Configuration

			Number of Items				
			Multiple-Choice		Constructed-Response		
Grade	Day	Session	Operational	Embedded	Operational	Embedded	Total
3	1	1	19	6	0	0	25
	2	2	5	0	8	0	13
		Total	24	6	8	0	38
4	1	1	23	7	0	0	30
	2	2	5	0	9	0	14
		Total	28	7	9	0	44
5	1	1	23	7	0	0	30
	2	2	5	0	9	0	14
		Total	28	7	9	0	44
6	1	1	23	7	0	0	30
	2	2	6	0	10	0	16
		Total	29	7	10	0	46
7	1	1	25	7	0	0	32
	2	2	6	0	10	0	16
		Total	31	7	10	0	48
8	1	1	25	7	0	0	32
	2	2	6	0	10	0	16
		Total	31	7	10	0	48

Table A3. ELA Estimated Time on Task by Session

Grade	Day	Session	Estimated Time on Task (min.)
3	1	1	38
	2	2	25
		Total	**63**
4	1	1	38
	2	2	50
		Total	**88**
5	1	1	40
	2	2	51
		Total	**91**
6	1	1	40
	2	2	51
		Total	**91**
7	1	1	40
	2	2	63
		Total	**103**
8	1	1	40
	2	2	63
		Total	**103**

Source: *2022 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 item—1 minute
- Average time to respond to a 2-point constructed-response item—3 minutes
- Average time to respond to a 4-point constructed-response item—20 minutes

Table A4. Mathematics Estimated Time on Task by Session

Grade	Day	Session	Estimated Time on Task (min.)
3	1	1	28.5
	2	2	45.5
		Total	**74**
4	1	1	34.5
	2	2	50.5
		Total	**85**
5	1	1	34.5
	2	2	50.5
		Total	**85**
6	1	1	34.5
	2	2	57
		Total	**91.5**
7	1	1	37.5
	2	2	57
		Total	**94.5**

Grade	Day	Session	Estimated Time on Task (min.)
8	1	1	37.5
	2	2	57
		Total	**94.5**

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

- Average time to respond to a multiple-choice item—1.5 minutes
- Average time to respond to a 1-point constructed-response item – 3 minutes
- Average time to respond to a 2-point constructed-response item—5 minutes
- Average time to respond to a 3-point constructed-response item—9 minutes

The testing times listed above do not include approximately 10 minutes reserved for preparation at the beginning of each session for handing out materials and reading directions. Additional details on security, scheduling, classroom organization and preparation, test materials, and administration can be found in the 2023 Teacher's Directions manuals located at https://www.nysed.gov/state-assessment/2023-grades-3-8-elementary-level-and-intermediate-level-field-tests-manuals-and and the 2023 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual* (*SAM*) located at https://www.nysed.gov/sites/default/files/programs/state-assessment/38-sam-2023.pdf.

Appendix B: ELA and Mathematics Test Blueprints

Table B1. ELA Test Blueprint

Grade	Total Points on OP Test	Strand	Point Range		% of Test	
			Target	Actual	Target	Actual
3	33	Literature	17	21	53%	64%
		Informational Text	16	12	47%	36%
4	37	Literature	17–20	12	47%–53%	32%
		Informational Text	17–20	25	47%–53%	68%
5	40	Literature	18–22	24	45%–55%	60%
		Informational Text	18–22	16	45%–55%	40%
6	40	Literature	18–22	18	45%–55%	45%
		Informational Text	18–22	22	45%–55%	55%
7	47	Literature	20–25	24	43%–57%	51%
		Informational Text	20–25	23	43%–57%	49%
8	47	Literature	20–25	24	43%–57%	51%
		Informational Text	20–25	23	43%–57%	49%

Table B2. Mathematics Test Blueprint

Grade	Total Points on OP Test	Domain	Point Range		% of Test	
			Target	Actual	Target	Actual
3	38	Operations and Algebraic Thinking	12–16	12	31%–43%	32%
		Number and Operations in Base Ten	3–5	4	7%–14%	10%
		Number and Operations – Fractions	7–11	8	18%–29%	21%
		Measurement and Data	8–12	11	21%–32%	29%
		Geometry [a]	1–3	3	2%–8%	8%
4	44	Operations and Algebraic Thinking	7–11	9	15%–25%	20%
		Number and Operations in Base Ten	9–13	10	20%–30%	23%
		Number and Operations – Fractions	9–13	11	20%–30%	25%
		Measurement and Data	4–6	6	9%–14%	14%
		Geometry	6–10	8	13%–23%	18%
		Number and Operations in Base Ten	11–15	13	25%–35%	30%
		Number and Operations – Fractions	15–19	16	34%–44%	36%
		Measurement and Data	10–14	13	22%–32%	30%
		Geometry [a]	1–3	2	2%–7%	5%
5	44	Ratios and Proportional Relationships	10–14	12	21%–30%	26%
		The Number System	8–12	9	17%–26%	19%
6	47	Expressions and Equations	12–20	18	25%–43%	38%
		Geometry	6–11	8	14%–24%	17%

Grade	Total Points on OP Test	Domain	Point Range		% of Test	
			Target	Actual	Target	Actual
7	49	Ratios and Proportional Relationships	12–16	14	24%–33%	29%
		The Number System	8–12	10	16%–25%	20%
		Expressions and Equations	13–19	15	26%–39%	31%
		Geometry	1–3	3	2%–7%	6%
		Statistics and Probability	6–10	7	12%–21%	14%
8	49	The Number System	1-4	2	2% - 9%	4%
		Expressions and Equations	14–20	13	28%–41%	27%
		Functions	8–12	12	16%–25%	25%
		Geometry	14–20	19	28%–41%	39%
		Statistics and Probability	2–5	3	4%–11%	6%

[a] There is a slight difference between the "Target % of Test" shown in these tables and the tables presented in the guides to the 2023 Mathematics Tests. The guides were intended to provide general guidance regarding content coverage of mathematics domains so that classroom instruction would continue to cover the depth and breadth of the mathematics standards.

Appendix C: Passage Selection Guidelines for Assessing ELA

General Guidelines

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

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

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

- Text complexity
- Text types
- Text suitability for specific standards

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

Appendix D: Universal Design Item Checklist

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

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

Universal Design Item Checklist	
√	The item is written with simplified sentences.
√	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 Grades 3–8 ELA and Mathematics Test items.

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 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 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 being tested
- presented at a reading level suitable for the grade being tested
- appropriate in context
- written so that a student possessing the 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 2023 Operational Test. The psychometrics department provides support, as necessary, and reviews the final item selection. The psychometrics department provides data files with parameters for all FT items eligible for the item pool. The pools of items eligible for 2023 item selection included 2019–2022 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. Keep in mind 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 provides 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 flagged DIF items that do not exhibit bias.
- Provide 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

Appendix G: Operational Item Maps

The following tables show the operational item maps for the 2023 NYSTP Grades 3–8 ELA and Mathematics Tests. Field test items that do not contribute to students' scores have been omitted. Additional details on the standards to which these items align may be found at http://www.nysed.gov/next-generation-learning-standards.

Table G1. ELA Grade 3 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Literature	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.4	Reading Standards for Literature	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.8	Reading Standards for Literature	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.5	Reading Standards for Literature	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Literature	Reading
7	Multiple Choice	1	NGLS.ELA.Content.NY-3.L.4	Language Standards	Reading
8	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Informational Text	Reading
9	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Informational Text	Reading
10	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.7	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Informational Text	Reading
12	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.5	Reading Standards for Informational Text	Reading
19	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.6	Reading Standards for Literature	Reading
20	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.8	Reading Standards for Literature	Reading
21	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.4	Reading Standards for Literature	Reading
23	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.9	Reading Standards for Literature	Reading
24	Constructed Response	2	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Writing to Sources
25	Constructed Response	2	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Literature	Writing to Sources
Session 2					
26	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Reading

Item	Type	Points	Standard	Strand	Subscore Category
27	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.5	Reading Standards for Literature	Reading
28	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.4	Reading Standards for Literature	Reading
29	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Literature	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.6	Reading Standards for Literature	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Literature	Reading
32	Constructed Response	2	NGLS.ELA.Content.NY-3.R.2	Reading Standards for Informational Text	Writing to Sources
33	Constructed Response	2	NGLS.ELA.Content.NY-3.R.3	Reading Standards for Informational Text	Writing to Sources
34	Constructed Response	2	NGLS.ELA.Content.NY-3.R.8	Reading Standards for Informational Text	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2023 Operational ELA Test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response items, please refer to the rubrics shown in the Educator Guides.

Table G2. ELA Grade 4 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.4	Reading Standards for Literature	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Literature	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.5	Reading Standards for Literature	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Literature	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Literature	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.8	Reading Standards for Literature	Reading
7	Multiple Choice	1	NGLS.ELA.Content.NY-4.L.4	Language Standards	Reading
8	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Literature	Reading
9	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Literature	Reading
10	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.5	Reading Standards for Literature	Reading
11	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Literature	Reading
12	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.9	Reading Standards for Literature	Reading

Item	Type	Points	Standard	Strand	Subscore Category
19	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.4	Reading Standards for Informational Text	Reading
20	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Informational Text	Reading
21	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Informational Text	Reading
23	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.5	Reading Standards for Informational Text	Reading
24	Constructed Response	2	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Informational Text	Writing to Sources
25	Constructed Response	2	NGLS.ELA.Content.NY-4.R.8	Reading Standards for Informational Text	Writing to Sources
Session 2					
26	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.7	Reading Standards for Informational Text	Reading
27	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.8	Reading Standards for Informational Text	Reading
28	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.5	Reading Standards for Informational Text	Reading
29	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Informational Text	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Informational Text	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Informational Text	Reading
32	Constructed Response	2	NGLS.ELA.Content.NY-4.R.9	Reading Standards for Informational Text	Writing to Sources
33	Constructed Response	2	NGLS.ELA.Content.NY-4.R.2	Reading Standards for Informational Text	Writing to Sources
34	Constructed Response	2	NGLS.ELA.Content.NY-4.R.4	Reading Standards for Informational Text	Writing to Sources
35	Constructed Response	4	NGLS.ELA.Content.NY-4.R.3	Reading Standards for Informational Text	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2023 Operational ELA Test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response items, please refer to the rubrics shown in the Educator Guides.

Table G3. ELA Grade 5 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Literature	Reading

Item	Type	Points	Standard	Strand	Subscore Category
3	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.8	Reading Standards for Literature	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.2	Reading Standards for Literature	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.5	Reading Standards for Literature	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Reading
7	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.6	Reading Standards for Literature	Reading
8	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Informational Text	Reading
9	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Informational Text	Reading
10	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.9	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Informational Text	Reading
12	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Informational Text	Reading
13	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.8	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.7	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Informational Text	Reading
23	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Informational Text	Reading
24	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.2	Reading Standards for Informational Text	Reading
25	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Informational Text	Reading
26	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.6	Reading Standards for Informational Text	Reading
27	Constructed Response	2	NGLS.ELA.Content.NY-5.R.8	Reading Standards for Informational Text	Writing to Sources
28	Constructed Response	2	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Informational Text	Writing to Sources
Session 2					
29	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Literature	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-5.L.4	Language Standards	Reading
32	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.5	Reading Standards for Literature	Reading
33	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.2	Reading Standards for Literature	Reading
34	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Reading
35	Multiple Choice	1	NGLS.ELA.Content.NY-5.R.3	Reading Standards for Literature	Reading
36	Constructed Response	2	NGLS.ELA.Content.NY-5.R.6	Reading Standards for Literature	Writing to Sources
37	Constructed Response	2	NGLS.ELA.Content.NY-5.R.4	Reading Standards for Literature	Writing to Sources

Item	Type	Points	Standard	Strand	Subscore Category
38	Constructed Response	2	NGLS.ELA.Content.NY-5.R.8	Reading Standards for Literature	Writing to Sources
39	Constructed Response	4	NGLS.ELA.Content.NY-5.R.2	Reading Standards for Literature	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2023 Operational ELA Test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response items, please refer to the rubrics shown in the Educator Guides.

Table G4. ELA Grade 6 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.8	Reading Standards for Literature	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Literature	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.5	Reading Standards for Literature	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.2	Reading Standards for Literature	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-6.L.4	Language Standards	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Literature	Reading
7	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.6	Reading Standards for Literature	Reading
8	Multiple Choice	1	NGLS.ELA.Content.NY-6.L.4	Language Standards	Reading
9	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.5	Reading Standards for Informational Text	Reading
10	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Informational Text	Reading
11	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.7	Reading Standards for Informational Text	Reading
12	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Informational Text	Reading
13	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.2	Reading Standards for Informational Text	Reading
14	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.9	Reading Standards for Informational Text	Reading
23	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.5	Reading Standards for Informational Text	Reading
24	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.7	Reading Standards for Informational Text	Reading
25	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Informational Text	Reading
26	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.2	Reading Standards for Informational Text	Reading

Item	Type	Points	Standard	Strand	Subscore Category
27	Constructed Response	2	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Informational Text	Writing to Sources
28	Constructed Response	2	NGLS.ELA.Content.NY-6.R.2	Reading Standards for Informational Text	Writing to Sources
Session 2					
29	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Informational Text	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Informational Text	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.9	Reading Standards for Informational Text	Reading
32	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.6	Reading Standards for Informational Text	Reading
33	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Informational Text	Reading
34	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.3	Reading Standards for Informational Text	Reading
35	Multiple Choice	1	NGLS.ELA.Content.NY-6.R.8	Reading Standards for Informational Text	Reading
36	Constructed Response	2	NGLS.ELA.Content.NY-6.R.6	Reading Standards for Literature	Writing to Sources
37	Constructed Response	2	NGLS.ELA.Content.NY-6.R.4	Reading Standards for Literature	Writing to Sources
38	Constructed Response	2	NGLS.ELA.Content.NY-6.R.2	Reading Standards for Literature	Writing to Sources
39	Constructed Response	4	NGLS.ELA.Content.NY-6.R.9	Reading Standards for Literature	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2023 Operational ELA Test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response items, please refer to the rubrics shown in the Educator Guides.

Table G5. ELA Grade 7 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Literature	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-7.L.4	Language Standards	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.6	Reading Standards for Literature	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.5	Reading Standards for Literature	Reading
7	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Literature	Reading

Item	Type	Points	Standard	Strand	Subscore Category
15	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Informational Text	Reading
16	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.9	Reading Standards for Informational Text	Reading
17	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.6	Reading Standards for Informational Text	Reading
18	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Informational Text	Reading
19	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Informational Text	Reading
20	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Informational Text	Reading
21	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.6	Reading Standards for Literature	Reading
23	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.5	Reading Standards for Literature	Reading
24	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Literature	Reading
25	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Literature	Reading
26	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading
27	Constructed Response	2	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Writing to Sources
28	Constructed Response	2	NGLS.ELA.Content.NY-7.R.8	Reading Standards for Literature	Writing to Sources
Session 2					
29	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Literature	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Literature	Reading
32	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Literature	Reading
33	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.9	Reading Standards for Literature	Reading
34	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Literature	Reading
35	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.6	Reading Standards for Literature	Reading
36	Multiple Choice	1	NGLS.ELA.Content.NY-7.L.4	Language Standards	Reading
37	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Informational Text	Reading
38	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.5	Reading Standards for Informational Text	Reading
39	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Informational Text	Reading
40	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.8	Reading Standards for Informational Text	Reading
41	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.4	Reading Standards for Informational Text	Reading
42	Multiple Choice	1	NGLS.ELA.Content.NY-7.R.6	Reading Standards for Informational Text	Reading

Item	Type	Points	Standard	Strand	Subscore Category
43	Constructed Response	2	NGLS.ELA.Content.NY-7.R.6	Reading Standards for Informational Text	Writing to Sources
44	Constructed Response	2	NGLS.ELA.Content.NY-7.R.2	Reading Standards for Informational Text	Writing to Sources
45	Constructed Response	2	NGLS.ELA.Content.NY-7.R.3	Reading Standards for Informational Text	Writing to Sources
46	Constructed Response	4	NGLS.ELA.Content.NY-7.R.8	Reading Standards for Informational Text	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2023 Operational ELA Test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response items, please refer to the rubrics shown in the Educator Guides.

Table G6. ELA Grade 8 Operational Item Map

Item	Type	Points	Standard	Strand	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Literature	Reading
2	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Literature	Reading
3	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Reading
4	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Reading
5	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Reading
6	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.6	Reading Standards for Literature	Reading
7	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.9	Reading Standards for Literature	Reading
15	Multiple Choice	1	NGLS.ELA.Content.NY-8.L.4	Language Standards	Reading
16	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Informational Text	Reading
17	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
18	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.8	Reading Standards for Informational Text	Reading
19	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Informational Text	Reading
20	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
21	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
22	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Literature	Reading
23	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Reading
24	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Literature	Reading

Item	Type	Points	Standard	Strand	Subscore Category
25	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.6	Reading Standards for Literature	Reading
26	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Literature	Reading
27	Constructed Response	2	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Writing to Sources
28	Constructed Response	2	NGLS.ELA.Content.NY-8.R.9	Reading Standards for Literature	Writing to Sources
Session 2					
29	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Literature	Reading
30	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Literature	Reading
31	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Literature	Reading
32	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Reading
33	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.9	Reading Standards for Literature	Reading
34	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Literature	Reading
35	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Literature	Reading
36	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Informational Text	Reading
37	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.8	Reading Standards for Informational Text	Reading
38	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.6	Reading Standards for Informational Text	Reading
39	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
40	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.9	Reading Standards for Informational Text	Reading
41	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Reading
42	Multiple Choice	1	NGLS.ELA.Content.NY-8.R.2	Reading Standards for Informational Text	Reading
43	Constructed Response	2	NGLS.ELA.Content.NY-8.R.4	Reading Standards for Informational Text	Writing to Sources
44	Constructed Response	2	NGLS.ELA.Content.NY-8.R.6	Reading Standards for Informational Text	Writing to Sources
45	Constructed Response	2	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Writing to Sources
46	Constructed Response	4	NGLS.ELA.Content.NY-8.R.3	Reading Standards for Informational Text	Writing to Sources

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each question on the 2023 Operational ELA Test. However, each constructed-response item measures proficiencies described in multiple standards, including writing and additional reading and language standards. For example, 2-point and 4-point constructed-response items require students to first conduct the analyses described in the mapped standard and then produce written responses that are rated based on writing standards. To gain greater insight into the measurement focus for constructed-response items, please refer to the rubrics shown in the Educator Guides.

Table G7. Mathematics Grade 3 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.4	Operations and Algebraic Thinking	Operations and Algebraic Thinking
2	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.6	Measurement and Data	Measurement and Data
3	Multiple Choice	1	NGLS.Math.Content.NY-3.NBT.3	Number and Operations in Base Ten	
5	Multiple Choice	1	NGLS.Math.Content.NY-3.NF.3a	Number and Operations - Fractions	Number and Operations - Fractions
6	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.2b	Measurement and Data	Measurement and Data
8	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.8a	Operations and Algebraic Thinking	Operations and Algebraic Thinking
9	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.1	Measurement and Data	Measurement and Data
10	Multiple Choice	1	NGLS.Math.Content.NY-3.NBT.1	Number and Operations in Base Ten	
12	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking
14	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.7a	Measurement and Data	Measurement and Data
15	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.9	Operations and Algebraic Thinking	Operations and Algebraic Thinking
16	Multiple Choice	1	NGLS.Math.Content.NY-3.NF.3c	Number and Operations - Fractions	Number and Operations - Fractions
18	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.7d	Measurement and Data	Measurement and Data
19	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.5	Operations and Algebraic Thinking	Operations and Algebraic Thinking
20	Multiple Choice	1	NGLS.Math.Content.NY-3.NF.2a	Number and Operations - Fractions	Number and Operations - Fractions

Item	Type	Points	Standard	Cluster	Subscore Category
22	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking
23	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.5a	Measurement and Data	Measurement and Data
24	Multiple Choice	1	NGLS.Math.Content.NY-3.NF.2b	Number and Operations - Fractions	Number and Operations - Fractions
25	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.6	Operations and Algebraic Thinking	Operations and Algebraic Thinking
Session 2					
26	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.8a	Operations and Algebraic Thinking	Operations and Algebraic Thinking
27	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.5b	Measurement and Data	Measurement and Data
28	Multiple Choice	1	NGLS.Math.Content.NY-3.OA.3	Operations and Algebraic Thinking	Operations and Algebraic Thinking
29	Multiple Choice	1	NGLS.Math.Content.NY-3.NF.3b	Number and Operations - Fractions	Number and Operations - Fractions
30	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.7c	Measurement and Data	Measurement and Data
31	Constructed Response	1	NGLS.Math.Content.NY-3.OA.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
32	Constructed Response	1	NGLS.Math.Content.NY-3.G.2	Geometry	
33	Constructed Response	1	NGLS.Math.Content.NY-3.MD.7b	Measurement and Data	Measurement and Data
34	Constructed Response	2	NGLS.Math.Content.NY-3.G.2	Geometry	
35	Constructed Response	2	NGLS.Math.Content.NY-3.OA.9	Operations and Algebraic Thinking	Operations and Algebraic Thinking
36	Constructed Response	2	NGLS.Math.Content.NY-3.MD.1	Measurement and Data	Measurement and Data
37	Constructed Response	2	NGLS.Math.Content.NY-3.NBT.4a	Number and Operations in Base Ten	

Item	Type	Points	Standard	Cluster	Subscore Category
38	Constructed Response	3	NGLS.Math.Content.NY-3.NF.3d	Number and Operations - Fractions	Number and Operations – Fractions

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each item. However, some items measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Table G8. Mathematics Grade 4 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.2a	Number and Operations in Base Ten	Number and Operations in Base Ten
2	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking
4	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.1	Number and Operations - Fractions	Number and Operations - Fractions
5	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.5	Number and Operations in Base Ten	Number and Operations in Base Ten
6	Multiple Choice	1	NGLS.Math.Content.NY-4.G.2a	Geometry	
8	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
9	Multiple Choice	1	NGLS.Math.Content.NY-3.MD.4	Measurement and Data	
10	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.3c	Number and Operations - Fractions	Number and Operations - Fractions
11	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.4	Operations and Algebraic Thinking	Operations and Algebraic Thinking
13	Multiple Choice	1	NGLS.Math.Content.NY-4.MD.4	Measurement and Data	
14	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking
15	Multiple Choice	1	NGLS.Math.Content.NY-4.G.2b	Geometry	

Item	Type	Points	Standard	Cluster	Subscore Category
17	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.5	Operations and Algebraic Thinking	Operations and Algebraic Thinking
18	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
19	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.4b	Number and Operations - Fractions	Number and Operations - Fractions
20	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
22	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.1	Number and Operations in Base Ten	Number and Operations in Base Ten
23	Multiple Choice	1	NGLS.Math.Content.NY-4.G.3	Geometry	
25	Multiple Choice	1	NGLS.Math.Content.NY-4.MD.6	Measurement and Data	
26	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.2b	Number and Operations in Base Ten	Number and Operations in Base Ten
28	Multiple Choice	1	NGLS.Math.Content.NY-3.G.1	Geometry	
29	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.1	Number and Operations - Fractions	Number and Operations - Fractions
30	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
Session 2					
31	Multiple Choice	1	NGLS.Math.Content.NY-4.G.3	Geometry	
32	Multiple Choice	1	NGLS.Math.Content.NY-4.NF.2	Number and Operations - Fractions	Number and Operations - Fractions
33	Multiple Choice	1	NGLS.Math.Content.NY-4.G.1	Geometry	
34	Multiple Choice	1	NGLS.Math.Content.NY-4.OA.3a	Operations and Algebraic Thinking	Operations and Algebraic Thinking
35	Multiple Choice	1	NGLS.Math.Content.NY-4.NBT.5	Number and Operations in Base Ten	Number and Operations in Base Ten
36	Constructed Response	1	NGLS.Math.Content.NY-4.NF.3d	Number and Operations - Fractions	Number and Operations - Fractions

Item	Type	Points	Standard	Cluster	Subscore Category
37	Constructed Response	1	NGLS.Math.Content.NY-4.NBT.3	Number and Operations in Base Ten	Number and Operations in Base Ten
38	Constructed Response	1	NGLS.Math.Content.NY-4.MD.5a	Measurement and Data	
39	Constructed Response	2	NGLS.Math.Content.NY-4.G.2c	Geometry	
40	Constructed Response	2	NGLS.Math.Content.NY-3.MD.8b	Measurement and Data	
41	Constructed Response	2	NGLS.Math.Content.NY-4.NF.3b	Number and Operations - Fractions	Number and Operations - Fractions
42	Constructed Response	2	NGLS.Math.Content.NY-4.OA.2	Operations and Algebraic Thinking	Operations and Algebraic Thinking
43	Constructed Response	2	NGLS.Math.Content.NY-4.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
44	Constructed Response	3	NGLS.Math.Content.NY-4.NF.4c	Number and Operations - Fractions	Number and Operations - Fractions

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each item. However, some items measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Table G9. Mathematics Grade 5 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.2	Number and Operations - Fractions	Number and Operations - Fractions
2	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.3a	Number and Operations in Base Ten	Number and Operations in Base Ten
3	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.5b	Measurement and Data	Measurement and Data
5	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.1	Number and Operations - Fractions	Number and Operations - Fractions
6	Multiple Choice	1	NGLS.Math.Content.NY-4.MD.1	Measurement and Data	Measurement and Data

Item	Type	Points	Standard	Cluster	Subscore Category
8	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.2	Number and Operations in Base Ten	Number and Operations in Base Ten
9	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.4b	Number and Operations - Fractions	Number and Operations - Fractions
10	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
11	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.3	Number and Operations - Fractions	Number and Operations - Fractions
13	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.6	Number and Operations - Fractions	Number and Operations - Fractions
14	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.1	Number and Operations in Base Ten	Number and Operations in Base Ten
16	Multiple Choice	1	NGLS.Math.Content.NY-5.G.4	Geometry	
17	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.5a	Number and Operations - Fractions	Number and Operations - Fractions
18	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.4	Measurement and Data	Measurement and Data
20	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.7	Number and Operations in Base Ten	Number and Operations in Base Ten
21	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.6	Number and Operations - Fractions	Number and Operations - Fractions
22	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.4	Number and Operations in Base Ten	Number and Operations in Base Ten
24	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.1	Number and Operations - Fractions	Number and Operations - Fractions
25	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.5a	Measurement and Data	Measurement and Data
26	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.7	Number and Operations in Base Ten	Number and Operations in Base Ten
28	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.7c	Number and Operations - Fractions	Number and Operations - Fractions
29	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.2	Number and Operations - Fractions	Number and Operations - Fractions

Item	Type	Points	Standard	Cluster	Subscore Category
30	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.1	Measurement and Data	Measurement and Data
Session 2					
31	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.4	Measurement and Data	Measurement and Data
32	Multiple Choice	1	NGLS.Math.Content.NY-5.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
33	Multiple Choice	1	NGLS.Math.Content.NY-5.NF.2	Number and Operations - Fractions	Number and Operations - Fractions
34	Multiple Choice	1	NGLS.Math.Content.NY-5.G.4	Geometry	
35	Multiple Choice	1	NGLS.Math.Content.NY-5.MD.1	Measurement and Data	Measurement and Data
36	Constructed Response	1	NGLS.Math.Content.NY-5.NBT.3b	Number and Operations in Base Ten	Number and Operations in Base Ten
37	Constructed Response	1	NGLS.Math.Content.NY-5.NF.7c	Number and Operations - Fractions	Number and Operations - Fractions
38	Constructed Response	1	NGLS.Math.Content.NY-5.MD.5b	Measurement and Data	Measurement and Data
39	Constructed Response	2	NGLS.Math.Content.NY-5.NBT.6	Number and Operations in Base Ten	Number and Operations in Base Ten
40	Constructed Response	2	NGLS.Math.Content.NY-5.NF.5b	Number and Operations - Fractions	Number and Operations - Fractions
41	Constructed Response	2	NGLS.Math.Content.NY-5.MD.5c	Measurement and Data	Measurement and Data
42	Constructed Response	2	NGLS.Math.Content.NY-5.NBT.3a	Number and Operations in Base Ten	Number and Operations in Base Ten
43	Constructed Response	2	NGLS.Math.Content.NY-5.NF.7c	Number and Operations - Fractions	Number and Operations - Fractions
44	Constructed Response	3	NGLS.Math.Content.NY-5.MD.2	Measurement and Data	Measurement and Data

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each item. However, some items measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Table G10. Mathematics Grade 6 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.5	Expressions and Equations	Expressions and Equations
2	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.6c	The Number System	The Number System
3	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.9	Expressions and Equations	Expressions and Equations
5	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.2	Ratios and Proportional Relationships	Ratios and Proportional Relationships
6	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.7a	The Number System	The Number System
7	Multiple Choice	1	NGLS.Math.Content.NY-6.G.4	Geometry	
9	Multiple Choice	1	NGLS.Math.Content.NY-5.OA.1	Expressions and Equations	Expressions and Equations
10	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.3d	Ratios and Proportional Relationships	Ratios and Proportional Relationships
12	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.2a	Expressions and Equations	Expressions and Equations
13	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.4	The Number System	The Number System
14	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.1	Expressions and Equations	Expressions and Equations
15	Multiple Choice	1	NGLS.Math.Content.NY-6.G.1	Geometry	
17	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.7c	The Number System	The Number System
18	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.2b	Expressions and Equations	Expressions and Equations
19	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.1	The Number System	The Number System
21	Multiple Choice	1	NGLS.Math.Content.NY-6.G.3	Geometry	

Item	Type	Points	Standard	Cluster	Subscore Category
22	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.2c	Expressions and Equations	Expressions and Equations
23	Multiple Choice	1	NGLS.Math.Content.NY-6.NS.7d	The Number System	The Number System
24	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.3	Expressions and Equations	Expressions and Equations
26	Multiple Choice	1	NGLS.Math.Content.NY-6.G.4	Geometry	
27	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.6	Expressions and Equations	Expressions and Equations
29	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.3a	Ratios and Proportional Relationships	Ratios and Proportional Relationships
30	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.8	Expressions and Equations	Expressions and Equations
Session 2					
31	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.3c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
32	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.3a	Ratios and Proportional Relationships	Ratios and Proportional Relationships
33	Multiple Choice	1	NGLS.Math.Content.NY-6.G.2	Geometry	
34	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.3b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
35	Multiple Choice	1	NGLS.Math.Content.NY-6.EE.3	Expressions and Equations	Expressions and Equations
36	Multiple Choice	1	NGLS.Math.Content.NY-6.RP.3b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
37	Constructed Response	1	NGLS.Math.Content.NY-6.G.1	Geometry	
38	Constructed Response	1	NGLS.Math.Content.NY-6.NS.7b	The Number System	The Number System
39	Constructed Response	1	NGLS.Math.Content.NY-6.RP.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships

Item	Type	Points	Standard	Cluster	Subscore Category
40	Constructed Response	2	NGLS.Math.Content.NY-6.EE.7	Expressions and Equations	Expressions and Equations
41	Constructed Response	2	NGLS.Math.Content.NY-6.RP.2	Ratios and Proportional Relationships	Ratios and Proportional Relationships
42	Constructed Response	2	NGLS.Math.Content.NY-6.NS.1	The Number System	The Number System
43	Constructed Response	2	NGLS.Math.Content.NY-6.EE.1	Expressions and Equations	Expressions and Equations
44	Constructed Response	2	NGLS.Math.Content.NY-6.RP.3b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
45	Constructed Response	2	NGLS.Math.Content.NY-6.G.2	Geometry	
46	Constructed Response	3	NGLS.Math.Content.NY-6.EE.9	Expressions and Equations	Expressions and Equations

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each item. However, some items measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Table G11. Mathematics Grade 7 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.2c	The Number System	The Number System
2	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.3	Expressions and Equations	Expressions and Equations
3	Multiple Choice	1	NGLS.Math.Content.NY-7.G.1	Geometry	
4	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.2b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
5	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.1d	The Number System	The Number System
6	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.4a	Expressions and Equations	Expressions and Equations

Item	Type	Points	Standard	Cluster	Subscore Category
8	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.3	The Number System	The Number System
9	Multiple Choice	1	NGLS.Math.Content.NY-6.SP.5c	Statistics and Probability	
10	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.1	Expressions and Equations	Expressions and Equations
12	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.1c	The Number System	The Number System
13	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.2a	Ratios and Proportional Relationships	Ratios and Proportional Relationships
15	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.1	Expressions and Equations	Expressions and Equations
16	Multiple Choice	1	NGLS.Math.Content.NY-7.SP.4	Statistics and Probability	
17	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.4a	Expressions and Equations	Expressions and Equations
18	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.3	The Number System	The Number System
20	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.2b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
21	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.2	Expressions and Equations	Expressions and Equations
23	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships
24	Multiple Choice	1	NGLS.Math.Content.NY-6.SP.3	Statistics and Probability	
25	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.2c	Ratios and Proportional Relationships	Ratios and Proportional Relationships
27	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.2	Expressions and Equations	Expressions and Equations
29	Multiple Choice	1	NGLS.Math.Content.NY-7.NS.2d	The Number System	The Number System
30	Multiple Choice	1	NGLS.Math.Content.NY-7.EE.4b	Expressions and Equations	Expressions and Equations

Item	Type	Points	Standard	Cluster	Subscore Category
31	Multiple Choice	1	NGLS.Math.Content.NY-6.SP.5a	Statistics and Probability	
32	Multiple Choice	1	NGLS.Math.Content.NY-7.RP.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships
Session 2					
33	Multiple Choice	1	CCSS.Math.Content.3.OA.A.1	Operations and Algebraic Thinking	Operations and Algebraic Thinking
34	Multiple Choice	1	CCSS.Math.Content.3.NF.A.2 b	Number and Operations - Fractions	Number and Operations - Fractions
35	Multiple Choice	1	CCSS.Math.Content.3.MD.C.7d	Measurement and Data	Measurement and Data
36	Multiple Choice	1	CCSS.Math.Content.3.MD.A.2	Measurement and Data	Measurement and Data
37	Multiple Choice	1	CCSS.Math.Content.3.G.A.2	Geometry	
38	Multiple Choice	1	CCSS.Math.Content.3.OA.D.8	Operations and Algebraic Thinking	Operations and Algebraic Thinking
39	Constructed Response	1	NGLS.Math.Content.NY-7.RP.1	Ratios and Proportional Relationships	Ratios and Proportional Relationships
40	Constructed Response	1	NGLS.Math.Content.NY-7.SP.8a	Statistics and Probability	
41	Constructed Response	1	NGLS.Math.Content.NY-7.EE.3	Expressions and Equations	Expressions and Equations
42	Constructed Response	1	NGLS.Math.Content.NY-7.NS.1b	The Number System	The Number System
43	Constructed Response	1	NGLS.Math.Content.NY-7.EE.4b	Expressions and Equations	Expressions and Equations
44	Constructed Response	1	NGLS.Math.Content.NY-7.RP.3	Ratios and Proportional Relationships	Ratios and Proportional Relationships
45	Constructed Response	1	NGLS.Math.Content.NY-7.RP.2b	Ratios and Proportional Relationships	Ratios and Proportional Relationships
46	Constructed Response	1	NGLS.Math.Content.NY-7.NS.3	The Number System	The Number System

Item	Type	Points	Standard	Cluster	Subscore Category
47	Constructed Response	1	NGLS.Math.Content.NY-7.EE.1	Expressions and Equations	Expressions and Equations
48	Constructed Response	2	NGLS.Math.Content.NY-7.RP.2c	Ratios and Proportional Relationships	Ratios and Proportional Relationships

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each item. However, some items measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Table G12. Mathematics Grade 8 Operational Item Map

Item	Type	Points	Standard	Cluster	Subscore Category
Session 1					
1	Multiple Choice	1	NGLS.Math.Content.NY-8.NS.2	The Number System	
2	Multiple Choice	1	NGLS.Math.Content.NY-8.SP.1	Statistics and Probability	
3	Multiple Choice	1	NGLS.Math.Content.NY-8.G.3	Geometry	Geometry
4	Multiple Choice	1	NGLS.Math.Content.NY-7.G.5	Geometry	Expressions and Equations
5	Multiple Choice	1	NGLS.Math.Content.NY-8.F.2	Functions	Functions
7	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.5	Expressions and Equations	Expressions and Equations
8	Multiple Choice	1	NGLS.Math.Content.NY-8.G.8	Geometry	Geometry
9	Multiple Choice	1	NGLS.Math.Content.NY-8.F.3	Functions	Functions
11	Multiple Choice	1	NGLS.Math.Content.NY-8.SP.2	Statistics and Probability	
12	Multiple Choice	1	NGLS.Math.Content.NY-8.F.1	Functions	Functions
14	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.7a	Expressions and Equations	Expressions and Equations
15	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.1	Expressions and Equations	Expressions and Equations
16	Multiple Choice	1	NGLS.Math.Content.NY-8.G.9	Geometry	Geometry
18	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.2	Expressions and Equations	Expressions and Equations

Item	Type	Points	Standard	Cluster	Subscore Category
19	Multiple Choice	1	NGLS.Math.Content.NY-8.G.2	Geometry	Geometry
20	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.6	Expressions and Equations	Expressions and Equations
22	Multiple Choice	1	NGLS.Math.Content.NY-7.G.2	Geometry	Expressions and Equations
23	Multiple Choice	1	NGLS.Math.Content.NY-8.SP.3	Statistics and Probability	
24	Multiple Choice	1	NGLS.Math.Content.NY-8.G.1a	Geometry	Geometry
25	Multiple Choice	1	NGLS.Math.Content.NY-8.G.5	Geometry	Geometry
26	Multiple Choice	1	NGLS.Math.Content.NY-8.F.2	Functions	Functions
28	Multiple Choice	1	NGLS.Math.Content.NY-8.F.4	Functions	Functions
29	Multiple Choice	1	NGLS.Math.Content.NY-8.G.4	Geometry	Geometry
31	Multiple Choice	1	NGLS.Math.Content.NY-8.F.4	Functions	Functions
32	Multiple Choice	1	NGLS.Math.Content.NY-8.G.1c	Geometry	Geometry
Session 2					
33	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.6	Expressions and Equations	Expressions and Equations
34	Multiple Choice	1	NGLS.Math.Content.NY-8.EE.5	Expressions and Equations	Expressions and Equations
35	Multiple Choice	1	NGLS.Math.Content.NY-8.NS.1	The Number System	
36	Multiple Choice	1	NGLS.Math.Content.NY-8.F.1	Functions	Functions
37	Multiple Choice	1	NGLS.Math.Content.NY-8.G.9	Geometry	Geometry
38	Multiple Choice	1	NGLS.Math.Content.NY-8.G.3	Geometry	Geometry
39	Constructed Response	1	NGLS.Math.Content.NY-8.EE.2	Expressions and Equations	Expressions and Equations
40	Constructed Response	1	NGLS.Math.Content.NY-8.G.7	Geometry	Geometry
41	Constructed Response	1	NGLS.Math.Content.NY-8.EE.7b	Expressions and Equations	Expressions and Equations

Item	Type	Points	Standard	Cluster	Subscore Category
42	Constructed Response	2	NGLS.Math.Content.NY-8.G.2	Geometry	Geometry
43	Constructed Response	2	NGLS.Math.Content.NY-8.EE.7b	Expressions and Equations	Expressions and Equations
44	Constructed Response	2	NGLS.Math.Content.NY-8.G.6	Geometry	Geometry
45	Constructed Response	2	NGLS.Math.Content.NY-8.F.4	Functions	Functions
46	Constructed Response	2	NGLS.Math.Content.NY-7.G.4	Geometry	Geometry
47	Constructed Response	2	NGLS.Math.Content.NY-8.EE.1	Expressions and Equations	Expressions and Equations
48	Constructed Response	3	NGLS.Math.Content.NY-8.F.3	Functions	Functions

Note. This item map is intended to identify the primary analytic skills necessary to successfully answer each item. However, some items measure proficiencies described in multiple standards, including a balanced combination of procedural and conceptual understanding.

Appendix H: ELA Short-Response Rubric

2-Credit Rubric–Short Response

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

Note. If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 1. Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response item in that session completely blank (no response attempted).

Appendix I: ELA Extended-Response Rubrics

New York State Grade 4-5 Expository Writing Evaluation Rubric

CRITERIA	NGLS	SCORE				
		4 **Essays at this level:**	**3** **Essays at this level:**	**2** **Essays at this level:**	**1** **Essays at this level**	**0** **Essays at this level:**
CONTENT AND ANALYSIS: the extent to which the essay conveys ideas and information clearly and accurately in order to support an analysis of topics or text(s)	W.2 R.1–9	–clearly introduce a topic in a manner that follows logically from the task and purpose –demonstrate insightful comprehension and analysis of the text(s)	–clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate comprehension and analysis of the text(s)	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a literal comprehension of the text(s)	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text(s)	–demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided text(s) to support analysis and reflection	W.2 R.1–8	–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

CRITERIA	NGLS	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:**
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2 L.3 L.6	–exhibit clear, purposeful organization –skillfully link ideas using grade-appropriate words and phrases –use grade-appropriate, stylistically sophisticated language and domain-specific vocabulary –provide a concluding statement that follows clearly from the topic and information presented	–exhibit clear organization –link ideas using grade-appropriate words and phrases –use grade-appropriate precise language and domain-specific vocabulary –provide a concluding statement that follows from the topic and information presented	–exhibit some attempt at organization –inconsistently link ideas using words and phrases –inconsistently use appropriate language and domain-specific vocabulary –provide a concluding statement that follows generally from the topic and information presented	–exhibit little attempt at organization, or attempts to organize are irrelevant to the task –lack the use of linking words and phrases –use language that is imprecise or inappropriate for the text(s) and task –provide a concluding statement that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –exhibit no use of linking words and phrases –use language that is predominantly incoherent or copied directly from the text(s) –do not provide a concluding statement
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2 L.1 L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

Note. Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response item 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	NGS	SCORE				
		4 **Essays at this level:**	**3** **Essays at this level:**	**2** **Essays at this level:**	**1** **Essays at this level:**	**0** **Essays at this level:**
CONTENT AND ANALYSIS: the extent to which the essay conveys complex ideas and information clearly and accurately in order to support claims in an analysis of topics or text(s)	W.2, R.1–9	–clearly introduce a topic in a manner that is compelling and follows logically from the task and purpose –demonstrate insightful analysis of the text(s)	–clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate analysis of the text(s)	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a literal comprehension of the text(s)	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text(s)	–demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided text(s) to support analysis and reflection	W.2, R.1–8	–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

CRITERIA	NGS	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:**
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2, L.3, L.6	–exhibit clear organization, with the skillful use of appropriate and varied transitions to create a unified whole and enhance meaning –establish and maintain a formal style, using grade-appropriate, stylistically sophisticated language and domain-specific vocabulary with a notable sense of voice –provide a concluding statement or section that is compelling and follows clearly from the topic and information presented	–exhibit clear organization, with the use of appropriate transitions to create a unified whole –establish and maintain a formal style using precise language and domain-specific vocabulary –provide a concluding statement or section that follows from the topic and information presented	–exhibit some attempt at organization, with inconsistent use of transitions –establish but fail to maintain a formal style, with inconsistent use of language and domain-specific vocabulary –provide a concluding statement or section that follows generally from the topic and information presented	–exhibit little attempt at organization, or attempts to organize are irrelevant to the task –lack a formal style, using language that is imprecise or inappropriate for the text(s) and task –provide a concluding statement or section that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –use language that is predominantly incoherent or copied directly from the text(s) –do not provide a concluding statement or section
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2, L.1, L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

Note. Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response item 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 Rubrics

1-Credit Constructed-Response Rubric

1 Credit	A 1-credit response is a correct answer to the question that indicates a thorough understanding of mathematical concepts and/or procedures.
0 Credits	A 0-credit response is incorrect, irrelevant, or incoherent.

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

2-Credit-Constructed Response Holistic Rubric

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

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

Appendix K: Mathematics Extended-Response Rubric

3-Credit Constructed-Response Holistic Rubric

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

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

Appendix L: Factor Analysis Results for Selected Subgroups

As described in Section 3: Validity, a principal component factor analysis was conducted on the 2023 Grades 3–8 ELA and Mathematics Tests data. The analyses were conducted for the total population of students and select subgroups: English Language Learners (ELLs), students with disabilities (SWDs), and students using test accommodations (SUAs). Tables L1 through L12 contain the results of factor analysis on the subpopulation data for the Grades 3–8 ELA and Mathematics Tests, respectively.

Table L1. ELA Grade 3 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**6.08**	**21.72**	**21.72**
	2	1.37	4.89	26.61
	3	1.12	4.01	30.62
	4	1.06	3.78	34.4
SWD	1	**6.38**	**22.77**	**22.77**
	2	1.46	5.21	27.98
	3	1.11	3.95	31.93
	4	1.02	3.65	35.59
SUA	1	**6.14**	**21.92**	**21.92**
	2	1.5	5.34	27.26
	3	1.11	3.98	31.23
	4	1.04	3.71	34.94

Table L2. ELA Grade 4 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**5.05**	**17.42**	**17.42**
	2	1.35	4.66	22.08
	3	1.11	3.83	25.9
	4	1.11	3.81	29.71
	5	1.05	3.61	33.33
	6	1.04	3.6	36.93
	7	1.02	3.51	40.44
	8	1.01	3.47	43.91
	9	1	3.45	47.37
SWD	1	**6.08**	**20.98**	**20.98**
	2	1.33	4.58	25.56
	3	1.04	3.6	29.15
	4	1.02	3.52	32.67

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
	5	1	3.45	36.12
SUA	1	**5.91**	**20.36**	**20.36**
	2	1.35	4.64	25
	3	1.06	3.64	28.65
	4	1.03	3.55	32.2
	5	1	3.45	35.66

Table L3. ELA Grade 5 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**5.54**	**17.31**	**17.31**
	2	1.5	4.7	22.01
	3	1.18	3.68	25.7
	4	1.12	3.49	29.18
	5	1.06	3.31	32.5
	6	1.05	3.28	35.77
	7	1.03	3.21	38.99
	8	1.02	3.19	42.17
SWD	1	**5.89**	**18.41**	**18.41**
	2	1.49	4.65	23.06
	3	1.24	3.87	26.93
	4	1.05	3.27	30.2
	5	1.02	3.19	33.39
SUA	1	**5.86**	**18.3**	**18.3**
	2	1.5	4.68	22.98
	3	1.23	3.85	26.83
	4	1.05	3.29	30.12
	5	1.02	3.18	33.29

Table L4. ELA Grade 6 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**5.56**	**17.36**	**17.36**
	2	1.5	4.67	22.04
	3	1.13	3.53	25.57
	4	1.1	3.43	29
	5	1.08	3.39	32.39

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
	6	1.06	3.31	35.7
	7	1.03	3.22	38.92
	8	1.01	3.17	42.09
	9	1	3.14	45.23
SWD	1	**5.87**	**18.33**	**18.33**
	2	1.47	4.6	22.93
	3	1.09	3.41	26.34
	4	1.08	3.38	29.71
	5	1.03	3.21	32.92
SWD/SUA	1	**6.02**	**18.8**	**18.8**
	2	1.45	4.53	23.33
	3	1.09	3.41	26.74
	4	1.06	3.33	30.06
	5	1.03	3.21	33.28

Table L5. ELA Grade 7 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**5.29**	**13.55**	**13.55**
	2	1.76	4.52	18.08
	3	1.18	3.02	21.1
	4	1.14	2.91	24.01
	5	1.12	2.86	26.87
	6	1.1	2.83	29.7
	7	1.08	2.78	32.48
	8	1.05	2.7	35.18
	9	1.04	2.67	37.84
	10	1.03	2.65	40.5
	11	1.02	2.62	43.12
	12	1.01	2.59	45.7
SWD	1	**6.54**	**16.76**	**16.76**
	2	1.7	4.35	21.11
	3	1.12	2.87	23.98
	4	1.07	2.73	26.72
	5	1.05	2.69	29.41
	6	1.03	2.65	32.06
	7	1.01	2.58	34.64
SUA	1	**6.68**	**17.12**	**17.12**

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
	2	1.68	4.3	21.42
	3	1.11	2.86	24.28
	4	1.06	2.71	27
	5	1.04	2.67	29.67
	6	1.02	2.62	32.29
	7	1	2.57	34.86

Table L6. ELA Grade 8 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**6.34**	**16.26**	**16.26**
	2	1.75	4.49	20.75
	3	1.16	2.96	23.71
	4	1.1	2.83	26.55
	5	1.09	2.81	29.35
	6	1.05	2.7	32.05
	7	1.05	2.68	34.73
	8	1.03	2.65	37.38
	9	1.01	2.59	39.98
SWD	1	**7.08**	**18.15**	**18.15**
	2	1.62	4.14	22.29
	3	1.11	2.86	25.15
	4	1.08	2.76	27.91
	5	1.07	2.74	30.65
SUA	1	**7.22**	**18.5**	**18.5**
	2	1.58	4.06	22.56
	3	1.11	2.86	25.42
	4	1.07	2.75	28.17
	5	1.06	2.71	30.88

Table L7. Mathematics Grade 3 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**8.05**	**25.15**	**25.15**
	2	1.61	5.02	30.17
	3	1.14	3.57	33.73
	4	1.09	3.41	37.14

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
	5	1.06	3.3	40.45
SWD	1	**8.66**	**27.05**	**27.05**
	2	1.34	4.17	31.22
	3	1.14	3.56	34.78
	4	1.01	3.16	37.94
SUA	1	**8.04**	**25.11**	**25.11**
	2	1.31	4.11	29.22
	3	1.16	3.62	32.84
	4	1.03	3.23	36.07
	5	1.02	3.19	39.26

Table L8. Mathematics Grade 4 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**7.57**	**20.45**	**20.45**
	2	1.57	4.23	24.68
	3	1.13	3.06	27.73
	4	1.08	2.93	30.66
	5	1.05	2.83	33.49
	6	1.01	2.73	36.22
SWD	1	**9.08**	**24.55**	**24.55**
	2	1.48	4.01	28.56
	3	1.09	2.95	31.51
	4	1.03	2.79	34.29
SUA	1	**8.66**	**23.41**	**23.41**
	2	1.49	4.02	27.43
	3	1.09	2.95	30.37
	4	1.04	2.81	33.18

Table L9. Mathematics Grade 5 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**8.03**	**21.71**	**21.71**
	2	1.67	4.5	26.22
	3	1.13	3.05	29.27
	4	1.08	2.91	32.18
	5	**1.03**	**2.79**	**34.97**

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
SWD	1	**9.44**	**25.51**	**25.51**
	2	1.64	4.42	29.93
	3	1.04	2.81	32.75
	4	1.02	2.75	35.5
SUA	1	**9.16**	**24.77**	**24.77**
	2	1.62	4.38	29.15
	3	1.04	2.82	31.97
	4	1.03	2.78	34.74

Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**7.44**	**19.08**	**19.08**
	2	1.31	3.36	22.44
	3	1.18	3.02	25.47
	4	1.12	2.86	28.33
	5	1.08	2.77	31.1
	6	1.05	2.7	33.79
	7	1.03	2.63	36.42
SWD	1	**8.66**	**22.2**	**22.2**
	2	1.27	3.26	25.46
	3	1.14	2.92	28.38
	4	1.01	2.6	30.98
	5	1	2.57	33.55
SUA	1	**9**	**23.08**	**23.08**
	2	1.28	3.27	26.35
	3	1.14	2.93	29.28

Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**8.07**	**19.69**	**19.69**
	2	1.64	3.99	23.68
	3	1.36	3.32	27
	4	1.12	2.72	29.72
	5	1.05	2.56	32.28
	6	1.04	2.54	34.82

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
	7	1	2.44	37.26
SWD	1	**9.45**	**23.04**	**23.04**
	2	1.35	3.3	26.34
	3	1.26	3.08	29.42
	4	1.04	2.53	31.95
	5	1.01	2.46	34.41
SUA	1	9.56	23.31	23.31
	2	1.33	3.25	26.56
	3	1.25	3.05	29.61
	4	1.04	2.54	32.15
	5	1.02	2.49	34.64

Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup

Demographic Category	Extracted Factor			
	N	Eigenvalue	Variance Accounted for	
			%	Cumulative %
ELL	1	**8.14**	**19.84**	**19.84**
	2	1.4	3.41	23.26
	3	1.19	2.9	26.16
	4	1.12	2.74	28.9
	5	1.1	2.68	31.58
	6	1.08	2.64	34.22
	7	1.04	2.54	36.75
	8	1.02	2.49	39.25
SWD	1	**7.95**	**19.39**	**19.39**
	2	1.19	2.9	22.29
	3	1.09	2.66	24.95
	4	1.08	2.64	27.59
	5	1.03	2.52	30.11
	6	1.01	2.47	32.58
	7	1	2.45	35.03
SUA	1	**8.01**	**19.53**	**19.53**
	2	1.2	2.92	22.46
	3	1.09	2.66	25.12
	4	1.08	2.64	27.76
	5	1.03	2.51	30.27
	6	1.02	2.48	32.74
	7	1	2.45	35.19

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, percent of omitted responses, *p* value, and the point-biserial correlations (PBis). Field test items that do not contribute to students' scores have been omitted.

Table M1. ELA Grade 3 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	163,076	0	0.78	0.53
2	MC	163,076	0	0.71	0.44
3	MC	163,076	0	0.74	0.50
4	MC	163,076	0	0.73	0.44
5	MC	163,076	0	0.41	0.33
6	MC	163,076	0	0.72	0.45
7	MC	163,076	0	0.84	0.51
8	MC	163,076	0	0.57	0.44
9	MC	163,076	0	0.68	0.45
10	MC	163,076	0	0.68	0.53
11	MC	163,076	0	0.63	0.49
12	MC	163,076	0	0.63	0.48
13	MC	163,076	0	0.53	0.37
14	MC	163,076	0	0.57	0.50
15	MC	163,076	0	0.50	0.38
16	MC	163,076	0	0.51	0.41
17	MC	163,076	0	0.49	0.36
18	CR	163,076	–	0.43	0.65
19	CR	163,076	–	0.47	0.65
20	MC	163,076	0	0.67	0.29
21	MC	163,076	0	0.52	0.43
22	MC	163,076	0	0.80	0.43
23	MC	163,076	0	0.76	0.51
24	MC	163,076	0	0.72	0.47
25	MC	163,076	0	0.57	0.49
26	CR	163,076	–	0.55	0.58
27	CR	163,076	–	0.65	0.59
28	CR	163,076	–	0.55	0.53

Table M2. ELA Grade 4 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	162,852	0	0.64	0.31
2	MC	162,852	0	0.63	0.37
3	MC	162,852	0	0.66	0.44
4	MC	162,852	0	0.72	0.45

Item	Type	N-Count	% Omit	*P* Value	PBis
5	MC	162,852	0	0.59	0.28
6	MC	162,852	0	0.74	0.42
7	MC	162,852	0	0.66	0.35
8	MC	162,852	0	0.71	0.40
9	MC	162,852	0	0.69	0.41
10	MC	162,852	0	0.65	0.43
11	MC	162,852	0	0.74	0.48
12	MC	162,852	0	0.49	0.33
13	MC	162,852	0	0.36	0.23
14	MC	162,852	0	0.54	0.38
15	MC	162,852	0	0.44	0.26
16	MC	162,852	0	0.47	0.48
17	MC	162,852	0	0.50	0.36
18	CR	162,852	–	0.52	0.66
19	CR	162,852	–	0.51	0.64
20	MC	162,852	0	0.59	0.43
21	MC	162,852	0	0.57	0.40
22	MC	162,852	0	0.62	0.43
23	MC	162,852	0	0.73	0.43
24	MC	162,852	0	0.62	0.50
25	MC	162,852	0	0.54	0.36
26	CR	162,852	–	0.54	0.61
27	CR	162,852	–	0.58	0.54
28	CR	162,852	–	0.61	0.58
29	CR	162,852	–	0.49	0.63

Table M3. ELA Grade 5 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	161,256	0	0.63	0.33
2	MC	161,256	0	0.65	0.27
3	MC	161,256	0	0.64	0.25
4	MC	161,256	0	0.63	0.32
5	MC	161,256	0	0.80	0.43
6	MC	161,256	0	0.85	0.40
7	MC	161,256	0	0.74	0.36
8	MC	161,256	0	0.61	0.23
9	MC	161,256	0	0.57	0.48
10	MC	161,256	0	0.60	0.36
11	MC	161,256	0	0.67	0.45
12	MC	161,256	0	0.69	0.43
13	MC	161,256	0	0.69	0.51
14	MC	161,256	0	0.33	0.24

Item	Type	N-Count	% Omit	*P* Value	PBis
15	MC	161,256	0	0.44	0.28
16	MC	161,256	0	0.35	0.25
17	MC	161,256	0	0.42	0.30
18	MC	161,256	0	0.53	0.44
19	MC	161,256	0	0.41	0.22
20	CR	161,256	–	0.63	0.57
21	CR	161,256	–	0.53	0.55
22	MC	161,256	0	0.61	0.38
23	MC	161,256	0	0.88	0.37
24	MC	161,256	0	0.81	0.40
25	MC	161,256	0	0.67	0.36
26	MC	161,256	0	0.42	0.37
27	MC	161,256	0	0.58	0.40
28	MC	161,256	0	0.66	0.40
29	CR	161,256	–	0.63	0.58
30	CR	161,256	–	0.71	0.61
31	CR	161,256	–	0.60	0.60
32	CR	161,256	–	0.39	0.64

Table M4. ELA Grade 6 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	159,541	0	0.86	0.42
2	MC	159,541	0	0.60	0.25
3	MC	159,541	0	0.58	0.26
4	MC	159,541	0	0.63	0.38
5	MC	159,541	0	0.61	0.36
6	MC	159,541	0	0.56	0.31
7	MC	159,541	0	0.41	0.16
8	MC	159,541	0	0.63	0.53
9	MC	159,541	0	0.65	0.43
10	MC	159,541	0	0.63	0.39
11	MC	159,541	0	0.81	0.48
12	MC	159,541	0	0.48	0.34
13	MC	159,541	0	0.77	0.44
14	MC	159,541	0	0.46	0.24
15	MC	159,541	0	0.38	0.14
16	MC	159,541	0	0.73	0.50
17	MC	159,541	0	0.61	0.46
18	MC	159,541	0	0.77	0.42
19	MC	159,541	0	0.37	0.20
20	CR	159,541	–	0.76	0.58
21	CR	159,541	–	0.75	0.59

Item	Type	N-Count	% Omit	*P* Value	PBis
22	MC	159,541	0	0.53	0.22
23	MC	159,541	0	0.82	0.43
24	MC	159,541	0	0.57	0.27
25	MC	159,541	0	0.75	0.46
26	MC	159,541	0	0.57	0.27
27	MC	159,541	0	0.62	0.30
28	MC	159,541	0	0.53	0.30
29	CR	159,541	–	0.77	0.61
30	CR	159,541	–	0.75	0.65
31	CR	159,541	–	0.68	0.57
32	CR	159,541	–	0.47	0.66

Table M5. ELA Grade 7 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	155,450	0	0.65	0.36
2	MC	155,450	0	0.77	0.50
3	MC	155,450	0	0.60	0.41
4	MC	155,450	0	0.64	0.30
5	MC	155,450	0	0.53	0.35
6	MC	155,450	0	0.51	0.32
7	MC	155,450	0	0.69	0.34
8	MC	155,450	0	0.60	0.25
9	MC	155,450	0	0.60	0.37
10	MC	155,450	0	0.74	0.55
11	MC	155,450	0	0.46	0.32
12	MC	155,450	0	0.55	0.29
13	MC	155,450	0	0.42	0.27
14	MC	155,450	0	0.63	0.48
15	MC	155,450	0	0.49	0.34
16	MC	155,450	0	0.47	0.25
17	MC	155,450	0	0.53	0.30
18	MC	155,450	0	0.70	0.44
19	MC	155,450	0	0.73	0.52
20	CR	155,450	–	0.79	0.63
21	CR	155,450	–	0.79	0.62
22	MC	155,450	0	0.72	0.38
23	MC	155,450	0	0.57	0.33
24	MC	155,450	0	0.51	0.30
25	MC	155,450	0	0.56	0.35
26	MC	155,450	0	0.59	0.30
27	MC	155,450	0	0.76	0.50
28	MC	155,450	0	0.58	0.33

Item	Type	N-Count	% Omit	*P* Value	PBis
29	MC	155,450	0	0.63	0.25
30	MC	155,450	0	0.66	0.41
31	MC	155,450	0	0.58	0.32
32	MC	155,450	0	0.53	0.32
33	MC	155,450	0	0.60	0.39
34	MC	155,450	0	0.70	0.40
35	MC	155,450	0	0.45	0.20
36	CR	155,450	–	0.76	0.52
37	CR	155,450	–	0.76	0.63
38	CR	155,450	–	0.74	0.62
39	CR	155,450	–	0.54	0.64

Table M6. ELA Grade 8 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	146,532	0	0.85	0.37
2	MC	146,532	0	0.80	0.38
3	MC	146,532	0	0.60	0.28
4	MC	146,532	0	0.51	0.26
5	MC	146,532	0	0.47	0.18
6	MC	146,532	0	0.51	0.21
7	MC	146,532	0	0.39	0.22
8	MC	146,532	0	0.86	0.46
9	MC	146,532	0	0.81	0.48
10	MC	146,532	0	0.68	0.38
11	MC	146,532	0	0.66	0.44
12	MC	146,532	0	0.79	0.54
13	MC	146,532	0	0.70	0.46
14	MC	146,532	0	0.70	0.43
15	MC	146,532	0	0.67	0.32
16	MC	146,532	0	0.76	0.52
17	MC	146,532	0	0.60	0.34
18	MC	146,532	0	0.67	0.38
19	MC	146,532	0	0.53	0.36
20	CR	146,532	–	0.69	0.63
21	CR	146,532	–	0.66	0.62
22	MC	146,532	0	0.68	0.36
23	MC	146,532	0	0.56	0.25
24	MC	146,532	0	0.73	0.47
25	MC	146,532	0	0.58	0.32
26	MC	146,532	0	0.52	0.23
27	MC	146,532	0	0.84	0.45
28	MC	146,532	0	0.70	0.38

Item	Type	N-Count	% Omit	*P* Value	PBis
29	MC	146,532	0	0.69	0.42
30	MC	146,532	0	0.48	0.33
31	MC	146,532	0	0.70	0.40
32	MC	146,532	0	0.48	0.24
33	MC	146,532	0	0.53	0.36
34	MC	146,532	0	0.46	0.27
35	MC	146,532	0	0.84	0.42
36	CR	146,532	–	0.80	0.63
37	CR	146,532	–	0.75	0.60
38	CR	146,532	–	0.69	0.59
39	CR	146,532	–	0.57	0.66

Table M7. Mathematics Grade 3 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	159,439	0	0.83	0.42
2	MC	159,439	0	0.88	0.33
3	MC	159,439	0	0.74	0.43
4	MC	159,439	0	0.53	0.46
5	MC	159,439	0	0.83	0.49
6	MC	159,439	0	0.69	0.45
7	MC	159,439	0	0.67	0.48
8	MC	159,439	0	0.58	0.44
9	MC	159,439	0	0.70	0.47
10	MC	159,439	0	0.58	0.54
11	MC	159,439	0	0.38	0.33
12	MC	159,439	0	0.71	0.45
13	MC	159,439	0	0.48	0.49
14	MC	159,439	0	0.77	0.47
15	MC	159,439	0	0.75	0.52
16	MC	159,439	0	0.73	0.57
17	MC	159,439	0	0.51	0.36
18	MC	159,439	0	0.78	0.49
19	MC	159,439	0	0.79	0.52
20	MC	159,439	0	0.71	0.55
21	MC	159,439	0	0.92	0.34
22	MC	159,439	0	0.63	0.61
23	MC	159,439	0	0.54	0.57
24	MC	159,439	0	0.73	0.45
25	CR	159,439	–	0.74	0.55
26	CR	159,439	–	0.63	0.59
27	CR	159,439	–	0.42	0.48
28	CR	159,439	–	0.46	0.60

Item	Type	N-Count	% Omit	*P* Value	PBis
29	CR	159,439	–	0.44	0.59
30	CR	159,439	–	0.59	0.65
31	CR	159,439	–	0.43	0.60
32	CR	159,439	–	0.49	0.57

Table M8. Mathematics Grade 4 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	159,198	0	0.85	0.40
2	MC	159,198	0	0.92	0.35
3	MC	159,198	0	0.69	0.56
4	MC	159,198	0	0.68	0.54
5	MC	159,198	0	0.78	0.45
6	MC	159,198	0	0.54	0.53
7	MC	159,198	0	0.50	0.30
8	MC	159,198	0	0.61	0.59
9	MC	159,198	0	0.49	0.36
10	MC	159,198	0	0.53	0.47
11	MC	159,198	0	0.42	0.48
12	MC	159,198	0	0.50	0.37
13	MC	159,198	0	0.48	0.52
14	MC	159,198	0	0.44	0.48
15	MC	159,198	0	0.70	0.38
16	MC	159,198	0	0.50	0.49
17	MC	159,198	0	0.53	0.43
18	MC	159,198	0	0.48	0.31
19	MC	159,198	0	0.74	0.44
20	MC	159,198	0	0.76	0.46
21	MC	159,198	0	0.70	0.31
22	MC	159,198	0	0.60	0.63
23	MC	159,198	0	0.68	0.48
24	MC	159,198	0	0.79	0.36
25	MC	159,198	0	0.64	0.54
26	MC	159,198	0	0.47	0.37
27	MC	159,198	0	0.80	0.39
28	MC	159,198	0	0.83	0.44
29	CR	159,198	–	0.66	0.52
30	CR	159,198	–	0.62	0.53
31	CR	159,198	–	0.59	0.55
32	CR	159,198	–	0.22	0.42
33	CR	159,198	–	0.44	0.63
34	CR	159,198	–	0.65	0.65
35	CR	159,198	–	0.60	0.66

Item	Type	N-Count	% Omit	*P* Value	PBis
36	CR	159,198	–	0.52	0.67
37	CR	159,198	–	0.33	0.66

Table M9. Mathematics Grade 5 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	157,029	0	0.87	0.32
2	MC	157,029	0	0.74	0.53
3	MC	157,029	0	0.81	0.45
4	MC	157,029	0	0.72	0.54
5	MC	157,029	0	0.87	0.36
6	MC	157,029	0	0.76	0.45
7	MC	157,029	0	0.51	0.42
8	MC	157,029	0	0.69	0.50
9	MC	157,029	0	0.55	0.52
10	MC	157,029	0	0.34	0.38
11	MC	157,029	0	0.58	0.42
12	MC	157,029	0	0.60	0.34
13	MC	157,029	0	0.55	0.49
14	MC	157,029	0	0.68	0.50
15	MC	157,029	0	0.52	0.53
16	MC	157,029	0	0.33	0.40
17	MC	157,029	0	0.55	0.54
18	MC	157,029	0	0.63	0.59
19	MC	157,029	0	0.80	0.45
20	MC	157,029	0	0.48	0.55
21	MC	157,029	0	0.48	0.45
22	MC	157,029	0	0.45	0.51
23	MC	157,029	0	0.75	0.55
24	MC	157,029	0	0.77	0.55
25	MC	157,029	0	0.77	0.51
26	MC	157,029	0	0.51	0.54
27	MC	157,029	0	0.60	0.34
28	MC	157,029	0	0.79	0.55
29	CR	157,029	–	0.57	0.51
30	CR	157,029	–	0.45	0.57
31	CR	157,029	–	0.49	0.62
32	CR	157,029	–	0.56	0.68
33	CR	157,029	–	0.26	0.59
34	CR	157,029	–	0.31	0.60
35	CR	157,029	–	0.46	0.65
36	CR	157,029	–	0.47	0.61
37	CR	157,029	–	0.40	0.74

Table M10. Mathematics Grade 6 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	152,136	0	0.72	0.50
2	MC	152,136	0	0.65	0.46
3	MC	152,136	0	0.46	0.55
4	MC	152,136	0	0.75	0.52
5	MC	152,136	0	0.61	0.46
6	MC	152,136	0	0.45	0.46
7	MC	152,136	0	0.56	0.47
8	MC	152,136	0	0.54	0.51
9	MC	152,136	0	0.51	0.49
10	MC	152,136	0	0.56	0.36
11	MC	152,136	0	0.65	0.55
12	MC	152,136	0	0.36	0.45
13	MC	152,136	0	0.67	0.47
14	MC	152,136	0	0.50	0.49
15	MC	152,136	0	0.56	0.46
16	MC	152,136	0	0.56	0.45
17	MC	152,136	0	0.63	0.64
18	MC	152,136	0	0.37	0.42
19	MC	152,136	0	0.48	0.42
20	MC	152,136	0	0.35	0.48
21	MC	152,136	0	0.68	0.50
22	MC	152,136	0	0.53	0.50
23	MC	152,136	0	0.45	0.41
24	MC	152,136	0	0.61	0.41
25	MC	152,136	0	0.60	0.40
26	MC	152,136	0	0.60	0.43
27	MC	152,136	0	0.59	0.60
28	MC	152,136	0	0.54	0.46
29	MC	152,136	0	0.67	0.58
30	CR	152,136	–	0.52	0.60
31	CR	152,136	–	0.60	0.44
32	CR	152,136	–	0.69	0.48
33	CR	152,136	–	0.13	0.45
34	CR	152,136	–	0.54	0.65
35	CR	152,136	–	0.55	0.64
36	CR	152,136	–	0.54	0.62
37	CR	152,136	–	0.47	0.71
38	CR	152,136	–	0.47	0.56
39	CR	152,136	–	0.34	0.71

Table M11. Mathematics Grade 7 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	145,907	0	0.89	0.30
2	MC	145,907	0	0.78	0.49
3	MC	145,907	0	0.76	0.46
4	MC	145,907	0	0.89	0.43
5	MC	145,907	0	0.67	0.47
6	MC	145,907	0	0.61	0.54
7	MC	145,907	0	0.74	0.57
8	MC	145,907	0	0.58	0.42
9	MC	145,907	0	0.59	0.39
10	MC	145,907	0	0.60	0.32
11	MC	145,907	0	0.49	0.33
12	MC	145,907	0	0.61	0.46
13	MC	145,907	0	0.52	0.31
14	MC	145,907	0	0.41	0.38
15	MC	145,907	0	0.83	0.48
16	MC	145,907	0	0.77	0.54
17	MC	145,907	0	0.64	0.43
18	MC	145,907	0	0.59	0.57
19	MC	145,907	0	0.54	0.55
20	MC	145,907	0	0.53	0.33
21	MC	145,907	0	0.37	0.32
22	MC	145,907	0	0.76	0.52
23	MC	145,907	0	0.49	0.54
24	MC	145,907	0	0.77	0.29
25	MC	145,907	0	0.64	0.57
26	MC	145,907	0	0.67	0.48
27	MC	145,907	0	0.43	0.45
28	MC	145,907	0	0.70	0.54
29	MC	145,907	0	0.82	0.41
30	MC	145,907	0	0.66	0.44
31	MC	145,907	0	0.70	0.58
32	CR	145,907	–	0.69	0.60
33	CR	145,907	–	0.57	0.65
34	CR	145,907	–	0.36	0.57
35	CR	145,907	–	0.40	0.60
36	CR	145,907	–	0.45	0.66
37	CR	145,907	–	0.43	0.64
38	CR	145,907	–	0.47	0.66
39	CR	145,907	–	0.66	0.62
40	CR	145,907	–	0.65	0.68
41	CR	145,907	–	0.48	0.74

Table M12. Mathematics Grade 8 Classical Item Analysis

Item	Type	N-Count	% Omit	*P* Value	PBis
1	MC	93,235	0	0.68	0.44
2	MC	93,235	0	0.66	0.26
3	MC	93,235	0	0.71	0.35
4	MC	93,235	0	0.65	0.43
5	MC	93,235	0	0.66	0.44
6	MC	93,235	0	0.45	0.36
7	MC	93,235	0	0.34	0.09
8	MC	93,235	0	0.59	0.39
9	MC	93,235	0	0.58	0.42
10	MC	93,235	0	0.55	0.41
11	MC	93,235	0	0.48	0.39
12	MC	93,235	0	0.51	0.36
13	MC	93,235	0	0.64	0.50
14	MC	93,235	0	0.73	0.54
15	MC	93,235	0	0.58	0.34
16	MC	93,235	0	0.27	0.34
17	MC	93,235	0	0.57	0.50
18	MC	93,235	0	0.49	0.25
19	MC	93,235	0	0.54	0.40
20	MC	93,235	0	0.58	0.40
21	MC	93,235	0	0.46	0.50
22	MC	93,235	0	0.48	0.44
23	MC	93,235	0	0.57	0.43
24	MC	93,235	0	0.58	0.57
25	MC	93,235	0	0.55	0.32
26	MC	93,235	0	0.71	0.43
27	MC	93,235	0	0.47	0.39
28	MC	93,235	0	0.47	0.27
29	MC	93,235	0	0.48	0.27
30	MC	93,235	0	0.51	0.56
31	MC	93,235	0	0.52	0.45
32	CR	93,235	–	0.59	0.52
33	CR	93,235	–	0.31	0.56
34	CR	93,235	–	0.50	0.59
35	CR	93,235	–	0.39	0.63
36	CR	93,235	–	0.46	0.68
37	CR	93,235	–	0.31	0.60
38	CR	93,235	–	0.40	0.69
39	CR	93,235	–	0.36	0.67
40	CR	93,235	–	0.32	0.56
41	CR	93,235	–	0.34	0.61

Appendix N: IRT Statistics

Tables N1 through N12 show item-calibration results for the operational items.

Table N1. ELA Grade 3 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	1.166	-1.065				
2	1	0.759	-0.944				
3	1	0.953	-0.937				
4	1	0.761	-1				
5	1	0.495	0.515				
6	1	0.764	-0.973				
7	1	1.397	-1.261				
8	1	0.678	-0.323				
9	1	0.749	-0.796				
10	1	0.976	-0.688				
11	1	0.829	-0.523				
12	1	0.804	-0.54				
13	1	0.536	-0.173				
14	1	0.852	-0.277				
15	1	0.547	0.013				
16	1	0.619	-0.054				
17	1	0.538	0.072				
18	2	1.024	0.29	0.726	-0.726		
19	2	0.959	0.125	0.616	-0.616		
20	1	0.409	-1.155				
21	1	0.648	-0.114				
22	1	0.848	-1.326				
23	1	1	-1.006				
24	1	0.824	-0.953				
25	1	0.791	-0.312				
26	2	0.719	-0.257	0.772	-0.772		
27	2	0.78	-0.664	0.746	-0.746		
28	2	0.658	-0.271	1.077	-1.077		

Table N2. ELA Grade 4 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.44	-0.898				
2	1	0.557	-0.681				
3	1	0.723	-0.698				
4	1	0.767	-0.941				
5	1	0.388	-0.581				
6	1	0.714	-1.092				
7	1	0.524	-0.902				

Item	MaxPts	a	b	d1	d2	d3	d4
8	1	0.641	-1.035				
9	1	0.672	-0.881				
10	1	0.682	-0.675				
11	1	0.884	-0.974				
12	1	0.46	0.071				
13	1	0.33	1.094				
14	1	0.568	-0.181				
15	1	0.355	0.459				
16	1	0.783	0.123				
17	1	0.523	-0.023				
18	2	1.013	-0.061	0.65	-0.65		
19	2	0.943	-0.047	0.627	-0.627		
20	1	0.664	-0.398				
21	1	0.594	-0.337				
22	1	0.657	-0.537				
23	1	0.743	-1.036				
24	1	0.866	-0.48				
25	1	0.505	-0.199				
26	2	0.863	-0.189	0.825	-0.825		
27	2	0.695	-0.412	0.916	-0.916		
28	2	0.878	-0.548	1.106	-1.106		
29	4	0.607	0.095	1.446	0.779	-0.471	-1.755

Table N3. ELA Grade 5 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.478	-0.772				
2	1	0.385	-1.053				
3	1	0.347	-1.045				
4	1	0.465	-0.783				
5	1	0.853	-1.328				
6	1	0.878	-1.597				
7	1	0.577	-1.253				
8	1	0.312	-0.894				
9	1	0.786	-0.289				
10	1	0.527	-0.532				
11	1	0.76	-0.71				
12	1	0.706	-0.871				
13	1	0.946	-0.71				
14	1	0.352	1.256				
15	1	0.382	0.385				
16	1	0.36	1.059				
17	1	0.417	0.484				

Item	MaxPts	a	b	d1	d2	d3	d4
18	1	0.676	-0.115				
19	1	0.29	0.763				
20	2	0.846	-0.644	0.99	-0.99		
21	2	0.781	-0.173	1.041	-1.041		
22	1	0.57	-0.532				
23	1	0.846	-1.808				
24	1	0.75	-1.443				
25	1	0.533	-0.906				
26	1	0.559	0.43				
27	1	0.598	-0.364				
28	1	0.63	-0.743				
29	2	0.856	-0.604	0.899	-0.899		
30	2	1.011	-0.919	0.801	-0.801		
31	2	0.897	-0.455	0.868	-0.868		
32	4	0.743	0.567	1.646	0.578	-0.653	-1.571

Table N4. ELA Grade 6 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.889	-1.668				
2	1	0.335	-0.76				
3	1	0.338	-0.571				
4	1	0.576	-0.663				
5	1	0.514	-0.618				
6	1	0.434	-0.341				
7	1	0.215	1.058				
8	1	0.929	-0.489				
9	1	0.69	-0.668				
10	1	0.577	-0.669				
11	1	0.94	-1.276				
12	1	0.498	0.109				
13	1	0.762	-1.207				
14	1	0.331	0.292				
15	1	0.187	1.621				
16	1	0.897	-0.938				
17	1	0.77	-0.433				
18	1	0.717	-1.249				
19	1	0.29	1.182				
20	2	0.963	-1.217	0.871	-0.871		
21	2	0.926	-1.127	0.808	-0.808		
22	1	0.302	-0.214				
23	1	0.819	-1.468				
24	1	0.352	-0.503				

Item	MaxPts	a	b	d1	d2	d3	d4
25	1	0.79	-1.093				
26	1	0.355	-0.477				
27	1	0.416	-0.762				
28	1	0.415	-0.181				
29	2	1.003	-1.177	0.693	-0.693		
30	2	1.12	-0.963	0.57	-0.57		
31	2	0.847	-0.868	0.913	-0.913		
32	4	0.77	0.186	1.338	0.804	-0.508	-1.634

Table N5. ELA Grade 7 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.521	-0.846				
2	1	0.963	-1.069				
3	1	0.62	-0.479				
4	1	0.413	-0.928				
5	1	0.5	-0.182				
6	1	0.458	-0.074				
7	1	0.506	-1.092				
8	1	0.319	-0.769				
9	1	0.539	-0.531				
10	1	1.125	-0.892				
11	1	0.45	0.263				
12	1	0.396	-0.343				
13	1	0.386	0.532				
14	1	0.776	-0.531				
15	1	0.494	0.056				
16	1	0.347	0.255				
17	1	0.413	-0.188				
18	1	0.722	-0.913				
19	1	0.986	-0.884				
20	2	1.18	-1.131	0.515	-0.515		
21	2	1.135	-1.145	0.564	-0.564		
22	1	0.579	-1.135				
23	1	0.463	-0.433				
24	1	0.416	-0.096				
25	1	0.493	-0.343				
26	1	0.411	-0.559				
27	1	0.945	-1.065				
28	1	0.474	-0.44				
29	1	0.34	-1				
30	1	0.627	-0.759				
31	1	0.44	-0.471				

Item	MaxPts	a	b	d1	d2	d3	d4
32	1	0.45	-0.167				
33	1	0.569	-0.506				
34	1	0.607	-0.983				
35	1	0.264	0.423				
36	2	0.736	-1.293	0.753	-0.753		
37	2	1.125	-1.054	0.618	-0.618		
38	2	1.039	-0.978	0.613	-0.613		
39	4	0.662	-0.169	1.334	0.69	-0.568	-1.456

Table N6. ELA Grade 8 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.697	-1.836				
2	1	0.637	-1.546				
3	1	0.372	-0.7				
4	1	0.355	-0.094				
5	1	0.226	0.276				
6	1	0.271	-0.078				
7	1	0.313	0.886				
8	1	0.985	-1.577				
9	1	0.908	-1.311				
10	1	0.582	-0.902				
11	1	0.71	-0.687				
12	1	1.113	-1.106				
13	1	0.762	-0.848				
14	1	0.709	-0.922				
15	1	0.452	-1.005				
16	1	0.984	-1.009				
17	1	0.484	-0.555				
18	1	0.572	-0.85				
19	1	0.545	-0.125				
20	2	0.905	-0.737	0.433	-0.433		
21	2	0.838	-0.623	0.372	-0.372		
22	1	0.521	-0.978				
23	1	0.322	-0.508				
24	1	0.787	-1.006				
25	1	0.443	-0.474				
26	1	0.3	-0.201				
27	1	0.89	-1.53				
28	1	0.561	-1.052				
29	1	0.641	-0.881				
30	1	0.48	0.126				
31	1	0.62	-1.001				

Item	MaxPts	a	b	d1	d2	d3	d4
32	1	0.315	0.169				
33	1	0.545	-0.129				
34	1	0.369	0.309				
35	1	0.758	-1.633				
36	2	1.059	-1.262	0.555	-0.555		
37	2	0.893	-1.095	0.679	-0.679		
38	2	0.824	-0.843	0.716	-0.716		
39	4	0.732	-0.296	1.278	0.862	-0.562	-1.577

Table N7. Mathematics Grade 3 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.948	-1.405				
2	1	0.801	-1.929				
3	1	0.768	-1.076				
4	1	0.736	-0.144				
5	1	1.28	-1.256				
6	1	0.76	-0.815				
7	1	0.821	-0.732				
8	1	0.703	-0.369				
9	1	0.846	-0.812				
10	1	0.96	-0.335				
11	1	0.513	0.65				
12	1	0.799	-0.914				
13	1	0.828	0.044				
14	1	0.941	-1.114				
15	1	1.07	-0.941				
16	1	1.289	-0.82				
17	1	0.534	-0.081				
18	1	1.043	-1.096				
19	1	1.199	-1.077				
20	1	1.137	-0.789				
21	1	1.07	-1.935				
22	1	1.247	-0.467				
23	1	1.053	-0.175				
24	1	0.826	-0.993				
25	1	1.172	-0.896				
26	1	1.127	-0.474				
27	1	0.79	0.319				
28	2	0.65	0.145	-0.163	0.163		
29	2	0.674	0.207	0.241	-0.241		
30	2	0.794	-0.335	0.003	-0.003		

Item	MaxPts	a	b	d1	d2	d3	d4
31	2	0.687	0.243	0.16	-0.16		
32	3	0.485	0.089	0.47	0.538	-1.008	

Table N8. Mathematics Grade 4 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.961	-1.531				
2	1	1.179	-1.843				
3	1	1.15	-0.703				
4	1	1.046	-0.676				
5	1	0.877	-1.186				
6	1	0.889	-0.184				
7	1	0.409	-0.033				
8	1	1.138	-0.401				
9	1	0.508	0.046				
10	1	0.726	-0.146				
11	1	0.767	0.31				
12	1	0.518	-0.014				
13	1	0.864	0.042				
14	1	0.757	0.241				
15	1	0.628	-1.006				
16	1	0.767	-0.011				
17	1	0.659	-0.154				
18	1	0.427	0.094				
19	1	0.79	-1.068				
20	1	0.891	-1.104				
21	1	0.468	-1.248				
22	1	1.322	-0.337				
23	1	0.829	-0.756				
24	1	0.654	-1.49				
25	1	0.989	-0.553				
26	1	0.53	0.134				
27	1	0.759	-1.389				
28	1	1.047	-1.361				
29	1	0.922	-0.636				
30	1	0.916	-0.477				
31	1	0.965	-0.371				
32	2	0.458	1.393	-0.382	0.382		
33	2	0.769	0.222	0.28	-0.28		
34	2	0.937	-0.581	0.403	-0.403		
35	2	0.862	-0.385	0.071	-0.071		
36	2	0.857	-0.104	0.055	-0.055		
37	3	0.698	0.566	-0.201	-0.355	0.556	

Table N9. Mathematics Grade 5 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.744	-1.892				
2	1	1.172	-0.879				
3	1	1.054	-1.242				
4	1	1.159	-0.824				
5	1	0.95	-1.673				
6	1	0.903	-1.068				
7	1	0.615	-0.046				
8	1	0.944	-0.738				
9	1	0.87	-0.209				
10	1	0.574	0.822				
11	1	0.635	-0.401				
12	1	0.485	-0.572				
13	1	0.787	-0.221				
14	1	0.935	-0.719				
15	1	0.883	-0.117				
16	1	0.636	0.806				
17	1	0.921	-0.206				
18	1	1.153	-0.466				
19	1	0.985	-1.197				
20	1	0.954	0.038				
21	1	0.688	0.068				
22	1	0.825	0.17				
23	1	1.313	-0.894				
24	1	1.421	-0.945				
25	1	1.19	-0.998				
26	1	0.903	-0.055				
27	1	0.483	-0.578				
28	1	1.536	-0.996				
29	1	0.826	-0.295				
30	1	1.001	0.144				
31	1	1.195	0.001				
32	2	0.913	-0.262	0.182	-0.182		
33	2	0.759	0.908	-0.18	0.18		
34	2	0.783	0.718	0.235	-0.235		
35	2	0.766	0.111	-0.012	0.012		
36	2	0.605	0.076	-1.048	1.048		
37	3	0.871	0.317	0.27	-0.001	-0.269	

Table N10. Mathematics Grade 6 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	1.022	-0.854				
2	1	0.793	-0.633				
3	1	0.918	0.112				
4	1	1.253	-0.909				
5	1	0.758	-0.482				
6	1	0.71	0.189				
7	1	0.738	-0.297				
8	1	0.838	-0.172				
9	1	0.77	-0.067				
10	1	0.527	-0.352				
11	1	1.085	-0.554				
12	1	0.72	0.574				
13	1	0.842	-0.726				
14	1	0.77	-0.031				
15	1	0.727	-0.287				
16	1	0.703	-0.303				
17	1	1.474	-0.458				
18	1	0.634	0.573				
19	1	0.607	0.07				
20	1	0.781	0.623				
21	1	0.941	-0.712				
22	1	0.801	-0.154				
23	1	0.585	0.235				
24	1	0.638	-0.529				
25	1	0.621	-0.501				
26	1	0.667	-0.491				
27	1	1.208	-0.355				
28	1	0.72	-0.207				
29	1	1.276	-0.625				
30	1	1.108	-0.13				
31	1	0.694	-0.463				
32	1	0.896	-0.776				
33	2	0.685	1.633	-0.634	0.634		
34	2	0.726	-0.198	-0.502	0.502		
35	2	0.715	-0.206	-0.498	0.498		
36	2	0.677	-0.174	-0.411	0.411		
37	2	0.94	0.058	-0.077	0.077		
38	2	0.509	0.094	-0.96	0.96		
39	3	0.833	0.663	0.652	0.326	-0.978	

Table N11. Mathematics Grade 7 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.741	-2.068				
2	1	1.066	-1.087				
3	1	0.908	-1.091				
4	1	1.546	-1.463				
5	1	0.809	-0.732				
6	1	0.963	-0.434				
7	1	1.309	-0.863				
8	1	0.651	-0.39				
9	1	0.59	-0.479				
10	1	0.469	-0.576				
11	1	0.465	0.027				
12	1	0.743	-0.496				
13	1	0.439	-0.143				
14	1	0.575	0.434				
15	1	1.268	-1.257				
16	1	1.318	-0.969				
17	1	0.699	-0.623				
18	1	1.057	-0.356				
19	1	0.946	-0.168				
20	1	0.469	-0.159				
21	1	0.481	0.709				
22	1	1.136	-0.959				
23	1	0.921	0.004				
24	1	0.492	-1.629				
25	1	1.117	-0.533				
26	1	0.855	-0.707				
27	1	0.717	0.266				
28	1	1.051	-0.751				
29	1	0.916	-1.395				
30	1	0.729	-0.71				
31	1	1.231	-0.709				
32	1	1.297	-0.663				
33	1	1.34	-0.28				
34	1	1.059	0.462				
35	2	0.77	0.411	0.641	-0.641		
36	2	0.979	0.165	0.791	-0.791		
37	2	0.738	0.219	-0.381	0.381		
38	2	0.866	0.076	0.27	-0.27		
39	2	0.713	-0.581	-0.662	0.662		
40	2	0.933	-0.529	-0.241	0.241		
41	3	0.828	0.019	-0.031	0.233	-0.202	

Table N12. Mathematics Grade 8 OP Item Parameter Estimates

Item	MaxPts	a	b	d1	d2	d3	d4
1	1	0.801	-0.783				
2	1	0.364	-1.153				
3	1	0.582	-1.074				
4	1	0.723	-0.674				
5	1	0.755	-0.7				
6	1	0.512	0.234				
7	1	0.139	2.87				
8	1	0.583	-0.441				
9	1	0.656	-0.373				
10	1	0.611	-0.237				
11	1	0.568	0.082				
12	1	0.519	-0.078				
13	1	0.941	-0.57				
14	1	1.359	-0.801				
15	1	0.478	-0.479				
16	1	0.525	1.306				
17	1	0.856	-0.305				
18	1	0.333	0.033				
19	1	0.581	-0.227				
20	1	0.601	-0.386				
21	1	0.793	0.123				
22	1	0.668	0.09				
23	1	0.662	-0.342				
24	1	1.069	-0.319				
25	1	0.448	-0.335				
26	1	0.789	-0.903				
27	1	0.55	0.12				
28	1	0.369	0.189				
29	1	0.357	0.1				
30	1	1.003	-0.077				
31	1	0.694	-0.115				
32	1	0.955	-0.367				
33	1	1.036	0.664				
34	1	1.107	-0.019				
35	2	0.802	0.409	0.438	-0.438		
36	2	0.9	0.107	0.034	-0.034		
37	2	0.651	0.657	-0.741	0.741		
38	2	0.933	0.307	0.018	-0.018		
39	2	0.907	0.458	-0.069	0.069		
40	2	0.768	0.945	1.073	-1.073		
41	3	0.526	0.642	0.196	-0.236	0.041	

Appendix O: Derivation and Estimation of Classification Consistency and Accuracy

Classification Consistency

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The agreement index (classification consistency) can be computed as

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

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

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

Then, kappa (Cohen, 1960) is

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

Classification Accuracy

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

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

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

Lee (2010) thoroughly discusses this IRT method for estimating decision indices, including the computational method used to estimate the results when integrating across the latent variable, θ.

Estimating Classification Indices

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

Files needed:

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

- Cut score file
    a. Contains the cut scores to be used
    b. Results are provided for all cut scores simultaneously (all performance levels), as well as the estimates based on each of the cut scores separately (Level 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 P: Raw-Score-to-Scale Score and Scale Score Frequency Tables

Tables P.1–P.12 show the raw-score-to-scale score conversion tables. Tables P.13–P.24 show the scale score distributions that include all students with valid scores by frequency (n-count), percent, cumulative frequency, and cumulative percent.

Table P1. ELA Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error
0	380	14
1	384	13
2	387	12
3	391	11
4	397	9
5	402	8
6	405	8
7	409	7
8	412	7
9	415	7
10	417	6
11	420	6
12	422	6
13	425	6
14	427	6
15	430	6
16	432	6
17	434	6
18	437	6
19	439	6
20	442	6
21	444	7
22	447	7
23	450	7
24	453	7
25	456	8
26	460	8
27	464	8
28	468	9
29	474	10
30	480	11
31	484	12
32	487	13
33	491	15

Table P2. ELA Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error
0	380	15
1	384	13
2	387	12
3	391	11
4	394	11
5	400	10
6	404	9
7	408	8
8	412	8
9	415	8
10	418	7
11	421	7
12	423	7
13	426	7
14	428	7
15	431	7
16	433	7
17	435	7
18	438	7
19	440	7
20	443	7
21	445	7
22	447	7
23	450	7
24	453	7
25	455	7
26	458	7
27	461	7
28	464	8
29	468	8
30	471	9
31	476	9
32	481	10
33	486	11
34	490	12
35	493	13
36	497	14
37	500	15

Table P3. ELA Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error
0	371	15
1	375	14
2	378	13
3	382	12
4	385	11
5	391	10
6	395	9
7	399	8
8	403	8
9	406	7
10	409	7
11	412	7
12	415	7
13	417	7
14	420	7
15	422	7
16	425	7
17	427	7
18	430	7
19	432	7
20	435	7
21	437	7
22	439	7
23	442	7
24	445	7
25	447	7
26	450	7
27	453	7
28	456	8
29	459	8
30	462	8
31	466	8
32	470	9
33	474	9
34	479	10
35	485	11
36	488	12
37	492	13
38	495	13
39	499	14
40	502	15

Table P4. ELA Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error
0	366	16
1	369	14
2	373	13
3	376	12
4	380	11
5	386	10
6	391	9
7	395	8
8	398	8
9	402	7
10	405	7
11	407	7
12	410	7
13	413	7
14	415	6
15	418	6
16	420	6
17	422	6
18	425	6
19	427	6
20	429	6
21	431	6
22	434	6
23	436	7
24	439	7
25	441	7
26	444	7
27	447	7
28	450	8
29	453	8
30	457	8
31	461	9
32	465	10
33	470	11
34	476	12
35	480	12
36	483	13
37	487	14
38	490	15
39	494	16
40	497	17

Table P5. ELA Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error
0	366	17
1	369	15
2	373	14
3	376	13
4	380	12
5	386	10
6	391	9
7	395	8
8	398	8
9	401	7
10	404	7
11	407	7
12	409	6
13	412	6
14	414	6
15	416	6
16	418	6
17	420	6
18	422	6
19	424	6
20	426	6
21	428	6
22	429	6
23	431	6
24	433	6
25	435	6
26	437	6
27	439	6
28	441	6
29	443	6
30	445	6
31	448	7
32	450	7
33	453	7
34	455	7
35	458	8
36	461	8
37	464	9
38	468	9
39	472	10
40	477	11

Raw Score	Scale Score	Standard Error
41	482	12
42	485	12
43	489	13
44	492	14
45	496	15
46	499	16
47	503	17

Table P6. ELA Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error
0	366	16
1	370	15
2	373	14
3	377	13
4	381	12
5	386	10
6	391	9
7	395	8
8	399	8
9	402	7
10	405	7
11	407	7
12	410	6
13	412	6
14	414	6
15	416	6
16	418	6
17	420	6
18	422	6
19	424	6
20	426	6
21	428	6
22	430	6
23	432	6
24	433	6
25	435	6
26	437	6
27	439	6
28	441	6
29	443	6
30	445	6
31	448	7

Raw Score	Scale Score	Standard Error
32	450	7
33	452	7
34	455	7
35	458	8
36	461	8
37	464	8
38	468	9
39	472	10
40	477	10
41	482	11
42	486	12
43	489	13
44	493	14
45	496	15
46	500	16
47	503	17

Table P7. Mathematics Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error
0	377	15
1	380	14
2	383	13
3	387	12
4	393	10
5	398	9
6	403	9
7	407	8
8	410	8
9	413	7
10	416	7
11	419	7
12	422	7
13	424	7
14	426	7
15	429	7
16	431	6
17	434	6
18	436	6
19	438	7
20	440	7
21	443	7
22	445	7

Raw Score	Scale Score	Standard Error
23	448	7
24	450	7
25	453	7
26	455	7
27	458	8
28	461	8
29	464	8
30	468	9
31	472	9
32	476	10
33	481	11
34	487	12
35	490	13
36	494	14
37	497	15
38	501	16

Table P8. Mathematics Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error
0	378	14
1	381	13
2	385	12
3	388	12
4	391	11
5	397	10
6	401	9
7	405	9
8	409	8
9	413	8
10	416	7
11	418	7
12	421	7
13	424	7
14	426	7
15	429	7
16	431	6
17	433	6
18	435	6
19	437	6
20	440	6
21	442	6
22	444	6

Raw Score	Scale Score	Standard Error
23	446	6
24	448	6
25	450	6
26	452	6
27	454	7
28	457	7
29	459	7
30	461	7
31	464	7
32	466	7
33	469	8
34	472	8
35	475	8
36	478	9
37	482	9
38	486	10
39	491	11
40	494	12
41	498	13
42	501	14
43	504	15
44	507	16

Table P9. Mathematics Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error
0	381	14
1	384	13
2	387	12
3	390	11
4	397	10
5	402	9
6	406	8
7	410	7
8	413	7
9	416	7
10	418	6
11	421	6
12	423	6
13	426	6
14	428	6
15	430	6
16	432	6

Raw Score	Scale Score	Standard Error
17	434	6
18	436	6
19	438	6
20	440	6
21	442	6
22	444	6
23	446	6
24	448	6
25	450	6
26	452	6
27	454	6
28	456	6
29	458	6
30	460	6
31	463	7
32	465	7
33	467	7
34	470	7
35	473	8
36	476	8
37	479	8
38	483	9
39	487	10
40	492	11
41	496	12
42	499	13
43	502	14
44	506	15

Table P10. Mathematics Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error
0	388	14
1	391	13
2	394	12
3	398	11
4	401	10
5	406	9
6	410	8
7	413	8
8	417	7
9	419	7
10	422	7

Raw Score	Scale Score	Standard Error
11	424	7
12	427	6
13	429	6
14	431	6
15	433	6
16	435	6
17	436	6
18	438	6
19	440	6
20	442	6
21	443	6
22	445	6
23	447	6
24	448	6
25	450	6
26	452	6
27	453	6
28	455	6
29	457	6
30	459	6
31	461	6
32	463	6
33	465	7
34	467	7
35	470	7
36	473	7
37	475	8
38	478	8
39	482	9
40	485	9
41	490	10
42	494	10
43	500	11
44	503	12
45	506	13
46	509	14
47	513	15

Table P11. Mathematics Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error
0	379	15
1	382	13
2	386	12
3	389	11
4	392	11
5	398	9
6	402	8
7	406	8
8	409	7
9	412	7
10	415	7
11	417	6
12	419	6
13	422	6
14	424	6
15	426	6
16	428	6
17	430	6
18	431	6
19	433	5
20	435	5
21	437	5
22	438	5
23	440	5
24	442	5
25	443	5
26	445	5
27	447	5
28	448	6
29	450	6
30	452	6
31	454	6
32	455	6
33	457	6
34	459	6
35	461	6
36	464	6
37	466	7
38	468	7
39	471	7
40	474	8

Raw Score	Scale Score	Standard Error
41	477	8
42	481	9
43	485	9
44	490	10
45	496	12
46	499	13
47	502	13
48	506	14
49	509	15

Table P12. Mathematics Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error
0	379	16
1	382	15
2	385	14
3	388	13
4	391	12
5	395	12
6	398	11
7	403	10
8	407	9
9	411	9
10	414	8
11	417	8
12	420	8
13	423	7
14	425	7
15	428	7
16	430	7
17	432	7
18	434	7
19	436	6
20	438	6
21	440	6
22	442	6
23	444	6
24	446	6
25	448	6
26	450	6
27	452	6
28	454	6
29	456	6

Raw Score	Scale Score	Standard Error
30	458	6
31	460	6
32	462	7
33	464	7
34	466	7
35	468	7
36	471	7
37	473	8
38	476	8
39	479	8
40	482	9
41	486	9
42	490	10
43	495	11
44	498	12
45	501	13
46	505	13
47	508	14
48	511	15
49	514	16

Table P13. ELA Grade 3 Scale Score Frequency Distribution

Scale			Cumulative	
Score	Freq.	%	Freq.	%
380	28	0.02	28	0.0
384	94	0.06	122	0.1
387	265	0.16	387	0.2
391	713	0.43	1,100	0.7
397	1,353	0.81	2,453	1.5
402	2,305	1.39	4,758	2.9
405	3,116	1.88	7,874	4.7
409	3,774	2.27	11,648	7.0
412	4,252	2.56	15,900	9.6
415	4,522	2.72	20,422	12.3
417	4,894	2.95	25,316	15.2
420	4,881	2.94	30,197	18.2
422	4,877	2.94	35,074	21.1
425	5,021	3.02	40,095	24.1
427	5,157	3.10	45,252	27.2
430	5,297	3.19	50,549	30.4
432	5,478	3.30	56,027	33.7
434	5,678	3.42	61,705	37.1

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
437	5,711	3.44	67,416	40.6
439	6,042	3.64	73,458	44.2
442	6,169	3.71	79,627	47.9
444	6,566	3.95	86,193	51.9
447	6,973	4.20	93,166	56.1
450	7,258	4.37	100,424	60.4
453	7,737	4.66	108,161	65.1
456	8,275	4.98	116,436	70.1
460	8,441	5.08	124,877	75.2
464	8,472	5.10	133,349	80.3
468	8,396	5.05	141,745	85.3
474	7,731	4.65	149,476	90.0
480	6,967	4.19	156,443	94.2
484	5,266	3.17	161,709	97.3
487	3,267	1.97	164,976	99.3
491	1,179	0.71	166,155	100

Table P14. ELA Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
380	19	0.01	19	0.0
384	62	0.04	81	0.0
387	190	0.11	271	0.2
391	581	0.35	852	0.5
394	1,059	0.64	1,911	1.2
400	1,674	1.01	3,585	2.2
404	2,347	1.41	5,932	3.6
408	2,831	1.70	8,763	5.3
412	3,292	1.98	12,055	7.3
415	3,657	2.20	15,712	9.5
418	3,905	2.35	19,617	11.8
421	4,323	2.60	23,940	14.4
423	4,530	2.73	28,470	17.1
426	4,831	2.91	33,301	20.0
428	5,178	3.12	38,479	23.2
431	5,263	3.17	43,742	26.3
433	5,551	3.34	49,293	29.7
435	5,849	3.52	55,142	33.2
438	6,090	3.66	61,232	36.8
440	6,247	3.76	67,479	40.6
443	6,372	3.83	73,851	44.4

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
445	6,471	3.89	80,322	48.3
447	6,545	3.94	86,867	52.3
450	6,813	4.10	93,680	56.4
453	7,062	4.25	100,742	60.6
455	7,075	4.26	107,817	64.9
458	7,077	4.26	114,894	69.1
461	7,220	4.34	122,114	73.5
464	6,949	4.18	129,063	77.7
468	6,979	4.20	136,042	81.9
471	6,629	3.99	142,671	85.9
476	6,276	3.78	148,947	89.6
481	5,453	3.28	154,400	92.9
486	4,669	2.81	159,069	95.7
490	3,331	2.00	162,400	97.7
493	2,271	1.37	164,671	99.1
497	1,136	0.68	165,807	99.8
500	366	0.22	166,173	100

Table P15. ELA Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
371	12	0.01	12	0.0
375	24	0.01	36	0.0
378	76	0.05	112	0.1
382	211	0.13	323	0.2
385	414	0.25	737	0.4
391	679	0.41	1,416	0.9
395	1,112	0.67	2,528	1.5
399	1,433	0.87	3,961	2.4
403	1,808	1.09	5,769	3.5
406	2,093	1.27	7,862	4.8
409	2,517	1.52	10,379	6.3
412	3,012	1.82	13,391	8.1
415	3,383	2.05	16,774	10.2
417	3,807	2.30	20,581	12.5
420	4,135	2.50	24,716	15.0
422	4,518	2.73	29,234	17.7
425	4,993	3.02	34,227	20.7
427	5,162	3.12	39,389	23.8
430	5,713	3.46	45,102	27.3
432	5,865	3.55	50,967	30.8

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
435	6,323	3.83	57,290	34.7
437	6,634	4.01	63,924	38.7
439	6,907	4.18	70,831	42.9
442	6,871	4.16	77,702	47.0
445	7,172	4.34	84,874	51.4
447	7,439	4.50	92,313	55.9
450	7,494	4.53	99,807	60.4
453	7,381	4.47	107,188	64.9
456	7,425	4.49	114,613	69.4
459	7,309	4.42	121,922	73.8
462	7,258	4.39	129,180	78.2
466	6,808	4.12	135,988	82.3
470	6,564	3.97	142,552	86.3
474	5,883	3.56	148,435	89.8
479	5,147	3.11	153,582	92.9
485	4,165	2.52	157,747	95.5
488	3,229	1.95	160,976	97.4
492	2,284	1.38	163,260	98.8
495	1,264	0.76	164,524	99.6
499	592	0.36	165,116	99.9
502	143	0.09	165,259	100

Table P16. ELA Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
366	3	0.00	3	0.0
369	21	0.01	24	0.0
373	57	0.03	81	0.0
376	114	0.07	195	0.1
380	278	0.17	473	0.3
386	535	0.32	1,008	0.6
391	732	0.44	1,740	1.1
395	1,128	0.68	2,868	1.7
398	1,511	0.92	4,379	2.7
402	1,832	1.11	6,211	3.8
405	2,033	1.23	8,244	5.0
407	2,421	1.47	10,665	6.5
410	2,685	1.63	13,350	8.1
413	2,966	1.80	16,316	9.9
415	3,177	1.92	19,493	11.8
418	3,471	2.10	22,964	13.9

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
420	3,546	2.15	26,510	16.1
422	3,914	2.37	30,424	18.4
425	4,282	2.59	34,706	21.0
427	4,649	2.82	39,355	23.8
429	4,811	2.91	44,166	26.8
431	5,274	3.20	49,440	30.0
434	5,663	3.43	55,103	33.4
436	6,164	3.73	61,267	37.1
439	6,543	3.96	67,810	41.1
441	7,048	4.27	74,858	45.4
444	7,641	4.63	82,499	50.0
447	8,038	4.87	90,537	54.9
450	8,483	5.14	99,020	60.0
453	8,871	5.37	107,891	65.4
457	9,109	5.52	117,000	70.9
461	9,321	5.65	126,321	76.5
465	9,055	5.49	135,376	82.0
470	8,272	5.01	143,648	87.0
476	7,321	4.44	150,969	91.5
480	5,792	3.51	156,761	95.0
483	4,141	2.51	160,902	97.5
487	2,524	1.53	163,426	99.0
490	1,203	0.73	164,629	99.7
494	357	0.22	164,986	100
497	65	0.04	165,051	100

Table P17. ELA Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
366	7	0.00	7	0.0
369	13	0.01	20	0.0
373	13	0.01	33	0.0
376	54	0.03	87	0.1
380	97	0.06	184	0.1
386	177	0.11	361	0.2
391	359	0.22	720	0.4
395	602	0.38	1,322	0.8
398	882	0.55	2,204	1.4
401	1,164	0.73	3,368	2.1
404	1,406	0.88	4,774	3.0
407	1,730	1.08	6,504	4.1

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
409	1,963	1.22	8,467	5.3
412	2,245	1.40	10,712	6.7
414	2,418	1.51	13,130	8.2
416	2,618	1.63	15,748	9.8
418	2,767	1.72	18,515	11.5
420	2,869	1.79	21,384	13.3
422	3,138	1.96	24,522	15.3
424	3,363	2.10	27,885	17.4
426	3,445	2.15	31,330	19.5
428	3,653	2.28	34,983	21.8
429	3,906	2.43	38,889	24.2
431	4,073	2.54	42,962	26.8
433	4,217	2.63	47,179	29.4
435	4,461	2.78	51,640	32.2
437	4,709	2.93	56,349	35.1
439	4,863	3.03	61,212	38.1
441	5,252	3.27	66,464	41.4
443	5,366	3.34	71,830	44.8
445	5,851	3.65	77,681	48.4
448	6,154	3.84	83,835	52.2
450	6,426	4.00	90,261	56.2
453	6,684	4.17	96,945	60.4
455	6,864	4.28	103,809	64.7
458	7,011	4.37	110,820	69.1
461	7,117	4.44	117,937	73.5
464	6,954	4.33	124,891	77.8
468	6,691	4.17	131,582	82.0
472	6,456	4.02	138,038	86.0
477	5,697	3.55	143,735	89.6
482	5,208	3.25	148,943	92.8
485	4,301	2.68	153,244	95.5
489	3,126	1.95	156,370	97.4
492	2,238	1.39	158,608	98.8
496	1,238	0.77	159,846	99.6
499	505	0.31	160,351	99.9
503	116	0.07	160,467	100

Table P18. ELA Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
366	8	0.01	8	0.0
370	11	0.01	19	0.0
373	19	0.01	38	0.0
377	47	0.03	85	0.1
381	119	0.08	204	0.1
386	210	0.14	414	0.3
391	353	0.23	767	0.5
395	610	0.40	1,377	0.9
399	813	0.53	2,190	1.4
402	1,024	0.67	3,214	2.1
405	1,256	0.83	4,470	2.9
407	1,390	0.91	5,860	3.8
410	1,580	1.04	7,440	4.9
412	1,781	1.17	9,221	6.1
414	1,736	1.14	10,957	7.2
416	1,910	1.25	12,867	8.5
418	1,974	1.30	14,841	9.8
420	2,158	1.42	16,999	11.2
422	2,297	1.51	19,296	12.7
424	2,412	1.58	21,708	14.3
426	2,488	1.63	24,196	15.9
428	2,757	1.81	26,953	17.7
430	2,894	1.90	29,847	19.6
432	3,121	2.05	32,968	21.7
433	3,431	2.25	36,399	23.9
435	3,653	2.40	40,052	26.3
437	3,977	2.61	44,029	28.9
439	4,132	2.71	48,161	31.6
441	4,594	3.02	52,755	34.7
443	5,088	3.34	57,843	38.0
445	5,226	3.43	63,069	41.4
448	5,674	3.73	68,743	45.2
450	6,365	4.18	75,108	49.3
452	6,553	4.31	81,661	53.6
455	7,096	4.66	88,757	58.3
458	7,297	4.79	96,054	63.1
461	7,653	5.03	103,707	68.1
464	7,866	5.17	111,573	73.3
468	7,666	5.04	119,239	78.3
472	7,562	4.97	126,801	83.3

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
477	6,863	4.51	133,664	87.8
482	6,111	4.01	139,775	91.8
486	4,827	3.17	144,602	95.0
489	3,581	2.35	148,183	97.4
493	2,325	1.53	150,508	98.9
496	1,133	0.74	151,641	99.6
500	484	0.32	152,125	99.9
503	87	0.06	152,212	100

Table P19. Mathematics Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
377	12	0.01	12	0.0
380	41	0.02	53	0.0
383	122	0.07	175	0.1
387	356	0.21	531	0.3
393	775	0.46	1,306	0.8
398	1,449	0.86	2,755	1.6
403	2,210	1.30	4,965	2.9
407	2,841	1.68	7,806	4.6
410	3,353	1.98	11,159	6.6
413	3,868	2.28	15,027	8.9
416	4,165	2.46	19,192	11.3
419	4,272	2.52	23,464	13.8
422	4,462	2.63	27,926	16.5
424	4,555	2.69	32,481	19.2
426	4,659	2.75	37,140	21.9
429	4,639	2.74	41,779	24.7
431	4,720	2.79	46,499	27.4
434	4,776	2.82	51,275	30.3
436	4,813	2.84	56,088	33.1
438	4,873	2.88	60,961	36.0
440	4,809	2.84	65,770	38.8
443	5,065	2.99	70,835	41.8
445	4,998	2.95	75,833	44.8
448	5,095	3.01	80,928	47.8
450	5,318	3.14	86,246	50.9
453	5,473	3.23	91,719	54.1
455	5,500	3.25	97,219	57.4
458	5,746	3.39	102,965	60.8
461	5,974	3.53	108,939	64.3

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
464	6,167	3.64	115,106	67.9
468	6,308	3.72	121,414	71.7
472	6,569	3.88	127,983	75.5
476	6,826	4.03	134,809	79.6
481	7,014	4.14	141,823	83.7
487	7,014	4.14	148,837	87.8
490	6,869	4.05	155,706	91.9
494	6,160	3.64	161,866	95.5
497	4,871	2.87	166,737	98.4
501	2,707	1.60	169,444	100

Table P20. Mathematics Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
378	13	0.01	13	0.0
381	38	0.02	51	0.0
385	93	0.05	144	0.1
388	303	0.18	447	0.3
391	747	0.44	1,194	0.7
397	1,291	0.76	2,485	1.5
401	2,045	1.21	4,530	2.7
405	2,841	1.68	7,371	4.4
409	3,432	2.03	10,803	6.4
413	3,955	2.34	14,758	8.7
416	4,152	2.45	18,910	11.2
418	4,303	2.54	23,213	13.7
421	4,281	2.53	27,494	16.2
424	4,376	2.58	31,870	18.8
426	4,305	2.54	36,175	21.4
429	4,346	2.57	40,521	23.9
431	4,490	2.65	45,011	26.6
433	4,356	2.57	49,367	29.2
435	4,424	2.61	53,791	31.8
437	4,491	2.65	58,282	34.4
440	4,506	2.66	62,788	37.1
442	4,488	2.65	67,276	39.7
444	4,516	2.67	71,792	42.4
446	4,583	2.71	76,375	45.1
448	4,437	2.62	80,812	47.7
450	4,537	2.68	85,349	50.4
452	4,591	2.71	89,940	53.1

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
454	4,574	2.70	94,514	55.8
457	4,801	2.84	99,315	58.7
459	4,728	2.79	104,043	61.5
461	4,936	2.92	108,979	64.4
464	4,869	2.88	113,848	67.2
466	4,994	2.95	118,842	70.2
469	5,080	3.00	123,922	73.2
472	5,113	3.02	129,035	76.2
475	4,975	2.94	134,010	79.2
478	5,112	3.02	139,122	82.2
482	5,094	3.01	144,216	85.2
486	4,978	2.94	149,194	88.1
491	4,857	2.87	154,051	91.0
494	4,524	2.67	158,575	93.7
498	3,903	2.31	162,478	96.0
501	3,241	1.91	165,719	97.9
504	2,284	1.35	168,003	99.2
507	1,290	0.76	169,293	100

Table P21. Mathematics Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
381	15	0.01	15	0.0
384	66	0.04	81	0.0
387	199	0.12	280	0.2
390	604	0.36	884	0.5
397	1,189	0.71	2,073	1.2
402	1,998	1.19	4,071	2.4
406	3,110	1.86	7,181	4.3
410	3,788	2.27	10,969	6.6
413	4,322	2.58	15,291	9.1
416	4,535	2.71	19,826	11.9
418	4,605	2.75	24,431	14.6
421	4,575	2.74	29,006	17.3
423	4,487	2.68	33,493	20.0
426	4,609	2.76	38,102	22.8
428	4,450	2.66	42,552	25.4
430	4,434	2.65	46,986	28.1
432	4,517	2.70	51,503	30.8
434	4,541	2.72	56,044	33.5
436	4,366	2.61	60,410	36.1

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
438	4,409	2.64	64,819	38.8
440	4,498	2.69	69,317	41.4
442	4,446	2.66	73,763	44.1
444	4,197	2.51	77,960	46.6
446	4,323	2.58	82,283	49.2
448	4,262	2.55	86,545	51.7
450	4,262	2.55	90,807	54.3
452	4,194	2.51	95,001	56.8
454	4,190	2.51	99,191	59.3
456	4,122	2.46	103,313	61.8
458	4,152	2.48	107,465	64.3
460	4,101	2.45	111,566	66.7
463	4,189	2.50	115,755	69.2
465	4,151	2.48	119,906	71.7
467	4,202	2.51	124,108	74.2
470	4,192	2.51	128,300	76.7
473	4,246	2.54	132,546	79.3
476	4,263	2.55	136,809	81.8
479	4,407	2.64	141,216	84.4
483	4,473	2.67	145,689	87.1
487	4,319	2.58	150,008	89.7
492	4,442	2.66	154,450	92.4
496	4,016	2.40	158,466	94.8
499	3,706	2.22	162,172	97.0
502	3,115	1.86	165,287	98.8
506	1,951	1.17	167,238	100

Table P22. Mathematics Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
388	16	0.01	16	0.0
391	74	0.04	90	0.1
394	245	0.15	335	0.2
398	571	0.35	906	0.5
401	1,445	0.88	2,351	1.4
406	2,538	1.54	4,889	3.0
410	3,637	2.21	8,526	5.2
413	4,767	2.89	13,293	8.1
417	5,456	3.31	18,749	11.4
419	5,699	3.46	24,448	14.8
422	5,553	3.37	30,001	18.2

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
424	5,227	3.17	35,228	21.4
427	4,973	3.02	40,201	24.4
429	4,580	2.78	44,781	27.2
431	4,307	2.61	49,088	29.8
433	4,160	2.52	53,248	32.3
435	4,001	2.43	57,249	34.7
436	3,865	2.35	61,114	37.1
438	3,843	2.33	64,957	39.4
440	3,777	2.29	68,734	41.7
442	3,734	2.27	72,468	44.0
443	3,747	2.27	76,215	46.2
445	3,756	2.28	79,971	48.5
447	3,554	2.16	83,525	50.7
448	3,690	2.24	87,215	52.9
450	3,624	2.20	90,839	55.1
452	3,580	2.17	94,419	57.3
453	3,641	2.21	98,060	59.5
455	3,670	2.23	101,730	61.7
457	3,599	2.18	105,329	63.9
459	3,735	2.27	109,064	66.2
461	3,775	2.29	112,839	68.5
463	3,808	2.31	116,647	70.8
465	3,801	2.31	120,448	73.1
467	3,832	2.33	124,280	75.4
470	3,836	2.33	128,116	77.7
473	3,707	2.25	131,823	80.0
475	3,777	2.29	135,600	82.3
478	3,770	2.29	139,370	84.6
482	3,523	2.14	142,893	86.7
485	3,700	2.25	146,593	89.0
490	3,474	2.11	150,067	91.1
494	3,335	2.02	153,402	93.1
500	3,165	1.92	156,567	95.0
503	2,778	1.69	159,345	96.7
506	2,339	1.42	161,684	98.1
509	1,924	1.17	163,608	99.3
513	1,184	0.72	164,792	100

Table P23. Mathematics Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
379	12	0.01	12	0.0
382	30	0.02	42	0.0
386	57	0.04	99	0.1
389	140	0.09	239	0.2
392	296	0.19	535	0.3
398	636	0.40	1,171	0.7
402	1,093	0.69	2,264	1.4
406	1,704	1.08	3,968	2.5
409	2,237	1.41	6,205	3.9
412	2,847	1.80	9,052	5.7
415	3,315	2.09	12,367	7.8
417	3,480	2.20	15,847	10.0
419	3,658	2.31	19,505	12.3
422	3,704	2.34	23,209	14.7
424	3,747	2.37	26,956	17.0
426	3,650	2.31	30,606	19.3
428	3,751	2.37	34,357	21.7
430	3,566	2.25	37,923	24.0
431	3,637	2.30	41,560	26.2
433	3,605	2.28	45,165	28.5
435	3,498	2.21	48,663	30.7
437	3,570	2.25	52,233	33.0
438	3,445	2.18	55,678	35.2
440	3,633	2.29	59,311	37.5
442	3,501	2.21	62,812	39.7
443	3,490	2.20	66,302	41.9
445	3,517	2.22	69,819	44.1
447	3,504	2.21	73,323	46.3
448	3,622	2.29	76,945	48.6
450	3,642	2.30	80,587	50.9
452	3,573	2.26	84,160	53.2
454	3,722	2.35	87,882	55.5
455	3,788	2.39	91,670	57.9
457	3,873	2.45	95,543	60.3
459	3,796	2.40	99,339	62.7
461	3,949	2.49	103,288	65.2
464	3,942	2.49	107,230	67.7
466	4,061	2.56	111,291	70.3
468	4,164	2.63	115,455	72.9
471	4,284	2.71	119,739	75.6

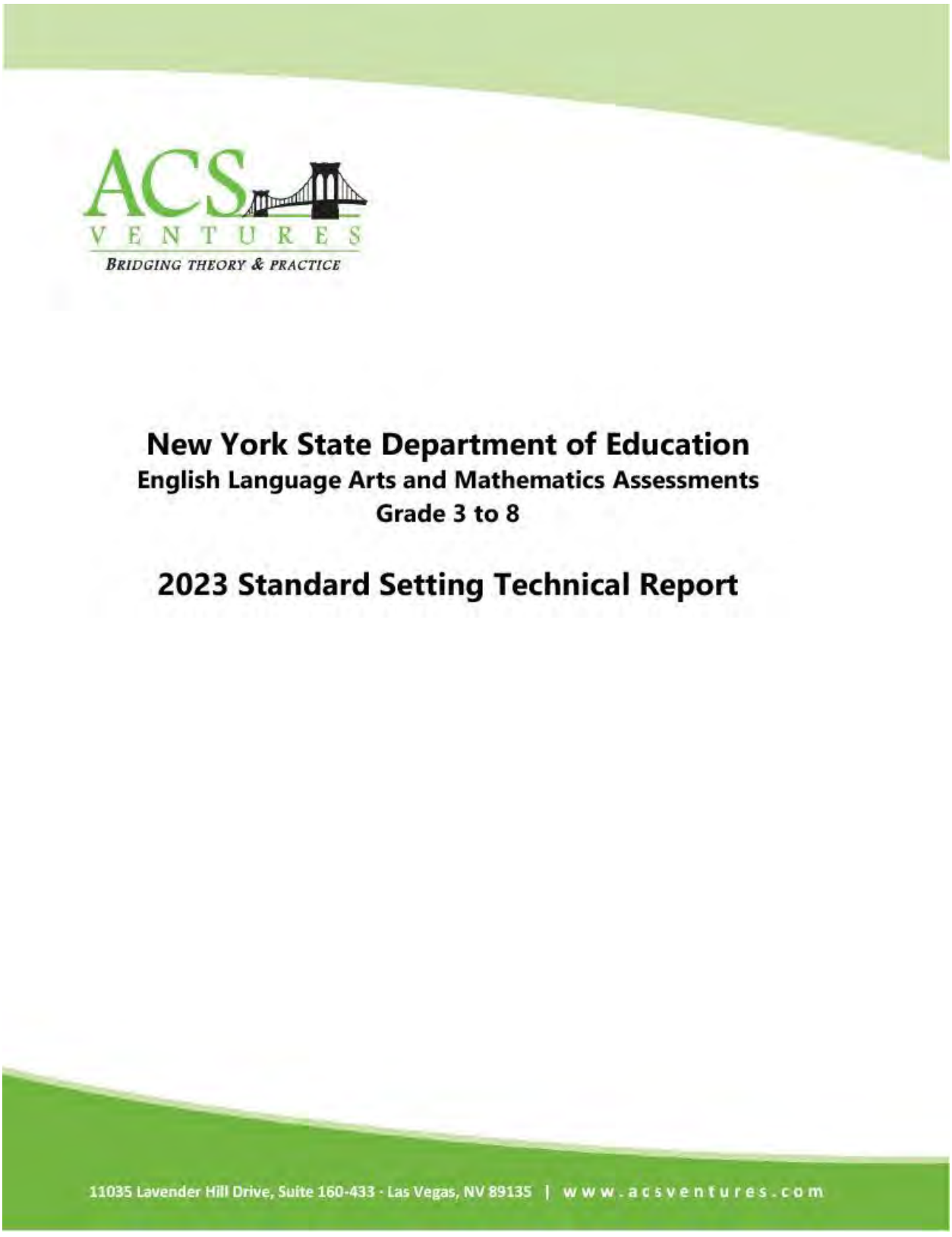
Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
474	4,253	2.69	123,992	78.3
477	4,299	2.72	128,291	81.0
481	4,308	2.72	132,599	83.7
485	4,456	2.81	137,055	86.6
490	4,324	2.73	141,379	89.3
496	4,162	2.63	145,541	91.9
499	4,096	2.59	149,637	94.5
502	3,684	2.33	153,321	96.8
506	3,120	1.97	156,441	98.8
509	1,898	1.20	158,339	100

Table P24. Mathematics Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
379	22	22	0.02	0.0
382	38	60	0.04	0.1
385	66	126	0.06	0.1
388	178	304	0.17	0.3
391	350	654	0.34	0.6
395	783	1,437	0.76	1.4
398	1,479	2,916	1.44	2.8
403	2,162	5,078	2.11	5.0
407	2,837	7,915	2.77	7.7
411	3,319	11,234	3.24	11.0
414	3,647	14,881	3.56	14.5
417	3,720	18,601	3.63	18.1
420	3,644	22,245	3.55	21.7
423	3,638	25,883	3.55	25.2
425	3,318	29,201	3.24	28.5
428	3,234	32,435	3.15	31.6
430	3,196	35,631	3.12	34.7
432	3,061	38,692	2.98	37.7
434	2,966	41,658	2.89	40.6
436	2,817	44,475	2.75	43.4
438	2,778	47,253	2.71	46.1
440	2,782	50,035	2.71	48.8
442	2,670	52,705	2.60	51.4
444	2,620	55,325	2.55	53.9
446	2,488	57,813	2.43	56.4
448	2,516	60,329	2.45	58.8
450	2,532	62,861	2.47	61.3

Scale Score	Freq.	%	Cumulative Freq.	Cumulative %
452	2,492	65,353	2.43	63.7
454	2,351	67,704	2.29	66.0
456	2,413	70,117	2.35	68.4
458	2,340	72,457	2.28	70.6
460	2,288	74,745	2.23	72.9
462	2,203	76,948	2.15	75.0
464	2,181	79,129	2.13	77.2
466	2,119	81,248	2.07	79.2
468	2,144	83,392	2.09	81.3
471	2,057	85,449	2.01	83.3
473	2,001	87,450	1.95	85.3
476	1,956	89,406	1.91	87.2
479	1,795	91,201	1.75	88.9
482	1,787	92,988	1.74	90.7
486	1,712	94,700	1.67	92.3
490	1,621	96,321	1.58	93.9
495	1,444	97,765	1.41	95.3
498	1,300	99,065	1.27	96.6
501	1,180	100,245	1.15	97.7
505	888	101,133	0.87	98.6
508	760	101,893	0.74	99.3
511	485	102,378	0.47	99.8
514	182	102,560	0.18	100

Appendix Q: Standard Setting Technical Report



Background

The New York State Department of Education (NYSED) has partnered with NWEA and ACS Ventures, LLC (ACS) to establish cut points for the New York English Language Arts (ELA) and Mathematics (Math) assessments for grades 3 to 8. Revisions were made to the blueprints of the assessments starting with the 2023 administration and as a result, updated cut scores for the assessments were required. The updated cut scores were designed to: (a) reflect the current New York State standards, (b) link students' scores on the assessments to the state's expectations for students in each performance level, and (c) are well articulated across grades and courses.

On August 1 – 3, 2023, NYSED partnered with NWEA and ACS to conduct a standard setting study for the assessments in ELA and Math. The bookmark standard setting methodology was used to establish the recommendations for three cut scores for each assessment. The definitions for the performance levels for the assessments are designed to represent students' knowledge and skills defined by the Performance Level Descriptors (PLDs). The performance levels are *Level 1, Level 2, Level 3, and Level 4.*

Standard Setting Plan and Methodology

During the 2022-23 academic year, NWEA, in partnership with ACS and NYSED, developed a plan for the standard setting study for the assessments. A standard setting plan was developed and reviewed with the NYSED Technical Advisory Committee in the fall of 2022 and again in the spring of 2023. Based on this plan, a total of 65 New York State educators engaged in a Bookmark standard setting procedure (Lewis, Mitzel, Mercado, & Schulz, 2012).

Panelists

The number of panelists in each panel, along with the assessments (content) they reviewed, is included in Table 1 below.

Table 1: Composition of Standard Setting Panels

Panel	Content	Grade Level	Panelists
Panel 1:	English Language Arts	Grades 3 to 4	11
Panel 2:	English Language Arts	Grades 5 to 6	12
Panel 3:	English Language Arts	Grades 7 to 8	10
Panel 4:	Mathematics	Grades 3 to 4	12
Panel 5:	Mathematics	Grades 5 to 6	10
Panel 6:	Mathematics	Grades 7 to 8	10

For the standard setting, NYSED gathered educators and stakeholders from across New York State. NYSED recruited and invited panelists to the workshop. NYSED took special care to invite workshop panelists who:

1. were well qualified and had knowledge of the current NYSED standards (e.g., had experience teaching in New York State classrooms),
2. were diverse in terms of demographic characteristics (e.g., gender),
3. were diverse in terms of geographic location within the state, and

4. had knowledge of the tested content and population, including panelists who had experience working with students with disabilities (SWDs) and English language learners (ELLs).

Prior to the workshop, panelists completed a brief demographic survey about their experience in education. The demographic survey and results can be found in Appendix A. As shown in Table 2 across all panels, slightly greater than 80% of panelists identified as classroom teachers. The remainder of panelists identified roles such as principal, test coordinator, or instructional coach.

Table 2: Demographic Survey Results for the standard setting panelists

		What is your current position?				
Content	Grade	Teacher	Teacher (Higher Education)	Administrator (School)	Administrator (District)	Other
ELA	Grade 3/4	7	–	1	1	--
ELA	Grade 5/6	10	–	–	–	--
ELA	Grade 7/8	9	–	--	–	1
Math	Grade 3/4	8	–	2	–	1
Math	Grade 5/6	7	–	2	–	--
Math	Grade 7/8	8	–	1	–	--
Total		49	0	6	1	2

Staffing

The standard setting workshop was staffed with a combination of personnel from NYSED, NWEA, and ACS. Staff members in attendance from NYSED included Zachary Warner Assistant Commissioner as well as Clara DeSorbo, Paul Anderson, Robin Parker, Sz-Shyan Wu, Chunmei Zheng; Tianli Li, and Laura Bogardus. The ACS team provided facilitation of the workshop. Andrew Wiley served as the lead facilitator throughout the workshop and served as a floating facilitator throughout the workshop and assisted with panels as needed. ACS supplied SIX additional facilitators, one for each panel. All facilitators had previously completed multiple standard setting workshops within the educational assessment arena. From NWEA, Monèt Phillips served as the project management lead throughout the planning for the workshop and at the workshop itself. Yanming Jiang and Chalie Patarapapichayatham from NWEA served as psychometricians for the project, completed all item calibration work prior to the workshop, and was available in-person throughout the workshop.

Workshop Materials

Panelists were provided five hard-copy documents that were used to make performance level recommendations: Performance Level Descriptions (PLDs), Ordered Item Booklets (OIBs), Practice Ordered Item Booklets, Rating sheets to record their ratings, and Passage booklets for the ELA panels.

The PLDs were created in collaboration with a committee of New York State educators and summarize the knowledge, skills, and abilities expected of students in each performance level. Panelists were supplied a copy of the range PLDs and developed threshold PLDs that were used to rate the test items. The threshold PLDs were developed, discussed, and reconciled during the workshop, and hard copies of the threshold PLDs were prepared and distributed to the panelists to refer to while completing their ratings.

The OIBs were hard copy versions of a sample of test items administered in the spring of 2023. In the spring of 2023, one core test form was administered at each grade level, including a core test form and an

alternate. All test forms included both single-point and multi-point items. The number of points in each test form ranged from 33 (ELA Grade 3) to 49 points (Math Grades 7 and 8). The OIBs were created by selecting a sample of items using the core test form as a starting point and supplementing the test form with additional items pretested in 2023. The selection of items for the OIBs was designed to ensure the OIBs represented a range of difficulty values and the difficulty level and item types in the operational test forms. The number of points on each test form and in the OIBs are reported in Table 3 below. In addition to the comprehensive item booklets, practice booklets were also developed for both assessments that contained a total of eight points to be rated. Practice OIBs were only created for the first of the two assessments that each panel reviewed (the lower of the two grade levels).

Table 3: Number of points for each core form and for the Ordered Item Booklet

		# of points	
Content	**Grade**	**Test Form**	**OIB**
ELA	3	33	45
ELA	4	37	50
ELA	5	40	50
ELA	6	40	50
ELA	7	47	55
ELA	8	47	55
Math	3	38	52
Math	4	44	55
Math	5	44	55
Math	6	47	60
Math	7	49	60
Math	8	49	60

Prior to the workshop, item difficulty estimates were calculated for all items using data from the Spring 2023 test administration. The item difficulty estimates for 1-point items were determined using the 2PL item response model (Birnbaum, 1968). For multi-point items, the parameters were determined using the Generalized Partial Credit Model (Muraki, 1992; Yen, 1993). Once all item difficulty estimates were determined, the response probability for each was calculated using a response probability of 0.67.

Workshop Overview

The workshop was completed over the course of three days. The agenda provided to panelists can be found in Appendix B. The workshop began with all panelists together and with NYSED staff providing a review of the current state of New York State assessments. Once this review was completed, the lead ACS facilitator reviewed the goals and purpose of the standard setting workshop, and also reviewed the methodology to be followed. Once this was completed, panelists were split into the grade specific panels defined in Table 1.

In the grade specific panels, panelists first reviewed the assessment for their assigned grade level. After this review, they reviewed the Performance Level Descriptions (PLDs) for the grade level and developed threshold PLDs for the grade level. After creating the threshold PLDs, panelists proceeded to complete the Bookmark methodology (further described below). Panelists completed three rounds of ratings in order to finalize their recommendations. Each panel covered two grade levels, with each panel starting with the lower of the two grade levels. Once all grade levels were completed, a vertical articulation was completed within each content area to review the results across all grades.

Performance Level Descriptions

Performance Level Descriptions (PLDs) summarize the knowledge, skills, and abilities expected of students in each performance level. As part of the initial development of the NYSED Grades 3 to 8 assessments, NWEA and NYSED had developed versions of the PLDs that were used to guide the development of test blueprints and test items for the assessments. The PLDs were developed in collaboration with New York State educators who provided initial recommendations for the PLDs, reviewed updated versions of the PLDs, and provided overall feedback on the structure and content of the PLDs. Egan, Schneider, and Ferrara (2012) suggest a framework of four types of PLDs, described below.

Policy PLDs summarize NYSED's definition for each performance level, providing information to stakeholders on the state's suggested interpretation of each level. They are typically not specific to a grade or content area. The policy PLDs for the NYSED grades 3 to 8 assessments in ELA and Math are shown in Table 4 below.

Table 4: Policy PLDs for the NYSED Assessments in ELA and Math

Non - Proficient		Proficient	
Needs Support	**Approaching Proficient**	**Proficient**	**Advanced**
Students performing at this level are below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices embodied by the Learning Standards that are considered insufficient for the expectations at this grade.	Students performing at this level are partially proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the Learning Standards that are considered partial but insufficient for the expectations at this grade. Students performing at Level 2 are considered on track to meet current New York high school graduation requirements but are not yet proficient in Learning Standards at this grade.	Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the Learning Standards that are considered sufficient for the expectations at this grade.	Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the Learning Standards that are considered more than sufficient for the expectations at this grade.

Range PLDs summarize the knowledge, skills, and abilities expected of students in each performance level on a specific test. The range PLDs show the types of content, as informed by the state content standards, which should be mastered by students in each performance level on the test. The range PLDs generally show these expectations for students across the range of performance for the performance level: for example, the Level 3 PLD for a test summarizes skills held by students who are just barely in Level 3 while also covering students that are close to demonstrating performance consistent with Level 4. Range PLDs are often shared with teachers and schools to help them understand the level of construct mastery expected of students in each performance level on each test.

Threshold PLDs are based on the range PLDs. They summarize the knowledge, skills, and abilities expected of students who are at the point-of-entry (the threshold) of each performance level. For any given test, these descriptors show the types of skills needed just to be classified in each level (e.g., just to be classified in Level 3). These PLDs specify the content expectations for students with performance analogous to the cut points. These descriptors are typically used by panelists at the standard setting workshops to help inform decisions they make about cut points.

Reporting PLDs are the version of the PLDs used for score reporting. Typically, the reporting PLDs comprise a version of the policy or range PLDs, and the language in the reporting PLDs is adjusted to be accessible to a wide audience that may not have in-depth content knowledge.

The current policy and range PLDs were provided to panelists at the standard setting workshop to inform their judgments. At the workshop, panelists discussed these PLDs, and panelists developed informal threshold PLDs to inform their cut score recommendations. The threshold PLDs developed by each panel are included in Appendix C. Reporting PLDs were not within the scope of the performance level setting.

Workshop Procedures

At the beginning of the workshop, all panelists convened in a single room for the opening session. The workshop started with an overview by NYSED staff, led by Zachary Warner. NYSED reviewed the development of the assessments as well as their expectations for appropriately rigorous cut score recommendations. NYSED also summarized the process that would be used to review and approve the cut points after the workshop.

The lead facilitator then provided an overview of the goals and objectives of the workshop and training on the procedures that would be followed during the workshop. Dr. Wiley restated the purpose of the workshop and described the standard setting method. Dr. Wiley also described the roles of the facilitators and panelists to the committee. Panelists were informed that they would be studying the OIBs, discussing the content-based expectations for students in each performance level, and providing item-level ratings using the Bookmark methodology. Panelists were reminded to keep the workshop materials confidential. Panelists then adjourned to their pre-assigned content groups.

Reviewing the Online Test

Within each breakout room, the facilitator welcomed panelists and led informal introductions. The first activity began when panelists were given the opportunity to review an online test form to better understand the nature of the content and item types on the assessment. Panelists were not required to answer the items but were encouraged to review and consider the expectations for items and how well prepared their students were for the assessment.

Discussion of Threshold Students

After reviewing the online test form, the group discussed the PLDs and behavior consistent with threshold performance. The individual PLD review activity then transitioned into small group discussion about the distinctions between performance levels. Each small group of panelists discussed and noted the differences in performance between students exhibiting behavior consistent with *Level 1* as compared to *Level 2*. Panelists also discussed the differences in behavior across levels 2 and 3, and finally with levels 3 and 4. The small groups organized and documented their threshold PLD recommendations and then shared these with the panel. Each small group then shared the rationale for the contents of their draft threshold PLDs and engaged in a whole group discussion. The panel facilitator captured the discussion and the reconciliation of the drafts into the panel's consensus threshold PLDs. These threshold PLDs were printed and distributed for use as an additional resource by the panelists during the standard setting activities. The threshold PLDs for all content areas can be found in Appendix C.

Item Ratings

After developing the threshold PLDs, panelists were provided further training on the Bookmark procedure and how they would review the test and complete their ratings. The training first explained the structure of the OIB and how it was created to arrange the items in the order of statistical difficulty. It was also explained how the 1-point items would appear once in each booklet (single score point possible) while multi-point items could appear on multiple pages (once for each score point possible). As the panelists reviewed each item in the OIB, they were instructed to take the following steps:

- Identify the knowledge and skills required to answer each item correctly.
- Determine the knowledge and/or skill requirements of later items that make them more difficult.
- Consider whether the student described in the threshold PLD should answer the item correctly.
- Place a bookmark at the first point where 2/3 of students at the threshold are NOT expected to respond correctly.

The process for recording bookmark placements for each cut score was described. For the practice and all three rounds of ratings, panelists provided their ratings via a hard-copy rating form submitted to the facilitators. Panelists were given the opportunity to ask questions or raise concerns about the process and all questions were addressed by the facilitator prior to the practice round.

To wrap up the training, panelists were provided an opportunity for a practice round with the Bookmark method with a set of 8 items from the spring test forms. The 8 items had been selected from the OIB and represented items across the full range of difficulty. Each set of 8 items also contained at least one constructed response (CR) multi-point item.

For the practice round, panelists were asked to determine a Bookmark rating for the threshold level between *Level 2 and Level 3*. As panelists completed their practice ratings, the facilitator checked in with panelists and ensured that panelists were comfortable with the process and understood how to record their ratings. After the practice ratings were completed, the panelists discussed how they completed their ratings and reviewed any questions or concerns they had.

After completing the practice round, panelists were asked to complete a readiness survey that asked the panelists if they felt comfortable with the process and were prepared to complete their operational ratings. The facilitators collected and reviewed the survey results to ensure that all panelists were prepared to begin the rating process. To begin their ratings, panelists were presented with the complete OIB and began their Round 1 ratings. An example of the rating forms that were used by panelists for one grade are provided in Appendix D. Upon completing their Round 1 ratings, panelists were provided multiple pieces of feedback. Samples of the feedback provided are included in Appendix E. The feedback included:

- The minimum, maximum, and median recommendation received from panelists
- The distribution of recommendations received across all panelists

The feedback was shared with panelists via tables and graphs that were projected on a large screen that allowed panelists to view the results. After presenting this feedback to panelists, the facilitator instructed the panelists to discuss the results with their fellow panelists. Panelists were encouraged to share their rationale for selecting their Bookmark location, and to discuss how they felt the Bookmark items could be linked to the threshold PLDs. Upon concluding these discussions, the panel moved forward to complete a second round of ratings following the same procedure as was done with Round 1.

After all ratings were completed and compiled, the facilitator reviewed the same feedback information that was provided after Round 1. In addition, after the discussion of this feedback was concluded, the facilitator also shared the estimated impact data with the panelists. The impact was based upon all test forms at each grade level and content and provided the percentage of students that would be classified into each of the four performance categories using the panel's current recommendation. Samples of this feedback are included in Appendix E.

After providing the impact data, the moderator facilitated a discussion of the impact data with all panelists. The moderators asked panelists to discuss whether they felt that the impact data was consistent with their expectations for student performance on the assessments. After completing the discussion of the impact data, panelists were instructed to complete a final round (Round 3) of ratings.

After completing these ratings, the results were compiled, and panelists were provided a summary of their final recommendations and the resulting impact. Panelists were not provided any new type of feedback, just the updated results based upon their Round 3 ratings. At that time, panelists completed a survey indicating their comfort level with the overall process and all materials for the first test were collected by the facilitators.

For each of the panels, the panels then moved forward with the standard setting for their second-grade level. Once all panels completed both grade-level ratings, a subset of panelists from each panel participated in a vertical articulation discussion.

Vertical Articulation

When setting cut scores, vertical articulation provides an opportunity for panelists to review and consider the cut scores being recommended across all grade levels, rather than just the grade level(s) their panel was assigned to review. It can provide a critical safeguard into the standard setting process and help prevent unusual or problematic spikes in impact within a single grade level that would be difficult to understand from a policy perspective.

During the workshop, a vertical articulation process was followed within each content area after all grade level recommendations had been completed. The articulation process was completed by a subset of panelists (3 panelists per panel) at the end of Day 3 of the workshop. During the vertical articulation, the panelists first discussed the knowledge and skills necessary to complete each of the assessments. They also discussed the expected challenges and new materials that were introduced each grade year and whether they would expect to observe consistent performance across years and content areas. After this discussion, the group reviewed the recommended cut scores and impact data across all grade levels. An example of the figure used to illustrate the impact data across all grades is included in Appendix F. In addition to a figure that indicated the estimated percentage of students in each of the four performance categories, panelists could recommend new cut points and immediately observe the impact on the percentage of students classified into each performance category. The panelists were able to make recommendations for changes in the Bookmark and observe any changes in the impact data in real-time. If changes to any cut scores were recommended, the changes needed to be reviewed and approved by the members of the panel that were part of that grade specific committee.

Results

Round 3

Tables 5 and 6 show the results of the ELA and Math standard setting activities after all three Rounds were completed within each panel. These results are the results BEFORE the vertical articulation activities were completed. Results for Round 1 and Round 2 are included in Appendix G.

Table 5: ELA recommendations after round 3 completed for all grades

ELA R3 recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	32.5%	21.7%	29.2%	28.6%	23.0%	16.8%
Level 2	Page #	7	6	13	11	9	10
	Theta	-0.500	-0.806	-0.531	-0.565	-0.832	-0.995
	Raw score cut	17	15	20	22	23	22
	% students	36.8%	29.3%	34.6%	30.4%	32.4%	27.4%
Level 3	Page #	27	22	27	25	28	27
	Theta cut point	0.605	0.097	0.333	0.298	0.193	-0.084
	Raw score cut	26	23	28	29	33	32
	% students	20.4%	30.2%	25.6%	27.6%	21.9%	33.6%
Level 4	Page #	38	38	40	37.5	44	41.5
	Theta cut point	1.318	1.06	1.423	1.327	0.821	0.888
	Raw score cut	30	30	34	34	38	39
	% students	10.4%	18.8%	10.5%	13.4%	22.6%	22.2%

Table 6: Math recommendations after round 3 completed for all grades

Math R3 recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	14.8%	21.0%	27.7%	24.5%	16.9%	38.0%
Level 2	Page #	7	6.5	10	5	5	5
	Theta	-1.049	-0.907	-0.648	-0.781	-1.013	-0.433
	Raw score cut	13	16	17	14	16	19
	% students	27.5%	23.6%	21.0%	29.9%	24.0%	18.8%
Level 3	Page #	25	23	22	21	19.5	16
	Theta cut point	-0.267	-0.132	-0.113	0.067	-0.316	0.102
	Raw score cut	23	25	25	27	27	26
	% students	40.4%	39.3%	37.3%	35.9%	35.6%	31.7%
Level 4	Page #	46	49	50	53	47	46
	Theta cut point	1.19	1.109	1.24	1.47	0.806	1.256
	Raw score cut	34	38	39	42	41	40
	% students	17.4%	16.1%	14.0%	9.7%	23.4%	11.6%

Vertical Articulation

During the vertical articulation discussion, both panels reviewed multiple potential changes and developed recommendations after reviewing both the impact data and the content of the items across grade levels. The Mathematics vertical articulation panel recommended no changes to the current cut scores. The ELA panel recommended the following changes to the Round 3 cut scores.

- ELA, Grade 3, the Level 3 cut score was modified from page 27 to page 25
- ELA, Grade 5, the Level 2 cut score was modified from page 13 to page 11
- ELA, Grade 5, the Level 3 cut score was modified from page 27 to page 25
- ELA, Grade 5, the Level 4 cut score was modified from page 40 to page 38
- ELA, Grade 7, the Level 3 cut score was modified from page 28 to page 25
- ELA, Grade 7, the Level 4 cut score was modified from page 44 to page 40

Final recommendations

Tables 7 and 8 show the final recommendations, with the results reflecting all changes recommended during vertical articulation.

Table 7: Final ELA cut score recommendations for all grades

Final ELA recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	32.5%	21.7%	25.7%	28.6%	23.0%	16.8%
Level 2	Page #	7	6	11	11	9	10
	Theta	-0.500	-0.806	-0.665	-0.565	-0.832	-0.995
	Raw score cut	17	15	19	22	23	22
	% students	31.7%	29.3%	33.7%	30.4%	28.4%	27.4%
Level 3	Page #	25	22	25	25	25	27
	Theta cut point	0.444	0.097	0.258	0.298	0.066	-0.084
	Raw score cut	25	23	27	29	32	32
	% students	25.4%	30.2%	26.5%	27.6%	21.6%	33.6%
Level 4	Page #	38	38	38	37.5	40	41.5
	Theta cut point	1.318	1.06	1.299	1.327	0.651	0.888
	Raw score cut	30	30	33	34	37	39
	% students	10.4%	18.8%	14.2%	13.4%	27.0%	22.2%

Table 8: Final Math recommendations for all grades

Math Final recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	14.8%	21.0%	27.7%	24.5%	16.9%	38.0%
Level 2	Page #	7	6.5	10	5	5	5
	Theta	-1.049	-0.907	-0.648	-0.781	-1.013	-0.433
	Raw score cut	13	16	17	14	16	19
	% students	27.5%	23.6%	21.0%	29.9%	24.0%	18.8%
Level 3	Page #	25	23	22	21	19.5	16
	Theta cut point	-0.267	-0.132	-0.113	0.067	-0.316	0.102
	Raw score cut	23	25	25	27	27	26
	% students	40.4%	39.3%	37.3%	35.9%	35.6%	31.7%
Level 4	Page #	46	49	50	53	47	46
	Theta cut point	1.19	1.109	1.24	1.47	0.806	1.256
	Raw score cut	34	38	39	42	41	40
	% students	17.4%	16.1%	14.0%	9.7%	23.4%	11.6%

Conditional standard error of measurement

To provide error bands around the recommended cut scores, the conditional standard error of measurement (CSEM) was calculated at each cut point recommendation. Using the CSEM values, the recommended theta cut point was increased by 1 or 2 CSEM to create error bands around the recommended cut scores. The theta values were used to determine the estimated percentage of students in each category. The estimated percentages and scale scores for each theta value are included in Tables 9 and 10 below.

Table 9. Recommended ELA Cut Scores with CSEM Estimates Around Cut Score Recommendations

		Recommended			Impact			
	Grade	Theta Cut 1/2	Theta Cut 2/3	Theta Cut 3/4	Level 1	Level 2	Level 3	Level 4
Plus 2 CSEM	3	0.050	1.097	2.214	50.8%	34.1%	12.4%	2.8%
	4	-0.196	0.709	1.800	43.1%	29.5%	22.9%	4.5%
	5	-0.057	0.919	2.122	45.7%	31.8%	19.8%	2.7%
	6	0.009	0.986	2.289	48.8%	32.7%	15.9%	2.6%
	7	-0.319	0.665	1.393	37.1%	35.9%	16.3%	10.7%
	8	-0.482	0.504	1.690	30.6%	36.8%	27.4%	5.2%
		Recommended			Impact			
	Grade	Theta Cut 1/2	Theta Cut 2/3	Theta Cut 3/4	Level 1	Level 2	Level 3	Level 4
Plus 1C SEM	3	-0.225	0.771	1.766	43.0%	31.4%	19.5%	6.1%
	4	-0.501	0.403	1.430	31.7%	32.1%	25.3%	10.8%
	5	-0.361	0.588	1.710	37.2%	31.3%	24.2%	7.3%
	6	-0.278	0.642	1.808	39.8%	30.3%	24.7%	5.2%
	7	-0.575	0.365	1.022	31.1%	33.0%	17.6%	18.4%
	8	-0.739	0.210	1.289	22.9%	34.6%	30.0%	12.5%

		Recommended			Impact			
	Grade	Theta Cut 1/2	Theta Cut 2/3	Theta Cut 3/4	Level 1	Level 2	Level 3	Level 4
Current	3	-0.500	0.444	1.318	32.5%	31.7%	25.4%	10.4%
	4	-0.806	0.097	1.060	21.7%	29.3%	30.2%	18.8%
	5	-0.665	0.258	1.299	25.7%	33.7%	26.5%	14.2%
	6	-0.565	0.298	1.327	28.6%	30.4%	27.6%	13.4%
	7	-0.832	0.066	0.651	23.0%	28.4%	21.6%	27.0%
	8	-0.995	-0.084	0.888	16.8%	27.4%	33.6%	22.2%

		Recommended			Impact			
	Grade	Theta Cut 1/2	Theta Cut 2/3	Theta Cut 3/4	Level 1	Level 2	Level 3	Level 4
Minus 1 CSEM	3	-0.775	0.117	0.870	26.0%	29.0%	24.6%	20.4%
	4	-1.112	-0.209	0.690	15.8%	27.3%	29.5%	27.4%
	5	-0.970	-0.073	0.887	16.0%	29.6%	31.8%	22.5%
	6	-0.852	-0.046	0.847	19.6%	29.2%	27.0%	24.2%
	7	-1.089	-0.234	0.281	16.2%	24.2%	19.3%	40.3%
	8	-1.252	-0.377	0.486	11.9%	21.7%	33.8%	32.5%

		Recommended			Impact			
	Grade	Theta Cut 1/2	Theta Cut 2/3	Theta Cut 3/4	Level 1	Level 2	Level 3	Level 4
Minus 2C SEM	3	-1.050	-0.209	0.422	17.0%	26.0%	21.2%	35.8%
	4	-1.417	-0.516	0.320	8.3%	23.5%	27.8%	40.5%
	5	-1.274	-0.403	0.475	10.9%	22.2%	35.4%	31.5%
	6	-1.139	-0.389	0.366	14.7%	21.1%	28.6%	35.6%
	7	-1.345	-0.533	-0.090	10.4%	20.7%	16.4%	52.5%
	8	-1.509	-0.671	0.085	7.8%	17.5%	27.5%	47.2%

Table10. Recommended Math Cut Scores with CSEM Estimates Around Cut Score Recommendations

		Recommended			Impact			
	Grade	Theta Cut 1/2	Theta Cut 2/3	Theta Cut 3/4	Level 1	Level 2	Level 3	Level 4
Plus 2 CSEM	3	-0.566	0.213	1.973	30.7%	27.9%	36.6%	4.8%
	4	-0.425	0.331	1.789	36.5%	25.3%	31.3%	6.9%
	5	-0.216	0.325	1.897	43.5%	18.2%	32.6%	5.7%
	6	-0.321	0.499	2.191	40.9%	27.6%	27.9%	3.6%
	7	-0.594	0.082	1.368	29.9%	25.2%	33.2%	11.6%
	8	0.061	0.570	1.884	54.2%	17.4%	23.5%	4.9%

		Recommended			Impact			
	Grade	Theta Cut 1/2	Theta Cut 2/3	Theta Cut 3/4	Level 1	Level 2	Level 3	Level 4
Plus 1 CSEM	3	-0.808	-0.027	1.581	22.5%	29.3%	39.5%	8.7%
	4	-0.666	0.100	1.449	26.0%	26.9%	37.3%	9.8%
	5	-0.432	0.106	1.569	35.5%	21.0%	35.1%	8.3%
	6	-0.551	0.283	1.830	34.1%	27.1%	33.3%	5.5%
	7	-0.803	-0.117	1.087	23.4%	24.6%	34.5%	17.6%
	8	-0.186	0.336	1.570	46.3%	18.0%	27.7%	8.0%

		Recommended			Impact			
	Grade	Theta Cut 1/2	Theta Cut 2/3	Theta Cut 3/4	Level 1	Level 2	Level 3	Level 4
Current	3	-1.049	-0.267	1.190	14.8%	27.5%	40.3%	17.4%
	4	-0.907	-0.132	1.109	21.0%	23.6%	39.3%	16.1%
	5	-0.648	-0.113	1.240	27.7%	21.0%	37.3%	14.0%
	6	-0.781	0.067	1.470	24.5%	29.9%	35.9%	9.7%
	7	-1.013	-0.316	0.806	16.9%	24.1%	35.6%	23.4%
	8	-0.433	0.102	1.256	38.0%	18.8%	31.7%	11.6%

		Recommended			Impact			
	Grade	Theta Cut 1/2	Theta Cut 2/3	Theta Cut 3/4	Level 1	Level 2	Level 3	Level 4
Minus 1 CSEM	3	-1.290	-0.507	0.798	10.0%	23.5%	40.5%	26.0%
	4	-1.149	-0.364	0.770	13.8%	22.8%	37.8%	25.7%
	5	-0.864	-0.332	0.912	20.0%	20.9%	39.4%	19.7%
	6	-1.012	-0.149	1.109	16.1%	31.5%	35.6%	16.8%
	7	-1.222	-0.515	0.525	10.4%	23.9%	36.5%	29.1%
	8	-0.680	-0.132	0.942	29.0%	19.9%	33.6%	17.4%

		Recommended			Impact			
	Grade	Theta Cut 1/2	Theta Cut 2/3	Theta Cut 3/4	Level 1	Level 2	Level 3	Level 4
Minus 2 CSEM	3	-1.532	-0.747	0.406	5.8%	19.4%	40.9%	34.0%
	4	-1.390	-0.596	0.430	7.1%	21.6%	36.2%	35.1%
	5	-1.080	-0.550	0.583	14.9%	18.0%	36.6%	30.5%
	6	-1.242	-0.364	0.748	10.0%	28.6%	37.3%	24.1%
	7	-1.432	-0.714	0.244	6.5%	21.3%	32.5%	39.8%
	8	-0.928	-0.366	0.628	19.5%	21.3%	33.0%	26.2%

Workshop Evaluation

During the workshop, panelists completed readiness surveys before each round of ratings and three evaluation surveys that focused on the purpose of the workshop, the preparedness of the panelists to complete their ratings, and their overall comfort with the cut score recommendations. Brief descriptions of the three evaluation surveys collected over the course of the workshop are described below:

- Evaluation #1 – Administered after the initial training for the first test to be reviewed and prior to panelists completing any ratings. This was collected for the first grade level only.
- Evaluation #2 – Administered after round 3 ratings were completed for each grade level and the overall recommendations had been reviewed. This survey was completed for all grade levels.
- Evaluation #3 – Administered after the vertical articulation was completed; administered only to panelists who participated in the vertical articulation.

Copies of all evaluation questions and a summary of the responses for each grade level can be found in Appendix H. On the readiness survey, all panelists indicated that they were prepared to begin the rating process. A brief review of the evaluation survey responses across all panels is provided below while also highlighting any grade-level results that may be of interest.

Evaluation #1

Evaluation #1 was administered at the conclusion of all training activities but before panelists completed operational ratings for test items. The expectation was that panelists would be somewhat comfortable with the training process and would feel like they had a sufficient amount of time to complete the activities. As can be seen in Appendix H, across all grade levels, panelists generally appeared to be comfortable with the process and were ready to begin ratings at that time. Almost all of the panelists either Agreed or Strongly Agreed that they were prepared to complete the Bookmark process prior to completing their first round of ratings. While some panelists expressed some concerns with the rating process, there was sufficient time between the beginning of Round 1 ratings and completing round 3 ratings to allow panelists to become more comfortable with the process. It should be noted that due to some miscommunication between the lead and grade 5 ELA facilitator, data was not collected for the grade 5 ELA group.

Table 11: Summary of evaluation #1 question across grade and content area

Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	N	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
8. After the practice exercise, I understood how to use the bookmark method.						
ELA Grade 3 - 4 (Grade 3)	11	--	--	18%	45%	36%
ELA Grade 5 - 6 (Grade 5, not collected)	NA					
ELA Grade 7 - 8 (Grade 7)	10	20%	--	10%	40%	30%
Math Grade 3- 4 (Grade 3)	12	--	--	8%	58%	33%
Math Grade 5 - 6 (Grade 5)	10	--	--	--	60%	40%
Math Grade 7 - 8 (Grade 7)	10	--	--	--	--	100%

Evaluation #2

Evaluation #2 was administered after completing all rounds for all grade levels and focused on the overall comfort with the process and the recommended cut scores for the grade. The evaluation includes questions about the panelists overall comfort level with the current cut score recommendation. Tables 12 and 13 provide a summary of the results of these questions for ELA and Math respectively. It should be noted here that due to some miscommunication between the lead facilitator and the ELA Grade 4 facilitator, the data for evaluation #2 was not collected for ELA Grade 4.

Table 12: Summary of panelists comfort level for ELA cut score recommendations

	N	Too Low	About Right	Too High
ELA Grade 3				
Approaching Proficient	11	--	100%	--
Proficient	11	9%	82%	9%
Advanced	11	--	100%	--
ELA Grade 4 (not collected)				
Approaching Proficient				
Proficient				
Advanced				
ELA Grade 5				
Approaching Proficient	12	--	83%	17%
Proficient	12	8%	67%	25%
Advanced	12	8%	50%	42%
ELA Grade 6				
Approaching Proficient	9	11%	67%	22%
Proficient	9	22%	44%	33%
Advanced	9	--	100%	--
ELA Grade 7				
Approaching Proficient	10	--	90%	10%
Proficient	10	10%	70%	20%
Advanced	10	20%	70%	10%
ELA Grade 8				
Approaching Proficient	9	--	100%	--
Proficient	9	11%	89%	--
Advanced	9	--	100%	--

Table 13: Summary of panelists comfort level for Math cut score recommendations

	N	Too Low	About Right	Too High
Math Grade 3				
Approaching Proficient	12	--	92%	8%
Proficient	12	--	92%	8%
Advanced	12	--	92%	8%
Math Grade 4				
Approaching Proficient	12	--	92%	8%
Proficient	12	8%	83%	8%
Advanced	12	--	92%	8%
Math Grade 5				
Approaching Proficient	11	9%	82%	9%
Proficient	11	--	73%	27%
Advanced	11	18%	82%	--
Math Grade 6				
Approaching Proficient	9	--	100%	--
Proficient	9	--	89%	11%
Advanced	9	--	89%	11%
Math Grade 7				
Approaching Proficient	10	10%	90%	--
Proficient	10	10%	90%	--
Advanced	10	--	100%	--
Math Grade 8				
Approaching Proficient	10	--	90%	10%
Proficient	10	--	100%	--
Advanced	10	--	100%	--

While the panels generally seemed comfortable with the ratings, there were a few areas where disagreements are noteworthy. For example, 42% of the ELA Grade 5 panelists felt that the Level 4 cut score was set *Too High*. Also in ELA Grade 5, 25% of panelists felt that the Level 3 was set *Too High*. It is noteworthy that the Grade 5 ELA is one grade that appears to have some discontent from the panelists, and this grade was the one grade that had all cut score recommendations shifted during the vertical articulation process. The ELA Grade 6 panelists also indicated some discomfort with the cut score recommendations. But in this grade, there does not appear to be a clear direction, with some panelists indicating they felt the cut scores were Too Low and others indicating Too High. A similar pattern is observed with ELA Grade 7, with some discomfort, but not a clear direction for the discomfort as either *Too Low* or *Too High*.

In Math, the panels seemed notably more comfortable with the ratings, with the panels almost always having 90% or more indicating that they felt the cut score recommendations were *About Right*.

Evaluation #3

Evaluation #3 was completed by the 16 of the 18 panelists who participated in the vertical articulation process. It focused on their overall comfort with the process and the final cut score recommendations. As can be seen in Tables 14 and 15, it appears that the vertical articulation participants were comfortable and agreed with the final cut score recommendations. Almost all panelists either agreed or strongly agreed that the vertical articulation orientation provided a reasonable review and that they were comfortable with the cut score recommendations determined during the vertical articulation process. One panelist (1 out 7 equals 14%) indicated that they did not feel like the final impact numbers reflected their expectations for students, but 15 of the 16 indicated they agreed or strongly agreed with the statement.

Table 14: Summary of ELA vertical articulation evaluation results

Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the four options for each statement.	Strongly Disagree	Disagree	Agree	Strongly Agree
1. The orientation to the vertical articulation process was comprehensive and allowed me to understand the purpose of the vertical articulation procedure.	--	--	43%	57%
2. I am comfortable with the final cut score recommendations that were completed as part of this workshop.	--	14%	71%	14%
3. I believe the final impact data reflects my expectations as far as the percent of students within each performance category.	--	14%	57%	29%

Table 15: Summary of Math vertical articulation evaluation results

Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the four options for each statement.	Strongly Disagree	Disagree	Agree	Strongly Agree
1. The orientation to the vertical articulation process was comprehensive and allowed me to understand the purpose of the vertical articulation procedure.	--	--	22%	78%
2. I am comfortable with the final cut score recommendations that were completed as part of this workshop.	--	--	11%	89%
3. I believe the final impact data reflects my expectations as far as the percent of students within each performance category.	--	--	22%	78%

Review of workshop

Throughout the duration of the project, and as the results were finalized, Kane's (2001) framework for validating standard setting activities was applied. Kane (1994) suggested three sources of evidence should be considered in the validation process: procedural, internal, and external. Evidence within each of these areas that was observed in this study is discussed here.

Procedural evidence can be viewed through the surveys and feedback provided throughout the standard setting workshop. During the workshop, in addition to the four evaluation surveys, all panelists completed readiness surveys prior to completing ratings, and all panelists indicated that they understood the task and were prepared to complete their ratings. In addition to the readiness survey, panelists completed surveys after completing their threshold PLDs and again after completing their final recommendations. As was reviewed above and is reported in Appendix H, the panelists consistently indicated they felt prepared to complete their ratings and overall were comfortable with the recommended cut scores.

The primary source of internal validity evidence can be observed when looking at the variability of the cut point recommendations. The standard error of the median was calculated for every round of ratings, for each of the cut point recommendations. Table 16 below provides the mean standard error value across each of the assessments as well as a summary within each content area and across all assessments. The variability of the standard error did decline as panelists moved from the first to the second round. However, when moving from the 2nd round to the 3rd round, the variability remains largely unchanged. A complete summary of all ratings for all three rounds is included in Appendix I.

Table 16: Standard error of median across rounds by grade and content area

	Round 1			Round 2			Round 3		
	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4	Level 2	Level 3	Level 4
ELA G3	1.60	1.79	1.40	0.16	0.46	0.62	0.28	0.89	0.55
ELA G4	1.15	0.78	0.87	0.28	0.35	0.60	0.36	0.51	0.62
ELA G5	1.80	2.13	1.72	0.37	0.56	1.06	0.36	0.48	0.94
ELA G6	0.96	0.95	1.20	0.33	0.30	0.38	0.46	0.41	0.38
ELA G7	1.60	2.49	1.54	0.50	1.20	1.27	0.53	2.09	1.93
ELA G8	0.99	1.78	1.32	0.63	1.17	0.97	0.57	1.10	0.99
Math G3	1.18	1.31	1.13	0.29	0.26	0.56	0.29	0.28	0.51
Math G4	1.03	1.82	1.62	1.25	1.41	0.58	1.14	0.78	1.14
Math G5	1.51	2.87	1.39	0.84	1.29	0.67	0.68	1.24	0.72
Math G6	0.66	1.65	1.16	0.31	0.38	0.48	0.21	1.40	0.43
Math G7	1.02	2.63	1.90	0.19	1.86	0.80	0.31	1.54	0.30
Math G8	0.95	2.71	1.03	0.17	0.54	0.00	0.16	0.10	0.00
ELA	1.20	1.91	1.36	0.44	0.82	0.67	0.45	0.90	0.71
Math	1.06	2.17	1.37	0.51	0.96	0.52	0.47	0.89	0.52
All	1.20	1.91	1.36	0.44	0.82	0.67	0.45	0.90	0.71

Conclusion

At the conclusion of the workshop, the cut score recommendations were provided to the NYSED for review. Based upon the evidence collected and the review of the performance of panelists, it appears that the cut point recommendations provide appropriate cut scores for the NYSED ELA and Math assessments.

Appendix A: Summary of panelist demographic information

Table A1: Current role for standard setting panelists

What is your current position?						
Content	Grade	Teacher	Teacher (Higher Education)	Administrator (School)	Administrator (District)	Other
ELA	Grade 3/4	7	--	1	1	--
ELA	Grade 5/6	10	--	--	--	--
ELA	Grade 7/8	9	--	--	--	1
Math	Grade 3/4	8	--	2	--	1
Math	Grade 5/6	7	--	2	--	--
Math	Grade 7/8	8	--	1	--	--
Total		49	0	6	1	2

Table A2: Gender identified for standard setting panelists

What is your gender?				
Content	**Grade**	**Male**	**Female**	**No answer**
ELA	Grade 3/4	--	8	1
ELA	Grade 5/6	2	8	--
ELA	Grade 7/8	1	8	1
Math	Grade 3/4	--	11	--
Math	Grade 5/6	1	8	--
Math	Grade 7/8	1	8	--
Total		5	51	2

Table A3: Race/ethnicity for standard setting panelists

What is your race?								
Content	Grade	American Indian or Alaska Native	Asian	Black or African American	Hispanic or Latino	Native Hawaiian or Pacific Islander	White	No answer
ELA	Grade 3/4	--	--	2	1	--	6	--
ELA	Grade 5/6	--	--	2	1	--	7	--
ELA	Grade 7/8	--	--	2	1	--	5	2
Math	Grade 3/4	--	1	1	--	--	9	--
Math	Grade 5/6	--	--	1	2	--	6	--
Math	Grade 7/8	--	1	--	--	8	--	--
Total		0	2	8	5	8	33	2

Table A4: New York State Region for standard setting panelists

Which region is your district in?									
		Capital District	Central	Long Island	Mid-Hudson	New York City	North Country / Adirondacks	Southern Tier	Western
ELA	Grade 3/4	--	--	2	1	2	--	2	2
ELA	Grade 5/6	1	--	1	3	2	2	--	1
ELA	Grade 7/8	1	--	3	--	3	1	1	1
Math	Grade 3/4	2	3	1	--	1	1	2	1
Math	Grade 5/6	1	2	1	--	2	--	2	1
Math	Grade 7/8	1	1	--	1	2	1	1	2
Total		6	6	8	5	12	5	8	8

Table A5: Experience with students for standard setting panelists

Experience with the following students			
Content	Grade	Students with Disabilities	English Language Learners
ELA	Grade 3/4	1	1
ELA	Grade 5/6	2	1
ELA	Grade 7/8	3	1
Math	Grade 3/4	1	1
Math	Grade 5/6	1	–
Math	Grade 7/8	3	1
Total		11	5

Appendix B: Agenda for standard setting workshop

nwea

New York State Grades 3-8
Standards Setting
August 1 – 3, 2023

Tuesday August 1st, 2023

8:30 - 9:00	Welcome & Orientation Introduction to the test Introduction to standard setting task
	Split into Breakout rooms
9:15 - 10:00	Take the first test (or a sample of items)
10:15 - 12:00	Develop threshold Performance Level Descriptions (PLDs)
12:00 - 12:45	Lunch
12:45 - 2:00	Develop threshold PLDs
2:00 - 2:15	Break
2:15 - 3:00	Practice ratings
3:00 - 4:30	Provide Round 1 ratings for test #1

Wednesday August 2, 2023

8:30 - 9:30	Review Round 1 ratings
9:30 - 10:15	Complete Round 2 ratings for test #1
10:15 - 11:00	Review Round 2 ratings
11:00 - 11:30	Complete Round 3 ratings
11:45 - 12:00	Review round 3 results
12:00 - 12:45	Lunch
12:45 - 1:30	Take the second test
1:30 - 3:15	Develop threshold PLDs
3:15 - 4:30	Provide Round 1 ratings for test #2

Thursday August 3, 2023

8:30 - 9:30	Review Round 1 results
9:30 - 10:15	Provide Round 2 results for test #2
10:30 - 11:30	Review Round 2 ratings
11:30 - 12:15	Lunch
12:15 - 12:45	Provide Round 3 ratings for test #2
1:00 - 1:15	Review Round 3 ratings
2:00 - 4:00	Vertical Articulation

Appendix C: Threshold PLDs for each grade and content area

ELA Grade 3 (Threshold Levels swapped to match the PLD document)

Grade Level Standard	Threshold Level 4	Threshold Level 3	Threshold Level 2
3R1: Develop and answer questions to locate relevant and specific details in a text to support an answer or inference. (RI&RL)	Clear and concise inferencing and use of details in reference to a text that connects to the claim.	Use of details in reference to a text that connects to the claim.	Use of detail / details in reference to the text.
3R2: Determine a theme or central idea and explain how it is supported by key details; summarize portions of a text (RI&RL)	State an accurate theme or central idea and describe how the relevant details support that theme.	State an accurate theme or central idea with relevant details for support.	State a theme or central idea with some details for support.
3R3: In literary texts, describe character traits, motivations, or feelings, drawing on specific details from the text. (RL)	Describe and analyze a character using specific and relevant details.	Describe a character accurately with at least one relevant detail.	Describe a character accurately.
In informational texts, describe the relationship among a series of events, ideas, concepts, or steps in a text, using language that pertains to time, sequence, and cause/effect. (RI)	Accurately describe and analyze the relationship between events, ideas or concepts.	Accurately describe the relationship between events, ideas or concepts.	Describe the relationship between events, ideas or concepts.
3R4: Determine the meaning of words, phrases, figurative language, and academic and content-specific words. (RI&RL)	Determines accurate meaning of words and phrases as used in the text with detail, particularly academic language, content specific words and figurative language.	Determines accurate meaning of words and phrases as used in the text, particularly academic language and content specific words.	Determines partial meaning of words and phrases, particularly academic language.

Grade Level Standard	Threshold Level 4	Threshold Level 3	Threshold Level 2
3R5: In literary texts, identify parts of stories, dramas, and poems using terms such as chapter, scene, and stanza. (RL) In informational texts, identify and use text features to build comprehension. (RI)	Accurately identify parts of stories, dramas and poems with correct use of terms; partially explains how they relate to the overall structure. Identifies text features accurately. Demonstrates inferential understanding between the text features and the text.	Accurately identify part of a story with the correct use of terms. Identifies some parts of dramas and poems with accuracy. Identifies text features accurately. Demonstrates the connection between the text features and the text.	Identify parts of a story. Identify and partially understand the connection between the text features and the text.
3R6: Discuss how the reader's point of view or perspective may differ from that of the author, narrator, or characters in a text. (RI&RL)	Explain one's own point of view and state the author's, narrator's etc. and explain the similarities and differences between the two with clarity and detail.	Explain one's own point of view and state the author's, narrator's etc. and explain the similarities or differences between the two.	Explain one's own point of view and state the author's, narrator's etc.
3R7: Explain how specific illustrations or text features contribute to what is conveyed by the words in a text (e.g., create mood, emphasize character or setting, or determine where, when, why, and how key events occur). (RI&RL)	Provides a detailed and accurate explanation of the illustration / text feature, and accurately describes how it relates to the text.	Provides a description of the illustration / text feature, and accurately describes how it relates to the text.	Provides a description of the illustration / text feature, with an attempt to relate it to the text.

Grade Level Standard	Threshold Level 4	Threshold Level 3	Threshold Level 2
3R8: Explain how claims in a text are supported by relevant reasons and evidence. (RI&RL)	Provides an explanation with some relevant details, reasons and evidence	Provides a simple explanation with minimal details, relevant reasons and evidence	Provides an insufficient explanation with limited reasons and evidence
Language Standards 1 and 2 are organized within grade bands. For the Core Conventions Skills and Core Punctuation and Spelling Skills for Grades 3–5, the student is expected to know and be able to use the skills by the end of 5th grade.	While taking some risks, there are few errors but do not hinder comprehension.	Frequent errors that do not hinder comprehension	Excessive errors that might hinder comprehension
3L3: Recognize differences between the conventions of spoken conversational English and academic English; signal this awareness by selecting conversational or academic forms when speaking or writing.	NA	NA	NA
3L4: Determine or clarify the meaning of unknown and multiple- meaning words and phrases, choosing flexibly from a range of strategies.	Can determine or clarify the meanings of unknown and multiple meaning words with detail using a range of strategies	Can determine or clarify the meanings of unknown and multiple meaning words with a range of strategies	May determine or clarify the meanings of unknown and multiple meaning words with a limited range of strategies
3L5: Demonstrate understanding of word relationships and nuances in word meanings.	NA	NA	NA

Grade Level Standard	Threshold Level 4	Threshold Level 3	Threshold Level 2
3L6: Acquire and accurately use conversational, general academic, and content- specific words and phrases, including those that signal spatial and temporal relationships (e.g., After dinner that night we went out for dessert).	NA	NA	NA
3W1: Write an argument to support claim(s), using clear reasons and relevant evidence.	Produce **mostly logical** arguments with **some insight** to support claim(s), incorporating content-specific vocabulary. Using **mostly relevant reasons** and evidence organized with linking words and phrases. It has to include a **concluding thought/idea.**	Produce **some logical** argument(s) to support claim(s). It may contain minor inaccuracies. Incorporates **some content-specific vocabulary**. Has sufficiently used **some relevant reasons and evidence**. Organized with **some linking words and phrases.**	Produce an argument **with a claim** that may be inaccurate. **Attempts to use** relevant reasons and evidence. **May contain linking words and phrases.**
3W2: Write informative/ explanatory texts to explore a topic and convey ideas and information relevant to the subject.	Produce **mostly insightful** informative/explanatory texts to explore topics and convey ideas and information. **Effectively incorporates** content-specific vocabulary and relevant reasons and evidence. **Effectively organized** with linking words and phrases. Provides a conclusion.	Produce informative/explanatory texts that **attempt to explore topics** using **some ideas** and information. Provides **some content-specific vocabulary** and relevant reasons and evidence. Organized with **some linking words and phrases**. Provides a conclusion.	Produce informative/explanatory texts that address a topic and may contain **some inaccuracies of ideas and information. Attempts to use** relevant reasons and evidence. **May contain linking words and phrases. May contain a concluding thought/idea.**
3W3: Write narratives to develop real or imagined experiences or events using effective techniques, descriptive details, and clear event sequences.	Write narratives to develop real or imagined experiences **that attempt to engage the reader.** Uses effective technique and descriptive details. **Attempts to clearly** establish a situation and narrator. Uses description to develop narrators and characters.	Write narratives to develop real or imagine experiences **with use of details** to describe and establish the situation and narrator. **Attempts to develop** narrators and characters. **Uses temporal words** to signal event order. **Provides a conclusion.**	Write narratives to develop real or imagine experiences **with some use of details** to describe and establish the situation and narrator. **May include the use of** temporal words to signal event order.

Grade Level Standard	Threshold Level 4	Threshold Level 3	Threshold Level 2
	Also uses temporal words to signal event order. Provides a conclusion.		
3W4: Create a response to a text, author, theme, or personal experience (e.g., poem, play, story, artwork, or other).	Create a response **that makes connections** to a text, author, theme, or person experience.	Create a response to a text, author, theme, or personal experience that **mostly addresses the task.**	Create a response that is **mostly coherent and somewhat relates** to a text, author, theme or a personal experience. **Attempts to address the task.**

ELA Grade 4

Grade Level Standard	Threshold Level 4	Threshold Level 3	Threshold Level 2
4R1: Locate and refer to relevant details and evidence when explaining what a text says explicitly/implicitly and make logical inferences. (RI&RL)	Generate inferences that are supported by sufficient, relevant details and evidence.	Generate inferences that are supported by sufficient details and evidence.	Generate inferences that may include insufficient or irrelevant details.
4R2: Determine a theme or central idea of a text and explain how it is supported by key details; summarize a text. (RI&RL)	Determine a theme or central idea by clearly explaining how it is supported with a detailed summary of the text.	Determine a theme or central idea by explaining how it is supported with a basic summary of the text.	Determine a theme or central idea by minimally explaining how it is supported with a limited summary of the text with some inaccuracies.
4R3: In literary texts, describe a character, setting, or event, drawing on specific details in the text. (RL) In informational texts, explain events, procedures, ideas, or concepts, including what happened and why, based on specific evidence from the text. (RI)	In literary texts, describe a character, setting or event drawing on key relevant details. In Informational texts, explain events, procedures, ideas or concepts including connections between what happened and why, using accurate details from the text.	In literary texts, describe a character, setting or event drawing on sufficient details. In Informational texts, explain events, procedures, ideas or concepts including connections between what happened and why, using details from the text.	In literary texts, describe a character, setting or event drawing on minimal details, with some inaccuracies. In Informational texts, explain events, procedures, ideas or concepts including connections between what happened and why, using minimal or limited details from the text with some inaccuracies
4R4: Determine the meaning of words, phrases, figurative language, academic and content-specific words. (RI&RL)	Determine with clarity the meaning of words, phrases, figurative language, academic and content-specific words or phrases in a text.	Determine the meaning of words, phrases, figurative language, academic and content-specific words or phrases in a text.	Determine the literal meaning of words, phrases, figurative language, academic and content-specific words or phrases in a text.
Grade Level Standard	**Threshold Level 4**	**Threshold Level 3**	**Threshold Level 2**

4R5: In literary texts, identify and analyze structural elements, using terms such as verse, rhythm, meter, characters, settings, dialogue, and stage directions. (RL) In informational texts, identify the overall structure using terms such as sequence, comparison, cause/effect, and problem/solution. (RI)	In literary texts, identify and analyze structural elements, and provide a detailed explanation as to how they relate to each other and the entire text using specific terminology. In informational texts, identify and analyze structures to provide clear descriptions using terms such as sequence, comparison, cause/effect and problem/solution.	In literary texts, identify and analyze structural elements, and provide an explanation as to how they relate to each other and the entire text. In informational texts, identify and analyze structures using basic terms.	In literary texts, identify basic structural elements with little to no analysis. In informational texts, identify and analyze structures using basic terms with minimal analysis.
4R6: In literary texts, compare and contrast the point of view from which different stories are narrated, including the difference between first- and third-person narrations. (RL) In informational texts, compare and contrast a primary and secondary source on the same event or topic. (RI)	L- Compares and contrasts the point of view, may include analysis between the first- and third-person narrations in different stories I- Compares and contrasts a primary and secondary source on the same event or topic, attempts to explain or analyze how it shapes the content/style of the text.	L- Compares and contrasts the point of view, may include the difference between first- and third-person narrations. I- Compares and contrasts a primary and secondary source.	L- Minimally compares and contrasts first- and third- person point of view, with some inaccuracies. I- Minimally compares and contrasts primary and secondary sources with some inaccuracies.
4R7: Identify information presented visually, orally, or quantitatively (e.g., in charts, graphs, diagrams, time lines, animations, or illustrations) and explain how the information contributes to an understanding of the text. (RI&RL)	Identify and **explain with some insights** how information presented visually, orally or quantitatively contributes to an understanding of a literary or informational text.	Identify and **minimally explain** how information presented visually, orally or quantitatively contributes to an understanding of a literary or informational text.	Identify and **attempt to explain** how information presented visually, orally or quantitatively contributes to an understanding of a literary or informational text.
Grade Level Standard	**Threshold Level 4**	**Threshold Level 3**	**Threshold Level 2**
4R8: Explain how claims in a text are supported by relevant reasons and evidence. (RI&RL)	Identify claims in a text. **Provides a detailed explanation** of how claims in a text are supported by relevant reasons and evidence.	Identify **a claim(s)** in a text **and explain** how they are supported by relevant reasons and evidence.	**Identify a claim in a text with little to no explanation** of how they are supported by relevant reason and evidence.

Page **35** of **168**

4R9: Recognize genres and make connections to other texts, ideas, cultural perspectives, eras, personal events, and situations.	Demonstrate **clear understanding** of how the introduced concept provides **insight** into the original text. **May or may not** provide **some thoughtful commentary** on the connections that are made.	Demonstrate **understanding** of how the introduced concept **relates** to information presented in the original text.	Demonstrate the ability to **identify** or make an **accurate** connection between the introduced concept and the original text.
Language Standards 1 and 2 are organized within grade bands. For the Core Conventions Skills and Core Punctuation and Spelling Skills for Grades 3–5, the student is expected to know and be able to use the skills by the end of Grade 5.	Write and/or speak with few errors in standard English grammar and usage with **clear comprehension**. Produce writing that **demonstrates comprehension and contains few or no errors.**	Write and/or speak with **some errors** in standard English grammar and usage that **do not hinder comprehension.** Produce writing with **some errors** and **does not hinder comprehension.**	Write and/or speak **with errors** in standard English capitalization, punctuation, and spelling **with limited comprehension.** Produce writing with **errors** and **limited comprehension**.
4L3: Use knowledge of language and its conventions when writing, speaking, reading, or listening.	Communicate using **some sophisticated and/or precise** language and conventions appropriate to the situation or task.	Communicate using **adequate** language appropriate to the situation or task.	Communicate using **accurate** language for the situation or task.
4L4: Determine or clarify the meaning of unknown and multiple-meaning words and phrases, choosing flexibly from a range of strategies.	Can determine or clarify the meanings of unknown and multiple meaning words with detail using a range of strategies.	Can determine or clarify the meanings of unknown and multiple meaning words with a limited range of strategies.	May determine or clarify the meanings of unknown and multiple meaning words with a limited range of strategies.
Grade Level Standard	**Threshold Level 4**	**Threshold Level 3**	**Threshold Level 2**

4W1: Write an argument to support claims, using clear reasons and relevant evidence.	Writes a well developed argument to support the claim with logical reasons and organized facts and details. Incorporates content-specific vocabulary, transitional words/phrases and a conclusion statement/section.	Writes a logical argument to support the claim. Has sufficiently used relevant reasons and evidence. Incorporates some organization, content-specific vocabulary, transitional words/phrases and a conclusion statement/section.	Write an argument with a claim that may be inaccurate. Attempts to use relevant reasons, evidence and organized facts/details. May contain transitional words/phrases, content-specific vocabulary and/or a conclusion.
Grade Level Standard	**Threshold Level 4**	**Threshold Level 3**	**Threshold Level 2**
4W4: Create a poem, story, play, artwork, or other response to a text, author, theme, or personal experience.	NA	NA	NA
4W5: Draw evidence from literary or informational texts to respond and support analysis, reflection, and research by applying Grade 4 reading standards.	Writes using evidence drawn from the text to support in-depth analysis, reflection and research.	Writes using some evidence drawn from the text to support analysis, reflection, and research.	Writes using minimal evidence drawn from the text to partially support analysis, reflection, and research.

ELA Grade 5

Grade Level Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
5R1: Locate and refer to relevant details and evidence when explaining what a text says explicitly/implicitly and make logical inferences. (RI&RL)	Makes an accurate inference	Accurate inference with relevant details from text	Provides an insightful inference with relevant details
5R2: Determine a theme or central idea and explain how it is supported by key details; summarize a text. (RI&RL)	Partially determines a correct theme	Determines a correct theme with summary based details	Theme with evidence and explanation

Grade Level Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
5R3: In literary texts, compare and contrast two or more characters, settings, and events, drawing on specific details in the text. (RL) In informational texts, explain the relationships or interactions between two or more individuals, events, ideas, or concepts based on specific evidence from the text. (RI)	Able to determine compare and contrast relationship in narrative and in nonfiction text	Able to determine, compare and contrast relationships and support with evidence.	Able to determine compare and contrast relationship and support with evidence Evidence is high quality
5R4: Determine the meaning of words, phrases, figurative language, and academic and content specific words and analyze their effect on meaning, tone, or mood. (RI&RL)	had a basic understanding of vocabulary	understands meaning and its effects, demonstrates part to whole thinking	understands meaning and its effect and is able to explain why with varied strategies

Grade Level	Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
5	R5: In literary texts, explain how a series of chapters, scenes, or stanzas fit together to determine the overall structure of a story, drama, or poem (RL) In informational texts, compare and contrast the overall structure in two or more texts using terms such as sequence, comparison, cause/effect, and problem/solution. (RI)	literal comprehension of structure	understanding of different structures	understanding different structures and their effects on the texts

5R6: In literary texts, explain how a narrator's or speaker's point of view influences how events are described. (RL) In informational texts, analyze multiple accounts of the same event or topic, noting important similarities and differences in the point of view they represent. (RI)	able to identify narrator's point of view identifies a similarity or difference	able to identify narrator's point of view and determine a vague connection identify similarities and differences of point of view	able to identify point of view, determine and connection, and explain why considers multiple points of view and their similarities and differences attempts analysis of the multiple accounts/points of view

Grade Level Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
5R7: Analyze how visual and multimedia elements contribute to meaning of literary and informational texts. (RI&RL)	Can identify elements and show partial understanding of how they contribute to the text	Can identity elements and show understanding of how they contribute to the text	Can identify elements, show how they contribute, and explain how they contribute to the text
5R8: Explain how claims in a text are supported by relevant reasons and evidence, identifying which reasons and evidence support which claims. (RI&RL)	identify a claim in the text with an attempt at evidence or minimal evidence	identify a claim within the text and support with accurate evidence from the text or article	identify a claim within the text, support with accurate evidence from the text and explain how the evidence supports the claim
Language Standards 1 and 2 are organized within grade bands. For the Core Conventions Skills and Core Punctuation and Spelling Skills for Grades 3–5, the student is expected to know and be able to use the skills by the end of 5th grade.	Response has many errors and the response is difficult to comprehend but not impossible to comprehend.	Some errors but the response does not hinder comprehension.	Response not only has few errors and is fully comprehensible, but there is some evidence of risk-taking and usage that enhance comprehension.

Grade Level	Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
5L3: Use knowledge of language and its conventions when writing, speaking, reading, or listening.		Communicate using some inaccurate language but begins to meet the task	Communicates using accurate and specific language to complete the task or situation.	Communicate using some sophisticated language and use of tone that are appropriate to the task but may not be consistent
5L4: Determine or clarify the meaning of unknown and multiple-meaning words and phrases, choosing flexibly from a range of strategies.		These are clear and concise as written Limited range of strategies while clarifying the meanings of unknown words and phrases.	These are clear and concise as written. Shows a range of strategies while clarifying the meanings of unknown words and phrases.	These are clear and concise as written. Students show use of at least two strategies while clarifying the meanings of unknown words and phrases.
5L5: Demonstrate understanding of figurative language, word relationships, and nuances in word meanings		Some ability to identify figurative language with some connection to the purpose.	Response has an example of identifying figurative language with a word(s) that suit the purpose.	Begins to identify and insightfully explain the meaning of figurative language, word relationships, and nuances in word meanings by choosing words precisely and purposefully, but may not do so consistently.

Grade Level	Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
5L6: Acquire and accurately use general academic and content-specific words and phrases, including those that signal contrast, addition, and other logical relationships (e.g., however, although, nevertheless, similarly, moreover, in addition).		Response begins to show correct application of academic and content-specific words and phrases	Applies general and content specific words and phrases with accuracy.	General academic and content-specific words are applied accurately, words are mostly applied precisely and skillfully
5W1: Write an argument to support claims with clear reasons and relevant evidence.		Write an argument to support claims with some relevant textual information and an attempt to use content-specific vocabulary in an organized way.	Write an argument to support claims with an attempt to use multiple sources, logical transitions, and a conclusion.	Write an argument to support claims with relevant supporting evidence, use of varied vocabulary and transitions, and clear organizational style including a conclusion.

Grade Level	Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
5W2: Write informative/ explanatory texts to explore a topic and convey ideas and information relevant to the subject.				
5W4: Create a poem, story, play, artwork, or other response to a text, author, theme, or personal experience.				
5W5: Draw evidence from literary or informational texts to respond and support analysis, reflection, and research by applying the Grade 5 Reading Standards.		Write using evidence from literary or informational texts with some relevance to respond to and support analysis, reflection, and research.	Write using evidence from literary or informational texts with accuracy and partially supports with relevant analysis.	Write using the best evidence from literary or informational texts with accuracy and partially have insightful support.

ELA Grade 6

Standard	Threshold Level 2	Threshold Level 3	Threshold 4
6R1: Cite textual evidence to support an analysis of what the text says explicitly/implicitly and make logical inferences. (RI&RL)	Partially relevant inference	Logical inference with basic support that explains rather than analyzes	Logical inference with a solid analysis
6R2: Determine a theme or central idea of a text and how it is developed by key supporting details over the course of a text; summarize a text. (RI&RL)	Partially relevant theme/summary	Logical and relevant theme or summary with minimal support	Logical and relevant theme or summary with accurate support
6R3: In literary texts, describe how events unfold, as well as how characters respond or change as the plot moves toward a resolution. (RL) In informational texts, analyze how individuals, events, and ideas are introduced, relate to each other, and are developed. (RI)	Partially describes events lacking in character response, or plot resolution Beginning to analyze how individuals, events, and ideas are introduced, relate to each other.	Describes events with inconsistent support Analyze how individuals, events, and ideas are introduced, relate to each other with inconsistent support	Describes events with accurate support Analyze how individuals, events, and ideas are introduced, relate to each other with accurate support

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
6R4: Determine the meaning of words and phrases as they are used in a text, including figurative and connotative meanings. Analyze the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings. (RI&RL)	Inconsistently determines the meaning of words/phrases; Basic or limited understanding of impact of word choices on meaning, tone, and mood.	Determines the meaning of words/phrases; Analysis or understanding of impact of word choices on meaning, tone, and mood with some inconsistency.	Determines the precise meaning of words/phrases; Detailed or insightful analysis or understanding of impact of word choices on meaning, tone, and mood.
6R5: In literary texts, analyze how a particular sentence, paragraph, stanza, chapter, scene, or section fits into the overall structure of a text and how it contributes to the development of theme/central idea, setting, or plot. (RL&RI)	Partially analyze how a particular sentence, paragraph, stanza, chapter, scene, or section fits into the overall structure of a text and how it contributes to the development of theme/central idea, setting, or plot.	Describe how a particular sentence, paragraph, stanza, chapter, scene, or section fits into the overall structure of a text and how it contributes to the development of theme/central idea, setting, or plot.	Accurately support how a particular sentence, paragraph, stanza, chapter, scene, or section fits into the overall structure of a text and how it contributes to the development of theme/central idea, setting, or plot.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
6R6: Identify the point of view and explain how it is developed and conveys meaning. (RI&RL)	Partially describes the point of view with limited explanation	Able to describe the point of view with partial explanation	Describes the point of view with detailed explanation
Explain how an author's geographic location or culture affects his or her perspective. (RI&RL)	Beginning to analyze the author's perspective	Able to analyze the author's perspective	Describes and analyzes the author's perspective
6R7: Compare and contrast how different formats, including print and digital media, contribute to the understanding of a subject. (RI&RL)	partially compares and contrasts different formats	Compares and contrasts different formats and begins to describe how the formats contribute to the understanding	Compares and contrasts different formats and describes how the formats contribute to the understanding (lacking sophistication)
6R8: Trace and evaluate the development of an argument and specific claims in texts, distinguishing claims that are supported by reasons and relevant evidence from claims that are not. (RI)	Determines claims with weak evidence to support	Trace the development of supported claims made; beginning to evaluate the support of said claims	Trace the development of supported claims made; evaluates the support of said claims (lacking sophistication)

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
Language Standards 1 and 2 are organized within grade band. For the Core Conventions Skills and Core Punctuation and Spelling Skills for Grades 6–8, the student is expected to know and be able to use the skills by the end of Grade 8.	basic command of language and conventions. Errors in conventions may hinder comprehension	Command of language and conventions, some errors that do not hinder comprehension	sophisticated command of language and conventions with few errors
6L3: Use knowledge of language and its conventions when writing, speaking, reading, or listening.	general use of language	starting to use grade level language but it is not consistent	consistent and skillful use of grade level language

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
6L4: Determine or clarify the meaning of unknown and multiple-meaning words and phrases, choosing flexibly from a range of strategies.	able to determine literal meanings of unknown words	able to determine and utilize a strategy to determine meaning of unknown words or phrases	Able to decipher meaning in multiple meaning words using connotation and denotation
6L5: Demonstrate understanding of figurative language, word relationships, and nuances in word meanings.	basic and inconsistent understanding of figurative language	general understanding of figurative language. Can identify usage of fl and can explain meaning	sophisticated understanding of figurative language. Can identify and explain FL using sophisticated language
6L6: Acquire and accurately use general academic and content-specific words and phrases; apply vocabulary knowledge when considering a word or phrase important to comprehension or expression.	inconsistently uses academic and content vocab	consistently uses academic and content vocab accurately	uses content and academic vocab with explanation and to show relationships and transitional language

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
6W1: Write arguments to support claims with clear reasons and relevant evidence.	Must have a concluding statement that partly explains argument, includes some relevant information, shows limited understanding,	Uses some precise language and adds some transitions; some relevant evidence	Well organized arguments, claims; more sophisticated; and varied transitions with more advanced style and tone. Shows some level of insight.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
6W5: Draw evidence from literary or informational texts to support analysis, reflection, and research. Apply Grade 6 reading standards to both literary and informational texts, where applicable.	Some relevant evidence that supports analysis	Relevant evidence that supports analysis	Key evidence; Insightful analysis, reflection

ELA Grade 7

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
7R1: Cite textual evidence to support an analysis of what the text says explicitly/implicitly and make logical inferences. (RI&RL)	Make some illogical inferences from a text; **attempts to** cite a text when explaining what the text says explicitly/ implicitly **but may struggle with relevance.** **ORIGINAL:** Make some illogical inferences from a text; inconsistently cite a text when explaining what the text says explicitly/ implicitly.	Make logical inferences from a text; cite textual evidence to support a **simple** analysis of what the text says explicitly/implicitly. **ORIGINAL:** Make logical inferences from a text; cite textual evidence to support an analysis of what the text says explicitly/implicitly.	Make **i**nferences, at least one of which is insightful, from a text; evaluate **and/or** cite precise textual evidence to support an analysis of what the text says explicitly/implicitly. **ORIGINAL:** Make insightful inferences from a text; evaluate and cite precise textual evidence to support an analysis of what the text says explicitly/implicitly.
7R2: Determine a theme or central idea of a text and analyze its development over the course of the text; summarize a text. (RI&RL)	Determine a theme or central idea of a text, to analyze its development over the course of the text; **and/or** insufficiently summarize the key supporting details and ideas. **Theme may be vague and the summary may lack depth.** **ORIGINAL:** Determine a theme or central idea of a text, to analyze its development over the course of the text; insufficiently summarize the key supporting details and ideas.	Determine a theme or central idea of a text and **provide a basic** analysis over the course of the text; summarize the key supporting details and ideas of the text. **May struggle with detail accuracy and not entirely in their own words.** **ORIGINAL:** Determine a theme or central idea of a text and analyze its development over the course of the text; summarize the key supporting details and ideas of the text.	Determine a theme or central idea of a text, and to **attempt to** insightfully analyze how it is supported by key details; **provide an accurate summary of the text.** **ORIGINAL:** Determine a theme or central idea of a text, and insightfully analyze how it is supported by key details; provide an accurate and sophisticated summary of the text.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
7R3: In literary texts, analyze how elements of plot are related, affect one another, and contribute to meaning. (RL) In informational texts, analyze how individuals, events, and ideas are introduced, relate to each other, and are developed. (RI)	In literary texts, minimally and/or inadequately analyze how particular elements of a plot are related and affect one another or contribute to meaning. **May struggle to analyze more complex elements.** In informational texts, minimally and/or inadequately analyze how individuals, events, and ideas are introduced, and/or relate to each other, and/or are developed. **May struggle with depth of development.** **ORIGINAL:** In literary texts, partially analyze how particular elements of a plot are related and affect one another or contribute to meaning. In informational texts, partially analyze how individuals, events, and ideas are introduced, and/or relate to each other, and/or are developed.	In literary texts, analyze how particular elements of a plot are related, affect one another, **and/or** contribute to meaning. In informational texts, analyze how individuals, events, and ideas are introduced, relate to each other, and are developed. **Analysis may lack details and/or contain some inaccuracies.** **ORIGINAL:** In literary texts, analyze how particular elements of a plot are related, affect one another, and contribute to meaning. In informational texts, analyze how individuals, events, and ideas are introduced, relate to each other, and are developed.	In literary texts, insightfully analyze how particular elements of a plot are related, affect one another, and contribute to meaning, **and/or** acknowledging the subtle ways in which the relationships are depicted. In informational texts, insightfully analyze how individuals, events, and ideas are introduced, relate to each other, and are developed, **and/or** acknowledging the subtlety in the relationships. **ORIGINAL:** In literary texts, insightfully analyze how particular elements of a plot are related, affect one another, and contribute to meaning, acknowledging the subtle ways in which the relationships are depicted. In informational texts, insightfully analyze how individuals, events, and ideas are introduced, relate to each other, and are developed, acknowledging the subtlety in the relationships.

7R4: Determine the meaning of words and phrases as they are used in a text, including figurative and connotative meanings. Analyze the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings. (RI&RL)	Inconsistently determine the meaning of words and phrases as used in a text, including figurative and connotative meanings; partially analyze the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings. **(Description is good as is)** **ORIGINAL:** Inconsistently determine the meaning of words and phrases as used in a text, including figurative and connotative meanings; partially analyze the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings.	Determine the meaning of **commonly used** words and phrases as they are used in a text, including figurative and connotative meanings; analyze the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings. **May struggle to determine obscure word meanings and depth of analysis.** **ORIGINAL:** Determine the meaning of words and phrases as they are used in a text, including figurative and connotative meanings; analyze the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings.	**Determine the meaning** of words and phrases as they are used in a text, including figurative **and/or** connotative meanings; provide a detailed **and/or** insightful analysis of the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings. **ORIGINAL:** Determine with precision the meaning of words and phrases as they are used in a text, including figurative and connotative meanings; provide a detailed, insightful analysis of the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
7R5: In literary texts, analyze how structure, including genre- specific features, contributes to the development of themes and central ideas. (RL) In informational texts, analyze the structure an author uses to organize a text, including how the sections contribute to the whole and to the development of themes and central ideas. (RI)	In literary texts, minimally and/or inadequately analyze how the structure, including genre-features, contributes to the development of themes and central ideas. **May struggle with more complex themes and ideas.** In informational texts, minimally and/or inadequately analyze the structure an author uses to organize a text, including how the sections contribute to the whole **and/or** to the development of themes and central ideas. **ORIGINAL:** In literary texts, partially analyze how the structure, including genre-features, contributes to the development of themes and central ideas. In informational texts, partially analyze the structure an author uses to organize a text, including how the sections contribute to the whole and to the development of themes and central ideas.	In literary texts, analyze how structure, including genre-specific features, contributes to the development of themes and central ideas. **May struggle with certain genres.** In informational texts, analyze the structure an author uses to organize a text, including how the sections contribute to the whole and to the development of themes and central ideas. **May lack depth and/or accuracy.** **ORIGINAL:** In literary texts, analyze how structure, including genre-specific features, contributes to the development of themes and central ideas. In informational texts, analyze the structure an author uses to organize a text, including how the sections contribute to the whole and to the development of themes and central ideas.	In literary texts, **provide a detailed analysis** of how structure, including genre-specific features, contributes to the development of themes and central ideas. In informational texts, **provide a detailed analysis** of the structure an author uses to organize a text, including how the sections contribute to the whole and to the development of themes and central ideas. **ORIGINAL:** In literary texts, analyze with precision how structure, including genre-specific features, contributes to the development of themes and central ideas. In informational texts, analyze with precision the structure an author uses to organize a text, including how the sections contribute to the whole and to the development of themes and central ideas.

7R6: In literary texts, analyze how an author develops and contrasts the point of view and the perspectives of different characters or narrators. (RL) In informational texts, analyze how the author distinguishes his or her position from that of others. (RI)	In literary texts, minimally and/or inadequately analyze how an author develops and contrasts the points of view or perspectives of different characters or narrators. In informational texts, provide minimal and/or inadequate analysis of how the author distinguishes his or her position from that of others. **Original**: In literary texts, partially analyze how an author develops and contrasts the points of view and/or perspectives of different characters or narrators. In informational texts, partially analyze how the author distinguishes his or her position from that of others.	In literary texts, briefly analyze how an author develops and contrasts the points of view and in perspectives of different characters or narrators to some degree but may struggle with supporting evidence. In informational texts, briefly analyze how the author distinguishes his or her position from that of others but may struggle with identifying supporting evidence. **Original**: In literary texts, analyze how an author develops and contrasts the points of view and perspectives of different characters or narrators. In informational texts, analyze how the author distinguishes his or her position from that of others.	In literary texts, sufficiently analyze how an author develops and contrasts the points of view and perspectives of different characters or narrators. In informational texts, sufficiently analyze how the author distinguishes his or her position from that of others. **Original**: In literary texts, analyze with precision how an author develops and contrasts the points of view and perspectives of different characters or narrators. In informational texts, analyze with precision how the author distinguishes his or her position from that of others.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
7R7: Compare and contrast a written text with audio, filmed, staged, or digital versions in order to analyze the effects of techniques unique to each media and each format's portrayal of a subject. (RI&RL)	SKIP	SKIP	SKIP
7R8: Trace and evaluate the development of an argument and specific claims in a text, assessing whether the reasoning is valid, and the evidence is relevant and sufficient and recognizing when irrelevant evidence is introduced. (RI)	Limited ability to trace and evaluate the development of an argument and specific claims in a text, may have difficulty assessing whether the reasoning is valid, and the evidence is relevant or irrelevant and sufficient to support the claims. **Original**: Partially trace and evaluate the development of an argument and specific claims in a text, minimally assessing whether the reasoning is valid, and the evidence is relevant or irrelevant and sufficient to support the claims.	Adequately trace and evaluate the development of an argument and specific claims in a text, generally assessing whether the reasoning is valid, and the evidence is relevant and sufficient to support the claims and recognizing irrelevant evidence. **Original**: Trace and evaluate the development of an argument and specific claims in a text, assessing whether the reasoning is valid, and the evidence is relevant and sufficient to support the claims and recognizing irrelevant evidence.	Successfully trace and evaluate the development of an argument and specific claims in a text, effectively assessing whether the reasoning is valid, and the evidence is relevant and sufficient to support the claims, and identifying relevant and irrelevant evidence. **Original**: Skillfully trace and evaluate the development of an argument and specific claims in a text, assessing with precision whether the reasoning is valid, and the evidence is relevant and sufficient to support the claims and differentiating between relevant and irrelevant evidence.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
Language Standards 1 and 2 are organized within grade band. For the Core Conventions Skills and Core Punctuation and Spelling Skills for Grades 6–8, the student is expected to know and be able to use the skills by the end of 8th grade.	7L1: Demonstrate an emerging command of the conventions of standard English grammar and usage to produce writing with frequent errors, some of which may hinder comprehension. 7L2: Demonstrate an emerging command of the conventions of standard English capitalization, punctuation, and spelling to produce writing with frequent errors, some of which may hinder comprehension.	7L1: Demonstrate a satisfactory command of the conventions of standard English grammar and usage to produce writing with some errors that do not hinder comprehension. 7L2: Demonstrate a satisfactory command of the conventions of standard English capitalization, punctuation, and spelling to produce writing with some errors that do not hinder comprehension.	7L1: Demonstrate consistent command, with some sophistication, of the conventions of standard English grammar and usage to produce writing with few errors. 7L2: Demonstrate a consistent command, with some sophistication, of the conventions of standard English capitalization, punctuation, and spelling to produce writing with few errors.
7L3: Use knowledge of language and its conventions when writing, speaking, reading, or listening.	Exhibit an inconsistent, imprecise, and/or inaccurate use of some grade-appropriate language and its conventions when reading and writing.	Exhibit accurate use of some grade-appropriate language and its conventions when reading and writing.	Exhibit a skillful use of grade appropriate language and its conventions that may be sophisticated when reading and writing.
7L4: Determine or clarify the meaning of unknown and multiple-meaning words and phrases, choosing flexibly	Partially determine or clarify the meaning of unknown and multiple-meaning words and phrases with some inaccuracy, demonstrating a limited range of strategies.	Determine or clarify the meaning of unknown and/or multiple-meaning words and phrases, choosing from a range of strategies.	Determine or clarify with precision and detail the meaning of unknown and multiple-meaning words and/or phrases, choosing from a range of strategies with some flexibility.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
7L5: Demonstrate understanding of figurative language, word relationships, and nuances in word meanings.	Demonstrate a limited understanding of figurative language, word relationships, and/or nuances in word meanings by attempting to choose words that partially connect to the purpose.	Demonstrate an understanding of figurative language, word relationships, and/or nuances in word meanings by choosing some words that suit the purpose.	Demonstrate some sophistication in their understanding of figurative language, word relationships, and nuances in word meanings by choosing some words precisely and purposefully.
7L6: Acquire and accurately use general academic and content-specific words and phrases; apply vocabulary knowledge when considering a word or phrase important to comprehension or expression.	Attempts to use general academic and content-specific words and phrases, some of which may be inaccurate, incomplete, and/or inconsistent.	Use general academic and content-specific words and phrases carefully and/or accurately.	Use general academic and content-specific words and phrases precisely with an attempt at sophistication.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
7W1: Write arguments to support claims with clear reasons and relevant evidence.	Write partial arguments to support a claim with reasoning and/or insufficient or irrelevant evidence; may include a counterclaim; use some content-specific vocabulary and transitions to connect ideas, though they may be repetitive; concluding statement partially or inconsistently explains the argument presented.	Write a cohesive argument that supports a claim with clear reasoning and relevant evidence, with an attempt to distinguish the claim from a counterclaim; use precise and content specific vocabulary and appropriate transitions with some variety to connect ideas while maintaining appropriate style and tone; concluding statement or section that gives a basic explanation of the argument presented.	Write a fairly compelling argument that supports a claim with valid reasoning and/or relevant, well-chosen evidence, distinguishing the claim from a counterclaim; use some sophisticated and content-specific vocabulary and purposeful and varied transitions to connect ideas while attempting to maintaining sophisticated style and tone; concluding statement or section that explains the argument presented.
7W2: Write informative/ explanatory texts to examine a topic and convey ideas, concepts, and information through the selection, organization, and analysis of relevant content.	Produce informative/ explanatory texts that partially and/or inconsistently introduce and address a topic, lacking clarity and coherence, demonstrating inconsistent control of the selection, organization, and basic analysis of content and evidence that may or may not be relevant; use some content-specific vocabulary and transitions to connect ideas, though they may be repetitive; concluding statement partially or inconsistently explains the argument presented.	Produce informative/ explanatory texts that introduce and examine a topic and convey ideas clearly through the selection, organization, and grade-level analysis of relevant content and evidence; use precise and content specific vocabulary and appropriate transitions with some variety to connect ideas while maintaining appropriate style and tone; concluding statement or section that gives a basic explanation of the argument presented.	Produce informative/ explanatory texts with some precision and insight that fully introduce and examine a topic and convey ideas clearly, coherently, and in an engaging way through the selection, organization, and analysis of relevant content and evidence; use some sophisticated and content-specific vocabulary and purposeful and varied transitions to connect ideas while attempting to maintaining sophisticated style and tone; concluding statement or section that explains the argument presented.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
7W5: Draw evidence from literary or informational texts to support analysis, reflection, and research, applying Grade 7 reading standards to both literary and informational text, where applicable.	Uses partial, inconsistent, and/or irrelevant evidence from literary or informational texts in an attempt to support a superficial analysis, reflection, and/OR research, applying Grade 7 reading standards where applicable	Uses some relevant evidence from literary or informational texts to support a grade-level (e.g., literal, basic) analysis, reflection, and/OR research, applying Grade 7 reading standards where applicable	Provides relevant and well-chosen evidence from literary or informational texts but may struggle to sustain a skillful analysis, reflection, and research, applying Grade 7 reading standards where applicable

ELA Grade 8

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
8R1: Cite textual evidence to strongly support an analysis of what the text says explicitly/implicitly and make logical inferences. (RI&RL)	Make **some** inferences; **limited ability to** cite textual evidence that partially supports analysis of what the text says explicitly/implicitly.	Make logical inferences; **briefly** cite textual evidence that logically supports a **simple** analysis of what the text says explicitly/implicitly.	Make sophisticated inferences; cite **at least one piece of** textual evidence that logically supports a **somewhat** insightful analysis of what the text says explicitly/implicitly.
8R2: Determine one or more themes or central ideas of a text and analyze their development over the course of the text; summarize a text. (RI&RL)	Determine a theme or central idea of a text and partially analyze its development over the course of the text; **and/or** insufficiently summarize the key supporting details and ideas. **Theme may be vague, and the summary may lack depth.**	Determine one or more themes or central ideas of a text and provide a **basic analysis** of its development over the course of the text; provide a **brief** summary of the text using key supporting details and ideas. **May struggle with detail accuracy.**	Determine one or more themes or central ideas of the text and **attempt to** provide an insightful analysis of its development over the course of the text; **provide an accurate summary of the text.**

s

1

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
8R3: In literary texts, analyze how particular lines of dialogue or events propel the action, reveal aspects of a character, or provoke a decision. (RL) In informational texts, analyze how individuals, events, and ideas are introduced, relate to each other, and are developed. (RI)	In literary texts, **minimally and/or inadequately analyze** how particular lines of dialogue or incidents in a story or drama propel the action, reveal aspects of a character, or provoke a decision. **May struggle to analyze more complex elements.** In informational texts, provide a partial analysis of how individuals, events, and ideas are introduced, relate to each other, and are developed. **May struggle with depth of development.**	In literary texts, provide a **brief** analysis of how particular lines of dialogue or incidents in a story or drama propel the action, reveal aspects of a character, or provoke a decision. **Analysis may lack details and/or complexity.** In informational texts, provide an analysis of how individuals, events, and ideas are introduced, relate to each other, and are developed. **Analysis may lack details and/or contain some inaccuracies.**	In literary texts, provide a detailed **and/or** insightful analysis of how particular lines of dialogue or incidents in a story or drama propel the action, reveal aspects of a character, or provoke a decision. In informational texts, provide a detailed **and/or** insightful analysis of the subtleties and interactions of how individuals, events, and ideas are introduced, relate to each other, and are developed.
8R4: Determine the meaning of words and phrases as they are used in a text, including figurative and connotative meanings. Analyze the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings. (RI&RL)	Inconsistently determine the meaning **of commonly used Tier 1** words and phrases as used in a text, including figurative and connotative meanings; provide a partial analysis of the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings. https://infercabulary.com/the-tiers-of-vocabulary/	Determine the meaning of **Tier 2** words and phrases as they are used in a text, including figurative and connotative meanings; analyze the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings. **May struggle to accurately identify the less commonly used definition of a Tier 2 word and depth of analysis.**	Determine with **some** precision the meaning of **Tier 2** words and phrases as they are used in a text, including figurative and connotative meanings; provide a detailed **and/or** nuanced analysis of the impact of specific word choices on meaning, tone, and mood, including words with multiple meanings.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
8R5: In literary texts, and informational texts, compare and contrast the structures of two or more texts in order to analyze how the differing structure of each text contributes to overall meaning, style, theme, or central idea. (RI&RL)	**Minimally and/or inadequately** compare and contrast the structure of two or more texts and attempt to provide an analysis of how the differing structure of each text contributes to its meaning, style, theme, or central idea. **May struggle with more complex themes and/or ideas.**	Compare and contrast the structure of two or more texts and analyze how the differing structure of each text contributes to its meaning, style, theme, or central idea. **May struggle with the connection between two or more texts.**	Provide a comparison and contrast, **with some detail**, of the structure of two or more texts with an insightful analysis of how the differing structure of each text contributes to its meaning, style, theme, or central idea, **but may struggle with nuances of complex texts and/or questions.**
8R6: In literary texts, analyze how the differences between the point of view, perspectives of the characters, the audience, or reader create effects such as mood and tone. (RL) In informational texts, analyze how the author addresses conflicting evidence or viewpoints. (RI)	In literary texts, has **difficulty analyzing and/or identifying** how differences in the points of view and the perspectives of the characters, the audience, or reader create such effects as mood and tone, **especially with complex texts.** In informational text, **has difficulty analyzing and/or identifying** how the author addresses conflicting evidence or viewpoints, **especially with complex texts.**	In literary texts, **provide a basic or literal analysis** of how differences in the points of view and the perspectives of the characters, the audience, or reader create such effects as mood and tone, **especially with less complex texts.** In informational texts, **provide a basic or literal analysis** of how the author addresses conflicting evidence or viewpoints, **especially with less complex texts.**	In literary texts, **provide an insightful analysis** of how differences in the points of view and the perspectives of the characters, the audience, or reader create such effects as mood and tone, **but may struggle with nuances of complex texts and/or questions.** In informational texts, **provide a detailed analysis** of how the author addresses conflicting evidence or viewpoints **but may struggle with nuances of complex texts and/or questions.**

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
8R7: Evaluate the advantages and disadvantages of using different media—text, audio, video, stage, or digital—to present a particular subject or idea and analyze the extent to which a production remains faithful to or departs from the written text. (RI&RL) **NOT ASSESSED**	Partially evaluate the advantages and disadvantages of using different media—text, audio, video, stage, or digital—to present a particular subject or idea and provide a simple analysis of the extent to which the production remains faithful or departs from the written text.	Evaluate the advantages and disadvantages of using different media—text, audio, video, stage, or digital—to present a particular subject or idea and provide an analysis of the extent to which the production remains faithful or departs from the written text.	Provide an insightful evaluation of the advantages and disadvantages of using different media—text, audio, video, stage, or digital—to present a particular subject or idea and provide a detailed analysis of the extent to which the production remains faithful or departs from the written text.
8R8: Trace and evaluate an argument and specific claims in a text, assessing whether the reasoning is valid, and the evidence is relevant and sufficient and recognizing when irrelevant evidence is introduced. (RI)	Partially trace and evaluate the argument and specific claims in an informational text, **with inconsistencies**; **may recognize** whether the reasoning is valid and the evidence is relevant and sufficient to support the **obvious claims**; may recognize when irrelevant evidence is introduced **with some inaccuracies.**	Trace and **provide a simple evaluation** of the argument and specific claims in an informational text and **recognize** whether the reasoning is valid and the evidence is relevant and sufficient to support the claims; recognize when irrelevant evidence is introduced. **May recognize** when irrelevant evidence is introduced.	**Effectively** trace and evaluate the argument and specific claims in an informational text, **accurately** assessing whether the reasoning is valid, and the evidence is relevant and sufficient to support the claims; recognize when irrelevant evidence is introduced.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
Language Standards 1 and 2 are organized within grade band. For the Core Conventions Skills and Core Punctuation and Spelling Skills for Grades 6–8, the student is expected to know and be able to use the skills by the end of Grade 8.	L1: Demonstrate **a beginning** command of the conventions of standard English grammar and usage when writing to produce writing with **frequent errors, some of which may hinder comprehension.** L2: Demonstrate **a beginning** command of the conventions of standard English capitalization, punctuation, and spelling to produce writing with **frequent errors, some of which** may hinder comprehension.	L1: Demonstrate a **satisfactory** command of the conventions of standard English grammar and usage when writing to produce writing with **some** errors that do not hinder comprehension. L2: Demonstrate a satisfactory command of the conventions of standard English capitalization, punctuation, and spelling to produce writing with **some** errors that do not hinder comprehension.	L1: Demonstrate a **consistent command, with some sophistication,** of the conventions of standard English grammar and usage when writing to produce few errors. L2: Demonstrate a **consistent** command, **with some sophistication,** of the conventions of standard English capitalization, punctuation, and spelling to produce writing with few errors.
8L3: Use knowledge of language and its conventions when writing, speaking, reading, or listening.	Exhibit **an inconsistent or a limited ability** to use grade-appropriate language and its conventions when reading and writing.	Exhibit an accurate use of grade-appropriate language **with an occasional misuse of conventions when reading and writing.**	Exhibit a skillful use of grade-appropriate language and its conventions when reading and writing, **with occasional misuse of sophisticated language.**

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
8L4: Determine or clarify the meaning of unknown and multiple-meaning words and phrases, choosing flexibly from a range of strategies.	Partially determine or clarify the meaning of unknown and multiple-meaning words and phrases **with some inaccuracy**, demonstrating **a** limited range of strategies.	Determine or clarify the meaning of unknown and**/or** multiple-meaning words and phrases, choosing ~~flexibly~~ from a range of strategies.	Determine or clarify with **some** precision and detail the meaning of unknown and multiple-meaning words and**/or** phrases, choosing from a range of strategies **with some flexibility.**
8L5: Demonstrate understanding of figurative language, word relationships, and nuances in word meanings.	Demonstrate a limited understanding of figurative language, word relationships, and**/or** nuances in word meanings by choosing words that partially connect to the purpose.	Demonstrate an understanding of figurative language, word relationships, and**/or** nuances in word meanings by choosing **some** words that suit the purpose.	Demonstrate **some sophistication in their** understanding of figurative language, word relationships, and nuances in word meanings by choosing **some** words precisely and purposefully.
8L6: Acquire and accurately use general academic and content-specific words and phrases; apply vocabulary knowledge when considering a word or phrase important to comprehension or expression.	**Attempts to** use general academic and content-specific words and phrases **some of which may be inaccurate, incomplete, and/or inconsistent.**	Use general academic and content-specific words and phrases carefully and**/or** accurately.	Use general academic and content-specific words and phrases precisely and with **an attempt at** sophistication.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
8W1: Write arguments to support claims with logical reasons and relevant evidence; acknowledge and distinguish the counterclaim; demonstrate an understanding of the topic or text.	Produce an argument that partially supports a claim with reasoning and **some** relevant evidence, may offer a counterclaim; use **some** content-specific vocabulary and **limited** transitions to connect ideas; provide a concluding statement that partially or **inconsistently** explains the argument presented.	Produce an argument that supports a claim with logical reasoning and relevant evidence, **with an attempt to** distinguish the claim from a counterclaim; use precise and content specific vocabulary, and appropriate transitions **with some variety** to connect ideas; provide a concluding statement or section that **gives a basic explanation of** the argument presented, while maintaining appropriate style and tone.	Produce **a somewhat i**nsightful argument that supports a claim with insightful reasoning and**/or** relevant evidence, distinguish the claim from a counterclaim; use **some** sophisticated and content-specific vocabulary, and purposeful and varied transitions to connect ideas; provide a concluding statement or section that explains the argument presented, **while attempting to maintain** sophisticated style and tone.
8W2: Write informative/ explanatory texts to examine a topic and convey ideas, concepts, and information through the selection, organization, and analysis of relevant content.	Produce informative/ explanatory texts that partially **and/or inconsistently** introduce and address a topic, lacking clarity and coherence; demonstrate inconsistent control of the selection, organization, and **basic** analysis of content and evidence **that may or may not be relevant**; use content-specific vocabulary and transitions to connect ideas, **though they may be repetitive**; provide a concluding statement that partially **or inconsistently** explains the argument presented.	Produce informative/ explanatory texts that ~~completely~~ introduce and examine a topic and convey ideas clearly through the selection, organization, and **grade-level** analysis of relevant content and evidence; use precise and content-specific vocabulary, and appropriate transitions **with some variety** to connect ideas; provide a concluding statement or section that **gives a basic explanation of** the argument presented, while **largely** maintaining appropriate style and tone.	Produce informative/ explanatory texts **with some insight and precision** that fully introduce and examine a topic and convey ideas clearly, coherently, and in a **mostly** engaging way through the selection, organization, and analysis of relevant content and evidence; use **some** advanced and content specific vocabulary, and purposeful and varied transitions to connect ideas, while **attempting to maintain sophisticated** style and tone; provide a concluding statement or section that explains the argument presented.

Standard	Threshold Level 2	Threshold Level 3	Threshold Level 4
8W5: Draw evidence from literary or informational texts to support analysis, reflection, and research. Apply Grade 8 reading standards to both literary and informational texts, where applicable.	Partially draw evidence from literary or informational texts **in an attempt** to support **a superficial** analysis, reflection, and**/or** research, applying Grade 8 reading standards to both literary and informational texts, where applicable. **Original:** Partially draw evidence from literary or informational texts to support analysis, reflection, and research, applying Grade 8 reading standards to both literary and informational texts, where applicable.	Draw **some** evidence from literary or informational texts to support **basic** analysis, reflection, and**/or** research, applying Grade 8 reading standards to both literary and informational texts, where applicable. **Original:** Draw evidence from literary or informational texts to support analysis, reflection, and research, applying Grade 8 reading standards to both literary and informational texts, where applicable.	Draw evidence from literary or informational texts **with some insight** to support analysis, reflection, and**/or** research, applying Grade 8 reading standards to both literary and informational texts, where applicable. **Original:** Insightfully draw evidence from literary or informational texts to support analysis, reflection, and research, applying Grade 8 reading standards to both literary and informational texts, where applicable.

Math Grade 3

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students represent and solve problems involving multiplication and division. (NY3.OA.1-4)	1. Number less than or equal to 10 but greater than 5 2. Products above 25 but less than or equal to 50 3. Equal groups and array problems. 4. Larger number for multiplication. 5. No visual support to interpret products. 6. Visual models for multiplication and division one step word problem. 7. Finding the unknown factor in a multiplication question. Finding factor versus finding products.	1. Use of variables(symbols) to represent the unknown factor 2. including 10 3. Word problems with measurement quantities. 4. Word problems within 100. 5. Multiplication equation less than or equal to 10 6. No visual models for word problems. 7. Unknown factor for both multiplication and division.	1. Two-step word problems. 2. Real world application. 3. Unknown whole number in multiplication and division in real-world problems. 4. More from the abstract understanding. 5. All whole numbers 6. Two-digit numbers

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students understand properties of multiplication and the relationship between multiplication and division. (NY3.OA.5,6)	• Apply the commutative property. • Apply the distributive property. • More challenging multiplication facts. • Division problems. • Factors less than or equal to 5.	• Apply the associative, commutative, and distributive properties. • Number less than 10. • Rewrite division problem as related multiplication problem. • Relate the unknown factors to the number of groups or the size of the group.	• Explain the properties (associative, commutative, distributive). • Explain relationships between multiplication and division problems with numbers less than 10. • Use multiplication to solve division problems. • Use division to solve multiplication problems.

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students solve problems involving the four operations and identify and explain patterns in arithmetic. (NY3.OA.8,9)	• Two step word problems using addition and subtraction. • Numbers less than 10. • Assess for reasonableness using mental computation and estimation strategies to solve two step word problems using numbers less than or equal to 5. • Identify the pattern and find missing numbers involving addition with numbers less than or equal to 10. • Identify the pattern and find missing numbers involving multiplication with 2, 5, and 10.	• Two step word problems using multiplication and division. • Numbers less than 10. • Assess for reasonableness using mental computation and estimation strategies to solve two step word problems using numbers within 100 and factors less than 10. • Extend the pattern involving addition with numbers less than or equal to 10 (includes use of table). • Extend the pattern involving multiplication with 2, 5, and 10 (includes use of table).	• Two step word problems using any two operations. • Unknown in any position. • Numbers greater than or equal to 10. • Assess and explain the reasonableness using mental computation and estimation strategies to solve two step word problems using numbers greater than or equal to 10. • Apply the pattern and explain the properties used to solve a real-world word problem.

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Use place value understanding and properties of operations to perform multi-digit arithmetic. (NY-3.NBT.1-4)	• Using visual aids to round three-digit numbers to nearest 10 or 100. • Multiply one digit by a multiple of 10 (range 10-50). • Three-digit numbers representing hundreds, tens, and ones. • Identifying that a group of 10 tens is equal to 100. • Read and write three-digit numbers using base-ten numerals and expanded form.	• No visual aid to round three-digit numbers to nearest 10 or 100. • Multiply one digit by a multiple of 10 (range 10-90) using associative and distributive properties. • Four-digit numbers representing thousands, hundreds, tens, and ones. • Identifying that a group of 10 hundreds is equal to 1, 000. • Read and write four-digit numbers using base-ten numerals, number names, and expanded form.	• No visual aid to round four- or five-digit numbers to nearest 10, 100, or 1000. • Use associative and distributive properties to explain the patterns when multiplying by multiples of 10. • Five-digit numbers representing ten thousands, thousands, hundreds, tens, and ones. • Identifying that a group of 10 thousand is equal to 10,000. • Read and write five-digit numbers using base-ten numerals, number names, and expanded form.

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students develop understanding of fractions as numbers. (NY-3.NF.1-3)	1. No visual models. 2. Represent unit fraction with denominator of 8 is added. 3. Able to partition the parts on the number line. 4. All fractions within 0-1 with denominators of 2, 4 and 8. 5. Recognize that b is the total number of parts in the whole. 6. Represent non-unit fractions with denominators of 8. 7. Equivalent fractions with denominators of 8 on the number line. 8. Generate equivalent fractions with denominators of 8. 9. No visual models for expression of fractions as whole numbers with denominators of 8. 10. Without visual models, use the <, > or = symbols to compare fractions and denominators of 8. 11. Compare fractions with the same numerator. and denominators of 8.	1. Understand unit fractions. 2. Understand that a/b fractions are made up of a, 1/b fraction pieces. ex. ¾ is 3 one-fourths. 3. denominators of 3 and 6 4. Recognize that each partition on the number line represents 1/b of the line. 5. Able to partition and label the number line to show that each partition is a, 1/b of the number line within 1. ex. ¼, 2/4, ¾, 4/4, that ¾ is 3 one-fourths on the number line. 6. Equivalents with denominators of 3 and 6 on the number line. 7. No visual models to generate equivalent fractions including denominators of 3 and 6. 8. Recognize fractions that are equivalent to a whole number. ex. 4/4 =1 9. Able to explain the use of <, > or = symbols to compare fractions including denominators of 3 and 6.	1. Application of unit fractions (1/b). 2. Explain unit fraction. 3. Including denominator to 10. 4. Generate and explain equivalent fractions. 5. Numbers greater than 1. ex 8/4. 6. Compare more than 2 fractions with the same numerator or denominator.

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students solve problems involving measurement and estimation of intervals of time, liquid volumes, and masses of objects. (NY-3.MD.1,2)	1. Read, write and measure time in intervals of one, five, fifteen, and thirty minutes. 2. Quarter past and quarter to 3. One step word problems. 4. Addition of time intervals in minutes. 5. Subtract Kg units included. one-step word problems of masses and volume.	1. Read, write and measure time in minutes 2. Solve one-step word problems with addition or subtraction 3. Crossing the new hour within word problems 4. Crossing am to pm in the word problems 5. Without models to measure or estimate liquid volumes and masses of objects. 6. All units (l, g, kg) 7. All operations and all units included. one-step word problems of masses and volume. 8. Unit in the word problems are the same. No conversions. 9. No visuals	1. Read, write, and measure time in minutes in two-step word problems. 2. Assess or explain a solution using estimation of two-step word problems.

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Solve problems involving measurement and estimation of intervals of time, liquid volumes, and masses of objects. (NY-3.MD.3-4, 8)	1. Solve one-step "how many more" and " how many less" problems. 2. Using information presented in a scaled picture graph or a scaled bar graph 3. Nearest half inch 4. Rulers marked with halves 5. Scale is marked off in halves for line plots	1. Draw or interpret a graph from data with up to five categories. 2. Two-step problems 3. Nearest quarter inch 4. Rulers marked with halves and fourths 5. Scale is marked off in halves or quarters for line plot. 6. Make line plot in whole, halves, quarters	1. Multi-step problems 2. Comparing problems with more than two categories 3. Nearest eighth inch 4. Rulers marked with eighths. 5. Scale is marked off in eighths for line plot. 6. Several categories (not just 5)

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students understand concepts of area and relate area to multiplication and to addition (geometric measurement). (NY-3.MD.5-7	• using the term "one square unit" vs. "unit square"	• moving away from visual models • use of multiplication to find area • use of real world problems • knowing products as square units • distributive property • can have one unknown side length	• more than one unknown side length • more of an emphasis on real-world problems • explain area vs. determining area • comparing the area of 2 or more figures • create real-world problems

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students recognize perimeter as an attribute of plane figures and distinguish between linear and area measures. (NY-3.MD.8)	• less visual models • understand that rectangles that have same area can have different perimeters • understand that rectangles that have same perimeter can have different area	• real-world problems • finding one unknown side given the perimeter • identify that rectangles that have same area can have different perimeters • identify that rectangles that have same perimeter can have different area	• solving problems involving perimeters of two different polygons given most of the side lengths

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Reason with shapes and their attributes. (NY-3.G.1-2)	• includes pentagons and hexagons • shapes divided into 8 parts vs. 2 or 4 parts and expressing each part as a fraction of the whole	• no visual models • shapes divided into 3 or 6 parts and expressing each part as a fraction of the whole • identify shapes that don't belong to a subgroup	• recognizing and classifying regular and irregular shapes • creating and partitioning shapes into equal parts and expressing as a fraction

Threshold PLDs worksheet
Math Grade 4

Cluster 1	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students use the four operations with whole numbers to solve problems. (NY-4.OA.1-3)	• Use visual model • Relating the comparison to multiplication • Use multiplication and division to distinguish between additive comparisons and multiplicative comparisons using whole number factors • One step word problems using drawings and manipulatives to represent the problem • Represent two step word problems using equations or expressions with a letter representing the unknown quantity • Solve two step word problems using any of the four operations with whole numbers and having whole number answers	• Interpret multiplication equations • Represent verbal statements of multiplicative comparisons as multiplication equations • Use multiplication and division to distinguish between additive comparisons and multiplicative comparisons • One step or two step problems using drawings or equations with a symbol for the unknown number • Represent multi step word problems using equations or expressions with a letter representing the unknown quantity • Solve multi step word problems using any of the four operations with whole numbers and having whole number answers, including problems in which remainders must be interpreted	• Use multiplication and division to solve multi step word problems involving multiplicative comparisons • Create real world problems that can be solved using multiplicative comparisons • Solve multi step word problems using any four operations with whole numbers and assess the reasonableness of answers using mental computation and estimation strategies including rounding by providing a valid mathematical explanation

Cluster 2	Threshold Level 2	Threshold Level 3	Threshold Level 4
Gain familiarity with factors and multiples. (NY-4.OA.4)	• Find all factor pairs for a whole number in the range of 26-100 • Given a model • Determine whether a given whole number is in the range of 26-100 is a multiple of a given one digit number • Identify multiples of a given whole number greater than 5 but less than 10 • Determine whether a given whole number is in the range of 26-50 is prime or composite	• Find all factor pairs for a whole number in the range of 1-100 • No model • Determine whether a given whole number is in the range of 1-100 is a multiple of a given one digit number • Determine whether a given whole number is in the range of 51-100 is prime or composite	• Explain the relationship between a multiple and a factor of a given whole number • No model • Explain how you know a given whole number is a multiple of a given one digit number • Explain why a given whole number is prime or composite

Cluster 3	Threshold Level 2	Threshold Level 3	Threshold Level 4
Generate and analyze patterns. (NY-4.OA.5)	• Generate a number or shape pattern that follows a given arithmetic rule involving addition or subtraction or identify the rule itself	• Generate a number or shape pattern that follows a given arithmetic rule involving multiplication or division • Identify the rule • Informally explain features of the pattern that were not explicit in the rule itself (ex. odd/even)	• Given a pattern that follows an arithmetic rule, generate a different pattern using the same rule • Identify the rule • Explain features of the pattern that were not explicit in the rule itself (ex. odd/even)

Cluster 4	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students generalize place value understanding for multi-digit whole numbers. (NY-4.NBT.1-3)	• May use drawings, models, manipulatives, and other aides • In whole numbers up to four digits, recognize that a digit in one place represents ten times as much as it represents in the place to its right. • With no visual models or manipulatives, read and write 4 digit numbers using base-ten numerals, numbers names, and expanded form. • With no visual models or manipulatives, compare two four digit numbers using >, <, =. • With no visual models or manipulatives, round four digit whole numbers to any place.	• No use of drawings, models, or manipulatives • In whole numbers up to multi digits, recognize that a digit in one place represents ten times as much as it represents in the place to its right. • With no visual models or manipulatives, read and write multi digit numbers using base-ten numerals, numbers, names, and expanded form. • With no visual models or manipulatives, compare two multi digit numbers using >, <, =. • Using place value understanding, round multi digit whole numbers to any place.	• Explain the relationship between place value and multiplication or division. • With no visual models or manipulatives, compare or order three or more multi digit numbers using >, <, =. • Explain the reasons for your comparisons. • Given a context, choose an appropriate rounded number.

Cluster 5	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students use place value understanding and properties of operations to perform multi-digit arithmetic. (NY-4.NBT.5,6)	1. Multiply three-digit numbers by a one-digit number. 2. Divide three-digit numbers by a one-digit number – no remainders! 3. With visuals (drawings or models).	1. Multiply four-digit numbers by a one-digit number. 2. Multiply a two-digit number by a two-digit number. 3. Divide four-digit numbers by a one-digit number – with or without a remainder. 4. Illustrate and explain the calculations (equations, arrays, area models).	1. Multiply multi-digit numbers by a one-digit or two-digit number. 2. Divide multi-digit numbers by a one-digit number – with or without a remainder. 3. Explore connection between strategy and standard algorithm.

Cluster 6	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students extend understanding of fraction equivalence and ordering. (NY-4.NF.1,2)	1. Generate two equivalent fractions (denominators 2, 3, 4, 6, 8). 2. May use a model, but not given. 3. Compare two fractions with like or unlike numerators or denominators. 4. Understand that fractions must refer to the same whole in order to compare.	1. Explain why fractions are equivalent using fraction models  2. Recognize and generate equivalent fractions (addition of denominators 5, 10, 12, 100). 3. Number and size of the parts can differ, even though the whole is the same size. 4. Compare fractions with like or unlike numerators or denominators. 5. Justify comparisons.	1. Generate equivalent fractions with any denominator (e.g. $\frac{3}{4} \times \frac{2}{2} = \frac{6}{8}$ because you can multiply the numerator and the denominator by $\frac{2}{2}$). 2. Order more than two fractions with different numerators and denominators.

Cluster 7	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students build fractions from unit fractions by applying and extending previous understandings of operations on whole numbers. (4.NF.3,4)	1. Add and subtract fractions with like denominators (no visuals). 2. Add and subtract mixed numbers. 3. Understand addition of fractions as joining of parts and subtraction as separating of parts. 4. Solve word problems. 5. Decompose fractions in more than one way (no model). Record using an expression/equation. 6. Recognize the relationship between a mixed number and an improper fraction with like denominators. 7. Recognize the equivalence of a mixed number and an improper fraction with like denominators. 8. Recognize non-unit fraction is equivalent to a unit fraction times a whole number (e.g. $\frac{1}{4} \times 5 = \frac{5}{4}$).	1. Add and subtract fractions (add denominators of 5, 10. 12, 100). 2. Decompose mixed numbers. 3. Justify decomposition. 4. add and subtract mixed numbers with like denominators using equivalent fractions. 5. Use multiples of fractions to solve real world problems (e.g. ¾ = 3 x ¼).	1. Create, solve, and explain mathematical and word problems that involve addition and subtraction of fractions. 2. Add and subtract mixed numbers using the properties of operations and the relationship between addition and subtraction. 3. Solve word problems with a whole number by a fraction (more than one strategy/pathway).

Cluster 8	Threshold Level 2	Threshold Level 3	Threshold Level 4
Understand decimal notation for fractions and compare decimal fractions. (NY-4.NF.5-7)	1. Explain the relationship between 10 and 100 to identify equivalent decimals and fractions (no visual model). 2. Use decimal notation with 10ths and 100ths. 3. Compare two decimals using a visual model to hundredths using symbols (<, >, =).	1. Add two fractions with denominators of 10 or 100. 2. Recognize an equivalent decimal on a number line and vice versa. 3. Compare two decimals in tenths or two decimals in hundredths (no visuals). 4. Justify the comparison.	1. Explain equivalent fractions with denominators of 10 or 100. Add them. 2. Compare more than two decimals in tenths and hundredths.

Cluster 9	Threshold Level 2	Threshold Level 3	Threshold Level 4
Solve problems involving measurement and conversion of measurements from a larger unit to a smaller unit. (NY-4.MD.1-3)	• select appropriate unit of measurement of given object • move from + and - of one-step story problems to multi. and division • given conversion and/or visual aid, convert from larger unit to smaller unit • find area or perimeter in real world problem	• know relative sizes of measurement units • know conversion and convert from larger to smaller • record conversions in a table • solve two step problems using the four operations • apply area and perimeter	• apply relative sizes of measurement to real world • know and convert smaller to larger • multi-step word problems • know conversion factor and convert larger to smaller • know that rectangles can have same area but different perimeter and vice versa

Cluster 10	Threshold Level 2	Threshold Level 3	Threshold Level 4
Represent and interpret data. (NY-4.MD.4)	• + and - fractions in line plots using ½ and 1/4	• make a line plot and solve using + and - using ½, ¼ and 1/8	• analyzing data and writing questions about it

Cluster 11	Threshold Level 2	Threshold Level 3	Threshold Level 4
Geometric measurement: understand concepts of angle and measure angles. (NY-4.MD.5-7)	• given a visual aid, recognize that a circle is composed of 360 one-degree angles • one degree = 1/360 of a circle • measure non-benchmark angles with a protractor • know that 1 angle can be decomposed into 2 smaller angles	• 1/360 of a circle is used to measure angles • recognize that a one-degree angle is used to make up angles • identify an angle of the same measure and/or measure or sketch angles with a protractor • add/subtract to find the missing angle	• identify and analyze angles in geometric shapes in the real world • decompose an angle and generate and solve + or - problem

Cluster 12	Threshold Level 2	Threshold Level 3	Threshold Level 4
Draw and identify lines and angles and classify shapes by properties of their lines and angles. (NY-4.G.1-3)	• moves to recognizing and identifying angles and types of lines • recognizing triangles based on angle size • recognizing that a quad with exactly 2 pairs of parallel sides is a square, rectangle, or parallelogram • recognize that a quad with 4 right angles is a square or rectangle • identify a symmetrical figure or draw a line of symmetry	• drawing lines and angles and identifying them in 2D shapes • identifying and naming triangles based on angle size • identify and name quads as parallelograms • identify and name quads with 4 right angles as rectangles • identify figures with lines of symmetry, the number of lines a figure has and drawing lines	• classifying triangles based on angle size • classify quads as parallelograms based on presence or absence of parallel lines • classify quads based on the presence or absence of right angles • explain what being symmetrical means using symmetrical and non-symmetrical figures

Math Grade 5

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students write and interpret numerical expressions. (NY-5.OA.1-2) NY-5.OA.2 Write simple expressions that record calculations with numbers, and interpret numerical expressions without evaluating them. e.g., Express the calculation "add 8 and 7, then multiply by 2" as $(8 + 7) \times 2$. Recognize that $3 \times (18{,}932 + 921)$ is three times as large as $18{,}932 + 921$, without having to calculate the indicated sum or product.	(OA.1) • Parentheses (possible) (OA. 2) • Write and interpret two steps, no grouping (same as 1)	(OA.1) • More than two operations (possible) (OA. 2) • Write simple expressions • Without evaluating	(OA.1) • Two or more types of grouping • Powers of ten (possible) (OA. 2) • Two or more types of grouping • Power of ten

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students analyze patterns and relationships. (NY- 5.OA.3)	(OA.3) • **Identify** relationships between corresponding terms • Form ordered pairs of corresponding terms from two patterns	(OA.3) • Generate two numerical patterns using two given rules	(OA.3) • Generate the rules • Explain relationships between corresponding terms • Explain why the ordered pair is graphed on the coordinate plane

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students understand the place value system. (NY-5.NBT.1-4) NY-5.NBT.2 Use whole-number exponents to denote powers of 10. Explain patterns in the number of zeros of the product when multiplying a number by powers of 10, and explain patterns in the placement of the decimal point when a decimal is multiplied or divided by a power of 10.	(NBT.1) • Recognize that in any multi-digit number • No manipulatives (NBT.2) • Use visual model (no Level 1) • Identify with no visual (NBT.3) • Read and write decimals to the **hundredths** using base-ten numerals, number names, expanded form, and inequality symbols (>, <, =) using visual models **No Manipulatives** (NBT.4) • Rounding to the hundredths with visual models no manipulatives	(NBT.1) • No visual (NBT.2) • Whole number exponents • No visual for powers of 10 • No visual for patterns in the zeros • Explain the patterns of the decimal point (NBT.3) • **Compare** decimals to the **thousandths** • No visual (NBT.4) • Rounding to any place value	(NBT.1) • Demonstrate or explain that in any multi-digit number (NBT.2) • Evaluating numerical expressions involving whole-number exponents • Analyze • Use patterns in the placement of the decimal point to evaluate numerical expressions (NBT.3) • (NBT.4)

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students perform operations with multi-digit whole numbers and with decimals to hundredths. (NY-5.NBT.6-7)	(NBT.6) • No visual and/or manipulatives • Divide whole numbers with up to **three-digit dividends** and **two-digit divisors, using strategies based on** place value, the properties of operations, and/or the relationship between multiplication and division. Calculate by using equations, rectangular arrays, and/or area models. (NBT.7) • Add and subtract two decimals involving tenths and/or hundredths (decimals with the same number of digits) using concrete models, drawings, or strategies based on place value, properties of operations, and/or the relationship between addition and subtraction. Relate the strategy to a written method. • Multiply and divide decimals involving tenths using concrete models or **drawings** and strategies based on place value, **properties of operations, and/or the relationship between multiplication and division. Relate the strategy to a written method.**	(NBT.6) • Find whole-number quotients of whole numbers with up to **four-digit dividends and two digit divisors,** • Illustrate and explain the calculation by using equations, rectangular arrays, and/or area models. (NBT.7) • Properties of operations, and/or the relationship between addition and subtraction. Relate the strategy to a written method, and explain the reasoning used.‡ (5.NBT.7) • Decimals may not have the same number of digits • **Hundredths** • Explain the reasoning used	(NBT.6) • Identify relationships between different approaches. • Check reasonableness of answers using a standard algorithm for multiplication. (NBT.7) • Apply these strategies to a real-world context. • Apply these strategies to a real-world context, relate the strategy to a written method, and explain the reasoning used.

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students use equivalent fractions as a strategy to add and subtract fractions. (NY-5.NF.1-2)	- Produce equivalent sum or difference with like denominators (5.NF.1) - Use benchmark fractions with unlike denominators (5.NF.2) - Use number lines or other visual models (5.NF.2)	- Include mixed numbers - Generate equivalent fractions (5.NF.1) - Use fractions that are not benchmark (5.NF.2) - Use mental estimation (5.NF.2) - Not showing of work	- N/A (5.NF.1) - Create the word problem (5.NF.2) - Extended response. Explanation of work.

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students apply and extend previous understandings of multiplication and division to multiply and divide fractions. (NY-5.NF.3-7)	- No Manipulatives (5.NF.3) - Use of mixed numbers - Does not use concrete models (5.NF.3) - Multiplying fractions not repeated addition - Use of fraction model (5.NF.4a) - Multiply - Student tiles with unit squares - No visual model (5.NF.5a) - N/A (5.NF.5b) - Multiplying fraction by fraction - Divisor can be a fraction (5.NF.6,5.NF.7c) - No concrete models-visual models (5.NF.7a-b)	- No visuals (5.NF.3) - No visuals (5.NF.3) - Interpreting the product as a part of a part - No use of models (5.NF.4a) (Use numerical example to explain) - Tiling with rectangles (factional parts) instead of unit squares (5.NF.4b) - Explain (5.NF.5a) - Explaining the product when multiplied by a fraction more than one or less than one (5.NF.5b) - Multiplying by mixed numbers - Interpreting the product and/ or quotient (5.NF. 6, 5.NF.7c) - - No visuals - Interpret the quotient (5.NF.7a-b)	-N/A (5.NF.3) -N/A (5.NF.3) -N/A (5.NF.4a) - Create area models and explain multiplication of fractions (5.NF.4b) - N/A (5.NF.5a) - N/A (5.NF.5b) - Create problems and creating context and equations including rectangular areas (5.NF.6, 5.NF.7c) - Solve real world problems (5.NF.7a-b)

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students convert like measurement units within a given measurement system. (NY-5.MD.1)	• Can apply conversion to solve one-step real-world problem	• Can apply conversion to solve multi-step real-world problem	• Can apply conversion to create and solve multi-step real-world problem

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students represent and interpret data. (NY-5.MD.2)	• Given a partially completed line plot students can complete it using fractional data	• Make a line plot from given fractional data	• Gather fractional data and create a line plot

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4

Students understand concepts of volume and relate volume to multiplication and to addition (geometric measurement). (NY-5.MD.3-5)	• Explain how volume is measured in cubic units • Identify volume of a right-rectangular prism • Properly label volume in terms of cubic cm, cubic in, or cubic ft • (Same standard as 1) • Use a visual model (instead of physical) and apply the formula Ixwxh to find volume • Composite figures added together	• Recognize how unit cubes can be used to measure volume • Recognize that the number of unit cubes in a solid figure has a volume measured in cubic units • Can use improvised units in addition to cubic cm, cubic in, or cubic ft • Understand that base times height is equal to length times width times height • Solve real-world problems of volume and recognize that volume equals base times height • Recognize that volume is additive and apply this to real-world problems	• Prove that using unit cubes can yield an equivalent volume as when using the formula base times height • N/A • Given a specific volume, students may identify a multitude of different possible dimensions • Apply volume formula to real-world problems • Create real-world problems and find a missing dimension • Find the volume of two or more composite figures in which simple shapes may overlap.

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students graph points on the coordinate plane to solve real-world and mathematical problems. (NY-5.G.1-2)	• Plot a coordinate on the coordinate plane • Mathematical problems graphed	• Recognize how coordinates are plotted on the coordinate plane • Apply coordinate plane graphing to real-world problems	• Plot multiple points or sets of coordinates on a coordinate plane • Create real-world problems involving a coordinate plane

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students classify two-dimensional figures into categories based on their properties. (NY-5.G.3-4)	• Use only visual models to classify two-dimensional figures based on attributes • (Same standard as 1)	• Explain attributes of two-dimensional figures without visual models • (Same standard as 1)	• Compare/contrast attributes of two-dimensional figures • (Same standard as 1)

Math Grade 6

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students understand ratio concepts and use ratio reasoning to solve problems. (NY-6.RP.1-3) 1. Understand the concept of a ratio and use ratio language to describe a ratio relationship between two quantities. 2. Understand the concept of a unit rate a/b associated with a ratio $a:b$ with $b \neq 0$ (b not equal to zero), and use rate language in the context of a ratio relationship. 3. Use ratio and rate reasoning to solve real-world and mathematical problems.	• (6.RP.1) Use ratio language to describe ratio relationships • (6.RP.2) Find the unit rate for two quantities in a ratio • (6.RP.3a, b, c, d) Use ratio reasoning to solve for percents and simple unit conversions	• (6.RP.1) Use concept of a ratio to describe ratio relationships • (6.RP.2) Use concept of a unit rate and use rate language • (6.RP.3a, b, c, d) Use ratio and rate reasoning to solve problems involving unit price and speed	• (6.RP.1) • (6.RP.2) Find Unit rates for two or more ratio relationships • (6.RP.3a, b, c, d) Use ratio and rate reasoning to solve problems across measurement systems and connect them with a variety of representations and strategies

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students apply and extend previous understandings of multiplication and division to divide fractions by fractions. (NY-6.NS.1)	(NS.1) • No visual models/manipulatives • Calculate quotients of fractions. • Solve scaffolded word problems involving division of fractions by fractions.	(NS.1) • Interpret and calculate quotients of fractions. • Solve word problems involving division of fractions by fractions. (6.NS.1)	(NS.1) • Create contextual problems involving the quotient of fractions. • Create word problems on division of fractions

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Compute fluently with multi-digit numbers and find common factors and multiples. (NY-6.NS.4)	(NS.4) • Find the **common factors of two whole numbers** less than or equal to 100. Find the **common multiples of two whole numbers** less than or equal to 12.	(NS.4) • Find the **greatest common** factor of two whole numbers less than or equal to 100. Use **the distributive property to express a sum of two whole numbers 1–100 with a common factor as a multiple of a sum of two whole numbers with no common factor other than 1. Find the least common multiple of two whole numbers less than or equal to 12.** (6.NS.4) e.g., Express 36 + 8 as 4 (9 + 2).	(NS.4) • Find the greatest common factor of two or more whole numbers **of any number**. Use the distributive property to express a sum of two whole numbers with a common factor as a multiple of a sum of two whole numbers with no common factor other than 1. Find the least common multiple **of any number** of two or more whole numbers.
Students can apply and extend previous understandings of numbers to the system of rational numbers. (NY-6.NS.5-8)	(NS.5) • Use positive and negative numbers to describe mathematical or **real-world quantities that have opposite values or directions** and **represent them on a number line (horizontal or vertical).**	(NS.5) • Use positive and negative numbers together to describe mathematical or real-world quantities that have opposite values or directions. Represent rational numbers on a number line (horizontal or vertical) and **compare them with or without the use of a number line. Explain the meaning of zero, 0, in a given situation. Explain ordering of rational numbers.**	(NS.5) • **Explain and apply** ordering of rational numbers.

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students can apply and extend previous understandings of numbers to the system of rational numbers. (NY-6.NS.5-8)	(NS.6) • Identify locations of given points in **quadrants** of the coordinate plane. Explore the relationship between the position of two points when **their coordinates differ only by signs.**	(NS.6) • **Use signs** of numbers in ordered pairs as indicating locations in quadrants of the coordinate plane. Recognize that when two ordered pairs differ only by signs, **the locations of the points are related by reflections across one or both axes.**	(NS.6) • Use signs of numbers in ordered pairs as indicating locations in quadrants of the coordinate plane. Recognize that when two ordered pairs differ only by signs, the locations of the points are **related by reflections**.
	(NS.7a,b) • Order and compare rational numbers.	(NS.7a,b) • **Write, interpret, and explain** • **Interpret** statements of inequality as statements about the relative position of two numbers on a number line.	(NS.7a,b) • **Apply** ordering to a set of rational numbers from a real-world situation.
	(NS.7c) • Understand the absolute value of **a rational number**	(NS.7c) • Interpret absolute value as magnitude for +/- quantity in a real-world situation.	(NS.7c) • N/A
	(NS.7d) Order and compare the absolute values of **rational numbers.**	(NS.7d) **Distinguish comparisons of absolute value from statements about order.**	NS.7d) Distinguish comparisons of absolute value from statements about order. **Using a real-world context, create comparison statements of absolute value.**
	(NS.8) Plot ordered pairs on a coordinate plane to solve mathematical problems. Use coordinates and absolute value to find distances between points **located in the same quadrant** that have the same first coordinate or the same second coordinate.	(NS.8) Plot ordered pairs on a coordinate plane **(all quadrants)** to solve **real-world** and mathematical problems. Use coordinates and absolute value to find distances between points with the same first coordinate or the **same second coordinate in all 4 quadrants.**	(NS.8) **Create real-world and mathematical problems** that involve plotting ordered pairs on a coordinate plane.

cv

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students can apply and extend previous understandings of arithmetic to algebraic expressions. (NY-6.EE.1-4)	6.EE.1 6.EE.2a -adds algebraic expressions not just numerical expressions 6.EE.2b -N/A 6.EE.2c - Applications to real world problems - No parentheses 6.EE.3, 6.EE.4 - Identify equivalent expressions	6.EE.1 6.EE.2a - writing with expressions which letters stand for numbers 6.EE.2b - adds algebraic expressions not just numerical expressions - View 1 or more parts of an expression as a single entity 6.EE.2c - May or may not contain parentheses - Order of Operations - Evaluate Expressions 6.EE.3, 6.EE.4 - Generate and apply - Use properties of operations	6.EE.1 6.EE.2a - Add nested grouping 6.EE.2b -N/A 6.EE.2c - Evaluate and Create Expressions - Could contain nested grouping symbols 6.EE.3, 6.EE.4 - Explain and rewrite expressions - Real World problems - Explain how quantities are related

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students can reason about and solve one-variable equations and inequalities. (NY-6.EE.5-8)	6.EE.5 - Add inequalities 6.EE.6 N/A 6.EE.7 - Equations with nonnegative rational numbers 6.EE.8 - write inequalities to represent a constraint in mathematical problems	6.EE.5 - Solve equation or inequality to determine if its true 6.EE.6 - Solving real world single step equations 6.EE.7 - Write and Solve real world problems using single step equations across all operations 6.EE.8 - Represent solutions of inequalities on a number line	6.EE.5 N/A 6.EE.6 N/A 6.EE.7 - Solve two step equations 6.EE.8 N/A

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students can represent and analyze quantitative relationships between dependent and independent variables. (NY-6.EE.9)	6.EE.9 - Determine the relationship between dependent and independent variable using graphs and tables	6.EE.9 - Real world problems - Analyzing the relationship between dependent and independent	6.EE.9 - Write the equation

Cluster	Threshold Level 2	Threshold Level 3	Threshold Level 4
Solve real-world and mathematical problems involving area, surface area and volume. (NY-6.G.1-5) 1. Find the area of triangles, trapezoids, and other polygons by composing into rectangles or decomposing into triangles and quadrilaterals. Apply these techniques in the context of solving real-world and mathematical problems. 2. Find volumes of right rectangular prisms with fractional edge lengths in the context of solving real-world and mathematical problems. 3. Draw polygons in the coordinate plane given coordinates for the vertices. Use coordinates to find the length of a side joining points with the same first coordinate or the same second coordinate. Apply these techniques in the context of solving real-world and mathematical problems. 4. represent three-dimensional figures using nets made up of rectangles and triangles, and use the nets to find the surface area of these figures. Apply these techniques in the context of solving real-world and mathematical problems. 5. Use area and volume models to explain perfect squares and perfect cubes.	• (6.G.1) Compose and decompose polygons into rectangles, triangles. and quadrilaterals to find area, with a visual aid • (6.G.2) Given a diagram, find the volume of a right rectangular prism with fractional edge lengths • (6.G.3) Draw polygons on a coordinate plane using multiple quadrants • (6.G.4) Find the surface area of a net composed of rectangles and triangles • (6.G.5) Use area and volume models to calculate perfect squares and cubes	• (6.G.1) Compose and decompose polygons into rectangles, triangles. and quadrilaterals to find area, without a visual aid, and apply these techniques to real-world problems • (6.G.2)Find the volume of a right rectangular prism with fractional edge lengths in the context of a real-world problem • (6.G.3) Draw polygons on a coordinate plane using multiple quadrants and apply it to solve real-world problems • (6.G.4) Represent 3d figures with nets, using rectangles and triangles, and apply these techniques to solve real-world problems • (6.G.5) Explain perfect squares and cubes using area and volume models	• (6.G.1) • (6.G.2) Find and apply volume of a right rectangular prism with fractional edge lengths • (6.G.3) • (6.G.4) • (6.G.5) Explain perfect squares and cubes in the context of real-world problems

Math Grade 7

Cluster - Table 1	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students analyze proportional relationships and use them to solve real-world and mathematical problems. (NY-7.RP.1-3)	• Calculate Unit Rate with at least one of the following ○ Ratios of fractions ○ Given a graph ○ Given verbal description • Excluding mixed numbers of at least one fraction description OR • Determine whether two quantities are in a proportional relationship with at least one of the following: ○ Given a graph ○ Given a table ○ Given verbal description OR • Identify constant of proportionality in ○ real world or ○ mathematical relationships OR • Verbal description using an equation OR • Identify point (0,0) representing the initial value OR • Identify unit rate using a point OR • Finding whole given part/percent	• Computes unit rates with ratios of fractions with at least one of the following ○ Like or different units ○ Across measurement systems OR • Determine whether two quantities are in a proportional relationship with at least one of the following: ○ Given an equation ○ Given a diagram OR • Identify constant of proportionality ○ verbal description of proportional relationships OR • Represent a proportional relationship using an equation OR • Explain what point (x,y) on the graph of a proportional relationship. OR • Use proportional relationships to solve multi-step ratio and percent problems	• Apply fractional ratios to describe and/or compare rates. OR • Explain whether two quantities are in a proportional relationship using multiple representations OR • Interpret the points (0,0) and (1, r) where r is the unit rate OR • Explain what point (x,y) on the graph of a proportional relationship means OR • Analyze and use proportional relationship to solve multi step real world and mathematical problems using ratio and/or percentages

Cluster - Table 2	Threshold Level 2	Threshold Level 3	Threshold Level 4
Apply and extend previous understanding of operations with fractions to add, subtract, multiply, and divide rational numbers. (NY-7.NS.1-3)	• not restricted to numbers (integers) between 10 and -10 • when a visual is not provided students can: 1. Apply properties of operations to rational numbers (add, multiply, divide) 2. Determine the additive inverse • Solve Two-Step real world problems with rational numbers • Convert fraction to decimal w/ repeating decimal	• majority of real world context • Solve multi-Step real world problems • apply properties of operations (mainly add, mult., divide) with rational numbers	• Identify and apply properties of operations • Create a real world problem • Justify the steps for the solution

Cluster - Table 3	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students use properties of operations to generate equivalent expressions. (NY-7.EE.1-2)	• Use properties to add and subtract. • Manipulate negative integers. • Expand linear expressions.	• Use rational coefficients • Factor • Real World Context • Explanation	• Identify Properties • Describe relationship of equivalent quantities • Explain equivalence

Cluster – Table 3	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students solve real-life and mathematical problems using numerical and algebraic expressions, equations, and inequalities.↔ NY-7.EE.3-4)	• Real world • Apply properties • Using variables, create linear expressions • Using variables, create linear inequalities • Determine the solution of an equation • Determine the solution of the inequality • Solving a 2-step inequality	• Solve Multi-step equation • Rational numbers • Apply properties of rational numbers • Convert numbers to different forms • Decide reasonableness • Solve Multi Step inequality • Identify the sequence of operations • Real World Problems • addition of $\leq$ and $\geq$ • Interpret the solution	• Estimate solution to real world problems • Check reasonableness of estimated solution • Explain the relationship between the steps used to solve a given equation • Explain the relationship between the steps used to solve a given inequality

Cluster - Table 1	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students draw, construct, and describe geometrical figures and describe the relationship between them. (NY-7.G.1-3)	• Solve a one-step problem involving scale factor OR • Classify a triangle based on angle size (right or acute or obtuse) OR • Identify the two-dimensional shape that results from slicing any solid other than a right rectangular pyramid parallel to the base.	• Compute actual area from a scale drawing OR • Reproduce a scale drawing at a different scale OR • When given measures of angles or sides, determine when the conditions determine: ○ a unique triangle ○ more than one triangle ○ no triangle OR • Describe the two-dimensional shape that results from slicing a right rectangular prism or right rectangular pyramid either: ○ parallel to the base ○ perpendicular to the base	• Recognize that a proportional relationship shows similarity OR • Explain the relationship between a scale factor and the volume of a three-dimensional object OR • Describe the two-dimensional shape that results from slicing a right rectangular prism or right rectangular pyramid in any direction.

Cluster - Table 3	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students solve real-life and mathematical problems involving angle measure, area, surface area, and volume. (NY-7.G.4-6)	• Apply the formula of circumference and area of a circle • Calculate the measures of unknown angles • Real world problems of composite 2-d figures made of triangles/rectangles • Solve real world surface area problems • Solve real world volume problems • Volume of triangular prism	• Given the circumference of a circle, calculate the radius and diameter. • Write and solve multi step equations for unknown angles. • Solve real world area of 2-dimensional objects composed of trapezoids. • Solve real world surface area of 2-dimensional objects composed of trapezoids. • Solve volume problems of 3-d objects composed of right rectangular prisms.	*Find the radius given area of a circle in terms of pi *Solving measures of unknown angles with variables on both sides *These are not 7th grade standards. • Derivation of volume of a cylinder formula • Use volume of a cylinder formula.

Cluster - Table 1	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students draw informal comparative inferences about two populations. (NY-7.SP.1,3,4)	• Identify the interquartile range of the given data set OR • Distinguish the similarities and differences between two data sets. OR • Apply the differences between the measures of center and variability to make informal inferences about a given set of data	• Construct and interpret box-plots, find the interquartile range and determine if a data point is an outlier. OR • Describe a visual comparison of visual overlap of two quantitative data distributions. OR • Draw informal comparative inference about the populations using measures of center, measures of variability from random samples.	• Explain how an outlier affects a set of data OR • Assess how an outlier affects the data set OR • Draw comparative inferences about populations including how an outlier affects the data set using measures of center, measures of variability from random samples.

Cluster - Table 2	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students investigate chance processes and develop, use, and evaluate probability models. (NY-7.SP.8)	• can recognize a compound event instead of a simple event (level 1) • given sample space- multiple choice with options listed can you interpret data • Calculate probability rather than identify	• Students need to figure out what the sample space looks like. • Students need to organize and interpret all of the possible outcomes from a sample space	• students must interpret based on results of a simulation • could be explaining in written response

Math Grade 8

Cluster - Table 1	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students know there are numbers that are not rational and approximate them by rational numbers. (NY-8.NS.1-2)	Decimal expansion repeats Approximate common irrational numbers on a number line • pi • square root of a rational number	Understand informally that every rational number has a: • decimal expansion • repeats What is not rational is called irrational. Rational approximation of irrational numbers and estimate the value of expressions. • compare size of irrational numbers	Distinguish between real numbers and non-real or imaginary numbers.

Cluster: Table 2	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students work with radicals and integer exponents. (NY-8.EE.1-4)	Apply one property of exponents to "real world context" positive exponents only Solve equations of perfect cubes, up to 125 (in addition to solving equations with perfect squares up to 225, which is PL1). Could be solved by properties of exponents or by guess-and-check, not necessarily use or understand inverse/root symbols.	Apply more than one exponent property OR Apply exponent property with neg. exponents to generate "equivalent expressions" A non-perfect square is called irrational Know squares up to 225, cubes up to 125. Know/use "square root" and "cube root" symbols to solve.	Apply multiple exponent properties with negative exponents OR real-world context

Cluster - Table 3	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students understand the connections between proportional relationships, lines, and linear equations. (NY-8.EE.5-6)	• identify unit rate as slope • given the equation, identify the graph • Given the graph of a line: derive the equation of a line that passes through the the vertical axis (y = mx and y = mx+b)	• graph and interpret slope • compare proportional relationships (slope) in different ways • explain slope using similar triangles and derive the equation of a line in the form y=mx and y = mx + b	• explain the slope in the context of the problem • Derive the equation of a line given 2 points that has a y-intercept and an x-intercept

Cluster - Table 1	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students analyze and solve linear equations and pairs of simultaneous linear equations. (NY-8.EE.7-8•)	Transform a given equation into a simpler equivalent equation by inspection. Solve linear equations in one variable with integer number coefficients • including those that require distributive property and/or combining like terms.	Recognize when a linear equation in one variable has • one solution • no solution • infinitely many solutions Solve linear equations in one variable with rational number coefficients • including those that require distributive property and/or combining like terms.	Solve "multi-layered" equations algebraically -This means multi-step and higher level situations per NYSED

Cluster - Table 2	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students define, evaluate, and compare functions.ᵩ(8.F.1-3)	Define a function as a rule that assigns each input to exactly one output. Recognize a graph and/or table of a function is a set of ordered pairs with an input and corresponding output Identify properties of linear (x), quadratic (x^2), cubic (x^3) functions given " y = …" from any representation (algebraic, graphical, numerical, OR verbal). Identify non-linear functions from a graph OR a table.	Understand each input has exactly one output Compare two functions represented in different ways. Know that y = mx + b makes a linear function, with a straight line graph. Recognize linear and nonlinear functions in the same form (table, graph, equation).	domain/range corresponds to input/output know and solve real life applications Compare more than two functions represented in different ways. Know the difference between linear and nonlinear functions when shown in different forms (table, graph, verbal, equation).

Cluster - Table 3	Threshold Level 2	Threshold Level 3	Threshold Level 4
Use functions to model relationships between quantities. (NY-8.F.4-5)	• Find initial value (y-intercept) and/or rate of change (slope) given a table **and/or** • Find initial value (y-intercept) and/or rate of change rate of change (slope) given 2 points (x,y) • Analyze the graph of a function	• construct a linear function • determine the slope and y-intercept of a function from a description, two points, table or graph • create a real world graph that represents a function • Analyze the graph of a function using two quantities	

Cluster - Table 3	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students understand congruence and similarity using physical models, transparencies, or geometry software. (NY-8.G.1-5)	• Verify the congruence of line segments on a coordinate plane after rotation, reflection, translation • Verify the congruence of angles on a coordinate plane after rotation, reflection, translation • Verify the congruence of parallel lines on a coordinate plane after rotation, reflection, translation • recognize that a two-dimensional figure is congruent to another if the corresponding angles are congruent and the corresponding sides are congruent • recognize that a two-dimensional figure is congruent to another after a sequence of rotations, reflections, translations • Recognize that size/shape does not change in translations, reflections, and translations. • Recognize that orientation changes with line reflections • recognize that a two-dimensional figure is congruent to another after a sequence of rotations, reflections, translation, and/or dilation • Determine the measurements of angles formed by two parallel lines cut by a transversal	• Verify through experiment: ○ the congruence of line segments on a coordinate plane after rotation, reflection, translation ○ the congruence of angles on a coordinate plane after rotation, reflection, translation ○ the congruence of parallel lines on a coordinate plane after rotation, reflection, translation • describe sequence of transformations (rotation, refection, translation) • describe the effect of dilations, translations, rotations, and reflections on two-dimensional figures using coordinates. • Know the definition of similarity • Use informal arguments to establish: ○ facts about angle sum and exterior angle of triangles ○ angles created when parallel lines are cut by a transversal ○ angle/angle criterion for similar triangles	• Describe sequence of transformations using coordinates • Determine the measurements of angles formed by 2 parallel lines cut by two transversals

Cluster- Table 2	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students understand and apply the Pythagorean Theorem. (NY-8.G.6-8)	Recognize that if $a^2 + b^2 = c^2$ works, it is a right triangle (or it is not a right triangle if the pythagorean theorem doesn't work.) given two sides of a right triangle students will solve for the missing side length using Pythagorean Theorem find the length of a leg segment for a triangle or the diagonal of a rectangle on the coordinate plane using PT, if you're NOT given a right triangle.	Understand a proof of the pythagorean theorem and its converse. Apply PT to solve for a missing side in real-world problems, and/or in three dimensions. Find distance between any two coordinate points using PT.	explain a proof of the pythagorean theorem Use PT to solve and model multi-step problems including cones, diag. of rectangular prisms, etc. Use PT to derive the distance formula

Cluster - Table 1	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students solve real-world and mathematical problems involving volume of cylinders, cones, and spheres. (NY-8.G.9)	Given real world context, select the appropriate equations for problems involving cones, cylinders and/or spheres.	Solve mathematical and real world volume problems.	Derive the formulas and show the relationship among them.

Cluster - Table 1	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students investigate patterns of association in bivariate data. (NY-8.SP.1-3)	Identify a scatter plot from a set of bivariate data. Describe patterns: • Outliers • positive or negative association • linear or nonlinear association	Construct and interpret scatter plots for bivariate measurement data	Analyze patterns of association between two quantities and use data to make and justify predictions.
	For scatter plots that suggest a linear association, informally fit a straight line.	Informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.	Determine the equation for a line of best fit.
	Identify the slope and intercept; use the equation in context of bivariate data.	Interpret the slope and intercept; use the equation in context of bivariate data.	Use the equation to make and justify predictions.

Cluster - Table 1 Grade 7)	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students draw, construct, and describe geometrical figures and describe the relationship between them. (NY-7.G.1-3)	• Solve a one-step problem involving scale factor OR • Classify a triangle based on angle size (right or acute or obtuse) OR • Identify the two-dimensional shape that results from slicing any solid other than a right rectangular pyramid parallel to the base.	• Compute actual area from a scale drawing OR • Reproduce a scale drawing at a different scale OR • When given measures of angles or sides, determine when the conditions determine: ○ a unique triangle ○ more than one triangle ○ no triangle OR • Describe the two-dimensional shape that results from slicing a right rectangular prism or right rectangular pyramid either: ○ parallel to the base ○ perpendicular to the base	• Recognize that a proportional relationship shows similarity OR • Explain the relationship between a scale factor and the volume of a three-dimensional object OR • Describe the two-dimensional shape that results from slicing a right rectangular prism or right rectangular pyramid in any direction.

Cluster - Table 3 (Post Test From Grade 7)	Threshold Level 2	Threshold Level 3	Threshold Level 4
Students solve real-life and mathematical problems involving angle measure, area, surface area, and volume. (NY-7.G.4-6)	• Apply the formula of circumference and area of a circle • Calculate the measures of unknown angles • Real world problems of composite 2-d figures made of triangles/rectangles • Solve real world surface area problems • Solve real world volume problems • Volume of triangular prism	• Given the circumference of a circle, calculate the radius and diameter. • Write and solve multi step equations for unknown angles. • Solve real world area of 2-dimensional objects composed of trapezoids. • Solve real world surface area of 2-dimensional objects composed of trapezoids. • Solve volume problems of 3-d objects composed of right rectangular prisms.	• Derivation of volume of a cylinder formula • Use volume of a cylinder formula.

Appendix D: Panelist Rating Form

NEW YORK STATE
EDUCATION DEPARTMENT
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023
Rating Sheet

NAME: _____

CONTENT: _____ GRADE LEVEL: _____

Instructions for Determining your Bookmarks:

- Review the booklet of items
- You will be recommending three cut scores:
  - Level 2
  - Level 3
  - Level 4
- For each item, ask yourself whether 2/3 of students demonstrating Level 2 threshold performance would be able to answer the multiple-choice item correctly or achieve that score or higher on a constructed-response item. Your Bookmark is the first item where your answer is "No" to this question. **Indicate your Bookmark placement by writing down the sequence number that represents your first *No* response.**
- Continue the process for placing Level 3 and Level 4 Bookmarks and indicate your ratings in the same manner as you did for Level 2. ***Practice will only be for Level 3.***

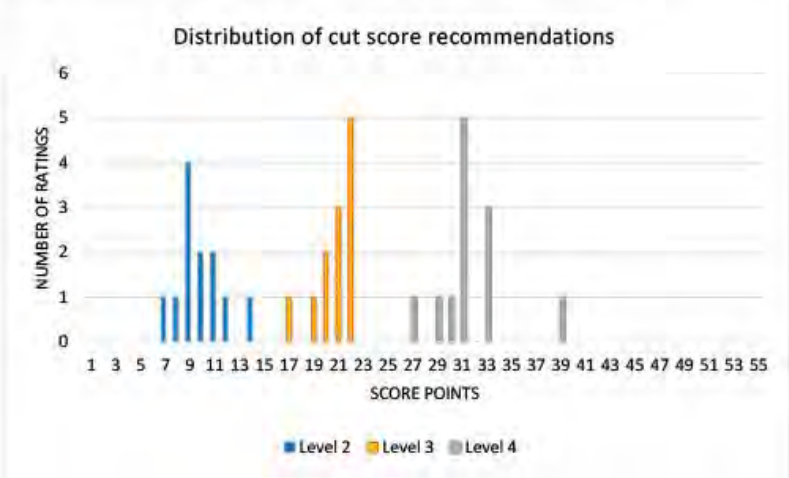
	OIB Sequence Number			
	Practice	**Round 1**	**Round 2**	**Round 3**
Level 2 Bookmark:				
Level 3 Bookmark:				
Level 4 Bookmark:				

Appendix E: Samples of feedback provided after each round of ratings

Sample 1: Summary of ratings feedback; provided after Round 1 and Round 2

# Panelists		12	
	Min	Max	Median
Level 1			
Level 2	7	14	9.5
Level 3	17	22	21.0
Level 4	27	39	31.0

Sample 2: Distribution of ratings received; provided after Rounds 1 and 2

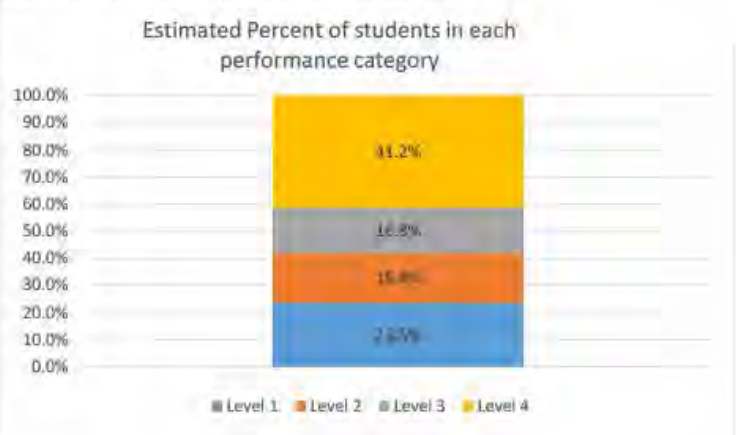


Sample 3: Summary of ratings plus impact data; provided after Rounds 2 and 3

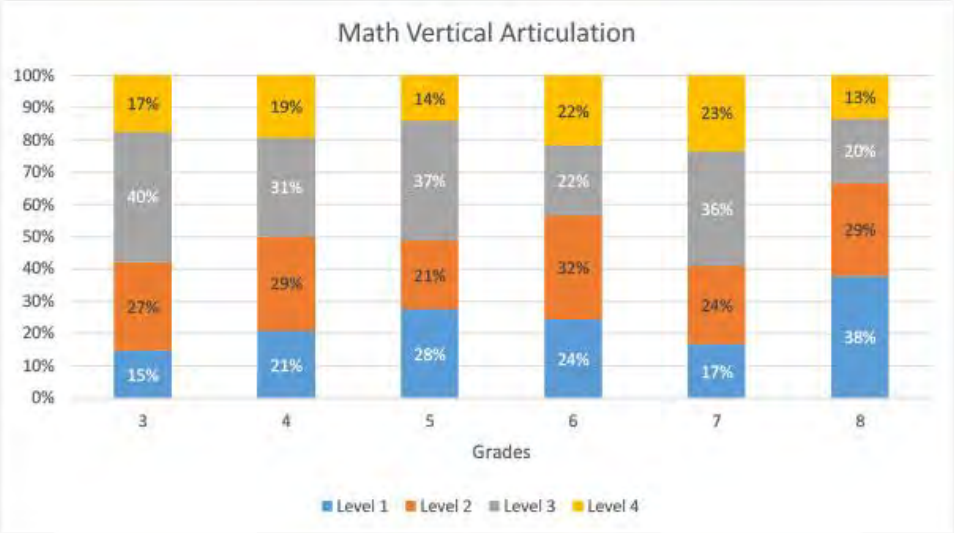
panelists 12

	Min	Max	Median	% students
Level 1				24%
Level 2	7	14	10	18%
Level 3	17	22	21	17%
Level 4	27	39	31	41%

Sample 4: Impact Data; provided after Rounds 2 and 3



Appendix F: Example of impact data presented during vertical articulation



Appendix G: Summary of cut score recommendations for Rounds 1 and 2

Table G1: Round 1 cut score recommendations for ELA assessments

Round 1 ELA recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	39.4%	31.7%	22.2%	35.8%	25.6%	18.7%
Level 2	Page #	12	9	10	13	10	12
	Theta	-0.316	-0.508	-0.762	-0.34	-0.711	-0.892
	Raw score cut	19	18	18	24	24	23
	% students	29.9%	27.8%	46.3%	28.6%	38.5%	25.5%
Level 3	Page #	29	25	30	27	32	27
	Theta cut point	0.663	0.3	0.501	0.485	0.37	-0.084
	Raw score cut	26	25	29	30	35	32
	% students	15.6%	25.8%	21.0%	26.7%	21.7%	43.3%
Level 4	Page #	37	40	42	40	51	45
	Theta cut point	1.212	1.147	1.483	1.551	1.279	1.19
	Raw score cut	29	31	34	35	40	41
	% students	15.2%	14.7%	10.5%	8.9%	14.3%	12.5%

Table G2: Round 2 cut score recommendations for ELA assessments

Final ELA recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	35.9%	31.7%	29.2%	35.8%	25.6%	16.8%
Level 2	Page #	11	9	13	13	10	10
	Theta	-0.395	-0.508	-0.531	-0.34	-0.711	-0.995
	Raw score cut	18	18	20	24	24	22
	% students	33.3%	27.8%	34.6%	28.6%	38.5%	27.4%
Level 3	Page #	29	24	28	27	32	27
	Theta cut point	0.663	0.23	0.4	0.485	0.37	-0.084
	Raw score cut	26	25	28	30	35	32
	% students	15.6%	21.7%	28.8%	26.7%	21.7%	33.6%
Level 4	Page #	37	38	43	40	51	41.5
	Theta cut point	1.212	1.06	1.57	1.551	1.279	0.888
	Raw score cut	29	30	35	34	40	39
	% students	15.2%	18.8%	7.3%	8.9%	14.3%	22.2%

Table G3: Round 1 cut score recommendations for Math assessments

Round 1 ELA recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	14.8%	13.8%	27.7%	40.9%	21.2%	18.7%
Level 2	Page #	7	5	10	9	8	8
	Theta	-1.049	-1.098	-0.648	-0.299	-0.855	-0.3
	Raw score cut	13	13	17	21	18	24
	% students	27.5%	28.2%	26.3%	13.5%	24.4%	25.5%
Level 3	Page #	25	20	26	21	22	20
	Theta cut point	-0.267	0.254	0.059	0.067	-0.22	0.485
	Raw score cut	23	24	27	27	29	30
	% students	40.4%	42.0%	29.2%	35.9%	33.9%	43.3%
Level 4	Page #	46	49	48	53	48	50
	Theta cut point	1.19	1.109	1.033	1.47	0.843	1.551
	Raw score cut	34	38	38	42	42	35
	% students	17.4%	16.1%	16.9%	9.7%	20.5%	12.5%

Table G4: Round 2 cut score recommendations for Math assessments

Final ELA recommendations		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Level 1	% students	14.8%	21.0%	27.7%	40.9%	19.1%	38.0%
Level 2	Page #	7	8	10	9	7	6
	Theta	-1.049	-0.853	-0.648	-0.299	-0.928	-0.375
	Raw score cut	13	16	17	21	17	19
	% students	27.5%	34.9%	21.0%	13.5%	28.8%	18.8%
Level 3	Page #	25	29	23	21	23	16
	Theta cut point	-0.267	0.172	-0.105	0.067	-0.158	0.102
	Raw score cut	23	29	25	27	30	26
	% students	40.4%	28.0%	28.9%	35.9%	31.6%	31.7%
Level 4	Page #	46	49	46	53	48	46
	Theta cut point	1.19	1.109	0.885	1.47	0.843	1.256
	Raw score cut	34	38	36	42	42	40
	% students	17.4%	16.1%	22.4%	9.7%	20.5%	11.6%

Appendix H: Evaluation surveys results for all grades

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #1 ELA Grade 3

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** ***Thank you!***

Content Area: _____ Grade Level: _____

Orientation, Test Review, and PLDs

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. I understood the purpose of this workshop.	0	0	0	5	6
	2. I understood the content measured by the assessment I reviewed.	0	0	0	7	4
	3. I understood how the assessment was administered.	0	0	0	5	6
	4. I understood the difference between the PLDs and the description of students demonstrating threshold performance.	0	0	0	6	5
	5. The PLDs were clear enough for me to describe the students demonstrating threshold performance.	0	0	0	5	6
	6. The instructions provided during the opening training session were clear.	0	0	0	9	2
	7. The instructions provided by the facilitator were clear.	1	0	2	7	1
	8. After the practice exercise, I understood how to use the bookmark method.	0	0	2	5	4
	9. Overall, I believe my opinions were considered and valued by my group.	0	0	0	4	7

	Please indicate your opinion regarding how much time you were given to complete the following activities. Please bubble *only one* of the three options for each activity.	Too Little Time	About Right	Too Much Time
Time	10. Reviewing the test.	0	10	1
	11. Reviewing the PLDs.	0	11	0
	12. Describing the students demonstrating threshold performance.	0	11	0

In the space provided below, please feel free to add comments about any of your responses. ***Thank you!***

- Setting the threshold based on performance level descriptions was a great way to communicate with committee and norm our review.
- Overall, the purpose and directions for the three days and for today's work were very clear. Thank you.
- Directions and practice was clear. Toggling between Questar app for tests and google drive was tricky at first.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #1 ELA Grade 5

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator. *Thank you!***

Content Area: _____ Grade Level: _____

Orientation, Test Review, and PLDs

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. I understood the purpose of this workshop.					
	2. I understood the content measured by the assessment I reviewed.					
	3. I understood how the assessment was administered.					
	4. I understood the difference between the PLDs and the description of students demonstrating threshold performance.					
	5. The PLDs were clear enough for me to describe the students demonstrating threshold performance.					
	6. The instructions provided during the opening training session were clear.					
	7. The instructions provided by the facilitator were clear.					
	8. After the practice exercise, I understood how to use the bookmark method.					
	9. Overall, I believe my opinions were considered and valued by my group.					

	Please indicate your opinion regarding <u>how much time</u> you were given to complete the following activities. Please bubble *only one* of the three options for each activity.	Too Little Time	About Right	Too Much Time
Time	10. Reviewing the test.			
	11. Reviewing the PLDs.			
	12. Describing the students demonstrating threshold performance.			

In the space provided below, please feel free to add comments about any of your responses. ***Thank you!***

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #1 ELA Grade 7

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator. *Thank you!***

Content Area: _____ Grade Level: _____

Orientation, Test Review, and PLDs

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. I understood the purpose of this workshop.	0	0	1	3	6
	2. I understood the content measured by the assessment I reviewed.	0	0	1	4	5
	3. I understood how the assessment was administered.	0	0	2	4	4
	4. I understood the difference between the PLDs and the description of students demonstrating threshold performance.	0	0	1	4	5
	5. The PLDs were clear enough for me to describe the students demonstrating threshold performance.	0	1	0	4	5
	6. The instructions provided during the opening training session were clear.	0	0	1	3	6
	7. The instructions provided by the facilitator were clear.	0	1	1	3	5
	8. After the practice exercise, I understood how to use the bookmark method.	2	0	1	4	3
	9. Overall, I believe my opinions were considered and valued by my group.	0	0	1	3	6

	Please indicate your opinion regarding how much time you were given to complete the following activities. Please bubble *only one* of the three options for each activity.	Too Little Time	About Right	Too Much Time
Time	10. Reviewing the test.	1	9	0
	11. Reviewing the PLDs.	1	8	1
	12. Describing the students demonstrating threshold performance.	1	9	0

In the space provided below, please feel free to add comments about any of your responses. ***Thank you!***

- Scott was so patient and thorough! Thanks!
- Formatting should be changed for creating PLD threshold - go form 4-2 rather 2-4 to maintain consistency of the rubric used.
- Scott Russell was very professional, helpful, and friendly. He was a wonderful facilitator.
- I would've liked to have had more information about what is expected BEFORE the workshop.
- I was pretty confused about the bookmarks, but I think I get it now. I do think our facilitator tried his best and was very patient with us. I would like to better understand why we are setting the standards after giving the assessment rather than before. Whenever I administer a test, I determine this information using rubrics and keys prior to giving the test. I do understand that standardized tests are different, but I am struggling to understand why we are doing this after the fact. Again, I'm sure there's a good reason; I just don't know what that is.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #1 Math Grade 3

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator. *Thank you!***

Content Area: _____ Grade Level: _____

Orientation, Test Review, and PLDs

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. I understood the purpose of this workshop.	0	0	0	5	7
	2. I understood the content measured by the assessment I reviewed.	0	0	0	4	8
	3. I understood how the assessment was administered.	0	0	0	4	8
	4. I understood the difference between the PLDs and the description of students demonstrating threshold performance.	0	0	1	7	4
	5. The PLDs were clear enough for me to describe the students demonstrating threshold performance.	0	0	0	7	5
	6. The instructions provided during the opening training session were clear.	0	1	2	5	4
	7. The instructions provided by the facilitator were clear.	0	0	2	6	4
	8. After the practice exercise, I understood how to use the bookmark method.	0	0	1	7	4
	9. Overall, I believe my opinions were considered and valued by my group.	0	0	1	5	6

	Please indicate your opinion regarding how much time you were given to complete the following activities. Please bubble *only one* of the three options for each activity.	Too Little Time	About Right	Too Much Time
Time	10. Reviewing the test.	0	12	0
	11. Reviewing the PLDs.	1	11	0
	12. Describing the students demonstrating threshold performance.	2	10	0

In the space provided below, please feel free to add comments about any of your responses. ***Thank you!***

- The pacing is perfect and the experience is enjoyable.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #1 Math Grade 5

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator. *Thank you!***

Content Area: _____ Grade Level: _____

Orientation, Test Review, and PLDs

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. I understood the purpose of this workshop.	0	0	0	7	3
	2. I understood the content measured by the assessment I reviewed.	0	0	0	5	5
	3. I understood how the assessment was administered.	0	0	2	1	7
	4. I understood the difference between the PLDs and the description of students demonstrating threshold performance.	0	0	2	6	2
	5. The PLDs were clear enough for me to describe the students demonstrating threshold performance.	0	0	0	9	1
	6. The instructions provided during the opening training session were clear.	0	0	1	3	6
	7. The instructions provided by the facilitator were clear.	0	0	0	4	6
	8. After the practice exercise, I understood how to use the bookmark method.	0	0	0	6	4
	9. Overall, I believe my opinions were considered and valued by my group.	0	0	0	4	6

	Please indicate your opinion regarding how much time you were given to complete the following activities. Please bubble *only one* of the three options for each activity.	Too Little Time	About Right	Too Much Time
Time	10. Reviewing the test.	0	10	0
	11. Reviewing the PLDs.	0	9	1
	12. Describing the students demonstrating threshold performance.	1	9	0

In the space provided below, please feel free to add comments about any of your responses. ***Thank you!***

- Great discussions between one another.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #1 Math Grade 7

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator. *Thank you!***

Content Area: _____ Grade Level: _____

Orientation, Test Review, and PLDs

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. I understood the purpose of this workshop.	0	0	0	0	10
	2. I understood the content measured by the assessment I reviewed.	0	0	0	1	9
	3. I understood how the assessment was administered.	0	0	0	1	9
	4. I understood the difference between the PLDs and the description of students demonstrating threshold performance.	0	0	1	3	6
	5. The PLDs were clear enough for me to describe the students demonstrating threshold performance.	0	0	1	5	4
	6. The instructions provided during the opening training session were clear.	0	0	0	2	8
	7. The instructions provided by the facilitator were clear.	0	0	0	3	7
	8. After the practice exercise, I understood how to use the bookmark method.	0	0	0	0	10
	9. Overall, I believe my opinions were considered and valued by my group.	0	0	1	1	8

	Please indicate your opinion regarding <u>how much time</u> you were given to complete the following activities. Please bubble *only one* of the three options for each activity.	Too Little Time	About Right	Too Much Time
Time	10. Reviewing the test.	2	8	0
	11. Reviewing the PLDs.	1	9	0
	12. Describing the students demonstrating threshold performance.	0	10	0

In the space provided below, please feel free to add comments about any of your responses. ***Thank you!***

- It is helpful when the facilitator tells us the end goal of the task before we begin, and also the timeframe that we have to complete the task.
- Table works well together and clear instruction!
- I thought there was adequate time for collaboration. My table worked very well together and our instructor was clear and patient with helping us understand directions and material.
- This is an excellent PD and collaborative time to work with other educators.
- It was great working with other educators from various location who brought their experiences and perspectives.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 ELA Grade 3

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	0	3	7	1
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	1	6	4
	3. I understood the impact data that were presented after Round 2.	0	0	1	5	5
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	0	0	4	7
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	4	7

	Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
Influence	6. The Performance Level Descriptions (PLDs).	2	3	3	3
	7. The descriptions of students demonstrating threshold performance.	0	3	5	3
	8. My perception of the difficulty of the items.	1	1	8	1
	9. My experiences with students.	0	1	5	5
	10. Discussion within my group.	0	2	4	5
	11. The item ratings of other participants.	0	5	5	1
	12. The percent of students in each performance level (the impact data).	0	5	5	1
	13. My sense of what a student needs to know to be identified at Level 2.	0	1	9	1
	14. My sense of what a student needs to know to be identified at Level 3.	0	1	9	1
	15. My sense of what a student needs to know to be identified at Level 4.	0	1	9	1

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	0	11	0
	17. *Level 3* cut score	1	9	1
	18. *Level 4* cut score	0	11	0

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

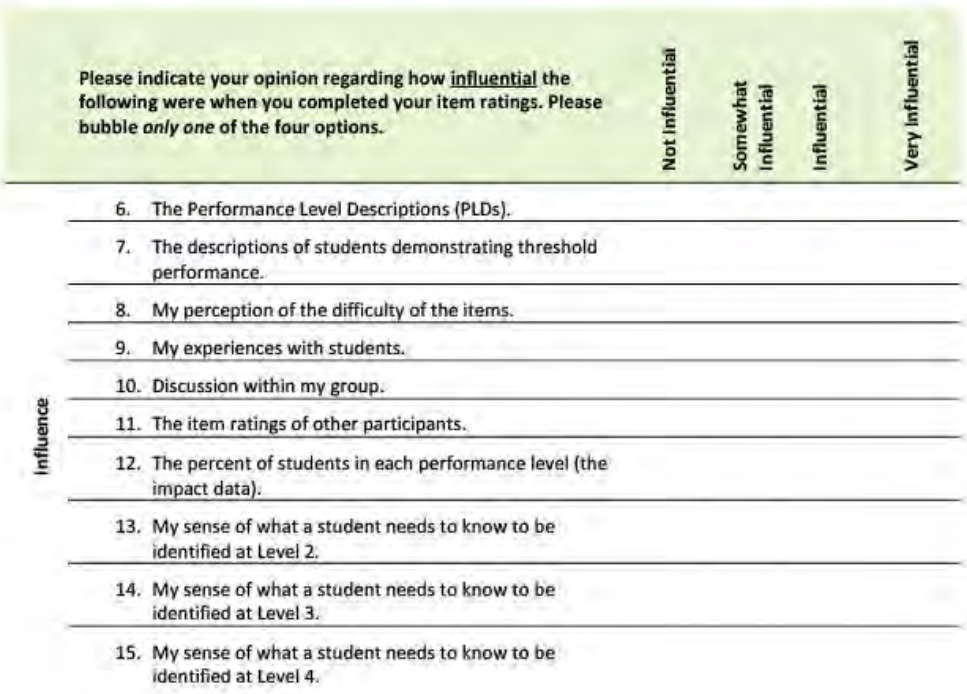
NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 ELA Grade 4

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.					
	2. I understood the cut-score summary data that was presented between the rounds.					
	3. I understood the impact data that were presented after Round 2.					
	4. By the end of round 3, I was comfortable with the item rating procedure.					
	5. Overall, I believe my opinions were considered and valued by my group.					



Influence	Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
	6. The Performance Level Descriptions (PLDs).				
	7. The descriptions of students demonstrating threshold performance.				
	8. My perception of the difficulty of the items.				
	9. My experiences with students.				
	10. Discussion within my group.				
	11. The item ratings of other participants.				
	12. The percent of students in each performance level (the impact data).				
	13. My sense of what a student needs to know to be identified at Level 2.				
	14. My sense of what a student needs to know to be identified at Level 3.				
	15. My sense of what a student needs to know to be identified at Level 4.				

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score			
	17. *Level 3* cut score			
	18. *Level 4* cut score			

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 ELA Grade 5

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	4	1	6	1
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	0	6	6
	3. I understood the impact data that were presented after Round 2.	0	0	0	6	6
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	0	0	6	6
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	5	7

Influence

Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
6. The Performance Level Descriptions (PLDs).	1	1	8	2
7. The descriptions of students demonstrating threshold performance.	2	0	8	2
8. My perception of the difficulty of the items.	0	1	7	4
9. My experiences with students.	0	0	2	10
10. Discussion within my group.	0	1	4	7
11. The item ratings of other participants.	0	6	4	2
12. The percent of students in each performance level (the impact data).	0	3	7	2
13. My sense of what a student needs to know to be identified at Level 2.	0	0	7	5
14. My sense of what a student needs to know to be identified at Level 3.	0	0	7	5
15. My sense of what a student needs to know to be identified at Level 4.	0	0	7	5

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	0	10	2
	17. *Level 3* cut score	1	8	3
	18. *Level 4* cut score	1	6	5

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

- I do not think the PLD were a huge factor in my decision making and we did not need to spend that much time on it.
- I found this process to be extremely informative about what happens that helps determine the scores. It is very beneficial to be able to discuss with colleagues.
- I feel like it would be helpful to take the whole test instead of spend just a half hour. I also feel that in the Ordered Item Booklet felt problematic when both writing and multiple choice questions were mixed together. I feel like it would be helpful to look at reading and then writing. I think schools should do more internal PD with the PLDs. The presenter was excellent and made the process feel accessible and manageable for someone to this work. I would love to have pencils instead of pens for establishing bookmarks for the different items. I am looking forward to the next step in the process.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 ELA Grade 6

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	0	1	2	6
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	0	3	6
	3. I understood the impact data that were presented after Round 2.	0	0	0	2	7
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	0	0	2	7
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	1	8

	Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
Influence	6. The Performance Level Descriptions (PLDs).	0	1	2	6
	7. The descriptions of students demonstrating threshold performance.	0	1	3	5
	8. My perception of the difficulty of the items.	0	1	5	3
	9. My experiences with students.	0	0	4	5
	10. Discussion within my group.	0	0	2	7
	11. The item ratings of other participants.	0	3	2	4
	12. The percent of students in each performance level (the impact data).	0	1	3	5
	13. My sense of what a student needs to know to be identified at Level 2.	0	0	4	5
	14. My sense of what a student needs to know to be identified at Level 3.	0	0	4	5
	15. My sense of what a student needs to know to be identified at Level 4.	0	0	4	5

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	1	6	2
	17. *Level 3* cut score	2	4	3
	18. *Level 4* cut score	0	9	0

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

- This was extremely helpful and I would love to be part of other NYSED events in the future.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 ELA Grade 7

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	0	2	5	3
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	0	2	8
	3. I understood the impact data that were presented after Round 2.	0	0	0	2	8
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	1	0	1	8
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	3	7

Influence

Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
6. The Performance Level Descriptions (PLDs).	0	1	5	4
7. The descriptions of students demonstrating threshold performance.	0	0	4	6
8. My perception of the difficulty of the items.	0	0	5	5
9. My experiences with students.	0	0	5	5
10. Discussion within my group.	0	0	3	7
11. The item ratings of other participants.	1	2	6	1
12. The percent of students in each performance level (the impact data).	0	4	4	2
13. My sense of what a student needs to know to be identified at Level 2.	0	1	3	6
14. My sense of what a student needs to know to be identified at Level 3.	0	1	3	6
15. My sense of what a student needs to know to be identified at Level 4.	0	1	3	6

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	0	9	1
	17. *Level 3* cut score	1	7	2
	18. *Level 4* cut score	2	7	1

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

- Great process.....I understand how important this is now. :)
- I appreciate having printed copies of the PLDs, but printed copies of the texts would be really helpful. It was hard to jump back and forth between screens from the questions. It also added a lot of time having to scroll through to find the different texts that corresponded to different questions. Even with "control F," it wouldn't find words within some of the titles, so it was kind of a pain.
- The item order data should be in the same direction on all of the documents, not some showing 1-4 left to right and others showing 1-4 going right to left. it was confusing and increased the liklihood of errors.
- Thank you for this opportunity. I was surprised by how difficult this procedure is -- the decisions that we make are important, and I appreciate the opportunity to share my experience with others. I also really like the way that the training was conducted. The way in which information was shared really encouraged us to use our professional knowledge and the resources provided to make decisions. It is good to be a part of this process!
- Thank you so much for the opportunity to be a part of this process!

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 ELA Grade 8

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	0	0	2	7
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	0	2	7
	3. I understood the impact data that were presented after Round 2.	0	0	0	2	7
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	0	0	3	6
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	2	7

	Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
Influence	6. The Performance Level Descriptions (PLDs).	0	0	6	3
	7. The descriptions of students demonstrating threshold performance.	0	0	6	3
	8. My perception of the difficulty of the items.	0	0	5	4
	9. My experiences with students.	0	0	2	7
	10. Discussion within my group.	0	0	3	6
	11. The item ratings of other participants.	0	2	4	3
	12. The percent of students in each performance level (the impact data).	0	2	4	3
	13. My sense of what a student needs to know to be identified at Level 2.	0	0	2	7
	14. My sense of what a student needs to know to be identified at Level 3.	0	0	2	7
	15. My sense of what a student needs to know to be identified at Level 4.	0	0	2	7

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	0	9	0
	17. *Level 3* cut score	1	8	0
	18. *Level 4* cut score	0	9	0

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

- Our facilator, Scott, was fantastic!!!
- This was a highly valuable experience - I learned a lot and feel like this can help inform my teaching.
- Thank you, NWEA, for the lovely clean hotel, and meals. I learned a lot from this experience and look forward to encouraging my colleagues to come to other events in the future. Scott Russell was a great facilitator.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 Math Grade 3

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	1	2	5	4
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	1	5	6
	3. I understood the impact data that were presented after Round 2.	0	0	0	4	8
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	0	0	4	8
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	3	9

	Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
Influence	6. The Performance Level Descriptions (PLDs).	0	0	5	7
	7. The descriptions of students demonstrating threshold performance.	1	4	3	4
	8. My perception of the difficulty of the items.	0	1	7	4
	9. My experiences with students.	0	1	3	8
	10. Discussion within my group.	0	0	3	9
	11. The item ratings of other participants.	1	4	5	2
	12. The percent of students in each performance level (the impact data).	2	4	5	1
	13. My sense of what a student needs to know to be identified at Level 2.	0	0	7	5
	14. My sense of what a student needs to know to be identified at Level 3.	0	0	7	5
	15. My sense of what a student needs to know to be identified at Level 4.	0	0	7	5

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	0	11	1
	17. *Level 3* cut score	0	11	1
	18. *Level 4* cut score	0	11	1

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

- Discussion helped with my understanding.
- Discussion time within our own tables was extremely helpful.
- I liked being able to have table discussions regarding certain questions.
- I thought this was very informative. I learned a lot about the standard setting process, felt like my District and my voice was heard. I am hopeful that the work in this will actually be used with the commissioner.
- I feel that this workshop has been a very valuable experience. The conversations at my table were invaluable!
- I love this workshop.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 Math Grade 4

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	1	0	5	6
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	0	5	7
	3. I understood the impact data that were presented after Round 2.	0	0	0	5	7
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	0	0	5	7
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	3	9

	Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
Influence	6. The Performance Level Descriptions (PLDs).	0	1	5	6
	7. The descriptions of students demonstrating threshold performance.	0	1	7	4
	8. My perception of the difficulty of the items.	0	0	7	5
	9. My experiences with students.	0	1	4	7
	10. Discussion within my group.	0	0	5	7
	11. The item ratings of other participants.	0	3	7	2
	12. The percent of students in each performance level (the impact data).	0	0	7	5
	13. My sense of what a student needs to know to be identified at Level 2.	0	1	5	6
	14. My sense of what a student needs to know to be identified at Level 3.	0	1	5	6
	15. My sense of what a student needs to know to be identified at Level 4.	0	1	5	6

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	0	11	1
	17. *Level 3* cut score	1	10	1
	18. *Level 4* cut score	0	11	1

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

- The discussions were very helpful in understanding the process.
- Group converastions were extremely beneficial.
- I greatly appreciate being included in this experience and would love to be a part of NYSED and NWEA workshops and conferences in the future. Our faciliatator, Lisa, was great to work with and very fair. I also really like that not just classroom teacher were included in this process. Speaking with AIS, math coaches, and even district level math professionals really helped me to see a broader perspective of the student body across NYS.
- It was perfect.
- This was a very valuable experience. I look forward sharing the positive experience with my peers.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 Math Grade 5

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	0	1	8	2
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	0	4	7
	3. I understood the impact data that were presented after Round 2.	0	0	0	2	9
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	0	1	3	7
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	2	9

	Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
Influence	6. The Performance Level Descriptions (PLDs).	0	0	8	3
	7. The descriptions of students demonstrating threshold performance.	0	0	9	2
	8. My perception of the difficulty of the items.	0	0	6	5
	9. My experiences with students.	0	1	5	5
	10. Discussion within my group.	0	0	7	4
	11. The item ratings of other participants.	0	5	6	0
	12. The percent of students in each performance level (the impact data).	0	2	8	1
	13. My sense of what a student needs to know to be identified at Level 2.	0	2	6	3
	14. My sense of what a student needs to know to be identified at Level 3.	0	1	7	3
	15. My sense of what a student needs to know to be identified at Level 4.	0	1	6	4

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	1	9	1
	17. *Level 3* cut score	0	8	3
	18. *Level 4* cut score	2	9	0

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

- We had some great discussions and I believe ultimately arrived at a fair conclusion.
- I appreciated and enjoyed the dialogue; I found the conversations to be stimulating and helpful.
- I enjoyed the conversation and perspectives each person provided. This is s great opportunity to learn what impacts there are on student performance in NYC and the larger school districts.
- The process makes sense, but it is extremely challenging to be okay with the results when it still varies greatly from your own recommendations.
- I believe Round 2 was a better representation of the cut scores and by looking at historical data.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 Math Grade 6

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	0	1	2	6
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	0	3	6
	3. I understood the impact data that were presented after Round 2.	0	0	0	3	6
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	0	0	3	6
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	2	7

Influence	Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
	6. The Performance Level Descriptions (PLDs).	0	0	4	5
	7. The descriptions of students demonstrating threshold performance.	0	1	6	2
	8. My perception of the difficulty of the items.	0	0	7	2
	9. My experiences with students.	0	0	5	4
	10. Discussion within my group.	0	0	2	7
	11. The item ratings of other participants.	0	2	6	1
	12. The percent of students in each performance level (the impact data).	0	1	8	0
	13. My sense of what a student needs to know to be identified at Level 2.	0	0	5	4
	14. My sense of what a student needs to know to be identified at Level 3.	0	0	5	4
	15. My sense of what a student needs to know to be identified at Level 4.	0	0	5	4

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	0	9	0
	17. *Level 3* cut score	0	8	1
	18. *Level 4* cut score	0	8	1

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

- The number of questions at performance levels 1 and 4 seemed to be lacking on the test. We wondered if this varied from year to year and how this influenced the number of students at each level.
- I found this process to be interesting and engaging. I appreciate the opportunity to have been a part of the process.
- In terms of the 6th grade test, I felt that there were not many level 1 or level 4 questions. This is reflected in our threshold scores.
- This has been a wonderful opportunity, thank you so much for including me in this very important work. Ryan was a terrific facilitator. In the future if more items could be written with PLDs of 1 and 4 in the 5th and 6th grades. Thank you.
- Biggest concern is basing this grading system for future exams and hoping for consistency with question types and levels.
- The best part about this workshop was the valuable input from teachers all over the state and their thoughts and ideas and comments."

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 Math Grade 7

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	0	0	5	5
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	0	0	10
	3. I understood the impact data that were presented after Round 2.	0	0	0	0	10
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	0	0	0	10
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	1	9

Influence

Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
6. The Performance Level Descriptions (PLDs).	0	0	2	8
7. The descriptions of students demonstrating threshold performance.	0	0	4	6
8. My perception of the difficulty of the items.	0	3	4	3
9. My experiences with students.	0	2	3	5
10. Discussion within my group.	0	0	3	7
11. The item ratings of other participants.	0	1	7	2
12. The percent of students in each performance level (the impact data).	1	3	2	4
13. My sense of what a student needs to know to be identified at Level 2.	0	0	5	5
14. My sense of what a student needs to know to be identified at Level 3.	0	0	5	5
15. My sense of what a student needs to know to be identified at Level 4.	0	0	5	5

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	1	9	0
	17. *Level 3* cut score	1	9	0
	18. *Level 4* cut score	0	10	0

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

- The conversations within the group were so valuable. Thank you for allowing me to be a part of this process.
- The participation in the discourse around the bookmarks was exceptional. The idea of placing a bookmark based on 'where I feel like 2/3 of my students would struggle' was a hard sell but using the performance level indicators was very helpful.
- We had an excellent discussion that helped me to change my view points from the beginning to end. This has probably been the most helpful math PD I have ever participated in. It is giving me a new light on how to approach questions in the future with my classes.
- Our group had really great discourse and valued everyone's experiences and thoughts. Certainly a great system of checks and balances within our group between the PLDs, the 2/3 threshold with students and our experiences.
- Excellent discussion within the small group and vital discussion whole group. Listening to the educators in the room brought new ideas to the forefront, aiding in a cohesive understanding of the material and process.
- Productive discussion and happy to hear colleagues' differing opinions. It's difficult to balance the many factors influencing our individual cut-score decisions; perceptions vary on what a "Level 2" student is and a "Level 2" question is, as well as the impact of questions written "for" a certain Performance Level. Overall I believe our group has come to a pretty good consensus on what standard setting means, but it's taken most of the full three rounds of ratings and discussions to come to a more confident sense about this.

NEW YORK STATE
DEPARTMENT OF EDUCATION
STANDARD SETTING WORKSHOP
AUGUST 1-3, 2023

Evaluation #2 Math Grade 8

The purpose of this evaluation is to help document the process used to recommend cut scores for New York State's ELA and Math assessments. Your opinions and comments are important, as they provide a basis for judging the quality of this process.

Please do not put your name on this form. While we need the information to examine the success of the various steps in the process, we want your comments to remain anonymous. This information will be reported only in the aggregate. **When you have completed the evaluation, please give it to the facilitator.** *Thank you!*

	Please consider the statements below and mark the level of agreement or disagreement you have with each. Please bubble only one of the five options for each statement.	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Training and Process	1. Before Round 1 began, I was comfortable with the item rating procedure.	0	0	0	1	9
	2. I understood the cut-score summary data that was presented between the rounds.	0	0	0	0	10
	3. I understood the impact data that were presented after Round 2.	0	0	0	0	10
	4. By the end of round 3, I was comfortable with the item rating procedure.	0	0	0	0	10
	5. Overall, I believe my opinions were considered and valued by my group.	0	0	0	2	8

	Please indicate your opinion regarding how influential the following were when you completed your item ratings. Please bubble *only one* of the four options.	Not Influential	Somewhat Influential	Influential	Very Influential
Influence	6. The Performance Level Descriptions (PLDs).	0	0	2	8
	7. The descriptions of students demonstrating threshold performance.	0	0	4	6
	8. My perception of the difficulty of the items.	0	1	5	4
	9. My experiences with students.	0	1	6	3
	10. Discussion within my group.	0	0	1	9
	11. The item ratings of other participants.	0	1	5	4
	12. The percent of students in each performance level (the impact data).	1	2	3	4
	13. My sense of what a student needs to know to be identified at Level 2.	0	2	2	6
	14. My sense of what a student needs to know to be identified at Level 3.	0	2	2	6
	15. My sense of what a student needs to know to be identified at Level 4.	0	2	2	6

Page 165 of 168

	Please indicate your opinion regarding whether you feel the group's final, recommended cut scores were <u>too low, about right, or too high</u> for each cut score. Please bubble *only one* of the three options for each cut score.	Too Low	About Right	Too High
Cut Score	16. *Level 2* cut score	0	9	1
	17. *Level 3* cut score	0	10	0
	18. *Level 4* cut score	0	10	0

In the space provided below, please feel free to add comments about any of your responses, make suggestions to improve future workshops, or tell us what you liked and did not like about this workshop. ***Thank you!***

- This workshop was well led and we were confident knowing what the task was. The entire confence was well run.
- Thank you for the opportunity to be a part of this!
- This was a very beneficial experience. This was also a great group of educators to work with!
- Please use historical data when comparing 8th grade previous year to current year.
- This experience was valuable on multiple levels. The discussions, the collaboration and the knowledge of this group was exemplary. The fascilitor- Russell- kept the group on task and was knowledgable.
- It was mentioned in our group that longitudinal impact data, as well as same-test comparisons, would be a useful tool to see. Overall, the process was great - though we were all less sure at the beginning, we grew together as a group and felt pretty comfortable with the process by the end! The discussions on this last day felt great!
- The reason I said that the Level 2 score was too high was truthfully because I thought the breakdown of the percentage of students at level 1 vs. level 2 was quite high. I don't necessarily disagree with these results because the questions in the OIB indicated that that was that was the appropriate cutoff. I think just overall we'd like to see a few more questions at each PLD level so we could've gotten a bit better understanding of the delination of Level 1 vs Level 2. Great discussion within our group!
- I am grateful for the opportunity to collaborate with other educators. Russell did an amazing job leading the group.

Appendix I: Cut score recommendations and standard error by round

Table II1: ELA cut score recommendations by round

	Level 2			Level 3			Level 4		
ELA G3	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	12	1.6	-0.316	29	1.79	0.663	37	1.4	1.212
Round 2	11	0.16	-0.395	29	0.46	0.663	37	0.62	1.212
Round 3	7	0.28	-0.500	27	0.89	0.605	38	0.55	1.318

	Level 2			Level 3			Level 4		
ELA G4	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	9	1.15	-0.508	25	0.78	0.300	40	0.87	1.147
Round 2	9	0.28	-0.508	24	0.35	0.230	38	0.60	1.060
Round 3	6	0.36	-0.806	22	0.51	0.097	38	0.62	1.060

	Level 2			Level 3			Level 4		
ELA G5	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	10	1.8	-0.762	29.5	2.13	0.501	42	1.72	1.483
Round 2	13	0.37	-0.531	28	0.56	0.400	43	1.06	1.570
Round 3	13	0.36	-0.531	27	0.48	0.333	39.5	0.94	1.423

	Level 2			Level 3			Level 4		
ELA G6	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	13	0.96	-0.340	27	0.95	0.485	40	1.20	1.551
Round 2	13	0.33	-0.340	27	0.30	0.485	39.5	0.38	1.551
Round 3	11	0.46	0.565	25	0.41	0.298	37.5	0.38	1.327

	Level 2			Level 3			Level 4		
ELA G7	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	10	1.60	-0.711	32	2.49	0.370	51	1.54	1.279
Round 2	10	0.50	-0.711	32	1.2	0.370	51	1.27	1.279
Round 3	9	0.53	-0.832	27.5	2.09	0.193	43.5	1.93	0.821

	Level 2			Level 3			Level 4		
ELA G8	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	12	0.99	-0.892	27	1.78	-0.084	44.5	1.32	1.190
Round 2	10	0.63	-0.995	27	1.17	-0.084	42	0.97	0.888
Round 3	10	0.57	-0.995	27	1.10	-0.084	41.5	0.99	0.888

Table I2: Math cut score recommendations by round

	Level 2			Level 3			Level 4		
Math G3	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	7	1.18	-1.049	25	1.31	-0.267	46	1.13	1.190
Round 2	7	0.29	-1.049	25	0.26	-0.267	46	0.56	1.190
Round 3	7	0.29	-1.049	25	0.28	-0.267	46	0.51	1.190

	Level 2			Level 3			Level 4		
Math G4	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	5	1.03	-1.098	19.5	1.82	-0.254	48.5	1.62	1.109
Round 2	8	1.25	-0.853	29	1.41	0.172	49	0.58	1.109
Round 3	6.5	1.14	-0.907	23	0.78	-0.132	49	1.14	1.109

	Level 2			Level 3			Level 4		
Math G5	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	10	1.51	-0.648	26	2.87	0.59	47.5	1.39	1.033
Round 2	9.5	0.84	-0.648	22.5	1.29	-0.105	46	0.67	0.885
Round 3	10	0.68	-0.648	22	1.24	-0.113	49.5	0.72	1.240

	Level 2			Level 3			Level 4		
Math G6	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	9	0.66	-0.299	20.5	1.65	0.067	52.5	1.16	1.470
Round 2	9	0.31	-0.299	21	0.38	0.067	53	0.48	1.470
Round 3	5	0.21	-0.781	21	1.4	0.067	53	0.43	1.470

	Level 2			Level 3			Level 4		
Math G7	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	7.5	1.02	-0.855	22	2.63	-0.220	47.5	1.9	0.843
Round 2	7	0.187	-0.928	23	1.86	-0.158	47.5	0.8	0.843
Round 3	5	0.31	-1.013	19.5	1.54	-0.316	47	0.3	0.806

	Level 2			Level 3			Level 4		
Math G8	Cut point	St Err	Theta	Cut point	St Err	Theta	Cut point	St Err	Theta
Round 1	8	0.95	-0.300	20	2.71	2.62	49.5	1.03	1.380
Round 2	5.5	0.17	-0.375	16	0.54	0.102	46	0.00	1.256
Round 3	5	0.16	-0.433	16	0.1	0.102	46	0.00	1.256