

**NEW YORK STATE
TESTING PROGRAM**

**ENGLISH AS A SECOND LANGUAGE ACHIEVEMENT TEST
(NYSESLAT)**

2007 ADMINISTRATION

TECHNICAL MANUAL

SUBMITTED BY

PEARSON

APRIL 2008



TABLE OF CONTENTS

TABLE OF CONTENTS	ii
TABLES.....	iv
FIGURES	vii
OVERVIEW OF THIS MANUAL	1
PART 1: INTRODUCTION.....	3
1.1 Background.....	3
1.2 Rationale and Purpose.....	4
1.3 Recommended Test Use.....	4
1.4 Test Accommodations	4
Large Type.....	5
Braille	5
PART 2: TEST DESIGN AND DEVELOPMENT.....	6
2.1 Overview.....	6
2.2 Test Specifications by Modality and Grade Span.....	6
2.3 Test Blueprint, Mapping, and Specification by New York State Learning Standards for English as a Second Language.....	8
2.4 Item Development and Review Processes	11
2.5 Field Test	11
2.6 Test Construction.....	12
Psychometric Guidelines for Operational Item Selection and Form Construction.....	13
Testing Written Language.....	14
Testing Oral Language	14
PART 3: SCORING	16
3.1 NYSESLAT Range Finding.....	16
3.2 Operational File.....	17
3.3 Rater Training.....	17
3.4 Inter-Rater and Intra-Rater Reliability.....	17
PART 4: CLASSICAL ITEM-LEVEL STATISTICS	19
4.1 Item-Level Descriptive Statistics.....	19
4.2 Differential Item Functioning	21
Summary of DIF Analyses.....	24
PART 5: RELIABILITY	25
5.1 Internal Consistency Reliability.....	25
5.2 Classical SEM (based on Classical Test Theory)	25
5.3 Standard Error of the Mean (SEMn).....	26
5.4 Conditional SEM (Based on Item Response Theory)	27
5.5 Summary of Descriptive and Reliability Statistics	28
5.6 Inter-Rater Reliability.....	32
Inter-Rater agreement.....	33
Intraclass Correlation.....	33
Kappa Coefficient	34
Summary.....	37
5.7 Reliability of Classification Decision at Proficient Cut	37
PART 6: VALIDITY	41
6.1 Content Validity	41
6.2 Internal Structure.....	42
Validity of the Scoring Structure via Confirmatory Factor Analyses (CFA)	46
6.3 External Structure (Relationship with the New York English Language Assessment)	48
Correlations and Other Descriptive Statistics	48

<i>Proficiency Level</i>	50
<i>Summary</i>	53
6.4 <i>External Structure (Relationship with the New York State Regents Examination in Comprehensive English)</i>	54
<i>Correlations and Other Descriptive Statistics</i>	54
<i>Proficiency Level Classification</i>	57
<i>Summary</i>	58
PART 7: CALIBRATION, EQUATING, AND SCALING	59
7.1 <i>The Rasch and Partial Credit Models</i>	59
<i>Description of the Rasch Model</i>	59
<i>Description of the Partial Credit Models (PCM)</i>	61
7.2 <i>Calibration and Equating of the NYSESLAT</i>	63
<i>Equating Method and Rationale</i>	63
<i>Common Item Linking Design</i>	64
<i>Description of Calibration Sample, Process, and Results</i>	64
<i>Calibration Sample</i>	64
<i>Calibration Process</i>	65
<i>Anchor Item Evaluation</i>	65
<i>Stability of Anchor Items</i>	65
<i>Calibration Results</i>	66
<i>TCC and Standard Error Curves Comparison</i>	66
7.3 <i>Scaling of the NYSESLAT</i>	72
<i>Raw Score to Scale Score Tables</i>	72
7.4 <i>Item Response Model and Rationale for Use</i>	73
7.5 <i>Evidence of Model Fit</i>	73
7.6 <i>Rasch Information</i>	74
PART 8: STANDARD SETTING	77
8.1 <i>Introduction</i>	77
8.2 <i>Standard-Setting Model</i>	77
8.3 <i>Committees of Panelists</i>	77
8.4 <i>Performance Levels and Cut Scores</i>	78
8.5 <i>Standard-Setting Process</i>	78
8.6 <i>Summary Statistics for the Three Rounds of Ratings</i>	80
8.7 <i>Post-Standard-Setting Analyses</i>	80
8.8 <i>Final Performance-Level Cut Points</i>	81
PART 9: SUMMARY OF OPERATIONAL TEST RESULTS	82
APPENDIX	87
APPENDIX A: TEST BLUEPRINT AND ITEM MAP	87
APPENDIX B: ITEM LEVEL STATISTICS BY GRADE SPAN	130
APPENDIX C: RAW SCORE TO SCALE SCORE CONVERSION TABLES	140
APPENDIX D: IRT STATISTICS	153
APPENDIX E: ITEM INFORMATION AT DIFFERENT CUT POINTS BY GRADES	163
APPENDIX F: CONFIRMATORY FACTOR ANALYSIS FOR THE 2007 NYSESLAT	189
APPENDIX G: SCALE SCORE SUMMARY BY SUBGROUPS	199
APPENDIX H: PROFICIENCY PERCENTAGES	225
APPENDIX I: EXIT RATE BY SUBGROUP	264
APPENDIX J: DIF CATEGORIES FOR OPERATIONAL ITEMS	277
APPENDIX K: REFERENCES	280

TABLES

Table 2.1	Test Specifications by Modality and Grade Span	7
Table 2.2	Maximum Number of Points by Modality and Grade Span.....	7
Table 2.3	Test Specifications by New York State ESL Learning Standards by Grade Span	8
Table 2.4	2007 NYSESLAT Item Mapping by New York State ESL Learning Standards.....	9
Table 2.5	Four-Year NYSESLAT Field Test Sampling Matrix.....	12
Table 4.1	Summary of Classical Item Difficulty and Item Discrimination Indices for Each Grade Span	20
Table 4.2	DIF Classification for Constructed-Response Items	22
Table 4.3	DIF Classification for Multiple-Choice Items	23
Table 4.4	Results of DIF Analyses for 2007 Operational Test.....	24
Table 5.1	Descriptive Statistics and Reliability by Grade Span and Modality	29
Table 5.2	Descriptive Statistics and Reliability by Grade and Modality	30
Table 5.3	Audit Sample.....	33
Table 5.4	Rater Agreement for Pre-writing and Writing Prompts.....	34
Table 5.5	Percentages of the Score Difference Between Raters	35
Table 5.6	Comparison Between Local and Pearson Raters.....	36
Table 5.7	Classification Accuracy and Consistency by Grade.....	40
Table 6.1	Intercorrelations Among the Modalities by Grade Span	43
Table 6.2	Intercorrelations Among the Modalities by Grade	43
Table 6.3	Global Fit Indices for the One- and Two-Factor Models (Grades K-1)	46
Table 6.4	Global Fit Indices for the One- and Two-Factor Models (Grades 2-4).....	47
Table 6.5	Global Fit Indices for the One- and Two-Factor Models (Grades 5-6).....	47
Table 6.6	Global Fit Indices for the One- and Two-Factor Models (Grades 7-8).....	47
Table 6.7	Global Fit Indices for the One- and Two-Factor Models (Grades 9-12).....	47
Table 6.8	Descriptive Statistics of the New York ELA and NYSESLAT Raw Scores of Reading, Writing, and Listening.....	48
Table 6.9	Descriptive Statistics of the New York ELA and NYSESLAT Combined Scale Scores of Listening/Speaking and Reading/Writing	49
Table 6.10	Grade 3: Percentage of ELLs Scoring at Each ELA level by NYSESLAT Proficiency Level.....	52
Table 6.11	Grade 4: Percentage of ELLs Scoring at Each ELA level by NYSESLAT Proficiency Level.....	52
Table 6.12	Grade 5: Percentage of ELLs Scoring at Each ELA level by NYSESLAT Proficiency Level.....	52
Table 6.13	Grade 6: Percentage of ELLs Scoring at Each ELA level by NYSESLAT Proficiency Level.....	53
Table 6.14	Grade 7: Percentage of ELLs Scoring at Each ELA level by NYSESLAT Proficiency Level.....	53
Table 6.15	Grade 8: Percentage of ELLs Scoring at Each ELA level by NYSESLAT Proficiency Level.....	53
Table 6.16	Descriptive Statistics of the NYSESLAT Raw Scores of Listening, Speaking, Reading, and Writing Modalities and their Correlations with the Regents English Exam Total Scale Scores	55
Table 6.17	Descriptive Statistics of the NYSESLAT Raw Scores of Listening/Speaking, Reading/Writing Modalities, and Total Scale Scores and their Correlations with the Regents English Exam Total Scale Scores	56
Table 6.18	Grade 10: Percentage of ELLs Scoring at each of the NYSESLAT Proficiency Level Passing/Failing the Regents English Exam.....	57
Table 6.19	Grade 11: Percentage of ELLs Scoring at each of the NYSESLAT Proficiency Level Passing/Failing the Regents English Exam.....	57
Table 6.20	Grade 12: Percentage of ELLs Scoring at each of the NYSESLAT Proficiency Level Passing/Failing the Regents English Exam.....	58
Table 7.1	Common Items as Anchor Items by Modality Combination and Grade Span.....	66
Table 7.2	Average Rasch Difficulty by Grade Span and Modality	75
Table 8.1	Panel Composition for Standard-setting Committees	79
Table 8.2	Final Performance Level Cut Points for Listening and Speaking.....	81
Table 8.3	Final Performance Level Cut Points for Reading and Writing.....	81
Table 9.1	Raw Score Summary by Grade.....	83
Table 9.2	Scale Score Summary by Grade.....	84
Table 9.3	Percentage of Students in Each Proficiency Level By Grade	85
Table 9.4	Exit Rate By Grade for 2005 and 2006 Administrations.....	86
APPENDIX A: TEST BLUEPRINT AND ITEM MAP		87
A.1: 2007 NYSESLAT Operational Test Blueprint.....		87
A.2: 2007 NYSESLAT Operational Test Information		89

<i>A.3: Item Specification</i>	119
APPENDIX B: ITEM LEVEL STATISTICS BY GRADE SPAN	130
<i>B.1: Kindergarten – Grade 1</i>	130
<i>B.2: Grade 2 – Grade 4</i>	132
<i>B.3: Grade 5 – Grade 6</i>	134
<i>B.4: Grade 7 – Grade 8</i>	136
<i>B.5: Grade 9 – Grade 12</i>	138
APPENDIX C: RAW SCORE TO SCALE SCORE CONVERSION TABLES	140
<i>C.1: Kindergarten – Grade 1</i>	140
<i>C.2: Grade 2 – Grade 4</i>	143
<i>C.3: Grade 5 – Grade 6</i>	145
<i>C.4: Grade 7 – Grade 8</i>	147
<i>C.5: Grade 9 – Grade 12</i>	150
APPENDIX D: IRT STATISTICS.....	153
<i>D.1: Kindergarten – Grade 1</i>	153
<i>D.2: Grade 2 – Grade 4</i>	155
<i>D.3: Grade 5 – Grade 6</i>	157
<i>D.4: Grade 7 – Grade 8</i>	159
<i>D.5: Grade 9 – Grade 12</i>	161
APPENDIX E: ITEM INFORMATION AT DIFFERENT CUT POINTS BY GRADES	163
<i>E.1: Kindergarten – Grade 1</i>	163
<i>E.2: Grade 2 – Grade 4</i>	167
<i>E.3: Grade 5 – Grade 6</i>	173
<i>E.4: Grade 7 – Grade 8</i>	177
<i>E.5: Grade 9 – Grade 12</i>	181
APPENDIX F: CONFIRMATORY FACTOR ANALYSIS FOR THE 2007 NYSESLAT	189
<i>Table F.1: Maximum Number of Points by Modality and Grade Span for the 2007 NYSESLAT</i>	189
<i>Table F.2a: Summary Statistics</i>	192
<i>Table F.2b: Factor Loading Estimates for the One- and Two-Factor Models</i>	193
<i>Table F.2c: Global Fit Indices for the One- and Two-Factor Models</i>	193
<i>Table F.3a: Summary Statistics</i>	194
<i>Table F.3b: Factor Loading Estimates for the One- and Two-Factor Models</i>	194
<i>Table F.3c: Global Fit Indices for the One- and Two-Factor Models</i>	194
<i>Table F.4a: Summary Statistics</i>	195
<i>Table F.4b: Factor Loading Estimates for the One- and Two-Factor Models</i>	195
<i>Table F.4c: Global Fit Indices for the One- and Two-Factor Models</i>	195
<i>Table F.5a: Summary Statistics</i>	196
<i>Table F.5b: Factor Loading Estimates for the One- and Two-Factor Models</i>	196
<i>Table F.5c: Global Fit Indices for the One- and Two-Factor Models</i>	196
<i>Table F.6a: Summary Statistics</i>	197
<i>Table F.6b: Factor Loading Estimates for the One- and Two-Factor Models</i>	197
<i>Table F.6c: Global Fit Indices for the One- and Two-Factor Models</i>	197
APPENDIX G: SCALE SCORE SUMMARY BY SUBGROUPS.....	199
<i>G.1: Kindergarten</i>	199
<i>G.2: Grade 1</i>	201
<i>G.3: Grade 2</i>	203
<i>G.4: Grade 3</i>	205
<i>G.5: Grade 4</i>	207
<i>G.6: Grade 5</i>	209
<i>G.7: Grade 6</i>	211
<i>G.8: Grade 7</i>	213
<i>G.9: Grade 8</i>	215
<i>G.10: Grade 9</i>	217

<i>G.11: Grade 10</i>	219
<i>G.12: Grade 11</i>	221
<i>G.13: Grade 12</i>	223
APPENDIX H: PROFICIENCY PERCENTAGES	225
<i>H.1: Kindergarten</i>	225
<i>H.2: Grade 1</i>	228
<i>H.3: Grade 2</i>	231
<i>H.4: Grade 3</i>	234
<i>H.5: Grade 4</i>	237
<i>H.6: Grade 5</i>	240
<i>H.7: Grade 6</i>	243
<i>H.8: Grade 7</i>	246
<i>H.9: Grade 8</i>	249
<i>H.10: Grade 9</i>	252
<i>H.11: Grade 10</i>	255
<i>H.12: Grade 11</i>	258
<i>H.13: Grade 12</i>	261
APPENDIX I: EXIT RATE BY SUBGROUP.....	264
<i>I.1: Kindergarten</i>	264
<i>I.2: Grade 1</i>	265
<i>I.3: Grade 2</i>	266
<i>I.4: Grade 3</i>	267
<i>I.5: Grade 4</i>	268
<i>I.6: Grade 5</i>	269
<i>I.7: Grade 6</i>	270
<i>I.8: Grade 7</i>	271
<i>I.9: Grade 8</i>	272
<i>I.10: Grade 9</i>	273
<i>I.11: Grade 10</i>	274
<i>I.12: Grade 11</i>	275
<i>I.13: Grade 12</i>	276
APPENDIX J: DIF CATEGORIES FOR OPERATIONAL ITEMS	277
<i>Table J.1.: Results of DIF Analyses</i>	277

FIGURES

Figure 5.1	Classification Accuracy.....	38
Figure 5.2	Classification Consistency	38
Figure 7.1	Sample Item Characteristic Curve	60
Figure 7.2	Category Response Curves for a One-step Item.....	61
Figure 7.3	Category Response Curves for a Two-step Item	62
Figure 7.4	2007 TCC Listening/Speaking Grade Span 1–5.....	67
Figure 7.5	2007 TCC Reading/Writing Grade Span 1–5.....	67
Figure 7.6	2005-06-07 TCC Listening/Speaking Grade Span 1	67
Figure 7.7	2005-06-07 TCC Listening/Speaking Grade Span 2	67
Figure 7.8	2005-06-07 TCC Listening/Speaking Grade Span 3	68
Figure 7.9	2005-06-07 TCC Listening/Speaking Grade Span 4	68
Figure 7.10	2005-06-07 TCC Listening/Speaking Grade Span 5	68
Figure 7.11	2005-06-07 SE Curves Listening/Speaking Grade Span 1	68
Figure 7.12	2005-06-07 SE Curves Listening/Speaking Grade Span 2	69
Figure 7.13	2005-06-07 SE Curves Listening/Speaking Grade Span 3	69
Figure 7.14	2005-06-07 SE Curves Listening/Speaking Grade Span 4	69
Figure 7.15	2005-06-07 SE Curves Listening/Speaking Grade Span 5	69
Figure 7.16	2005-06-07 TCC Reading/Writing Grade Span 1	70
Figure 7.17	2005-06-07 TCC Reading/Writing Grade Span 2	70
Figure 7.18	2005-06-07 TCC Reading/Writing Grade Span 3	70
Figure 7.19	2005-06-07 TCC Reading/Writing Grade Span 4	70
Figure 7.20	2005-06-07 TCC Reading/Writing Grade Span 5	71
Figure 7.21	2005-06-07 SE Curves Reading/Writing Grade Span 1	71
Figure 7.22	2005-06-07 SE Curves Reading/Writing Grade Span 2	71
Figure 7.23	2005-06-07 SE Curves Reading/Writing Grade Span 3	71
Figure 7.24	2005-06-07 SE Curves Reading/Writing Grade Span 4	72
Figure 7.25	2005-06-07 SE Curves Reading/Writing Grade Span 5	72
Figure F.1:	Path Diagram for the One-Factor Model.....	191
Figure F.2:	Path Diagram for the Two-Factor Model.....	192

OVERVIEW OF THIS MANUAL

This New York State English as a Second Language Achievement Test (NYSESLAT) Technical Manual for the 2007 administration is organized around nine major parts: Introduction; Test Design and Development; Scoring; Classical Item-Level Statistics; Reliability; Validity; Calibration, Equating, and Scaling (CES); Item Response Theory (IRT) Statistics; Standard Setting; and Summary of Operational Test Results. An overview of this manual is provided below.

Part 1

Part 1 presents the background for the NYSESLAT, its rationale and purpose, recommended test use, and test accommodations. Test accommodations include large type and Braille.

Part 2

Part 2 describes the test development process of the NYSESLAT. It includes test specifications, item development and review processes, item field testing, and test construction.

Part 3

Part 3 provides a description of the scoring process. It includes the description of the range finding meeting that was held in Albany, New York, in 2007. It also provides information about the audit scoring process that was conducted on a ten-percent sample, description of rater training, and inter-rater reliability.

Part 4

Part 4 provides item-level descriptive statistics based on Classical Test Theory (CTT).

Part 5

Part 5 explains the internal consistency reliability, classical Standard Error of Measurement (SEM), conditional SEM based on IRT, and inter-rater reliability. It also provides results of the inter-rater and intra-rater reliability, the rater agreement analyses, the reliability of each of the four modalities, and the reliability of classification decision at the proficient cut.

Part 6

Part 6 describes the validity studies that were conducted. It includes evidence of validity based on test content, internal structure, and relationships to other variables.

Part 7

Part 7 explains the Rasch and Partial Credit Models and provides sample item characteristic curves for a one-step item and a two-step item. It also includes the process of the calibration, equating, and scaling of the 2007 administration of the NYSESLAT. Part 7 also explains the rationale for the use of the IRT model. It includes the IRT model fit statistics and the average Rasch difficulty of the subtests.

Part 8

Part 8 presents the standard-setting process that was followed to establish the performance level cuts. It includes the standard-setting model, the standard-setting process, summary statistics for the round-by-round ratings, evaluation results, post-standard-setting analyses, and final performance-level cut points.

Part 9

Part 9 presents the raw score summary, scale score summary, percentage of students in each performance category, and exit rate for the 2005, 2006, and the 2007 administrations of the NYSESLAT.

PART 1: INTRODUCTION

1.1 Background

Title III of the Federal *No Child Left Behind* (NCLB) Act of 2001 requires annual assessment of the English language skills of limited English proficient (LEP) students. Section 3121(d)(1) explains that each state must use evaluation measures designed to assess “the progress of children in attaining English proficiency, including a child’s level of comprehension, speaking, listening, reading, and writing skills in English.” NCLB requires demonstrated annual improvement in English proficiency for such students in order for them to meet “challenging State academic content and student academic achievement standards.” New York State regulations also require annual assessment of LEP students using a state-approved assessment.

NCLB requires that the annual assessment of LEP students be based on specific student achievement objectives. Section 3122(a)(1) states that “each State educational agency or specially qualified agency receiving a grant under subpart 1 shall develop annual measurable achievement objectives for limited English proficient children served under this part that relate to such children’s development and attainment of English proficiency while meeting challenging State academic content and student academic achievement standards.” Section 3113(b)(2) explains that the “standards and objectives for raising the level of English proficiency [will be] derived from the four recognized domains of speaking, listening, reading, and writing, and [will be] aligned with achievement of the challenging State academic content and student academic achievement standards” set out in Title I of the Act.

The New York State Education Department (NYSED) developed *Learning Standards for English as a Second Language* (ESL) to meet the requirements of NCLB. New York’s learning standards and performance indicators (achievement objectives) for ESL are derived from the domains of speaking, listening, reading, and writing, and they align with the State’s English Language Arts standards. New York’s ESL learning standards and performance indicators are organized in four grade spans: Pre-K–1, 2–4, 5–8, and 9–12. To meet Federal and State requirements regarding the assessment of LEP students, NYSED requested test development, research, and scoring based on the State’s *Learning Standards for English as a Second Language*. Accordingly, the New York State English as a Second Language Achievement Test (NYSESLAT) was developed with four subtests—Speaking, Listening, Reading, and Writing—in each of five grade spans: K–1, 2–4, 5–6, 7–8, and 9–12. Individual test items match to specific ESL learning standards and performance indicators. In addition, the test was developed in accordance with the Standards for Educational and Psychological Testing (American Educational Research Association, 1999) and New York State testing requirements. The test is consistent with the principles of Universal Design, which means that the test is as accessible as possible to all populations, including special needs students. The test is also consistent with applicable federal and state testing requirements.

In response to NYSED’s request, Pearson¹ offered a solution with two distinct phases. With respect to the first phase, custom items were developed and used to create the 2007 NYSESLAT

¹Harcourt Assessment, Inc. has been purchased by Pearson. Pearson is now the vendor for the NYSESLAT.

operational forms. Anchor items were pulled from the 2006 test and also used to produce custom forms for the 2007 test administration. With respect to the second phase, brand new custom items will be developed for the 2008 test administration.

1.2 Rationale and Purpose

The New York State Board of Regents has established learning standards for all English language learners attending New York State schools. NCLB mandates that all English language learners from kindergarten through grade 12 be assessed every year to measure their English language proficiency in speaking, listening, reading, and writing, and that their annual progress toward proficiency be tracked. In compliance with this directive, NYSED developed an annual test that measures student progress toward meeting the state's ESL learning standards. This test is the NYSESLAT. The NYSESLAT helps schools determine which instructional standards teachers must devote time to in order to ensure their English language learners fully acquire the language proficiency that will prepare them for success in the classroom. The purpose of the test is to measure annual student improvement in achieving English language proficiency in order to ultimately exit an ESL or bilingual education program, move into an English Language Arts classroom, and function successfully without any additional support.

1.3 Recommended Test Use

The NYSESLAT is designed to assess students at all proficiency levels within each grade span. This vertical development of the language tested allows the test to discriminate more finely among students at different stages of language acquisition. Because test results provide students, teachers, and parents with an objective report of each student's strengths and weaknesses in the English language skills of speaking, listening, reading, and writing, the NYSESLAT helps determine whether these students are making adequate progress toward English language proficiency.

The test results can also help schools focus on ways to improve instruction so that English language learners become proficient in English, thereby allowing more time for content-based materials, such as mathematics and science.

1.4 Test Accommodations

All test items were developed following the guidelines of Universal Design. Adherence to these guidelines ensured that the assessments were accessible and valid for the widest range of students, including students with disabilities. Applying universal test design during the development process helped eliminate the need to address after-the-fact accommodations, and provided a better assessment for all students. Checklists were used to review every item to ensure that each was built with consideration of the following: equitable use, flexibility in use, simple intuitive design, perceptible information, tolerance for error, low physical effort, and size and span for approach and use. During forms construction, Pearson utilized in-house content and fairness experts to ensure that the forms were pulled with concepts of Universal Design in mind. Pearson stringently reviewed forms for special populations—such as visually or hearing-impaired students—to ensure that items were fair, reliable, and accessible to all.

Large Type

Pearson has standardized large-type product specifications that serve to ease the test-taking experience for visually impaired children who require large type. One form in large type (i.e., minimum 18-point type, and no larger than 24-point type for titles) was produced for each grade span. Pages are printed in black only and on a cream-colored, non-glare vellum stock to ease readability of pages. Covers are printed on heavier stock to provide stiffness to the booklets, which protects interior text pages. Plastic spiral binding makes turning of pages easy to accomplish.

Braille

Pearson produced a Braille version of the NYSESLAT for blind children. Pearson created the Braille version of the NYSESLAT using certified and experienced transcribers who can work with the multiple codes, rules, and guidelines. Pearson produced Braille forms for each NYSESLAT subtest and grade span. For the K–1 grade span, a checklist was provided rather than a Braille test.

If a content area was difficult to Braille, Pearson determined with content specialists if there were other ways that the construct could be worded or measured. To adapt some items, pictures were described. Care was taken not to convey the correct answer but to give a description that would enable the student to ascertain the correct answer.

PART 2: TEST DESIGN AND DEVELOPMENT

2.1 Overview

To meet the requirements of Title III of the Federal *No Child Left Behind* (NCLB) Act and of New York State regulations regarding the assessment of limited English proficient students, NYSED developed *Learning Standards for English as a Second Language*, a comprehensive set of ESL learning standards and performance indicators in the four domains of speaking, listening, reading, and writing. NYSED then requested test development, research, and scoring based on these standards. The NYSESLAT was developed with four subtests (Speaking, Listening, Reading, and Writing) in each of five grade spans (K–1, 2–4, 5–6, 7–8, 9–12) to assess the English language skills and the progress toward achieving full English language proficiency of students in kindergarten through grade 12 who are English language learners. The test was developed in accordance with the Standards for Educational and Psychological Testing (American Educational Research Association, 1999) and New York State testing requirements. The test is consistent with the principles of Universal Design, making it as accessible as possible to all populations, including special needs students, and is also consistent with applicable federal and state testing requirements.

2.2 Test Specifications by Modality and Grade Span

The NYSESLAT is divided into the four basic domains or modalities—Speaking, Listening, Reading, and Writing—for grades K–12. It includes multiple-choice, constructed-response, short-response, and extended-response items. The total number of items per grade span varies. For grade span K–1, there are a total of 70 items; for grade span 2–4, there are a total of 80 items; and for grade spans 5–6, 7–8, and 9–12, there are a total of 87 items each.

The Speaking modality has 16 constructed-response items for all grade spans. The Listening and Reading modalities consist of only multiple-choice items. The number of items for the Listening modality varies from 24–25 for the different grade spans. The number of items for the Reading modality varies from 15–27 for the different grade spans. The number of items for the Writing modality ranges from 15–19 for the various grade spans. The Writing modality is composed of the following parts:

- Multiple-choice section that assesses English language learners’ understanding of the principles of written English at the phoneme, word, and sentence levels;
- Developmental writing items (K–1 only);
- Pre-writing activity (grades 2–12);
- Extended responses to graphics-based prompts (grades K–12).

Test specifications by modality by grade span for the 2007 administration of the NYSESLAT are shown in Table 2.1. Table 2.2 provides the maximum number of points by modality by grade span. Table 2.3 provides a breakdown of the number of items based on New York State’s ESL learning standards and grade span. The 2007 test design consists of all custom developed items for the NYSESLAT.

Table 2.1
Test Specifications by Modality and Grade Span

Number of Items and Passages in NYSESLAT Subtests							
Grade Span	Speaking	Listening	Reading	Writing			Total Number of Items per Grade Span
				Writing Conventions	Pre-writing	Writing Prompt	
	CR	MC	MC/Passages	MC	SR	ER	
K-1	16	24	15/0	6	0	8 CR (Dev. Writing) + 1 ER	70
2-4	16	24	24/5	12	3	1	80
5-6	16	25	27/5	15	3	1	87
7-8	16	25	27/5	15	3	1	87
9-12	16	25	27/5	15	3	1	87

Table 2.2
Maximum Number of Points by Modality and Grade Span

Maximum Number of Points in NYSESLAT Subtests							
Grade Span	Speaking	Listening	Reading	Writing			Total Number of Points per Grade Span
				Writing Conventions	Pre-writing	Writing Prompt	
	CR	MC	MC/Passages	MC	SR	ER	
K-1	34	24	15/0	6	0	16 CR (Dev. Writing) + 2 ER	97
2-4	34	24	24/5	12	6	4	104
5-6	34	25	27/5	15	6	4	111
7-8	34	25	27/5	15	6	4	111
9-12	34	25	27/5	15	6	4	111

Table 2.3
Test Specifications by New York State ESL Learning Standards by Grade Span

Grade Span	Standard	Number of Items	Percent of Total Items	Number of Score	
				Points	Percent of Total Points
K-1	1	45	64	57	59
	2	4	6	8	8
	3	7	10	8	8
	4	10	14	18	19
	5	4	6	6	6
	Total Test		70	100	97
2-4	1	38	48	44	42
	2	11	14	15	14
	3	12	15	16	15
	4	15	19	22	21
	5	4	5	7	7
	Total Test		80	100	104
5-6	1	32	37	39	35
	2	13	15	17	15
	3	18	21	22	20
	4	19	22	25	23
	5	5	6	8	7
	Total Test		87	100	111
7-8	1	34	39	41	37
	2	16	18	20	18
	3	16	18	20	18
	4	20	23	28	25
	5	1	1	2	2
	Total test		87	100	111
9-12	1	26	30	32	29
	2	11	13	15	14
	3	29	33	35	32
	4	15	17	21	19
	5	6	7	8	7
	Total Test		87	100	111

Note. Standard 1. Students will listen, speak, read, and write in English for information and understanding; Standard 2. Students will listen, speak, read, and write in English for literary response, enjoyment, and expression; Standard 3. Students will listen, speak, read, and write in English for critical analysis and evaluation. Standard 4. Students will listen, speak, read, and write in English for classroom and social interaction. Standard 5. Students will demonstrate cross-cultural knowledge and understanding.

2.3 Test Blueprint, Mapping, and Specification by New York State Learning Standards for English as a Second Language

Appendix A.1 provides the test blueprint. Table 2.4 provides an item mapping summary, and A.2 provides in detail the item mapping by New York State Learning Standards for each item within each grade span and modality. Item specifications are presented in A.3.

Table 2.4
2007 NYSESLAT Item Mapping by New York State ESL Learning Standards

Grade Span	Subtest	Standards	Performance Indicators	Items in Subtest	Percentage of Subtest Score	
K-1	Speaking	Standard 1	PI 9, 12	1, 3, 7, 11	24	
		Standard 2	PI 6, 8	2, 8	18	
		Standard 4	PI 2, 3, 5, 11	4, 6, 9, 12-16	47	
		Standard 5	PI 2, 6	5, 10	12	
	Listening	Standard 1	PI 1, 3, 12, 16	1-3, 5, 6, 8-12, 15-18	58	
		Standard 3	PI 1	4, 7, 13, 14, 22, 24	25	
		Standard 4	PI 7, 9	19, 21	8	
		Standard 5	PI 1	20, 23	8	
	Reading	Standard 1	PI 1, 4, 5, 16	1-13	87	
		Standard 2	PI 2	14, 15	13	
	WC	Standard 1	PI 12	1-6	40	
		Writing	Standard 1	PI 2, 11, 12	7-14	53
			Standard 3	PI 1	15	7
	2-4	Speaking	Standard 1	PI 9, 12	1, 4, 7	18
Standard 2			PI 6, 8	2, 8	18	
Standard 3			PI 1	11	6	
Standard 4			PI 2, 3, 5	5, 9, 12-16	41	
Standard 5			PI 2	3, 6, 10	18	
Listening		Standard 1	PI 1, 6, 16	1-8, 12, 13, 16	46	
		Standard 2	PI 2	14, 19	8	
		Standard 3	PI 1, 9	9, 10, 15	13	
		Standard 4	PI 3, 7	11, 17, 18, 20-24	33	
Reading		Standard 1	PI 1, 16	1-7, 12-15, 18, 19	54	
		Standard 2	PI 2, 3, 16	8, 10, 11, 21-24	29	
		Standard 3	PI 3, 16	16, 17, 20	13	
		Standard 5	PI 6	9	4	
WC		Standard 1	PI 5, 12, 16	1-8	36	
		Standard 3	PI 4	9-12	18	
Writing		Standard 1	PI 11	Pre-writing 1, 2, 3	28	
		Standard 3	PI 1	Writing 1	18	
5-6		Speaking	Standard 1	PI 9, 12	1, 3, 6, 7	24
			Standard 2	PI 6, 8	2, 8	18
			Standard 3	PI 1	11	6
	Standard 4		PI 3, 5, 8	9, 12-16	35	
	Standard 5		PI 2, 6	4, 5, 10	18	
	Listening	Standard 1	PI 1, 6, 16	1-5, 9, 12-14	36	
		Standard 2	PI 2	8	4	
		Standard 4	PI 3, 7, 8	6, 11, 15-25	52	
		Standard 5	PI 1	7, 10	8	

Table 2.4 (Continued)
2007 NYSESLAT Item Mapping by New York State ESL Learning Standards

Grade Span	Subtest	Standards	Performance Indicators	Items in Subtest	Percentage of Subtest Score	
	Reading	Standard 1	PI 1, 5, 6	1-5, 7, 14-16	33	
		Standard 2	PI 2, 4, 5	8, 11, 12, 18, 21-25, 27	37	
		Standard 3	PI 1, 3, 9	6, 9, 10, 13, 17, 19, 20, 26	30	
	WC	Standard 1	PI 1, 12	1-7	28	
		Standard 3	PI 4	8-15	32	
	Writing	Standard 1	PI 11	Pre-writing 1, 2, 3	24	
		Standard 3	PI 1	Writing 1	16	
	7-8	Speaking	Standard 1	PI 9, 12	1, 3, 6, 7	24
			Standard 2	PI 6, 8	2, 8	18
			Standard 3	PI 1	16	6
			Standard 4	PI 3, 5	4, 5, 9, 11-15	47
			Standard 5	PI 2	10	6
		Listening	Standard 1	PI 1, 3, 6, 16	1-9, 12-15	52
			Standard 4	PI 3, 7, 8	10, 11, 16-25	48
		Reading	Standard 1	PI 1, 6, 16	1-4, 14-16	26
Standard 2			PI 2, 3, 4, 5, 12	9-12, 17, 18, 20-27	52	
Standard 3			PI 5, 9	5-8, 13, 19	22	
WC		Standard 1	PI 12	1-7	28	
		Standard 3	PI 4	8-15	32	
Writing		Standard 1	PI 11	Pre-writing 1, 2, 3	24	
		Standard 3	PI 1	Writing 1	16	
9-12		Speaking	Standard 1	PI 9, 12	1, 3, 7	18
	Standard 2		PI 6, 8	2, 8	18	
	Standard 3		PI 1	11, 14, 15	18	
	Standard 4		PI 3, 5	4, 5, 9, 12, 13, 16	35	
	Standard 5		PI 2, 6	6, 10	12	
	Listening	Standard 1	PI 1, 16	1-3, 14, 15, 19	24	
		Standard 3	PI 5, 9	4-6, 10-12	24	
		Standard 4	PI 7, 8	16-25	36	
		Standard 5	PI 1, 2	7-9, 13	16	
	Reading	Standard 1	PI 1, 6	1-3, 5, 6, 9, 18	26	
		Standard 2	PI 2, 4, 5, 12	13-17, 23, 25-27	33	
		Standard 3	PI 2, 5, 9	4, 7, 8, 10-12, 19-22, 24	41	
	WC	Standard 1	PI 12	1-7	28	
		Standard 3	PI 4	8-15	32	
	Writing	Standard 1	PI 11	Pre-writing 1, 2, 3	24	
Standard 3		PI 6	Writing 1	16		

2.4 Item Development and Review Processes

With the exception of the anchor items, all items on the 2007 NYSESLAT were written by educators of English language learners, including a number of New York State teachers. An Item Writers' Training Guide was sent out to all item writers to assist them in developing the items. In addition, Pearson Specialists were available via phone and/or e-mail to answer any questions the item writers had. Assessment specialists at Pearson reviewed the passages and items, and in accordance with the item specifications, the assessment specialists ensured the following:

- Absence of bias and sensitive topics in passages
- Item soundness
Item soundness could be interpreted as item validity. An item's soundness is based on the idea that the information derived from an item and its consequences are true or valid
- Absence of bias in items
- Appropriateness of topic, vocabulary, and language structure for each grade span
- Match to the intended New York State Learning Standards for English as a Second Language

The Reading passages were then reviewed by a committee of New York State ESL teachers. Each test question was also reviewed by grade-level ESL educators from around New York State. Only those passages and test questions judged to be of acceptable quality and fair to students from the various ethnic/cultural backgrounds of those who presently live in New York State were approved for use. In addition to the committee reviews, a "Final Eyes" review was conducted by a separate group of New York teachers, to ensure items were error free and appropriate. After items were field tested, New York teachers were invited to a range-finding session, during which they reviewed potential anchor papers and select practice sets, which were used to aid teachers in scoring the 2007 NYSESLAT test.

Anchor items serve the purpose of providing statistics that can help equate tests from one year to the next, so that the scaled scores between them are comparable. Anchor item selection criteria were described on p. 59.

2.5 Field Test

Pearson constructed, and NYSED reviewed and approved, stand-alone field test forms split by modalities (Listening, Reading, Writing, and Speaking) that were administered to a representative sample of limited English proficient students around New York State. There were a total of two field test forms at each grade span, and each form was split by four modalities (Listening, Speaking, Reading, and Writing). Each modality was administered to the schools selected in the fall of 2006. Thus, a district that had schools that may be selected to administer the Listening test in 2006 would have schools that may be selected to administer the Writing test in 2007. The sampling matrix (Table 2.5) used for the sampling assignments for NYSESLAT was provided by NYSED.

Table 2.5
Four-Year NYSESLAT Field Test Sampling Matrix

Field test Form	2006	2007	2008	2009
Listening	S1	S2	S3	S4
Speaking	S2	S3	S4	S1
Reading	S3	S4	S1	S2
Writing	S4	S1	S2	S3

In order to select a representative and sufficient sample of students to participate in the fall 2006 field test, there were three major processes for the fall 2006 sample. The first process was the assignment of a code to each district which will be used over the next several years to indicate which field test strand the district will be assigned. The second step was to project the fall 2007 enrollments so selection could be made early for schools to participate. The third part of the process is the selection of schools to participate in the actual fall 2006 field test. Data from this field test administration informed construction of the 2007 NYSESLAT operational test. The 2006 field-testing sampling plan and statistics were discussed in a separate field test report, while the 2007 sampling plan presented to NYSED by Pearson is also presented in that report.

2.6 Test Construction

Items selected for the 2007 NYSESLAT represented a complete range of difficulty at all grade levels from K–12. Items ranged from easier ones with high p-values, primarily aimed at measuring the skills of students with very limited ability in English, to items with fairly low p-values, aimed at measuring the skills of students with advanced ability in English. The number of both multiple-choice and constructed-response items was increased at each proficiency level, meeting the requirements of the NYSED.

The selection criteria for choosing items for the 2007 NYSESLAT operational forms included: item content, skill measured, item difficulty, gender balance, and ethnic balance. Content considerations include having a variety of settings and activities represented in items, having a mix of people, animals, and objects represented, and ensuring that there was no content overlap within a subtest or across subtests in the same grade span. Skills measured varied by subtest and by sections within a subtest. For example, the Word/Sentence Reading items in the first section of the Reading subtest for grades 2–12 tested ability to read a mixture of nouns, verbs, noun phrases, verb phrases, sometimes adjectives and adjective phrases, and occasionally adverbs. The Reading Comprehension items in the second section of the 2–12 Reading subtest tested ability to identify the main idea or the topic of a passage, to understand details in a passage, to make inferences based on information in a passage, to infer the meaning of a vocabulary word from context, to infer an author’s purpose, and to understand why information was organized or presented in a particular way in a passage. Similar specific content criteria applied for the other subtests. Items selected for the 2007 NYSESLAT represented a complete range of difficulty at all grade levels from K–12. As explained above, items ranged from easier ones with high p-values to items with fairly low p-values. Difficulty level of Reading passages, as distinct from the difficulty of the Reading Comprehension items on the passages, was a criterion as well for the Reading subtest. Finally, gender and ethnic balance were considered in item selection. The people represented in the items across a subtest represented a balance of males and females and a variety of ethnic backgrounds. Differential Item Functioning (DIF) analyses, described in Section 6.1, were performed to identify potential items that needed to be scrutinized for item

bias. None of the items that showed DIF was considered by content specialists to be biased either on the basis of gender or ethnicity.

The process for constructing the 2007 NYSESLAT operational forms was as follows: As soon as Pearson had clean statistics from the fall 2006 field test, Pearson ESL assessment specialists reviewed all items from the field test and pulled forms based on the above criteria. In addition to the fall 2006 field test items, items were pulled from the 2006 operational test to be used as anchor items in the 2007 operational test. The anchor items were necessary for placing the 2007 operational items on the NYSESLAT established scale and constituted roughly 25% of the items in each subtest. The above selection criteria for the non-anchor items (unique items) applied to anchor items as well. Pearson produced proposed draft operational forms which were then presented to NYSED. A joint Pearson-NYSED team reviewed these forms, following all item-selection criteria, and made various changes to the draft forms. Final forms were then reviewed by Pearson ESL assessment specialists, editorial staff, psychometricians, and quality control staff, as well as by NYSED.

Psychometric Guidelines for Operational Item Selection and Form Construction

Statistical considerations included item difficulty and other statistical characteristics of the items. P-values were to be distributed between approximately 0.30 to 0.95, with fewer items at the extremes of difficulty and more items of moderate difficulty. In addition to selecting items with appropriate p-values, staff members were given the following instructions during the item selection and form construction process:

- A. Item Statistics Check
 - 1. Item difficulty check;
 - 2. Range of item difficulties, the items were flagged if p-value < 0.30 or > 0.95;
 - 3. Point biserial range: try to avoid items with pt. biserial < 0.30;
 - 4. Omit rate: watching items with omit rate > 5%;
 - 5. Avoid items with DIF bias flag. If it is necessary to select an item with flag, then it needs to be reviewed carefully during the operational test construction;
- B. There are no changes to items once the item is field tested;
- C. Tests are built to the statistical targets:
 - 1. Average p-value comparison between spring '07 form and spring '06 form. The average p-values should be similar to one another;
 - 2. Average p-value comparison between anchor item set for spring 2007 and the entire spring 2006 test. The average p-values should be similar to one another;
- D. The total number of items at each level and the number of items within each strand must follow the test blueprint.

In general, flagged items should be avoided. However, the match to blueprint and content considerations should take priority over statistical targets in most circumstances.

Testing Written Language

A fundamental consideration in constructing the NYSESLAT is the language that is being tested. While the test developer's native speaker intuition provides choices about what language is tested, more rigorous methods in language choice need to be applied to provide consistency across the forms of the five grade spans and to create a vertical structure within each form, wherein language ranges from the most simple—that which is first acquired by non-native speakers—to advanced language that would indicate a level of English proficiency sufficient for participation in regular academic classes.

For the NYSESLAT, a test designed to assess students at all proficiency levels—beginning, intermediate, and advanced—within each grade span, this vertical development of the language tested allows the test to discriminate more finely among students at different stages of language acquisition. Being able to accurately identify students at different levels of language development provides better information to classroom teachers, who must find the most effective way to help their students reach proficiency. It also provides the very important evidence of students' progress toward proficiency, which is required by the NCLB legislation.

To determine the appropriate language for Reading passages, Pearson assessment specialists, editors, and item and passage writers applied the Flesch-Kincaid grade level readability analyses to all Reading passages. Readability measures are primarily based on factors such as the number of words in the sentences and the number of letters or syllables per word. Additionally, ESL assessment specialists also evaluate the coherence of a passage, the number of anaphora, vocabulary difficulty, sentence and text structure, and concreteness and abstractness. Finally, input from the New York State teacher committee that reviewed the passages was considered. It is the sum of these analyses and evaluations that determines the appropriateness of the language of a passage.

There is a gradual increase in difficulty from passage to passage at every grade span, so that each form includes beginner-level passages as well as passages that are representative of on-grade Reading passages found on English Language Arts reading tests. Pearson also uses the *Educational Developmental Laboratory (EDL) Core Vocabularies in Reading, Mathematics, Science, and Social Studies*, published by Steck-Vaughn, to help determine age- and grade-appropriate language for English language learner items and stimuli for the oral-language subtests. Furthermore, Pearson ESL assessment specialists and editors ensure that the language in all stimuli and items, from kindergarten through grade 12, is both topic- and age-appropriate for test takers.

Testing Oral Language

Recognizing that oral language structure and vocabulary in English differ vastly from the written language, issues of oral language assessment among kindergarten through grade 12 English language learners have been the subject of special investigation at Pearson. Pearson's English language proficiency professionals have conducted research on the item types that appear in the NYSESLAT Speaking and Listening subtests by presenting examples of these item types to English language learners in cognitive laboratories. This setting allows for careful observation and recording of student responses and their reactions to items. Outcomes of this procedure led to important design decisions regarding:

- Item types
- Number of items
- Length of pauses between items
- Use of recorded stimuli
- Recording student spoken responses

The Speaking and Listening subtests of the NYSESLAT are based on these decisions. To ensure that the language in the Speaking and Listening stimuli and items reflect current spoken language as much as possible, Speaking and Listening scripts were submitted to a read-aloud proofing process with English language learner assessment specialists and editors. Additionally, for the oral components of the NYSESLAT to be relevant, the Speaking and Listening subtests had to have predictive validity for academic achievement. Therefore, both academic and social languages are integral components of the Speaking and Listening subtests of the NYSESLAT.

PART 3: SCORING

This part describes the process for scoring the Field Test, Operational Test, and Audit. All Field Test scoring is done by Pearson at the San Antonio site. For field test scoring, each grade span had at least one team of five readers scoring. A 10 percent check score (2nd reading) was done by team leaders. Anchors, training sets, and rubrics were used as scoring guides. Readers scored Pre-writing (PW) 1, 2, then 3 successively. If questions arose during scoring, usually the problem was discussed by the group, in order to maintain consistency in scoring. The operational test scoring was done by New York State teachers. The final operational test data set was provided by the New York State Data Repository, with rubrics and training material as scoring aids. An audit of the Operational Test was done by Pearson readers at the San Antonio site. Ten percent of the total Operational responses are scored in this audit. The same check-score procedures used in scoring the field test are used in the audit scoring. The details of the scoring process for operational items are described below.

3.1 NYSESLAT Range Finding

Range finding was held in Albany on July 9–12, 2007, after field test items were scored by Pearson. The participants included:

- One full-time Pearson Supervisor and five temporary Pearson Performance Assessment Scoring Center (PASC) facilitators (one for each grade span: K–1, 2–4, 5–6, 7–8, and 9–12);
- Three New York State Education Department representatives, who were there to greet participants and show support for the range-finding process. Two of the three NYSED representatives participated in range finding;
- Four teachers and/or education department representatives for each grade span.

Teachers were informed of the selection process for “paper-pulling.” At Pearson, facilitators were teamed up with a second developer, and this team read several hundred papers to find clear-cut, typical examples of score points to share with New York teachers. This range of papers also contained exemplars that would be helpful to include in training sets to make scoring clear.

Either the facilitators or their partners had participated in scoring the field test prior to range finding and were well-acquainted with the rubric, prompts, and hundreds of papers reviewed during scoring. Sample responses for each item were sorted into preliminary range sets. These sets were presented at the range finding meeting in Albany.

Three Pre-writing items and one writing prompt were reviewed per grade span. Each group of teachers read three assembled sets of sample papers per item. Two sets ranged from possible low to high responses, and one set was a mixed range of papers. Each set included at least 15 papers.

Teachers read and assigned scores to each paper and then, as a group, discussed the scores they gave. The group came to a consensus of how each paper should be scored. After coming to agreement about the scores, the group discussed the merits of each paper and selected which would be used as training sets. They used the rubric as their scoring guide.

Pearson's PASC facilitators documented discussions and decisions made at each grade-span session. This documentation helped guide the facilitator in later preparing annotations for each anchor paper selected within each grade span.

The anchor sets contained three examples of each score point, and the annotations explained the reasoning that was used to assign the given score point. Training sets included papers that helped to discriminate the difference between "line papers." A variety of examples were used to show other types of responses different from the anchors, as well as those similar to anchor papers. Through this process, the papers chosen were carefully reviewed and compared through this process to assure consistency.

3.2 Operational File

The Operational File came from the New York State Data Repository after scoring was completed. These files were verified by Pearson's Quality Assurance (QA) Department to ensure data accuracy based on the description values in the file layout. Once all individual files were verified, a concatenated file was built. This file then became the Operational File for the spring 2007 NYSESLAT. Below are the procedures that Pearson conducted to assure reliable and accurate scoring of the items.

3.3 Rater Training

The accuracy of scoring was monitored by training team leaders who are experienced, proficient readers. These team leaders successfully completed a two-day general team leader training workshop and have had the experience of training as Room Directors for many custom projects. These trainers are seasoned Performance Assessment Scoring Center (PASC) readers who have vast experience in all facets of scoring. They carefully monitored the scoring and accuracy of their teams of readers. All responses received a single reading with at least 10 percent getting a second reading to monitor for reliability and accuracy. Team leaders independently gave the check score. All readers maintained at least an 88 percent agreement rate.

All PASC readers have a minimum of a bachelor's degree and have successfully completed generalized workshops in performance assessment scoring before ever being considered as a potential reader for a specific project such as NYSESLAT. Training of readers is based on anchors and training sets that are part of the sampler incorporated into the training procedures for operational scoring done by New York teachers. The sets were developed by New York teachers during range finding in March.

3.4 Inter-Rater and Intra-Rater Reliability

All readers were trained to score to the same scale to ensure accurate, consistent, reliable scoring. PASC adhered to stringent criteria in its general screening, training, and qualifying procedures as preliminary measures for obtaining high levels of consistency and reliability. Team leaders conducted "read behinds," reading the same booklets after readers to check for accuracy of scores. Also, 10 percent of all booklets were "second" read by the training team leader to check accuracy. If individuals were not "on track," retraining ensued. (These readers were perhaps being systematically too lenient or too harsh in their ratings or varied

unsystematically and unpredictably from other raters and deviated from the training standards in their scores). The statistics of the rater reliability are presented in Section 5.6.

PART 4: CLASSICAL ITEM-LEVEL STATISTICS

4.1 Item-Level Descriptive Statistics

This section presents the raw score summary statistics for all items in the spring 2007 administration of the NYSESLAT within the framework of Classical Test Theory. The concatenated file from the New York State Data Repository was used to obtain all raw score statistics. The p-value for each item is defined as the proportion of students that answer an item correctly for the multiple-choice items. A high p-value means that an item is easy; a low p-value means that an item is difficult. For the constructed-response items, the p-value is reported as the average number of points out of the maximum number of possible points.

The point biserial correlation for each item is an index of the association between the item-score and the total-test score. It shows the ability of the item to discriminate between low-ability and high-ability students. An item with a high point biserial correlation discriminates more effectively between the low and the high ability students than a low point biserial correlation.

The item-level statistics for the operational 2007 NYSESLAT are presented in Appendices B.1–B.5 by grade span. The tables are grouped by Listening/Speaking and Reading/Writing modality combinations. The following item information and statistics are presented for each item:

- Item number
- Item format (multiple-choice, constructed-response, short-response, or extended-response)
- Maximum number of possible points
- N-count (number of students)
- Response options for multiple-choice items and percentage of students obtaining each score point for constructed-response items
- Omits (percentage of students omitting an item)
- P-value for multiple-choice items (percentage of examinees that answered the item correctly)
- Item mean for constructed-response items (average number of points earned out of the maximum number of possible points)
- Point Biserial (index of discrimination between high and low scoring students)

Items that are too easy or too difficult are flagged on the basis of their p-values. Such items do not provide adequate information and their inclusion serves a limited purpose in the measurement process. However, the NYSESLAT, being a standard-referenced examination, allows exception to the rule if the item is deemed absolutely necessary by content experts. Although there is no consensus as to what is an acceptable p-value, the guiding policy during item review and form building is a p-value ≤ 0.90 and ≥ 0.30 .

As explained above, the point biserial is another index that indicates the statistical suitability of an item for inclusion in the examination. Since it is an index of correlation of the item to the total test, one would expect a larger point biserial to indicate a desirable value for item inclusion. However, in the context of form building, items are selected not only for their adherence in measuring the underlying trait, i.e., unidimensionality (high correlation), but also on the basis of local independence of the item (low correlation). Very high correlation, in fact, may mean a

redundancy of the item vis-à-vis some other items that perform the same function. On the other hand, too low a value for the point biserial may indicate a “shaky” association of the item with the underlying trait.

For the NYSESLAT, the upper limit of the point biserial was fixed at 0.80, while the lower limit was fixed at 0.25. Although there is seldom an occurrence of the upper limit violation for the NYSESLAT examination, the negative value of the point biserial is stringently observed, since this may indicate a wrong key during the scoring process or an item with a very bad distractor or distractors. These items are never used in the form-building process. Once again, the setting of the upper and lower limits are subjective issues and may be acceptable in many cases, especially if the examination of the response pattern for each distractor shows a negative point biserial with the positive value being only for the correct answer.

As can be seen from Appendix B, with the exception of some high p-values, all the items fall well within our pre-set level of acceptance both in terms of the p-value and the point biserial. In Table 4.1, the mean p-values and point-biserial indices at each grade span for the combined modalities of Listening/Speaking and Reading/Writing are in the middle range and acceptable.

Table 4.1
Summary of Classical Item Difficulty and Item Discrimination Indices for Each Grade Span

Grade Span	Modality	Item Difficulty		Item Discrimination	
		Mean	SD	Mean	SD
K-1	Listening	0.85	0.10	0.39	0.06
	Speaking	0.78	0.08	0.71	0.05
	Reading	0.63	0.09	0.50	0.09
	Writing	0.76	0.16	0.57	0.17
	Listening and Speaking	0.82	0.10	0.52	0.17
	Reading and Writing	0.70	0.15	0.53	0.14
2-4	Listening	0.77	0.13	0.39	0.10
	Speaking	0.86	0.06	0.72	0.05
	Reading	0.68	0.18	0.48	0.06
	Writing	0.70	0.15	0.54	0.12
	Listening and Speaking	0.81	0.12	0.52	0.18
	Reading and Writing	0.69	0.16	0.50	0.09
5-6	Listening	0.77	0.11	0.45	0.09
	Speaking	0.84	0.05	0.76	0.06
	Reading	0.66	0.16	0.44	0.10
	Writing	0.73	0.13	0.50	0.10
	Listening and Speaking	0.80	0.10	0.57	0.17
	Reading and Writing	0.69	0.15	0.47	0.10
7-8	Listening	0.72	0.12	0.45	0.09
	Speaking	0.81	0.06	0.79	0.05
	Reading	0.60	0.15	0.44	0.09
	Writing	0.71	0.13	0.51	0.12
	Listening and Speaking	0.75	0.11	0.58	0.19
	Reading and Writing	0.64	0.15	0.47	0.11
9-12	Listening	0.70	0.14	0.37	0.08
	Speaking	0.76	0.05	0.77	0.05
	Reading	0.66	0.12	0.40	0.09
	Writing	0.65	0.09	0.51	0.11
	Listening and Speaking	0.72	0.12	0.52	0.21
	Reading and Writing	0.66	0.11	0.45	0.11

4.2 Differential Item Functioning

Differential Item Functioning (DIF) analyses are often used to identify potential items that need to be scrutinized for item bias. DIF statistical procedures compute the probability that one demographic group is more likely to answer an item correctly than another group, when the groups are equally able. This information is useful in reviewing items and tests for potential bias in items. However, DIF does not necessarily indicate the existence of bias. As such, the flagging of an item for DIF must be evaluated substantively for bias indication

Since the NYSESLAT included constructed-response items that were polytomously scored, the Mantel-Haenszel odds ratio α could not be used as a DIF index for all the items in the form. Instead, a generalization of the Mantel-Haenszel (1959) procedure for ordered categories, the *Mantel Statistic* (Mantel, 1963), was used for the assessment of DIF in the mixed-format examinations. The Mantel chi-square involves comparing the mean for two groups, conditional on a matching variable. It has one degree of freedom under the null hypothesis of no conditional association between group membership and response. For dichotomous items, the Mantel statistic reduces to the usual Mantel-Haenszel chi-square statistic (without continuity correction). The Mantel statistic has the following mathematical formulation:

$$\text{Mantel Chi-square} = \frac{\left(\sum_K F_K - \sum_K E(F_K)\right)^2}{\sum_K \text{Var}(F_K)}, \quad (1)$$

where F_K represents the sum of scores for the focal group at the k th level of the matching variable, E represents the expected, and Var represents the variance of F_K .

$$F_K = \sum_T y_T n_{FTK}, \quad (2)$$

where y_T represents the T scores that can be obtained on the item, while n_{FTK} denotes the number of focal group members who are on the k th level of the matching variable and received an item score of y_T . The expectation of F_K under the hypothesis of no association is

$$E(F_K) = \frac{n_{F+K}}{n_{++K}} \sum_T y_T n_{+TK}. \quad (3)$$

The Mantel-Haenszel and the Mantel statistic, however, offer a significance test of the presence of DIF without an indication of the direction of DIF, i.e., whether in favor of the reference or the comparison group. The statistic has low power in detecting an association in which the pattern of association for some of the strata is in the opposite direction of the patterns displayed by other strata. On the other hand, as a significance test, its power increases with the number of responses in the two groups of comparison.

Standardized Mean Difference (SMD) is a summary statistic to accompany the Mantel approach (Dorans and Schmitt, 1991). This statistic compares the means of the reference and focal groups, adjusting for differences in the distribution of the reference and focal group members across the values of the matching variable.

Mathematically, SMD is defined as follows:

$$SMD = \sum_k p_{Fk} m_{Fk} - \sum_k p_{Rk} m_{Rk}, \quad (4)$$

where

$$p_{Fk} = \frac{n_{F+k}}{n_{F++}} \quad (5)$$

is the proportion of the focal group members who are at the k^{th} level of the matching variable and

$$m_{Fk} = \frac{1}{n_{F+k} (\sum_t y_t n_{Ftk})} \quad (6)$$

is the mean item score of the focal group members at the k^{th} level, and m_{Rk} is the analogous value for the reference group. As can be seen from the equation above, the *SMD* is the difference between the unweighted item mean of the focal group and the weighted item mean of the reference group. The weights for the reference group are applied to compensate for differences in the number of students in the reference and focal groups (within each level of ability). A negative *SMD* value implies that the focal group has a lower mean item score than the reference group, conditional on the matching variable.

For the DIF classification of constructed-response items, the *SMD* is divided by the total group item standard deviation to obtain an effect-size value for the *SMD*. This effect-size *SMD* is then examined in conjunction with the Mantel χ^2 to obtain *DIF* classifications as shown in Table 4.2 below. DIF is categorized as “no DIF” (AA), “mild DIF” (BB) or “extreme DIF” (CC).

Table 4.2
DIF Classification for Constructed-Response Items

Category	Description	Criterion
AA	No <i>DIF</i>	Non-significant Mantel χ^2 or Significant Mantel χ^2 and $ SMD/SD \leq .17$
BB	Intermediate <i>DIF</i>	Significant Mantel χ^2 and $.17 < SMD/SD \leq .25$
CC	Large <i>DIF</i>	Significant Mantel χ^2 and $.25 < SMD/SD $

Note. SD is the total group standard deviation of the item score in its original metric.

For multiple-choice items, the Mantel-Haenszel Chi-square (M-H χ^2) is used in conjunction with the M-H odds ratio (transformed to what ETS calls the Delta Scale (D)). To calculate the delta, the odds ratio should be obtained first. The odds of a correct response (proportion passing

divided by proportion failing) is P/Q (i.e., $P/[1-P]$). The odds ratio is simply the odds of a correct response of the reference group divided by the odds of a correct response of the focal group. For a given item, the odds ratio is defined as follows:

$$\alpha_{M-H} = \frac{P_r / Q_r}{P_f / Q_f}. \quad (7)$$

The corresponding null hypothesis is that the odds of getting the item correct are equal for the two groups (the odds ratio is equal to 1):

$$H_0: \alpha_{M-H} = \frac{P_r / Q_r}{P_f / Q_f} = 1. \quad (8)$$

In order to make the odds ratio symmetrical around zero with its range being in the interval $-\infty$ to $+\infty$, the odds ratio is transformed into a log odds ratio as per the following:

$$\beta_{M-H} = \ln(\alpha_{M-H}). \quad (9)$$

The simple natural logarithm transformation of this odds ratio is symmetrical about zero (zero has the interpretation of equal odds). This DIF measure is a signed index where a positive value signifies DIF in favor of the reference group, while a negative value indicates DIF in favor of the focal group. β_{M-H} is amenable to linear transformations to other interval scale metrics (Camilli & Shepard, 1994). This fact is utilized by ETS to transform β_{M-H} to their Delta Scale metric (D), via:

$$D = -2.35 \cdot \beta_{M-H}. \quad (10)$$

The quantity D represents the apparent difference in the difficulty of the items in the Delta metric for the two groups whose performance is being compared. The following table depicts *DIF* classifications for MC items based on the M-H χ^2 and the item Delta Scale difference value (D):

Table 4.3
DIF Classification for Multiple-Choice Items

Category	Description	Criterion
A	No DIF	Non-significant M-H χ^2 or $ D < 1.0$
B	Intermediate DIF	Neither A nor C
C	Large DIF	Significant M-H χ^2 and $ D \geq 1.5$

For NYSESLAT, we compared the performance of male with female students and of Hispanic students with the rest of the ethnic groups including white students. In these comparisons, male students and Hispanic students were considered the reference groups, while females and the other ethnicities were considered the focal groups. However, all groups could not be analyzed for DIF because of the low n-counts in some of the groups. For the field test items, Spanish-speaking students were compared with students whose home language was something other than Spanish, provided enough n-counts were available for the comparison.

In the separate 2007 Field Test Technical Report, DIF analysis results for each grade span by form for the field test items were provided, while Appendix J of this report provides the same for the operational items. The > sign next to DIF category indicates that the item is in favor of the reference group, while < indicates in favor of the focal group.

Summary of DIF Analyses

Tables 4.4 provides a summary of DIF analysis of operational test items, including n-counts for each DIF analysis group and number of items flagged by both SMD and/or Mantel for the constructed items and ETS Delta Scale metric (D) and M-H criteria for multiple-choice items (see details in Table 4.2 and Table 4.3). All items flagged for DIF were carefully reviewed during operational test construction. Only those items that passed the reviews were included in the operational tests. It is important to note that DIF does not indicate bias. None of the items showing moderate or large DIF in Table 4.4 were identified as “biased” by the content specialists.

Table 4.4
2007 Results of DIF Analyses for 2007 Operational Test

Grade Span	Modality	N-Counts	Total Number of DIF Items	Total Number of Moderate DIF Items	Total Number of Large DIF Items
K-1	Listen/Speak	50421	18	14	4
	Read/Write	50177	11	9	2
2-4	Listen/Speak	57450	5	3	2
	Read/Write	57224	2	2	0
5-6	Listen/Speak	24767	10	7	3
	Read/Write	24620	5	2	3
7-8	Listen/Speak	22602	5	3	2
	Read/Write	22415	19	14	5
9-12	Listen/Speak	35473	14	11	3
	Read/Write	34654	17	13	4

Note. Detail results can be found in Appendix J.

PART 5: RELIABILITY

5.1 Internal Consistency Reliability

The internal consistency of a test is measured by the stability of scores from one sample of content to another. Several methods can be used to estimate the internal consistency of a test. One approach is to split all test questions into two groups and then correlate student scores on the two half-tests. This is known as a split-half estimate of reliability. This method avoids the implications of any changes in the individual by administering only a single test. If scores have a high rate of correlation on the two half-tests, it can be concluded that the test questions complement one another, function well as a group, and measure similar concepts. This also suggests that measurement error is minimal.

The split-half method's decision about which questions contribute to each half-test's score can have an impact on the resulting correlation. Pearson uses Cronbach's coefficient alpha statistic (Cronbach, 1951) to avoid this concern about the split-half method. The coefficient alpha is the average split-half correlation based on all possible divisions of a test into two parts. The coefficient alpha can be used to estimate the internal consistency of both dichotomously (right or wrong, 0 or 1 score values) and polytomously (a wide range of score values) scored test items. Coefficient alpha is computed by the following formula:

$$\alpha = \frac{I}{I-1} \left(1 - \frac{\sum_i s_i^2}{S_x^2} \right) \quad (11)$$

where

I is the number of items on the test,

s_i^2 is the variance of item i , and

S_x^2 is the total test variance.

Cronbach's coefficient alpha reliability statistic was calculated and presented in Table 5.1 and 5.2. Reliability values of close to or more than 0.90 are considered to be high, and the test is considered a reliable test.

5.2 Classical SEM (based on Classical Test Theory)

Since no assessment measures ability with perfect consistency, it is useful to take into account the likely size of measurement errors. One way to describe the inconsistency of assessment results is to assess a student on multiple occasions and note how much the scores vary. Repeatedly measuring a student can only be done hypothetically. However, if you could assess a student on multiple occasions, you would obtain a collection of the student's obtained scores. The scores would cluster around an average value. The standard deviation, or spread, of these obtained scores is known as the standard error of measurement (SEM).

The SEM is another index of reliability and provides an estimate of the amount of error in an individual's observed test score. The individual's observed total score is considered the estimate of the person's true score. Because the standard error of measurement is inversely related to the reliability of a test, the greater the reliability, the less the standard error of measurement, and the more confidence one may have in the accuracy, or precision, of the observed test score. The measurement error is commonly expressed in terms of standard deviation units; that is, the standard error of measurement is the standard deviation of the measurement error distribution. The standard error of measurement is calculated with the following equation:

$$SEM = SD\sqrt{1-r_{xx}} \Leftrightarrow s_e = s_x \sqrt{1 - \frac{s_t^2}{s_x^2}} \quad (12)$$

where

$SEM (=s_e)$ refers to the standard error of measurement,

$SD (=s_x)$ is the standard deviation unit of the scale for a test,

r_{xx} is the reliability coefficient for a sample test (or estimate of ρ_{XX} , which is a population reliability coefficient),

s_t^2 is the estimate of σ_T^2 , and

s_x^2 is the estimate of σ_X^2 .

The SEMs are presented in Tables 5.1 and 5.2 in Section 5.5. The smaller the SEMs (close to 0), the higher the quality of the test will be.

5.3 Standard Error of the Mean (SEMn)

The standard error of the mean, on the other hand, is an estimate of the magnitude of sampling error associated with the sample mean in the estimation of the population mean. This expected standard mean of sampling errors of the mean is called the standard error of the mean (SEMn) and is defined as follows:

$$SEMn = \frac{\sigma}{\sqrt{n}} \quad (13)$$

where

SEMn = standard error of the mean

σ = standard deviation of the population

n = number of responses in each sample

The SEMns are presented in Tables 5.1 and 5.2 in Section 5.5. The more accurate the estimation of the population mean, the smaller the SEMn values will be.

5.4 Conditional SEM (Based on Item Response Theory)

Unlike the SEM based on the classical test theory, the SEM based on the item response theory (IRT) is not the same for all persons. For example, if a person gets either a few or a large number of items correct (extreme score), the standard error is greater than if the person gets a moderate number of items correct. This implies that the standard error of measurement depends on the total score (Andrich & Luo, 2004).

Under the Rasch model, the SEM for each person is as follows:

$$\sigma_{\hat{\beta}} = \frac{1}{\sqrt{\sum_{i=1}^L p_{vi}(1-p_{vi})}} \quad (14)$$

where

v is subscript for a person,

i is subscript for an item,

L is length of the test,

$\hat{\beta}$ is ability estimate, and

p_{vi} is the probability that a person answers an item correctly and is defined as follows:

$$P_{vi} = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}} \quad (15)$$

where β_v is person v 's ability and δ_i is the item's difficulty.

A confidence band can be used in interpreting the ability estimate. For example, an approximate 68% confidence interval for $\hat{\beta}$ is given by

$$\hat{\beta} \pm SEM \quad (16)$$

Note that the standard error for item difficulty is smallest when the probability of passing is close to the probability of failing. That is, when an item is near the threshold level for many persons in the sample, the standard error is small (Embretson & Reise, 2000).

According to the general consensus in measurement, an aspect for the popularity of IRT methods in analyzing data is based on the fact that classical stats assume equivalency of students and item measurements for all examinees and items in a test. IRT methods allow for the differentiation of varied student and item performances in estimating the reliability of the measurement (Crocker & Algina, 1986). As evidenced by Equation 5 above, one reason for the fluctuation in the standard errors of students is that they are a function of the n-counts. As such, the standard errors for each of the ability score estimates are smallest in the middle of the score distribution (where most examinees perform) and greatest for estimates in the extreme where subsequently lower

numbers of students perform on a test, and thus produce less precise estimates. It is for this reason that IRT estimates with individual standard errors at score points, i.e., conditioned on theta (student ability estimates), are preferred to classical SEMs which do not differentiate between the precision of student estimates at different levels of performances.

The conditional standard errors of measurement are presented in the raw score to scale score conversion tables in Appendix C and for IRT statistics in Appendix D.

5.5 Summary of Descriptive and Reliability Statistics

Tables 5.1 and 5.2 provide the raw score descriptive statistics and reliabilities by grade span and grade level. The tables include the following information:

- Number of items
- Maximum number of possible points
- Number of students
- Means and standard deviations in raw scores
- Mean p-values
- Standard error of the mean (SEMn)
- Cronbach's alpha internal consistency reliability
- Standard error of measurement (SEM)

In analyzing the information provided in Tables 5.1 and 5.2, the measures of reliability should be examined by keeping to the general trend that associates greater reliability with a lower error index for that particular measure. In Table 1, the test reliability values of Listening/Speaking and Reading/Writing across the five grade spans are all above 0.90, which is a strong indication that the test forms were constructed in good quality. The SEMn values of Listening/Speaking and Reading/Writing across the five grade spans ranged from 0.04 to 0.09, which is very small and acceptable. The SEM values of Listening/Speaking and Reading/Writing across the five grade spans ranged from 2.46 to 3.16, which is also acceptable. Of course, if an item has, for example, a SEMn greater than the standard deviation of the mean, then one would definitely question the relevancy of the information, which might be the result of an outlier or outliers.

In Table 5.2, most test reliability values of Listening/Speaking and Reading/Writing across the twelve grades are all above 0.90. The reliability for Grade K Reading/Writing may seem a little low in relation to the other reliability values, but this could partially be because of their ages. They just started learning reading/writing even in their mother language. This may also be explained by the relatively erratic learning to read and write English in the kindergarten grades, which could be based on students being in a new environment, missing their homes, the level of learning to read and write at home (which may not be uniform across students), etc. This kind of reasoning seems plausible by the fact that it is by the same type of reasoning that kindergarten students are separated from other grade levels and tested on their own by other states (e.g., MI). The SEMn and SEM values are small and within acceptable ranges at each grade level.

Table 5.1
Descriptive Statistics and Reliability by Grade Span and Modality

Grade	Test	Number of Items	Max Points	N-Count	RS Mean	SD	Mean P-value	SEMn	Reliability	SEM
K-1	Listening	24	24	50387	20.45	3.77	0.85	0.02	0.84	1.52
	Speaking	16	34	50430	26.15	7.76	0.77	0.03	0.94	1.88
	Reading	15	15	50146	9.41	3.93	0.63	0.02	0.84	1.59
	Writing	15	24	50445	17.60	5.38	0.73	0.02	0.89	1.80
	Listening and Speaking	40	58	50387	46.60	10.54	0.80	0.05	0.93	2.69
	Reading and Writing	30	39	50146	27.06	8.58	0.69	0.04	0.92	2.46
2-4	Listening	24	24	57397	18.46	4.26	0.77	0.02	0.82	1.80
	Speaking	16	34	57430	29.01	6.85	0.85	0.03	0.95	1.57
	Reading	24	24	57172	16.43	5.10	0.68	0.02	0.86	1.88
	Writing	16	22	57434	14.98	5.01	0.68	0.02	0.86	1.86
	Listening and Speaking	40	58	57397	47.48	10.06	0.82	0.04	0.93	2.60
	Reading and Writing	40	46	57172	31.45	9.50	0.68	0.04	0.92	2.68
5-6	Listening	25	25	24751	19.33	4.97	0.77	0.03	0.87	1.79
	Speaking	16	34	24781	28.48	7.73	0.84	0.05	0.96	1.58
	Reading	27	27	24605	17.72	5.65	0.66	0.04	0.86	2.10
	Writing	19	25	24776	18.00	5.32	0.72	0.03	0.86	1.97
	Listening and Speaking	41	59	24751	47.84	11.66	0.81	0.07	0.95	2.64
	Reading and Writing	46	52	24605	35.77	10.15	0.69	0.06	0.92	2.91
7-8	Listening	25	25	22588	17.91	5.51	0.72	0.04	0.88	1.92
	Speaking	16	34	22621	27.27	8.88	0.80	0.06	0.97	1.65
	Reading	27	27	22401	16.15	5.97	0.60	0.04	0.87	2.18
	Writing	19	25	22624	17.00	5.79	0.68	0.04	0.88	2.04
	Listening and Speaking	41	59	22588	45.22	13.18	0.77	0.09	0.95	2.85
	Reading and Writing	46	52	22401	33.21	10.79	0.64	0.07	0.92	3.04
9-12	Listening	25	25	35428	17.49	4.85	0.70	0.03	0.83	1.99
	Speaking	16	34	35513	25.60	8.96	0.75	0.05	0.96	1.74
	Reading	27	27	34613	17.76	5.43	0.66	0.03	0.84	2.19
	Writing	19	25	35522	15.91	6.29	0.64	0.03	0.88	2.21
	Listening and Speaking	41	59	35428	43.12	12.38	0.73	0.07	0.94	3.01
	Reading and Writing	46	52	34613	33.83	10.70	0.65	0.06	0.91	3.16

Note. The data file used in the table above is the final research file. The total n-counts for grades K–12 are 192,053. The stats presented in the table are based on the n-counts after cleaning (remove duplicate cases, off-level cases, and all records with Strand Raw Score = 0 or Greater than 900).

Table 5.2
Descriptive Statistics and Reliability by Grade and Modality

Grade	Test	Number of Items	Max Points	N-Count	RS		Mean			
					Mean	SD	P-value	SEMn	Reliability	SEM
K	Listening	24	24	24422	19.13	4.01	0.80	0.03	0.81	1.74
	Speaking	16	34	24445	23.99	8.16	0.71	0.05	0.94	2.00
	Reading	15	15	24258	7.19	3.18	0.48	0.02	0.69	1.76
	Writing	15	24	24454	14.98	5.18	0.62	0.03	0.87	1.85
	Listening and Speaking	40	58	24422	43.14	10.97	0.74	0.07	0.93	2.94
	Reading and Writing	30	39	24258	22.22	7.39	0.57	0.05	0.87	2.63
1	Listening	24	24	25965	21.68	3.04	0.90	0.02	0.82	1.28
	Speaking	16	34	25985	28.17	6.76	0.83	0.04	0.93	1.73
	Reading	15	15	25888	11.48	3.39	0.77	0.02	0.83	1.40
	Writing	15	24	25991	20.06	4.30	0.84	0.03	0.86	1.58
	Listening and Speaking	40	58	25965	49.86	8.98	0.86	0.06	0.93	2.39
	Reading and Writing	30	39	25888	31.59	7.00	0.81	0.04	0.91	2.15
2	Listening	24	24	21933	17.11	4.23	0.71	0.03	0.79	1.93
	Speaking	16	34	21951	28.57	6.79	0.84	0.05	0.94	1.64
	Reading	24	24	21818	14.48	4.79	0.60	0.03	0.83	1.98
	Writing	16	22	21950	13.29	5.00	0.60	0.03	0.85	1.95
	Listening and Speaking	40	58	21933	45.69	9.84	0.79	0.07	0.92	2.73
	Reading and Writing	40	46	21818	27.82	9.14	0.60	0.06	0.91	2.81
3	Listening	24	24	19336	18.98	4.06	0.79	0.03	0.81	1.75
	Speaking	16	34	19341	29.30	6.76	0.86	0.05	0.95	1.53
	Reading	24	24	19271	17.10	4.87	0.71	0.04	0.86	1.85
	Writing	16	22	19343	15.56	4.76	0.71	0.03	0.85	1.81
	Listening and Speaking	40	58	19336	48.29	9.89	0.83	0.07	0.93	2.53
	Reading and Writing	40	46	19271	32.69	9.00	0.71	0.06	0.91	2.62
4	Listening	24	24	16128	19.68	4.01	0.82	0.03	0.83	1.65
	Speaking	16	34	16138	29.27	7.03	0.86	0.06	0.95	1.51
	Reading	24	24	16083	18.28	4.88	0.76	0.04	0.87	1.75
	Writing	16	22	16141	16.57	4.61	0.75	0.04	0.86	1.75
	Listening and Speaking	40	58	16128	48.96	10.19	0.84	0.08	0.94	2.46
	Reading and Writing	40	46	16083	34.89	8.88	0.76	0.07	0.92	2.50
5	Listening	25	25	13222	19.30	4.77	0.77	0.04	0.86	1.81
	Speaking	16	34	13235	28.87	7.21	0.85	0.06	0.95	1.56
	Reading	27	27	13149	17.58	5.60	0.65	0.05	0.86	2.11
	Writing	19	25	13230	18.03	5.20	0.72	0.05	0.86	1.96
	Listening and Speaking	41	59	13222	48.19	10.95	0.82	0.10	0.94	2.61
	Reading and Writing	46	52	13149	35.65	10.00	0.69	0.09	0.92	2.91
6	Listening	25	25	11529	19.36	5.20	0.77	0.05	0.88	1.77
	Speaking	16	34	11546	28.04	8.27	0.82	0.08	0.96	1.61
	Reading	27	27	11456	17.89	5.71	0.66	0.05	0.87	2.09
	Writing	19	25	11546	17.96	5.44	0.72	0.05	0.87	1.98
	Listening and Speaking	41	59	11529	47.44	12.41	0.80	0.12	0.95	2.67
	Reading and Writing	46	52	11456	35.92	10.33	0.69	0.10	0.92	2.91

Table 5.2 (Continued)
Descriptive Statistics and Reliability by Grade by Modality

Grade	Test	Number of Items	Max Points	N-Count	RS		Mean			
					Mean	SD	P-value	SEMn	Reliability	SEM
7	Listening	25	25	10944	17.61	5.50	0.70	0.05	0.87	1.95
	Speaking	16	34	10955	27.13	8.94	0.80	0.09	0.97	1.66
	Reading	27	27	10865	15.49	5.89	0.57	0.06	0.86	2.21
	Writing	19	25	10957	16.79	5.80	0.67	0.06	0.88	2.04
	Listening and Speaking	41	59	10944	44.76	13.27	0.76	0.13	0.95	2.88
	Reading and Writing	46	52	10865	32.32	10.71	0.62	0.10	0.92	3.06
8	Listening	25	25	11644	18.20	5.50	0.73	0.05	0.88	1.89
	Speaking	16	34	11666	27.40	8.82	0.81	0.08	0.97	1.63
	Reading	27	27	11536	16.78	5.99	0.62	0.06	0.87	2.16
	Writing	19	25	11667	17.20	5.78	0.69	0.05	0.88	2.04
	Listening and Speaking	41	59	11644	45.65	13.08	0.77	0.12	0.95	2.82
	Reading and Writing	46	52	11536	34.06	10.80	0.65	0.10	0.92	3.02
9	Listening	25	25	13325	16.48	5.22	0.66	0.05	0.85	2.05
	Speaking	16	34	13359	24.03	9.94	0.71	0.09	0.97	1.79
	Reading	27	27	13011	16.37	5.65	0.61	0.05	0.84	2.24
	Writing	19	25	13360	14.51	6.66	0.58	0.06	0.89	2.25
	Listening and Speaking	41	59	13325	40.53	13.83	0.69	0.12	0.95	3.14
	Reading and Writing	46	52	13011	31.06	11.33	0.60	0.10	0.92	3.25
10	Listening	25	25	11734	17.41	4.74	0.70	0.04	0.82	2.01
	Speaking	16	34	11751	25.52	8.62	0.75	0.08	0.96	1.76
	Reading	27	27	11471	17.80	5.28	0.66	0.05	0.83	2.19
	Writing	19	25	11755	15.93	6.08	0.64	0.06	0.87	2.20
	Listening and Speaking	41	59	11734	42.95	11.90	0.73	0.11	0.94	3.01
	Reading and Writing	46	52	11471	33.86	10.39	0.65	0.10	0.91	3.16
11	Listening	25	25	6998	18.72	4.18	0.75	0.05	0.79	1.91
	Speaking	16	34	7014	27.43	7.45	0.81	0.09	0.95	1.66
	Reading	27	27	6860	19.32	4.78	0.72	0.06	0.81	2.11
	Writing	19	25	7018	17.55	5.58	0.70	0.07	0.86	2.11
	Listening and Speaking	41	59	6998	46.20	9.99	0.78	0.12	0.92	2.82
	Reading and Writing	46	52	6860	37.03	9.22	0.71	0.11	0.89	3.04
12	Listening	25	25	3371	19.23	3.87	0.77	0.07	0.77	1.87
	Speaking	16	34	3389	28.31	7.46	0.83	0.13	0.96	1.56
	Reading	27	27	3271	19.83	4.69	0.73	0.08	0.81	2.06
	Writing	19	25	3389	17.92	5.43	0.72	0.09	0.85	2.07
	Listening and Speaking	41	59	3371	47.60	9.43	0.81	0.16	0.91	2.77
	Reading and Writing	46	52	3271	38.06	8.59	0.73	0.15	0.88	2.97

Note. The data file used in the table above is the final research file. The total n-counts for grades K-12 are 192,053. The stats presented in the table are based on the n-counts after cleaning (remove duplicate cases, off-level cases, and all records with Strand Raw Score = 0 or Greater than 900).

5.6 Inter-Rater Reliability

Another source of measurement error results in the evaluation of student work. Inter-rater reliability investigates the extent to which examinees would obtain the same score if the assessment task is scored two or more times by the same rater or different raters. One way to estimate this type of reliability is to have two raters score each student's paper and then obtain the correlation. In this case, consistency is defined as similarity of students' rank orderings by two raters. Another way to obtain evidence of inter-rater reliability is to calculate the percent agreement between raters. If raters always agree in their assignment of scores, there is 100% agreement. If raters never agree in their assignment of scores, there is 0% agreement. The choice between using a correlation coefficient or percent agreement depends upon whether students' absolute (actual) or relative (rank order) score level is important for a particular interpretation and use.

Audit scoring of the operational test conducted by PASC readers consists of only open-ended items. Personnel selected as team leaders are usually the same people who have participated in the range finding process and are familiar with the New York State Test. Training consists of a review of the same scoring materials used by the New York State teachers in scoring the full Operational Test. Readers are tested by scoring sample papers prior to actual audit scoring. Once qualified, readers are check-scored by team leaders to maintain their accuracy.

To ensure that local teachers apply the same rigorous scoring standards as intended by the New York State Education Department (NYSED), and to provide evidence of inter-rater reliability, the Department requires that Pearson rescore 10% of all test papers after each test administration.

The selection of the 10% audit sample for each grade span followed a stratified random selection procedure based on the Need/Resource Category, which divided the State schools into six categories: New York City, Big 4 Cities (Buffalo, Rochester, Syracuse, and Yonkers), High Need Urban/Suburbs, High Need Rural, Average Need, and Low Need.

Target values were calculated for each level using the percentage of enrollment data for each of the above classifications applied to the level targets for '10% Plus 40% Over Sample.' Since the sample selection level was by school, enrollment data collected for the spring 2007 administration was used to select schools to be used for the sampling and analysis, with the goals of acquiring at least 95% but no more than 105% of the target count for each grade span.

These schools returned their booklets to Pearson for rescoring. Below are the procedures that Pearson conducted to assure reliability and accurate scoring of the items.

Table 5.3 provides the audit sample information regarding the number of students obtained for the rater analyses after merging the auditing file (Accudata) and final research file. The Accudata and Research file were merged by student's ID (SID). The Ns in the table are based on the 88% of the Accudata file from auditing to match to the research file data base. The percentage of the audit sample ranged from 8% to 19% across five grade spans.

Table 5.3
Audit Sample

Grade Span	Number of Students in Audit Sample after Merging ¹	Number of Students in Population by Grade Span	Percentage of Total Population ²
K-1	4457	50539	9
2-4	4657	57672	8
5-6	3666	25094	15
7-8	4412	22921	19
9-12	3781	35827	11
Total	20973	192053	11

Note 1: There are 3008 Accudata students that did not merge with the final research file and 4 Accudata students that did merge with the final research file but had missing data.

Note 2: Percentage of Total Population = Number of Students in Audit Sample/Number of students in final research file.

Inter-Rater Agreement

Table 5.4 provides the rater agreement for the Pre-writing and Writing constructed-response items between local raters and Pearson raters. The Pearson raters scored the items independently. When the two raters assigned the same score to a student's paper, the agreement rating was denoted as *exact*, i.e., perfect agreement. Ratings that differed by exactly one score point were denoted as *adjacent*. Ratings that differed by two or more score points were denoted as *non-adjacent*. The results are presented by grade span and items. For Grade K–1, items 1–9 were scored 2. For Grade 2–12, items 1–3 were scored 0, 1, or 2, and item number 4 was scored 0, 1, 2, 3, or 4. The following are the descriptions of the rater-agreement variable:

Exact: 0 Score Point Difference between Local and Pearson Raters

Adjacent: +/-1 Score Point Difference between Local and Pearson Raters

Non-Adjacent: +/-2 Score Point Difference between Local and Pearson Raters

In Table 5.4, for the 2-point items, the percent of exact rating ranged from 62.87 to 97.82 across all the grade spans. And the percent of adjacent rating ranged from 1.59 to 33.91. The total percent of the exact and adjacent ratings ranged from 95.32 to 99.97. The percent of exact rating is relatively low for item 4 (Grade 2–12) because it has a greater range (0–4). The percent of exact rating ranged from 48.53 to 57.68 across the four grade spans. And the percent of adjacent rating ranged from 37.72 to 43.69. The total percent of the exact and adjacent ratings for the four point items ranged from 92.22 to 96.65.

The table also provides the intraclass correlations between the local raters and Pearson raters, Kappa coefficient, the asymptotic standard error (ASE) of the Kappa coefficient, and its lower and upper 95 percent confidence limits.

Intraclass Correlation

The Intraclass Correlation assesses rating reliability by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects. The intra-class correlations in Table 5.4 were calculated by using the random and fixed effects model introduced by Shrout and Fleiss (1979) to measure the agreement of the local and Pearson raters. The SAS

program from Douglas Steinley and Philip Karl Wood (2000) was modified so that large student samples could be used for the analyses. The intraclass correlations ranged from 0.62 to 0.93 across all the grade spans. The lowest correlations occurred at Grade Span 9-12.

Kappa coefficient

Kappa coefficient is an index of measuring the rater agreement with the value always less than or equal to 1. A value of 1 implies perfect agreement, and values less than 1 imply less than perfect agreement. In rare situations, Kappa can be negative. This is a sign that the two observers agreed less than would be expected just by chance. It is rare that we get perfect agreement. Different people have different interpretations as to what is a good level of agreement.

Here is one interpretation of Kappa (Altman DG. Practical Statistics for Medical Research, 1991, page 404).

- Poor agreement = less than 0.20
- Fair agreement = 0.20 to 0.40
- Moderate agreement = 0.40 to 0.60
- Good agreement = 0.60 to 0.80
- Very good agreement = 0.80 to 1.00

In Table 5.4, the range of Weighted Kappa across grade spans is from 0.51 to 0.91, which indicates that the rater agreement falls in from moderate agreement to very good agreement using the interpretation rule above. The Weighted Kappa was lower at higher grade span such as grade span 9-12. The asymptotic standard error (ASE) of the Kappa coefficient and its lower and upper 95 percent confidence limits indicate that the errors of the Kappa coefficient are very small and within limited range.

Table 5.4
Rater Agreement for Pre-writing and Writing Prompts

Grade Span	Item	Max Point	N-count	Exact	Adjacent	Non-Adjacent	Intraclass Correlation	Weighted Kappa	ASE	95% Confidence Limits	
										Lower	Upper
K-1	1	2	4457	87.41	12.52	0.07	0.77	0.74	0.01	0.72	0.76
	2	2	4457	85.12	14.72	0.16	0.75	0.72	0.01	0.70	0.74
	3	2	4457	97.82	1.59	0.58	0.92	0.89	0.01	0.87	0.92
	4	2	4457	97.51	1.66	0.83	0.93	0.91	0.01	0.89	0.93
	5	2	4457	88.29	11.38	0.34	0.87	0.83	0.01	0.82	0.85
	6	2	4457	85.44	14.29	0.27	0.85	0.80	0.01	0.79	0.82
	7	2	4457	83.19	16.40	0.40	0.87	0.80	0.01	0.78	0.81
	8	2	4457	76.04	23.58	0.38	0.82	0.73	0.01	0.71	0.74
	9	2	4457	75.99	23.51	0.49	0.81	0.73	0.01	0.71	0.74
2-4	1	2	4657	83.40	15.07	1.52	0.72	0.65	0.01	0.63	0.67
	2	2	4657	81.23	17.63	1.14	0.78	0.71	0.01	0.69	0.73
	3	2	4657	79.92	18.72	1.35	0.77	0.70	0.01	0.68	0.72
	4	4	4657	55.64	41.01	3.35	0.74	0.57	0.01	0.56	0.59

Table 5.4 (Continued)
Rater Agreement for Pre-writing and Writing Prompts

Grade Span	Item	Max Point	N-count	Exact	Adjacent	Non-Adjacent	Intraclass Correlation	Weighted Kappa	ASE	95% Confidence Limits	
										Lower	Upper
5-6	1	2	3666	82.62	17.13	0.25	0.75	0.70	0.01	0.68	0.72
	2	2	3666	84.86	14.76	0.38	0.79	0.73	0.01	0.71	0.75
	3	2	3666	88.54	10.80	0.65	0.82	0.76	0.01	0.74	0.79
	4	4	3666	52.24	42.61	5.16	0.68	0.51	0.01	0.49	0.53
7-8	1	2	4412	85.13	14.14	0.73	0.83	0.79	0.01	0.77	0.81
	2	2	4412	85.65	13.64	0.70	0.84	0.80	0.01	0.78	0.81
	3	2	4412	67.95	28.24	3.81	0.66	0.58	0.01	0.56	0.60
	4	4	4412	57.68	37.72	4.60	0.74	0.59	0.01	0.57	0.61
9-12	1	2	3781	77.23	21.53	1.24	0.76	0.67	0.01	0.64	0.69
	2	2	3781	67.23	28.09	4.68	0.66	0.58	0.01	0.56	0.60
	3	2	3781	62.87	33.91	3.23	0.62	0.53	0.01	0.50	0.55
	4	4	3781	48.53	43.69	7.78	0.72	0.54	0.01	0.53	0.56

Note: The stats in the table are based on the merged file between Accudata (auditing) and Pearson’s data base (final research file). The Accudata and Research file were merged by SIDs. The Ns in the table are based on the 88% of the Accudata file from auditing to match to the Pearson’s data base.

Table 5.5 provides the frequency distribution of the score point differences between the local raters’ and Pearson raters’ scores by grade span. For the 2-point items, the percentage of 0 differences ranged from 63 to 98. The percentage of 1 score point difference ranged from 95 to 100. For the 4-point items, the percentage of 0 differences ranged from 49 to 58. The percentage of 1 score point difference ranged from 93 to 97. The above evidence indicates that the rater agreement percentage is higher for the 2-point item than the 4-point item. In general, the rater agreement percentage is high and acceptable for all the open-ended items.

Table 5.5
Percentages of the Score Difference Between Raters

Grade Span	Item	Percentage of Point Difference								
		Score Difference (Local score minus Pearson score)								
		-4	-3	-2	-1	0	1	2	3	4
K-1	1			0	6	87	6	0		
	2			0	7	85	7	0		
	3			0	1	98	1	0		
	4			1	1	98	1	0		
	5			0	3	88	8	0		
	6			0	4	85	10	0		
	7			0	5	83	11	0		
	8			0	7	76	17	0		
	9			0	10	76	14	0		

Table 5.5 (Continued)
Percentages of the Score Difference Between Raters

Grade Span	Item	Percentage of Point Difference								
		Score Difference (Local score minus Pearson score)								
		-4	-3	-2	-1	0	1	2	3	4
2-4	1			1	11	83	4	0		
	2			1	13	81	5	0		
	3			1	13	80	6	0		
	4			1	15	56	26	2	0	
5-6	1			0	7	83	10	0		
	2			0	8	85	7	0		
	3			1	6	89	5	0		
	4			0	12	52	31	5	0	
7-8	1			0	4	85	11	1		
	2			0	4	86	10	0		
	3			2	9	68	19	2		
	4	0		1	11	58	26	4	0	0
9-12	1			0	10	77	12	1		
	2			1	7	67	21	4		
	3			1	13	63	21	2		
	4		0	2	18	49	25	5	0	0

Table 5.6 provides the mean and standard deviation of each item for both the local raters and the Pearson raters. There is a column for the local raters, a column for Pearson raters, and a third column for the difference of the mean scores and ratio of standard deviations between the local and Pearson raters. The largest mean difference (0.28) occurred at Grade Span 5-6 (item 4), and SD ratio ranged from 0.94 to 1.28. For most items, the mean difference is close to 0, and SD ratio is close to 1, which is a strong indication of rater agreement.

Table 5.6
Comparison Between Local and Pearson Raters

Grade Span	Item	Local		Pearson		Differences	
		Mean	SD	Mean	SD	Mean	SD Ratio
K-1	1	1.66	0.53	1.66	0.53	0.00	1.00
	2	1.59	0.56	1.59	0.56	0.00	1.00
	3	1.86	0.49	1.86	0.50	0.01	0.98
	4	1.79	0.60	1.80	0.59	-0.01	1.02
	5	1.36	0.68	1.31	0.72	0.05	0.94
	6	1.13	0.70	1.07	0.73	0.05	0.96
	7	1.35	0.81	1.29	0.82	0.07	0.99
	8	1.09	0.83	0.99	0.81	0.11	1.02
	9	0.96	0.84	0.92	0.80	0.04	1.05
2-4	1	1.63	0.65	1.72	0.58	-0.09	1.12
	2	1.46	0.72	1.55	0.68	-0.09	1.06
	3	1.42	0.74	1.51	0.70	-0.09	1.06
	4	1.96	1.10	1.82	0.93	0.14	1.18

Table 5.6 (Continued)
Comparison Between Local and Pearson Raters

Grade Span	Item	Local		Pearson		Differences	
		Mean	SD	Mean	SD	Mean	SD Ratio
5-6	1	1.53	0.61	1.51	0.59	0.03	1.03
	2	1.61	0.63	1.61	0.62	0.00	1.02
	3	1.67	0.63	1.69	0.60	-0.02	1.05
	4	2.36	1.10	2.07	0.86	0.28	1.28
7-8	1	1.38	0.71	1.31	0.72	0.08	0.99
	2	1.37	0.71	1.31	0.70	0.06	1.01
	3	1.32	0.79	1.22	0.80	0.10	0.99
	4	1.95	1.14	1.73	0.94	0.21	1.21
9-12	1	1.49	0.72	1.47	0.75	0.02	0.96
	2	1.32	0.79	1.12	0.83	0.21	0.95
	3	1.21	0.79	1.10	0.78	0.10	1.01
	4	2.14	1.27	2.01	1.09	0.13	1.17

Summary

All the evidence of inter-rater reliability statistics (percentage of agreement, intraclass correlation, weighted kappa coefficient, percentage of score point difference, and mean difference) indicated that the Local raters and Pearson raters have good agreement regarding the NYSESLAT scores.

5.7 Reliability of Classification Decision at Proficient Cut

Based on the NYSESLAT scale scores, student performance is classified into one of four proficiency levels. While it is always important to know the reliability of student scores in any examination, it is of even greater importance to assess the reliability of the decisions based on these scores. Evaluation of the reliability of classification decisions is performed through estimation of the probabilities of correct and consistent classification of student performance. Procedures from Livingston and Lewis (1995) were applied to derive measures of the accuracy and consistency of the classifications. Brief descriptions of the procedures used and results obtained are presented in this section.

The accuracy of decisions is the extent to which decisions would agree with those that would be made if each student could somehow be tested with all possible forms of the examination. The consistency of decisions is the extent to which decisions would agree with the decisions that would have been made if the students had taken a parallel form of the NYSESLAT, equal in difficulty and covering the same content as the form they actually took. These ideas are shown schematically in Figures 5.1 and 5.2. Please note that the term Achieves Proficient Status refers to the proficient category on the Listening/Speaking and Reading/Writing combinations score and the term Does Not Achieve Proficient Status refers to all categories below proficient status.

		Decision made on a form actually taken	
		Does Not Achieve Proficient Status	Achieves Proficient Status
True status made on all-forms average	Does Not Achieve Proficient Status	Correct Classification	Misclassification
	Achieves Proficient Status	Misclassification	Correct Classification

Note: Adapted from Young and Yoon (1998).

Figure 5.1: Classification Accuracy

		Decision made on the 2nd form taken	
		Does Not Achieve Proficient Status	Achieves Proficient Status
Decision made on the 1st form taken	Does Not Achieve Proficient Status	Correct Classification	Misclassification
	Achieves Proficient Status	Misclassification	Correct Classification

Note: Adapted from Young and Yoon (1998).

Figure 5.2: Classification Consistency

In Figure 5.1, accurate classifications occur when the decision made on the basis of the all-forms average agrees with the decision made on the basis of the form actually taken.

Misclassifications occur when, for example, a student who actually accomplished Does Not Achieve Proficient Status on the basis of his or her all-forms average is classified incorrectly as accomplishing Achieves Proficient Status. Consistent classification occurs (Figure 5.2) when two forms agree on the classification of a student as either Achieves Proficient Status or Does Not Achieve Proficient Status, whereas inconsistent classification occurs when the decisions made by the forms differ.

These analyses make use of the techniques outlined and implemented by Hanson (1991), Haertel (1996), Livingston and Lewis (1995), and Young and Yoon (1998). The software developed by Hanson (1995) was used for the analyses. Estimates of decision accuracy and consistency were made for the Achieves Proficient Status cut points on the Listening/Speaking and Reading/Writing scores reported in the NYSESLAT.

The table also includes the proportions of False Positive and False Negative classifications. The sum of values of Accuracy, False Positive, and False Negative is equal to 1.00, but due to rounding, the table values may or may not equal 1.00. False Positive and False Negative classifications refer to the mismatch between student true scores and observed scores. The False Positive value is the proportion of student scores misclassified to the category Achieves Proficient Status when student scores do not meet proficient status. The False Negative value is the proportion of student scores misclassified to the category Does Not Achieve Proficient Status when student scores actually do meet proficient status.

Table 5.7 presents the results of the decision accuracy and consistency of the Achieves Proficient Status cut scores for the Listening/Speaking and Reading/Writing scores. The table contains the following:

- Accuracy
- False positives
- False negatives
- Consistency

Decision accuracy refers to the agreement between the classifications based on the form actually taken and the classifications that would be made if each student could somehow be tested with all possible forms (or true score) of the examination. For example, for K students, 77% of the classifications for Listening/Speaking at the proficiency level were accurate, while 23% percent of the students were classified as proficient when they were not. No students were classified as proficient when they were not. For Listening/Speaking combination proficiency level classification, the accuracy of the classification for Listening/Speaking at the proficiency level ranged from 69% to 92% across all grade levels. For Reading/Writing combination proficiency level classification, the accuracy ranged from 87% to 97% across all the grade levels.

Decision consistency refers to the agreement between the classifications based on the form actually taken and the classifications that would be made on the basis of an alternate form. For example, for K students, the chance that students were classified the same was 69%. For the Listening/Speaking combination proficiency level classifications, the consistency ranged from 65% to 88% across all the grade levels. For the Reading/Writing combination proficiency level classifications, the consistency ranged from 82% to 96% across all the grade levels. Both decision accuracy and consistency for the proficiency level were higher for the Reading/Writing combination than those for the Listening/Speaking. The table also illustrates the general rule that decision consistency is less than decision accuracy. This is understandable, since decision consistency is dependent on a single alternate form, while accuracy relates to the agreement of the assessment classification with the classification on multiple forms, where the variation on a single form would be subsumed under the overall variation based on countless multiple forms to produce true scores.

The false positive ranged from 4% to 23% for Listening/Speaking and 2% to 9% for Reading/Writing across all the grade levels. The false positive values were very low for all grade levels (4% to 12%) except K (23%). The false negative ranged from 0% to 25% for Listening/Speaking and 1% to 7% for Reading/Writing across all the grade levels. Relatively speaking, more attention should be paid to the high false positives values, because we don't like to see high percentages of students who were not proficient classified as proficient and exit the program. Generally speaking, the NYSESLAT shows much higher percentages of accuracy and consistency in comparison to the false positive and negative percentages.

Table 5.7
Classification Accuracy and Consistency by Grade

Grade	Test	Accuracy	False Positives	False Negatives	Consistency
K	Listening and Speaking	0.77	0.23	0.00	0.69
	Reading and Writing	0.90	0.05	0.05	0.87
1	Listening and Speaking	0.81	0.07	0.13	0.74
	Reading and Writing	0.93	0.04	0.03	0.90
2	Listening and Speaking	0.70	0.09	0.20	0.65
	Reading and Writing	0.91	0.05	0.04	0.88
3	Listening and Speaking	0.69	0.06	0.25	0.67
	Reading and Writing	0.87	0.06	0.07	0.82
4	Listening and Speaking	0.74	0.12	0.14	0.68
	Reading and Writing	0.92	0.04	0.04	0.89
5	Listening and Speaking	0.74	0.10	0.16	0.71
	Reading and Writing	0.90	0.06	0.04	0.86
6	Listening and Speaking	0.78	0.10	0.12	0.72
	Reading and Writing	0.94	0.04	0.02	0.92
7	Listening and Speaking	0.86	0.05	0.09	0.80
	Reading and Writing	0.90	0.07	0.03	0.87
8	Listening and Speaking	0.89	0.06	0.05	0.85
	Reading and Writing	0.93	0.03	0.03	0.91
9	Listening and Speaking	0.92	0.04	0.05	0.88
	Reading and Writing	0.91	0.04	0.05	0.88
10	Listening and Speaking	0.85	0.06	0.09	0.81
	Reading and Writing	0.90	0.06	0.04	0.87
11	Listening and Speaking	0.84	0.07	0.08	0.79
	Reading and Writing	0.88	0.09	0.04	0.84
12	Listening and Speaking	0.87	0.07	0.06	0.83
	Reading and Writing	0.97	0.02	0.01	0.96

Note: The data file used in the table above is the final research file. The total n-counts for grades K–12 are 192,053. The stats presented in the table are based on the n-counts after cleaning (remove duplicate cases, off-level cases, and all records with Strand Raw Score = 0 or Greater than 900). Additionally, cases with missing grade are also removed.

PART 6: VALIDITY

Assessments constructed from Pearson support the validity-related standards set forth in the Standards for Educational and Psychological Testing. Our judgments about test validity are based on the following sources of evidence of validity²:

- Test content—“an analysis of the relationship between a test’s content and the construct it is intended to measure” (p. 11)
- Internal structure—“the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are made” (p. 13)
- Relationships to other variables—“analyses of the relationship of test scores to variables external to the test” (p. 13)

6.1 Content Validity

Evidence of validity based on test content is revealed by the extent to which the material on the test represents an appropriate sampling of skills, knowledge, and understanding of the domain tested. The *New York State Learning Standards for English as a Second Language* defines the skills, knowledge, and understanding which are expected for English language learner (ELL) children in New York State. There are several steps involved in aligning NYSESLAT items and test forms to these standards in order to ensure the content validity of the test.

Pearson psychometricians and ELL assessment specialists, with input from NYSED assessment and content specialists, developed a test blueprint that includes a range of item types that can measure all those standards and performance indicators from the *New York State Learning Standards for English as a Second Language* that are measurable on a standardized test. (Pearson ELL assessment specialists, working with committees of New York State ESL teachers, identified performance indicators that were designed for classroom use only, and thus could not be measured in a standardized test.)

When NYSESLAT item writers received item writing assignments, they were instructed to use NYSESLAT item type specifications in conjunction with the *New York State Learning Standards for English as a Second Language* to guide their writing, and then to identify a particular standard and performance indicator that best matches what each item is measuring. When items were reviewed by committees of New York State teachers, an integral part of the review process was discussion of and agreement on the standard and performance indicator assigned to each item.

²The page number in the parenthesis is the page number in the Standards for Educational and Psychological Testing, 1999.

The assigned standards and performance indicators were one of several criteria (e.g., item statistics, content diversity, gender and ethnic balance, etc.) that were then considered by Pearson ELL assessment specialists and NYSED assessment and content specialists during forms construction to ensure that forms reflected broad coverage of the *Learning Standards*. After forms construction was completed, item maps were developed for each form to show this coverage item by item. ELL assessment specialists at Pearson and content specialists at NYSED reviewed the item maps to confirm the alignment of a given form to the *New York State Learning Standards for English as a Second Language*. The item mapping provided in Table 6.1 below and Appendices A.2–A.8 gives concrete evidence for the alignment to the *New York State Learning Standards for English as a Second Language*.

6.2 Internal Structure

An assessment procedure should not be a random collection of assessment tasks or test questions. Each task in the assessment should contribute positively to the total result. The interrelationship among the tasks on an assessment is known as the internal structure of the assessment. Typical questions that investigate the relationships among assessment parts (Nitko, 2004) include:

- Do all of the assessment tasks “work together” so that each task contributes positively toward assessing the quality of interest?
- If different parts of the assessment procedure are to provide unique information, do the results support this uniqueness?
- If different parts of the assessment procedure are to provide the same or similar information, do the results support this?

In order to investigate the answers to these questions, correlations based on raw scores were obtained between the four modalities. Tables 6.1 and 6.2 present the intercorrelations among the four modalities by grade span and grade. The evidence of internal structure of the 2007 NYSESLAT is also depicted by the point biserial correlation coefficient and fit statistics. Appendices B.1–B.5 and D.1–D.5 provide these statistics for the 2007 NYSESLAT administration.

Table 6.1
Intercorrelations Among the Modalities by Grade Span

Grade Span	Test	Correlation		Coefficient	
		Listening	Speaking	Reading	Writing
K-1	Listening	1.00			
	Speaking	0.63	1.00		
	Reading	0.51	0.43	1.00	
	Writing	0.61	0.53	0.71	1.00
2-4	Listening	1.00			
	Speaking	0.62	1.00		
	Reading	0.67	0.51	1.00	
	Writing	0.67	0.57	0.78	1.00
5-6	Listening	1.00			
	Speaking	0.68	1.00		
	Reading	0.69	0.55	1.00	
	Writing	0.73	0.69	0.74	1.00
7-8	Listening	1.00			
	Speaking	0.67	1.00		
	Reading	0.70	0.54	1.00	
	Writing	0.75	0.72	0.70	1.00
9-12	Listening	1.00			
	Speaking	0.59	1.00		
	Reading	0.68	0.53	1.00	
	Writing	0.68	0.64	0.71	1.00

Note: The data file used in the table above is the final research file. The total n-counts for grades K-12 are 192,053. The stats presented in the table are based on raw scores with the n-counts after cleaning (remove duplicate cases, off-level cases, and all records with Strand Raw Score = 0 or Greater than 900). Additionally, cases with missing grade are also removed.

Table 6.2
Intercorrelations Among the Modalities by Grade

Grade	Test	Correlation		Coefficient	
		Listening	Speaking	Reading	Writing
K	Listening	1.00			
	Speaking	0.58	1.00		
	Reading	0.36	0.26	1.00	
	Writing	0.53	0.43	0.55	1.00
1	Listening	1.00			
	Speaking	0.63	1.00		
	Reading	0.49	0.45	1.00	
	Writing	0.56	0.53	0.69	1.00
2	Listening	1.00			
	Speaking	0.57	1.00		
	Reading	0.61	0.47	1.00	
	Writing	0.62	0.53	0.76	1.00
3	Listening	1.00			
	Speaking	0.65	1.00		
	Reading	0.65	0.53	1.00	
	Writing	0.66	0.60	0.76	1.00

Table 6.2 (Continued)
Intercorrelations Among the Modalities by Grade

Grade	Test	Correlation		Coefficient	
		Listening	Speaking	Reading	Writing
4	Listening	1.00			
	Speaking	0.69	1.00		
	Reading	0.67	0.56	1.00	
	Writing	0.68	0.63	0.77	1.00
5	Listening	1.00			
	Speaking	0.67	1.00		
	Reading	0.68	0.53	1.00	
	Writing	0.72	0.67	0.73	1.00
6	Listening	1.00			
	Speaking	0.70	1.00		
	Reading	0.70	0.58	1.00	
	Writing	0.74	0.71	0.75	1.00
7	Listening	1.00			
	Speaking	0.68	1.00		
	Reading	0.69	0.52	1.00	
	Writing	0.75	0.73	0.69	1.00
8	Listening	1.00			
	Speaking	0.67	1.00		
	Reading	0.72	0.56	1.00	
	Writing	0.74	0.71	0.71	1.00
9	Listening	1.00			
	Speaking	0.64	1.00		
	Reading	0.69	0.58	1.00	
	Writing	0.72	0.70	0.73	1.00
10	Listening	1.00			
	Speaking	0.56	1.00		
	Reading	0.66	0.51	1.00	
	Writing	0.67	0.63	0.70	1.00
11	Listening	1.00			
	Speaking	0.47	1.00		
	Reading	0.62	0.42	1.00	
	Writing	0.60	0.54	0.64	1.00
12	Listening	1.00			
	Speaking	0.34	1.00		
	Reading	0.57	0.32	1.00	
	Writing	0.53	0.39	0.59	1.00

Note: The data file used in the table above is the final research file. The total n-counts for grades K–12 are 192,053. The stats presented in the table are based on the n-counts after cleaning (remove duplicate cases, off-level cases, and all records with Strand Raw Score = 0 or Greater than 900). Additionally, cases with missing grade are also removed.

Observations of language proficiency assessment subtests in Table 6.1:

- Listening and Speaking are moderately correlated for Grades 2–12 ranging from 0.59–0.68.
- Listening and Reading are moderately correlated for Grades 2–12 ranging from 0.51–0.70.
- Reading and Writing are moderately correlated for Grades 2–12 ranging from 0.70–0.78.
- Speaking and Writing are moderately correlated for Grades 2–12 ranging from 0.53–0.72.
- The language domain pairs of Listening and Speaking and Reading and Writing are moderately to highly correlated, while Speaking and Reading are not highly correlated.
- Listening and Reading and Listening and Writing correlations are higher than Listening and Speaking for Grades 2–12. One possible explanation is that the Speaking subtest is more related to social context, while Listening, Reading, and Writing are more related to academic context. It is possible that the higher the grades, the more the contents of Listening, Reading, and Writing were related to academic setting.

Generally speaking, the internal correlations based on grade spans are within expectation. The Reading and Writing subtests have the highest correlation range (from 0.70 to 0.78) across the five grade spans. Speaking and Reading subtests have the lowest correlation range (from 0.26 to 0.58). Listening and Speaking, Listening and Reading, Listening and Writing, and Speaking and Writing have correlation ranges between 0.51 and 0.75.

Observations of language proficiency assessment subtests appear in Table 6.2:

Kindergarten

- Students in this age group do not usually read or write yet, but can have Listening and Speaking skills. The expected outcome is that neither Reading nor Writing will correlate highly with Listening or Speaking.
- In Table 6.2, for kindergarten students, Speaking and Listening have a correlation of 0.58, and Reading and Writing have a correlation of 0.55. Speaking and Reading have a low correlation of 0.26, Listening and Reading have a low correlation of 0.36, Speaking and Writing have a correlation of 0.43, and Listening and Writing have a correlation of 0.53.

Grades 1–8

- A steady increase in the correlation (from 0.56 to 0.75) between Writing and Listening is observed.
- A possible explanation (personal communication with Pearson content specialists) is that, in general, students during this age span experience expanding use and development in their Writing skills. At the same time, demands on the Listening skills of this age group remain fairly static with only moderate development.

Grades 9–12

- A steady decrease in the correlation (from 0.72 to 0.53) between Writing and Listening is observed.
- A possible explanation (personal communication with Pearson content specialists) is that by high school, there is an increased focus on use of writing skills, especially an increased focus on academic content. Requirements of high-school age student listening skills also increase, but not nearly at as steep a curve as writing.
- Similar arguments may be made for the correlational behavior between Speaking and Writing in grades 1–12. There is a steady increase in the correlation (from 0.55 to 0.73) for Grades 1–8, and there is a steady decrease in the correlation from (0.70 to 0.39) for Grades 9–12.
-

Generally speaking, the internal correlations based on grade levels are within expectation. The Reading and Writing subtests have the highest correlation range (from 0.55 to 0.77) across the five grade spans. Speaking and Reading have the lowest correlations range (from 0.26 to 0.58). Listening and Speaking, Listening and Reading, Listening and Writing, and Speaking and Writing have correlation ranges between 0.34 and 0.75.

Validity of the Scoring Structure via Confirmatory Factor Analyses (CFA)

In order to assess the validity of the scoring structure, i.e., Speaking and Listening as one score and Reading and Writing as the other, a confirmatory factor analysis was undertaken with the 2007 full population data in order to compare the two-factor modality combinations with the unidimensional model (i.e., the total test providing a single score).

The 2007 NYSESLAT scoring dimensionality tests produced results that substantiate the use of the two-factor model in reporting IRT scores. With the exception of the 7–8 and 9–12 grade spans, the two-factor model provided a similar fit of the model to the data as the one-factor unidimensional model as measured by global indices of fit. The unidimensional scoring model for the 7–8 and 9–12 grade spans, however, did not outweigh the use of the two-factor model in terms of model-fit and, as such, there was no clear preference of one model over the other for these two grade spans. This pattern, however, was justifiable by content experts as a distinct demarcation of modalities for younger students for whom a marked differentiation did exist between their Speaking and Listening ability as compared to their Reading and Writing ability. This differentiation, according to content experts, blurs for students in the higher grades. (See Tables 6.3-6.7 below; Appendix F provides the 2007 CFA analyses in their entirety.)

Table 6.3
Global Fit Indices for the One- and Two-Factor Models (Grades K-1)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	8070.680	1158.310
GFI	0.940	0.990
AGFI	0.820	0.960
RMR	0.057	0.015
RMSEA	0.180	0.077

Note: χ^2 with 5 *df*, $p=0.00$ for the one-factor model and χ^2 with 4 *df*, $p=0.00$ for the two-factor model.

Table 6.4
Global Fit Indices for the One- and Two-Factor Models (Grades 2-4)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	6712.14	2161.190
GFI	0.950	0.980
AGFI	0.860	0.940
RMR	0.041	0.025
RMSEA	0.150	0.098

Note: χ^2 with 5 *df*, *p*=0.00 for the one-factor model and χ^2 with 4 *df*, *p*=0.00 for the two-factor model.

Table 6.5
Global Fit Indices for the One- and Two-Factor Models (Grades 5-6)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	1931.250	1148.990
GFI	0.970	0.980
AGFI	0.920	0.940
RMR	0.027	0.023
RMSEA	0.120	0.100

Note: χ^2 with 5 *df*, *p*=0.00 for the one-factor model and χ^2 with 4 *df*, *p*=0.00 for the two-factor model.

Table 6.6
Global Fit Indices for the One- and Two-Factor Models (Grades 7-8)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	2003.490	2222.930
GFI	0.970	0.970
AGFI	0.900	0.880
RMR	0.031	0.031
RMSEA	0.130	0.150

Note: χ^2 with 5 *df*, *p*=0.00 for the one-factor model and χ^2 with 4 *df*, *p*=0.00 for the two-factor model.

Table 6.7
Global Fit Indices for the One- and Two-Factor Models (Grades 9-12)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	2172.820	2108.660
GFI	0.980	0.980
AGFI	0.930	0.910
RMR	0.029	0.029
RMSEA	0.110	0.120

Note: χ^2 with 2 *df*, *p*=0.00 for the one-factor model and χ^2 with 1 *df*, *p*=0.00 for the two-factor model.

Overall, there seemed to be a decrease in the precision of model-fit for the two-factor model regarding the unidimensional model as one progressed from lower grade spans to the higher ones.

6.3 External Structure (Relationship with the New York English Language Assessment)

In order to ascertain the external structure of the NYSESLAT vis-à-vis an external criterion, data on grades 3–8 English Language Learners (ELLs) who took the NYSESLAT in May 2007 and

the New York State English Language Assessment (ELA) in Jan. 2007 were analyzed. In New York State, all ELLs except for the first year ESL students are required to take the ELA test.

The purpose of the NYSESLAT is to measure annual student progress in achieving English language proficiency in order to ultimately exit an ESL or bilingual program. It is also expected that ELLs who score proficient on the NYSESLAT will be able to function successfully in mainstream education programs and have the same chance as their English proficient counterparts to pass the required state examinations. It is therefore hypothesized that a relationship exists between the NYSESLAT and the ELA test, i.e., those who perform well on the NYSESLAT are generally expected to perform well on the ELA test.

Correlations and Other Descriptive Statistics

Table 6.8 gives the sample size, minimum and maximum observed scores, means, standard deviations, and correlation between the New York ELA test and NYSESLAT raw score total of Reading, Writing, and Listening for each grade from Grade 3 to Grade 8. Table 6.9 gives the sample size, minimum and maximum observed scores, means, standard deviations, and correlation between the New York ELA test and NYSESLAT combined scale scores of Listening and Speaking and Reading and Writing for each grade from Grade 3 to Grade 8.

Table 6.8
Descriptive Statistics of the New York ELA and NYSESLAT Total Raw Scores of Reading, Writing, and Listening

Grade	N	Test	Mean	SD	Sample Min.	Sample Max.	Correlation
3	16854	NYSESLAT Raw	53.85	9.95	9	70	0.73
		NY ELA Scale Score	634.06	34.12	475	780	
4	13707	NYSESLAT Raw	57.03	9.34	13	70	0.75
		NY ELA Scale Score	625.57	37.97	430	775	
5	10857	NYSESLAT Raw	58.06	11.07	11	77	0.72
		NY ELA Scale Score	622.74	38.51	495	795	
6	9157	NYSESLAT Raw	59.11	11.36	6	77	0.72
		NY ELA Scale Score	621.78	27.49	480	785	
7	8393	NYSESLAT Raw	53.75	13.16	6	77	0.74
		NY ELA Scale Score	600.18	44.93	470	716	
8	9038	NYSESLAT Raw	56.01	13.30	6	77	0.73
		NY ELA Scale Score	606.25	33.91	430	726	

Table 6.9
Descriptive Statistics of the New York ELA and NYSESLAT Combined Scale Scores of Listening/Speaking and Reading/Writing

Grade	N	Test	Mean	SD	Sample Min.	Sample Max.	Correlation
3	16854	NYSESLAT Scale Score	1328.64	75.53	1024	1603	0.69
		NY ELA Scale Score	634.06	34.12	475	780	
4	13707	NYSESLAT Scale Score	1353.62	79.59	1036	1603	0.70
		NY ELA Scale Score	625.57	37.97	430	775	
5	10857	NYSESLAT Scale Score	1357.11	71.08	1040	1633	0.67
		NY ELA Scale Score	622.74	38.51	495	795	
6	9157	NYSESLAT Scale Score	1364.10	76.42	1040	1633	0.68
		NY ELA Scale Score	621.78	27.49	480	785	
7	8393	NYSESLAT Scale Score	1364.14	80.37	1012	1664	0.71
		NY ELA Scale Score	600.18	44.93	470	716	
8	9038	NYSESLAT Scale Score	1377.72	84.77	994	1664	0.71
		NY ELA Scale Score	606.25	33.91	430	726	

The correlations between New York ELA test and NYSESLAT raw score total of the three modalities (Reading, Writing, and Listening) ranged from 0.72 to 0.75 across the six grade spans. The correlations are positive between the two tests, which are within expectation because the modalities tested are Reading, Writing, and Listening for the New York ELA test, and the total raw scores include those three modalities also. The correlations between New York ELA test and NYSESLAT combined scale scores of the four modalities ranged from 0.67 to 0.71 across the five grade levels. The correlations in Table 6.9 are positive, but slightly lower than the correlations in Table 6.8, which is within expectation because the NYSESLAT total combined scale scores include the Speaking modality, whereas the New York ELA test does not include the speaking part.

Generally speaking, the correlation analyses provide evidence that the external structure check of the NYSESLAT with respect to the New York State ELA ascertains a positive relationship (moderate high) between the two examinations (see Table 6.8 and Table 6.9 above). This indicates that higher scores on the NYSESLAT are associated with higher scores on the ELA for each grade span.

Proficiency Level

In order to know the classification distribution at each proficiency level of both ELA and NYSESLAT tests, the percentage distributions of ELA proficiency level by NYSESLAT proficiency level at each grade level were presented in Tables 6.10–6.15.

Table 6.10 shows that for those Grade 3 students who were classified as beginning students on the NYSESLAT, 98.61% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as intermediate students on the NYSESLAT, 95.73% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 69.38% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 30.63% of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 30.47% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 69.52% of them were classified as proficient on the ELA test.

Table 6.11 shows that for those Grade 4 students who were classified as beginning students on the NYSESLAT, 98.57% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as intermediate students on the NYSESLAT, 97.12% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 79.95% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 20.25% of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 45.68% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 54.31% of them were classified as proficient on the ELA test.

Table 6.12 shows that for those Grade 5 students who were classified as beginning students on the NYSESLAT, 98.86% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as intermediate students on the NYSESLAT, 97.83% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 81.96% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 18.04% of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 43.92% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 56.08% of them were classified as proficient on the ELA test.

Table 6.13 shows that for those Grade 6 students who were classified as beginning students on the NYSESLAT, 98.47% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as intermediate students on the NYSESLAT, 98.93% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 89.62% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 10.38% of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 64.37% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 35.62% of them were classified as proficient on the ELA test.

Table 6.14 shows that for those Grade 7 students who were classified as beginning students on the NYSESLAT, 99.80% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as intermediate students on the NYSESLAT, 99.48% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 90.54% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 9.46% of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 61.24% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 38.76% of them were classified as proficient on the ELA test.

Table 6.15 shows that for those Grade 8 students who were classified as beginning students on the NYSESLAT, 99.50% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as intermediate students on the NYSESLAT, 98.96% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 93.49% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 6.51% of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 74.70% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 25.31% of them were classified as proficient on the ELA test.

Generally speaking, about 96%–99% of students who were classified as beginning and intermediate on the NYSESLAT test were also classified as not proficient (Level 1 and Level 2) by the ELA cut scores across the six grade levels. About 69%–93% of students who were classified as advanced on the NYSESLAT test were also classified as not proficient (Level 1 and Level 2) by the ELA cut scores across the six grade levels. About 7%–30% of students who were classified as advanced on the NYSESLAT test were classified as proficient by the ELA cut scores across the six grade levels. The higher the grade level, the lower the percentage. About 30%–75% of students who were classified as proficient on the NYSESLAT test were classified as not proficient by the ELA cut scores across the six grade levels. About 25%–69% of students who were classified as proficient on the NYSESLAT test were also classified as proficient by the ELA cut scores across the six grade levels. The higher the grade level, the lower the percentage. In general, all the evidence above indicates that those ELLs who scored well on the NYSESLAT also scored well on the ELA test, and those who did not score well on the NYSESLAT also did not perform well on the ELA test.

Table 6.10**Grade 3: Percentage of ELLs Scoring at Each ELA Level by NYSESLAT Proficiency Level**

	NYSESLAT Proficiency Levels			
	Beginning (N=1011)	Intermediate (N=3608)	Advanced (N=8750)	Proficient (N=3485)
Level 1 on ELA	90.6%	57.43%	12.21%	1.35%
Level 2 on ELA	8.01%	38.3%	57.17%	29.12%
Level 3 on ELA	1.38%	4.27%	30.34%	66.74%
Level 4 on ELA	0%	0%	0.29%	2.78%
Total	100%	100%	100%	100%

Table 6.11**Grade 4: Percentage of ELLs Scoring at Each ELA Level by NYSESLAT Proficiency Level**

	NYSESLAT Proficiency Levels			
	Beginning (N=770)	Intermediate (N=2959)	Advanced (N=6618)	Proficient (N=3360)
Level 1 on ELA	91.69%	71.54%	16.8%	2.32%
Level 2 on ELA	6.88%	25.58%	63.15%	43.36%
Level 3 on ELA	1.43%	2.84%	19.96%	53.48%
Level 4 on ELA	0%	0.03%	0.09%	0.83%
Total	100%	100%	100%	100%

Table 6.12**Grade 5: Percentage of ELLs Scoring at Each ELA Level by NYSESLAT Proficiency Level**

	NYSESLAT Proficiency Levels			
	Beginning (N=874)	Intermediate (N=2813)	Advanced (N=5322)	Proficient (N=1849)
Level 1 on ELA	87.53%	54.53%	10.6%	0.87%
Level 2 on ELA	11.33%	43.3%	71.36%	43.05%
Level 3 on ELA	1.03%	2.17%	18.04%	55.65%
Level 4 on ELA	0.11%	0%	0%	0.43%
Total	100%	100%	100%	100%

Table 6.13**Grade 6: Percentage of ELLs Scoring at Each ELA Level by NYSESLAT Proficiency Level**

	NYSESLAT Proficiency Levels			
	Beginning (N=848)	Intermediate (N=2704)	Advanced (N=4221)	Proficient (N=1384)
Level 1 on ELA	69.22%	24.67%	2.82%	0.14%
Level 2 on ELA	29.25%	74.26%	86.8%	64.23%
Level 3 on ELA	1.53%	1.04%	10.33%	34.83%
Level 4 on ELA	0%	0.04%	0.05%	0.79%
Total	100%	100%	100%	100%

Table 6.14**Grade 7: Percentage of ELLs Scoring at Each ELA Level by NYSESLAT Proficiency Level**

	NYSESLAT Proficiency Levels			
	Beginning (N=1510)	Intermediate (N=3088)	Advanced (N=2990)	Proficient (N=805)
Level 1 on ELA	90.26%	51.55%	9.7%	0.87%
Level 2 on ELA	9.54%	47.93%	80.84%	60.37%
Level 3 on ELA	0.2%	0.52%	9.46%	38.26%
Level 4 on ELA	0%	0%	0%	0.5%
Total	100%	100%	100%	100%

Table 6.15**Grade 8: Percentage of ELLs Scoring at Each ELA Level by NYSESLAT Proficiency Level**

	NYSESLAT Proficiency Levels			
	Beginning (N=1582)	Intermediate (N=3266)	Advanced (N=2795)	Proficient (N=1395)
Level 1 on ELA	88.94%	50.55%	10.81%	1.29%
Level 2 on ELA	10.56%	48.41%	82.68%	73.41%
Level 3 on ELA	0.51%	1.04%	6.51%	24.95%
Level 4 on ELA	0%	0%	0%	0.36%
Total	100%	100%	100%	100%

Summary

The correlation analyses provide evidence of a positive relationship between the NYSESLAT and the ELA assessment (ranging from 0.67 to 0.75). This indicates that higher scores on the NYSESLAT are associated with higher scores on the ELA. The ELA test and the NYSESLAT tests were developed to serve different populations of students. Thus, it is not surprising that we see only a moderate, rather than a high, positive correlation (not high) between the two tests.

The analysis also confirms the validity of the NYSESLAT cut score in that about 96% to 99% (across six grade levels) of those who scored at the proficiency level on NYSESLAT test also

were classified as Level 2 or Level 3 (proficient) by the ELA test. On the other hand, the high percentages of students who were classified as not proficient by both tests is a strong indication that students who cannot do well on one test also did not perform well on the other test.

6.4 External Structure (Relationship with the New York State Regents Examination in Comprehensive English)

In order to ascertain the external “structure” of the New York State English as a Second Language Achievement Test (NYSESLAT) vis-à-vis an external criterion, data for grades 10 through 12 English Language Learners (ELLs) who took the NYSESLAT in May 2007 and the New York State Regents Examination in Comprehensive English (Regents English exam) in June 2007, were analyzed. The Regents English exam is an end of the course exam. All high school students, including ELLs, are required to pass the Regents English exam for high school graduation requirement. Students can take the exam in any grade, although most students take the exam in grade 11.

The purpose of the NYSESLAT is to measure annual student progress in achieving English language proficiency in order to ultimately exit an ESL or bilingual program. It is also expected that ELLs who score proficient on the NYSESLAT will be able to function successfully in mainstream education programs and to have a comparable chance as their English proficient counterparts to pass the required State examinations. It is, therefore, hypothesized that a relationship exists between the NYSESLAT and the Regents English exam, i.e., those who perform well on the NYSESLAT, are generally expected to perform well on the Regents English exam.

Correlations and Other Descriptive Statistics

Table 6.16 shows the sample size, maximum score points, minimum and maximum observed scores, means, standard deviations, and correlation between the New York Regents English exam and each of the NYSESLAT modalities raw score, and the total raw scores of Reading, Writing, and Listening for grades 10 through 12. Table 6.17 presents the sample size, minimum and maximum observed scores, means, standard deviations, and correlation between the New York Regents English exam and the combined scale scores of Listening/Speaking and Reading/Writing, and total scale scores of the four modalities (Listening/Speaking/Reading/Writing) for grades 10 through 12.

Table 6.16
Descriptive Statistics of the NYSESLAT Raw Scores of Listening, Speaking,
Reading, and Writing Modalities and their Correlations with the Regents English
Exam Total Scale Scores

Grade	N	Test	Max Points	Mean	SD	Sample Min.	Sample Max.	Correlation with Regents English
10	2240	Listening	25	18.91	3.99	1	25	0.49
		Speaking	34	28.20	7.27	0	34	0.34
		Reading	27	19.21	4.58	2	27	0.54
		Writing	25	17.76	5.12	0	25	0.53
		LI/RD/WR Total	77	55.88	11.69	13	77	0.61
11	6728	Listening	25	19.45	3.69	2	25	0.41
		Speaking	34	28.80	6.34	0	34	0.28
		Reading	27	20.04	4.33	2	27	0.51
		Writing	25	18.64	4.73	0	25	0.51
		LI/RD/WR Total	77	58.12	10.68	10	77	0.58
12	3827	Listening	25	19.08	3.64	2	25	0.36
		Speaking	34	28.22	6.98	0	34	0.20
		Reading	27	19.61	4.37	3	27	0.41
		Writing	25	17.90	4.67	0	25	0.41
		LI/RD/WR Total	77	56.58	10.38	20	76	0.48
10-12	12848	Listening	25	19.25	3.73	1	25	0.41
		Speaking	34	28.52	6.71	0	34	0.27
		Reading	27	19.77	4.40	2	27	0.49
		Writing	25	18.26	4.80	0	25	0.48
		LI/RD/WR Total	77	57.27	10.82	10	77	0.55

Table 6.17
Descriptive Statistics of the NYSESLAT Scale Scores of Listening/Speaking, Reading/Writing Modalities, and Total Scale Scores and their Correlations with the Regents English Exam Total Scale Scores

Grade	N	Test	Mean	SD	Sample Min.	Sample Max.	Correlation with Regents English
10	2240	Regents English	40.93	20.54	0	93	-
		Listening and Speaking	704.39	45.95	471	842	0.47
		Reading and Writing	704.26	35.10	593	859	0.59
		Total NYSESLAT	1408.65	73.39	1126	1652	0.57
11	6728	Regents English	51.03	19.28	0	98	-
		Listening and Speaking	708.32	41.62	534	842	0.39
		Reading and Writing	711.39	34.71	559	859	0.58
		Total NYSESLAT	1419.71	67.50	1155	1701	0.54
12	3827	Regents English	51.96	17.23	0	95	-
		Listening and Speaking	703.68	42.35	497	842	0.32
		Reading and Writing	705.81	32.60	599	859	0.47
		Total NYSESLAT	1409.49	66.06	1152	1652	0.44
10-12	12848	Regents English	49.54	19.34	0	98	-
		Listening and Speaking	706.26	42.69	471	842	0.38
		Reading and Writing	708.47	34.29	559	859	0.55
		Total NYSESLAT	1414.73	68.35	1126	1701	0.51

The correlations between the scale scores of the Regents English exam and the NYSESLAT raw score total of the three modalities (Reading, Writing, and Listening) ranged from 0.48 to 0.61 across the three grade levels, which is higher than the correlations between the scale scores of the Regents English exam and any modality of the NYSESLAT raw scores. Additionally it is within expectation that the correlations between the scale scores of the Regents English exam and the NYSESLAT combined raw scores of Reading/Writing are higher than the correlations between the scale scores of the Regents English exam and the NYSESLAT combined raw scores of Listening/Speaking.

The correlations between the scale scores of the Regents English exam and the NYSESLAT combined scale scores of the four modalities ranged from 0.44 to 0.57 across the three grade levels. The correlations in Table 6.17 are positive, but slightly lower than the correlations in Table 6.16, which are within expectation because the NYSESLAT total combined scale scores include the Speaking modality, whereas the Regents English exam does not include the speaking component. Additionally it is within expectation that the correlations between the scale scores of the Regents English exam and the NYSESLAT combined scale scores of Reading/Writing are higher than the correlations between the scale scores of the Regents English exam and the NYSESLAT combined scale scores of Listening/Speaking.

Generally speaking, the correlation analyses provide evidence that the external structure check of the NYSESLAT, with respect to the Regents English exam, ascertains a positive relationship (moderate) between the two examinations for grades 10–12.

Proficiency Level Classification

The classification distribution in relation to passing/failing the Regents English exam by NYSESLAT proficiency level was also examined. The results are presented in Tables 6.18–6.20.

Table 6.18 shows that 49.4 percent of the grade 10 students who scored proficient on the NYSESLAT passed the Regents English exam. The percentage of students who passed the Regents English exam decreased to 20.2 percent for students at the advanced level, and around five percent for students at the intermediate and beginning level

Table 6.18
Grade 10: Percentage of ELLs Scoring at each of NYSESLAT Proficiency Level Passing/Failing the Regents English Exam

NYSESLAT Proficiency Levels				
Regents English Exam	Beginning (N=180)	Intermediate (N=948)	Advanced (N=794)	Proficient (N=318)
Pass	5.60%	5.20%	20.20%	49.40%
Fail	94.40%	94.80%	79.80%	50.60%
Total	100%	100%	100%	100%

Table 6.19 shows that the grade 11 students who were classified as proficient on the NYSESLAT, 66.1 percent passed the Regents English exam. For those students who were classified as advanced students on the NYSESLAT test, 43.3 percent passed. For those students who were classified as beginning and intermediate levels on the NYSESLAT, the percentage of students passing the Regents English were 11.8 percent and 17.2 percent respectively

Table 6.19
Grade 11: Percentage of ELLs Scoring at each of NYSESLAT Proficiency Level Passing/Failing the Regents English Exam

NYSESLAT Proficiency Levels				
Regents English Exam	Beginning (N=408)	Intermediate (N=3504)	Advanced (N=1984)	Proficient (N=832)
Pass	11.80%	17.20%	43.30%	66.10%
Fail	88.20%	82.80%	56.70%	33.90%
Total	100%	100%	100%	100%

Table 6.20 shows that the grade 12 students who were classified as proficient on the NYSESLAT, 56.5 percent passed the Regents English exam. For those students who were classified as advanced students on the NYSESLAT test, 41.3 percent passed. For those students who were classified as beginning and intermediate levels on the NYSESLAT, the percentage of students passing the Regents English were 13.5 percent and 22.9 percent respectively

Table 6.20
Grade 12: Percentage of ELLs Scoring at each of NYSESLAT Proficiency Level
Passing/Failing the Regents English Exam

Regents English Exam	NYSESLAT Proficiency Levels			
	Beginning (N=342)	Intermediate (N=2275)	Advanced (N=858)	Proficient (N=352)
Pass	13.50%	22.90%	41.30%	56.50%
Fail	86.50%	77.10%	58.70%	43.50%
Total	100%	100%	100%	100%

Summary

The correlation analyses provide evidence of a moderate and positive relationship between the NYSESLAT and the Regents English exam. The evidence suggests that the higher NYSESLAT scores are associated with higher scores on the Regents English exam required for high school graduation.

The analyses also confirm the validity of the NYSESLAT cut score in that ELLs who were classified as proficient on the NYSESLAT had higher chance of passing the Regents English exam than ELLs who were below the proficient level on the NYSESLAT.

PART 7: CALIBRATION, EQUATING, AND SCALING

This part introduces the IRT models and the processes that were used to calibrate, equate, and scale the NYSESLAT using these IRT models.

The Rasch model (Rasch, 1960) for dichotomous items and the Partial Credit Model (PCM) (Masters, 1982) for polytomous items were used to calibrate, equate, and scale the NYSESLAT. These measurement models are regularly used to construct test forms, for scaling and equating, and to develop and maintain large item banks. All item and test analyses, including item-fit analysis, scaling, equating, diagnosis, and performance prediction were accomplished within this framework. The statistical software used to calibrate and scale the NYSESLAT was *Winsteps* Version 3.29 (Linacre & Wright, 2000).

7.1 The Rasch and Partial Credit Models

Description of the Rasch Model

Mathematically, the Rasch model (Rasch, 1980), also known as the one-parameter logistic (1-PL) model can be formulated as:

$$P(U_{ij} = 1/\theta) = \frac{1}{1 + \exp\{-(\theta - \delta_i)\}} \quad (17)$$

Where

δ_i is item difficulty, and

θ is examinee ability.

The most basic expression of the Rasch model, however, is in the Item Characteristic Curve (ICC). It shows the probability of a correct response to an item as a function of the ability level. The probability of a correct response is bounded by 1 (correct response) and 0 (incorrect response). The ability scale is, in theory, unbounded. In practice, the ability scale typically ranges from -4 to +4 logits for heterogeneous ability groups where the logit (log-odds unit) is nothing more than a unit of interval measurement which is well-defined within the context of a single homogeneous test in the same manner that the height of a person is measured in feet or meters.

As an example, consider Figure 7.1, which depicts an item that falls at approximately 0.85 on the ability (horizontal) scale. When a person answers an item at the same level as their ability, then that person has a probability of roughly 50% of answering the item correctly. Another way of expressing Figure 7.1 is that if we have a group of 100 people, all of whom have an ability of 0.85, we would expect about 50% of them to answer the item correctly. A person whose ability was above 0.85 would have a higher probability of getting the item right, while a person whose ability is below 0.85 would have a lower probability of getting the item right. This is the basic formulation of Rasch measurement for test items having only 2 possible categories (i.e., wrong or correct).

Figure 7.2 extends this formulation to show the probabilities of obtaining a wrong answer or a correct answer. The curve on the left ($j = 0$) shows the probability of getting a score of “0,” while the curve on the right ($j = 1$) shows the probability of getting a score of “1.” The point at which

the two curves cross indicates the transition point on the ability scale where the most likely score on the item changes from a “0” to a “1.” Here, the probability of answering the item correctly is 50%.

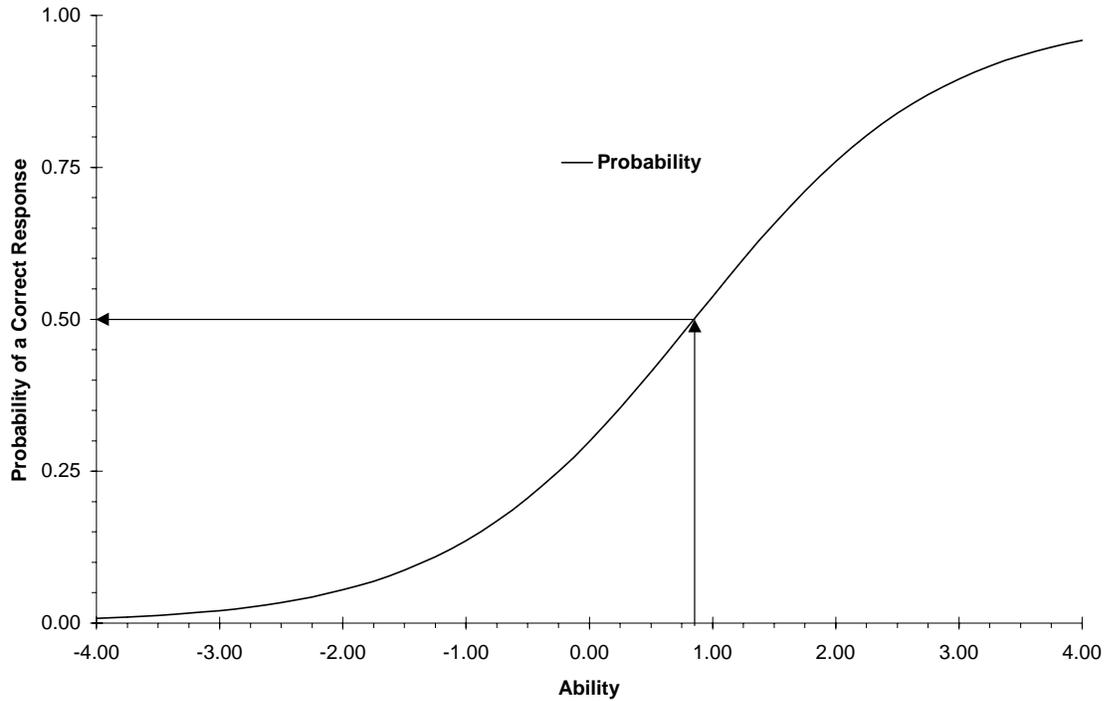


Figure 7.1: Sample Item Characteristic Curve

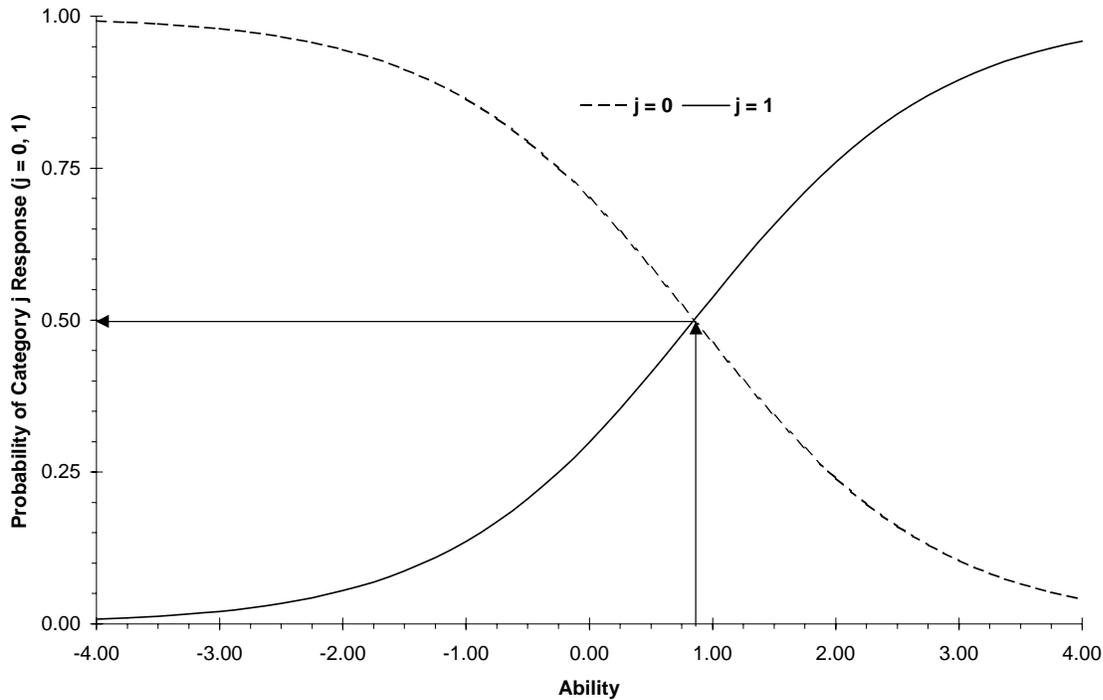


Figure 7.2: Category Response Curves for a One-step Item

Description of the Partial Credit Models (PCM)

The PCM is a direct extension of the dichotomous one-parameter IRT model developed by Rasch in the 1950s (Rasch, 1980). For an item/task involving m_i score categories, one general expression for the probability of scoring x on item/task i is given by

$$P_{xi} = \frac{\exp \sum_{j=0}^x (\theta - D_{ij})}{\sum_{k=0}^{m_i} \left[\exp \sum_{j=0}^k (\theta - D_{ij}) \right]} \quad (18)$$

where $x = 0, 1, \dots, m_i$, and by definition,

$$\sum_{j=0}^0 (\theta - D_{ij}) = 0$$

The above equation gives the probability of scoring x on the i -th test item as a function of ability (θ) and the difficulty of the m_i steps of the task (Masters, 1982).

According to this model, the probability of an examinee scoring in a particular category (step) is the sum of the logit (log-odds) differences between θ and D_{ij} of all the completed steps, divided by the sum of the differences of all the steps of a task. Thissen and Steinberg (1983) referred to this model as a divide-by-total model. The parameters estimated by this model are (a) an ability estimate for each person (or ability estimate at each raw score level) and (b) m_i threshold (difficulty) estimates for each task with $m_i + 1$ score categories.

The key step in the formulation and the point at which the Rasch dichotomous model merges with the PCM, requires us to assume an additional score category. Suppose that, rather than scoring items as completely wrong or completely correct, we add a category representing

answers that, though not totally correct, are still clearly not totally incorrect. These relationships are shown in Figure 7.3.

The left-most curve ($j = 0$) in Figure 7.3 represents the probability for all examinees getting a score of “0” (completely incorrect) on the item, given their ability. Those of very low ability (e.g., below -2) are very likely to be in this category and, in fact, are more likely to be in this category than the other two. Those receiving a “1” (partial credit) tend to fall in the middle range of abilities (the middle curve, $j = 1$). The final, right-most curve ($j = 2$) represents the probability for those receiving scores of “2” (completely correct). Anyone has a chance of getting full credit for the item, however, we may not expect everyone to receive full credit based on our data.

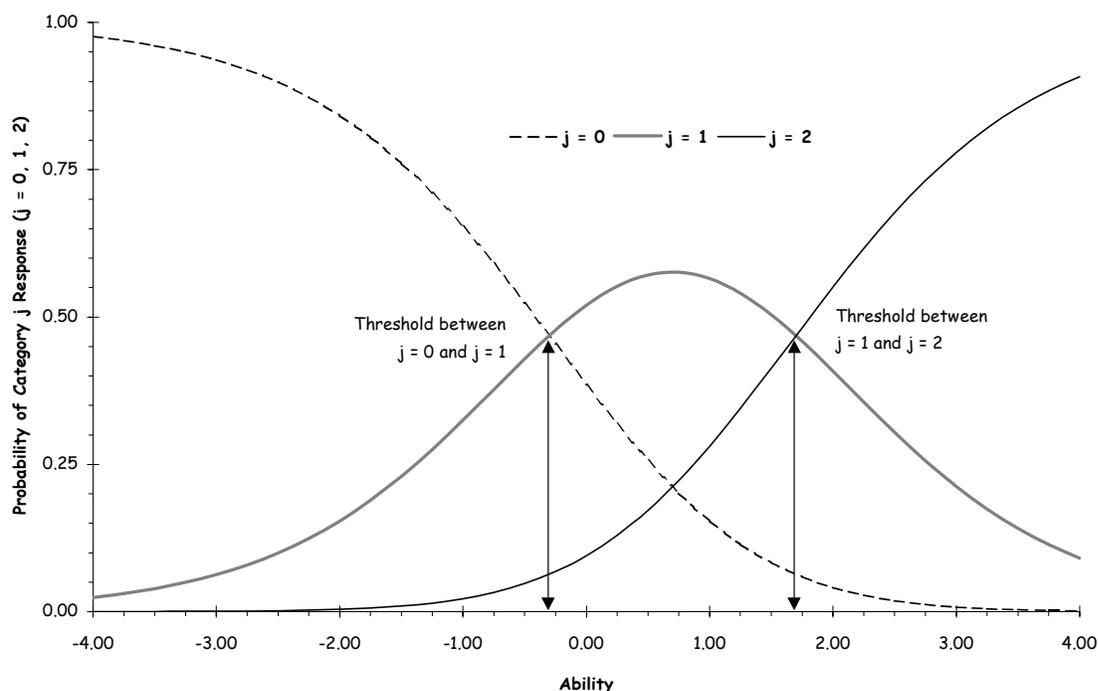


Figure 7.3: Category Response Curves for a Two-step Item

Although the actual computations are quite complex, the points at which lines cross each other have a similar interpretation as for the dichotomous case. Consider the point at which the $j = 0$ line crosses the $j = 1$ line, indicated by the left arrow. For abilities to the left of (or less than) this point, the probability is greatest for a “0” response. To the right of (or above) this point, and up to the point at which the $j = 1$ and $j = 2$ lines cross (marked by the right arrow), the most likely response is a “1.” For abilities to the right of this point, the most likely score is a “2.”

Scaling multiple-choice items together with polytomous items, whether they have three or more response categories, is a straightforward process of applying the appropriate measurement model, i.e., the Rasch model for the dichotomously scored items and the PCM for the polytomously scored items. The quality of the scaling then can be assessed in terms of known procedures that are applicable to the two measurement models discussed above.

One important property of these models is their ability to separate the estimation of item/task parameters from the person parameters. The total score given by the sum of the categories in which a person responds is a sufficient statistic for estimating person ability (i.e., no additional

information need be estimated). The total number of scores across examinees in a particular category is a sufficient statistic for estimating the step difficulty for that category. Thus with the Rasch and the PCM, the same total score will yield the same ability estimate for different examinees.

7.2 Calibration and Equating of the NYSESLAT

Equating Method and Rationale

In 2007, the 2006 NYSESLAT vertical scale was used to establish the 2007 NYSESLAT scale. In 2006, the 2005 NYSESLAT vertical scale was used to establish the 2006 NYSESLAT scale. The 2005 NYSESLAT administration was selected from the Stanford English Language Proficiency Test (SELP) examinations and the SELP scale has a well-established record as being a stable scale. The common item, non-equivalent group design, was used for 2007 and 2006 equating because it is one of the most practical of the various designs used for equating. This design does not, for example, require large sample sizes necessary for the random groups design or a concern toward student fatigue by using the single group design (Kolen and Brennan, 1995), which makes the common items non-equivalent group design a very popular choice for large scale assessments.

The common-item non-equivalent group design necessitates the inclusion of items that are common to both administrations that would allow the linking of the tests and/or forms to be placed on the same scale. There are several methods used for this item response theory (IRT) scale transformation. Kolen and Brennan (2004) provide several IRT scale transformation methods that include the Mean/Mean, Mean/Sigma, Haebara, and the Stocking-Lord methods. For the Rasch model, the equating constant of scale transformation is often used by researchers and practitioners alike. A fairly common practice among psychometricians is to use a method called “the anchor or the fixed parameter method.” In this method, the common items whose calibrations are known (either through a previous-year calibration or a separate calibration of the form) are anchored or fixed to their known estimates during calibration of the other forms that are to be put on the scale of the first form. In treating these common item parameters as known, they are not estimated, and the remaining item parameters (for the uncommon items) are forced onto the same scale as the anchored (fixed) items (Hanson & Béguin, 1999; Linacre, 2003; Taherbhai & Seo, 2007). For the NYSESLAT, the fixed method of scale transformation was used. This transformation was achieved with the *Winsteps* program that calibrated the items with the use of the two IRT models discussed above.

As Kolen and Brennan (1995) state, “After the item parameters are on the same scale, IRT true score equating can be used to relate number-correct scores on Form X and Form Y. In this process, the true score associated on one form associated with a given θ is considered to be equivalent to the true score on another form associated with that θ (p. 175). Furthermore, observed scores can easily be substituted for true scores because doing so has been justified in IRT by the observance that the true score conversions are similar to observed score conversions (Lord & Wingersky, 1984). True score equating was used for the NYSESLAT by the application of the Newton-Raphson method for finding the roots of the non-linear function associated in the true score equating process. This procedure allows observed scores to be matched to their respective theta values, which in turn are converted to more acceptable scale score values (see scale score transformation to the SELP scale equations on p. 56). These thetas for the different forms/tests are on the same scale as established by the scale transformation method discussed

above. By applying the true score equating process, the 2007 NYSESLAT was successfully equated to the 2006 NYSESLAT while maintaining the established 2006 common vertical scale, which was based on the 2005 SELP scale.

In establishing the 2007 vertical scale, the data file used for the calibration and equating of the NYSESLAT had 161,358 cases, which was 84% of the final operational research file (total N=192,053) that Pearson received from New York. The calibration/equating data was representative of the population covering New York City, and the Big Four Cities (Buffalo, Syracuse, Rochester, and Yonkers).

Thus by fixing the known parameters of the common set of items between the two years, the items on the 2007 operational form were calibrated, and the newly administered items were then located on the 2006 scale. Once the scale locations of the 2007 NYSESLAT were known, IRT true score equating was used to relate the raw scores on the 2007 NYSESLAT to the 2006 scale. In this process, the true score on the NYSESLAT with a given level of examinee ability was considered to be an equivalent true score on the 2006 test associated with that level of examinee ability (Kolen and Brennan, 2004, p. 178).

Common Item Linking Design

As explained in the previous section, the common item design is contingent upon the use of items that are common on two administrations, i.e., between the examination that is already placed on the established scale and the one that is to be placed on that scale. Alternatively, one can use items on the current examination that have already been placed on an established scale. These items then act as anchor items to place the new form on the existing scale.

For the NYSESLAT, Pearson used the pre-existing 2006 vertical scale³ that was used to create the NYSESLAT vertical scale of 2007. For the 2007 administration, the common items between 2006 and 2007 were the anchor items on the 2007 NYSESLAT and were, as explained earlier, fixed to the parameter values from the pre-existing vertical scale of 2006. That is, the common items between 2006 and 2007 were used as anchor items between the two years. Any remaining non-common items on the NYSESLAT were calibrated together with the anchor items using the Rasch and Partial Credit models. Fixing the values of the anchor items prior to calibration resulted in the item difficulty and step parameters of all the items being placed on the same ability metric.

Description of Calibration Sample, Process, and Results

Calibration Sample

An equating sample (80% of the population) that was used for the equating analyses represented the ELL population in all six need/resource categories: NYC, Big Four, and High Need Urban/Suburban, Rural, Average Need, and Low Need districts.

³ For additional details of how the 2006 vertical scale was established, please see the *2006 NYSESLAT Technical Manual*. Harcourt Assessment, Inc.

Calibration Process

The separate scales, one for Speaking/Listening and one for Reading/Writing, were then obtained by taking the item parameters for those separate combinations and using them to create raw score-to-scale score tables. Finally, when these calibrations and scales were completed, the items field tested for the 2007 administration was calibrated to the 2006 vertical scale.

A more detailed outline of the procedure follows:

- The *Winsteps* software program was used to conduct the item calibration, fixing the item parameters for the anchor set items to the 2006 scale values.
- The results of this calibration were used as the operational item parameters to create the preliminary scales for the NYSESLAT spring 2007 administration. Item estimates of the common items from the 2006 administration were used as anchors to place the 2007 scale to the established 2006 vertical scale. The final reporting scale was derived from the vertical scale using a linear transformation.

Anchor Item Evaluation

There are some general guidelines (2007 NYSESLAT Form Construction Specifications) in the use of anchor items. These guidelines are as follows:

- Use an adequate number of anchor items.
- Anchor items selected for linking the two forms need to represent a miniature version of the test in terms of content.
- Anchor items need to appear in the same or similar location as they do in the test that we are equating to.
- Anchor items are expected to have strong statistical properties (e.g., $p_{bis} > 0.30$) and span the range of difficulty including easy, medium, and difficult items. The mean difficulty (p-value/Rasch difficulty values) of the 2007 anchor items should be very similar to the mean difficulty of the 2006 operational test.

Stability of Anchor Items

The stability of the common item parameters was examined prior to their use as anchor items in the equating. There are various methods of evaluating anchor item stability. Most of them are based on how much of a difference exists between the parameters on the established scale and those on the transformed scale. Since the *Winsteps* program provides “Displacement” values for the common items that are anchored, a check of these values can be used to exclude items from being anchor items. Items that exceed the displacement value of 0.30, for example, can be excluded as anchor items and the process is repeated until all displacement values are less than or equal to 0.30. Another criteria is that the total number of anchor items need to be at least 20% of the total number of items for each modality combination. If the 20% anchor item criteria are not met, then a slightly higher displacement value of 0.50 will be used. There are no true or correct procedures that can be applied in excluding common items as anchor items. A balance has to be determined between the minimum items desired in the anchor pool and the number of common items that are excluded from being anchor items. However, as a general rule, since the number of anchor items used in equating is an important criterion in the stability of the linking procedure, many psychometricians recommend no more than 20% of the common items be excluded from being anchor items, e.g., the PACT Assessment Huynh (2000).

For the 2007 NYSESLAT, we used the displacement value of 0.30 (as is done by many large-scale assessments) for an item to be considered as an outlier. We used this criteria consistently across all the grade spans. For all grade spans, the overall numbers of anchor items were around the 20% or close to 20% of the total items. See Table 7.1 for the number of items excluded for each modality combination by grade span.

Items that were excluded from being anchor items had their parameters estimated and the new estimate became the adjusted item parameter in the item bank for future use.

Table 7.1
Common Items as Anchor Items by Modality Combination and Grade Span

Grade Span	Modality Combination	Total Number of Items	Common Items ¹ (2006 and 2007)	Percentage of Total Items	Anchor Items (2006 and 2007)	Percentage of Total Items
K-1	Listening/Speaking	40	10	25	9	23
	Reading/Writing	30	6	20	5	17
2-4	Listening/Speaking	40	11	28	10	25
	Reading/Writing	40	9	23	7	18
5-6	Listening/Speaking	41	10	24	8	20
	Reading/Writing	46	10	22	9	20
7-8	Listening/Speaking	41	12	29	11	27
	Reading/Writing	46	10	22	8	17
9-12	Listening/Speaking	41	11	27	11	27
	Reading/Writing	46	10	22	9	20

Note 1. Common items are the same items appearing in both 2006 and 2007 operational tests.

Calibration Results

TCC and Standard Error Curves Comparison

In order to ascertain the maintenance of the scale across years, comparisons between the 2005, 2006, and 2007 administrations' Test Characteristic Curves (TCC) and Standard Error Curves (SE Curves) were examined with respect to their alignment at each level of administration. For TCCs, the x-axis represents the scale scores, with the y-axis representing the expected number correct or proportion correct. TCCs represent the relative difficulty of a given test form, with TCCs on the right representing more difficult test forms. Measurement error is associated with every test score. The standard error of measurement (SEM) can be used to obtain a range within which a student's true score is likely to fall. The x-axis represents the scale scores, with the y-axis representing the SEM.

All the 2005, 2006, and 2007 TCC plots (See Figures 7.4 & 7.5) indicate that the TCC shifts to the right with each progression of grade span across the Listening/Speaking and Reading/Writing vertical scales. The shift to the right with the corresponding increase in grade span clearly indicates that both the Listening/Speaking and Reading/Writing tests are ordinal across the grade spans (see Figures 7.4 & 7.5). The TCC alignment at each grade span seems within our realm of acceptance (See Figures 7.6-7.10 & 7.16-7.20 for all the 2005, 2006, and 2007 plots). Overall, the alignment of TCC among 2005, 2006, and 2007 align well and the TCCs align better for the middle and higher portion of the curve than at the tail portion of the curves.

The differences in terms of the SE curves among 2005, 2006, and 2007 for both the Listening/Speaking and Reading/Writing tests are very minimal across all the grade spans (See Figures 7.11–7.15 & 7.21–7.25). The noticeable differences occur only at the tail portion of the SE curves.

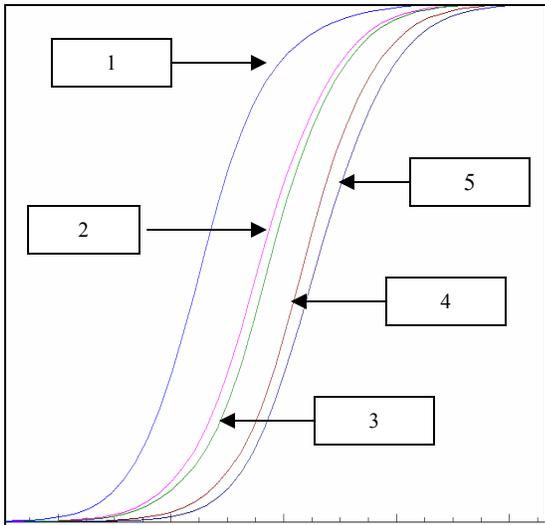


Figure 7.4
2007 TCC LI/SP Grade Span 1-5.

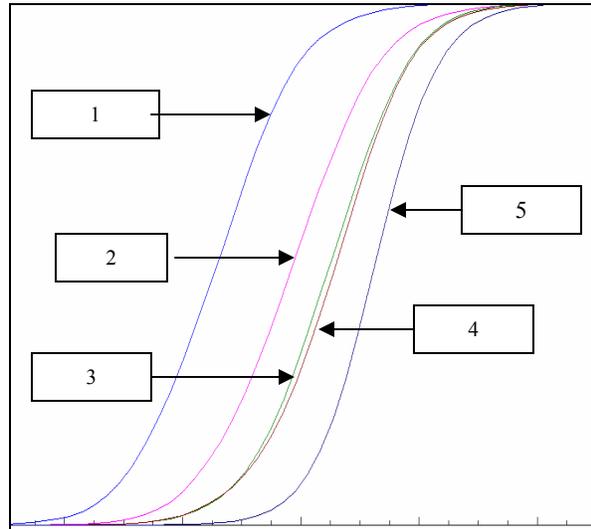


Figure 7.5
2007 TCC RD/WR Grade Span 1-5

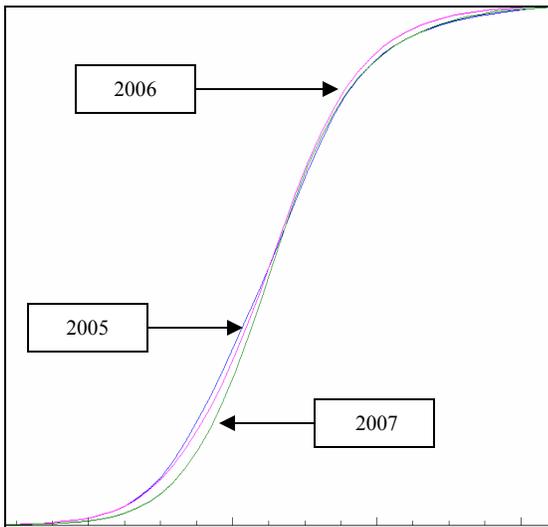


Figure 7.6
2005-06-07 TCC LI/SP Grade Span 1

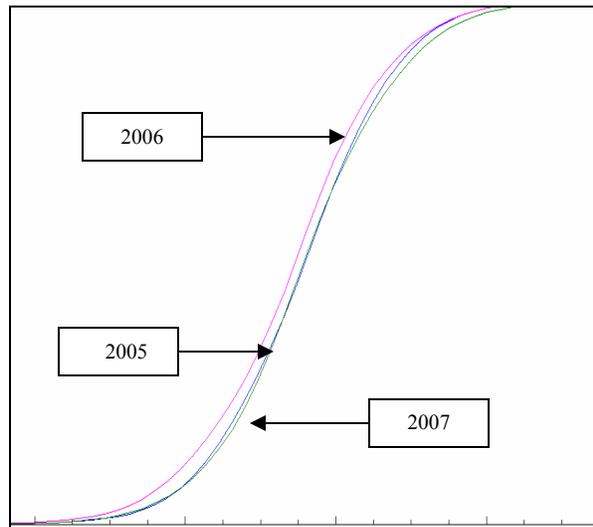


Figure 7.7
2005-06-07 TCC LI/SP Grade Span 2

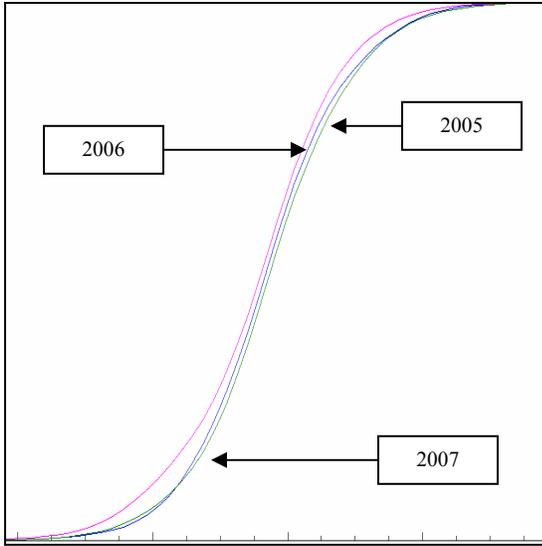


Figure 7.8
2005-06-07 TCC LI/SP Grade Span 3

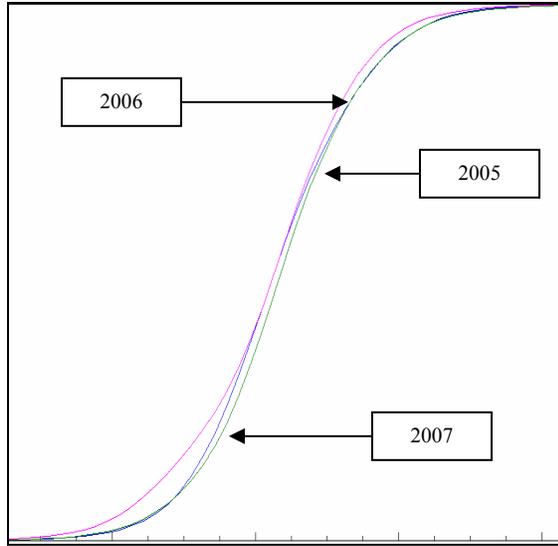


Figure 7.9
2005-06-07 TCC LI/SP Grade Span 4

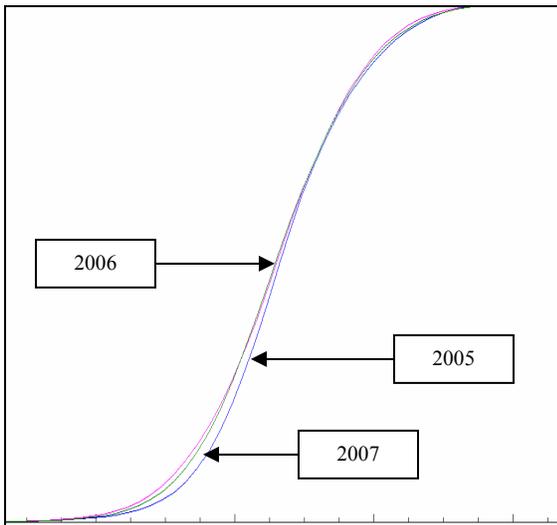


Figure 7.10
2005-06-07 TCC LI/SP Grade Span 5

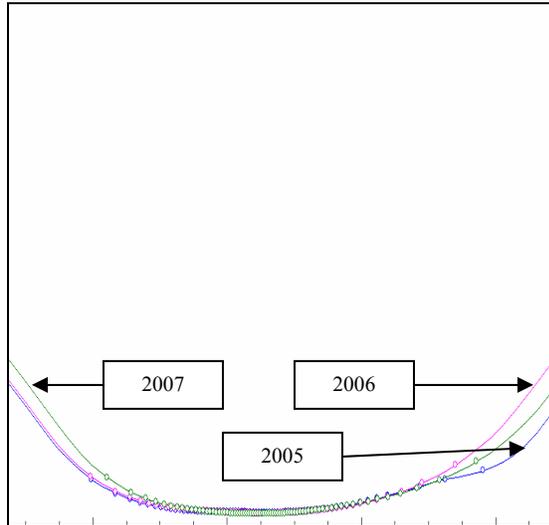


Figure 7.11
2005-06-07 SE Curves LI/SP Grade Span 1

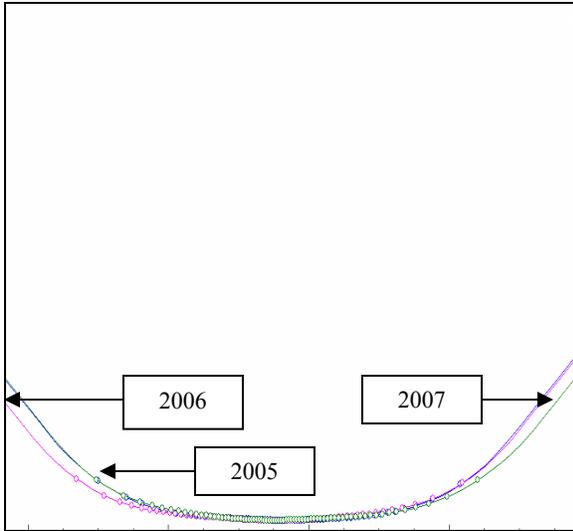


Figure 7.12
2005-06-07 SE Curves LI/SP Grade Span 2

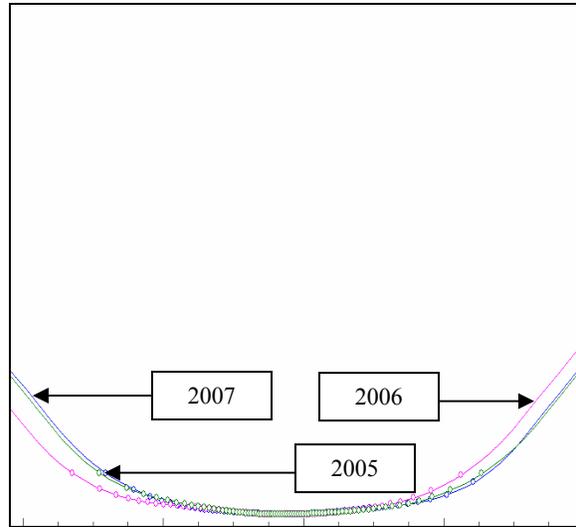


Figure 7.13
2005-06-07 SE Curves LI/SP Grade Span 3

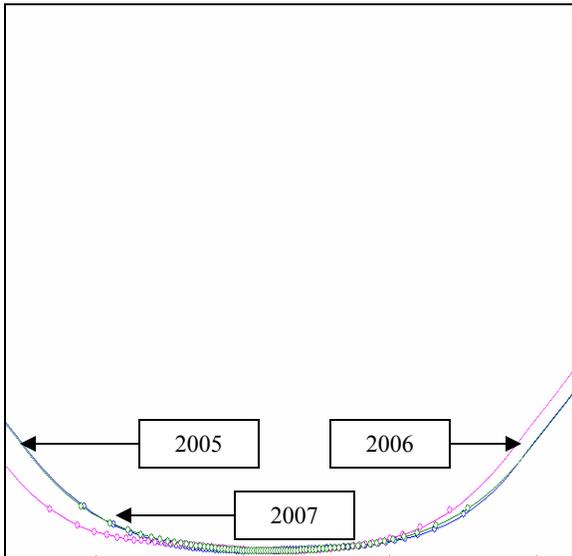


Figure 7.14
2005-06-07 SE Curves LI/SP Grade Span 4

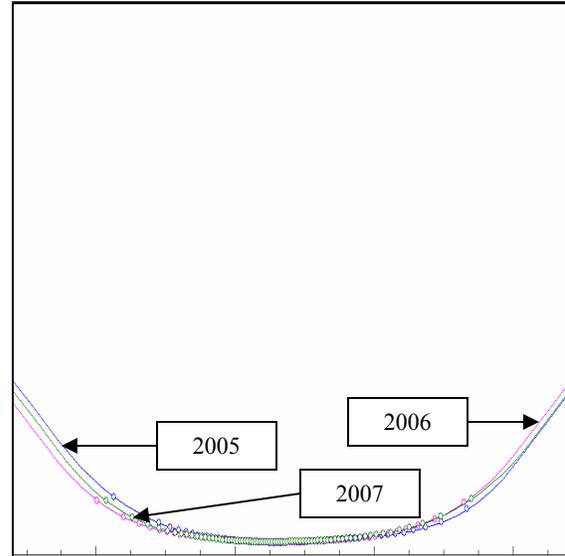


Figure 7.15
2005-06-07 SE Curves LI/SP Grade Span 5

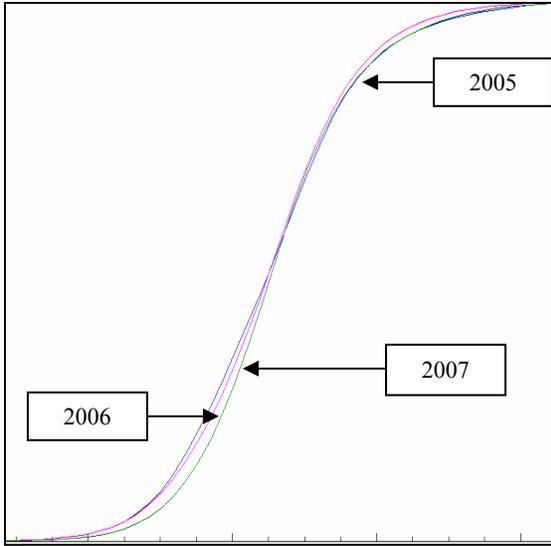


Figure 7.16
2005-06-07 TCC RD/WR Grade Span 1

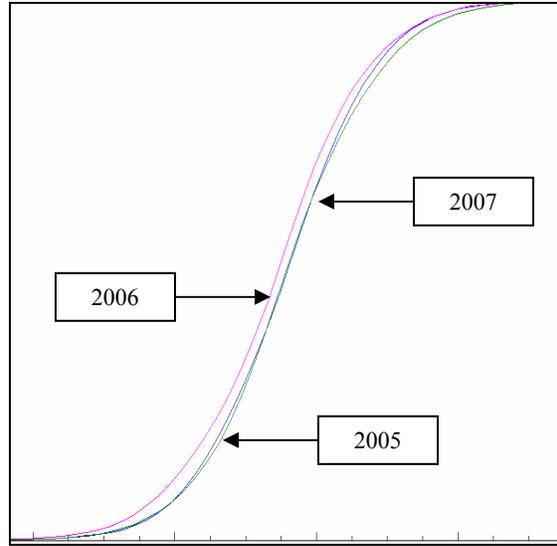


Figure 7.17
2005-06-07 TCC RD/WR Grade Span 2

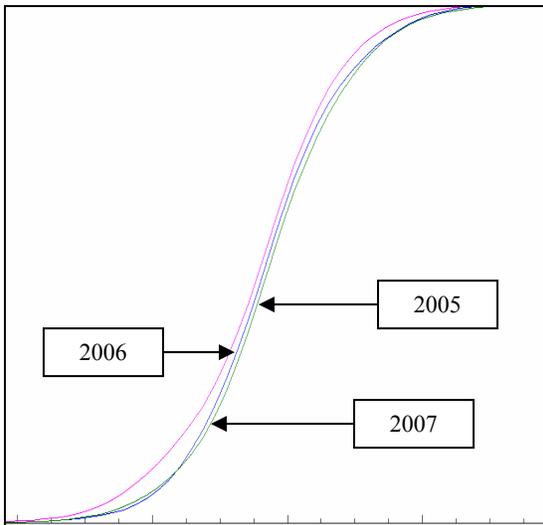


Figure 7.18
2005-06-07 TCC RD/WR Grade Span 3

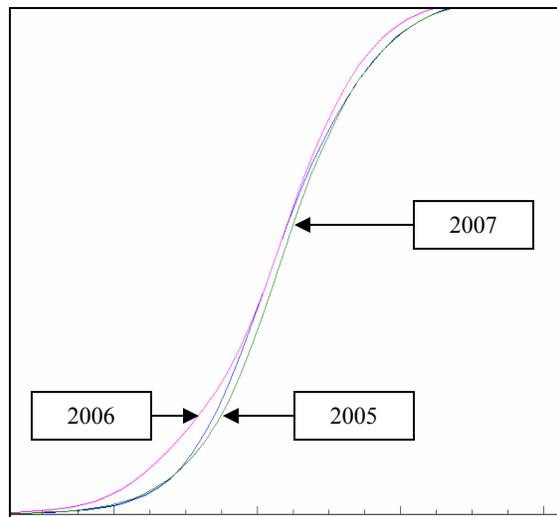


Figure 7.19
2005-06-07 TCC RD/WR Grade Span 4

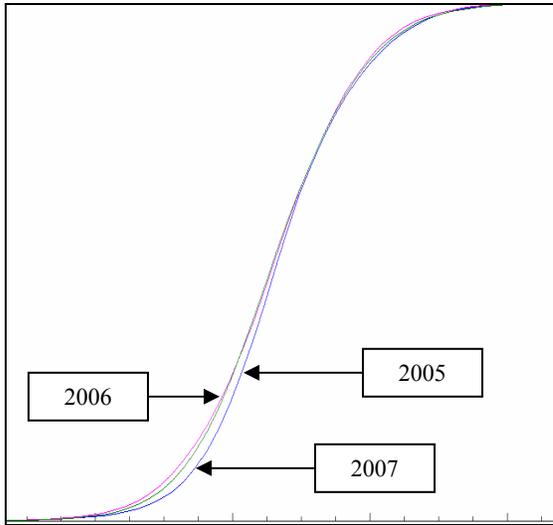


Figure 7.20
2005-06-07 TCC RD/WR Grade Span 5

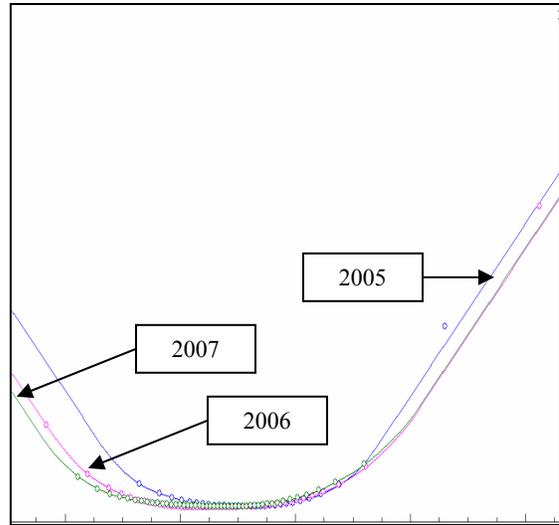


Figure 7.21
2005-06-07 SE Curves RD/WR Grade Span 1

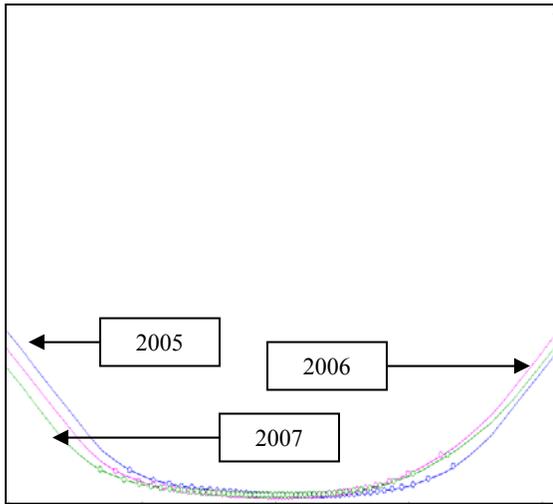


Figure 7.22
2005-06-07 SE Curves RD/WR Grade Span 2

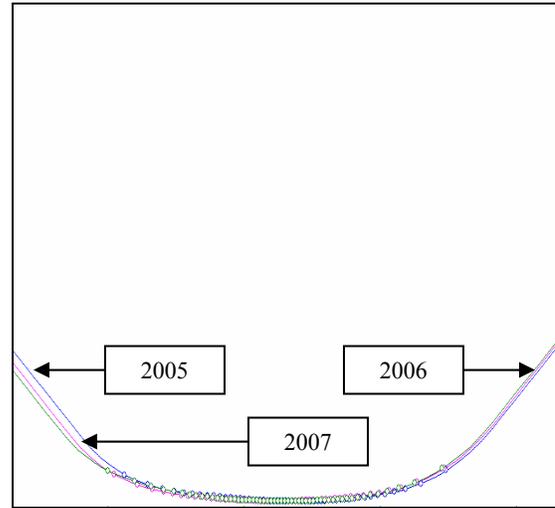


Figure 7.23
2005-06-07 SE Curves RD/WR Grade Span 3

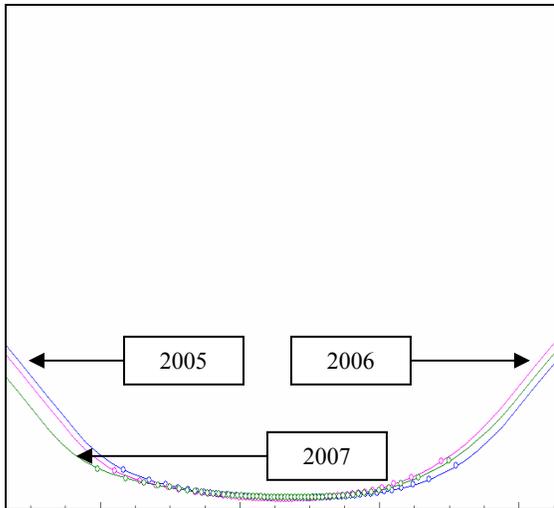


Figure 7.24
2005-06-07 SE Curves RD/WR Grade Span 4

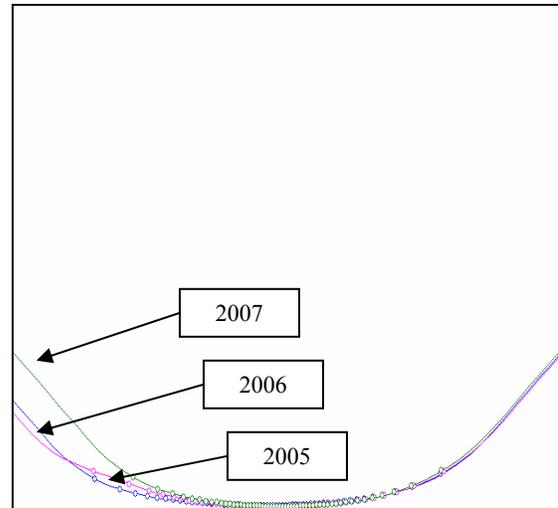


Figure 7.25
2005-06-07 SE Curves RD/WR Grade Span 5

7.3 Scaling of the NYSESLAT

As explained earlier, the NYSESLAT vertical scale was created based on the pre-existing 2006 vertical scale. The measurement of students' progress across years is augmented by the use of such a vertical scale. Since student abilities and item parameters are placed on an *equated* vertical scale, year-to-year progress in language proficiency for students who are administered the NYSESLAT for the three consecutive years can also be measured and documented. On this scale, a scale score of 500, for example, is greater than 490 and less than 510, irrespective of the students' grade levels and the year of administration. The scale score estimate of students, therefore, becomes an indication of the progress they have made over the previous year's estimate.

The NYSESLAT scaling procedure involves linear transformations of the raw score points into scale score points. These transformations do not give more weight to particular subtests, and they change neither the rank ordering of students nor their performance level classification. Linear transformation constants are utilized. The following equation was used to derive the scale scores:

$$\text{Scale Score} = 35 * (\text{theta}) + 600 \quad (19)$$

In the above equation, theta was derived from item parameters that have been adjusted for the test form and grade span level.

Raw Score-to-Scale Score Tables

The final reporting scales were used to produce raw score-to-scale score conversion tables for the Speaking/Listening and Reading/Writing combinations. Appendices C.1–C.5 provides the raw-to-scale score conversion tables for the Speaking/Listening and Reading/Writing combinations by grade span.

7.4 Item Response Model and Rational for Use

In addition to reporting raw score summary statistics and item level statistics using the classical test theory (CTT), the items on the NYSESLAT test were also analyzed within the framework of Item Response Theory (IRT). The Rasch model (Rasch, 1960) for dichotomous items and the Partial Credit Model (Masters, 1982) for polytomous items were used for developing, scoring, and reporting the NYSESLAT assessment. These models were recommended for several reasons.

First, the NYSESLAT vertical scale was created based on the pre-existing SELP vertical scale that was developed using the Rasch model. By using SELP items with known Rasch item difficulties, Pearson was able to create the NYSESLAT vertical scale in a timely fashion.

Second, the sample size requirements for calibration, scaling, and equating under the Rasch and Partial Credit models are significantly smaller than they are for other IRT models. For example, the Rasch model requires on the order of 400 examinees per form for equating versus approximately 1,500 examinees per form under the 3PL IRT model (Kolen and Brennan, 2004, p. 288).

Finally, for the requirements of the NYSESLAT program, the Rasch model has one characteristic that makes it very useful. There exists a one-to-one relationship between raw scores and scale scores. That is, a student who answers a certain number of items correctly will receive the same scale score as a second student with the same raw score, regardless of which particular items within the test form were answered correctly. These reasons lead Pearson to recommend that for the NYSESLAT the Rasch model be adopted as the IRT methodology.

7.5 Evidence of Model Fit

Fit statistics are used for evaluating the goodness-of-fit of a model to the data. Fit statistics are calculated by comparing the observed and expected trace lines obtained for an item after parameter estimates are obtained using a particular model. *Winsteps* provides two kinds of fit statistics called mean-squares that show the size of the randomness or amount of distortion of the measurement system.

The OUTFIT and the INFIT statistics are used in order to ascertain the suitability of the data for constructing variables and making measures with the Rasch model. These fit statistics are mean square standardized residuals for item by person responses averaged over persons and partitioned between ability groups (OUTFIT) and within ability groups (INFIT). When the observed item characteristic curve (ICC) departs from the expected ICC from a reference value of 1, there is an expectation of high ability students failing on an easy item or low ability students succeeding on a difficult one. The OUTFIT mean square evaluates the agreement between the observed ICC and the best fitting Rasch model curve over the ability sub-groups. It is a standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behavior by persons on items far from the person's ability level. The INFIT, on the other hand, is a within-group mean square, which summarizes the degree of misfit remaining within ability groups after the between-group misfit has been removed from the total. The INFIT, therefore, is a standardized information-weighted mean square statistic, which is more sensitive to unexpected responses to items near the person's ability level.

OUTFIT mean-squares are influenced by outliers and are usually easy to diagnose and remedy. INFIT mean-squares, on the other hand, are influenced by response patterns and are harder to diagnose and remedy. In general, mean-squares near 1.0 indicate little distortion of the measurement system, while values less than 1.0 indicate that observations are too predictable (redundancy, model overfit). Values greater than 1.0 indicate unpredictability (unmodeled noise, model underfit).

Generally speaking, when item fit indices are lower than 0.6, they do not discriminate well and show a greater than expected degree of consistency. Similarly, a fit value higher than 1.5 indicates inconsistency in examinee scores on the item (e.g., some unexpectedly high scores for low-ability candidates and low scores for high-ability candidates).

To an extent, no data fits the IRT model perfectly. Items that degraded or distorted the measurement system were also not included as anchor items and were excluded from the anchor pool. For NYSESLAT, no items that had an INFIT/OUTFIT value greater than or equal to 2.0 were included as anchor items. The percentage of items that were flagged for INFIT and OUTFIT varied depending on grade spans and subtests. In general, around 0–2% of items were flagged for INFIT. However, the percent of misfit was greater for OUTFIT across grade spans.

The OUTFIT and the INFIT statistics are presented in the item statistics tables in Appendices D.1–D.5.

7.6 Rasch Information

Table 8.1 presents the grade span, the modality, the number of items in each modality, the maximum number of points attainable for each modality, and the average Rasch difficulty for each modality. Appendix E includes item information at different cut points at each grade level.

Table 7.2
Average Rasch Difficulty by Grade Span and Modality

Grade Span	Test	Number of Items	Max Points	Average Rasch Difficulty
K-1	Listening	24	24	-2.40
	Speaking	16	34	-1.77
	Reading	15	15	-1.46
	Writing	15	24	-2.62
	Listening and Speaking	40	58	-2.14
	Reading and Writing	30	39	-2.04
	Total	70	97	-2.10
2-4	Listening	24	24	-0.03
	Speaking	16	34	-0.85
	Reading	24	24	-0.20
	Writing	16	22	-0.22
	Listening and Speaking	40	58	-0.36
	Reading and Writing	40	46	-0.21
	Total	80	104	-0.29
5-6	Listening	25	25	0.17
	Speaking	16	34	-0.62
	Reading	27	27	0.95
	Writing	19	25	0.52
	Listening and Speaking	41	59	-0.14
	Reading and Writing	46	52	0.77
	Total	87	111	0.34
7-8	Listening	25	25	1.06
	Speaking	16	34	0.25
	Reading	27	27	1.13
	Writing	19	25	0.31
	Listening and Speaking	41	59	0.74
	Reading and Writing	46	51	0.79
	Total	87	111	0.77
9-12	Listening	25	25	1.30
	Speaking	16	34	0.70
	Reading	27	27	1.85
	Writing	19	25	1.76
	Listening and Speaking	41	59	1.06
	Reading and Writing	46	51	1.81
	Total	87	111	1.46

Note: The statistics above are based on the equating sample (invalid cases were removed after cleaning procedures).

Appendices D.1–D.5 contain the results of the operational items for the NYSESLAT, which includes the Rasch item parameters. The following IRT item parameters are presented for each item grouped by Listening/Speaking and Reading/Writing combinations:

- Number of students
- Rasch difficulty value
- Standard error (SE) of Rasch difficulty
- MNSQ Infit: Standardized information-weighted mean square statistic, which is sensitive to unexpected behavior affecting responses to items near the person’s ability level
- MNSQ Outfit: Standardized outlier-sensitive mean-square fit statistic that is sensitive to unexpected behavior by persons on items far from the person’s ability level
- Flag of Misfit

PART 8: STANDARD SETTING

8.1 Introduction

As the contractor for the New York State English as a Second Language Achievement Test (NYSESLAT), Pearson organized a performance standard-setting meeting. The standard-setting sessions were conducted in Albany, New York, from March 28 to April 1, 2005. The purpose of this meeting was to provide preliminary recommendations on performance cut scores for the NYSESLAT.

For each group, there was one psychometrics staff member from Pearson to facilitate the technical part of the standard setting. In addition, a content specialist from Pearson and a NYSED official were present to provide support during the standard-setting sessions.

8.2 Standard-Setting Model

Item mapping is a well-established method available for establishing performance standards. The item-mapping procedure is capable of incorporating both multiple-choice and constructed-response items into the same process (Mitzel, H.C., Lewis, D.M., & Green, D.R., 2001). It has several other favorable characteristics, including:

- simplifying the judgment task by reducing the cognitive load required by panelists.
- connecting the judgment task of setting cut scores with the measurement model.
- connecting content with performance level descriptors.

The item-mapping procedure orders items for each test into a booklet according to the difficulty of the items, which is determined by IRT scaling techniques. Easy items are placed in the beginning of the booklet, and subsequent items become increasingly more difficult to the end of the booklet. Passages, rubrics, and sample student responses are placed in the booklet for reference purposes. Panelists are also provided with test blueprints and various other materials, including performance level descriptors, scoring rubrics, test booklets, and anchor papers.

8.3 Committees of Panelists

Five standard-setting committees were established to set the cut scores for the five grade spans of the NYSESLAT. As indicated in Table 8.1, the first group recommended standards on grades K–1; the second group recommended standards on grades 2–4; the third group recommended standards on grades 5–6; the fourth group recommended standards on grades 7–8; and the fifth group recommended standards on grades 9–12.

The panel members were New York State certified ESL, English language arts, bilingual education, and bilingual special education teachers who were familiar with the content standards. The NYSESLAT panelists were recruited by NYSED to participate in the standard-setting meeting. The panelists possess knowledge of working with students with limited English proficiency and also represent all regions of the state of New York. It was also recommended to have other educational stakeholders as well, such as administrators, curriculum specialists, NYSED members, and members of the professional community. These additional members tend to provide valuable insights from their area of expertise and help strengthen the consequential validity argument during panel discussions.

Table 8.1
Panel Composition for Standard-setting Committees

Grade	Group	Number of Judges
K-1	1	20
2-4	2	20
5-6	3	18
7-8	4	13
9-12	5	16

8.4 Performance Levels and Cut Scores

For the NYSESLAT, four performance levels are required, which correspond to three cut scores. The four performance levels are:

- Beginning
- Intermediate
- Advanced
- Proficient

The three cut scores are:

- Intermediate (between the Beginning and Intermediate performance levels)
- Advanced (between the Intermediate and Advanced performance levels)
- Proficient (between the Advanced and Proficient performance levels)

To set the three cut points, the item-mapping procedure was utilized. The standard-setting process is described below.

8.5 Standard-Setting Process

The standard setting began with introductions from the NYSED, Pearson, and panelists. This was followed by a presentation by the lead facilitator on the role of the panelists in the standard-setting process, setting performance standards, and placing cut scores. The goal was to familiarize panelists with the standard-setting process and the item-mapping procedure. This session took place in a large group setting (all five groups together).

After the orientation, the panelists were separated into specific breakout rooms according to their group assignments. Each group/room was led by a facilitator who was an expert in the standard-setting methodology. Assessment specialists rotated from group to group in order to provide content support. In addition, the panel members were further divided into three smaller table groups within their grade spans, each composed of five to seven members. These small groups

worked independently but had the opportunity to collaborate with the other table groups in their grade span during the standard-setting process. The following sequences of tasks were followed.

Review of the Assessment

The first task was to review the assessment blueprint. This was done in order for the panelists to gain an understanding of what the assessment was intended to measure. Discussions about the assessment content, the use of different item types, and number of questions were conducted. The panel members further defined the general performance level descriptors into specific descriptors to help the panel members come to a shared understanding about what it meant to be performing at each of the performance levels. The facilitator led this discussion with support from the assessment specialist who floated between the rooms.

Experience the Assessment

Next, the panel members had an opportunity to experience the assessment administered at the grade span assigned to them. This was an effective way to demonstrate to the panelists the knowledge and skills that students must possess to obtain a high score. It is assumed that panelists are likely to set more realistic performance standards if they experience the assessment themselves.

Scoring the Assessment

After the panelists finished taking the assessment, they were provided with an answer key to grade their test. The panelists scored their own assessments using the scoring rubrics and answer key provided. The scoring process offered an opportunity for the panelist to develop an understanding of the scoring of open-ended responses. They were provided with exemplars of score points. A discussion session followed the scoring of the assessment.

Review of Student Performance Levels

Panelists reviewed the previously established definitions of performance levels (Appendices F.3–F.4, 2006 NYSESLAT Technical Report). Then they discussed the performance levels. The goal was to help panelists clearly distinguish between student performance levels. Panelists' suggestions were related to the performance standards and content frameworks. The suggestions were retained for reference during the standard-setting process. Panelists reviewed definitions and offered illustrative suggestions for the Beginning, Intermediate, Advanced, and Proficient performance standards. After all the performance levels were reviewed, a discussion section was held. The focus was on the characteristics and interrelationships between and among performance standards.

Three Rounds of Ratings

The actual standard setting proceeded in three rounds. Each round was designed to foster increased consensus among panelists, although reaching consensus was not necessary. Panelists expressed their cut score judgment by placing a marker on the item that a student at that threshold of a performance level should master. One marker was placed for each cut score. There were three cut points.

During the Round 1 ratings, each panelist began by setting his/her three cut scores. The data was captured for each panelist. Before the Round 2 ratings, panelists were provided feedback on the Round 1 cut score positions of all panelists and their group medians. The panelists then discussed the Round 1 results. After the discussions, the Round 2 cuts were made, followed by further discussions. At this point, the panelists were provided with information about the percentage of students who would be classified in each of the performance levels, if those cuts were to be implemented.

In order to promote consistency across the grade spans, the groups came together to discuss the process and results of their assigned grades between all grade spans. Panelists then got back into their breakout groups and proceeded to make their Round 3 ratings. The median cut scores of the panelists then served as the starting point for the decision-makers on establishing the cut scores for the assessment.

8.6 Summary Statistics for the Three Rounds of Ratings

Appendix F.2 (2006 NYSESLAT Technical Report) provides the summary statistics for the round-by-round results by grade of the three performance level cuts. The statistics include the range, the raw score mean and standard deviation, the standard error of the mean, the median, standard error of the median, and the interquartile range.

8.7 Post-Standard-Setting Analyses

The median scores from the standard-setting committees were used as the recommended cuts. The cut scores were based on the total NYSESLAT score. After the standard-setting meetings, Pearson performed several post-standard-setting analyses. The first step was to look up the equivalent scale scores corresponding to the raw score cuts recommended by the committees. Graphs were then plotted using the grades as the independent variable and scale score as the dependent variable. The three cut points were then plotted on the same graph to show that the cuts were monotonically increasing from the lower cuts to the higher cuts. Some technical adjustments were made to the recommended cuts. Any adjustments made were within one standard error of the median. A quadratic equation was then applied to the scaled cut scores across all 13 grades, K–12. Further analyses were performed to smooth the cuts. This was done in order to ensure that there are no reversals.

Next, the proportion of the three cut points in raw score points based on the total test was applied to the Listening/Speaking and Reading/Writing combinations for each of the grades. Impact analysis was conducted on the 2005 operational data. The percentage of students falling into each of the performance levels was calculated for each grade if those cut points were adopted. This information was provided to the NYSED to make their final decisions on the cut points for the Listening/Speaking and Reading/Writing combinations.

8.8 Final Performance-Level Cut Points

The final cut points adopted by NYSED for the 2007 administration of the NYSESLAT for the Listening/Speaking and Reading/Writing combinations in raw score points, scale score, and theta metric are presented in Table 8.2 and Table 8.3. There are three cut points which correspond to four performance levels. Any score below the Intermediate cut point is the Beginning performance level.

Table 8.2
Final Performance Level Cut Points for Listening and Speaking

Grade	Listening & Speaking								
	Raw Score			Scale Score			Theta		
	I	A	P	I	A	P	I	A	P
K	23	43	52	511	573	616	-2.55	-0.76	0.47
1	26	45	55	520	582	649	-2.28	-0.52	1.39
2	16	36	51	543	602	661	-1.63	0.06	1.73
3	20	39	53	555	611	675	-1.27	0.31	2.14
4	24	42	54	567	621	684	-0.95	0.59	2.39
5	25	43	54	579	631	686	-0.59	0.89	2.46
6	28	45	55	587	639	696	-0.37	1.11	2.74
7	21	40	53	597	651	703	-0.08	1.45	2.95
8	24	43	53	605	661	703	0.15	1.74	2.95
9	26	44	52	623	678	715	0.66	2.23	3.28
10	27	46	53	628	686	721	0.80	2.44	3.46
11	30	48	54	635	694	728	1.00	2.68	3.66
12	32	50	54	640	704	728	1.14	2.96	3.66

Table 8.3
Final Performance Level Cut Points for Reading and Writing

Grade	Reading & Writing								
	Raw Score			Scale Score			Theta		
	I	A	P	I	A	P	I	A	P
K	22	30	34	537	582	609	-1.79	-0.52	0.25
1	26	34	38	559	609	655	-1.17	0.25	1.58
2	18	30	38	571	623	663	-0.82	0.64	1.81
3	20	31	40	580	630	679	-0.57	0.86	2.25
4	22	34	41	589	642	691	-0.33	1.19	2.61
5	24	36	45	620	660	704	0.56	1.73	2.98
6	25	38	46	623	668	710	0.66	1.94	3.15
7	26	38	46	629	674	710	0.83	2.13	3.15
8	27	40	46	632	682	710	0.92	2.34	3.15
9	19	37	44	643	698	726	1.21	2.80	3.60
10	20	38	45	646	702	732	1.30	2.91	3.78
11	21	41	46	649	711	739	1.39	3.16	3.97
12	22	41	46	652	715	739	1.48	3.30	3.97

Note. The statistics in the above two tables were based on the valid cases in the equating sample.

I = Intermediate, A = Advanced, P = Proficient.

PART 9: SUMMARY OF OPERATIONAL TEST RESULTS

This part presents both the raw score and scale score summaries for each of the reporting combinations (Listening/Speaking and Reading/Writing). Table 9.1 presents the raw score summary by grade. Table 9.2 presents the scale score summary by grade. Tables 9.1 and 9.2 include the sample size, the mean, median, interquartile range, range, and the standard deviation. Table 9.3 presents the percentage of students in each of the proficiency levels by grade. The overall proficiency level is defined by the lower of the two proficiency level designations (e.g., Listening/Speaking level=3, Reading/Writing level=4, Overall level=3). Table 9.4 presents the percentage of the Exit Rate for 2005, 2006, and 2007, including n- counts and the percentage of English language learners receiving exit status when they score at the Proficient Level (level 4) on both the LS and RW modality combinations.

Additionally, similar types of analyses (See Appendices G, H, and I) were done according to the following subgroups:

- Gender
- Ethnicity (American Indian or Alaska Native, Black or African-American, Asian and Pacific Islander, Hispanic or Latino, White, and Multi-Racial [not Hispanic origin])
- Needs/Resource Group (NYC, Big Four Cities, High Need Urban/Suburban, High Need Rural, Average Need, Low Need, Charter Schools)
- Five major languages background (Arabic, Bengali, Chinese, English, Spanish), language missing, and other languages
- Number of years with LEP
- Students with disability (Autism, Emotional Disturbance, Learning Disability, Mental Retardation, Deafness, Hearing Impairment, Speech or Language Impairment, Visual Impairment, Orthopedic Impairment, other health impairment, multiple disabilities, Traumatic Brain Injury).

Table 9.1
Raw Score Summary by Grade

Grade	Test	N-count	Mean	Median	IQR	Range	SD
K	Listening and Speaking	24239	43.22	46	13	0~58	10.90
	Reading and Writing	24239	22.22	22	10	0~39	7.38
1	Listening and Speaking	25871	49.91	53	8	0~58	8.90
	Reading and Writing	25871	31.59	34	9	0~39	6.99
2	Listening and Speaking	21810	45.77	48	10	0~58	9.74
	Reading and Writing	21810	27.82	29	13	0~46	9.13
3	Listening and Speaking	19269	48.34	51	8	0~58	9.81
	Reading and Writing	19269	32.69	35	12	0~46	9.00
4	Listening and Speaking	16074	48.99	53	8	0~58	10.15
	Reading and Writing	16074	34.89	38	10	0~46	8.87
5	Listening and Speaking	13145	48.23	52	9	0~59	10.90
	Reading and Writing	13145	35.65	38	13	0~52	10.00
6	Listening and Speaking	11455	47.50	52	11	0~59	12.32
	Reading and Writing	11455	35.92	39	15	0~52	10.33
7	Listening and Speaking	10858	44.79	50	15	0~59	13.24
	Reading and Writing	10858	32.32	34	17	0~52	10.70
8	Listening and Speaking	11531	45.71	51	15	0~59	13.01
	Reading and Writing	11531	34.06	36	17	2~52	10.79
9	Listening and Speaking	12997	40.71	45	21	0~59	13.69
	Reading and Writing	12997	31.07	32	19	1~52	11.32
10	Listening and Speaking	11463	43.06	46	17	1~59	11.81
	Reading and Writing	11463	33.87	36	16	1~52	10.38
11	Listening and Speaking	6854	46.35	49	12	0~59	9.80
	Reading and Writing	6854	37.03	39	13	3~52	9.22
12	Listening and Speaking	3269	47.84	50	10	2~59	9.09
	Reading and Writing	3269	38.06	40	12	2~52	8.59

Note: The data file used in Table 9.1 above is the final research file. The total n-counts for grades K–12 are 192,053. The stats presented in the table are based on the n-counts after cleaning (remove duplicate cases, off-level cases, and all records with Strand Raw Score = 0 or Greater than 900).

Table 9.2
Scale Score Summary by Grade

Grade	Test	N-count	Mean	Median	IQR	Range	SD
K	Listening and Speaking	24239	586.39	587	58	323~747	51.11
	Reading and Writing	24239	544.79	540	49	327~728	43.79
1	Listening and Speaking	25871	629.28	630	73	323~747	58.26
	Reading and Writing	25871	612.76	612	80	327~728	60.39
2	Listening and Speaking	21810	649.10	650	50	370~793	44.97
	Reading and Writing	21810	618.26	621	58	380~810	46.61
3	Listening and Speaking	19269	666.07	667	51	370~793	50.16
	Reading and Writing	19269	645.30	649	66	380~810	50.65
4	Listening and Speaking	16074	672.56	681	57	370~793	53.76
	Reading and Writing	16074	660.53	667	61	380~810	54.59
5	Listening and Speaking	13145	668.93	674	54	376~800	52.12
	Reading and Writing	13145	667.07	671	53	411~833	44.44
6	Listening and Speaking	11455	668.77	674	69	376~800	58.46
	Reading and Writing	11455	668.88	675	62	411~833	45.89
7	Listening and Speaking	10858	683.05	691	68	414~828	56.44
	Reading and Writing	10858	655.91	658	63	411~836	44.90
8	Listening and Speaking	11531	688.98	696	74	414~828	58.31
	Reading and Writing	11531	663.84	666	67	489~836	46.59
9	Listening and Speaking	12997	678.60	684	78	421~842	55.31
	Reading and Writing	12997	683.24	682	63	518~859	41.88
10	Listening and Speaking	11463	687.13	688	70	471~842	49.81
	Reading and Writing	11463	693.33	695	55	518~859	39.92
11	Listening and Speaking	6854	700.14	701	58	421~842	44.91
	Reading and Writing	6854	705.33	706	50	559~859	37.76
12	Listening and Speaking	3269	707.43	706	51	497~842	44.70
	Reading and Writing	3269	709.27	710	50	544~859	36.59

Note: Generally speaking, the mean for each grade should increase from one grade to the next higher grade in a similar manner as shown in Table 8.2 and Table 8.3 of this manual, which depicts increases across grade levels. However, due to artifacts of the population whereby some grades may have a greater percentage of higher scoring students than the next higher grade, the mean for the lower grade can be higher than the next higher grades. The statistics above are based on the final research file. The total n-counts for grade K–12 are 192,053. The stats presented in the table are based on the n-counts after cleaning (remove duplicate cases, off-level cases, and all records with Strand Raw Score = 0 or Greater than 900).

Table 9.3
Percentage of Students in Each Proficiency Level by Grade

Grade	Test	Proficiency Levels			
		Beginning	Intermediate	Advanced	Proficient
K	Listening and Speaking	6.55	30.96	38.67	23.82
	Reading and Writing	47.45	34.93	10.03	7.59
	Overall	48.30	36.40	10.89	4.41
1	Listening and Speaking	3.47	13.30	47.72	35.51
	Reading and Writing	18.33	29.74	31.78	20.14
	Overall	18.93	31.52	36.93	12.62
2	Listening and Speaking	2.42	9.61	50.55	37.42
	Reading and Writing	15.47	36.33	33.68	14.52
	Overall	15.76	36.54	36.17	11.53
3	Listening and Speaking	3.49	8.01	46.78	41.72
	Reading and Writing	10.59	22.07	42.16	25.18
	Overall	11.03	22.59	47.87	18.51
4	Listening and Speaking	4.71	9.65	44.08	41.56
	Reading and Writing	10.22	22.50	35.51	31.77
	Overall	11.04	23.24	44.03	21.69
5	Listening and Speaking	5.87	12.46	43.10	38.57
	Reading and Writing	14.30	26.61	39.79	19.30
	Overall	15.05	26.88	43.35	14.72
6	Listening and Speaking	9.89	14.72	40.96	34.43
	Reading and Writing	16.74	29.45	35.91	17.90
	Overall	18.25	29.53	39.59	12.63
7	Listening and Speaking	8.69	16.39	38.14	36.77
	Reading and Writing	27.44	34.32	28.84	9.40
	Overall	27.88	34.30	29.85	7.98
8	Listening and Speaking	9.46	18.66	29.38	42.49
	Reading and Writing	26.16	34.18	25.29	14.37
	Overall	26.94	34.02	26.27	12.77
9	Listening and Speaking	16.92	30.61	24.73	27.74
	Reading and Writing	17.52	43.98	22.25	16.25
	Overall	23.71	41.40	22.66	12.23
10	Listening and Speaking	10.93	37.22	26.21	25.63
	Reading and Writing	11.73	44.21	27.78	16.28
	Overall	16.53	46.52	26.28	10.66
11	Listening and Speaking	6.83	36.96	30.06	26.16
	Reading and Writing	6.51	49.49	25.84	18.16
	Overall	10.36	53.06	26.00	10.58
12	Listening and Speaking	6.06	39.83	23.77	30.35
	Reading and Writing	5.60	47.57	26.64	20.19
	Overall	9.64	53.50	24.41	12.45

Note: The statistics above are based on the final research file. The total n-counts for grade K–12 are 192,053. The stats presented in the table are based on the n-counts after cleaning (remove duplicate cases, off-level cases, and all records with Strand Raw Score = 0 or Greater than 900).

Table 9.4
Exit Rate by Grade for 2005, 2006 and 2007 Administrations

Grade	2007		2006		2005	
	N-Count	Percent	N-Count	Percent	N-Count	Percent
K	1069	4.41	1265	5.21	225	2.66
1	3265	12.62	4166	16.17	2953	12.07
2	2514	11.53	4081	17.59	3582	17.22
3	3567	18.51	3692	19.41	2518	15.20
4	3486	21.69	3124	20.86	3088	20.99
5	1935	14.72	2564	19.17	2268	19.53
6	1447	12.63	1765	15.88	1979	15.86
7	866	7.98	1706	13.79	1479	12.05
8	1473	12.77	2150	17.52	2070	17.26
9	1589	12.23	1909	13.05	1634	11.33
10	1222	10.66	1822	14.15	1588	11.40
11	725	10.58	1021	14.78	894	11.62
12	407	12.45	345	12.50	445	10.36
Total	23565	12.48	29610	15.27	24723	14.41

APPENDIX A: TEST BLUEPRINT AND ITEM MAP

A.1: 2007 NYSESLAT Operational Test Blueprint

NYSESLAT 2007 Operational Test Blueprint							
Strand	Cluster	Number of Items					Notes
		K-1	2-4	5-6	7-8	9-12	
Listening	Word/Sentence Comprehension	14	10	3	3	3	All items in this strand are multiple-choice items with 0-1 score points.
	Comprehension of Conversational Language	10	9	12	12	12	
	Synthesizing Information (Task-Based)		5	10	10	10	
	Number of Items	24	24	25	25	25	
Number of Points	24	24	25	25	25		
Reading	Word Reading	11					
	Sentence Reading	2					
	Short Passages with Questions	2					
	Word/Sentence Reading		7	3	3	3	
	Comprehension		17	24	24	24	
Number of Items	15	24	27	27	27	All items in this strand are multiple-choice items with 0-1 score points.	
Number of Points	15	24	27	27	27		

A.1: 2007 NYSESLAT Operational Test Blueprint (Continued)

NYSESLAT 2007 Operational Test Blueprint							
Strand	Cluster	Number of Items					Notes
		K-1	2-4	5-6	7-8	9-12	
Writing Conventions	Phonemic Understanding	6	4	2	2	2	All items in this strand are multiple-choice items with 0-1 score points.
	Mechanics & Structure A		4	5	5	5	
	Mechanics & Structure B		4	8	8	8	
Writing	Developmental Writing	8					Level K-1 has 8 CR items with 0-2 score points
	Pre-writing		3	3	3	3	Each of the Levels 2-4, 5-6, 7-8, 9-12 has 3 CR items with 0-2 score points
	Extended Response	1	1	1	1	1	Each of the five levels has one ER item. The ER item for Level K-1 has 0-2 score points. The ER items for the Levels 2-4, 5-6, 7-8, and 9-12 each has 0-4 score points.
Number of Items		15	16	19	19	19	
Number of Points		24	22	25	25	25	
Speaking	Repeat/Read Aloud	2	2	2	2	2	Each of the five levels has 2 SS items with 0-2 score points in this cluster.
	Sentence Completion	5	5	5	5	5	Each of the five levels has 5 SS items with 0-2 score points in this cluster.
	Storytelling	1	1	1	1	1	Each of the five levels has 1 SE item with 0-4 score points in this cluster.
	Picture Description	3	3	3	3	3	Each of the five levels has 3 SS items with 0-2 score points in this cluster.
	Social Interaction	5	5	5	5	5	Each of the five levels has 5 SS items with 0-2 score points in this cluster.
Number of Items		16	16	16	16	16	
Number of Points		34	34	34	34	34	
Total Number of Items		70	80	87	87	87	
Total Number of Points		97	104	111	111	111	

A.2: 2007 NYSESLAT Operational Test Information

Grade K-1

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
1	SS	2	1	Become familiar with some conventions of American English.	Repeat/Read Aloud
2	SS	2	2	Develop comprehension of text to prepare to read aloud.	Repeat/Read Aloud
3	SS	2	1	Convey information, using some organizational patterns and structures.	Sentence Completion
4	SS	2	4	Explain actions, choices, and decisions in social and classroom situations.	Sentence Completion
5	SS	2	5	Learn about and demonstrate an appreciation of some commonalities and distinctions across cultures and generations, including the students' own.	Sentence Completion
6	SS	2	4	Explain actions, choices, and decisions in social and classroom situations.	Sentence Completion
7	SS	2	1	Convey information, using some organizational patterns and structures.	Sentence Completion
8	SE	4	2	Create personal stories, using appropriate vocabulary and elements of the literature students have read or heard.	Storytelling
9	SS	2	4	Explain actions, choices, and decisions in social and classroom situations.	Picture Description
10	SS	2	5	Demonstrate familiarity with some U.S. cultural referents at the local and national levels	Picture Description
11	SS	2	1	Convey information, using some organizational patterns and structures.	Picture Description
12	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
13	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
14	SS	2	4	Discover alternative ways of saying things in social and classroom interactions.	Social Interaction
15	SS	2	4	Describe, read about, participate in, or recommend a favorite activity, book, song, or other interest to various audiences.	Social Interaction
16	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade K-1**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Listening					
1	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
2	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
3	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
4	MC	1	3	Form and express responses to ideas through reading, listening, viewing, discussing, and writing.	Word/Sentence Comprehension
5	MC	1	1	Become familiar with some conventions of American English.	Word/Sentence Comprehension
6	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful.	Word/Sentence Comprehension
7	MC	1	3	Form and express responses to ideas through reading, listening, viewing, discussing, and writing.	Word/Sentence Comprehension
8	MC	1	1	Become familiar with some conventions of American English.	Word/Sentence Comprehension
9	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
10	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
11	MC	1	1	Select information appropriate to the purpose of the investigation.	Word/Sentence Comprehension
12	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
13	MC	1	3	Form and express responses to ideas through reading, listening, viewing, discussing, and writing.	Word/Sentence Comprehension
14	MC	1	3	Form and express responses to ideas through reading, listening, viewing, discussing, and writing.	Word/Sentence Comprehension
15	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
16	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade K-1**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Listening					
17	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
18	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful.	Comprehension of Conversational Language
19	MC	1	4	Use appropriate vocabulary, language, and interaction styles for various audiences and social or school situations.	Comprehension of Conversational Language
20	MC	1	5	Demonstrate familiarity with some cultural and language patterns and norms in American English.	Comprehension of Conversational Language
21	MC	1	4	Follow oral directions and written directions to participate in classroom and social activities.	Comprehension of Conversational Language
22	MC	1	3	Form and express responses to ideas through reading, listening, viewing, discussing, and writing.	Comprehension of Conversational Language
23	MC	1	5	Demonstrate familiarity with some cultural and language patterns and norms in American English.	Comprehension of Conversational Language
24	MC	1	3	Form and express responses to ideas through reading, listening, viewing, discussing, and writing.	Comprehension of Conversational Language
Reading					
1	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word Reading
2	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word Reading
3	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word Reading
4	MC	1	1	Formulate, ask, and respond to questions to obtain and provide information and meaning.	Word Reading
5	MC	1	1	Formulate, ask, and respond to questions to obtain and provide information and meaning.	Word Reading
6	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word Reading
7	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word Reading

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade K-1**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Reading					
8	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word Reading
9	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful.	Word Reading
10	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful.	Word Reading
11	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful.	Word Reading
12	MC	1	1	Compare, contrast, and categorize to gain a deeper understanding of information and objects.	Sentence Reading
13	MC	1	1	Compare, contrast, and categorize to gain a deeper understanding of information and objects.	Sentence Reading
14	MC	1	2	Use basic reading and listening strategies to make literary text comprehensible and meaningful.	Short Passage with Question
15	MC	1	2	Use basic reading and listening strategies to make literary text comprehensible and meaningful.	Short Passage with Question
Writing					
1	MC	1	1	Become familiar with some conventions of American English.	Phonemic Understanding
2	MC	1	1	Become familiar with some conventions of American English.	Phonemic Understanding
3	MC	1	1	Become familiar with some conventions of American English.	Phonemic Understanding
4	MC	1	1	Become familiar with some conventions of American English.	Phonemic Understanding
5	MC	1	1	Become familiar with some conventions of American English.	Phonemic Understanding
6	MC	1	1	Become familiar with some conventions of American English.	Phonemic Understanding
7	CR	2	1	Become familiar with some conventions of American English.	Developmental Writing: Sentence Copying
8	CR	2	1	Become familiar with some conventions of American English.	Developmental Writing: Sentence Copying
9	CR	2	1	Express and develop ideas and understanding, using some elements of the "writing process."	Developmental Writing: Letter Writing

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade K-1**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Writing					
10	CR	2	1	Express and develop ideas and understanding, using some elements of the "writing process."	Developmental Writing: Letter Writing
11	CR	2	1	Become familiar with some conventions of American English.	Developmental Writing: Word Writing
12	CR	2	1	View, listen to, read, gather, organize, and discuss information from various sources.	Developmental Writing: Word Writing
13	CR	2	1	Become familiar with some conventions of American English.	Developmental Writing: Sentence Writing
14	CR	2	1	Become familiar with some conventions of American English.	Developmental Writing: Sentence Writing
15	ER	2	3	Form and express responses to ideas through reading, listening, viewing, discussing, and writing.	Extended Response

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 2-4**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
1	SS	2	1	Become familiar with some conventions of American English.	Repeat/Read Aloud
2	SS	2	2	Read aloud with confidence, accuracy, and fluency.	Repeat/Read Aloud
3	SS	2	5	Recognize and demonstrate an appreciation of some commonalities and distinctions across cultures and groups (differentiated by gender, ability, generations, etc.) including the student's own.	Sentence Completion
4	SS	2	1	Convey information, using a variety of organizational patterns and structures.	Sentence Completion
5	SS	2	4	Explain actions, choices, and decisions in social and academic situations.	Sentence Completion
6	SS	2	5	Demonstrate familiarity with a broad range of U.S. cultural and political referents through institutions, functions, and processes at the local and national levels.	Sentence Completion
7	SS	2	1	Convey information, using a variety of organizational patterns and structures.	Sentence Completion
8	SE	4	2	Create personal stories, poems, and songs, including those that reflect traditional and popular American culture; use appropriate vocabulary and elements of the literature students have read or heard.	Storytelling

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 2-4**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
9	SS	2	4	Explain actions, choices, and decisions in social and academic situations.	Picture Description
10	SS	2	5	Demonstrate familiarity with a broad range of U.S. cultural and political referents through institutions, functions, and processes at the local and national levels.	Picture Description
11	SS	2	3	Form and express responses to a variety of literary, informational, and persuasive material through reading, listening, viewing, discussing, and writing; use details and evidence as support.	Picture Description
12	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
13	SS	2	4	Describe, read about, participate in, or recommend a favorite activity, book, song, or other interest to various audiences.	Social Interaction
14	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
15	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
16	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
Listening					
1	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
2	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
3	MC	1	1	Apply learning strategies to acquire information and make oral texts comprehensible and meaningful.	Word/Sentence Comprehension
4	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
5	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
6	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
7	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 2-4**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Listening					
8	MC	1	1	Apply learning strategies to acquire information and make oral texts comprehensible and meaningful.	Word/Sentence Comprehension
9	MC	1	3	Apply learning strategies to explore a variety of materials.	Word/Sentence Comprehension
10	MC	1	3	Apply learning strategies to explore a variety of materials.	Word/Sentence Comprehension
11	MC	1	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Comprehension of Conversational Language
12	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
13	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
14	MC	1	2	Identify and use listening strategies to make literary text comprehensible and meaningful.	Comprehension of Conversational Language
15	MC	1	3	Form and express responses to a variety of literary, informational, and persuasive material through reading, listening, viewing, discussing, and writing; use details and evidence for support.	Comprehension of Conversational Language
16	MC	1	1	Formulate and share opinions about information and ideas with reference to features in oral and written text such as details and facts.	Comprehension of Conversational Language
17	MC	1	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Comprehension of Conversational Language
18	MC	1	4	Follow oral directions to participate in classroom and social activities.	Comprehension of Conversational Language
19	MC	1	2	Identify and use listening strategies to make literary text comprehensible and meaningful.	Comprehension of Conversational Language
20	MC	1	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Task-based Listening
21	MC	1	4	Follow oral directions to participate in classroom and social activities.	Task-based Listening
22	MC	1	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Task-based Listening
23	MC	1	4	Follow oral directions to participate in classroom and social activities.	Task-based Listening
24	MC	1	4	Follow oral directions to participate in classroom and social activities.	Task-based Listening

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 2-4**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Reading					
1	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading
2	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading
3	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading
4	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading
5	MC	1	1	Apply learning strategies to acquire information and make written texts comprehensible and meaningful.	Word/Sentence Reading
6	MC	1	1	Apply learning strategies to acquire information and make written texts comprehensible and meaningful.	Word/Sentence Reading
7	MC	1	1	Apply learning strategies to acquire information and make written texts comprehensible and meaningful.	Word/Sentence Reading
8	MC	1	2	Identify and use reading strategies to make literary text comprehensible and meaningful.	Comprehension
9	MC	1	5	Learn about and demonstrate an appreciation of some commonalities and distinctions across cultures and generations, including the students' own.	Comprehension
10	MC	1	2	Identify and use reading and listening strategies to make literary text comprehensible and meaningful.	Comprehension
11	MC	1	2	Apply learning strategies to comprehend and make inferences about literature and produce literary responses.	Comprehension
12	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
13	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
14	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
15	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
16	MC	1	3	Recognize personal point of view in self and others in discussing, interpreting, and evaluating information.	Comprehension
17	MC	1	3	Apply learning strategies to examine, interpret, and evaluate a variety of materials.	Comprehension
18	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 2-4**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Reading					
19	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
20	MC	1	3	Recognize personal point of view in self and others in discussing information.	Comprehension
21	MC	1	2	Recognize some features that distinguish some genres and use those features to aid comprehension.	Comprehension
22	MC	1	2	Identify and use reading strategies to make literary text comprehensible and meaningful.	Comprehension
23	MC	1	2	Identify and use reading strategies to make literary text comprehensible and meaningful.	Comprehension
24	MC	1	2	Identify and use reading strategies to make literary text comprehensible and meaningful.	Comprehension
Writing					
1	MC	1	1	Convey information and ideas through written language, using conventions and features of American English.	Phonemic Understanding
2	MC	1	1	Apply learning strategies to acquire information and make oral texts comprehensible and meaningful.	Phonemic Understanding
3	MC	1	1	Convey information and ideas through written language, using conventions and features of American English.	Phonemic Understanding
4	MC	1	1	Convey information and ideas through written language, using conventions and features of American English.	Phonemic Understanding
5	MC	1	1	Convey information and ideas through written language, using conventions and features of American English.	Mechanics & Structure
6	MC	1	1	Formulate, ask, and respond to questions to obtain and provide information and meaning.	Mechanics & Structure
7	MC	1	1	Convey information and ideas through written language, using conventions and features of American English.	Mechanics & Structure
8	MC	1	1	Convey information and ideas through written language, using conventions and features of American English.	Mechanics & Structure
9	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
10	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 2-4**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Writing					
11	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
12	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
13	CR	2	1	Use the process of prewriting, drafting, revising, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
14	CR	2	1	Use the process of prewriting, drafting, revising, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
15	CR	2	1	Use the process of prewriting, drafting, revising, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
16	ER	4	3	Form and express responses to a variety of literary, informational, and persuasive material through reading, listening, viewing, discussing, and writing; use details and evidence as support.	Writing

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 5-6**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
1	SS	2	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose.	Repeat/Read Aloud
2	SS	2	2	Read aloud with confidence, accuracy, and fluency.	Repeat/Read Aloud
3	SS	2	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures.	Sentence Completion
4	SS	2	5	Demonstrate an understanding of a broad range of U.S. cultural and political referents through institutions, functions, and processes at the local and national levels, and compare/contrast these with parallels in the students' native community.	Sentence Completion

A.2: 2007 NYSESLAT Operational Test Information (Continued)
Grade 5-6

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
5	SS	2	5	Recognize and demonstrate an appreciation of some commonalities and distinctions across cultures and groups (differentiated by gender, ability, generations, etc.), including the student's own.	SentenceCompletion
6	SS	2	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures.	Sentence Completion
7	SS	2	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures.	Sentence Completion
8	SE	4	2	Create stories, poems, songs, and plays, including those that reflect traditional and popular American culture, observing the conventions of the genre; create an effective voice, using a variety of writing or speaking styles appropriate to different audiences, purposes, and settings.	Storytelling
9	SS	2	4	Explain actions, choices, and decisions in social and academic situations.	Picture Description
10	SS	2	5	Demonstrate an understanding of a broad range of U.S. cultural and political referents through institutions, functions, and processes at the local and national levels, and compare/contrast these with parallels in the students' native community.	Picture Description
11	SS	2	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences, supporting positions with well-developed arguments.	Picture Description
12	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
13	SS	2	4	Negotiate and manage interactions to accomplish social and classroom tasks.	Social Interaction
14	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
15	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
16	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 5-6**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Listening					
1	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
2	MC	1	1	Apply learning strategies to acquire information and make oral texts comprehensible and meaningful.	Word/Sentence Comprehension
3	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
4	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
5	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
6	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Comprehension of Conversational Language
7	MC	1	5	Demonstrate an understanding of cultural and language patterns and norms in American English, including different regional and social varieties of English.	Comprehension of Conversational Language
8	MC	1	2	Identify and use listening strategies to make literary text comprehensible and meaningful.	Comprehension of Conversational Language
9	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
10	MC	1	5	Demonstrate an understanding of cultural and language patterns and norms in American English, including different regional and social varieties of English.	Comprehension of Conversational Language
11	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Comprehension of Conversational Language
12	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful.	Comprehension of Conversational Language
13	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
14	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text.	Comprehension of Conversational Language
15	MC	1	4	Negotiate and manage interactions to accomplish social and classroom tasks.	Comprehension of Conversational Language
16	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 5-6**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Listening					
17	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
18	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
19	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
20	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
21	MC	1	4	Negotiate and manage interactions to accomplish social and classroom tasks.	Task-based Listening
22	MC	1	4	Request and provide information and assistance, orally, for personal, social, and academic purposes.	Task-based Listening
23	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
24	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
25	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
Reading					
1	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading
2	MC	1	1	Formulate, ask, and respond to various question forms to obtain, clarify, and extend information and meaning.	Word/Sentence Reading
3	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading
4	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
5	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
6	MC	1	3	Apply learning strategies to examine and interpret a variety of materials.	Comprehension
7	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text.	Comprehension

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 5-6**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Reading					
8	MC	1	2	Identify and use reading strategies to make literary text comprehensible and meaningful.	Comprehension
9	MC	1	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences, supporting positions with well-developed arguments.	Comprehension
10	MC	1	3	Apply learning strategies to examine and interpret a variety of materials.	Comprehension
11	MC	1	2	Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings.	Comprehension
12	MC	1	2	Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings.	Comprehension
13	MC	1	3	Apply learning strategies to examine and interpret a variety of materials.	Comprehension
14	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
15	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text.	Comprehension
16	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
17	MC	1	3	Apply learning strategies to examine and interpret a variety of materials.	Comprehension
18	MC	1	2	Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings.	Comprehension
19	MC	1	3	Recognize personal point of view in self and others in discussing information.	Comprehension
20	MC	1	3	Apply learning strategies to examine and interpret a variety of materials.	Comprehension
21	MC	1	2	Identify and use reading strategies to make literary text comprehensible and meaningful.	Comprehension
22	MC	1	2	Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings.	Comprehension

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 5-6**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Reading					
23	MC	1	2	Identify and use reading strategies to make literary text comprehensible and meaningful.	Comprehension
24	MC	1	2	Identify key literary elements in texts and relate those features to students' own experience.	Comprehension
25	MC	1	2	Identify key literary elements in texts and relate those features to students' own experience.	Comprehension
26	MC	1	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences, supporting positions with well-developed arguments.	Comprehension
27	MC	1	2	Identify key literary elements in texts and relate those features to students' own experience.	Comprehension
Writing					
1	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Phonemic Understanding
2	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Phonemic Understanding
3	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
4	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
5	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
6	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
7	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 5-6**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Writing					
8	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
9	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
10	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
11	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
12	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
13	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
14	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
15	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
16	CR	2	1	Use the process of prewriting, drafting, revising, peer editing, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
17	CR	2	1	Use the process of prewriting, drafting, revising, peer editing, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
18	CR	2	1	Use the process of prewriting, drafting, revising, peer editing, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
19	ER	4	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences, supporting positions with well-developed arguments.	Writing

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 7-8**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
1	SS	2	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose.	Repeat/Read Aloud
2	SS	2	2	Read aloud with confidence, accuracy, and fluency.	Repeat/Read Aloud
3	SS	2	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures.	Sentence Completion
4	SS	2	4	Explain actions, choices, and decisions in social and academic contexts.	Sentence Completion
5	SS	2	4	Explain actions, choices, and decisions in social and academic contexts.	Sentence Completion
6	SS	2	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures.	Sentence Completion
7	SS	2	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures.	Sentence Completion
8	SE	4	2	Create stories, poems, songs, and plays, including those that reflect traditional and popular American culture, observing the conventions of the genre; create an effective voice, using a variety of writing or speaking styles appropriate to different audiences, purposes, and settings.	Storytelling
9	SS	2	4	Explain actions, choices, and decisions in social and academic contexts.	Picture Description
10	SS	2	5	Demonstrate an understanding of a broad range of U.S. cultural and political referents through institutions, functions, and processes at the local and national levels, and compare/contrast these with parallels in the students' native community.	Picture Description
11	SS	2	4	Explain actions, choices, and decisions in social and academic situations.	Picture Description
12	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
13	SS	2	4	Explain actions, choices, and decisions in social and academic situations.	Social Interaction
14	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 7-8**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
15	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
16	SS	2	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences, supporting positions with well-developed arguments.	Social Interaction
Listening					
1	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful.	Word/Sentence Comprehension
2	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
3	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful.	Word/Sentence Comprehension
4	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
5	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
6	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
7	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text.	Comprehension of Conversational Language
8	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
9	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
10	MC	1	4	Follow oral and written directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Comprehension of Conversational Language
11	MC	1	4	Follow oral and written directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Comprehension of Conversational Language
12	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text.	Comprehension of Conversational Language
13	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 7-8**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Listening					
14	MC	1	1	Select information appropriate to the purpose of the investigation, relate ideas from one written or spoken source to another, and exclude nonessential information.	Comprehension of Conversational Language
15	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
16	MC	1	4	Negotiate and manage interactions to accomplish social and classroom tasks.	Task-based Listening
17	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
18	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
19	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
20	MC	1	4	Request and provide information and assistance, orally, for personal, social, and academic purposes.	Task-based Listening
21	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
22	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
23	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
24	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
25	MC	1	4	Follow oral directions to participate in classroom and social activities, and provide directions to peers in selected interactions.	Task-based Listening
Reading					
1	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful.	Word/Sentence Reading
2	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading
3	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 7-8**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Reading					
4	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
5	MC	1	3	Recognize, explain, and evaluate how structural features affect readers' and listeners' understanding and appreciation of text.	Comprehension
6	MC	1	3	Recognize, explain, and evaluate how structural features affect readers' and listeners' understanding and appreciation of text.	Comprehension
7	MC	1	3	Recognize, explain, and evaluate how structural features affect readers' and listeners' understanding and appreciation of text.	Comprehension
8	MC	1	3	Apply learning strategies to examine and interpret a variety of materials.	Comprehension
9	MC	1	2	Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings.	Comprehension
10	MC	1	2	Locate and identify selected literary elements and techniques in texts and relate those elements to those in other works and to students' own experiences.	Comprehension
11	MC	1	2	Locate and identify selected literary elements and techniques in texts and relate those elements to those in other works and to students' own experiences.	Comprehension
12	MC	1	2	Identify and explain the distinguishing features of the major genres, and use those features to aid comprehension, interpretation, and discussion of literature.	Comprehension
13	MC	1	3	Apply learning strategies to examine and interpret a variety of materials.	Comprehension
14	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
15	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
16	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text.	Comprehension
17	MC	1	2	Locate and identify selected literary elements and techniques in texts and relate those elements to those in other works and to students' own experiences.	Comprehension

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 7-8**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Reading					
18	MC	1	2	Identify and use reading strategies to make literary text comprehensible and meaningful.	Comprehension
19	MC	1	3	Apply learning strategies to examine and interpret a variety of materials.	Comprehension
20	MC	1	2	Identify and use reading strategies to make literary text comprehensible and meaningful.	Comprehension
21	MC	1	2	Identify and use reading strategies to make literary text comprehensible and meaningful.	Comprehension
22	MC	1	2	Apply learning strategies to comprehend and make inferences about literature and produce literary responses.	Comprehension
23	MC	1	2	Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings.	Comprehension
24	MC	1	2	Locate and identify selected literary elements and techniques in texts and relate those elements to those in other works and to students' own experiences.	Comprehension
25	MC	1	2	Locate and identify selected literary elements and techniques in texts and relate those elements to those in other works and to students' own experiences.	Comprehension
26	MC	1	2	Locate and identify selected literary elements and techniques in texts and relate those elements to those in other works and to students' own experiences.	Comprehension
27	MC	1	2	Locate and identify selected literary elements and techniques in texts and relate those elements to those in other works and to students' own experiences.	Comprehension
Writing					
1	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Phonemic Understanding
2	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Phonemic Understanding
3	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 7-8**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Writing					
4	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
5	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
6	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
7	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
8	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
9	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
10	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
11	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
12	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
13	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
14	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
15	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
16	CR	2	1	Use the process of prewriting, drafting, revising, peer editing, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 7-8**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Writing					
17	CR	2	1	Use the process of prewriting, drafting, revising, peer editing, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
18	CR	2	1	Use the process of prewriting, drafting, revising, peer editing, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
19	ER	4	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences, supporting positions with well-developed arguments.	Writing

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 9-12**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
1	SS	2	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose.	Repeat/Read Aloud
2	SS	2	2	Read aloud with confidence, accuracy, fluency, and expression to demonstrate understanding and to convey an interpretation of meaning.	Repeat/Read Aloud
3	SS	2	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures.	Sentence Completion
4	SS	2	4	Explain actions, choices, and decisions in social and academic situations.	Sentence Completion
5	SS	2	4	Explain actions, choices, and decisions in social and academic situations.	Sentence Completion
6	SS	2	5	Recognize and demonstrate an appreciation of some commonalities and distinctions across cultures and groups (differentiated by gender, ability, generations, etc.), including the students' own.	Sentence Completion
7	SS	2	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures.	Sentence Completion

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 9-12**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
8	SE	4	2	Create stories, poems, and plays, including those that reflect traditional and popular American culture, observing the conventions of the genre; create an effective voice, using typical features of a given genre; create an effective voice, using a variety of writing styles appropriate to different audiences, purposes, and settings.	Storytelling
9	SS	2	4	Explain actions, choices, and decisions in social and academic situations.	Picture Description
10	SS	2	5	Demonstrate an understanding of a broad range of U.S. cultural and political referents through institutions, functions, and processes at the local and national levels, and compare/contrast these with parallels in the students' native community.	Picture Description
11	SS	2	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences, supporting positions with well-developed arguments.	Picture Description
12	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
13	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction
14	SS	2	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences, supporting positions with well-developed arguments.	Social Interaction
15	SS	2	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences, supporting positions with well-developed arguments.	Social Interaction
16	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes.	Social Interaction

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 9-12**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Listening					
1	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
2	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
3	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful.	Word/Sentence Comprehension
4	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension of Conversational Language
5	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension of Conversational Language
6	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension of Conversational Language
7	MC	1	5	Demonstrate an understanding of cultural and language patterns and norms in American English, including different regional and social varieties of English, and identify and interpret how these patterns and norms are used.	Comprehension of Conversational Language
8	MC	1	5	Demonstrate an understanding of cultural and language patterns and norms in American English, including different regional and social varieties of English, and identify and interpret how these patterns and norms are used.	Comprehension of Conversational Language
9	MC	1	5	Demonstrate an understanding of a broad range of U.S. cultural and political referents through institutions, functions, and processes at the local and national levels, and compare/contrast these with parallels in the student's native community.	Comprehension of Conversational Language
10	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension of Conversational Language
11	MC	1	3	Recognize, explain, evaluate, and analyze how structural features affect readers' and listeners' understanding and appreciation of text.	Comprehension of Conversational Language

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 9-12**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Listening					
12	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension of Conversational Language
13	MC	1	5	Demonstrate an understanding of cultural and language patterns and norms in American English, including different regional and social varieties of English, and identify and interpret how these patterns and norms are used.	Comprehension of Conversational Language
14	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
15	MC	1	1	Identify and use listening strategies to make text comprehensible and meaningful.	Comprehension of Conversational Language
16	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions.	Task-based Listening
17	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions.	Task-based Listening
18	MC	1	4	Negotiate and manage interactions to accomplish social and classroom tasks.	Task-based Listening
19	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful.	Task-based Listening
20	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions.	Task-based Listening
21	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions.	Task-based Listening
22	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions.	Task-based Listening
23	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions.	Task-based Listening

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 9-12**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Listening					
24	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions.	Task-based Listening
25	MC	1	4	Negotiate and manage interactions to accomplish social and classroom tasks.	Task-based Listening
Reading					
1	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading
2	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading
3	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Word/Sentence Reading
4	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension
5	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
6	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text.	Comprehension
7	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension
8	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension
9	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text.	Comprehension
10	MC	1	3	Recognize, explain, evaluate, and analyze how structural features affect readers' and listeners' understanding and appreciation of text.	Comprehension
11	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension
12	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 9-12**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Reading					
13	MC	1	2	Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings.	Comprehension
14	MC	1	2	Apply reading strategies to make literary text comprehensible and meaningful.	Comprehension
15	MC	1	2	Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings.	Comprehension
16	MC	1	2	Apply learning strategies to comprehend, make inferences about, and analyze literature, and to produce literary responses.	Comprehension
17	MC	1	2	Locate and identify a wide range of significant literary elements and techniques in texts and use those elements to interpret the work, comparing and contrasting the work to other works and to students' own experiences.	Comprehension
18	MC	1	1	Identify and use reading strategies to make text comprehensible and meaningful.	Comprehension
19	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension
20	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension
21	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension
22	MC	1	3	Assess, compare, and evaluate the quality of spoken or written texts and visual presentations, using different criteria related to the organization, subject area, and purpose of text.	Comprehension
23	MC	1	2	Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings.	Comprehension
24	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials.	Comprehension

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 9-12**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Reading					
25	MC	1	2	Make predictions, inferences, and deductions, and describe different levels of meaning of literary works presented orally and in written form, including literal and implied meanings.	Comprehension
26	MC	1	2	Locate and identify a wide range of significant literary elements and techniques in texts and use those elements to interpret the work, comparing and contrasting the work to other works and to students' own experiences.	Comprehension
27	MC	1	2	Locate and identify a wide range of significant literary elements and techniques in texts and use those elements to interpret the work, comparing and contrasting the work to other works and to students' own experiences.	Comprehension
Writing					
1	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Phonemic Understanding
2	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Phonemic Understanding
3	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
4	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
5	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
6	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure
7	MC	1	1	Convey information and ideas through written language, using conventions and features of American English appropriate to audience and purpose.	Mechanics & Structure

A.2: 2007 NYSESLAT Operational Test Information (Continued)**Grade 9-12**

Modality/ Position	Item Type	Score Point	Learning Standard	Performance Indicator	Cluster
Writing					
8	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
9	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
10	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
11	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
12	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
13	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
14	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
15	MC	1	3	Evaluate students' own and others' work, individually and collaboratively, on the basis of a variety of criteria.	Mechanics & Structure
16	CR	2	1	Use the process of prewriting, drafting, revising, peer editing, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
17	CR	2	1	Use the process of prewriting, drafting, revising, peer editing, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
18	CR	2	1	Use the process of prewriting, drafting, revising, peer editing, and proofreading (the "writing process") to produce well-constructed informational texts.	Pre-writing
19	ER	4	3	Speak and write, using the conventions and features of American English, to effectively influence an audience (e.g., to persuade, negotiate, argue).	Writing

A.3: Item Specification**LISTENING: Word/Sentence Comprehension, grades K–1**

The test taker must choose which of three letters represents the first sound in a word. These items measure phonemic discrimination, and are designed especially to test students who are just beginning to learn the sound system in English.

Item Component	Description of Component
Stimulus	--
Stem	closed stem question: “What is the first sound in the word _____?”
Answer options	3 answer options consisting of 3 individual letter graphics
Graphic	--

LISTENING: Word/Sentence Comprehension, all grades

The test taker hears a question and must choose which of three small graphics represents the correct answer to the question. These items measure comprehension of words, phrases, and sentences, and are designed especially to test students at lower levels of proficiency.

Item Component	Description of Component
Stimulus	--
Stem	closed stem question, maximum 15 words: “Which picture shows _____?” , “Which girl is _____?” , “Which building is where _____?” , etc.
Answer options	3 answer options consisting of 3 small individual graphics lettered A-C
Graphic	--

A.3: Item Specification (Continued)**LISTENING: Comprehension of Conversational Language, all grades**

The test taker listens to someone speaking (a monologue) in a situation where a person would naturally be speaking without expecting a response from the listener, e.g., a teacher talking to a class, someone making an announcement over a loudspeaker, or someone briefly describing an event. These items measure comprehension of main ideas and details and the ability to make inferences. The language tested should simulate authentic conversational English. The language level should be at the upper level of the grade span.

Item Component	Description of Component
Stimulus	simulated authentic conversation with one person speaking; 20–30 words (grades K–1), 40–50 words (grades 2–4), 50–60 words (grades 5–8), 55–70 words (grades 9–12); there is one item per stimulus (grades K–6) or two items per stimulus (grades 7–12)
Stem	closed stem question, maximum 15 words
Answer options	3 answer options consisting of 3 small individual graphics lettered A–C (grades K–4) or 4 text answer options of one to eight words (grades 2–12; both formats are used in grades 2–4)
Graphic	--

LISTENING: Synthesizing Information (Task-based Listening), grades 2–12

The test taker is required to perform—through imagined manipulation of the graphics provided—real-world tasks. The test taker must comprehend discrete information and details and then synthesize the information to perform the tasks. The language tested should simulate authentic conversational English. The language level should be at the upper level of the grade span.

Item Component	Description of Component
Advance organizer	1–2 sentence description of an authentic context for the set of items; maximum 25 words; description will appear in test booklet with a large overview graphic depicting the scenario
Stimulus	simulated authentic conversation with one person speaking; 40–50 words (grades 2–4), 50–60 words (grades 5–8), 55–70 words (grades 9–12), with distractor material embedded; there is one item per stimulus
Stem	closed stem question, maximum 15 words

A.3: Item Specification (Continued)**LISTENING: Synthesizing Information (Task-based Listening), grades 2–12**

The test taker is required to perform—through imagined manipulation of the graphics provided—real-world tasks. The test taker must comprehend discrete information and details and then synthesize the information to perform the tasks. The language tested should simulate authentic conversational English. The language level should be at the upper level of the grade span.

Item Component	Description of Component
Answer options	4 answer options consisting of either 4 small individual graphics lettered A–D or 1 large graphic with 4 options lettered A–D
Graphic	large overview graphic depicting the scenario and containing elements of the items

READING: Word Reading, grades K–1

The test taker reads a question, looks at a small graphic, and must choose which of three words correctly identifies the graphic. These items measure word recognition skills. This item type is especially designed to test students at lower levels of proficiency.

Item Component	Description of Component
Stimulus	graphic (below)
Stem	closed stem question, maximum 10 words: “What do you see? ,” “What is the boy doing?” (grades K–1), “Which word tells what is in the picture?”, “Which word tells what the girl is holding?” (grades 2–4), etc.
Answer options	3 answer options consisting of 1 word each
Graphic	graphic depicting correct answer option

A.3: Item Specification (Continued)**READING: Sentence Reading, grades K–1**

The test taker reads a question that includes a descriptive phrase, or several descriptive sentences followed by a question, and must choose which of three graphics correctly represents what is described. These items measure basic phrase and sentence comprehension.

Item Component	Description of Component
Stimulus	(optional) description of context with distractor material embedded; maximum 25 words
Stem	closed stem question, maximum 10 words: “Which one tells you _____?”; “Which one can you _____?”; etc.
Answer options	3 answer options consisting of 3 small individual graphics
Graphic	--

READING: Short Passage with Question, grades K–1

The test taker reads a brief passage and answers one question about the passage. These items measure comprehension of main ideas and details and the ability to make inferences.

Item Component	Description of Component
Stimulus	fictional, descriptive, informational, or functional passage; maximum 30 words
Stem	closed stem question, maximum 10 words
Answer options	3 answer options consisting of 3 small individual graphics
Graphic	--

A.3: Item Specification (Continued)**READING: Word/Sentence Reading, grades 2–12**

The test taker reads a question that includes a key word or a descriptive phrase, or several descriptive sentences followed by a question, and must choose which of three graphics correctly represents what is named or described. These items measure word recognition skills and basic phrase and sentence comprehension.

Item Component	Description of Component
Stimulus	(optional) description of context with distractor material embedded; maximum 25 words
Stem	closed stem question, maximum 10 words: “Which picture shows _____?”; “Which person is _____?”; etc.
Answer options	3 answer options consisting of small individual graphics
Graphic	--

READING: Comprehension, grades 2–12

The test taker reads a passage and answers several questions about the passage. These items measure comprehension of main ideas and details, the ability to make inferences, literary analysis skills, the ability to understand how and why information is organized in a passage, and other typical reading comprehension skills.

Item Component	Description of Component
Stimulus	fictional, descriptive, informational, or functional passage; maximum 200 words (grades 2–4), 250 words (grades 5–8), or 350 words (grades 9–12)
Stem	closed stem question, maximum 20 words
Answer options	4 answer options consisting of 1 word, phrase, or sentence each, maximum 15 words
Graphic	--

A.3: Item Specification (Continued)**WRITING CONVENTIONS: Phonemic Understanding, grades K–1**

The test taker listens to a question, looks at a small graphic, and must choose which of three letters or letter blends correctly relates to the graphic. These items measure understanding of phonemes as they relate to writing. This item type is especially designed to test students at beginning levels of proficiency, who often read and write based on phonics rules.

Item Component	Description of Component
Stimulus	--
Stem	closed stem question: “Which letter(s) begin(s)/end(s) the word _____?”
Answer options	3 answer options consisting of 1 letter each or 1 letter blend each
Graphic	graphic depicting the word that begins or ends with the correct answer option

WRITING CONVENTIONS: Phonemic Understanding, grades 2–12

The test taker reads a question, looks at a small graphic, and must choose which of three words correctly identifies the graphic. These items measure understanding of phonemes as they relate to writing. This item type is especially designed to test students at lower levels of proficiency, who often read and write based on phonics rules.

Item Component	Description of Component
Stimulus	--
Stem	closed stem question: “Which word goes with the picture?”
Answer options	3 answer options consisting of 1 word each
Graphic	graphic depicting correct answer option

A.3: Item Specification (Continued)**WRITING CONVENTIONS: Mechanics and Structure Type 1, grades 2–12**

The test taker must choose which of the options is a correct sentence. These items measure understanding of sentence-level grammar and other writing conventions in English.

Item Component	Description of Component
Stimulus	--
Stem	closed stem question: “Which sentence is correct?”
Answer options	3 answer options (grades K–1) or 4 answer options (grades 2–12) consisting of 1 sentence (or sentence fragment) each, maximum 20 words
Graphic	--

WRITING CONVENTIONS: Mechanics and Structure Type 2, grades 2–12

The test taker must determine if the underlined portion of the stimulus sentence is correct as is or can be corrected by replacing it with one of three answer options. These items measure understanding of sentence-level grammar and other writing conventions in English.

Item Component	Description of Component
Stimulus	1–2 sentences with part of the sentence underlined; maximum 15 words (grades 2–8) or 25 words (grades 9–12)
Stem	closed stem question: “Which answer is correct?”
Answer options	4 answer options consisting of 3 phrases to replace underlined part of stimulus and 1 “Correct as is” option
Graphic	--

WRITING: Sentence Copying, grades K–1

The test taker must copy a full sentence. (The sentence may be a question.) These items measure understanding of basic mechanical skills.

Item Component	Description of Component
Prompt	a complete sentence five to seven words in length
Graphic	--

A.3: Item Specification (Continued)**WRITING: Letter Writing, grades K–1**

The test taker must write a spoken letter. These items measure understanding of basic mechanical skills.

Item Component	Description of Component
Prompt	a word that starts with the letter to be written
Graphic	--

WRITING: Word Writing, grades K–1

The test taker must write a word that is dictated. Word dictations are accompanied by a graphic representation of the dictated word. These items measure understanding of basic mechanical skills.

Item Component	Description of Component
Prompt	a word that can be clearly represented in graphic form
Graphic	graphic depicting the dictated word

WRITING: Sentence Writing, grades K–1

The test taker must write a sentence that is dictated. (The sentence may be a question.) These items measure understanding of basic mechanical skills.

Item Component	Description of Component
Prompt	a complete sentence four to six words in length
Graphic	--

A.3: Item Specification (Continued)**PRE-WRITING: grades 2–12**

The test taker briefly responds to three short questions related to a subsequent Writing item in order to generate ideas for the Writing item.

Item Component	Description of Component
Prompt	3 questions that relate to the topic of the subsequent Writing item
Graphic	--

WRITING: Extended Response: Description and/or Narrative, grades K–2

The test taker describes or tells a story about a picture. This item measures the ability to use vocabulary, grammar, features of extended discourse, and other writing conventions correctly and appropriately in order to describe or to tell a story.

Item Component	Description of Component
Prompt	graphic and directions: “Look at the picture. Write about what you see in the picture. Tell a story about this picture.”
Graphic	1 large graphic (drawing for K–1, photograph for 2–6)

WRITING: Extended Response: Exposition, grades 2–12

The test taker compares and/or discusses activities of a particular type. This item measures the ability to use vocabulary, grammar, features of extended discourse, and other writing conventions correctly and appropriately to write an expository essay.

Item Component	Description of Component
Prompt	graphics and directions: “Write about Write an essay about”
Graphic	2 graphics (photographs) depicting related activities or situations

A.3: Item Specification (Continued)**SPEAKING: Repeat/Read Aloud, all grades**

The test taker hears and reads a word, phrase, or sentence and then repeats it. These items measure pronunciation and fluency.

Item Component	Description of Component
Prompt	a word, phrase, or sentence, one to six words in length
Graphic	--

SPEAKING: Sentence Completion, all grades

The test taker hears and reads the beginning of a sentence, and must complete the sentence based on a scenario depicted in an accompanying graphic. These items measure the ability to use vocabulary and grammar correctly and appropriately.

Item Component	Description of Component
Prompt	graphic and beginning of a sentence, maximum 10 words
Graphic	graphic depicting a recognizable scene or activity

SPEAKING: Story-Telling, all grades

The test taker must present a brief narrative based on three graphics that show a sequence of events. This item measures pronunciation, intonation, rhythm, fluency, and the ability to use vocabulary, grammar, and discourse organizational features correctly and appropriately.

Item Component	Description of Component
Prompt	graphics
Graphic	3 sequenced graphics labeled First (1), Next (2), Last (3)

A.3: Item Specification (Continued)**SPEAKING: Social Interaction, all grades**

The test taker responds to a question or statement as though in an actual conversation. These items measure appropriateness of response, as well as pronunciation, intonation, rhythm, fluency, and the ability to use vocabulary and grammar correctly.

Item Component	Description of Component
Prompt	a question or statement, maximum 10 words, requiring a rejoinder
Graphic	--

APPENDIX B: ITEM LEVEL STATISTICS BY GRADE SPAN

B.1: Kindergarten–Grade 1

Listening and Speaking

	Item Number	Item Format	Max Points	N-count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
LI	1	MC	1	50421	0	94	3	2	0	1	0.94	0.32
	2	MC	1	50421	0	3	3	92	0	1	0.92	0.39
	3	MC	1	50421	0	94	2	2	0	1	0.94	0.40
	4	MC	1	50421	0	4	92	3	0	1	0.92	0.46
	5	MC	1	50421	0	3	4	92	0	1	0.92	0.51
	6	MC	1	50421	0	2	5	93	0	1	0.93	0.39
	7	MC	1	50421	0	5	91	3	0	1	0.91	0.47
	8	MC	1	50421	0	2	9	89	0	1	0.89	0.37
	9	MC	1	50421	0	84	10	5	0	1	0.84	0.33
	10	MC	1	50421	0	88	7	4	0	1	0.88	0.40
	11	MC	1	50421	0	2	17	81	0	1	0.81	0.50
	12	MC	1	50421	0	2	86	12	0	1	0.86	0.32
	13	MC	1	50421	0	78	11	10	0	1	0.78	0.38
	14	MC	1	50421	0	10	13	77	0	1	0.77	0.44
	15	MC	1	50421	0	93	4	3	0	1	0.93	0.32
	16	MC	1	50421	0	4	94	2	0	1	0.94	0.33
	17	MC	1	50421	0	12	86	1	0	1	0.86	0.40
	18	MC	1	50421	0	5	6	88	0	1	0.88	0.48
	19	MC	1	50421	0	87	6	5	0	1	0.87	0.41
	20	MC	1	50421	0	8	88	3	0	1	0.88	0.41
	21	MC	1	50421	0	16	80	3	0	1	0.80	0.39
	22	MC	1	50421	0	7	17	75	0	1	0.75	0.37
	23	MC	1	50421	0	73	8	18	0	1	0.73	0.29
	24	MC	1	50421	0	19	51	29	0	1	0.51	0.28
SP	1	SS	2	50421	2	11	87	0	0	0	0.92	0.58
	2	SS	2	50421	3	20	77	0	0	0	0.87	0.62
	3	SS	2	50421	10	42	48	0	0	0	0.69	0.67
	4	SS	2	50421	6	14	80	0	0	0	0.87	0.68
	5	SS	2	50421	10	43	48	0	0	0	0.69	0.69
	6	SS	2	50421	6	22	71	0	0	0	0.83	0.73
	7	SS	2	50421	8	36	56	0	0	0	0.74	0.72
	8	SE	4	50421	6	9	28	35	22	0	0.65	0.81
	9	SS	2	50421	6	31	64	0	0	0	0.79	0.74
	10	SS	2	50421	6	35	59	0	0	0	0.76	0.73
	11	SS	2	50421	7	43	50	0	0	0	0.71	0.73
	12	SS	2	50421	10	23	67	0	0	0	0.79	0.75
	13	SS	2	50421	17	24	59	0	0	0	0.71	0.70
	14	SS	2	50421	14	22	64	0	0	0	0.75	0.76
	15	SS	2	50421	11	23	66	0	0	0	0.77	0.72
	16	SS	2	50421	7	8	86	0	0	0	0.89	0.70

Note. The data file used in Table B.1–B.5 were based on the final research file (N=192,053) after cleaning (remove duplicate cases, off-level cases, and all records with Strand Raw Score = 0 or Greater than 900).

B.1: Kindergarten–Grade 1**Reading and Writing**

	Item Number	Item Format	Max Points	N- count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
RD	1	MC	1	50177	0	19	76	3	0	1	0.76	0.54
	2	MC	1	50177	0	22	17	60	0	1	0.60	0.61
	3	MC	1	50177	0	76	18	6	0	1	0.76	0.47
	4	MC	1	50177	0	21	72	6	0	1	0.72	0.46
	5	MC	1	50177	0	72	14	12	0	1	0.72	0.44
	6	MC	1	50177	0	24	22	53	0	1	0.53	0.56
	7	MC	1	50177	0	61	18	19	0	2	0.61	0.41
	8	MC	1	50177	0	26	63	9	0	1	0.63	0.58
	9	MC	1	50177	0	29	63	8	0	1	0.63	0.49
	10	MC	1	50177	0	25	15	59	0	1	0.59	0.58
	11	MC	1	50177	0	67	20	12	0	2	0.67	0.49
	12	MC	1	50177	0	56	27	15	0	2	0.56	0.55
	13	MC	1	50177	0	20	34	44	0	2	0.44	0.51
	14	MC	1	50177	0	20	66	12	0	2	0.66	0.26
	15	MC	1	50177	0	54	29	15	0	3	0.54	0.49
WR	1	MC	1	50177	0	7	4	88	0	1	0.88	0.49
	2	MC	1	50177	0	93	3	4	0	1	0.93	0.37
	3	MC	1	50177	0	88	7	5	0	1	0.88	0.43
	4	MC	1	50177	0	14	81	4	0	1	0.81	0.54
	5	MC	1	50177	0	5	4	90	0	1	0.90	0.49
	6	MC	1	50177	0	6	89	4	0	1	0.89	0.45
	7	CR	2	50177	3	31	66	0	0	0	0.82	0.49
	8	CR	2	50177	3	37	59	0	0	0	0.78	0.47
	9	CR	2	50177	7	2	92	0	0	0	0.92	0.41
	10	CR	2	50177	11	2	87	0	0	0	0.88	0.49
	11	CR	2	50177	15	44	41	0	0	0	0.63	0.73
	12	CR	2	50177	21	51	27	0	0	0	0.53	0.76
	13	CR	2	50177	24	23	53	0	0	0	0.65	0.82
	14	CR	2	50177	33	31	36	0	0	0	0.51	0.84
	15	ER	2	50177	39	31	30	0	0	0	0.46	0.83

B.2: Grade 2–Grade 4**Listening and Speaking**

	Item Number	Item Format	Max Points	N- count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
LI	1	MC	1	57450	0	4	95	1	0	0	0.95	0.47
	2	MC	1	57450	0	94	4	2	0	0	0.94	0.47
	3	MC	1	57450	0	6	7	87	0	0	0.87	0.41
	4	MC	1	57450	0	10	88	2	0	0	0.88	0.38
	5	MC	1	57450	0	3	21	75	0	0	0.75	0.54
	6	MC	1	57450	0	94	3	3	0	0	0.94	0.40
	7	MC	1	57450	0	96	1	3	0	0	0.96	0.37
	8	MC	1	57450	0	12	20	68	0	1	0.68	0.15
	9	MC	1	57450	0	90	5	5	0	0	0.90	0.42
	10	MC	1	57450	0	59	30	11	0	0	0.59	0.27
	11	MC	1	57450	0	5	84	7	4	1	0.84	0.38
	12	MC	1	57450	0	5	8	5	82	1	0.82	0.38
	13	MC	1	57450	0	14	7	66	12	1	0.66	0.43
	14	MC	1	57450	0	72	15	7	5	1	0.72	0.39
	15	MC	1	57450	0	8	77	7	8	1	0.77	0.46
	16	MC	1	57450	0	11	8	12	69	1	0.69	0.31
	17	MC	1	57450	0	7	13	63	17	1	0.63	0.36
	18	MC	1	57450	0	67	13	15	5	1	0.67	0.42
	19	MC	1	57450	0	12	13	49	26	1	0.49	0.29
	20	MC	1	57450	0	4	63	6	27	1	0.63	0.23
	21	MC	1	57450	0	7	10	6	77	1	0.77	0.48
	22	MC	1	57450	0	1	90	5	3	1	0.90	0.35
	23	MC	1	57450	0	3	12	79	6	1	0.79	0.60
	24	MC	1	57450	0	66	7	23	4	1	0.66	0.44
SP	1	SS	2	57450	2	6	92	0	0	0	0.95	0.62
	2	SS	2	57450	1	12	87	0	0	0	0.93	0.62
	3	SS	2	57450	5	19	76	0	0	0	0.85	0.74
	4	SS	2	57450	4	15	82	0	0	0	0.89	0.75
	5	SS	2	57450	5	19	76	0	0	0	0.85	0.73
	6	SS	2	57450	5	19	75	0	0	0	0.85	0.76
	7	SS	2	57450	5	21	75	0	0	0	0.85	0.71
	8	SE	4	57450	4	4	18	35	39	0	0.76	0.80
	9	SS	2	57450	4	14	83	0	0	0	0.89	0.77
	10	SS	2	57450	5	25	70	0	0	0	0.83	0.74
	11	SS	2	57450	7	41	51	0	0	0	0.72	0.68
	12	SS	2	57450	4	11	86	0	0	0	0.91	0.66
	13	SS	2	57450	6	14	79	0	0	0	0.87	0.73
	14	SS	2	57450	6	13	81	0	0	0	0.87	0.73
	15	SS	2	57450	4	16	80	0	0	0	0.88	0.73
	16	SS	2	57450	7	16	77	0	0	0	0.85	0.72

B.2: Grade 2–Grade 4**Reading and Writing**

	Item Number	Item Format	Max Points	N-count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
RD	1	MC	1	57224	0	93	5	2	0	0	0.93	0.47
	2	MC	1	57224	0	3	93	3	0	0	0.93	0.46
	3	MC	1	57224	0	90	7	3	0	0	0.90	0.51
	4	MC	1	57224	0	3	95	3	0	0	0.95	0.45
	5	MC	1	57224	0	3	5	92	0	0	0.92	0.47
	6	MC	1	57224	0	90	5	5	0	0	0.90	0.42
	7	MC	1	57224	0	13	6	80	0	0	0.80	0.45
	8	MC	1	57224	0	7	6	4	83	1	0.83	0.60
	9	MC	1	57224	0	78	5	9	6	1	0.78	0.54
	10	MC	1	57224	0	67	20	5	7	1	0.67	0.46
	11	MC	1	57224	0	13	58	12	16	1	0.58	0.53
	12	MC	1	57224	0	8	11	71	9	1	0.71	0.49
	13	MC	1	57224	0	16	20	17	46	1	0.46	0.41
	14	MC	1	57224	0	14	38	38	9	1	0.38	0.34
	15	MC	1	57224	0	11	55	14	19	1	0.55	0.53
	16	MC	1	57224	0	12	60	17	9	1	0.60	0.41
	17	MC	1	57224	0	25	15	46	13	1	0.46	0.40
	18	MC	1	57224	0	10	11	9	68	1	0.68	0.54
	19	MC	1	57224	0	64	15	11	8	2	0.64	0.45
	20	MC	1	57224	0	10	56	13	19	1	0.56	0.50
	21	MC	1	57224	0	22	9	53	15	2	0.53	0.49
	22	MC	1	57224	0	15	58	15	11	1	0.58	0.52
	23	MC	1	57224	0	51	22	12	13	1	0.51	0.50
	24	MC	1	57224	0	18	10	12	58	1	0.58	0.50
WR	1	MC	1	57224	0	92	4	3	0	0	0.92	0.48
	2	MC	1	57224	0	9	86	5	0	0	0.86	0.48
	3	MC	1	57224	0	8	5	87	0	0	0.87	0.54
	4	MC	1	57224	0	15	8	77	0	1	0.77	0.50
	5	MC	1	57224	0	6	9	7	78	1	0.78	0.60
	6	MC	1	57224	0	5	4	13	78	1	0.78	0.54
	7	MC	1	57224	0	53	8	12	26	1	0.53	0.37
	8	MC	1	57224	0	19	41	20	19	1	0.41	0.36
	9	MC	1	57224	0	6	10	5	78	1	0.78	0.57
	10	MC	1	57224	0	11	64	7	17	1	0.64	0.46
	11	MC	1	57224	0	6	13	68	12	1	0.68	0.51
	12	MC	1	57224	0	50	7	18	24	1	0.50	0.40
	13	CR	2	57224	9	18	73	0	0	0	0.82	0.67
	14	CR	2	57224	12	28	59	0	0	0	0.73	0.69
	15	CR	2	57224	14	28	58	0	0	0	0.72	0.69
	16	ER	4	57224	11	21	38	23	8	0	0.49	0.77

B.3: Grade 5–Grade 6**Listening and Speaking**

	Item Number	Item Format	Max Points	N- count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
LI	1	MC	1	24767	0	93	5	1	0	0	0.93	0.52
	2	MC	1	24767	0	3	94	3	0	0	0.94	0.52
	3	MC	1	24767	0	90	8	2	0	0	0.90	0.46
	4	MC	1	24767	0	4	5	5	84	1	0.84	0.45
	5	MC	1	24767	0	3	7	87	3	0	0.87	0.50
	6	MC	1	24767	0	5	3	6	86	0	0.86	0.63
	7	MC	1	24767	0	9	84	3	4	0	0.84	0.53
	8	MC	1	24767	0	74	7	8	11	0	0.74	0.41
	9	MC	1	24767	0	5	69	13	12	0	0.69	0.34
	10	MC	1	24767	0	59	23	10	7	0	0.59	0.38
	11	MC	1	24767	0	8	14	12	66	1	0.66	0.43
	12	MC	1	24767	0	13	63	19	5	0	0.63	0.34
	13	MC	1	24767	0	3	88	4	5	0	0.88	0.61
	14	MC	1	24767	0	13	10	60	17	0	0.60	0.26
	15	MC	1	24767	0	3	6	10	81	0	0.81	0.50
	16	MC	1	24767	0	84	7	5	3	1	0.84	0.48
	17	MC	1	24767	0	5	11	6	78	1	0.78	0.49
	18	MC	1	24767	0	10	73	13	3	0	0.73	0.50
	19	MC	1	24767	0	5	16	73	6	1	0.73	0.39
	20	MC	1	24767	0	82	7	3	7	0	0.82	0.45
	21	MC	1	24767	0	4	5	86	5	0	0.86	0.53
	22	MC	1	24767	0	13	67	6	14	0	0.67	0.45
	23	MC	1	24767	0	20	26	1	52	0	0.52	0.33
	24	MC	1	24767	0	7	6	84	3	0	0.84	0.49
	25	MC	1	24767	0	3	2	75	19	1	0.75	0.29
SP	1	SS	2	24767	2	8	90	0	0	0	0.94	0.62
	2	SS	2	24767	2	9	89	0	0	0	0.93	0.65
	3	SS	2	24767	7	24	70	0	0	0	0.81	0.79
	4	SS	2	24767	4	28	67	0	0	0	0.82	0.68
	5	SS	2	24767	5	21	74	0	0	0	0.85	0.71
	6	SS	2	24767	8	27	65	0	0	0	0.78	0.77
	7	SS	2	24767	6	28	66	0	0	0	0.80	0.73
	8	SE	4	24767	6	5	16	33	40	0	0.74	0.85
	9	SS	2	24767	7	23	70	0	0	0	0.82	0.81
	10	SS	2	24767	5	20	75	0	0	0	0.85	0.80
	11	SS	2	24767	6	25	69	0	0	0	0.81	0.78
	12	SS	2	24767	6	11	84	0	0	0	0.89	0.78
	13	SS	2	24767	9	12	79	0	0	0	0.85	0.81
	14	SS	2	24767	6	13	82	0	0	0	0.88	0.78
	15	SS	2	24767	8	15	77	0	0	0	0.84	0.76
	16	SS	2	24767	8	8	85	0	0	0	0.89	0.79

B.3: Grade 5–Grade 6

Reading and Writing

	Item Number	Item Format	Max Points	N-count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
RD	1	MC	1	24620	0	1	2	96	0	0	0.96	0.36
	2	MC	1	24620	0	2	96	2	0	0	0.96	0.33
	3	MC	1	24620	0	4	91	4	0	0	0.91	0.46
	4	MC	1	24620	0	18	3	4	74	0	0.74	0.38
	5	MC	1	24620	0	6	77	10	7	0	0.77	0.53
	6	MC	1	24620	0	74	10	9	7	1	0.74	0.52
	7	MC	1	24620	0	22	13	8	56	0	0.56	0.51
	8	MC	1	24620	0	5	87	5	4	0	0.87	0.54
	9	MC	1	24620	0	78	6	6	9	0	0.78	0.57
	10	MC	1	24620	0	27	10	56	6	0	0.56	0.25
	11	MC	1	24620	0	28	14	15	43	0	0.43	0.43
	12	MC	1	24620	0	11	5	76	7	1	0.76	0.31
	13	MC	1	24620	0	37	21	27	16	1	0.37	0.31
	14	MC	1	24620	0	7	7	13	73	0	0.73	0.52
	15	MC	1	24620	0	8	10	12	69	0	0.69	0.59
	16	MC	1	24620	0	14	16	58	12	1	0.58	0.49
	17	MC	1	24620	0	13	9	22	57	1	0.57	0.42
	18	MC	1	24620	0	62	11	11	15	1	0.62	0.48
	19	MC	1	24620	0	59	10	22	9	1	0.59	0.51
	20	MC	1	24620	0	23	42	15	19	1	0.42	0.33
	21	MC	1	24620	0	56	12	23	8	1	0.56	0.52
	22	MC	1	24620	0	11	59	14	15	1	0.59	0.55
	23	MC	1	24620	0	8	9	73	9	1	0.73	0.54
	24	MC	1	24620	0	21	46	17	15	1	0.46	0.33
	25	MC	1	24620	0	11	7	64	18	1	0.64	0.35
	26	MC	1	24620	0	57	21	13	8	1	0.57	0.46
	27	MC	1	24620	0	17	13	13	57	1	0.57	0.43
WR	1	MC	1	24620	0	8	5	87	0	0	0.87	0.46
	2	MC	1	24620	0	12	85	3	0	0	0.85	0.41
	3	MC	1	24620	0	6	2	88	4	0	0.88	0.45
	4	MC	1	24620	0	3	14	11	71	1	0.71	0.40
	5	MC	1	24620	0	9	74	12	5	1	0.74	0.55
	6	MC	1	24620	0	16	7	57	19	1	0.57	0.39
	7	MC	1	24620	0	23	48	13	16	1	0.48	0.34
	8	MC	1	24620	0	85	4	7	3	1	0.85	0.54
	9	MC	1	24620	0	3	83	5	8	1	0.83	0.43
	10	MC	1	24620	0	82	4	5	8	1	0.82	0.59
	11	MC	1	24620	0	5	7	18	69	1	0.69	0.51
	12	MC	1	24620	0	81	4	5	10	1	0.81	0.54
	13	MC	1	24620	0	7	14	62	17	1	0.62	0.54
	14	MC	1	24620	0	13	15	6	64	1	0.64	0.42
	15	MC	1	24620	0	21	48	16	14	1	0.48	0.41
	16	CR	2	24620	6	35	59	0	0	0	0.76	0.51
	17	CR	2	24620	8	25	68	0	0	0	0.80	0.62
	18	CR	2	24620	8	15	76	0	0	0	0.84	0.63
	19	ER	4	24620	7	12	33	32	16	0	0.60	0.76

B.4: Grade 7–Grade 8

Listening and Speaking

	Item Number	Item Format	Max Points	N-count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
LI	1	MC	1	22602	0	3	90	7	0	0	0.90	0.40
	2	MC	1	22602	0	2	4	93	0	0	0.93	0.38
	3	MC	1	22602	0	4	91	5	0	0	0.91	0.40
	4	MC	1	22602	0	6	74	9	10	0	0.74	0.52
	5	MC	1	22602	0	3	15	66	17	0	0.66	0.37
	6	MC	1	22602	0	9	5	80	5	0	0.80	0.49
	7	MC	1	22602	0	81	9	5	5	0	0.81	0.51
	8	MC	1	22602	0	74	15	5	6	0	0.74	0.63
	9	MC	1	22602	0	7	55	25	13	0	0.55	0.23
	10	MC	1	22602	0	10	11	5	73	0	0.73	0.54
	11	MC	1	22602	0	4	9	6	81	0	0.81	0.54
	12	MC	1	22602	0	48	12	19	21	1	0.48	0.37
	13	MC	1	22602	0	59	26	7	8	0	0.59	0.50
	14	MC	1	22602	0	5	11	15	69	0	0.69	0.41
	15	MC	1	22602	0	5	12	63	20	0	0.63	0.39
	16	MC	1	22602	0	77	10	8	4	0	0.77	0.48
	17	MC	1	22602	0	7	72	16	5	0	0.72	0.35
	18	MC	1	22602	0	17	16	54	13	1	0.54	0.36
	19	MC	1	22602	0	22	18	6	54	0	0.54	0.36
	20	MC	1	22602	0	10	7	73	10	1	0.73	0.52
	21	MC	1	22602	0	12	69	12	6	1	0.69	0.40
	22	MC	1	22602	0	7	80	4	8	0	0.80	0.56
	23	MC	1	22602	0	71	18	6	5	0	0.71	0.40
	24	MC	1	22602	0	16	9	7	68	0	0.68	0.49
	25	MC	1	22602	0	3	4	15	78	1	0.78	0.51
SP	1	SS	2	22602	3	8	89	0	0	0	0.93	0.63
	2	SS	2	22602	4	15	82	0	0	0	0.89	0.70
	3	SS	2	22602	9	25	66	0	0	0	0.78	0.81
	4	SS	2	22602	10	24	66	0	0	0	0.78	0.80
	5	SS	2	22602	10	25	65	0	0	0	0.77	0.80
	6	SS	2	22602	7	25	67	0	0	0	0.80	0.79
	7	SS	2	22602	11	27	62	0	0	0	0.76	0.79
	8	SE	4	22602	9	8	18	29	36	0	0.69	0.87
	9	SS	2	22602	8	27	65	0	0	0	0.79	0.81
	10	SS	2	22602	9	26	65	0	0	0	0.78	0.80
	11	SS	2	22602	8	20	72	0	0	0	0.82	0.81
	12	SS	2	22602	9	14	77	0	0	0	0.84	0.76
	13	SS	2	22602	10	11	78	0	0	0	0.84	0.81
	14	SS	2	22602	8	12	79	0	0	0	0.86	0.79
	15	SS	2	22602	8	13	79	0	0	0	0.85	0.79
	16	SS	2	22602	13	20	68	0	0	0	0.77	0.81

B.4: Grade 7–Grade 8**Reading and Writing**

	Item Number	Item Format	Max Points	N- count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
RD	1	MC	1	22415	0	3	5	93	0	0	0.93	0.35
	2	MC	1	22415	0	8	88	4	0	0	0.88	0.46
	3	MC	1	22415	0	88	6	7	0	0	0.88	0.45
	4	MC	1	22415	0	4	8	82	5	0	0.82	0.49
	5	MC	1	22415	0	64	13	12	10	0	0.64	0.46
	6	MC	1	22415	0	25	53	11	11	0	0.53	0.43
	7	MC	1	22415	0	33	11	45	10	0	0.45	0.39
	8	MC	1	22415	0	18	49	12	21	0	0.49	0.46
	9	MC	1	22415	0	78	11	4	7	0	0.78	0.52
	10	MC	1	22415	0	9	14	16	61	0	0.61	0.65
	11	MC	1	22415	0	7	15	16	61	0	0.61	0.49
	12	MC	1	22415	0	9	22	45	24	0	0.45	0.29
	13	MC	1	22415	0	9	14	11	66	1	0.66	0.46
	14	MC	1	22415	0	53	11	23	13	0	0.53	0.43
	15	MC	1	22415	0	22	11	57	10	0	0.57	0.50
	16	MC	1	22415	0	14	21	13	51	0	0.51	0.46
	17	MC	1	22415	0	53	13	16	17	0	0.53	0.34
	18	MC	1	22415	0	22	56	12	10	0	0.56	0.49
	19	MC	1	22415	0	11	15	20	53	0	0.53	0.39
	20	MC	1	22415	0	16	16	19	48	1	0.48	0.46
	21	MC	1	22415	0	9	12	72	7	1	0.72	0.57
	22	MC	1	22415	0	66	14	10	10	0	0.66	0.58
	23	MC	1	22415	0	15	17	7	60	1	0.60	0.41
	24	MC	1	22415	0	22	16	44	18	1	0.44	0.32
	25	MC	1	22415	0	13	41	26	19	1	0.41	0.34
	26	MC	1	22415	0	42	21	22	14	1	0.42	0.30
	27	MC	1	22415	0	7	45	21	26	1	0.45	0.44
WR	1	MC	1	22415	0	96	2	2	0	1	0.96	0.33
	2	MC	1	22415	0	2	2	95	0	1	0.95	0.32
	3	MC	1	22415	0	5	84	3	6	1	0.84	0.46
	4	MC	1	22415	0	3	6	9	81	1	0.81	0.33
	5	MC	1	22415	0	83	3	9	4	1	0.83	0.54
	6	MC	1	22415	0	69	13	12	5	1	0.69	0.50
	7	MC	1	22415	0	8	11	57	23	1	0.57	0.38
	8	MC	1	22415	0	54	10	17	17	1	0.54	0.47
	9	MC	1	22415	0	16	5	9	70	1	0.70	0.57
	10	MC	1	22415	0	14	77	4	3	1	0.77	0.58
	11	MC	1	22415	0	4	4	28	63	1	0.63	0.43
	12	MC	1	22415	0	8	14	67	11	1	0.67	0.54
	13	MC	1	22415	0	15	66	14	4	1	0.66	0.50
	14	MC	1	22415	0	14	5	74	6	1	0.74	0.51
	15	MC	1	22415	0	15	49	20	15	1	0.49	0.44
	16	CR	2	22415	12	32	56	0	0	0	0.72	0.65
	17	CR	2	22415	13	32	55	0	0	0	0.71	0.66
	18	CR	2	22415	20	27	53	0	0	0	0.67	0.65
	19	ER	4	22415	13	17	35	26	9	0	0.50	0.77

B.5: Grade 9–Grade 12**Listening and Speaking**

	Item Number	Item Format	Max Points	N- count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
LI	1	MC	1	35473	0	5	2	93	0	0	0.93	0.37
	2	MC	1	35473	0	92	4	4	0	0	0.92	0.44
	3	MC	1	35473	0	86	3	10	0	0	0.86	0.41
	4	MC	1	35473	0	6	86	5	3	1	0.86	0.40
	5	MC	1	35473	0	10	6	17	66	1	0.66	0.37
	6	MC	1	35473	0	9	3	85	3	0	0.85	0.36
	7	MC	1	35473	0	14	61	12	12	1	0.61	0.43
	8	MC	1	35473	0	24	8	62	6	0	0.62	0.42
	9	MC	1	35473	0	6	10	7	77	1	0.77	0.52
	10	MC	1	35473	0	53	31	9	8	0	0.53	0.44
	11	MC	1	35473	0	16	55	18	11	1	0.55	0.37
	12	MC	1	35473	0	34	47	14	5	1	0.47	0.23
	13	MC	1	35473	0	8	48	41	2	1	0.41	0.18
	14	MC	1	35473	0	9	8	68	15	1	0.68	0.33
	15	MC	1	35473	0	62	13	7	17	0	0.62	0.38
	16	MC	1	35473	0	7	3	7	82	0	0.82	0.39
	17	MC	1	35473	0	2	11	4	83	0	0.83	0.41
	18	MC	1	35473	0	11	4	77	7	0	0.77	0.46
	19	MC	1	35473	0	72	11	6	10	1	0.72	0.25
	20	MC	1	35473	0	86	6	5	4	0	0.86	0.41
	21	MC	1	35473	0	18	70	4	7	1	0.70	0.27
	22	MC	1	35473	0	8	6	26	59	1	0.59	0.29
	23	MC	1	35473	0	69	8	8	14	1	0.69	0.48
	24	MC	1	35473	0	20	13	9	57	1	0.57	0.38
	25	MC	1	35473	0	23	61	11	5	1	0.61	0.25
SP	1	SS	2	35473	5	18	77	0	0	0	0.86	0.67
	2	SS	2	35473	5	26	69	0	0	0	0.82	0.67
	3	SS	2	35473	9	32	59	0	0	0	0.75	0.77
	4	SS	2	35473	9	39	52	0	0	0	0.71	0.76
	5	SS	2	35473	12	32	56	0	0	0	0.72	0.78
	6	SS	2	35473	14	36	51	0	0	0	0.69	0.76
	7	SS	2	35473	7	35	57	0	0	0	0.75	0.73
	8	SE	4	35473	7	8	20	31	34	0	0.69	0.87
	9	SS	2	35473	8	30	62	0	0	0	0.77	0.78
	10	SS	2	35473	7	36	57	0	0	0	0.75	0.78
	11	SS	2	35473	10	32	59	0	0	0	0.75	0.78
	12	SS	2	35473	7	22	71	0	0	0	0.82	0.76
	13	SS	2	35473	9	25	66	0	0	0	0.79	0.74
	14	SS	2	35473	12	30	58	0	0	0	0.73	0.78
	15	SS	2	35473	11	26	63	0	0	0	0.76	0.81
	16	SS	2	35473	13	19	68	0	0	0	0.78	0.80

B.5: Grade 9–Grade 12

Reading and Writing

	Item Number	Item Format	Max Points	N-count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
RD	1	MC	1	34654	0	87	9	4	0	0	0.87	0.37
	2	MC	1	34654	0	7	75	18	0	0	0.75	0.36
	3	MC	1	34654	0	18	8	73	0	0	0.73	0.22
	4	MC	1	34654	0	8	6	5	81	0	0.81	0.31
	5	MC	1	34654	0	6	83	5	5	0	0.83	0.44
	6	MC	1	34654	0	5	19	59	17	1	0.59	0.44
	7	MC	1	34654	0	24	9	12	55	0	0.55	0.41
	8	MC	1	34654	0	16	68	9	6	0	0.68	0.48
	9	MC	1	34654	0	48	19	22	10	1	0.48	0.43
	10	MC	1	34654	0	5	15	72	7	1	0.72	0.32
	11	MC	1	34654	0	22	8	25	45	0	0.45	0.26
	12	MC	1	34654	0	55	30	5	9	0	0.55	0.35
	13	MC	1	34654	0	59	16	13	11	1	0.59	0.54
	14	MC	1	34654	0	3	5	6	85	1	0.85	0.45
	15	MC	1	34654	0	12	70	10	8	0	0.70	0.46
	16	MC	1	34654	0	13	12	10	64	1	0.64	0.63
	17	MC	1	34654	0	23	12	49	15	1	0.49	0.38
	18	MC	1	34654	0	14	23	17	45	1	0.45	0.45
	19	MC	1	34654	0	74	14	6	5	1	0.74	0.35
	20	MC	1	34654	0	8	19	66	6	1	0.66	0.39
	21	MC	1	34654	0	12	65	8	14	1	0.65	0.49
	22	MC	1	34654	0	72	8	13	6	1	0.72	0.45
	23	MC	1	34654	0	74	7	9	9	1	0.74	0.48
	24	MC	1	34654	0	19	8	65	7	1	0.65	0.36
	25	MC	1	34654	0	7	11	64	16	2	0.64	0.32
	26	MC	1	34654	0	74	9	10	5	1	0.74	0.34
	27	MC	1	34654	0	20	49	18	11	1	0.49	0.36
WR	1	MC	1	34654	0	87	6	5	0	2	0.87	0.47
	2	MC	1	34654	0	14	15	69	0	2	0.69	0.43
	3	MC	1	34654	0	12	7	67	11	2	0.67	0.46
	4	MC	1	34654	0	72	17	2	7	2	0.72	0.46
	5	MC	1	34654	0	4	8	10	76	2	0.76	0.48
	6	MC	1	34654	0	20	10	55	13	2	0.55	0.50
	7	MC	1	34654	0	20	63	6	8	2	0.63	0.46
	8	MC	1	34654	0	9	74	9	6	2	0.74	0.43
	9	MC	1	34654	0	5	11	67	14	2	0.67	0.50
	10	MC	1	34654	0	59	16	12	10	2	0.59	0.46
	11	MC	1	34654	0	19	69	6	5	2	0.69	0.50
	12	MC	1	34654	0	12	13	12	62	2	0.62	0.50
	13	MC	1	34654	0	13	48	6	30	2	0.48	0.28
	14	MC	1	34654	0	61	10	14	14	2	0.61	0.59
	15	MC	1	34654	0	11	13	15	59	2	0.59	0.47
	16	CR	2	34654	14	22	64	0	0	0	0.75	0.65
	17	CR	2	34654	20	27	52	0	0	0	0.66	0.67
	18	CR	2	34654	23	34	42	0	0	0	0.60	0.67
	19	ER	4	34654	13	16	28	28	16	0	0.54	0.78

APPENDIX C: RAW SCORE TO SCALE SCORE CONVERSION TABLES

C.1: Kindergarten–Grade 1

Listening and Speaking

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-7.92	2.01	323	70
1	-6.51	1.02	372	36
2	-5.78	0.73	398	26
3	-5.33	0.61	413	21
4	-5.01	0.54	425	19
5	-4.75	0.49	434	17
6	-4.53	0.45	441	16
7	-4.34	0.42	448	15
8	-4.17	0.40	454	14
9	-4.02	0.38	459	13
10	-3.87	0.37	464	13
11	-3.74	0.36	469	12
12	-3.62	0.35	473	12
13	-3.50	0.34	477	12
14	-3.39	0.33	481	12
15	-3.28	0.32	485	11
16	-3.18	0.32	489	11
17	-3.08	0.31	492	11
18	-2.99	0.31	495	11
19	-2.89	0.30	499	11
20	-2.80	0.30	502	10
21	-2.71	0.30	505	10
22	-2.63	0.29	508	10
23	-2.54	0.29	511	10
24	-2.45	0.29	514	10
25	-2.37	0.29	517	10
26	-2.29	0.29	520	10
27	-2.20	0.29	523	10
28	-2.12	0.29	526	10
29	-2.04	0.29	529	10
30	-1.95	0.29	532	10
31	-1.87	0.29	535	10
32	-1.79	0.29	538	10
33	-1.70	0.29	540	10
34	-1.61	0.29	543	10
35	-1.53	0.30	547	10
36	-1.44	0.30	550	11
37	-1.35	0.30	553	11
38	-1.25	0.31	556	11

C.1: Kindergarten–Grade 1**Listening and Speaking**

Raw Score	Measure	SE	Scale Score	SE(SS)
39	-1.16	0.31	559	11
40	-1.06	0.32	563	11
41	-0.96	0.32	566	11
42	-0.85	0.33	570	11
43	-0.74	0.33	574	12
44	-0.63	0.34	578	12
45	-0.51	0.35	582	12
46	-0.38	0.36	587	13
47	-0.25	0.37	591	13
48	-0.11	0.39	596	14
49	0.05	0.40	602	14
50	0.22	0.42	608	15
51	0.41	0.44	614	16
52	0.61	0.47	622	17
53	0.85	0.51	630	18
54	1.14	0.56	640	20
55	1.49	0.63	652	22
56	1.97	0.76	669	26
57	2.74	1.04	696	36
58	4.19	2.02	747	71

Note. The statistics in Tables C.1–C.5 were based on the equating sample (Total N=191,642). Invalid cases were removed after cleaning procedures.

*C.1: Kindergarten–Grade 1***Reading and Writing**

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-7.81	2.03	327	71
1	-6.34	1.05	378	37
2	-5.54	0.77	406	27
3	-5.04	0.65	423	23
4	-4.67	0.58	437	20
5	-4.37	0.52	447	18
6	-4.12	0.48	456	17
7	-3.90	0.45	464	16
8	-3.71	0.43	470	15
9	-3.53	0.41	476	14
10	-3.37	0.40	482	14
11	-3.21	0.39	487	14
12	-3.07	0.38	493	13
13	-2.93	0.37	498	13
14	-2.79	0.37	502	13
15	-2.65	0.37	507	13
16	-2.52	0.37	512	13
17	-2.38	0.37	517	13
18	-2.25	0.37	521	13
19	-2.11	0.37	526	13
20	-1.98	0.37	531	13
21	-1.84	0.37	536	13
22	-1.70	0.37	540	13
23	-1.56	0.37	545	13
24	-1.42	0.38	550	13
25	-1.28	0.38	555	13
26	-1.13	0.39	560	13
27	-0.98	0.39	566	14
28	-0.83	0.40	571	14
29	-0.66	0.41	577	14
30	-0.49	0.42	583	15
31	-0.31	0.44	589	15
32	-0.11	0.46	596	16
33	0.11	0.48	604	17
34	0.35	0.51	612	18
35	0.64	0.56	622	20
36	0.99	0.63	635	22
37	1.47	0.75	651	26
38	2.23	1.03	678	36
39	3.66	2.02	728	71

C.2: Grade 2–Grade 4**Listening and Speaking**

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-6.57	2.02	370	71
1	-5.12	1.04	421	36
2	-4.36	0.75	448	26
3	-3.89	0.63	464	22
4	-3.54	0.55	476	19
5	-3.26	0.50	486	18
6	-3.03	0.46	494	16
7	-2.83	0.43	501	15
8	-2.65	0.41	507	14
9	-2.49	0.39	513	14
10	-2.34	0.38	518	13
11	-2.21	0.36	523	13
12	-2.08	0.35	527	12
13	-1.96	0.34	531	12
14	-1.85	0.33	535	12
15	-1.74	0.32	539	11
16	-1.64	0.32	543	11
17	-1.54	0.31	546	11
18	-1.44	0.31	549	11
19	-1.35	0.30	553	11
20	-1.26	0.30	556	10
21	-1.17	0.30	559	10
22	-1.08	0.29	562	10
23	-1.00	0.29	565	10
24	-0.91	0.29	568	10
25	-0.83	0.29	571	10
26	-0.75	0.29	574	10
27	-0.66	0.29	577	10
28	-0.58	0.29	580	10
29	-0.50	0.29	583	10
30	-0.41	0.29	586	10
31	-0.33	0.29	588	10
32	-0.25	0.29	591	10
33	-0.16	0.29	594	10
34	-0.07	0.29	597	10
35	0.01	0.30	600	10
36	0.10	0.30	604	10
37	0.19	0.30	607	11
38	0.29	0.31	610	11
39	0.38	0.31	613	11
40	0.48	0.32	617	11
41	0.58	0.32	620	11
42	0.68	0.33	624	11
43	0.79	0.33	628	12
44	0.90	0.34	632	12
45	1.02	0.35	636	12
46	1.15	0.36	640	12
47	1.28	0.37	645	13
48	1.42	0.38	650	13
49	1.56	0.39	655	14
50	1.73	0.41	660	14

C.2: Grade 2–Grade 4**Listening and Speaking**

Raw Score	Measure	SE	Scale Score	SE(SS)
51	1.90	0.43	667	15
52	2.10	0.46	674	16
53	2.33	0.49	681	17
54	2.59	0.54	691	19
55	2.92	0.61	702	21
56	3.37	0.74	718	26
57	4.11	1.02	744	36
58	5.53	2.01	793	70

Reading and Writing

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-6.27	2.02	380	71
1	-4.84	1.03	431	36
2	-4.07	0.75	457	26
3	-3.60	0.63	474	22
4	-3.24	0.56	486	20
5	-2.95	0.52	497	18
6	-2.70	0.48	505	17
7	-2.48	0.46	513	16
8	-2.28	0.44	520	15
9	-2.09	0.42	527	15
10	-1.92	0.41	533	14
11	-1.75	0.40	539	14
12	-1.60	0.39	544	14
13	-1.45	0.38	549	13
14	-1.30	0.38	554	13
15	-1.16	0.37	559	13
16	-1.03	0.37	564	13
17	-0.90	0.36	569	13
18	-0.77	0.36	573	12
19	-0.64	0.35	578	12
20	-0.51	0.35	582	12
21	-0.39	0.35	586	12
22	-0.27	0.35	591	12
23	-0.15	0.35	595	12
24	-0.03	0.35	599	12
25	0.09	0.35	603	12
26	0.22	0.35	608	12
27	0.34	0.35	612	12
28	0.46	0.35	616	12
29	0.59	0.36	621	12
30	0.72	0.36	625	13
31	0.85	0.36	630	13
32	0.98	0.37	634	13
33	1.12	0.38	639	13
34	1.26	0.38	644	13
35	1.41	0.39	649	14
36	1.57	0.40	655	14
37	1.74	0.42	661	15
38	1.92	0.44	667	15

C.2: Grade 2–Grade 4**Reading and Writing**

Raw Score	Measure	SE	Scale Score	SE(SS)
39	2.12	0.46	674	16
40	2.34	0.48	682	17
41	2.60	0.52	691	18
42	2.89	0.57	701	20
43	3.26	0.64	714	23
44	3.75	0.77	731	27
45	4.54	1.05	759	37
46	6.01	2.03	810	71

C.3: Grade 5–Grade 6**Listening and Speaking**

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-6.40	2.02	376	71
1	-4.95	1.04	427	37
2	-4.17	0.76	454	27
3	-3.69	0.64	471	22
4	-3.34	0.56	483	20
5	-3.05	0.51	493	18
6	-2.81	0.47	502	16
7	-2.61	0.44	509	15
8	-2.43	0.42	515	15
9	-2.26	0.40	521	14
10	-2.11	0.38	526	13
11	-1.97	0.36	531	13
12	-1.85	0.35	535	12
13	-1.73	0.34	540	12
14	-1.61	0.33	544	12
15	-1.50	0.32	547	11
16	-1.40	0.32	551	11
17	-1.30	0.31	554	11
18	-1.21	0.31	558	11
19	-1.12	0.30	561	11
20	-1.03	0.30	564	10
21	-0.94	0.29	567	10
22	-0.85	0.29	570	10
23	-0.77	0.29	573	10
24	-0.69	0.29	576	10
25	-0.61	0.28	579	10
26	-0.52	0.28	582	10
27	-0.44	0.28	584	10
28	-0.36	0.28	587	10
29	-0.28	0.28	590	10
30	-0.20	0.28	593	10
31	-0.12	0.28	596	10
32	-0.04	0.28	598	10
33	0.04	0.29	601	10
34	0.12	0.29	604	10
35	0.20	0.29	607	10
36	0.29	0.29	610	10

C.3: Grade 5–Grade 6**Listening and Speaking**

Raw Score	Measure	SE	Scale Score	SE(SS)
37	0.37	0.30	613	10
38	0.46	0.30	616	10
39	0.55	0.30	619	11
40	0.64	0.31	623	11
41	0.74	0.31	626	11
42	0.84	0.32	629	11
43	0.94	0.32	633	11
44	1.04	0.33	636	11
45	1.15	0.33	640	12
46	1.27	0.34	644	12
47	1.39	0.35	648	12
48	1.51	0.36	653	13
49	1.65	0.37	658	13
50	1.79	0.39	663	14
51	1.95	0.41	668	14
52	2.12	0.43	674	15
53	2.32	0.45	681	16
54	2.54	0.49	689	17
55	2.80	0.54	698	19
56	3.13	0.61	709	21
57	3.57	0.73	725	26
58	4.31	1.02	751	36
59	5.72	2.01	800	70

Reading and Writing

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-5.40	2.02	411	71
1	-3.95	1.04	462	36
2	-3.18	0.76	489	27
3	-2.70	0.64	506	22
4	-2.34	0.56	518	20
5	-2.05	0.51	528	18
6	-1.80	0.48	537	17
7	-1.59	0.45	544	16
8	-1.40	0.43	551	15
9	-1.22	0.41	557	14
10	-1.06	0.39	563	14
11	-0.91	0.38	568	13
12	-0.77	0.37	573	13
13	-0.63	0.36	578	13
14	-0.51	0.35	582	12
15	-0.38	0.35	587	12
16	-0.27	0.34	591	12
17	-0.15	0.34	595	12
18	-0.04	0.33	599	12
19	0.07	0.33	602	11
20	0.17	0.32	606	11
21	0.28	0.32	610	11
22	0.38	0.32	613	11
23	0.48	0.32	617	11

C.3: Grade 5–Grade 6**Reading and Writing**

Raw Score	Measure	SE	Scale Score	SE(SS)
24	0.58	0.32	620	11
25	0.68	0.32	624	11
26	0.78	0.31	627	11
27	0.88	0.31	631	11
28	0.98	0.31	634	11
29	1.08	0.32	638	11
30	1.17	0.32	641	11
31	1.28	0.32	645	11
32	1.38	0.32	648	11
33	1.48	0.32	652	11
34	1.59	0.33	655	11
35	1.69	0.33	659	11
36	1.80	0.33	663	12
37	1.91	0.34	667	12
38	2.03	0.34	671	12
39	2.15	0.35	675	12
40	2.28	0.36	680	13
41	2.41	0.37	684	13
42	2.55	0.38	689	13
43	2.70	0.39	694	14
44	2.86	0.41	700	14
45	3.04	0.43	706	15
46	3.23	0.46	713	16
47	3.46	0.49	721	17
48	3.72	0.54	730	19
49	4.05	0.61	742	21
50	4.50	0.74	758	26
51	5.24	1.02	783	36
52	6.65	2.01	833	70

C.4: Grade 7–Grade 8**Listening and Speaking**

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-5.31	2.01	414	70
1	-3.88	1.03	464	36
2	-3.13	0.74	490	26
3	-2.67	0.62	506	22
4	-2.34	0.54	518	19
5	-2.07	0.49	528	17
6	-1.85	0.46	535	16
7	-1.65	0.43	542	15
8	-1.48	0.40	548	14
9	-1.32	0.39	554	13
10	-1.18	0.37	559	13
11	-1.05	0.36	563	12
12	-0.93	0.34	568	12
13	-0.81	0.33	572	12
14	-0.70	0.33	575	11
15	-0.60	0.32	579	11

C.4: Grade 7–Grade 8**Listening and Speaking**

Raw Score	Measure	SE	Scale Score	SE(SS)
16	-0.50	0.31	583	11
17	-0.40	0.31	586	11
18	-0.31	0.30	589	11
19	-0.22	0.30	592	10
20	-0.14	0.29	595	10
21	-0.05	0.29	598	10
22	0.03	0.29	601	10
23	0.11	0.28	604	10
24	0.19	0.28	607	10
25	0.27	0.28	609	10
26	0.35	0.28	612	10
27	0.42	0.28	615	10
28	0.50	0.28	618	10
29	0.58	0.28	620	10
30	0.66	0.28	623	10
31	0.73	0.28	626	10
32	0.81	0.28	628	10
33	0.89	0.28	631	10
34	0.97	0.28	634	10
35	1.05	0.29	637	10
36	1.13	0.29	640	10
37	1.22	0.29	643	10
38	1.30	0.29	646	10
39	1.39	0.30	649	10
40	1.48	0.30	652	11
41	1.57	0.31	655	11
42	1.67	0.31	658	11
43	1.77	0.32	662	11
44	1.87	0.32	665	11
45	1.97	0.33	669	12
46	2.09	0.34	673	12
47	2.20	0.35	677	12
48	2.33	0.36	681	12
49	2.46	0.37	686	13
50	2.60	0.39	691	13
51	2.76	0.40	696	14
52	2.93	0.42	702	15
53	3.12	0.45	709	16
54	3.34	0.49	717	17
55	3.60	0.54	726	19
56	3.92	0.61	737	21
57	4.37	0.73	753	26
58	5.10	1.02	778	36
59	6.51	2.01	828	70

C.4: Grade 7–Grade 8**Reading and Writing**

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-5.41	2.02	411	71
1	-3.96	1.04	461	36
2	-3.18	0.76	489	27
3	-2.69	0.64	506	22
4	-2.33	0.57	518	20
5	-2.03	0.52	529	18
6	-1.78	0.49	538	17
7	-1.56	0.46	546	16
8	-1.36	0.44	553	15
9	-1.17	0.42	559	15
10	-1.01	0.40	565	14
11	-0.85	0.39	570	14
12	-0.70	0.38	575	13
13	-0.56	0.37	580	13
14	-0.43	0.36	585	13
15	-0.30	0.35	589	12
16	-0.18	0.35	594	12
17	-0.07	0.34	598	12
18	0.05	0.33	602	12
19	0.16	0.33	605	12
20	0.26	0.33	609	11
21	0.37	0.32	613	11
22	0.47	0.32	617	11
23	0.57	0.32	620	11
24	0.67	0.32	624	11
25	0.77	0.31	627	11
26	0.87	0.31	630	11
27	0.97	0.31	634	11
28	1.07	0.31	637	11
29	1.16	0.31	641	11
30	1.26	0.31	644	11
31	1.36	0.31	648	11
32	1.46	0.32	651	11
33	1.56	0.32	655	11
34	1.66	0.32	658	11
35	1.77	0.33	662	11
36	1.88	0.33	666	12
37	1.99	0.33	670	12
38	2.10	0.34	674	12
39	2.22	0.35	678	12
40	2.34	0.36	682	12
41	2.47	0.37	687	13
42	2.61	0.38	691	13
43	2.76	0.39	697	14
44	2.92	0.41	702	14
45	3.10	0.43	708	15
46	3.29	0.46	715	16
47	3.52	0.49	723	17
48	3.79	0.54	733	19
49	4.12	0.62	744	22
50	4.57	0.74	760	26

C.4: Grade 7–Grade 8**Reading and Writing**

Raw Score	Measure	SE	Scale Score	SE(SS)
51	5.32	1.03	786	36
52	6.74	2.01	836	70

C.5: Grade 9–Grade 12**Listening and Speaking**

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-5.11	2.01	421	70
1	-3.69	1.03	471	36
2	-2.94	0.74	497	26
3	-2.49	0.62	513	22
4	-2.15	0.55	525	19
5	-1.88	0.50	534	17
6	-1.65	0.46	542	16
7	-1.45	0.43	549	15
8	-1.27	0.41	555	14
9	-1.11	0.40	561	14
10	-0.96	0.38	566	13
11	-0.82	0.37	571	13
12	-0.69	0.36	576	12
13	-0.56	0.35	580	12
14	-0.45	0.34	584	12
15	-0.33	0.33	588	12
16	-0.23	0.33	592	11
17	-0.12	0.32	596	11
18	-0.02	0.32	599	11
19	0.08	0.31	603	11
20	0.17	0.31	606	11
21	0.27	0.30	609	11
22	0.36	0.30	612	11
23	0.45	0.30	616	10
24	0.53	0.30	619	10
25	0.62	0.29	622	10
26	0.71	0.29	625	10
27	0.79	0.29	628	10
28	0.88	0.29	631	10
29	0.96	0.29	634	10
30	1.04	0.29	637	10
31	1.13	0.29	639	10
32	1.21	0.29	642	10
33	1.30	0.29	645	10
34	1.38	0.29	648	10
35	1.47	0.29	651	10
36	1.55	0.29	654	10
37	1.64	0.30	657	10
38	1.73	0.30	661	10
39	1.82	0.30	664	11
40	1.91	0.30	667	11
41	2.00	0.31	670	11

C.5: Grade 9–Grade 12**Listening and Speaking**

Raw Score	Measure	SE	Scale Score	SE(SS)
42	2.10	0.31	674	11
43	2.20	0.32	677	11
44	2.30	0.32	681	11
45	2.41	0.33	684	12
46	2.52	0.34	688	12
47	2.64	0.35	692	12
48	2.76	0.36	697	12
49	2.89	0.37	701	13
50	3.03	0.38	706	13
51	3.19	0.40	712	14
52	3.35	0.42	717	15
53	3.54	0.45	724	16
54	3.76	0.48	732	17
55	4.02	0.53	741	19
56	4.34	0.61	752	21
57	4.78	0.73	767	26
58	5.51	1.02	793	36
59	6.92	2.01	842	70

Reading and Writing

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-3.75	2.01	469	70
1	-2.34	1.02	518	36
2	-1.61	0.73	544	26
3	-1.17	0.60	559	21
4	-0.86	0.53	570	19
5	-0.60	0.48	579	17
6	-0.39	0.45	586	16
7	-0.20	0.42	593	15
8	-0.03	0.40	599	14
9	0.12	0.38	604	13
10	0.25	0.37	609	13
11	0.38	0.35	613	12
12	0.50	0.34	618	12
13	0.62	0.33	622	12
14	0.73	0.33	625	11
15	0.83	0.32	629	11
16	0.93	0.31	633	11
17	1.03	0.31	636	11
18	1.12	0.31	639	11
19	1.22	0.30	643	11
20	1.31	0.30	646	10
21	1.40	0.30	649	10
22	1.48	0.30	652	10
23	1.57	0.29	655	10
24	1.66	0.29	658	10
25	1.74	0.29	661	10
26	1.83	0.29	664	10

C.5: Grade 9–Grade 12**Reading and Writing**

Raw Score	Measure	SE	Scale Score	SE(SS)
27	1.91	0.29	667	10
28	2.00	0.29	670	10
29	2.08	0.29	673	10
30	2.17	0.30	676	10
31	2.26	0.30	679	10
32	2.35	0.30	682	10
33	2.44	0.30	685	11
34	2.53	0.31	689	11
35	2.62	0.31	692	11
36	2.72	0.31	695	11
37	2.82	0.32	699	11
38	2.93	0.33	702	11
39	3.04	0.33	706	12
40	3.15	0.34	710	12
41	3.27	0.35	715	12
42	3.40	0.37	719	13
43	3.54	0.38	724	13
44	3.69	0.40	729	14
45	3.86	0.42	735	15
46	4.04	0.45	741	16
47	4.26	0.48	749	17
48	4.51	0.53	758	19
49	4.83	0.60	769	21
50	5.27	0.73	784	25
51	5.99	1.02	810	36
52	7.40	2.01	859	70

APPENDIX D: IRT STATISTICS

D.1: Kindergarten–Grade 1

Listening and Speaking

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
LI	1	33362	-3.64	0.03	1.09	1.19	--
	2	33244	-3.29	0.02	1.06	1.11	--
	3	33205	-3.75	0.03	0.99	1.02	--
	4	33388	-3.23	0.02	0.96	0.91	--
	5	33281	-3.02	0.02	0.83	0.58	--
	6	33409	-3.23	0.02	1.05	1.31	--
	7	33410	-2.97	0.02	0.97	0.86	--
	8	33499	-2.57	0.02	1.18	1.80	--
	9	33279	-2.19	0.02	1.21	1.46	--
	10	33451	-2.45	0.02	1.09	1.12	--
	11	33464	-1.65	0.02	1.00	0.97	--
	12	33478	-2.34	0.02	1.22	2.03	√
	13	33369	-1.58	0.02	1.19	1.36	--
	14	33447	-1.41	0.01	1.11	1.15	--
	15	33479	-3.28	0.02	1.10	1.82	--
	16	33407	-3.45	0.03	1.08	1.63	--
	17	33449	-2.23	0.02	1.12	1.34	--
	18	33381	-2.60	0.02	0.99	0.98	--
	19	33395	-2.48	0.02	1.07	1.23	--
	20	33457	-2.54	0.02	1.09	1.19	--
	21	33433	-1.52	0.01	1.10	1.23	--
	22	33324	-1.35	0.01	1.23	1.45	--
	23	33350	-0.91	0.01	1.25	1.40	--
	24	33326	0.15	0.01	1.35	1.72	--
SP	1	33620	-3.52	0.02	1.01	0.76	--
	2	33619	-2.73	0.01	0.92	0.82	--
	3	33619	-1.17	0.01	0.95	0.93	--
	4	33619	-2.32	0.01	0.88	0.84	--
	5	33619	-1.22	0.01	0.89	0.88	--
	6	33619	-2.06	0.01	0.79	0.76	--
	7	33619	-1.51	0.01	0.85	0.82	--
	8	33617	-0.81	0.01	0.80	0.80	--
	9	33617	-1.92	0.01	0.78	0.73	--
	10	33617	-1.81	0.01	0.81	0.76	--
	11	33617	-1.46	0.01	0.77	0.75	--
	12	33617	-1.61	0.01	0.78	0.71	--
	13	33617	-0.99	0.01	0.94	0.92	--
	14	33617	-1.39	0.01	0.83	0.76	--
	15	33616	-1.44	0.01	0.85	0.81	--
	16	33614	-2.30	0.01	0.80	0.65	--

Note: “√” indicate that the item was flagged as misfit; “--” indicates that the item was not flagged.

Reading and Writing

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
RD	1	32663	-2.38	0.01	0.94	0.84	--
	2	32792	-1.31	0.01	0.90	0.83	--
	3	32837	-2.30	0.01	1.03	0.88	--
	4	32742	-2.05	0.01	1.12	1.28	--
	5	32713	-1.91	0.01	1.08	0.94	--
	6	32727	-0.82	0.01	1.01	1.03	--
	7	32573	-1.31	0.01	1.23	1.27	--
	8	32784	-1.49	0.01	0.96	0.85	--
	9	32785	-1.42	0.01	1.12	1.24	--
	10	32656	-1.21	0.01	0.95	0.87	--
	11	32475	-1.77	0.01	1.08	0.95	--
	12	32552	-1.10	0.01	1.01	1.00	--
	13	32389	-0.23	0.01	1.15	1.29	--
	14	32344	-1.60	0.01	1.49	2.25	√
	15	32285	-0.95	0.01	1.12	1.13	--
WR	1	32935	-3.48	0.02	0.88	0.73	--
	2	32969	-4.24	0.02	0.95	0.79	--
	3	32994	-3.44	0.02	0.96	0.80	--
	4	32954	-2.85	0.02	0.92	0.74	--
	5	32964	-3.76	0.02	0.85	0.60	--
	6	32915	-3.67	0.02	0.91	0.90	--
	7	33047	-3.35	0.01	1.19	1.41	--
	8	33047	-3.09	0.01	1.28	1.56	--
	9	33047	-3.29	0.01	1.03	8.60	√
	10	33047	-2.88	0.01	1.07	2.69	√
	11	33046	-1.63	0.01	0.86	0.86	--
	12	33046	-0.89	0.01	0.80	0.80	--
	13	33046	-1.57	0.01	0.63	0.59	--
	14	33045	-0.79	0.01	0.65	0.61	--
	15	33037	-0.40	0.01	0.65	0.60	--

D.2: Grade 2–Grade 4**Listening and Speaking**

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
LI	1	40576	-1.98	0.02	0.92	0.61	--
	2	40556	-1.80	0.02	0.93	0.71	--
	3	40548	-0.54	0.02	1.03	0.87	--
	4	40557	-0.80	0.02	1.12	1.08	--
	5	40562	0.38	0.01	0.91	0.84	--
	6	40584	-1.75	0.02	1.04	0.96	--
	7	40582	-2.24	0.03	1.03	1.07	--
	8	40499	0.80	0.01	1.42	1.71	--
	9	40571	-1.01	0.02	1.07	0.93	--
	10	40558	1.42	0.01	1.24	1.53	--
	11	40517	-0.36	0.02	1.16	1.26	--
	12	40512	-0.26	0.01	1.13	1.25	--
	13	40479	0.97	0.01	1.04	1.09	--
	14	40500	0.60	0.01	1.12	1.24	--
	15	40516	0.16	0.01	1.07	1.10	--
	16	40487	0.77	0.01	1.20	1.31	--
	17	40491	1.13	0.01	1.11	1.22	--
	18	40513	0.93	0.01	1.07	1.12	--
	19	40451	1.74	0.01	1.15	1.42	--
	20	40518	1.17	0.01	1.31	1.58	--
	21	40509	0.01	0.01	1.10	1.10	--
	22	40490	-1.16	0.02	1.14	1.32	--
	23	40488	0.04	0.01	0.84	0.70	--
	24	40404	0.98	0.01	1.03	1.06	--
SP	1	40642	-2.51	0.02	0.90	0.57	--
	2	40642	-2.36	0.02	0.98	0.75	--
	3	40642	-0.61	0.01	0.79	0.71	--
	4	40642	-1.16	0.01	0.74	0.63	--
	5	40642	-0.58	0.01	0.78	0.74	--
	6	40642	-0.59	0.01	0.75	0.67	--
	7	40642	-0.74	0.01	0.82	0.78	--
	8	40642	0.07	0.01	0.80	0.81	--
	9	40642	-1.05	0.01	0.69	0.57	--
	10	40642	-0.58	0.01	0.76	0.69	--
	11	40642	0.18	0.01	0.84	0.83	--
	12	40641	-1.23	0.01	0.91	0.86	--
	13	40641	-0.58	0.01	0.82	0.72	--
	14	40641	-0.60	0.01	0.79	0.72	--
	15	40641	-0.88	0.01	0.78	0.73	--
	16	40637	-0.44	0.01	0.84	0.77	--

Reading and Writing

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
RD	1	38766	-2.48	0.03	0.88	0.50	--
	2	38764	-2.55	0.03	0.89	0.51	--
	3	38754	-1.87	0.02	0.89	0.61	--
	4	38749	-3.07	0.03	0.90	0.56	--
	5	38766	-2.34	0.02	0.92	0.92	--
	6	38747	-1.90	0.02	0.99	1.23	--
	7	38745	-0.73	0.01	1.02	1.16	--
	8	38674	-1.04	0.02	0.84	0.62	--
	9	38630	-0.62	0.01	0.93	0.82	--
	10	38658	0.20	0.01	1.06	1.04	--
	11	38622	0.73	0.01	0.94	0.91	--
	12	38615	-0.10	0.01	1.04	1.07	--
	13	38565	1.41	0.01	1.06	1.15	--
	14	38571	1.90	0.01	1.12	1.36	--
	15	38594	0.87	0.01	0.94	0.93	--
	16	38504	0.62	0.01	1.13	1.14	--
	17	38530	1.41	0.01	1.11	1.21	--
	18	38558	0.06	0.01	0.99	0.92	--
	19	38446	0.29	0.01	1.10	1.15	--
	20	38507	0.81	0.01	1.02	1.04	--
	21	38440	0.95	0.01	0.98	1.01	--
	22	38560	0.74	0.01	0.96	0.93	--
	23	38524	1.12	0.01	0.97	0.98	--
	24	38474	0.68	0.01	0.99	0.97	--
WR	1	38761	-2.34	0.02	0.89	0.53	--
	2	38786	-1.00	0.02	0.79	0.63	--
	3	38774	-1.48	0.02	0.88	0.65	--
	4	38762	-0.47	0.01	0.98	0.93	--
	5	38754	-0.65	0.01	0.86	0.71	--
	6	38718	-0.49	0.01	0.92	0.84	--
	7	38746	1.06	0.01	1.15	1.25	--
	8	38731	1.75	0.01	1.12	1.29	--
	9	38627	-0.68	0.01	0.90	0.76	--
	10	38693	0.41	0.01	1.07	1.13	--
	11	38649	0.32	0.01	0.97	0.96	--
	12	38593	1.22	0.01	1.13	1.24	--
	13	38798	-1.32	0.01	0.99	0.99	--
	14	38798	-0.54	0.01	0.98	0.99	--
	15	38795	-0.36	0.01	0.98	0.97	--
	16	38798	1.05	0.01	0.93	0.93	--

D.3: Grade 5–Grade 6**Listening and Speaking**

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
LI	1	17715	-1.68	0.03	0.91	0.76	--
	2	17710	-1.88	0.04	0.89	0.47	--
	3	17709	-1.08	0.03	1.05	1.35	--
	4	17469	-0.41	0.02	1.12	1.16	--
	5	17705	-0.58	0.03	1.04	1.05	--
	6	17709	-0.48	0.03	0.85	0.70	--
	7	17701	0.02	0.02	0.89	0.77	--
	8	17697	0.52	0.02	1.16	1.21	--
	9	17689	0.91	0.02	1.24	1.44	--
	10	17697	1.58	0.02	1.17	1.37	--
	11	17697	1.12	0.02	1.12	1.23	--
	12	17692	1.35	0.02	1.23	1.47	--
	13	17697	-0.71	0.03	0.86	0.61	--
	14	17695	1.49	0.02	1.33	1.70	--
	15	17701	0.02	0.02	1.05	1.01	--
	16	17695	-0.22	0.02	1.10	1.15	--
	17	17673	0.31	0.02	1.07	1.07	--
	18	17696	0.64	0.02	1.03	1.05	--
	19	17682	0.68	0.02	1.20	1.32	--
	20	17691	0.11	0.02	1.04	0.92	--
	21	17691	-0.70	0.03	1.11	0.99	--
	22	17685	1.09	0.02	1.09	1.20	--
	23	17695	1.93	0.02	1.19	1.48	--
	24	17681	-0.29	0.02	1.05	1.16	--
	25	17671	0.53	0.02	1.39	1.64	--
SP	1	17729	-2.40	0.03	0.99	0.69	--
	2	17729	-2.37	0.03	0.97	0.64	--
	3	17729	-0.27	0.02	0.74	0.66	--
	4	17729	-0.51	0.02	0.88	0.85	--
	5	17729	-0.71	0.02	0.90	0.90	--
	6	17729	0.05	0.02	0.77	0.72	--
	7	17729	-0.17	0.02	0.82	0.84	--
	8	17729	0.45	0.01	0.72	0.73	--
	9	17729	-0.23	0.02	0.68	0.61	--
	10	17729	-0.74	0.02	0.69	0.57	--
	11	17729	-0.28	0.02	0.75	0.70	--
	12	17728	-0.83	0.02	0.73	0.58	--
	13	17728	-0.28	0.02	0.72	0.58	--
	14	17727	-0.71	0.02	0.73	0.61	--
	15	17727	-0.30	0.02	0.84	0.77	--
	16	17727	-0.58	0.02	0.73	0.61	--

Reading and Writing

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
RD	1	17475	-2.44	0.05	0.93	0.63	--
	2	17477	-1.93	0.04	0.97	0.80	--
	3	17472	-0.97	0.03	0.89	0.58	--
	4	17453	0.61	0.02	1.10	1.15	--
	5	17450	0.44	0.02	0.91	0.80	--
	6	17438	0.61	0.02	0.94	0.86	--
	7	17450	1.70	0.02	0.94	0.92	--
	8	17460	-0.48	0.03	0.86	0.66	--
	9	17445	0.31	0.02	0.85	0.74	--
	10	17449	1.69	0.02	1.26	1.36	--
	11	17445	2.37	0.02	1.00	1.01	--
	12	17445	0.53	0.02	1.18	1.33	--
	13	17429	2.71	0.02	1.09	1.27	--
	14	17446	0.67	0.02	0.94	0.88	--
	15	17443	0.94	0.02	0.84	0.75	--
	16	17440	1.56	0.02	0.96	0.97	--
	17	17440	1.65	0.02	1.05	1.07	--
	18	17428	1.42	0.02	0.97	0.96	--
	19	17436	1.44	0.02	0.95	0.90	--
	20	17432	2.37	0.02	1.08	1.21	--
	21	17428	1.60	0.02	0.94	0.92	--
	22	17422	1.36	0.02	0.91	0.85	--
	23	17404	0.65	0.02	0.91	0.80	--
	24	17384	2.21	0.02	1.15	1.25	--
	25	17408	1.26	0.02	1.15	1.20	--
	26	17397	1.61	0.02	1.01	1.00	--
	27	17408	1.64	0.02	1.05	1.07	--
WR	1	17469	-0.47	0.03	0.95	0.78	--
	2	17469	-0.25	0.02	1.01	0.98	--
	3	17472	-0.61	0.03	0.95	0.92	--
	4	17463	0.83	0.02	1.07	1.15	--
	5	17468	0.64	0.02	0.89	0.83	--
	6	17464	1.63	0.02	1.10	1.15	--
	7	17470	2.13	0.02	1.13	1.22	--
	8	17468	-0.33	0.02	0.87	0.69	--
	9	17476	-0.02	0.02	0.95	1.07	--
	10	17473	-0.03	0.02	0.82	0.68	--
	11	17452	0.89	0.02	0.96	0.92	--
	12	17435	0.35	0.02	0.79	0.67	--
	13	17456	1.33	0.02	0.91	0.87	--
	14	17452	1.22	0.02	1.09	1.16	--
	15	17441	2.07	0.02	1.05	1.10	--
	16	17480	-0.29	0.02	1.17	1.29	--
	17	17480	-0.17	0.02	0.97	1.02	--
	18	17480	-0.19	0.02	0.95	0.98	--
	19	17480	1.23	0.01	0.91	0.91	--

D.4: Grade 7–Grade 8

Listening and Speaking

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
LI	1	16592	-0.56	0.03	1.10	1.05	--
	2	16591	-1.20	0.04	1.06	0.66	--
	3	16587	-0.69	0.03	1.09	0.89	--
	4	16537	0.95	0.02	1.02	1.01	--
	5	16581	1.60	0.02	1.28	1.42	--
	6	16578	0.37	0.02	1.12	1.10	--
	7	16578	0.59	0.02	1.00	1.01	--
	8	16585	1.04	0.02	0.86	0.75	--
	9	16584	2.30	0.02	1.46	1.88	--
	10	16573	1.03	0.02	1.03	0.96	--
	11	16572	0.31	0.02	1.03	0.92	--
	12	16554	2.66	0.02	1.20	1.45	--
	13	16581	2.02	0.02	1.00	1.04	--
	14	16563	1.36	0.02	1.21	1.28	--
	15	16586	1.77	0.02	1.21	1.34	--
	16	16584	0.76	0.02	1.09	1.07	--
	17	16575	1.20	0.02	1.35	1.70	--
	18	16551	2.35	0.02	1.23	1.50	--
	19	16557	2.35	0.02	1.23	1.46	--
	20	16549	0.91	0.02	1.09	1.09	--
	21	16551	1.44	0.02	1.26	1.39	--
	22	16567	0.52	0.02	0.96	0.86	--
	23	16559	1.29	0.02	1.23	1.38	--
	24	16559	1.48	0.02	1.07	1.10	--
	25	16553	0.63	0.02	1.07	1.08	--
SP	1	16612	-1.59	0.03	0.91	0.52	--
	2	16612	-0.87	0.02	0.84	0.67	--
	3	16612	0.48	0.02	0.71	0.66	--
	4	16612	0.61	0.02	0.71	0.71	--
	5	16612	0.64	0.02	0.73	0.68	--
	6	16612	0.24	0.02	0.75	0.73	--
	7	16612	0.77	0.02	0.78	0.73	--
	8	16612	1.35	0.01	0.72	0.73	--
	9	16612	0.39	0.02	0.70	0.64	--
	10	16612	0.48	0.02	0.71	0.67	--
	11	16612	0.16	0.02	0.68	0.63	--
	12	16612	0.12	0.02	0.81	0.72	--
	13	16612	0.33	0.02	0.65	0.50	--
	14	16612	-0.02	0.02	0.70	0.54	--
	15	16612	0.09	0.02	0.67	0.58	--
	16	16612	0.74	0.02	0.73	0.65	--

Reading and Writing

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
RD	1	15667	-1.58	0.04	0.96	0.99	--
	2	15666	-0.87	0.03	0.91	0.72	--
	3	15659	-0.83	0.03	0.92	0.72	--
	4	15653	-0.32	0.02	0.93	0.83	--
	5	15646	0.99	0.02	1.02	0.97	--
	6	15635	1.60	0.02	1.04	1.05	--
	7	15635	2.02	0.02	1.05	1.11	--
	8	15648	1.82	0.02	0.96	0.97	--
	9	15644	0.04	0.02	0.92	0.75	--
	10	15643	1.13	0.02	0.76	0.67	--
	11	15626	1.20	0.02	0.95	0.90	--
	12	15623	2.08	0.02	1.21	1.31	--
	13	15612	0.93	0.02	1.01	0.97	--
	14	15635	1.62	0.02	1.03	1.04	--
	15	15644	1.39	0.02	0.94	0.90	--
	16	15635	1.70	0.02	1.01	1.01	--
	17	15633	1.64	0.02	1.14	1.17	--
	18	15640	1.52	0.02	0.95	0.93	--
	19	15637	1.56	0.02	1.09	1.12	--
	20	15620	1.77	0.02	1.00	1.00	--
	21	15621	0.50	0.02	0.86	0.71	--
	22	15627	0.81	0.02	0.87	0.79	--
	23	15625	1.18	0.02	1.10	1.13	--
	24	15607	2.12	0.02	1.17	1.26	--
	25	15617	2.33	0.02	1.09	1.20	--
	26	15603	2.20	0.02	1.17	1.28	--
	27	15608	2.04	0.02	0.99	1.01	--
WR	1	15666	-2.37	0.05	0.96	0.62	--
	2	15664	-2.35	0.05	0.96	0.72	--
	3	15662	-0.60	0.03	0.95	0.97	--
	4	15656	-0.12	0.02	1.13	1.39	--
	5	15663	-0.44	0.03	0.87	0.69	--
	6	15663	0.70	0.02	0.97	0.95	--
	7	15657	1.42	0.02	1.12	1.15	--
	8	15614	1.52	0.02	1.01	1.02	--
	9	15658	0.79	0.02	0.83	0.77	--
	10	15660	0.02	0.02	0.86	0.74	--
	11	15664	1.08	0.02	1.09	1.14	--
	12	15634	0.80	0.02	0.92	0.89	--
	13	15657	0.84	0.02	0.96	0.94	--
	14	15652	0.44	0.02	0.91	0.87	--
	15	15649	1.84	0.02	1.01	1.04	--
	16	15669	0.00	0.02	1.01	1.04	--
	17	15669	0.05	0.02	0.99	1.01	--
	18	15669	0.68	0.01	1.06	1.15	--
	19	15669	1.59	0.01	0.93	0.93	--

D.5: Grade 9–Grade 12

Listening and Speaking

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
LI	1	23892	-0.77	0.03	0.99	0.75	--
	2	23893	-0.95	0.03	1.13	0.75	--
	3	23889	0.10	0.02	1.02	0.82	--
	4	23888	0.12	0.02	1.05	1.20	--
	5	23775	1.69	0.02	1.15	1.18	--
	6	23897	0.32	0.02	1.09	1.30	--
	7	23859	1.99	0.02	1.08	1.12	--
	8	23898	1.93	0.02	1.10	1.17	--
	9	23909	0.94	0.02	0.94	0.82	--
	10	23916	2.18	0.02	1.09	1.12	--
	11	23849	2.42	0.02	1.12	1.23	--
	12	23838	2.83	0.02	1.29	1.65	--
	13	23895	3.18	0.02	1.38	1.94	--
	14	23907	1.63	0.02	1.22	1.37	--
	15	23930	1.69	0.02	1.22	1.33	--
	16	23918	0.49	0.02	1.08	1.13	--
	17	23918	0.48	0.02	1.03	0.98	--
	18	23898	0.98	0.02	1.02	1.04	--
	19	23889	1.34	0.02	1.33	1.62	--
	20	23913	0.45	0.02	0.92	0.97	--
	21	23879	1.67	0.02	1.26	1.57	--
	22	23887	2.10	0.02	1.27	1.49	--
	23	23886	1.51	0.02	0.99	0.95	--
	24	23865	2.20	0.02	1.17	1.25	--
	25	23868	2.02	0.02	1.33	1.55	--
SP	1	23982	-0.51	0.02	0.92	0.73	--
	2	23982	-0.25	0.02	0.94	0.87	--
	3	23982	0.82	0.01	0.76	0.70	--
	4	23982	0.97	0.01	0.76	0.72	--
	5	23982	1.17	0.01	0.69	0.65	--
	6	23982	1.45	0.01	0.72	0.69	--
	7	23982	0.55	0.01	0.82	0.77	--
	8	23982	1.33	0.01	0.67	0.67	--
	9	23982	0.62	0.01	0.72	0.67	--
	10	23982	0.58	0.01	0.73	0.68	--
	11	23982	0.84	0.01	0.74	0.69	--
	12	23982	0.21	0.01	0.75	0.61	--
	13	23982	0.56	0.01	0.82	0.73	--
	14	23982	1.00	0.01	0.73	0.68	--
	15	23982	0.87	0.01	0.68	0.59	--
	16	23980	0.89	0.01	0.71	0.57	--

Reading and Writing

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
RD	1	22377	0.30	0.02	0.97	0.82	--
	2	22353	1.39	0.02	1.04	1.08	--
	3	22341	1.52	0.02	1.19	1.32	--
	4	22393	0.98	0.02	1.07	1.19	--
	5	22365	0.68	0.02	0.94	0.85	--
	6	22326	2.25	0.02	1.01	1.03	--
	7	22382	2.44	0.02	1.02	1.01	--
	8	22357	1.76	0.02	0.93	0.86	--
	9	22273	2.86	0.01	0.98	0.98	--
	10	22340	1.51	0.02	1.11	1.16	--
	11	22353	3.01	0.01	1.17	1.23	--
	12	22361	2.49	0.01	1.09	1.12	--
	13	22335	2.25	0.02	0.86	0.81	--
	14	22340	0.45	0.02	0.93	0.71	--
	15	22349	1.68	0.02	0.96	0.93	--
	16	22334	1.92	0.02	0.78	0.68	--
	17	22313	2.77	0.01	1.05	1.07	--
	18	22248	3.02	0.01	0.96	0.96	--
	19	22304	1.41	0.02	1.11	1.19	--
	20	22246	2.06	0.02	1.02	1.02	--
	21	22278	1.94	0.02	0.93	0.89	--
	22	22252	1.64	0.02	0.95	0.92	--
	23	22222	1.36	0.02	0.96	0.91	--
	24	22205	1.91	0.02	1.10	1.17	--
	25	22161	2.06	0.02	1.14	1.20	--
	26	22191	1.39	0.02	1.09	1.20	--
	27	22181	2.78	0.01	1.09	1.13	--
WR	1	22362	0.15	0.02	0.92	0.64	--
	2	22333	1.57	0.02	1.02	1.03	--
	3	22334	1.74	0.02	0.99	0.99	--
	4	22401	1.49	0.02	0.99	1.00	--
	5	22400	0.99	0.02	1.01	0.99	--
	6	22361	2.38	0.02	0.93	0.91	--
	7	22402	2.04	0.02	0.99	0.98	--
	8	22385	1.35	0.02	1.02	1.09	--
	9	22365	1.73	0.02	0.95	0.90	--
	10	22384	2.19	0.02	0.99	0.97	--
	11	22392	1.67	0.02	0.92	0.84	--
	12	22369	2.08	0.02	0.97	0.95	--
	13	22367	2.73	0.01	1.19	1.26	--
	14	22366	2.03	0.02	0.86	0.81	--
	15	22357	2.16	0.02	1.00	1.00	--
	16	22418	1.07	0.01	0.97	0.94	--
	17	22418	1.75	0.01	0.97	0.96	--
	18	22417	2.03	0.01	0.96	0.96	--
	19	22418	2.29	0.01	0.90	0.90	--

APPENDIX E: ITEM INFORMATION AT DIFFERENT CUT POINTS BY GRADES

E.1: Kindergarten–Grade 1

Grade K

Listening and Speaking

	Item #	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-3.64	0.19	0.05	0.02
	2	-3.29	0.22	0.07	0.02
	3	-3.75	0.18	0.05	0.01
	4	-3.23	0.22	0.07	0.02
	5	-3.02	0.24	0.09	0.03
	6	-3.23	0.22	0.07	0.02
	7	-2.97	0.24	0.09	0.03
	8	-2.57	0.25	0.12	0.04
	9	-2.19	0.24	0.16	0.06
	10	-2.45	0.25	0.13	0.05
	11	-1.65	0.21	0.21	0.10
	12	-2.34	0.25	0.14	0.05
	13	-1.58	0.20	0.21	0.10
	14	-1.41	0.18	0.23	0.11
	15	-3.28	0.22	0.07	0.02
	16	-3.45	0.21	0.06	0.02
	17	-2.23	0.24	0.15	0.06
	18	-2.60	0.25	0.12	0.04
	19	-2.48	0.25	0.13	0.05
	20	-2.54	0.25	0.12	0.04
	21	-1.52	0.19	0.22	0.11
	22	-1.35	0.18	0.23	0.12
	23	-0.91	0.14	0.25	0.16
	24	0.15	0.06	0.20	0.24
SP	1	-3.52	0.20	0.06	0.02
	2	-2.73	0.25	0.11	0.04
	3	-1.17	0.16	0.24	0.14
	4	-2.32	0.25	0.14	0.05
	5	-1.22	0.17	0.24	0.13
	6	-2.06	0.24	0.17	0.07
	7	-1.51	0.19	0.22	0.11
	8	-0.81	0.13	0.25	0.17
	9	-1.92	0.23	0.18	0.08
	10	-1.81	0.22	0.19	0.08
	11	-1.46	0.19	0.22	0.11
	12	-1.61	0.20	0.21	0.10
	13	-0.99	0.14	0.25	0.15
	14	-1.39	0.18	0.23	0.12

Grade K
Listening and Speaking

Item #	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
15	-1.44	0.19	0.22	0.11
16	-2.30	0.25	0.15	0.06

Note. The statistics in Tables E.1–E.5 were based on the equating sample (total N=161,358). Invalid cases were removed after cleaning procedures.

Grade K

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	-2.38	0.23	0.12	0.06
	2	-1.31	0.24	0.21	0.14
	3	-2.30	0.23	0.12	0.07
	4	-2.05	0.25	0.15	0.08
	5	-1.91	0.25	0.16	0.09
	6	-0.82	0.20	0.24	0.19
	7	-1.31	0.24	0.21	0.14
	8	-1.49	0.24	0.20	0.13
	9	-1.42	0.24	0.21	0.13
	10	-1.21	0.23	0.22	0.15
	11	-1.77	0.25	0.17	0.10
	12	-1.10	0.22	0.23	0.16
	13	-0.23	0.14	0.24	0.24
	14	-1.60	0.25	0.19	0.12
	15	-0.95	0.21	0.24	0.18
WR	1	-3.48	0.13	0.05	0.02
	2	-4.24	0.07	0.02	0.01
	3	-3.44	0.14	0.05	0.02
	4	-2.85	0.19	0.08	0.04
	5	-3.76	0.11	0.04	0.02
	6	-3.67	0.11	0.04	0.02
	7	-3.35	0.14	0.05	0.03
	8	-3.09	0.17	0.07	0.03
	9	-3.29	0.15	0.06	0.03
	10	-2.88	0.19	0.08	0.04
	11	-1.63	0.25	0.19	0.11
	12	-0.89	0.21	0.24	0.18
	13	-1.57	0.25	0.19	0.12
	14	-0.79	0.20	0.25	0.19
	15	-0.40	0.16	0.25	0.23

Grade 1**Listening and Speaking**

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-3.64	0.16	0.04	0.01
	2	-3.29	0.20	0.06	0.01
	3	-3.75	0.15	0.04	0.01
	4	-3.23	0.20	0.06	0.01
	5	-3.02	0.22	0.07	0.01
	6	-3.23	0.20	0.06	0.01
	7	-2.97	0.22	0.07	0.01
	8	-2.57	0.24	0.10	0.02
	9	-2.19	0.25	0.13	0.03
	10	-2.45	0.25	0.11	0.02
	11	-1.65	0.23	0.18	0.04
	12	-2.34	0.25	0.12	0.02
	13	-1.58	0.22	0.19	0.05
	14	-1.41	0.21	0.21	0.05
	15	-3.28	0.20	0.06	0.01
	16	-3.45	0.18	0.05	0.01
	17	-2.23	0.25	0.13	0.03
	18	-2.60	0.24	0.10	0.02
	19	-2.48	0.25	0.11	0.02
	20	-2.54	0.25	0.10	0.02
	21	-1.52	0.22	0.20	0.05
	22	-1.35	0.20	0.21	0.06
	23	-0.91	0.16	0.24	0.08
	24	0.15	0.07	0.22	0.17
SP	1	-3.52	0.17	0.05	0.01
	2	-2.73	0.24	0.09	0.02
	3	-1.17	0.19	0.23	0.07
	4	-2.32	0.25	0.12	0.02
	5	-1.22	0.19	0.22	0.06
	6	-2.06	0.25	0.15	0.03
	7	-1.51	0.22	0.20	0.05
	8	-0.81	0.15	0.24	0.09
	9	-1.92	0.24	0.16	0.03
	10	-1.81	0.24	0.17	0.04
	11	-1.46	0.21	0.20	0.05
	12	-1.61	0.22	0.19	0.05
	13	-0.99	0.17	0.24	0.08
	14	-1.39	0.21	0.21	0.05
	15	-1.44	0.21	0.20	0.05
	16	-2.30	0.25	0.12	0.02

Grade 1**Reading and Writing**

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	-2.38	0.18	0.06	0.02
	2	-1.31	0.25	0.14	0.05
	3	-2.30	0.18	0.07	0.02
	4	-2.05	0.21	0.08	0.03
	5	-1.91	0.22	0.09	0.03
	6	-0.82	0.24	0.19	0.08
	7	-1.31	0.25	0.14	0.05
	8	-1.49	0.24	0.13	0.04
	9	-1.42	0.25	0.13	0.04
	10	-1.21	0.25	0.15	0.05
	11	-1.77	0.23	0.10	0.03
	12	-1.10	0.25	0.16	0.06
	13	-0.23	0.20	0.24	0.12
	14	-1.60	0.24	0.12	0.04
	15	-0.95	0.25	0.18	0.07
WR	1	-3.48	0.08	0.02	0.01
	2	-4.24	0.04	0.01	0.00
	3	-3.44	0.08	0.02	0.01
	4	-2.85	0.13	0.04	0.01
	5	-3.76	0.07	0.02	0.00
	6	-3.67	0.07	0.02	0.01
	7	-3.35	0.09	0.03	0.01
	8	-3.09	0.11	0.03	0.01
	9	-3.29	0.10	0.03	0.01
	10	-2.88	0.13	0.04	0.01
	11	-1.63	0.24	0.11	0.04
	12	-0.89	0.25	0.18	0.07
	13	-1.57	0.24	0.12	0.04
	14	-0.79	0.24	0.19	0.08
	15	-0.40	0.22	0.23	0.11

E.2: Grade 2–Grade 4**Grade 2****Listening and Speaking**

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-1.98	0.24	0.10	0.02
	2	-1.80	0.25	0.12	0.03
	3	-0.54	0.19	0.23	0.08
	4	-0.80	0.21	0.21	0.07
	5	0.38	0.10	0.24	0.16
	6	-1.75	0.25	0.12	0.03
	7	-2.24	0.23	0.08	0.02
	8	0.80	0.07	0.22	0.20
	9	-1.01	0.23	0.19	0.06
	10	1.42	0.04	0.16	0.24
	11	-0.36	0.17	0.24	0.10
	12	-0.26	0.16	0.24	0.11
	13	0.97	0.06	0.20	0.22
	14	0.60	0.09	0.23	0.18
	15	0.16	0.12	0.25	0.14
	16	0.77	0.08	0.22	0.20
	17	1.13	0.06	0.19	0.23
	18	0.93	0.07	0.21	0.21
	19	1.74	0.03	0.13	0.25
	20	1.17	0.05	0.19	0.23
	21	0.01	0.14	0.25	0.13
	22	-1.16	0.24	0.18	0.05
	23	0.04	0.13	0.25	0.13
	24	0.98	0.06	0.20	0.22
SP	1	-2.51	0.21	0.07	0.01
	2	-2.36	0.22	0.07	0.02
	3	-0.61	0.19	0.22	0.08
	4	-1.16	0.24	0.18	0.05
	5	-0.58	0.19	0.23	0.08
	6	-0.59	0.19	0.23	0.08
	7	-0.74	0.21	0.21	0.07
	8	0.07	0.13	0.25	0.13
	9	-1.05	0.23	0.19	0.06
	10	-0.58	0.19	0.23	0.08
	11	0.18	0.12	0.25	0.14
	12	-1.23	0.24	0.17	0.05
	13	-0.58	0.19	0.23	0.08
	14	-0.60	0.19	0.22	0.08
	15	-0.88	0.22	0.20	0.06
	16	-0.44	0.18	0.24	0.09

Grade 2

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	-2.48	0.13	0.04	0.01
	2	-2.55	0.13	0.04	0.01
	3	-1.87	0.19	0.07	0.02
	4	-3.07	0.09	0.02	0.01
	5	-2.34	0.15	0.05	0.02
	6	-1.90	0.19	0.07	0.02
	7	-0.73	0.25	0.16	0.07
	8	-1.04	0.25	0.13	0.05
	9	-0.62	0.25	0.17	0.07
	10	0.20	0.19	0.24	0.14
	11	0.73	0.14	0.25	0.19
	12	-0.10	0.22	0.22	0.11
	13	1.41	0.09	0.22	0.24
	14	1.90	0.06	0.17	0.25
	15	0.87	0.13	0.25	0.20
	16	0.62	0.16	0.25	0.18
	17	1.41	0.09	0.22	0.24
	18	0.06	0.21	0.23	0.13
	19	0.29	0.19	0.24	0.15
	20	0.81	0.14	0.25	0.20
	21	0.95	0.12	0.24	0.21
	22	0.74	0.14	0.25	0.19
	23	1.12	0.11	0.24	0.22
	24	0.68	0.15	0.25	0.18
WR	1	-2.34	0.15	0.05	0.02
	2	-1.00	0.25	0.14	0.05
	3	-1.48	0.22	0.10	0.03
	4	-0.47	0.24	0.19	0.08
	5	-0.65	0.25	0.17	0.07
	6	-0.49	0.24	0.18	0.08
	7	1.06	0.11	0.24	0.22
	8	1.75	0.07	0.19	0.25
	9	-0.68	0.25	0.17	0.07
	10	0.41	0.17	0.25	0.16
	11	0.32	0.18	0.24	0.15
	12	1.22	0.10	0.23	0.23
	13	-1.32	0.24	0.11	0.04
	14	-0.54	0.25	0.18	0.08
	15	-0.36	0.24	0.20	0.09
	16	1.05	0.12	0.24	0.22

Grade 3**Listening and Speaking**

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-1.98	0.22	0.08	0.02
	2	-1.80	0.23	0.10	0.02
	3	-0.54	0.22	0.21	0.06
	4	-0.80	0.24	0.19	0.05
	5	0.38	0.13	0.25	0.13
	6	-1.75	0.24	0.10	0.02
	7	-2.24	0.20	0.07	0.01
	8	0.80	0.10	0.24	0.16
	9	-1.01	0.25	0.17	0.04
	10	1.42	0.06	0.19	0.22
	11	-0.36	0.20	0.22	0.07
	12	-0.26	0.20	0.23	0.08
	13	0.97	0.09	0.22	0.18
	14	0.60	0.12	0.24	0.15
	15	0.16	0.16	0.25	0.11
	16	0.77	0.10	0.24	0.16
	17	1.13	0.08	0.21	0.20
	18	0.93	0.09	0.23	0.18
	19	1.74	0.04	0.16	0.24
	20	1.17	0.07	0.21	0.20
	21	0.01	0.17	0.24	0.09
	22	-1.16	0.25	0.15	0.03
	23	0.04	0.17	0.25	0.10
	24	0.98	0.09	0.22	0.18
SP	1	-2.51	0.17	0.05	0.01
	2	-2.36	0.19	0.06	0.01
	3	-0.61	0.22	0.20	0.06
	4	-1.16	0.25	0.15	0.03
	5	-0.58	0.22	0.21	0.06
	6	-0.59	0.22	0.21	0.06
	7	-0.74	0.23	0.19	0.05
	8	0.07	0.16	0.25	0.10
	9	-1.05	0.25	0.16	0.04
	10	-0.58	0.22	0.21	0.06
	11	0.18	0.15	0.25	0.11
	12	-1.23	0.25	0.15	0.03
	13	-0.58	0.22	0.21	0.06
	14	-0.60	0.22	0.20	0.06
	15	-0.88	0.24	0.18	0.04
	16	-0.44	0.21	0.22	0.07

Grade 3

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	-2.48	0.11	0.03	0.01
	2	-2.55	0.11	0.03	0.01
	3	-1.87	0.17	0.06	0.02
	4	-3.07	0.07	0.02	0.00
	5	-2.34	0.12	0.04	0.01
	6	-1.90	0.17	0.06	0.02
	7	-0.73	0.25	0.14	0.05
	8	-1.04	0.24	0.11	0.03
	9	-0.62	0.25	0.15	0.05
	10	0.20	0.22	0.22	0.10
	11	0.73	0.17	0.25	0.15
	12	-0.10	0.24	0.20	0.08
	13	1.41	0.11	0.23	0.21
	14	1.90	0.07	0.19	0.24
	15	0.87	0.16	0.25	0.16
	16	0.62	0.18	0.25	0.14
	17	1.41	0.11	0.23	0.21
	18	0.06	0.23	0.21	0.09
	19	0.29	0.21	0.23	0.11
	20	0.81	0.16	0.25	0.16
	21	0.95	0.15	0.25	0.17
	22	0.74	0.17	0.25	0.15
	23	1.12	0.13	0.25	0.18
	24	0.68	0.17	0.25	0.14
WR	1	-2.34	0.12	0.04	0.01
	2	-1.00	0.24	0.12	0.04
	3	-1.48	0.20	0.08	0.02
	4	-0.47	0.25	0.17	0.06
	5	-0.65	0.25	0.15	0.05
	6	-0.49	0.25	0.16	0.06
	7	1.06	0.14	0.25	0.18
	8	1.75	0.08	0.21	0.23
	9	-0.68	0.25	0.15	0.05
	10	0.41	0.20	0.24	0.12
	11	0.32	0.21	0.23	0.11
	12	1.22	0.12	0.24	0.19
	13	-1.32	0.22	0.09	0.03
	14	-0.54	0.25	0.16	0.05
	15	-0.36	0.25	0.18	0.06
	16	1.05	0.14	0.25	0.18

Grade 4

Listening and Speaking

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-1.98	0.19	0.07	0.01
	2	-1.80	0.21	0.08	0.01
	3	-0.54	0.24	0.18	0.05
	4	-0.80	0.25	0.16	0.04
	5	0.38	0.16	0.25	0.10
	6	-1.75	0.21	0.08	0.02
	7	-2.24	0.17	0.05	0.01
	8	0.80	0.13	0.25	0.14
	9	-1.01	0.25	0.14	0.03
	10	1.42	0.08	0.21	0.20
	11	-0.36	0.23	0.20	0.06
	12	-0.26	0.22	0.21	0.06
	13	0.97	0.11	0.24	0.16
	14	0.60	0.14	0.25	0.12
	15	0.16	0.19	0.24	0.09
	16	0.77	0.13	0.25	0.14
	17	1.13	0.10	0.23	0.17
	18	0.93	0.11	0.24	0.15
	19	1.74	0.06	0.18	0.23
	20	1.17	0.10	0.23	0.18
	21	0.01	0.20	0.23	0.08
	22	-1.16	0.25	0.13	0.03
	23	0.04	0.20	0.23	0.08
	24	0.98	0.11	0.24	0.16
SP	1	-2.51	0.14	0.04	0.01
	2	-2.36	0.16	0.05	0.01
	3	-0.61	0.24	0.18	0.05
	4	-1.16	0.25	0.13	0.03
	5	-0.58	0.24	0.18	0.05
	6	-0.59	0.24	0.18	0.05
	7	-0.74	0.25	0.17	0.04
	8	0.07	0.19	0.23	0.08
	9	-1.05	0.25	0.14	0.03
	10	-0.58	0.24	0.18	0.05
	11	0.18	0.18	0.24	0.09
	12	-1.23	0.25	0.12	0.03
	13	-0.58	0.24	0.18	0.05
	14	-0.60	0.24	0.18	0.05
	15	-0.88	0.25	0.15	0.04
	16	-0.44	0.23	0.19	0.05

Grade 4

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	-2.48	0.09	0.02	0.01
	2	-2.55	0.09	0.02	0.01
	3	-1.87	0.15	0.04	0.01
	4	-3.07	0.06	0.01	0.00
	5	-2.34	0.10	0.03	0.01
	6	-1.90	0.14	0.04	0.01
	7	-0.73	0.24	0.11	0.03
	8	-1.04	0.22	0.09	0.02
	9	-0.62	0.24	0.12	0.04
	10	0.20	0.23	0.20	0.08
	11	0.73	0.19	0.24	0.11
	12	-0.10	0.25	0.17	0.06
	13	1.41	0.13	0.25	0.18
	14	1.90	0.09	0.22	0.22
	15	0.87	0.18	0.24	0.13
	16	0.62	0.20	0.23	0.11
	17	1.41	0.13	0.25	0.18
	18	0.06	0.24	0.18	0.07
	19	0.29	0.23	0.21	0.08
	20	0.81	0.18	0.24	0.12
	21	0.95	0.17	0.25	0.13
	22	0.74	0.19	0.24	0.12
	23	1.12	0.15	0.25	0.15
	24	0.68	0.20	0.23	0.11
WR	1	-2.34	0.10	0.03	0.01
	2	-1.00	0.22	0.09	0.03
	3	-1.48	0.18	0.06	0.02
	4	-0.47	0.25	0.13	0.04
	5	-0.65	0.24	0.12	0.04
	6	-0.49	0.25	0.13	0.04
	7	1.06	0.16	0.25	0.14
	8	1.75	0.10	0.23	0.21
	9	-0.68	0.24	0.12	0.03
	10	0.41	0.22	0.22	0.09
	11	0.32	0.23	0.21	0.08
	12	1.22	0.14	0.25	0.16
	13	-1.32	0.20	0.07	0.02
	14	-0.54	0.25	0.13	0.04
	15	-0.36	0.25	0.14	0.05
	16	1.05	0.16	0.25	0.14

E.3: Grade 5–Grade 6**Grade 5****Listening and Speaking**

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-1.68	0.19	0.07	0.02
	2	-1.88	0.17	0.06	0.01
	3	-1.08	0.24	0.11	0.03
	4	-0.41	0.25	0.17	0.05
	5	-0.58	0.25	0.15	0.04
	6	-0.48	0.25	0.16	0.05
	7	0.02	0.23	0.21	0.07
	8	0.52	0.19	0.24	0.11
	9	0.91	0.15	0.25	0.14
	10	1.58	0.09	0.22	0.21
	11	1.12	0.13	0.25	0.16
	12	1.35	0.11	0.24	0.19
	13	-0.71	0.25	0.14	0.04
	14	1.49	0.10	0.23	0.20
	15	0.02	0.23	0.21	0.07
	16	-0.22	0.24	0.19	0.06
	17	0.31	0.21	0.23	0.09
	18	0.64	0.17	0.25	0.12
	19	0.68	0.17	0.25	0.12
	20	0.11	0.22	0.22	0.08
	21	-0.70	0.25	0.14	0.04
	22	1.09	0.13	0.25	0.16
	23	1.93	0.07	0.19	0.23
	24	-0.29	0.24	0.18	0.06
	25	0.53	0.19	0.24	0.11
SP	1	-2.40	0.12	0.03	0.01
	2	-2.37	0.12	0.04	0.01
	3	-0.27	0.24	0.18	0.06
	4	-0.51	0.25	0.16	0.05
	5	-0.71	0.25	0.14	0.04
	6	0.05	0.23	0.21	0.08
	7	-0.17	0.24	0.19	0.06
	8	0.45	0.19	0.24	0.10
	9	-0.23	0.24	0.19	0.06
	10	-0.74	0.25	0.14	0.04
	11	-0.28	0.24	0.18	0.06
	12	-0.83	0.25	0.13	0.03
	13	-0.28	0.24	0.18	0.06
	14	-0.71	0.25	0.14	0.04
	15	-0.30	0.24	0.18	0.06
	16	-0.58	0.25	0.15	0.04

Grade 5

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	-2.44	0.05	0.02	0.00
	2	-1.93	0.07	0.02	0.01
	3	-0.97	0.15	0.06	0.02
	4	0.61	0.25	0.19	0.08
	5	0.44	0.25	0.17	0.07
	6	0.61	0.25	0.19	0.08
	7	1.70	0.18	0.25	0.17
	8	-0.48	0.19	0.09	0.03
	9	0.31	0.25	0.16	0.06
	10	1.69	0.18	0.25	0.17
	11	2.37	0.12	0.23	0.23
	12	0.53	0.25	0.18	0.07
	13	2.71	0.09	0.20	0.25
	14	0.67	0.25	0.19	0.08
	15	0.94	0.24	0.21	0.10
	16	1.56	0.20	0.25	0.16
	17	1.65	0.19	0.25	0.17
	18	1.42	0.21	0.24	0.14
	19	1.44	0.21	0.24	0.15
	20	2.37	0.12	0.23	0.23
	21	1.60	0.19	0.25	0.16
	22	1.36	0.21	0.24	0.14
	23	0.65	0.25	0.19	0.08
	24	2.21	0.13	0.24	0.22
	25	1.26	0.22	0.24	0.13
	26	1.61	0.19	0.25	0.16
	27	1.64	0.19	0.25	0.16
WR	1	-0.47	0.19	0.09	0.03
	2	-0.25	0.21	0.11	0.04
	3	-0.61	0.18	0.08	0.03
	4	0.83	0.25	0.21	0.09
	5	0.64	0.25	0.19	0.08
	6	1.63	0.19	0.25	0.16
	7	2.13	0.14	0.24	0.21
	8	-0.33	0.21	0.10	0.03
	9	-0.02	0.23	0.13	0.05
	10	-0.03	0.23	0.12	0.04
	11	0.89	0.24	0.21	0.10
	12	0.35	0.25	0.16	0.06
	13	1.33	0.22	0.24	0.14
	14	1.22	0.22	0.23	0.13
	15	2.07	0.15	0.24	0.20
	16	-0.29	0.21	0.10	0.04
	17	-0.17	0.22	0.11	0.04
	18	-0.19	0.22	0.11	0.04
	19	1.23	0.22	0.23	0.13

Grade 6

Listening and Speaking

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-1.68	0.17	0.05	0.01
	2	-1.88	0.15	0.05	0.01
	3	-1.08	0.22	0.09	0.02
	4	-0.41	0.25	0.15	0.04
	5	-0.58	0.25	0.13	0.03
	6	-0.48	0.25	0.14	0.04
	7	0.02	0.24	0.19	0.06
	8	0.52	0.21	0.23	0.09
	9	0.91	0.17	0.25	0.12
	10	1.58	0.11	0.24	0.18
	11	1.12	0.15	0.25	0.14
	12	1.35	0.13	0.25	0.16
	13	-0.71	0.24	0.12	0.03
	14	1.49	0.12	0.24	0.17
	15	0.02	0.24	0.19	0.06
	16	-0.22	0.25	0.17	0.05
	17	0.31	0.22	0.21	0.07
	18	0.64	0.20	0.24	0.10
	19	0.68	0.19	0.24	0.10
	20	0.11	0.24	0.20	0.06
	21	-0.70	0.24	0.12	0.03
	22	1.09	0.15	0.25	0.14
	23	1.93	0.08	0.21	0.21
	24	-0.29	0.25	0.16	0.04
	25	0.53	0.21	0.23	0.09
SP	1	-2.40	0.10	0.03	0.01
	2	-2.37	0.11	0.03	0.01
	3	-0.27	0.25	0.16	0.04
	4	-0.51	0.25	0.14	0.04
	5	-0.71	0.24	0.12	0.03
	6	0.05	0.24	0.19	0.06
	7	-0.17	0.25	0.17	0.05
	8	0.45	0.21	0.22	0.08
	9	-0.23	0.25	0.16	0.05
	10	-0.74	0.24	0.12	0.03
	11	-0.28	0.25	0.16	0.04
	12	-0.83	0.24	0.11	0.03
	13	-0.28	0.25	0.16	0.04
	14	-0.71	0.24	0.12	0.03
	15	-0.30	0.25	0.16	0.04
	16	-0.58	0.25	0.13	0.03

Grade 6

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	-2.44	0.04	0.01	0.00
	2	-1.93	0.07	0.02	0.01
	3	-0.97	0.14	0.05	0.02
	4	0.61	0.25	0.17	0.07
	5	0.44	0.25	0.15	0.06
	6	0.61	0.25	0.17	0.07
	7	1.70	0.19	0.25	0.15
	8	-0.48	0.18	0.08	0.03
	9	0.31	0.24	0.14	0.05
	10	1.69	0.19	0.25	0.15
	11	2.37	0.13	0.24	0.22
	12	0.53	0.25	0.16	0.06
	13	2.71	0.10	0.22	0.24
	14	0.67	0.25	0.17	0.07
	15	0.94	0.25	0.20	0.09
	16	1.56	0.21	0.24	0.14
	17	1.65	0.20	0.24	0.15
	18	1.42	0.22	0.23	0.13
	19	1.44	0.22	0.23	0.13
	20	2.37	0.13	0.24	0.22
	21	1.60	0.20	0.24	0.14
	22	1.36	0.22	0.23	0.12
	23	0.65	0.25	0.17	0.07
	24	2.21	0.14	0.25	0.20
	25	1.26	0.23	0.22	0.11
	26	1.61	0.20	0.24	0.15
	27	1.64	0.20	0.24	0.15
WR	1	-0.47	0.18	0.08	0.03
	2	-0.25	0.20	0.09	0.03
	3	-0.61	0.17	0.07	0.02
	4	0.83	0.25	0.19	0.08
	5	0.64	0.25	0.17	0.07
	6	1.63	0.20	0.24	0.15
	7	2.13	0.15	0.25	0.19
	8	-0.33	0.20	0.08	0.03
	9	-0.02	0.22	0.11	0.04
	10	-0.03	0.22	0.11	0.04
	11	0.89	0.25	0.19	0.09
	12	0.35	0.24	0.14	0.05
	13	1.33	0.22	0.23	0.12
	14	1.22	0.23	0.22	0.11
	15	2.07	0.16	0.25	0.19
	16	-0.29	0.20	0.09	0.03
	17	-0.17	0.21	0.10	0.03
	18	-0.19	0.21	0.10	0.03
	19	1.23	0.23	0.22	0.11

E.4: Grade 7–Grade 8**Grade 7****Listening and Speaking**

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-0.56	0.24	0.10	0.03
	2	-1.20	0.19	0.06	0.02
	3	-0.69	0.23	0.09	0.02
	4	0.95	0.19	0.23	0.10
	5	1.60	0.13	0.25	0.16
	6	0.37	0.24	0.19	0.07
	7	0.59	0.22	0.21	0.08
	8	1.04	0.19	0.24	0.11
	9	2.30	0.08	0.21	0.23
	10	1.03	0.19	0.24	0.11
	11	0.31	0.24	0.18	0.06
	12	2.66	0.06	0.18	0.24
	13	2.02	0.10	0.23	0.20
	14	1.36	0.15	0.25	0.14
	15	1.77	0.12	0.24	0.18
	16	0.76	0.21	0.22	0.09
	17	1.20	0.17	0.25	0.13
	18	2.35	0.07	0.21	0.23
	19	2.35	0.07	0.21	0.23
	20	0.91	0.20	0.23	0.10
	21	1.44	0.15	0.25	0.15
	22	0.52	0.23	0.20	0.07
	23	1.29	0.16	0.25	0.13
	24	1.48	0.14	0.25	0.15
	25	0.63	0.22	0.21	0.08
SP	1	-1.59	0.15	0.04	0.01
	2	-0.87	0.21	0.08	0.02
	3	0.48	0.23	0.20	0.07
	4	0.61	0.22	0.21	0.08
	5	0.64	0.22	0.21	0.08
	6	0.24	0.24	0.18	0.06
	7	0.77	0.21	0.22	0.09
	8	1.35	0.16	0.25	0.14
	9	0.39	0.24	0.19	0.07
	10	0.48	0.23	0.20	0.07
	11	0.16	0.25	0.17	0.05
	12	0.12	0.25	0.17	0.05
	13	0.33	0.24	0.19	0.06
	14	-0.02	0.25	0.15	0.05
	15	0.09	0.25	0.16	0.05
	16	0.74	0.21	0.22	0.09

Grade 7

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	-1.58	0.08	0.02	0.01
	2	-0.87	0.13	0.05	0.02
	3	-0.83	0.13	0.05	0.02
	4	-0.32	0.18	0.07	0.03
	5	0.99	0.25	0.18	0.09
	6	1.60	0.22	0.23	0.14
	7	2.02	0.18	0.25	0.18
	8	1.82	0.20	0.24	0.17
	9	0.04	0.21	0.10	0.04
	10	1.13	0.24	0.20	0.10
	11	1.20	0.24	0.20	0.11
	12	2.08	0.17	0.25	0.19
	13	0.93	0.25	0.18	0.09
	14	1.62	0.21	0.23	0.15
	15	1.39	0.23	0.22	0.13
	16	1.70	0.21	0.24	0.15
	17	1.64	0.21	0.24	0.15
	18	1.52	0.22	0.23	0.14
	19	1.56	0.22	0.23	0.14
	20	1.77	0.20	0.24	0.16
	21	0.50	0.24	0.14	0.06
	22	0.81	0.25	0.17	0.08
	23	1.18	0.24	0.20	0.11
	24	2.12	0.17	0.25	0.19
	25	2.33	0.15	0.25	0.21
	26	2.20	0.16	0.25	0.20
	27	2.04	0.18	0.25	0.19
WR	1	-2.37	0.04	0.01	0.00
	2	-2.35	0.04	0.01	0.00
	3	-0.60	0.16	0.06	0.02
	4	-0.12	0.20	0.09	0.04
	5	-0.44	0.17	0.07	0.03
	6	0.70	0.25	0.16	0.07
	7	1.42	0.23	0.22	0.13
	8	1.52	0.22	0.23	0.14
	9	0.79	0.25	0.16	0.08
	10	0.02	0.21	0.10	0.04
	11	1.08	0.25	0.19	0.10
	12	0.80	0.25	0.17	0.08
	13	0.84	0.25	0.17	0.08
	14	0.44	0.24	0.13	0.06
	15	1.84	0.20	0.24	0.17
	16	0.00	0.21	0.10	0.04
	17	0.05	0.22	0.10	0.04
	18	0.68	0.25	0.15	0.07
	19	1.59	0.22	0.23	0.14

Grade 8

Listening and Speaking

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-0.56	0.22	0.08	0.03
	2	-1.20	0.16	0.05	0.02
	3	-0.69	0.21	0.07	0.02
	4	0.95	0.21	0.21	0.10
	5	1.60	0.15	0.25	0.16
	6	0.37	0.25	0.16	0.07
	7	0.59	0.24	0.18	0.08
	8	1.04	0.21	0.22	0.11
	9	2.30	0.09	0.23	0.23
	10	1.03	0.21	0.22	0.11
	11	0.31	0.25	0.16	0.06
	12	2.66	0.07	0.20	0.24
	13	2.02	0.12	0.25	0.20
	14	1.36	0.18	0.24	0.14
	15	1.77	0.14	0.25	0.18
	16	0.76	0.23	0.20	0.09
	17	1.20	0.19	0.23	0.13
	18	2.35	0.09	0.23	0.23
	19	2.35	0.09	0.23	0.23
	20	0.91	0.22	0.21	0.10
	21	1.44	0.17	0.24	0.15
	22	0.52	0.24	0.18	0.07
	23	1.29	0.18	0.24	0.13
	24	1.48	0.17	0.25	0.15
	25	0.63	0.24	0.19	0.08
SP	1	-1.59	0.13	0.03	0.01
	2	-0.87	0.20	0.06	0.02
	3	0.48	0.24	0.17	0.07
	4	0.61	0.24	0.18	0.08
	5	0.64	0.24	0.19	0.08
	6	0.24	0.25	0.15	0.06
	7	0.77	0.23	0.20	0.09
	8	1.35	0.18	0.24	0.14
	9	0.39	0.25	0.16	0.07
	10	0.48	0.24	0.17	0.07
	11	0.16	0.25	0.14	0.05
	12	0.12	0.25	0.14	0.05
	13	0.33	0.25	0.16	0.06
	14	-0.02	0.25	0.13	0.05
	15	0.09	0.25	0.14	0.05
	16	0.74	0.23	0.20	0.09

Grade 8

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	-1.58	0.07	0.02	0.01
	2	-0.87	0.12	0.04	0.02
	3	-0.83	0.13	0.04	0.02
	4	-0.32	0.17	0.06	0.03
	5	0.99	0.25	0.16	0.09
	6	1.60	0.22	0.22	0.14
	7	2.02	0.19	0.24	0.18
	8	1.82	0.21	0.23	0.17
	9	0.04	0.21	0.08	0.04
	10	1.13	0.25	0.18	0.10
	11	1.20	0.25	0.18	0.11
	12	2.08	0.18	0.25	0.19
	13	0.93	0.25	0.16	0.09
	14	1.62	0.22	0.22	0.15
	15	1.39	0.24	0.20	0.13
	16	1.70	0.22	0.23	0.15
	17	1.64	0.22	0.22	0.15
	18	1.52	0.23	0.21	0.14
	19	1.56	0.23	0.22	0.14
	20	1.77	0.21	0.23	0.16
	21	0.50	0.24	0.12	0.06
	22	0.81	0.25	0.15	0.08
	23	1.18	0.25	0.18	0.11
	24	2.12	0.18	0.25	0.19
	25	2.33	0.16	0.25	0.21
	26	2.20	0.17	0.25	0.20
	27	2.04	0.19	0.24	0.19
WR	1	-2.37	0.03	0.01	0.00
	2	-2.35	0.04	0.01	0.00
	3	-0.60	0.15	0.05	0.02
	4	-0.12	0.19	0.07	0.04
	5	-0.44	0.16	0.06	0.03
	6	0.70	0.25	0.14	0.07
	7	1.42	0.23	0.20	0.13
	8	1.52	0.23	0.21	0.14
	9	0.79	0.25	0.14	0.08
	10	0.02	0.21	0.08	0.04
	11	1.08	0.25	0.17	0.10
	12	0.80	0.25	0.15	0.08
	13	0.84	0.25	0.15	0.08
	14	0.44	0.24	0.11	0.06
	15	1.84	0.20	0.24	0.17
	16	0.00	0.20	0.08	0.04
	17	0.05	0.21	0.08	0.04
	18	0.68	0.25	0.13	0.07
	19	1.59	0.22	0.22	0.14

E.5: Grade 9–Grade 12**Grade 9****Listening and Speaking**

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-0.77	0.16	0.05	0.02
	2	-0.95	0.14	0.04	0.01
	3	0.10	0.23	0.10	0.04
	4	0.12	0.23	0.10	0.04
	5	1.69	0.19	0.23	0.14
	6	0.32	0.24	0.11	0.05
	7	1.99	0.17	0.25	0.17
	8	1.93	0.17	0.24	0.16
	9	0.94	0.25	0.17	0.08
	10	2.18	0.15	0.25	0.19
	11	2.42	0.13	0.25	0.21
	12	2.83	0.09	0.23	0.24
	13	3.18	0.07	0.20	0.25
	14	1.63	0.20	0.23	0.13
	15	1.69	0.19	0.23	0.14
	16	0.49	0.25	0.13	0.05
	17	0.48	0.25	0.13	0.05
	18	0.98	0.24	0.17	0.08
	19	1.34	0.22	0.21	0.11
	20	0.45	0.25	0.12	0.05
	21	1.67	0.20	0.23	0.14
	22	2.10	0.16	0.25	0.18
	23	1.51	0.21	0.22	0.12
	24	2.20	0.15	0.25	0.19
	25	2.02	0.16	0.25	0.17
SP	1	-0.51	0.18	0.06	0.02
	2	-0.25	0.21	0.07	0.03
	3	0.82	0.25	0.16	0.07
	4	0.97	0.24	0.17	0.08
	5	1.17	0.23	0.19	0.10
	6	1.45	0.21	0.22	0.12
	7	0.55	0.25	0.13	0.06
	8	1.33	0.22	0.21	0.11
	9	0.62	0.25	0.14	0.06
	10	0.58	0.25	0.14	0.06
	11	0.84	0.25	0.16	0.07
	12	0.21	0.24	0.10	0.04
	13	0.56	0.25	0.13	0.06
	14	1.00	0.24	0.17	0.08
	15	0.87	0.25	0.16	0.08
	16	0.89	0.25	0.16	0.08

Grade 9

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	0.30	0.20	0.07	0.03
	2	1.39	0.25	0.16	0.09
	3	1.52	0.24	0.17	0.10
	4	0.98	0.25	0.12	0.06
	5	0.68	0.23	0.10	0.05
	6	2.25	0.19	0.23	0.16
	7	2.44	0.18	0.24	0.18
	8	1.76	0.23	0.19	0.12
	9	2.86	0.14	0.25	0.22
	10	1.51	0.24	0.17	0.10
	11	3.01	0.12	0.25	0.23
	12	2.49	0.17	0.24	0.19
	13	2.25	0.19	0.23	0.16
	14	0.45	0.22	0.08	0.04
	15	1.68	0.24	0.19	0.11
	16	1.92	0.22	0.21	0.13
	17	2.77	0.14	0.25	0.21
	18	3.02	0.12	0.25	0.23
	19	1.41	0.25	0.16	0.09
	20	2.06	0.21	0.22	0.15
	21	1.94	0.22	0.21	0.13
	22	1.64	0.24	0.18	0.11
	23	1.36	0.25	0.15	0.09
	24	1.91	0.22	0.21	0.13
	25	2.06	0.21	0.22	0.14
	26	1.39	0.25	0.16	0.09
	27	2.78	0.14	0.25	0.21
WR	1	0.15	0.19	0.06	0.03
	2	1.57	0.24	0.18	0.10
	3	1.74	0.23	0.19	0.12
	4	1.49	0.25	0.17	0.10
	5	0.99	0.25	0.12	0.06
	6	2.38	0.18	0.24	0.18
	7	2.04	0.21	0.22	0.14
	8	1.35	0.25	0.15	0.09
	9	1.73	0.23	0.19	0.12
	10	2.19	0.20	0.23	0.16
	11	1.67	0.24	0.18	0.11
	12	2.08	0.21	0.22	0.15
	13	2.73	0.15	0.25	0.21
	14	2.03	0.21	0.22	0.14
	15	2.16	0.20	0.23	0.16
	16	1.07	0.25	0.13	0.07
	17	1.75	0.23	0.19	0.12
	18	2.03	0.21	0.22	0.14
	19	2.29	0.19	0.23	0.17

Grade 10

Listening and Speaking

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-0.77	0.14	0.04	0.01
	2	-0.95	0.13	0.03	0.01
	3	0.10	0.22	0.08	0.03
	4	0.12	0.22	0.08	0.03
	5	1.69	0.21	0.22	0.12
	6	0.32	0.24	0.10	0.04
	7	1.99	0.18	0.24	0.15
	8	1.93	0.18	0.23	0.15
	9	0.94	0.25	0.15	0.07
	10	2.18	0.16	0.25	0.17
	11	2.42	0.14	0.25	0.19
	12	2.83	0.10	0.24	0.23
	13	3.18	0.08	0.22	0.25
	14	1.63	0.21	0.21	0.12
	15	1.69	0.21	0.22	0.12
	16	0.49	0.24	0.11	0.05
	17	0.48	0.24	0.11	0.05
	18	0.98	0.25	0.15	0.07
	19	1.34	0.23	0.19	0.10
	20	0.45	0.24	0.11	0.04
	21	1.67	0.21	0.22	0.12
	22	2.10	0.17	0.24	0.16
	23	1.51	0.22	0.20	0.11
	24	2.20	0.16	0.25	0.17
	25	2.02	0.18	0.24	0.15
SP	1	-0.51	0.17	0.05	0.02
	2	-0.25	0.19	0.06	0.02
	3	0.82	0.25	0.14	0.06
	4	0.97	0.25	0.15	0.07
	5	1.17	0.24	0.17	0.08
	6	1.45	0.23	0.20	0.10
	7	0.55	0.25	0.11	0.05
	8	1.33	0.23	0.19	0.10
	9	0.62	0.25	0.12	0.05
	10	0.58	0.25	0.12	0.05
	11	0.84	0.25	0.14	0.06
	12	0.21	0.23	0.09	0.04
	13	0.56	0.25	0.11	0.05
	14	1.00	0.25	0.15	0.07
	15	0.87	0.25	0.14	0.06
	16	0.89	0.25	0.14	0.07

Grade 10

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	0.30	0.20	0.06	0.03
	2	1.39	0.25	0.15	0.08
	3	1.52	0.25	0.16	0.09
	4	0.98	0.24	0.11	0.05
	5	0.68	0.23	0.09	0.04
	6	2.25	0.20	0.22	0.15
	7	2.44	0.18	0.24	0.16
	8	1.76	0.24	0.18	0.10
	9	2.86	0.14	0.25	0.20
	10	1.51	0.25	0.16	0.09
	11	3.01	0.13	0.25	0.22
	12	2.49	0.18	0.24	0.17
	13	2.25	0.20	0.22	0.15
	14	0.45	0.21	0.07	0.03
	15	1.68	0.24	0.18	0.10
	16	1.92	0.23	0.20	0.12
	17	2.77	0.15	0.25	0.20
	18	3.02	0.13	0.25	0.22
	19	1.41	0.25	0.15	0.08
	20	2.06	0.22	0.21	0.13
	21	1.94	0.23	0.20	0.12
	22	1.64	0.24	0.17	0.09
	23	1.36	0.25	0.14	0.07
	24	1.91	0.23	0.20	0.12
	25	2.06	0.22	0.21	0.13
	26	1.39	0.25	0.15	0.08
	27	2.78	0.15	0.25	0.20
WR	1	0.15	0.18	0.06	0.03
	2	1.57	0.25	0.16	0.09
	3	1.74	0.24	0.18	0.10
	4	1.49	0.25	0.16	0.08
	5	0.99	0.24	0.11	0.05
	6	2.38	0.19	0.23	0.16
	7	2.04	0.22	0.21	0.13
	8	1.35	0.25	0.14	0.07
	9	1.73	0.24	0.18	0.10
	10	2.19	0.21	0.22	0.14
	11	1.67	0.24	0.17	0.10
	12	2.08	0.22	0.21	0.13
	13	2.73	0.16	0.25	0.19
	14	2.03	0.22	0.21	0.13
	15	2.16	0.21	0.22	0.14
	16	1.07	0.25	0.12	0.06
	17	1.75	0.24	0.18	0.10
	18	2.03	0.22	0.21	0.13
	19	2.29	0.20	0.23	0.15

Grade 11**Listening and Speaking**

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-0.77	0.12	0.03	0.01
	2	-0.95	0.11	0.03	0.01
	3	0.10	0.21	0.07	0.03
	4	0.12	0.21	0.07	0.03
	5	1.69	0.22	0.20	0.11
	6	0.32	0.22	0.08	0.03
	7	1.99	0.20	0.22	0.13
	8	1.93	0.20	0.22	0.13
	9	0.94	0.25	0.13	0.06
	10	2.18	0.18	0.24	0.15
	11	2.42	0.16	0.25	0.17
	12	2.83	0.12	0.25	0.21
	13	3.18	0.09	0.23	0.24
	14	1.63	0.23	0.19	0.10
	15	1.69	0.22	0.20	0.11
	16	0.49	0.23	0.09	0.04
	17	0.48	0.23	0.09	0.04
	18	0.98	0.25	0.13	0.06
	19	1.34	0.24	0.16	0.08
	20	0.45	0.23	0.09	0.04
	21	1.67	0.22	0.20	0.11
	22	2.10	0.19	0.23	0.14
	23	1.51	0.23	0.18	0.09
	24	2.20	0.18	0.24	0.15
	25	2.02	0.19	0.22	0.14
SP	1	-0.51	0.15	0.04	0.02
	2	-0.25	0.17	0.05	0.02
	3	0.82	0.25	0.12	0.05
	4	0.97	0.25	0.13	0.06
	5	1.17	0.25	0.15	0.07
	6	1.45	0.24	0.17	0.09
	7	0.55	0.24	0.10	0.04
	8	1.33	0.24	0.16	0.08
	9	0.62	0.24	0.10	0.04
	10	0.58	0.24	0.10	0.04
	11	0.84	0.25	0.12	0.05
	12	0.21	0.21	0.07	0.03
	13	0.56	0.24	0.10	0.04
	14	1.00	0.25	0.13	0.06
	15	0.87	0.25	0.12	0.05
	16	0.89	0.25	0.12	0.06

Grade 11

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	0.30	0.19	0.05	0.02
	2	1.39	0.25	0.12	0.07
	3	1.52	0.25	0.14	0.07
	4	0.98	0.24	0.09	0.05
	5	0.68	0.22	0.07	0.03
	6	2.25	0.21	0.20	0.13
	7	2.44	0.19	0.22	0.15
	8	1.76	0.24	0.16	0.09
	9	2.86	0.15	0.24	0.19
	10	1.51	0.25	0.14	0.07
	11	3.01	0.14	0.25	0.20
	12	2.49	0.19	0.22	0.15
	13	2.25	0.21	0.20	0.13
	14	0.45	0.20	0.06	0.03
	15	1.68	0.24	0.15	0.08
	16	1.92	0.23	0.17	0.10
	17	2.77	0.16	0.24	0.18
	18	3.02	0.14	0.25	0.20
	19	1.41	0.25	0.13	0.07
	20	2.06	0.22	0.19	0.11
	21	1.94	0.23	0.18	0.10
	22	1.64	0.25	0.15	0.08
	23	1.36	0.25	0.12	0.06
	24	1.91	0.23	0.17	0.10
	25	2.06	0.22	0.19	0.11
	26	1.39	0.25	0.12	0.07
	27	2.78	0.16	0.24	0.18
WR	1	0.15	0.17	0.04	0.02
	2	1.57	0.25	0.14	0.08
	3	1.74	0.24	0.16	0.09
	4	1.49	0.25	0.13	0.07
	5	0.99	0.24	0.09	0.05
	6	2.38	0.20	0.22	0.14
	7	2.04	0.23	0.19	0.11
	8	1.35	0.25	0.12	0.06
	9	1.73	0.24	0.16	0.09
	10	2.19	0.21	0.20	0.12
	11	1.67	0.25	0.15	0.08
	12	2.08	0.22	0.19	0.11
	13	2.73	0.16	0.24	0.17
	14	2.03	0.23	0.18	0.11
	15	2.16	0.22	0.20	0.12
	16	1.07	0.24	0.10	0.05
	17	1.75	0.24	0.16	0.09
	18	2.03	0.23	0.18	0.11
	19	2.29	0.21	0.21	0.13

Grade 12

Listening and Speaking

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
LI	1	-0.77	0.11	0.02	0.01
	2	-0.95	0.10	0.02	0.01
	3	0.10	0.19	0.05	0.03
	4	0.12	0.19	0.05	0.03
	5	1.69	0.23	0.17	0.11
	6	0.32	0.21	0.06	0.03
	7	1.99	0.21	0.20	0.13
	8	1.93	0.21	0.19	0.13
	9	0.94	0.25	0.10	0.06
	10	2.18	0.19	0.22	0.15
	11	2.42	0.17	0.23	0.17
	12	2.83	0.13	0.25	0.21
	13	3.18	0.10	0.25	0.24
	14	1.63	0.24	0.17	0.10
	15	1.69	0.23	0.17	0.11
	16	0.49	0.23	0.07	0.04
	17	0.48	0.22	0.07	0.04
	18	0.98	0.25	0.11	0.06
	19	1.34	0.25	0.14	0.08
	20	0.45	0.22	0.07	0.04
	21	1.67	0.23	0.17	0.11
	22	2.10	0.20	0.21	0.14
	23	1.51	0.24	0.15	0.09
	24	2.20	0.19	0.22	0.15
	25	2.02	0.21	0.20	0.14
SP	1	-0.51	0.14	0.03	0.02
	2	-0.25	0.16	0.04	0.02
	3	0.82	0.24	0.09	0.05
	4	0.97	0.25	0.11	0.06
	5	1.17	0.25	0.12	0.07
	6	1.45	0.24	0.15	0.09
	7	0.55	0.23	0.08	0.04
	8	1.33	0.25	0.14	0.08
	9	0.62	0.23	0.08	0.04
	10	0.58	0.23	0.08	0.04
	11	0.84	0.24	0.10	0.05
	12	0.21	0.20	0.06	0.03
	13	0.56	0.23	0.08	0.04
	14	1.00	0.25	0.11	0.06
	15	0.87	0.25	0.10	0.05
	16	0.89	0.25	0.10	0.06

Grade 12

Reading and Writing

	Item	Rasch Difficulty	Item Information at Intermediate Cut	Item Information at Advanced Cut	Item Information at Proficient Cut
RD	1	0.30	0.18	0.05	0.02
	2	1.39	0.25	0.11	0.07
	3	1.52	0.25	0.12	0.07
	4	0.98	0.23	0.08	0.05
	5	0.68	0.21	0.06	0.03
	6	2.25	0.22	0.19	0.13
	7	2.44	0.20	0.21	0.15
	8	1.76	0.25	0.15	0.09
	9	2.86	0.16	0.24	0.19
	10	1.51	0.25	0.12	0.07
	11	3.01	0.15	0.24	0.20
	12	2.49	0.20	0.21	0.15
	13	2.25	0.22	0.19	0.13
	14	0.45	0.19	0.05	0.03
	15	1.68	0.25	0.14	0.08
	16	1.92	0.24	0.16	0.10
	17	2.77	0.17	0.23	0.18
	18	3.02	0.15	0.25	0.20
	19	1.41	0.25	0.11	0.07
	20	2.06	0.23	0.17	0.11
	21	1.94	0.24	0.16	0.10
	22	1.64	0.25	0.13	0.08
	23	1.36	0.25	0.11	0.06
	24	1.91	0.24	0.16	0.10
	25	2.06	0.23	0.17	0.11
	26	1.39	0.25	0.11	0.07
	27	2.78	0.17	0.23	0.18
WR	1	0.15	0.16	0.04	0.02
	2	1.57	0.25	0.13	0.08
	3	1.74	0.25	0.14	0.09
	4	1.49	0.25	0.12	0.07
	5	0.99	0.24	0.08	0.05
	6	2.38	0.21	0.20	0.14
	7	2.04	0.23	0.17	0.11
	8	1.35	0.25	0.11	0.06
	9	1.73	0.25	0.14	0.09
	10	2.19	0.22	0.19	0.12
	11	1.67	0.25	0.14	0.08
	12	2.08	0.23	0.18	0.11
	13	2.73	0.17	0.23	0.17
	14	2.03	0.23	0.17	0.11
	15	2.16	0.22	0.18	0.12
	16	1.07	0.24	0.09	0.05
	17	1.75	0.25	0.14	0.09
	18	2.03	0.23	0.17	0.11
	19	2.29	0.21	0.20	0.13

APPENDIX F: CONFIRMATORY FACTOR ANALYSIS FOR THE 2007 NYSESLAT

New York State regulations, under the Federal Title III of the *No Child Left Behind* (NCLB) Act of 2001, require annual assessment of limited English proficient students using the New York State English as a Second Language Achievement Test (NYSESLAT) as a state approved assessment tool. This test has been developed in accordance with the Standards for Educational and Psychological Testing (American Educational Research Association, 1999) and the New York State testing requirements, within the principles and consistency of Universal Design and applicable federal and state testing requirements. To meet these requirements, the test was developed for five grade spans (K–1, 2–4, 5–6, 7–8, and 9–12) in four modalities, i.e., Speaking, Listening, Reading, and Writing, to assess the English language proficiency of students from kindergarten through grade 12 who are English language learners.

The maximum number of points by modalities and grade span for 2007 Administration is depicted in the following table:

Table F.1.
Maximum Number of Points by Modality and Grade Span for the 2007 NYSESLAT

Maximum number of Points in NYSESLAT Subtests							
Grade Span	Speaking	Listening	Reading	Writing			Total Number of Points
	CR	MC	MC	Writing Conventions	Pre-Writing	Writing Prompt	
				MC	SR	ER	
K–1	34	24	15	6	0	16 CR (Dev. Writing) + 2 ER	97
2–4	34	24	24	12	6	4	104
5–6	34	25	27	15	6	4	111
7–8	34	25	27	15	6	4	111
9–12	34	25	27	15	6	4	111

F.1: Reporting of the NYSESLAT Scores

The NYSESLAT scores are reported on two different scales, i.e., a scale that is a combination of Speaking and Listening modalities and another which is a combination of the Reading and Writing modalities. In 2007, the New York State Department of Education (NYSDE) professed an interest in examining the dimensionality of the data, with an eye toward the reporting of scores on a single English Language Performance Achievement (ELPA) score.

In order to test the assumption of unidimensionality and compare it with the currently used two-dimensional model, a confirmatory factor analysis was used to examine the factor loading based on the different combination of the observed variables.

F.2: Method and Process

Since the primary interest of this study was to compare the current reporting structure with an alternate pre-conceived unidimensional structure, a confirmatory factor model was the obvious choice for analysis from the various other factor models available.

This study, therefore, tested the following two structures:

- A one factor model with the latent variable being ELPA
- A two factor model that tests the current structure of Speaking and Listening as one factor (SPKLIS), and Reading, Writing Conventions (Wrcon), and Writing Constructed-Response items (Wrcr, Pre-writing + Writing Prompt) as the other factor (WRTRD)

The results of these confirmatory factor analyses are based on the 2007 data set, and it is based on the full population responses.

Since this study was not intended to assess the appropriateness of individual item loading on the different modality-factors, all items were collapsed within their respective modality to produce five scores for each individual student, i.e., for Speaking, Listening, Reading, Writing Conventions (Wrcon), and Writing Constructed-Response Items (Wrcr).

The factor loadings for the one- and the two-factor models were therefore based on subtests of Speaking, Listening, Reading, Wrcon, and Wrcr.

For the NYSESLAT, since the items were collapsed to produce the subtests discussed above, there were five modality measures that extend from a low of 6 scoring categories for the Writing Conventions to a high of 34 categories for the Speaking subtest across all grade spans. The subtests can therefore be looked upon as polytomously scored items.

Furthermore, as Jöreskog and Sörbom (1995a) have stated, for ordered categorical data that are not normally distributed but assumed to reflect normally distributed underlying variables, a better way to estimate the population correlations is by using polychloric correlation coefficients. With this in mind, polychloric correlations were used with the ML estimation procedure in this study.

Although there are various methods to evaluate the confirmatory factor model, a more global approach was used to assess the relative performance of the different models. The most commonly used indices were referenced in this study, i.e., the goodness-of-fit index (GFI), the adjusted-goodness-of-fit index (AGFI), the root-mean-squared residual (RMR) and the root-mean-squared error of approximation (RMSEA) and the χ^2 index. However, a general consensus as to which of these methods is best does not exist, and no special weight is given to any of the indices outlined above in this study.

According to Swygert, et al. (2001), in the case of the overall fit of the model, the χ^2 index should not be seen as much as a test of significance for the null of no difference in the fit of the model to the data but should be used as an indication of how far the data deviate from the model. High values indicate large deviance and small values indicate small deviance. Furthermore, large sample sizes will generally produce larger Chi-square values.

The GFI and AGFI indices have a range of 0 to 1, where values closer to 1 indicate a better fit of the model to the data. On the other hand, the RMR is an index produced by the averaging of the fitted residuals. Hu and Bentler (1999) suggest $RMR < 0.06$ as acceptable. Similarly, Hu and Bentler (1999) suggest < 0.05 as acceptable for the RMSEA, which is a measure of the goodness of fit with known distribution properties. On the other hand, Browne and Cudeck (1993) and MacCallum, Browne, and Sugawara suggest $RMSEA \leq 0.05$ as close model fit, values ranging from 0.06 to 0.08 as adequate fit, those between 0.09 and 0.10 as mediocre fit, and values greater than 0.10 as poor fit.

The path diagrams for each of the one- and the two-factor models are shown in Figures F.1 and F.2:

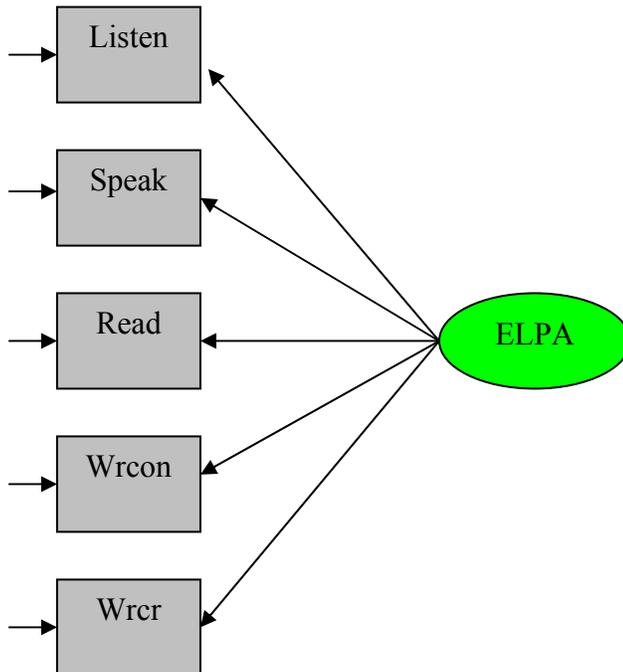


Figure F.1. Path Diagram for the One-Factor Model

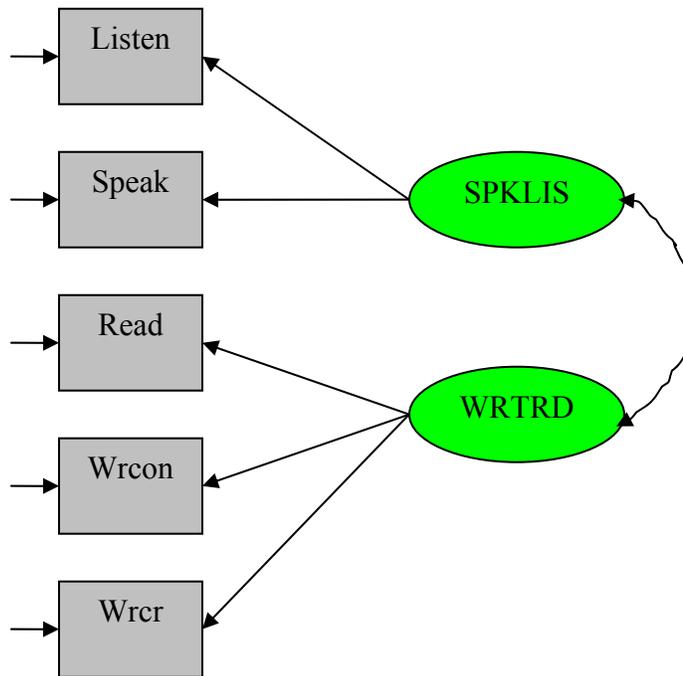


Figure F.2. Path Diagram for the Two-Factor Model

The structural equation modeling together with the polychoric correlation coefficients and ML estimation procedures were implemented by the use of Jöreskog's and Sörbom's Liseral 8.71 computer program. And the latent variable is standardized.

F.3: Results and Comments

Grade Span K-1

The summary statistics are provided in Table F.2a. Factor loadings for the one and the two factor models are provided in Table F.2b with the fit indices for the two models displayed in Table F.2c.

**Table F.2a.
Summary Statistics**

Variable	N-count	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Maximum
Listen	49845	20.48	3.72	-1.72	3.45	0.00	24.00
Speak	49845	26.07	7.88	-1.50	1.98	0.00	34.00
Read	49845	9.41	3.93	-0.08	-1.23	0.00	15.00
Wrcon	49845	5.31	1.22	-1.96	3.28	0.00	6.00
Wrcr	49845	12.34	4.44	-0.68	-0.26	0.00	18.00

Table F.2b.
Factor Loading Estimates for the One- and Two-Factor Models

Modalities	One-Factor Model	Two-Factor Model	
	Λ	Λ_1	Λ_2
Listen	0.67	0.83	–
Speak	0.60	0.73	–
Read	0.79	–	0.80
Wrcon	0.83	–	0.83
Wrcr	0.89	–	0.89
		Correlation	0.78

Table F.2c.
Global Fit Indices for the One- and Two-Factor Models

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	8070.68	1158.31
GFI	0.94	0.99
AGFI	0.82	0.96
RMR	0.06	0.02
RMSEA	0.18	0.08

Note: χ^2 with 5 *df*, $p=0.00$ for the one-factor model and χ^2 with 4 *df*, $p=0.00$ for the two-factor model.

For the one-factor model, the values of GFI and AGFI are 0.940 and 0.820 respectively, and the RMR = 0.057 and the RMSEA = 0.180 are not sufficiently small, indicating that the model fit leaves much to be desired. The Chi-square value is also sufficiently large and significant with 5 degrees of freedom, suggesting a revision of the model for a better fit.

In keeping with the current dimensionality of the test, the two-factor model with the loading of Speak and Listen on one factor (SPKLIS), and the Writing Conventions and Writing Constructed-Response and Read on the second factor (WRTRD) was modeled. The results indicate a great improvement in fit of the model to the data. The rather weak loading of the Listen modality (0.67) is considerably improved with the two-factor model to 0.83. Although the *p*-value of the Chi-square is still in the range of significance at the 0.05 level of rejection ($\chi^2 = 1158.310$, $p = 0.00$), the Chi-square itself is reduced tremendously. The RMR and RMSEA values have decreased a great deal to 0.015 and 0.077 respectively, and the GFI as well as the AGFI are at their peak of 0.990 and 0.960.

Grade Span 2–4

The summary statistics are provided in Table F.3a. Factor loadings for the one- and the two-factor models are provided in Table F.3b with the fit indices for the two models displayed in Table F.3c.

Table F.3a.
Summary Statistics

Variable	N-count	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Maximum
Listen	56980	18.49	4.23	-1.13	0.98	0.00	24.00
Speak	56980	28.94	7.00	-2.40	5.94	0.00	34.00
Read	56980	16.44	5.09	-0.51	-0.47	0.00	24.00
Wrcon	56980	8.54	2.73	-0.80	-0.16	0.00	12.00
Wrcr	56980	6.49	2.69	-1.01	0.26	0.00	10.00

Table F.3b.
Factor Loading Estimates for the One- and Two-Factor Models

Modalities	One-Factor Model	Two-Factor Model	
	Λ	Λ_1	Λ_2
Listen	0.77	0.88	-
Speak	0.61	0.67	-
Read	0.87	-	0.89
Wrcon	0.81	-	0.82
Wrcr	0.77	-	0.76
		Correlation	0.85

Table F.3c.
Global Fit Indices for the One- and Two-Factor Models

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	6712.14	2161.19
GFI	0.95	0.98
AGFI	0.86	0.94
RMR	0.04	0.03
RMSEA	0.15	0.10

Note: χ^2 with 5 *df*, $p=0.00$ for the one-factor model and χ^2 with 4 *df*, $p=0.00$ for the two-factor model.

The fit of the model to the data is very poor with the one-factor model. The Chi-square value is extremely large ($\chi^2 = 6712.14$ with 5 *df*, $p = 0.00$). The value of GFI is 0.950 with a poor value for the AGFI, i.e., 0.860. The RMR is acceptable at 0.041 while the RMSEA is large at 0.150. A summary of the fit indices suggest that an alternate model is justifiable.

With the two-factor model, the Chi-square reduces substantially to 2161.190 (with 4 *df*, $p = 0.00$). All the other fit indices also react positively with the GFI almost closer to the expected value of 0.980. Although the two-factor model shows substantial improvement over the one-factor model, the low values of RMR = 0.025 and RMSEA = 0.098 and the acceptable value of AGFI = 0.940 suggest that the two-factor model is a substantial improvement over the one-factor model and is an acceptable model.

Grade Span 5–6

The summary statistics are provided in Table F.4a. Factor loadings for the one- and the two-factor models are provided in Table F.4b with the fit indices for the two models displayed in Table F.4c.

Table F.4a.
Summary Statistics

Variable	N-count	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Maximum
Listen	24543	19.35	4.95	-1.33	1.15	0.00	25.00
Speak	24543	28.42	7.83	-2.08	3.92	0.00	34.00
Read	24543	17.72	5.65	-0.46	-0.71	0.00	27.00
Wrcon	24543	10.89	3.29	-0.91	0.03	0.00	15.00
Wrcr	24543	7.18	2.46	-1.39	1.61	0.00	10.00

Table F.4b.
Factor Loading Estimates for the One- and Two-Factor Models

Modalities	One-Factor Model	Two-Factor Model	
	Λ	Λ_1	Λ_2
Listen	0.84	0.88	-
Speak	0.73	0.75	-
Read	0.82	-	0.83
Wrcon	0.85	-	0.86
Wrcr	0.71	-	0.71
		Correlation	0.93

Table F.4c.
Global Fit Indices for the One- and Two-Factor Models

Fit Indices	One-Factor Model	Two-Factor Model
Listen	1931.25	1148.99
Speak	0.97	0.98
Read	0.92	0.94
Wrcon	0.03	0.02
Wrcr	0.12	0.10

Note: χ^2 with 5 *df*, $p=0.00$ for the one-factor model and χ^2 with 4 *df*, $p=0.00$ for the two-factor model.

The fit of the models to the data for the 5–6 grade span follows the same trend as the 2–4 grade span dimensionality tests with substantial reduction in the Chi-square from the one- to the two-factor model ($\chi^2 = 1148.990$ with 4 *df*, $p = 0.00$ for the two-factor model). The RMR and the RMSEA also show a reduction. Similarly, there is an improvement for the GFI and the AGFI with the GFI showing a healthy value of 0.980.

Grade Span 7–8

The summary statistics are provided in Table F.5a. Factor loadings for the one- and the two-factor models are provided in Table F.5b with the fit indices for the two models displayed in Table F.5c.

Table F.5a.
Summary Statistics

Variable	N-count	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Maximum
Listen	22295	17.93	5.50	-0.86	-0.26	0.00	25.00
Speak	22295	27.23	8.93	-1.64	1.83	0.00	34.00
Read	22295	16.16	5.97	-0.12	-1.07	0.00	27.00
Wrcon	22295	10.91	3.33	-0.67	-0.50	0.00	15.00
Wrer	22295	6.21	2.88	-0.83	-0.26	0.00	10.00

Table F.5b.
Factor Loading Estimates for the One- and Two-Factor Models

Modalities	One-Factor Model	Two-Factor Model	
	Λ	Λ_1	Λ_2
Listen	0.86	0.87	-
Speak	0.75	0.75	-
Read	0.79	-	0.80
Wrcon	0.84	-	0.84
Wrer	0.75	-	0.75
		Correlation	0.98

Table F.5c.
Global Fit Indices for the One- and Two-Factor Models

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	2003.49	2222.93
GFI	0.97	0.97
AGFI	0.90	0.88
RMR	0.03	0.03
RMSEA	0.13	0.15

Note: χ^2 with 5 *df*, *p*=0.00 for the one-factor model and χ^2 with 4 *df*, *p*=0.00 for the two-factor model.

There does not seem to be much improvement in the 7–8 grade span for the two-factor model when compared to the one-factor model. The Chi-square increases from the one-factor value of 2003.490 with 5 *df* to 2222.930 with 4 degrees of freedom. However, mixed results are obtained for the other fit indices. For example, GFI and RMR both show same results over their counterparts in the one-factor model while the AGFI and RMSEA show slightly poorer fit.

Grade Span 9–12

The summary statistics are provided in Table F.6a. Factor loadings for the one- and the two-factor models are provided in Table F.6b with the fit indices for the two models displayed in Table F.6c.

Table F.6a.
Summary Statistics

Variable	N-count	Mean	St. Dev.	Skewness	Kurtosis	Minimum	Maximum
Listen	34094	17.54	4.82	-0.71	-0.18	0.00	25.00
Speak	34094	25.75	8.77	-1.29	1.04	0.00	34.00
Read	34094	17.79	5.42	-0.45	-0.66	0.00	27.00
Wrcon	34094	10.03	3.57	-0.53	-0.73	0.00	15.00
Wrwr	34094	6.24	3.02	-0.68	-0.59	0.00	10.00

Table F.6b.
Factor Loading Estimates for the One- and Two-Factor Models

Modalities	One-Factor Model	Two-Factor Model	
	Λ	Λ_1	Λ_2
Listen	0.82	0.83	-
Speak	0.69	0.70	-
Read	0.82	-	0.83
Wrcon	0.82	-	0.82
Wrwr	0.74	-	0.74
		Correlation	0.98

Table F.6c.
Global Fit Indices for the One- and Two-Factor Models

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	2172.82	2108.66
GFI	0.98	0.98
AGFI	0.93	0.91
RMR	0.03	0.03
RMSEA	0.11	0.12

Note: χ^2 with 2 *df*, *p*=0.00 for the one-factor model and χ^2 with 1 *df*, *p*=0.00 for the two-factor model.

Similar to the 2005 and 2006 data, 7–8 grade span results, and the 9–12 grade span data also did not show much improvement with the two-factor model over the one-factor model. The factor loadings remain almost identical with the use of the two-factor model in comparison to their loadings on the one-factor model. Although, the Chi-square reduced ($\chi^2 = 2172.820$ with 5 *df* for the one-factor model and $\chi^2 = 2108.660$ with 4 *df* for the two-factor model), most other fit indices remained about the same from the one-factor to the two-factor model. The one-factor model produced slightly better results, but the AGFI increased slightly in comparison to their values under the one-factor model.

F.4: Conclusion

Overall, the 2007 NYSESLAT dimensionality tests with the use of Confirmatory Factor Analysis produced results that substantiate the use of the two-factor model in reporting IRT scores, i.e., Speaking and Listening as one score and Writing Conventions and Writing Constructed-Response Items and Reading as the other. With the exception of the 7–8 and the 9–12 grade spans, the two-factor model provided a better fit of the model to the data than the one-factor unidimensional model as measured by global indices of fit.

If one were to accept the recommendations of Browne et al. (1993) and MacCallum et al. (1996) for the evaluation of the RMSEA as a fit index, and Hu and Bentler (1999) for the RMR index, in conjunction with acceptable values of other fit indices, then the two-factor model for K–1 grade span has the most acceptable fit of the model to the data. Besides having acceptable RMR and RMSEA under the various criteria discussed above, the GFI and the AGFI for grade span K–1 were 0.99 and 0.96.

In the 2–4 and the 5–6 grade spans, the unidimensionality of the NYSESLAT scoring cannot be justifiable because of substantial improvement of the two-factor model fit to the data over the unidimensional model.

With respect to the analysis of the 7–8 and the 9–12 grade span tests, the factor loadings based on the one-factor model remain much the same when they are loaded on two-factors. Since the one-factor model does not show any improvement over the fit of the two-factor model, it could be left alone to preserve the status quo of the NYSESLAT, at least for the next few years. Extra care could also be taken with an eye towards content and substantive issues in selecting items for the future.

APPENDIX G: SCALE SCORE SUMMARY BY SUBGROUPS

G.1: Kindergarten Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
K	LS	Female	11543	588.88	587	51.85	63
K	LS	Male	12696	584.12	582	50.32	58
K	LS	American Indian or Alaska Native	41	572.00	578	51.17	52
K	LS	Black or African-American	858	579.53	582	51.64	64
K	LS	Asian or Pacific Islander	4694	585.51	582	52.62	58
K	LS	Hispanic or Latino	16282	584.99	585	49.85	58
K	LS	White	2329	600.46	596	54.19	60
K	LS	Multi-Racial (not Hispanic origin)	35	603.57	596	49.99	48
K	LS	NYC	15711	586.55	587	53.06	58
K	LS	Big 4 Cities	1174	580.99	582	47.37	55
K	LS	High Need Urban/Suburban	2936	580.45	578	46.88	55
K	LS	High Need Rural	161	587.06	596	37.19	45
K	LS	Average Need	2675	591.21	591	47.61	51
K	LS	Low Need	1425	590.35	591	46.88	48
K	LS	Charter Schools	136	599.99	596	44.61	56
K	LS	0 Years LEP	2611	567.98	574	60.85	76
K	LS	1 Years LEP	21239	588.84	587	49.34	55
K	LS	Arabic	499	584.72	587	56.89	58
K	LS	Bengali	639	588.14	587	48.10	63
K	LS	Chinese	843	586.55	582	52.72	58
K	LS	English	1883	590.02	587	47.21	51
K	LS	Spanish	14532	584.19	582	50.04	58
K	LS	Other Language	4866	590.97	591	54.24	63
K	LS	Students with Disabilities	2060	567.32	566	45.80	56

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
K	RW	Female	11543	548.79	545	43.61	50
K	RW	Male	12696	541.15	536	43.63	54
K	RW	American Indian or Alaska Native	41	538.76	540	34.40	29
K	RW	Black or African-American	858	534.50	531	44.05	53
K	RW	Asian or Pacific Islander	4694	565.98	560	50.22	65
K	RW	Hispanic or Latino	16282	538.76	536	39.55	48
K	RW	White	2329	547.88	545	44.38	50
K	RW	Multi-Racial (not Hispanic origin)	35	560.94	555	37.66	47
K	RW	NYC	15711	549.74	545	45.94	56
K	RW	Big 4 Cities	1174	529.96	526	36.53	43
K	RW	High Need Urban/Suburban	2936	531.45	531	35.05	43
K	RW	High Need Rural	161	537.16	536	30.04	38
K	RW	Average Need	2675	538.34	536	37.51	43
K	RW	Low Need	1425	542.04	536	43.03	49
K	RW	Charter Schools	136	549.92	550	43.61	50
K	RW	0 Years LEP	2611	529.82	526	42.62	53
K	RW	1 Years LEP	21239	546.86	540	43.66	54
K	RW	Arabic	499	541.69	536	42.00	49
K	RW	Bengali	639	563.45	560	47.59	58
K	RW	Chinese	843	573.01	566	51.65	60
K	RW	English	1883	537.31	536	36.94	48
K	RW	Spanish	14532	538.97	536	39.76	48
K	RW	Other Language	4866	559.77	555	49.88	57
K	RW	Students with Disabilities	2060	526.63	526	38.37	43

G.2: Grade 1**Listening and Speaking**

Grade	Test	Group	N-count	Mean	Median	SD	IQR
1	LS	Female	12125	631.41	630	58.59	73
1	LS	Male	13746	627.40	622	57.91	78
1	LS	American Indian or Alaska Native	32	636.25	630	52.95	73
1	LS	Black or African-American	1084	622.72	622	62.35	61
1	LS	Asian or Pacific Islander	5275	634.88	630	58.15	67
1	LS	Hispanic or Latino	17098	625.51	622	56.95	61
1	LS	White	2357	647.28	640	61.42	88
1	LS	Multi-Racial (not Hispanic origin)	25	601.80	596	66.16	99
1	LS	NYC	16913	629.23	630	60.00	78
1	LS	Big 4 Cities	1272	616.07	614	53.73	65
1	LS	High Need Urban/Suburban	3147	622.63	622	53.72	61
1	LS	High Need Rural	191	638.97	630	44.64	61
1	LS	Average Need	2769	635.49	630	54.31	67
1	LS	Low Need	1464	640.82	640	56.25	67
1	LS	Charter Schools	101	652.00	640	49.22	74
1	LS	0 Years LEP	1684	595.53	602	76.14	114
1	LS	1 Years LEP	6189	613.24	614	59.89	74
1	LS	2 Years LEP	17627	638.09	630	53.11	67
1	LS	Arabic	578	624.16	622	62.98	87
1	LS	Bengali	661	634.16	630	57.66	73
1	LS	Chinese	756	635.61	630	63.40	73
1	LS	English	1751	631.54	630	54.13	73
1	LS	Spanish	15900	625.31	622	57.11	61
1	LS	Other Language	5243	637.53	630	60.69	67
1	LS	Students with Disabilities	2682	608.91	608	54.65	62

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
1	RW	Female	12125	618.19	612	59.06	74
1	RW	Male	13746	607.97	604	61.15	85
1	RW	American Indian or Alaska Native	32	627.31	622	56.59	82
1	RW	Black or African-American	1084	596.36	596	63.17	85
1	RW	Asian or Pacific Islander	5275	641.39	635	57.71	74
1	RW	Hispanic or Latino	17098	603.26	604	57.83	69
1	RW	White	2357	625.19	622	60.18	89
1	RW	Multi-Racial (not Hispanic origin)	25	582.68	571	64.17	72
1	RW	NYC	16913	613.27	612	60.80	80
1	RW	Big 4 Cities	1272	593.20	589	61.22	85
1	RW	High Need Urban/Suburban	3147	605.18	604	57.88	69
1	RW	High Need Rural	191	612.07	612	50.93	58
1	RW	Average Need	2769	617.88	612	57.76	74
1	RW	Low Need	1464	629.60	622	59.73	89
1	RW	Charter Schools	101	622.81	622	60.94	68
1	RW	0 Years LEP	1684	583.23	577	63.49	82
1	RW	1 Years LEP	6189	599.74	596	61.03	75
1	RW	2 Years LEP	17627	620.07	622	58.18	68
1	RW	Arabic	578	607.22	604	61.26	85
1	RW	Bengali	661	632.25	635	55.30	82
1	RW	Chinese	756	650.60	651	60.03	66
1	RW	English	1751	611.02	604	59.83	80
1	RW	Spanish	15900	603.49	604	57.62	69
1	RW	Other Language	5243	632.73	635	61.73	89
1	RW	Students with Disabilities	2682	582.91	577	59.53	82

G.3: Grade 2

Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
2	LS	Female	10035	649.64	655	45.08	46
2	LS	Male	11775	648.64	650	44.87	50
2	LS	American Indian or Alaska Native	31	644.03	645	44.00	50
2	LS	Black or African-American	1000	643.29	645	46.41	54
2	LS	Asian or Pacific Islander	3778	655.72	660	48.03	49
2	LS	Hispanic or Latino	15126	646.75	650	43.41	50
2	LS	White	1866	658.03	660	47.62	49
2	LS	Multi-Racial (not Hispanic origin)	9	643.00	628	36.60	50
2	LS	NYC	14680	647.16	650	45.09	50
2	LS	Big 4 Cities	1183	640.92	645	42.68	50
2	LS	High Need Urban/Suburban	2589	648.93	655	43.62	46
2	LS	High Need Rural	117	656.32	667	42.00	49
2	LS	Average Need	2070	659.34	660	44.44	45
2	LS	Low Need	1077	664.15	660	45.28	51
2	LS	Charter Schools	88	652.47	655	30.62	37
2	LS	0 Years LEP	1281	607.81	620	64.96	99
2	LS	1 Years LEP	2925	625.53	628	52.41	69
2	LS	2 Years LEP	4147	653.39	655	40.01	46
2	LS	3 Years LEP	13215	656.94	655	37.41	45
2	LS	Arabic	476	640.42	645	46.18	50
2	LS	Bengali	508	657.21	660	41.35	49
2	LS	Chinese	467	652.73	655	51.24	53
2	LS	English	1447	651.28	655	40.49	53
2	LS	Spanish	13846	646.43	650	43.73	50
2	LS	Other Language	4314	654.82	660	48.34	49
2	LS	Students with Disabilities	3123	634.54	636	40.27	50

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
2	RW	Female	10035	621.98	625	46.01	54
2	RW	Male	11775	615.09	616	46.89	58
2	RW	American Indian or Alaska Native	31	605.06	603	52.41	57
2	RW	Black or African-American	1000	607.22	608	47.80	61
2	RW	Asian or Pacific Islander	3778	636.63	639	48.50	55
2	RW	Hispanic or Latino	15126	613.37	616	44.50	58
2	RW	White	1866	627.03	630	47.69	56
2	RW	Multi-Racial (not Hispanic origin)	9	595.33	578	51.13	80
2	RW	NYC	14680	616.82	621	45.61	58
2	RW	Big 4 Cities	1183	604.48	608	46.56	66
2	RW	High Need Urban/Suburban	2589	617.28	621	45.88	58
2	RW	High Need Rural	117	623.72	630	52.13	54
2	RW	Average Need	2070	625.82	630	48.22	56
2	RW	Low Need	1077	639.09	639	49.84	55
2	RW	Charter Schools	88	627.73	625	39.29	56
2	RW	0 Years LEP	1281	588.40	586	52.54	76
2	RW	1 Years LEP	2925	601.72	603	51.20	70
2	RW	2 Years LEP	4147	620.88	625	46.36	54
2	RW	3 Years LEP	13215	623.87	625	42.66	50
2	RW	Arabic	476	607.39	612	47.99	61
2	RW	Bengali	508	636.19	639	45.31	45
2	RW	Chinese	467	637.37	639	53.15	55
2	RW	English	1447	617.96	621	45.49	58
2	RW	Spanish	13846	613.17	616	44.58	58
2	RW	Other Language	4314	629.80	630	48.59	58
2	RW	Students with Disabilities	3123	590.53	591	44.26	62

G.4: Grade 3

Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
3	LS	Female	8794	666.41	674	49.99	46
3	LS	Male	10475	665.78	667	50.30	51
3	LS	American Indian or Alaska Native	36	666.78	667	35.87	41
3	LS	Black or African-American	927	660.38	660	52.65	55
3	LS	Asian or Pacific Islander	3014	668.18	674	55.03	62
3	LS	Hispanic or Latino	13732	665.02	667	48.28	51
3	LS	White	1555	674.78	681	53.88	52
3	LS	Multi-Racial (not Hispanic origin)	5	596.20	574	64.68	83
3	LS	NYC	13395	665.01	667	50.18	51
3	LS	Big 4 Cities	1000	656.41	660	50.61	49
3	LS	High Need Urban/Suburban	2198	666.03	667	48.63	51
3	LS	High Need Rural	107	676.29	674	49.28	57
3	LS	Average Need	1667	675.20	674	48.58	52
3	LS	Low Need	837	673.65	681	53.59	57
3	LS	Charter Schools	59	683.39	681	36.14	42
3	LS	0 Years LEP	1243	623.25	640	72.58	112
3	LS	1 Years LEP	2350	623.44	624	57.64	74
3	LS	2 Years LEP	2177	665.98	667	46.69	55
3	LS	3 Years LEP	3018	676.96	674	41.68	47
3	LS	4 Years LEP	10293	677.89	674	38.37	47
3	LS	Arabic	478	653.06	664	58.38	71
3	LS	Bengali	369	672.56	674	48.65	52
3	LS	Chinese	336	668.20	674	55.98	57
3	LS	English	1131	669.95	667	46.77	46
3	LS	Spanish	12683	664.58	667	48.32	51
3	LS	Other Language	3652	669.95	674	54.76	57
3	LS	Students with Disabilities	3314	656.02	655	42.48	49

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
3	RW	Female	8794	649.84	655	49.58	61
3	RW	Male	10475	641.50	644	51.23	62
3	RW	American Indian or Alaska Native	36	639.94	655	42.18	55
3	RW	Black or African-American	927	632.57	639	53.84	68
3	RW	Asian or Pacific Islander	3014	658.52	661	52.01	61
3	RW	Hispanic or Latino	13732	642.46	644	48.98	62
3	RW	White	1555	652.77	655	54.85	70
3	RW	Multi-Racial (not Hispanic origin)	5	580.40	533	78.96	94
3	RW	NYC	13395	643.92	649	50.15	58
3	RW	Big 4 Cities	1000	630.85	639	52.10	68
3	RW	High Need Urban/Suburban	2198	648.36	655	49.99	61
3	RW	High Need Rural	107	647.34	649	50.62	70
3	RW	Average Need	1667	654.23	655	51.62	57
3	RW	Low Need	837	657.81	661	51.06	61
3	RW	Charter Schools	59	649.31	649	38.16	66
3	RW	0 Years LEP	1243	609.47	612	58.80	85
3	RW	1 Years LEP	2350	616.83	616	57.10	82
3	RW	2 Years LEP	2177	647.06	649	53.63	70
3	RW	3 Years LEP	3018	653.13	655	47.03	57
3	RW	4 Years LEP	10293	653.41	655	43.67	57
3	RW	Arabic	478	628.66	634	57.10	76
3	RW	Bengali	369	662.84	667	49.42	71
3	RW	Chinese	336	661.77	667	54.34	67
3	RW	English	1131	645.32	649	50.57	58
3	RW	Spanish	12683	642.12	644	48.92	62
3	RW	Other Language	3652	653.51	655	52.80	66
3	RW	Students with Disabilities	3314	617.97	621	47.68	63

G.5: Grade 4

Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
4	LS	Female	7337	672.83	674	53.83	57
4	LS	Male	8737	672.34	681	53.70	57
4	LS	American Indian or Alaska Native	23	682.96	681	41.82	63
4	LS	Black or African-American	859	663.19	667	54.38	55
4	LS	Asian or Pacific Islander	2499	672.03	674	58.13	62
4	LS	Hispanic or Latino	11402	672.30	674	52.72	52
4	LS	White	1287	682.22	691	52.2	63
4	LS	Multi-Racial (not Hispanic origin)	4	590.00	583	72.49	111
4	LS	NYC	11546	672.16	681	53.62	57
4	LS	Big 4 Cities	839	662.85	667	50.53	55
4	LS	High Need Urban/Suburban	1640	669.31	674	54.41	57
4	LS	High Need Rural	92	679.60	681	48.82	47
4	LS	Average Need	1244	680.48	681	54.12	66
4	LS	Low Need	665	683.42	691	56.11	68
4	LS	Charter Schools	48	682.13	681	32.74	45
4	LS	0 Years LEP	1214	628.89	645	74.03	113
4	LS	1 Years LEP	2258	628.81	628	61.62	88
4	LS	2 Years LEP	1719	666.31	667	50.27	70
4	LS	3 Years LEP	1572	683.86	681	44.31	47
4	LS	4 Years LEP	2203	688.72	691	40.41	51
4	LS	5 Years LEP	6972	688.20	681	38.88	51
4	LS	Arabic	353	660.14	667	59.24	78
4	LS	Bengali	278	682.52	681	53.95	63
4	LS	Chinese	305	669.39	674	61.74	90
4	LS	English	880	679.67	681	49.70	50
4	LS	Spanish	10574	671.62	674	52.77	57
4	LS	Other Language	3196	673.67	681	55.66	57
4	LS	Students with Disabilities	3155	671.01	674	43.09	46

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
4	RW	Female	7337	665.47	667	54.20	67
4	RW	Male	8737	656.38	661	54.56	66
4	RW	American Indian or Alaska Native	23	671.26	674	39.06	46
4	RW	Black or African-American	859	644.81	649	55.20	74
4	RW	Asian or Pacific Islander	2499	670.34	674	56.80	62
4	RW	Hispanic or Latino	11402	658.40	661	53.26	66
4	RW	White	1287	670.85	674	56.84	62
4	RW	Multi-Racial (not Hispanic origin)	4	581.25	549	90.48	111
4	RW	NYC	11546	659.73	667	54.17	61
4	RW	Big 4 Cities	839	644.54	649	56.70	74
4	RW	High Need Urban/Suburban	1640	662.15	667	55.38	71
4	RW	High Need Rural	92	657.98	661	42.27	48
4	RW	Average Need	1244	668.70	674	53.86	62
4	RW	Low Need	665	674.62	674	54.93	70
4	RW	Charter Schools	48	674.81	674	37.38	39
4	RW	0 Years LEP	1214	620.95	621	61.57	94
4	RW	1 Years LEP	2258	630.39	630	60.48	88
4	RW	2 Years LEP	1719	658.66	661	58.23	85
4	RW	3 Years LEP	1572	670.78	674	51.98	62
4	RW	4 Years LEP	2203	671.69	674	47.84	57
4	RW	5 Years LEP	6972	671.67	674	45.18	57
4	RW	Arabic	353	644.09	655	63.24	92
4	RW	Bengali	278	675.15	674	53.64	65
4	RW	Chinese	305	670.86	674	59.76	80
4	RW	English	880	666.32	667	53.07	67
4	RW	Spanish	10574	657.76	661	53.37	66
4	RW	Other Language	3196	666.83	674	56.35	67
4	RW	Students with Disabilities	3155	638.11	639	48.11	66

G.6: Grade 5

Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
5	LS	Female	5993	669.46	674	52.86	54
5	LS	Male	7152	668.48	674	51.49	54
5	LS	American Indian or Alaska Native	15	654.93	653	52.15	45
5	LS	Black or African-American	809	663.24	668	50.58	62
5	LS	Asian or Pacific Islander	2012	666.76	671	55.41	62
5	LS	Hispanic or Latino	9285	668.65	674	51.62	54
5	LS	White	1024	680.35	681	49.50	56
5	LS	NYC	9162	668.20	674	52.00	54
5	LS	Big 4 Cities	888	660.28	668	52.81	53
5	LS	High Need Urban/Suburban	1348	667.58	674	50.75	54
5	LS	High Need Rural	71	669.06	663	48.76	54
5	LS	Average Need	1107	677.17	681	54.33	61
5	LS	Low Need	534	681.13	681	48.95	56
5	LS	Charter Schools	18	685.83	685	42.94	62
5	LS	0 Years LEP	1101	634.48	648	72.21	110
5	LS	1 Years LEP	2211	628.87	629	56.72	78
5	LS	2 Years LEP	1613	664.59	663	47.35	56
5	LS	3 Years LEP	1226	677.98	674	41.92	50
5	LS	4 Years LEP	1126	686.84	689	38.40	46
5	LS	5 Years LEP	1568	688.73	689	37.36	41
5	LS	6 Years LEP or More	4140	686.18	681	36.67	46
5	LS	Arabic	303	657.26	658	50.02	72
5	LS	Bengali	202	671.06	674	46.40	54
5	LS	Chinese	199	651.85	658	65.26	85
5	LS	English	697	681.70	681	42.83	51
5	LS	Spanish	8681	667.44	674	51.87	54
5	LS	Other Language	2677	671.68	674	53.66	54
5	LS	Students with Disabilities	2720	670.80	668	40.44	50

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
5	RW	Female	5993	670.23	675	44.24	55
5	RW	Male	7152	664.42	667	44.44	56
5	RW	American Indian or Alaska Native	15	659.40	663	47.44	70
5	RW	Black or African-American	809	657.16	663	48.90	62
5	RW	Asian or Pacific Islander	2012	673.72	680	47.40	61
5	RW	Hispanic or Latino	9285	665.40	671	42.96	56
5	RW	White	1024	677.12	680	44.67	58
5	RW	NYC	9162	665.30	671	43.30	56
5	RW	Big 4 Cities	888	658.37	663	47.75	58
5	RW	High Need Urban/Suburban	1348	671.06	675	45.34	55
5	RW	High Need Rural	71	665.92	671	43.59	56
5	RW	Average Need	1107	675.80	675	46.09	58
5	RW	Low Need	534	683.44	687	45.67	58
5	RW	Charter Schools	18	684.78	687	39.49	68
5	RW	0 Years LEP	1101	640.44	641	52.28	74
5	RW	1 Years LEP	2211	640.75	634	48.80	65
5	RW	2 Years LEP	1613	665.32	667	44.44	60
5	RW	3 Years LEP	1226	675.36	675	41.26	52
5	RW	4 Years LEP	1126	679.67	680	40.51	51
5	RW	5 Years LEP	1568	680.21	682	36.69	49
5	RW	6 Years LEP or More	4140	677.86	680	33.68	41
5	RW	Arabic	303	650.63	652	45.89	67
5	RW	Bengali	202	680.88	680	44.04	54
5	RW	Chinese	199	661.72	667	52.04	67
5	RW	English	697	678.79	680	41.65	54
5	RW	Spanish	8681	664.49	667	42.98	56
5	RW	Other Language	2677	672.38	675	47.23	55
5	RW	Students with Disabilities	2720	655.17	659	37.16	49

G.7: Grade 6

Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
6	LS	Female	5197	669.39	674	58.72	69
6	LS	Male	6258	668.25	674	58.24	69
6	LS	American Indian or Alaska Native	23	676.30	689	44.41	51
6	LS	Black or African-American	689	663.32	668	53.90	62
6	LS	Asian or Pacific Islander	1729	661.87	668	59.77	69
6	LS	Hispanic or Latino	8135	669.40	681	58.67	69
6	LS	White	870	681.01	681	55.05	61
6	LS	Multi-Racial (not Hispanic origin)	9	637.67	640	77.39	97
6	LS	NYC	8137	666.50	674	57.67	62
6	LS	Big 4 Cities	721	671.27	681	54.87	69
6	LS	High Need Urban/Suburban	1163	668.96	681	65.94	76
6	LS	High Need Rural	86	676.85	674	54.87	61
6	LS	Average Need	879	677.45	681	57.19	61
6	LS	Low Need	447	685.72	689	56.86	72
6	LS	Charter Schools	18	700.06	698	37.47	51
6	LS	0 Years LEP	1274	630.77	638	70.01	105
6	LS	1 Years LEP	2053	624.98	626	61.94	89
6	LS	2 Years LEP	1497	660.63	663	51.47	60
6	LS	3 Years LEP	1095	679.15	681	44.86	61
6	LS	4 Years LEP	920	691.16	689	41.47	46
6	LS	5 Years LEP	833	695.58	698	40.61	41
6	LS	6 Years LEP or More	3610	694.96	689	39.88	41
6	LS	Arabic	280	656.96	658	54.63	70
6	LS	Bengali	188	668.47	674	53.82	65
6	LS	Chinese	214	642.08	648	68.2	85
6	LS	English	551	681.49	689	53.71	51
6	LS	Spanish	7615	668.53	674	58.70	69
6	LS	Other Language	2243	669.25	674	57.44	69
6	LS	Students with Disabilities	2057	677.73	681	44.74	45

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
6	RW	Female	5197	672.89	680	45.81	61
6	RW	Male	6258	665.56	671	45.69	66
6	RW	American Indian or Alaska Native	23	665.22	689	49.81	76
6	RW	Black or African-American	689	663.04	667	46.30	63
6	RW	Asian or Pacific Islander	1729	672.98	680	48.13	65
6	RW	Hispanic or Latino	8135	667.45	671	44.76	62
6	RW	White	870	679.18	680	48.71	65
6	RW	Multi-Racial (not Hispanic origin)	9	637.56	627	84.32	79
6	RW	NYC	8137	666.20	671	44.50	62
6	RW	Big 4 Cities	721	667.58	671	47.58	62
6	RW	High Need Urban/Suburban	1163	671.65	680	47.93	68
6	RW	High Need Rural	86	674.22	680	51.31	61
6	RW	Average Need	879	679.16	684	47.52	65
6	RW	Low Need	447	690.62	694	49.55	62
6	RW	Charter Schools	18	692.39	694	31.38	26
6	RW	0 Years LEP	1274	644.36	645	49.75	70
6	RW	1 Years LEP	2053	641.73	634	49.66	69
6	RW	2 Years LEP	1497	663.60	663	45.75	63
6	RW	3 Years LEP	1095	676.02	675	43.33	58
6	RW	4 Years LEP	920	682.13	684	37.03	51
6	RW	5 Years LEP	833	688.20	689	37.25	46
6	RW	6 Years LEP or More	3610	684.57	684	33.76	43
6	RW	Arabic	280	649.66	648	46.54	71
6	RW	Bengali	188	671.56	675	42.17	55
6	RW	Chinese	214	661.91	661	53.54	86
6	RW	English	551	679.02	680	43.59	51
6	RW	Spanish	7615	666.89	671	44.63	62
6	RW	Other Language	2243	675.31	680	48.19	65
6	RW	Students with Disabilities	2057	661.85	663	39.19	51

G.8: Grade 7

Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
7	LS	Female	4963	683.31	691	56.68	68
7	LS	Male	5895	682.83	691	56.23	68
7	LS	American Indian or Alaska Native	18	666.61	669	57.92	62
7	LS	Black or African-American	656	681.88	686	47.98	62
7	LS	Asian or Pacific Islander	1781	677.75	681	58.50	77
7	LS	Hispanic or Latino	7573	682.70	691	56.68	65
7	LS	White	824	699.31	702	52.85	57
7	LS	Multi-Racial (not Hispanic origin)	6	643.33	644	47.52	64
7	LS	NYC	7916	682.42	691	56.33	68
7	LS	Big 4 Cities	674	680.57	691	52.79	65
7	LS	High Need Urban/Suburban	949	679.17	686	59.66	77
7	LS	High Need Rural	80	693.80	696	68.74	63
7	LS	Average Need	845	689.73	696	55.47	71
7	LS	Low Need	379	692.60	696	54.40	64
7	LS	Charter Schools	13	691.54	696	38.00	48
7	LS	0 Years LEP	1286	653.44	658	63.94	101
7	LS	1 Years LEP	2154	643.72	640	57.24	80
7	LS	2 Years LEP	1557	674.65	673	52.43	69
7	LS	3 Years LEP	1108	694.33	691	47.06	61
7	LS	4 Years LEP	953	703.85	702	42.95	49
7	LS	5 Years LEP	759	710.91	709	40.94	40
7	LS	6 Years LEP or More	2910	711.28	709	36.99	46
7	LS	Arabic	273	673.68	681	55.67	77
7	LS	Bengali	178	689.67	696	57.25	71
7	LS	Chinese	198	659.88	658	60.07	84
7	LS	English	513	697.18	702	49.99	53
7	LS	Spanish	7096	681.87	691	56.53	68
7	LS	Other Language	2276	685.69	691	55.75	74
7	LS	Students with Disabilities	1641	692.66	691	43.32	48

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
7	RW	Female	4963	659.32	662	45.10	64
7	RW	Male	5895	653.04	655	44.52	63
7	RW	American Indian or Alaska Native	18	636.56	639	41.18	61
7	RW	Black or African-American	656	651.95	655	41.05	58
7	RW	Asian or Pacific Islander	1781	657.55	658	46.14	67
7	RW	Hispanic or Latino	7573	654.47	655	44.47	63
7	RW	White	824	669.48	670	46.51	61
7	RW	Multi-Racial (not Hispanic origin)	6	614.67	624	34.48	52
7	RW	NYC	7916	654.82	655	44.03	63
7	RW	Big 4 Cities	674	649.68	655	46.46	65
7	RW	High Need Urban/Suburban	949	653.17	658	45.58	67
7	RW	High Need Rural	80	662.80	666	59.38	55
7	RW	Average Need	845	664.57	666	45.81	63
7	RW	Low Need	379	675.58	682	45.53	64
7	RW	Charter Schools	13	663.08	655	42.33	34
7	RW	0 Years LEP	1286	636.00	634	46.38	64
7	RW	1 Years LEP	2154	631.38	624	45.15	64
7	RW	2 Years LEP	1557	650.72	648	46.11	65
7	RW	3 Years LEP	1108	663.17	666	43.94	61
7	RW	4 Years LEP	953	671.03	674	39.04	49
7	RW	5 Years LEP	759	675.45	678	37.96	47
7	RW	6 Years LEP or More	2910	672.60	674	33.21	40
7	RW	Arabic	273	639.89	641	47.11	69
7	RW	Bengali	178	665.27	666	44.62	67
7	RW	Chinese	198	646.58	637	48.91	73
7	RW	English	513	668.02	674	40.62	53
7	RW	Spanish	7096	653.65	655	44.19	63
7	RW	Other Language	2276	661.53	662	45.24	61
7	RW	Students with Disabilities	1641	649.86	651	37.16	44

G.9: Grade 8
Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
8	LS	Female	5267	688.69	696	58.56	74
8	LS	Male	6264	689.22	696	58.10	74
8	LS	American Indian or Alaska Native	10	697.40	713	86.25	133
8	LS	Black or African-American	722	695.69	702	49.75	61
8	LS	Asian or Pacific Islander	1873	686.13	686	58.92	80
8	LS	Hispanic or Latino	8055	686.96	696	58.80	77
8	LS	White	859	708.88	717	54.63	60
8	LS	Multi-Racial (not Hispanic origin)	12	656.58	652	36.61	45
8	LS	NYC	8751	687.72	696	58.74	74
8	LS	Big 4 Cities	563	693.17	702	53.60	64
8	LS	High Need Urban/Suburban	944	688.19	696	59.22	77
8	LS	High Need Rural	66	685.27	696	49.82	59
8	LS	Average Need	832	696.14	702	55.96	64
8	LS	Low Need	365	700.02	702	56.95	72
8	LS	Charter Schools	10	653.40	639	44.22	53
8	LS	0 Years LEP	1221	659.53	665	68.81	113
8	LS	1 Years LEP	2144	648.47	646	56.73	77
8	LS	2 Years LEP	1750	672.83	673	52.65	72
8	LS	3 Years LEP	1247	693.52	691	49.26	64
8	LS	4 Years LEP	1046	708.30	709	47.80	56
8	LS	5 Years LEP	862	717.04	717	41.72	46
8	LS	6 Years LEP or More	3136	720.73	717	39.76	41
8	LS	Arabic	268	679.21	681	57.24	76
8	LS	Bengali	192	698.95	702	53.88	75
8	LS	Chinese	183	674.42	673	56.27	80
8	LS	English	487	705.66	709	53.34	64
8	LS	Spanish	7579	686.03	696	58.99	77
8	LS	Other Language	2476	695.26	702	56.85	79
8	LS	Students with Disabilities	1464	699.92	702	45.13	49

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
8	RW	Female	5267	666.16	670	47.05	67
8	RW	Male	6264	661.90	666	46.12	70
8	RW	American Indian or Alaska Native	10	672.90	673	46.76	64
8	RW	Black or African-American	722	665.93	670	42.20	60
8	RW	Asian or Pacific Islander	1873	667.99	670	48.40	72
8	RW	Hispanic or Latino	8055	660.86	662	46.02	64
8	RW	White	859	681.29	687	46.94	64
8	RW	Multi-Racial (not Hispanic origin)	12	639.50	646	34.84	42
8	RW	NYC	8751	662.09	666	45.57	67
8	RW	Big 4 Cities	563	664.33	670	49.94	72
8	RW	High Need Urban/Suburban	944	662.90	666	47.83	70
8	RW	High Need Rural	66	657.85	668	51.74	71
8	RW	Average Need	832	675.89	678	47.52	64
8	RW	Low Need	365	681.56	691	51.36	71
8	RW	Charter Schools	10	648.30	651	43.28	82
8	RW	0 Years LEP	1221	643.96	641	49.10	77
8	RW	1 Years LEP	2144	638.09	630	45.79	61
8	RW	2 Years LEP	1750	654.16	651	47.16	67
8	RW	3 Years LEP	1247	665.72	666	44.43	63
8	RW	4 Years LEP	1046	676.54	678	41.80	51
8	RW	5 Years LEP	862	683.19	687	40.46	50
8	RW	6 Years LEP or More	3136	684.10	687	34.98	46
8	RW	Arabic	268	652.84	648	46.30	58
8	RW	Bengali	192	671.49	672	41.68	61
8	RW	Chinese	183	665.47	662	50.91	67
8	RW	English	487	675.73	678	49.87	64
8	RW	Spanish	7579	660.11	662	45.69	64
8	RW	Other Language	2476	672.09	678	46.55	61
8	RW	Students with Disabilities	1464	657.16	658	41.75	53

G.10: Grade 9
Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
9	LS	Female	5866	679.16	684	54.46	75
9	LS	Male	7131	678.15	681	56.00	78
9	LS	American Indian or Alaska Native	25	684.04	677	69.70	72
9	LS	Black or African-American	1108	679.12	683	52.19	74
9	LS	Asian or Pacific Islander	2646	673.09	674	51.79	73
9	LS	Hispanic or Latino	8103	678.23	684	56.72	78
9	LS	White	1057	695.65	701	52.39	71
9	LS	Multi-Racial (not Hispanic origin)	58	659.55	648	52.29	90
9	LS	NYC	9548	677.70	681	55.45	78
9	LS	Big 4 Cities	583	692.77	701	51.46	67
9	LS	High Need Urban/Suburban	1107	670.29	677	57.62	81
9	LS	High Need Rural	77	691.30	692	46.87	42
9	LS	Average Need	1125	682.22	684	54.60	69
9	LS	Low Need	524	685.37	688	50.95	63
9	LS	Charter Schools	7	708.14	717	43.09	91
9	LS	0 Years LEP	1547	651.63	648	60.19	91
9	LS	1 Years LEP	3500	647.71	648	48.80	65
9	LS	2 Years LEP	1957	671.33	670	48.70	67
9	LS	3 Years LEP	1350	684.67	684	44.96	63
9	LS	4 Years LEP	1047	698.60	701	41.90	50
9	LS	5 Years LEP	825	711.27	712	40.39	44
9	LS	6 Years LEP or More	2574	720.83	724	39.23	40
9	LS	Arabic	340	672.73	670	50.20	75
9	LS	Bengali	270	675.73	677	47.94	70
9	LS	Chinese	436	650.32	651	52.42	68
9	LS	English	637	695.52	701	55.14	68
9	LS	Spanish	7574	677.79	684	56.64	80
9	LS	Other Language	3326	681.15	684	52.44	72
9	LS	Students with Disabilities	196	697.05	701	50.20	51

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
9	RW	Female	5866	686.19	685	41.56	60
9	RW	Male	7131	680.81	679	41.98	61
9	RW	American Indian or Alaska Native	25	687.32	692	43.14	64
9	RW	Black or African-American	1108	683.21	682	41.24	63
9	RW	Asian or Pacific Islander	2646	686.33	685	42.74	64
9	RW	Hispanic or Latino	8103	680.64	679	41.10	61
9	RW	White	1057	696.00	702	43.19	65
9	RW	Multi-Racial (not Hispanic origin)	58	673.00	664	45.48	59
9	RW	NYC	9548	682.99	682	41.36	63
9	RW	Big 4 Cities	583	682.61	685	42.61	69
9	RW	High Need Urban/Suburban	1107	676.03	673	43.07	63
9	RW	High Need Rural	77	684.82	689	39.60	63
9	RW	Average Need	1125	686.50	685	42.80	64
9	RW	Low Need	524	695.52	695	43.22	65
9	RW	Charter Schools	7	711.00	729	50.06	91
9	RW	0 Years LEP	1547	667.57	661	42.33	63
9	RW	1 Years LEP	3500	665.34	658	38.83	53
9	RW	2 Years LEP	1957	678.95	676	41.62	57
9	RW	3 Years LEP	1350	686.13	685	39.03	57
9	RW	4 Years LEP	1047	693.65	695	37.78	52
9	RW	5 Years LEP	825	704.65	710	36.90	47
9	RW	6 Years LEP or More	2574	707.80	710	32.31	40
9	RW	Arabic	340	671.73	664	40.62	63
9	RW	Bengali	270	681.47	685	39.48	61
9	RW	Chinese	436	674.32	670	42.46	59
9	RW	English	637	692.24	695	43.07	66
9	RW	Spanish	7574	680.51	679	41.14	61
9	RW	Other Language	3326	689.75	689	42.09	61
9	RW	Students with Disabilities	196	679.61	684	39.95	61

G.11: Grade 10
Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
10	LS	Female	5509	686.10	688	50.24	63
10	LS	Male	5954	688.08	688	49.40	70
10	LS	American Indian or Alaska Native	19	695.37	688	45.44	27
10	LS	Black or African-American	1019	693.53	692	45.56	60
10	LS	Asian or Pacific Islander	2942	678.43	677	46.21	64
10	LS	Hispanic or Latino	6626	688.10	692	51.78	70
10	LS	White	817	703.39	701	44.56	58
10	LS	Multi-Racial (not Hispanic origin)	40	667.18	677	55.33	69
10	LS	NYC	8751	685.82	688	50.09	73
10	LS	Big 4 Cities	407	688.91	688	48.79	67
10	LS	High Need Urban/Suburban	933	685.86	688	51.45	70
10	LS	High Need Rural	55	692.60	697	49.41	54
10	LS	Average Need	831	695.09	697	46.60	60
10	LS	Low Need	474	698.03	701	44.88	54
10	LS	Charter Schools	12	676.00	674	51.81	70
10	LS	0 Years LEP	1239	665.34	664	58.53	81
10	LS	1 Years LEP	2625	664.24	664	45.50	55
10	LS	2 Years LEP	2265	679.09	677	43.52	55
10	LS	3 Years LEP	1342	690.71	692	42.70	53
10	LS	4 Years LEP	1011	699.71	701	43.73	58
10	LS	5 Years LEP	767	709.13	712	38.52	48
10	LS	6 Years LEP or More	2067	721.27	724	40.95	40
10	LS	Arabic	205	692.89	697	42.72	47
10	LS	Bengali	266	687.52	688	42.87	60
10	LS	Chinese	520	667.15	667	41.04	55
10	LS	English	519	697.42	697	47.61	57
10	LS	Spanish	6206	687.74	692	52.12	73
10	LS	Other Language	3414	686.15	684	47.22	63
10	LS	Students with Disabilities	135	699.05	712	52.30	65

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
10	RW	Female	5509	694.64	695	39.81	57
10	RW	Male	5954	692.13	694	39.99	55
10	RW	American Indian or Alaska Native	19	699.00	692	27.64	37
10	RW	Black or African-American	1019	696.31	695	38.85	54
10	RW	Asian or Pacific Islander	2942	694.30	695	41.37	57
10	RW	Hispanic or Latino	6626	690.75	692	38.69	55
10	RW	White	817	707.39	710	42.58	56
10	RW	Multi-Racial (not Hispanic origin)	40	683.70	682	40.91	72
10	RW	NYC	8751	693.11	695	39.35	55
10	RW	Big 4 Cities	407	681.48	679	43.45	63
10	RW	High Need Urban/Suburban	933	690.42	689	41.30	58
10	RW	High Need Rural	55	694.91	699	42.45	51
10	RW	Average Need	831	698.51	699	40.06	62
10	RW	Low Need	474	704.45	706	40.06	53
10	RW	Charter Schools	12	678.58	673	48.62	91
10	RW	0 Years LEP	1239	680.86	676	42.37	61
10	RW	1 Years LEP	2625	681.23	676	41.19	54
10	RW	2 Years LEP	2265	689.48	689	39.38	54
10	RW	3 Years LEP	1342	694.48	695	37.42	52
10	RW	4 Years LEP	1011	699.55	702	35.39	48
10	RW	5 Years LEP	767	704.68	706	35.76	47
10	RW	6 Years LEP or More	2067	712.34	715	32.64	43
10	RW	Arabic	205	690.73	692	35.72	48
10	RW	Bengali	266	693.49	695	37.99	52
10	RW	Chinese	520	692.82	692	41.15	55
10	RW	English	519	695.18	695	42.41	60
10	RW	Spanish	6206	690.67	692	38.70	58
10	RW	Other Language	3414	697.66	699	41.21	54
10	RW	Students with Disabilities	135	682.30	689	42.12	69

G.12: Grade 11
Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
11	LS	Female	3476	699.15	701	44.89	62
11	LS	Male	3378	701.16	701	44.92	58
11	LS	American Indian or Alaska Native	14	710.57	706	49.59	23
11	LS	Black or African-American	617	703.65	706	40.37	55
11	LS	Asian or Pacific Islander	1870	696.41	692	41.90	57
11	LS	Hispanic or Latino	3809	699.55	701	47.00	62
11	LS	White	509	714.89	712	41.74	53
11	LS	Multi-Racial (not Hispanic origin)	35	683.94	692	40.14	63
11	LS	NYC	4792	699.56	701	45.34	62
11	LS	Big 4 Cities	226	698.00	697	43.67	54
11	LS	High Need Urban/Suburban	634	697.09	697	46.07	54
11	LS	High Need Rural	40	696.08	701	43.06	68
11	LS	Average Need	747	705.90	706	43.19	55
11	LS	Low Need	411	702.65	701	41.33	43
11	LS	Charter Schools	4	702.50	706	25.48	43
11	LS	0 Years LEP	536	690.06	692	50.09	63
11	LS	1 Years LEP	1207	681.13	681	46.34	61
11	LS	2 Years LEP	1361	692.18	692	41.23	50
11	LS	3 Years LEP	1128	698.52	697	41.25	54
11	LS	4 Years LEP	719	708.45	706	40.51	51
11	LS	5 Years LEP	532	713.51	717	40.09	53
11	LS	6 Years LEP or More	1262	722.37	724	39.91	51
11	LS	Arabic	130	705.62	706	38.63	55
11	LS	Bengali	155	698.50	697	43.45	57
11	LS	Chinese	338	691.07	688	37.38	50
11	LS	English	391	709.47	712	44.51	48
11	LS	Spanish	3474	699.29	701	46.91	62
11	LS	Other Language	2104	701.08	701	42.85	58
11	LS	Students with Disabilities	74	710.62	712	41.41	44

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
11	RW	Female	3476	706.08	706	37.63	47
11	RW	Male	3378	704.55	706	37.88	50
11	RW	American Indian or Alaska Native	14	703.14	705	25.53	35
11	RW	Black or African-American	617	703.34	702	37.55	50
11	RW	Asian or Pacific Islander	1870	710.13	710	38.56	50
11	RW	Hispanic or Latino	3809	701.82	702	36.66	53
11	RW	White	509	716.81	719	39.04	52
11	RW	Multi-Racial (not Hispanic origin)	35	698.71	706	42.39	68
11	RW	NYC	4792	705.38	706	37.50	50
11	RW	Big 4 Cities	226	688.06	692	37.28	51
11	RW	High Need Urban/Suburban	634	701.39	699	37.81	48
11	RW	High Need Rural	40	696.65	692	33.52	52
11	RW	Average Need	747	710.64	710	37.72	50
11	RW	Low Need	411	711.38	715	37.89	46
11	RW	Charter Schools	4	714.25	722	36.29	50
11	RW	0 Years LEP	536	696.54	699	38.38	54
11	RW	1 Years LEP	1207	694.95	695	40.12	57
11	RW	2 Years LEP	1361	703.32	702	38.98	53
11	RW	3 Years LEP	1128	704.86	702	37.24	50
11	RW	4 Years LEP	719	708.83	710	34.77	40
11	RW	5 Years LEP	532	714.02	715	34.63	49
11	RW	6 Years LEP or More	1262	715.94	719	32.88	36
11	RW	Arabic	130	706.24	706	34.11	47
11	RW	Bengali	155	703.32	706	36.78	47
11	RW	Chinese	338	711.07	715	35.43	46
11	RW	English	391	704.07	706	39.54	50
11	RW	Spanish	3474	701.86	702	36.59	53
11	RW	Other Language	2104	710.15	710	39.24	50
11	RW	Students with Disabilities	74	700.42	704	33.14	43

G.13: Grade 12 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
12	LS	Female	1691	705.91	706	44.56	55
12	LS	Male	1578	709.06	712	44.82	48
12	LS	American Indian or Alaska Native	6	699.67	681	49.06	42
12	LS	Black or African-American	291	708.73	712	40.72	44
12	LS	Asian or Pacific Islander	888	702.97	697	43.60	57
12	LS	Hispanic or Latino	1807	707.79	712	45.57	51
12	LS	White	264	720.00	724	44.15	55
12	LS	Multi-Racial (not Hispanic origin)	13	680.54	684	40.07	51
12	LS	NYC	2233	703.71	701	46.09	55
12	LS	Big 4 Cities	134	702.95	706	43.70	51
12	LS	High Need Urban/Suburban	270	715.49	717	39.44	49
12	LS	High Need Rural	26	714.12	715	40.22	44
12	LS	Average Need	395	717.98	717	39.62	49
12	LS	Low Need	210	718.31	717	39.75	44
12	LS	Charter Schools	1	793.00	793	–	0
12	LS	0 Years LEP	392	715.16	717	51.47	57
12	LS	1 Years LEP	302	701.77	706	43.95	58
12	LS	2 Years LEP	612	696.79	697	43.48	50
12	LS	3 Years LEP	510	704.07	701	42.50	55
12	LS	4 Years LEP	436	705.90	706	42.58	55
12	LS	5 Years LEP	310	712.58	712	42.68	53
12	LS	6 Years LEP or More	657	715.54	717	43.82	49
12	LS	Arabic	49	713.00	717	38.79	44
12	LS	Bengali	60	695.15	688	40.52	57
12	LS	Chinese	140	699.84	697	41.30	54
12	LS	English	250	717.48	717	39.87	44
12	LS	Spanish	1596	706.11	706	46.20	51
12	LS	Other Language	1048	706.81	706	44.19	51
12	LS	Students with Disabilities	72	683.19	686	61.24	67

Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
12	RW	Female	1691	709.78	710	36.25	50
12	RW	Male	1578	708.73	710	36.96	44
12	RW	American Indian or Alaska Native	6	718.00	717	42.86	63
12	RW	Black or African-American	291	708.68	706	41.92	53
12	RW	Asian or Pacific Islander	888	714.71	715	36.39	43
12	RW	Hispanic or Latino	1807	705.49	706	35.42	47
12	RW	White	264	717.95	719	35.53	46
12	RW	Multi-Racial (not Hispanic origin)	13	696.46	689	29.38	40
12	RW	NYC	2233	705.95	706	35.04	47
12	RW	Big 4 Cities	134	694.95	699	36.80	51
12	RW	High Need Urban/Suburban	270	712.20	715	37.11	49
12	RW	High Need Rural	26	704.12	706	37.33	46
12	RW	Average Need	395	720.77	719	35.15	46
12	RW	Low Need	210	728.88	729	42.33	47
12	RW	Charter Schools	1	729.00	729	–	0
12	RW	0 Years LEP	392	720.48	724	40.16	57
12	RW	1 Years LEP	302	713.95	715	41.19	52
12	RW	2 Years LEP	612	705.82	706	37.93	47
12	RW	3 Years LEP	510	711.21	710	34.18	40
12	RW	4 Years LEP	436	704.64	706	35.98	50
12	RW	5 Years LEP	310	710.64	710	33.31	40
12	RW	6 Years LEP or More	657	704.07	706	32.50	44
12	RW	Arabic	49	707.78	706	30.73	37
12	RW	Bengali	60	703.45	708	37.91	48
12	RW	Chinese	140	716.24	715	29.01	36
12	RW	English	250	713.00	715	38.55	46
12	RW	Spanish	1596	704.15	706	35.15	47
12	RW	Other Language	1048	714.04	715	37.94	46
12	RW	Students with Disabilities	72	667.19	669	48.13	79

APPENDIX H: PROFICIENCY PERCENTAGES

H.1: Kindergarten **Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
K	LS	Female	11543	6.30	29.56	38.09	26.05
K	LS	Male	12696	6.78	32.24	39.19	21.79
K	LS	American Indian or Alaska Native	41	9.76	36.59	41.46	12.20
K	LS	Black or African-American	858	10.02	30.89	38.81	20.28
K	LS	Asian or Pacific Islander	4694	7.18	31.40	38.13	23.29
K	LS	Hispanic or Latino	16282	6.36	32.28	38.57	22.79
K	LS	White	2329	5.37	20.95	40.15	33.53
K	LS	Multi-Racial (not Hispanic origin)	35	–	20.00	51.43	28.57
K	LS	NYC	15711	7.23	30.80	37.14	24.83
K	LS	Big 4 Cities	1174	7.24	33.13	39.95	19.68
K	LS	High Need Urban/Suburban	2936	6.37	35.35	39.51	18.77
K	LS	High Need Rural	161	3.73	29.81	47.20	19.25
K	LS	Average Need	2675	4.00	29.01	42.13	24.86
K	LS	Low Need	1425	4.42	26.88	44.91	23.79
K	LS	Charter Schools	136	2.94	21.32	44.12	31.62
K	LS	0 Years LEP	2611	18.84	29.76	32.21	19.19
K	LS	1 Years LEP	21239	5.01	31.02	39.44	24.54
K	LS	Arabic	499	10.22	28.06	37.07	24.65
K	LS	Bengali	639	5.95	29.42	38.81	25.82
K	LS	Chinese	843	5.69	32.38	39.98	21.95
K	LS	English	1883	4.20	29.16	42.33	24.32
K	LS	Spanish	14532	6.70	32.62	38.26	22.42
K	LS	Other Language	4866	6.80	27.66	38.16	27.37
K	LS	Students with Disabilities	2060	9.66	45.19	34.08	11.07

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
K	RW	Female	11543	42.68	37.76	11.14	8.41
K	RW	Male	12696	51.78	32.36	9.03	6.84
K	RW	American Indian or Alaska Native	41	43.90	43.90	9.76	2.44
K	RW	Black or African-American	858	58.74	29.14	6.76	5.36
K	RW	Asian or Pacific Islander	4694	29.91	35.85	16.53	17.70
K	RW	Hispanic or Latino	16282	52.69	34.45	8.17	4.69
K	RW	White	2329	42.38	38.21	11.03	8.37
K	RW	Multi-Racial (not Hispanic origin)	35	25.71	48.57	17.14	8.57
K	RW	NYC	15711	42.63	35.73	11.90	9.74
K	RW	Big 4 Cities	1174	62.95	29.90	4.09	3.07
K	RW	High Need Urban/Suburban	2936	60.73	31.57	5.59	2.11
K	RW	High Need Rural	161	50.93	44.10	4.35	0.62
K	RW	Average Need	2675	53.12	35.93	7.03	3.93
K	RW	Low Need	1425	50.67	33.26	9.75	6.32
K	RW	Charter Schools	136	38.24	44.85	9.56	7.35
K	RW	0 Years LEP	2611	60.32	29.57	6.74	3.37
K	RW	1 Years LEP	21239	45.70	35.59	10.51	8.20
K	RW	Arabic	499	50.10	33.47	8.22	8.22
K	RW	Bengali	639	29.42	37.87	15.81	16.90
K	RW	Chinese	843	25.74	36.54	16.25	21.47
K	RW	English	1883	53.69	35.00	7.49	3.82
K	RW	Spanish	14532	52.39	34.53	8.37	4.71
K	RW	Other Language	4866	34.32	36.25	14.82	14.61
K	RW	Students with Disabilities	2060	65.83	26.21	4.95	3.01

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
K	Overall	Female	11543	43.65	39.35	11.91	5.09
K	Overall	Male	12696	52.52	33.72	9.96	3.80
K	Overall	American Indian or Alaska Native	41	43.90	48.78	7.32	–
K	Overall	Black or African-American	858	59.44	29.72	7.58	3.26
K	Overall	Asian or Pacific Islander	4694	31.21	40.18	19.56	9.05
K	Overall	Hispanic or Latino	16282	53.45	35.30	8.40	2.86
K	Overall	White	2329	43.07	38.51	12.02	6.40
K	Overall	Multi-Racial (not Hispanic origin)	35	25.71	48.57	20.00	5.71
K	Overall	NYC	15711	43.72	37.57	13.17	5.54
K	Overall	Big 4 Cities	1174	63.03	30.92	4.68	1.36
K	Overall	High Need Urban/Suburban	2936	61.14	32.08	5.21	1.57
K	Overall	High Need Rural	161	50.93	44.72	4.35	–
K	Overall	Average Need	2675	53.72	36.37	6.84	3.07
K	Overall	Low Need	1425	51.02	35.09	10.74	3.16
K	Overall	Charter Schools	136	38.97	44.85	11.03	5.15
K	Overall	0 Years LEP	2611	62.05	29.45	6.20	2.30
K	Overall	1 Years LEP	21239	46.45	37.26	11.57	4.72
K	Overall	Arabic	499	50.50	34.47	9.42	5.61
K	Overall	Bengali	639	30.36	42.25	18.00	9.39
K	Overall	Chinese	843	26.45	39.62	22.54	11.39
K	Overall	English	1883	53.80	35.74	7.91	2.55
K	Overall	Spanish	14532	53.25	35.41	8.51	2.83
K	Overall	Other Language	4866	35.57	39.23	17.02	8.18
K	Overall	Students with Disabilities	2060	66.89	26.84	5.15	1.12

H.2: Grade 1**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
1	LS	Female	12125	3.32	12.86	46.23	37.59
1	LS	Male	13746	3.60	13.69	49.04	33.67
1	LS	American Indian or Alaska Native	32	–	12.50	46.88	40.63
1	LS	Black or African-American	1084	5.90	14.48	47.23	32.38
1	LS	Asian or Pacific Islander	5275	2.75	11.41	47.13	38.71
1	LS	Hispanic or Latino	17098	3.68	14.35	49.01	32.95
1	LS	White	2357	2.33	9.29	40.05	48.32
1	LS	Multi-Racial (not Hispanic origin)	25	16.00	20.00	36.00	28.00
1	LS	NYC	16913	3.88	14.11	45.89	36.12
1	LS	Big 4 Cities	1272	4.56	15.49	54.87	25.08
1	LS	High Need Urban/Suburban	3147	3.37	13.76	53.42	29.46
1	LS	High Need Rural	191	–	6.28	54.97	38.74
1	LS	Average Need	2769	1.84	10.00	49.95	38.21
1	LS	Low Need	1464	1.78	8.88	45.63	43.72
1	LS	Charter Schools	101	–	5.94	44.55	49.50
1	LS	0 Years LEP	1684	19.89	18.47	36.16	25.48
1	LS	1 Years LEP	6189	6.33	19.52	48.17	25.98
1	LS	2 Years LEP	17627	0.90	10.61	48.74	39.75
1	LS	Arabic	578	4.84	17.99	44.46	32.70
1	LS	Bengali	661	2.87	10.44	46.90	39.79
1	LS	Chinese	756	3.70	11.64	45.37	39.29
1	LS	English	1751	2.17	11.71	50.60	35.52
1	LS	Spanish	15900	3.76	14.51	48.75	32.98
1	LS	Other Language	5243	3.11	11.12	44.67	41.10
1	LS	Students with Disabilities	2682	4.36	22.63	51.30	21.70

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
1	RW	Female	12125	15.04	29.27	33.55	22.14
1	RW	Male	13746	21.23	30.16	30.23	18.38
1	RW	American Indian or Alaska Native	32	6.25	34.38	31.25	28.13
1	RW	Black or African-American	1084	28.69	30.07	26.01	15.22
1	RW	Asian or Pacific Islander	5275	7.64	19.41	37.46	35.49
1	RW	Hispanic or Latino	17098	21.85	33.15	30.07	14.93
1	RW	White	2357	11.88	27.92	34.32	25.88
1	RW	Multi-Racial (not Hispanic origin)	25	40.00	32.00	16.00	12.00
1	RW	NYC	16913	18.57	28.81	32.01	20.62
1	RW	Big 4 Cities	1272	30.35	30.42	26.89	12.34
1	RW	High Need Urban/Suburban	3147	20.15	34.51	29.17	16.17
1	RW	High Need Rural	191	10.47	38.74	35.60	15.18
1	RW	Average Need	2769	14.16	31.27	33.41	21.16
1	RW	Low Need	1464	10.59	25.89	35.18	28.35
1	RW	Charter Schools	101	13.86	26.73	34.65	24.75
1	RW	0 Years LEP	1684	37.59	29.04	21.50	11.88
1	RW	1 Years LEP	6189	24.93	32.59	27.21	15.27
1	RW	2 Years LEP	17627	14.23	28.80	34.37	22.60
1	RW	Arabic	578	21.11	30.97	30.45	17.47
1	RW	Bengali	661	9.08	21.33	42.21	27.38
1	RW	Chinese	756	7.01	16.01	33.60	43.39
1	RW	English	1751	17.53	32.50	31.47	18.50
1	RW	Spanish	15900	21.75	32.86	30.40	14.99
1	RW	Other Language	5243	11.46	22.18	34.79	31.57
1	RW	Students with Disabilities	2682	35.83	32.36	22.22	9.58

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
1	Overall	Female	12125	15.65	31.12	38.87	14.37
1	Overall	Male	13746	21.82	31.87	35.22	11.08
1	Overall	American Indian or Alaska Native	32	6.25	40.63	37.50	15.63
1	Overall	Black or African-American	1084	29.24	31.37	29.34	10.06
1	Overall	Asian or Pacific Islander	5275	8.30	22.92	48.25	20.53
1	Overall	Hispanic or Latino	17098	22.46	34.48	33.67	9.39
1	Overall	White	2357	12.22	29.23	38.99	19.56
1	Overall	Multi-Racial (not Hispanic origin)	25	44.00	32.00	16.00	8.00
1	Overall	NYC	16913	19.32	30.75	37.07	12.87
1	Overall	Big 4 Cities	1272	30.66	31.76	30.03	7.55
1	Overall	High Need Urban/Suburban	3147	20.46	35.62	34.51	9.41
1	Overall	High Need Rural	191	10.47	39.79	39.27	10.47
1	Overall	Average Need	2769	14.48	32.83	38.82	13.87
1	Overall	Low Need	1464	10.93	28.14	42.28	18.65
1	Overall	Charter Schools	101	13.86	27.72	42.57	15.84
1	Overall	0 Years LEP	1684	40.38	29.93	22.27	7.42
1	Overall	1 Years LEP	6189	26.05	34.40	31.31	8.24
1	Overall	2 Years LEP	17627	14.45	30.64	40.28	14.63
1	Overall	Arabic	578	21.97	33.04	33.39	11.59
1	Overall	Bengali	661	9.53	24.51	49.17	16.79
1	Overall	Chinese	756	7.80	20.37	48.02	23.81
1	Overall	English	1751	18.22	33.41	37.01	11.36
1	Overall	Spanish	15900	22.36	34.27	33.96	9.40
1	Overall	Other Language	5243	12.05	24.87	43.07	20.01
1	Overall	Students with Disabilities	2682	36.39	34.41	24.16	5.03

H.3: Grade 2**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
2	LS	Female	10035	2.42	9.50	50.07	38.01
2	LS	Male	11775	2.42	9.71	50.95	36.93
2	LS	American Indian or Alaska Native	31	3.23	16.13	48.39	32.26
2	LS	Black or African-American	1000	3.40	11.30	53.80	31.50
2	LS	Asian or Pacific Islander	3778	2.12	8.66	43.70	45.53
2	LS	Hispanic or Latino	15126	2.43	10.06	52.80	34.70
2	LS	White	1866	2.41	6.91	44.32	46.36
2	LS	Multi-Racial (not Hispanic origin)	9	–	–	66.67	33.33
2	LS	NYC	14680	2.76	10.16	51.07	36.01
2	LS	Big 4 Cities	1183	2.37	12.51	56.97	28.15
2	LS	High Need Urban/Suburban	2589	2.12	9.46	51.37	37.04
2	LS	High Need Rural	117	–	8.55	39.32	52.14
2	LS	Average Need	2070	1.30	6.76	46.04	45.89
2	LS	Low Need	1077	1.21	5.01	43.83	49.95
2	LS	Charter Schools	88	–	7.95	56.82	35.23
2	LS	0 Years LEP	1281	19.59	21.23	38.10	21.08
2	LS	1 Years LEP	2925	6.50	24.51	45.78	23.21
2	LS	2 Years LEP	4147	0.58	8.17	52.45	38.80
2	LS	3 Years LEP	13215	0.41	5.63	52.31	41.65
2	LS	Arabic	476	3.78	13.87	54.83	27.52
2	LS	Bengali	508	0.79	7.68	44.88	46.65
2	LS	Chinese	467	3.00	9.42	42.83	44.75
2	LS	English	1447	0.76	9.33	51.28	38.63
2	LS	Spanish	13846	2.61	10.05	52.88	34.45
2	LS	Other Language	4314	2.46	8.39	45.02	44.14
2	LS	Students with Disabilities	3123	2.02	16.65	59.24	22.09

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
2	RW	Female	10035	13.09	35.25	35.76	15.89
2	RW	Male	11775	17.50	37.25	31.90	13.35
2	RW	American Indian or Alaska Native	31	19.35	41.94	25.81	12.90
2	RW	Black or African-American	1000	22.30	39.50	27.70	10.50
2	RW	Asian or Pacific Islander	3778	8.76	25.70	39.20	26.34
2	RW	Hispanic or Latino	15126	17.12	39.40	32.30	11.19
2	RW	White	1866	11.95	31.19	37.03	19.83
2	RW	Multi-Racial (not Hispanic origin)	9	33.33	33.33	22.22	11.11
2	RW	NYC	14680	15.79	37.72	33.04	13.45
2	RW	Big 4 Cities	1183	24.01	39.22	26.88	9.89
2	RW	High Need Urban/Suburban	2589	16.84	33.84	35.46	13.87
2	RW	High Need Rural	117	15.38	26.50	41.03	17.09
2	RW	Average Need	2070	11.88	32.03	37.54	18.55
2	RW	Low Need	1077	6.31	29.53	36.95	27.21
2	RW	Charter Schools	88	5.68	38.64	40.91	14.77
2	RW	0 Years LEP	1281	40.12	32.71	20.14	7.03
2	RW	1 Years LEP	2925	28.82	36.99	23.62	10.56
2	RW	2 Years LEP	4147	14.25	35.57	34.22	15.96
2	RW	3 Years LEP	13215	10.47	36.93	37.12	15.48
2	RW	Arabic	476	21.22	42.44	26.26	10.08
2	RW	Bengali	508	9.45	22.44	44.69	23.43
2	RW	Chinese	467	8.14	24.63	39.83	27.41
2	RW	English	1447	15.13	36.63	34.14	14.10
2	RW	Spanish	13846	17.24	39.47	32.19	11.10
2	RW	Other Language	4314	11.50	29.18	36.93	22.39
2	RW	Students with Disabilities	3123	33.37	43.13	19.12	4.39

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
2	Overall	Female	10035	13.39	35.53	38.67	12.41
2	Overall	Male	11775	17.78	37.41	34.03	10.78
2	Overall	American Indian or Alaska Native	31	19.35	41.94	29.03	9.68
2	Overall	Black or African-American	1000	22.90	39.30	29.70	8.10
2	Overall	Asian or Pacific Islander	3778	8.84	26.42	43.86	20.88
2	Overall	Hispanic or Latino	15126	17.43	39.57	34.19	8.81
2	Overall	White	1866	12.27	30.98	40.30	16.45
2	Overall	Multi-Racial (not Hispanic origin)	9	33.33	33.33	22.22	11.11
2	Overall	NYC	14680	16.19	37.85	35.27	10.69
2	Overall	Big 4 Cities	1183	24.01	39.48	30.68	5.83
2	Overall	High Need Urban/Suburban	2589	16.88	34.30	37.89	10.93
2	Overall	High Need Rural	117	15.38	26.50	42.74	15.38
2	Overall	Average Need	2070	12.03	32.17	39.76	16.04
2	Overall	Low Need	1077	6.31	30.36	42.06	21.26
2	Overall	Charter Schools	88	5.68	38.64	44.32	11.36
2	Overall	0 Years LEP	1281	42.47	31.69	20.37	5.46
2	Overall	1 Years LEP	2925	29.44	37.20	25.23	8.14
2	Overall	2 Years LEP	4147	14.30	36.03	37.06	12.61
2	Overall	3 Years LEP	13215	10.56	37.19	39.89	12.36
2	Overall	Arabic	476	21.64	42.02	29.41	6.93
2	Overall	Bengali	508	9.45	23.23	48.43	18.90
2	Overall	Chinese	467	8.14	25.48	43.04	23.34
2	Overall	English	1447	15.20	37.25	36.28	11.26
2	Overall	Spanish	13846	17.59	39.59	34.12	8.70
2	Overall	Other Language	4314	11.78	29.37	41.08	17.78
2	Overall	Students with Disabilities	3123	33.53	43.42	19.69	3.36

H.4: Grade 3**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
3	LS	Female	8794	3.59	7.58	47.03	41.79
3	LS	Male	10475	3.40	8.36	46.58	41.66
3	LS	American Indian or Alaska Native	36	–	11.11	47.22	41.67
3	LS	Black or African-American	927	3.56	12.19	48.00	36.25
3	LS	Asian or Pacific Islander	3014	4.31	9.12	40.28	46.28
3	LS	Hispanic or Latino	13732	3.31	7.54	49.20	39.95
3	LS	White	1555	3.41	7.33	37.43	51.83
3	LS	Multi-Racial (not Hispanic origin)	5	40.00	20.00	20.00	20.00
3	LS	NYC	13395	3.72	7.98	47.64	40.66
3	LS	Big 4 Cities	1000	4.90	10.90	49.10	35.10
3	LS	High Need Urban/Suburban	2198	2.78	8.33	47.95	40.95
3	LS	High Need Rural	107	0.93	7.48	45.79	45.79
3	LS	Average Need	1667	2.28	6.66	41.69	49.37
3	LS	Low Need	837	2.99	7.41	37.87	51.73
3	LS	Charter Schools	59	–	1.69	44.07	54.24
3	LS	0 Years LEP	1243	23.17	14.96	37.73	24.14
3	LS	1 Years LEP	2350	12.89	29.28	40.30	17.53
3	LS	2 Years LEP	2177	1.10	10.01	50.67	38.22
3	LS	3 Years LEP	3018	0.43	4.57	48.67	46.32
3	LS	4 Years LEP	10293	0.33	2.86	48.11	48.70
3	LS	Arabic	478	7.53	14.02	42.05	36.40
3	LS	Bengali	369	2.98	7.32	41.19	48.51
3	LS	Chinese	336	4.17	11.01	37.50	47.32
3	LS	English	1131	1.77	6.28	49.78	42.18
3	LS	Spanish	12683	3.43	7.64	49.25	39.68
3	LS	Other Language	3652	3.72	8.84	40.55	46.88
3	LS	Students with Disabilities	3314	1.96	10.29	59.02	28.73

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
3	RW	Female	8794	8.72	20.40	43.21	27.67
3	RW	Male	10475	12.16	23.47	41.27	23.09
3	RW	American Indian or Alaska Native	36	8.33	30.56	44.44	16.67
3	RW	Black or African-American	927	14.67	27.08	40.56	17.69
3	RW	Asian or Pacific Islander	3014	8.13	16.16	39.18	36.53
3	RW	Hispanic or Latino	13732	10.92	23.56	43.22	22.31
3	RW	White	1555	9.97	17.23	39.55	33.25
3	RW	Multi-Racial (not Hispanic origin)	5	60.00	20.00	–	20.00
3	RW	NYC	13395	10.94	22.48	42.53	24.05
3	RW	Big 4 Cities	1000	16.80	26.60	40.90	15.70
3	RW	High Need Urban/Suburban	2198	10.05	19.79	42.31	27.84
3	RW	High Need Rural	107	8.41	29.91	35.51	26.17
3	RW	Average Need	1667	7.62	20.64	40.61	31.13
3	RW	Low Need	837	5.97	17.44	41.58	35.01
3	RW	Charter Schools	59	1.69	32.20	40.68	25.42
3	RW	0 Years LEP	1243	34.75	23.65	29.12	12.47
3	RW	1 Years LEP	2350	27.83	30.30	27.28	14.60
3	RW	2 Years LEP	2177	10.20	23.89	38.40	27.51
3	RW	3 Years LEP	3018	6.63	20.05	45.36	27.97
3	RW	4 Years LEP	10293	4.95	20.28	47.03	27.74
3	RW	Arabic	478	20.92	24.06	35.15	19.87
3	RW	Bengali	369	6.23	16.26	35.50	42.01
3	RW	Chinese	336	8.63	13.99	37.20	40.18
3	RW	English	1131	9.37	21.57	45.36	23.70
3	RW	Spanish	12683	11.07	23.71	43.07	22.15
3	RW	Other Language	3652	8.90	17.94	40.83	32.34
3	RW	Students with Disabilities	3314	20.61	36.06	33.80	9.54

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
3	Overall	Female	8794	9.27	21.01	49.86	19.85
3	Overall	Male	10475	12.52	23.90	46.20	17.38
3	Overall	American Indian or Alaska Native	36	8.33	33.33	47.22	11.11
3	Overall	Black or African-American	927	14.78	27.51	44.23	13.48
3	Overall	Asian or Pacific Islander	3014	8.69	17.35	46.65	27.31
3	Overall	Hispanic or Latino	13732	11.37	23.91	48.73	15.98
3	Overall	White	1555	10.29	17.81	44.95	26.95
3	Overall	Multi-Racial (not Hispanic origin)	5	60.00	20.00	–	20.00
3	Overall	NYC	13395	11.39	22.96	47.97	17.68
3	Overall	Big 4 Cities	1000	17.20	27.10	45.00	10.70
3	Overall	High Need Urban/Suburban	2198	10.42	20.38	49.73	19.47
3	Overall	High Need Rural	107	8.41	30.84	41.12	19.63
3	Overall	Average Need	1667	8.10	20.82	47.33	23.76
3	Overall	Low Need	837	6.45	18.88	47.07	27.60
3	Overall	Charter Schools	59	1.69	32.20	45.76	20.34
3	Overall	0 Years LEP	1243	37.09	23.41	30.09	9.41
3	Overall	1 Years LEP	2350	29.57	31.40	29.79	9.23
3	Overall	2 Years LEP	2177	10.20	25.08	44.60	20.12
3	Overall	3 Years LEP	3018	6.69	20.51	52.05	20.74
3	Overall	4 Years LEP	10293	5.05	20.64	53.69	20.63
3	Overall	Arabic	478	21.55	24.48	38.08	15.90
3	Overall	Bengali	369	6.78	16.53	46.07	30.62
3	Overall	Chinese	336	9.52	15.18	44.35	30.95
3	Overall	English	1131	9.55	22.28	51.02	17.15
3	Overall	Spanish	12683	11.52	24.04	48.59	15.85
3	Overall	Other Language	3652	9.26	19.00	46.80	24.95
3	Overall	Students with Disabilities	3314	20.76	36.39	36.06	6.79

H.5: Grade 4**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
4	LS	Female	7337	4.73	9.58	43.90	41.79
4	LS	Male	8737	4.69	9.71	44.24	41.36
4	LS	American Indian or Alaska Native	23	–	8.70	43.48	47.83
4	LS	Black or African-American	859	4.66	14.90	47.38	33.06
4	LS	Asian or Pacific Islander	2499	5.16	11.92	40.34	42.58
4	LS	Hispanic or Latino	11402	4.79	8.87	45.54	40.81
4	LS	White	1287	3.11	8.62	36.36	51.90
4	LS	Multi-Racial (not Hispanic origin)	4	50.00	25.00	25.00	–
4	LS	NYC	11546	5.01	9.48	44.02	41.49
4	LS	Big 4 Cities	839	4.41	12.75	50.18	32.66
4	LS	High Need Urban/Suburban	1640	4.88	10.91	45.91	38.29
4	LS	High Need Rural	92	2.17	6.52	44.57	46.74
4	LS	Average Need	1244	2.97	8.04	42.77	46.22
4	LS	Low Need	665	3.31	9.47	34.89	52.33
4	LS	Charter Schools	48	–	2.08	52.08	45.83
4	LS	0 Years LEP	1214	24.79	15.90	34.51	24.79
4	LS	1 Years LEP	2258	15.94	30.03	36.89	17.14
4	LS	2 Years LEP	1719	2.33	16.64	46.31	34.73
4	LS	3 Years LEP	1572	0.64	6.36	45.67	47.33
4	LS	4 Years LEP	2203	0.32	3.50	45.17	51.02
4	LS	5 Years LEP	6972	0.44	2.98	46.76	49.81
4	LS	Arabic	353	6.52	18.41	39.66	35.41
4	LS	Bengali	278	2.88	8.63	41.37	47.12
4	LS	Chinese	305	6.23	16.39	33.44	43.93
4	LS	English	880	2.16	7.73	45.11	45.00
4	LS	Spanish	10574	4.94	9.08	45.62	40.36
4	LS	Other Language	3196	4.44	10.89	40.89	43.77
4	LS	Students with Disabilities	3155	1.65	9.00	55.59	33.76

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
4	RW	Female	7337	8.68	20.79	35.36	35.18
4	RW	Male	8737	11.51	23.94	35.64	28.90
4	RW	American Indian or Alaska Native	23	4.35	13.04	43.48	39.13
4	RW	Black or African-American	859	16.65	26.89	33.99	22.47
4	RW	Asian or Pacific Islander	2499	8.68	18.37	31.57	41.38
4	RW	Hispanic or Latino	11402	10.32	23.41	36.94	29.33
4	RW	White	1287	7.93	19.81	31.47	40.79
4	RW	Multi-Racial (not Hispanic origin)	4	75.00	–	–	25.00
4	RW	NYC	11546	10.33	22.93	35.64	31.10
4	RW	Big 4 Cities	839	16.57	27.65	34.45	21.33
4	RW	High Need Urban/Suburban	1640	10.30	21.10	34.02	34.57
4	RW	High Need Rural	92	7.61	18.48	48.91	25.00
4	RW	Average Need	1244	7.07	20.26	36.74	35.93
4	RW	Low Need	665	7.07	17.44	32.93	42.56
4	RW	Charter Schools	48	–	14.58	52.08	33.33
4	RW	0 Years LEP	1214	31.80	29.16	24.05	14.99
4	RW	1 Years LEP	2258	26.62	30.65	23.87	18.87
4	RW	2 Years LEP	1719	11.40	27.92	27.87	32.81
4	RW	3 Years LEP	1572	6.11	19.59	37.15	37.15
4	RW	4 Years LEP	2203	4.49	18.88	39.99	36.63
4	RW	5 Years LEP	6972	3.63	19.23	41.47	35.67
4	RW	Arabic	353	20.68	22.66	30.03	26.63
4	RW	Bengali	278	5.40	16.91	37.41	40.29
4	RW	Chinese	305	10.49	17.38	30.82	41.31
4	RW	English	880	8.07	20.11	36.02	35.80
4	RW	Spanish	10574	10.59	23.72	36.75	28.94
4	RW	Other Language	3196	9.01	20.74	32.01	38.24
4	RW	Students with Disabilities	3155	15.06	36.70	33.63	14.61

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
4	Overall	Female	7337	9.62	21.74	45.09	23.55
4	Overall	Male	8737	12.22	24.50	43.15	20.12
4	Overall	American Indian or Alaska Native	23	4.35	13.04	65.22	17.39
4	Overall	Black or African-American	859	17.23	28.52	39.70	14.55
4	Overall	Asian or Pacific Islander	2499	9.48	19.73	42.70	28.09
4	Overall	Hispanic or Latino	11402	11.18	23.97	45.11	19.73
4	Overall	White	1287	8.55	20.36	39.63	31.47
4	Overall	Multi-Racial (not Hispanic origin)	4	75.00	–	25.00	–
4	Overall	NYC	11546	11.28	23.52	43.90	21.30
4	Overall	Big 4 Cities	839	17.52	27.41	42.07	12.99
4	Overall	High Need Urban/Suburban	1640	10.79	22.99	44.09	22.13
4	Overall	High Need Rural	92	7.61	19.57	55.43	17.39
4	Overall	Average Need	1244	7.48	21.22	45.90	25.40
4	Overall	Low Need	665	7.22	18.65	42.26	31.88
4	Overall	Charter Schools	48	–	14.58	62.50	22.92
4	Overall	0 Years LEP	1214	34.93	27.76	27.02	10.30
4	Overall	1 Years LEP	2258	28.92	32.60	27.72	10.76
4	Overall	2 Years LEP	1719	11.98	30.08	35.78	22.16
4	Overall	3 Years LEP	1572	6.30	20.42	45.93	27.35
4	Overall	4 Years LEP	2203	4.68	19.43	49.84	26.06
4	Overall	5 Years LEP	6972	3.94	19.65	52.05	24.35
4	Overall	Arabic	353	20.68	23.80	36.54	18.98
4	Overall	Bengali	278	5.76	18.35	47.84	28.06
4	Overall	Chinese	305	11.48	20.33	38.36	29.84
4	Overall	English	880	8.30	20.91	45.80	25.00
4	Overall	Spanish	10574	11.50	24.27	44.90	19.33
4	Overall	Other Language	3196	9.89	21.87	41.30	26.94
4	Overall	Students with Disabilities	3155	15.53	36.96	37.46	10.05

H.6: Grade 5**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
5	LS	Female	5993	6.11	12.25	42.53	39.11
5	LS	Male	7152	5.66	12.64	43.58	38.12
5	LS	American Indian or Alaska Native	15	13.33	6.67	60.00	20.00
5	LS	Black or African-American	809	5.69	15.45	46.23	32.63
5	LS	Asian or Pacific Islander	2012	6.16	14.96	42.10	36.78
5	LS	Hispanic or Latino	9285	6.10	11.83	43.56	38.51
5	LS	White	1024	3.22	11.04	38.18	47.56
5	LS	NYC	9162	5.97	12.57	43.30	38.16
5	LS	Big 4 Cities	888	7.77	14.75	46.06	31.42
5	LS	High Need Urban/Suburban	1348	6.31	11.28	44.73	37.69
5	LS	High Need Rural	71	2.82	14.08	52.11	30.99
5	LS	Average Need	1107	4.79	12.83	37.22	45.17
5	LS	Low Need	534	2.62	9.55	41.57	46.25
5	LS	Charter Schools	18	5.56	–	44.44	50.00
5	LS	0 Years LEP	1101	24.34	16.53	31.52	27.61
5	LS	1 Years LEP	2211	18.18	32.20	34.55	15.06
5	LS	2 Years LEP	1613	2.98	20.09	46.62	30.32
5	LS	3 Years LEP	1226	0.98	10.28	47.23	41.52
5	LS	4 Years LEP	1126	0.09	5.77	44.05	50.09
5	LS	5 Years LEP	1568	0.45	3.25	45.98	50.32
5	LS	6 Years LEP or More	4140	0.43	4.01	46.79	48.77
5	LS	Arabic	303	7.26	21.12	42.57	29.04
5	LS	Bengali	202	2.48	14.36	46.04	37.13
5	LS	Chinese	199	12.56	23.12	33.17	31.16
5	LS	English	697	1.87	7.60	43.62	46.92
5	LS	Spanish	8681	6.36	12.18	43.87	37.60
5	LS	Other Language	2677	5.12	13.15	41.39	40.34
5	LS	Students with Disabilities	2720	1.51	11.25	53.42	33.82

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
5	RW	Female	5993	12.53	25.63	40.63	21.21
5	RW	Male	7152	15.79	27.43	39.08	17.70
5	RW	American Indian or Alaska Native	15	26.67	20.00	46.67	6.67
5	RW	Black or African-American	809	19.65	27.94	37.21	15.20
5	RW	Asian or Pacific Islander	2012	12.72	22.96	38.22	26.09
5	RW	Hispanic or Latino	9285	14.66	27.52	40.52	17.31
5	RW	White	1024	9.77	24.61	38.18	27.44
5	RW	NYC	9162	14.68	27.68	39.94	17.70
5	RW	Big 4 Cities	888	19.82	28.49	36.37	15.32
5	RW	High Need Urban/Suburban	1348	12.76	24.55	40.21	22.48
5	RW	High Need Rural	71	18.31	21.13	42.25	18.31
5	RW	Average Need	1107	11.56	22.31	40.65	25.47
5	RW	Low Need	534	8.05	19.85	39.89	32.21
5	RW	Charter Schools	18	11.11	16.67	33.33	38.89
5	RW	0 Years LEP	1101	34.24	30.06	25.16	10.54
5	RW	1 Years LEP	2211	35.59	32.07	21.21	11.13
5	RW	2 Years LEP	1613	15.62	30.38	35.03	18.97
5	RW	3 Years LEP	1226	8.81	26.59	40.86	23.74
5	RW	4 Years LEP	1126	6.57	22.47	44.85	26.11
5	RW	5 Years LEP	1568	5.29	22.83	45.73	26.15
5	RW	6 Years LEP or More	4140	4.28	24.03	51.57	20.12
5	RW	Arabic	303	27.06	30.03	30.03	12.87
5	RW	Bengali	202	6.44	26.24	40.59	26.73
5	RW	Chinese	199	21.61	24.12	34.17	20.10
5	RW	English	697	7.17	24.68	42.47	25.68
5	RW	Spanish	8681	15.21	27.65	40.26	16.89
5	RW	Other Language	2677	12.40	24.02	38.74	24.84
5	RW	Students with Disabilities	2720	16.40	38.35	36.95	8.31

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
5	Overall	Female	5993	13.38	25.96	44.39	16.27
5	Overall	Male	7152	16.44	27.66	42.48	13.42
5	Overall	American Indian or Alaska Native	15	33.33	20.00	40.00	6.67
5	Overall	Black or African-American	809	20.27	28.06	41.16	10.51
5	Overall	Asian or Pacific Islander	2012	13.47	23.56	43.84	19.14
5	Overall	Hispanic or Latino	9285	15.42	27.79	43.56	13.23
5	Overall	White	1024	10.35	24.41	42.19	23.05
5	Overall	NYC	9162	15.46	27.89	43.00	13.65
5	Overall	Big 4 Cities	888	20.27	29.17	38.96	11.60
5	Overall	High Need Urban/Suburban	1348	13.65	25.00	46.51	14.84
5	Overall	High Need Rural	71	18.31	21.13	45.07	15.49
5	Overall	Average Need	1107	12.38	22.40	43.54	21.68
5	Overall	Low Need	534	8.43	20.60	47.75	23.22
5	Overall	Charter Schools	18	11.11	16.67	44.44	27.78
5	Overall	0 Years LEP	1101	37.15	28.25	26.16	8.45
5	Overall	1 Years LEP	2211	37.13	32.25	23.25	7.37
5	Overall	2 Years LEP	1613	16.24	31.68	37.69	14.38
5	Overall	3 Years LEP	1226	9.05	27.16	45.35	18.43
5	Overall	4 Years LEP	1126	6.57	23.18	49.64	20.60
5	Overall	5 Years LEP	1568	5.74	22.70	51.28	20.28
5	Overall	6 Years LEP or More	4140	4.57	24.44	55.36	15.63
5	Overall	Arabic	303	28.05	29.37	33.33	9.24
5	Overall	Bengali	202	6.44	27.23	45.54	20.79
5	Overall	Chinese	199	23.12	23.62	39.20	14.07
5	Overall	English	697	7.60	25.25	48.21	18.94
5	Overall	Spanish	8681	15.99	27.87	43.27	12.88
5	Overall	Other Language	2677	13.19	24.43	43.11	19.28
5	Overall	Students with Disabilities	2720	16.69	38.53	38.71	6.07

H.7: Grade 6**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
6	LS	Female	5197	10.06	14.49	39.81	35.64
6	LS	Male	6258	9.75	14.91	41.91	33.43
6	LS	American Indian or Alaska Native	23	4.35	13.04	39.13	43.48
6	LS	Black or African-American	689	9.87	15.97	44.70	29.46
6	LS	Asian or Pacific Islander	1729	11.51	18.28	41.30	28.92
6	LS	Hispanic or Latino	8135	10.07	13.90	40.81	35.22
6	LS	White	870	5.06	14.25	38.97	41.72
6	LS	Multi-Racial (not Hispanic origin)	9	22.22	22.22	22.22	33.33
6	LS	NYC	8137	10.20	15.14	42.28	32.38
6	LS	Big 4 Cities	721	7.77	16.64	38.70	36.89
6	LS	High Need Urban/Suburban	1163	13.41	13.33	35.17	38.09
6	LS	High Need Rural	86	3.49	10.47	51.16	34.88
6	LS	Average Need	879	7.05	13.31	39.82	39.82
6	LS	Low Need	447	5.82	11.86	35.57	46.76
6	LS	Charter Schools	18	–	–	38.89	61.11
6	LS	0 Years LEP	1274	28.41	21.59	31.16	18.84
6	LS	1 Years LEP	2053	28.64	28.30	29.76	13.30
6	LS	2 Years LEP	1497	7.41	24.38	44.09	24.11
6	LS	3 Years LEP	1095	1.46	15.98	47.03	35.53
6	LS	4 Years LEP	920	0.76	7.07	48.26	43.91
6	LS	5 Years LEP	833	0.60	5.52	42.74	51.14
6	LS	6 Years LEP or More	3610	0.83	4.54	45.48	49.14
6	LS	Arabic	280	9.29	26.07	40.71	23.93
6	LS	Bengali	188	5.85	21.28	40.43	32.45
6	LS	Chinese	214	21.03	23.36	35.51	20.09
6	LS	English	551	5.63	11.80	40.29	42.29
6	LS	Spanish	7615	10.32	14.00	41.05	34.63
6	LS	Other Language	2243	9.14	15.29	41.51	34.06
6	LS	Students with Disabilities	2057	2.87	12.25	52.16	32.72

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
6	RW	Female	5197	14.76	27.52	37.04	20.69
6	RW	Male	6258	18.39	31.05	34.96	15.60
6	RW	American Indian or Alaska Native	23	21.74	17.39	39.13	21.74
6	RW	Black or African-American	689	20.61	30.19	34.98	14.22
6	RW	Asian or Pacific Islander	1729	15.91	27.24	34.47	22.38
6	RW	Hispanic or Latino	8135	17.03	30.08	36.66	16.24
6	RW	White	870	12.30	27.82	32.53	27.36
6	RW	Multi-Racial (not Hispanic origin)	9	44.44	11.11	22.22	22.22
6	RW	NYC	8137	17.54	30.95	35.93	15.58
6	RW	Big 4 Cities	721	17.34	29.26	35.09	18.31
6	RW	High Need Urban/Suburban	1163	18.14	23.47	37.75	20.64
6	RW	High Need Rural	86	11.63	31.40	38.37	18.60
6	RW	Average Need	879	12.51	25.94	35.04	26.51
6	RW	Low Need	447	7.83	24.16	32.66	35.35
6	RW	Charter Schools	18	–	22.22	55.56	22.22
6	RW	0 Years LEP	1274	34.85	32.97	22.76	9.42
6	RW	1 Years LEP	2053	38.97	32.68	19.29	9.06
6	RW	2 Years LEP	1497	18.64	36.94	28.66	15.76
6	RW	3 Years LEP	1095	11.60	30.96	35.71	21.74
6	RW	4 Years LEP	920	6.85	27.50	43.37	22.28
6	RW	5 Years LEP	833	4.44	23.41	44.54	27.61
6	RW	6 Years LEP or More	3610	4.24	24.88	49.00	21.88
6	RW	Arabic	280	29.64	36.43	23.57	10.36
6	RW	Bengali	188	12.23	30.32	38.83	18.62
6	RW	Chinese	214	25.23	27.10	27.10	20.56
6	RW	English	551	11.80	25.77	39.20	23.23
6	RW	Spanish	7615	17.20	30.37	36.60	15.82
6	RW	Other Language	2243	14.40	26.88	34.95	23.76
6	RW	Students with Disabilities	2057	15.70	39.52	35.34	9.43

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
6	Overall	Female	5197	16.62	27.79	40.93	14.66
6	Overall	Male	6258	19.59	30.98	38.48	10.95
6	Overall	American Indian or Alaska Native	23	21.74	17.39	47.83	13.04
6	Overall	Black or African-American	689	21.92	29.61	39.48	9.00
6	Overall	Asian or Pacific Islander	1729	17.18	28.57	39.68	14.57
6	Overall	Hispanic or Latino	8135	18.73	29.92	39.79	11.56
6	Overall	White	870	12.53	28.28	37.59	21.61
6	Overall	Multi-Racial (not Hispanic origin)	9	44.44	11.11	22.22	22.22
6	Overall	NYC	8137	19.11	31.04	39.23	10.62
6	Overall	Big 4 Cities	721	18.86	28.71	38.70	13.73
6	Overall	High Need Urban/Suburban	1163	19.78	23.22	41.62	15.39
6	Overall	High Need Rural	86	11.63	33.72	39.53	15.12
6	Overall	Average Need	879	13.65	26.05	41.07	19.23
6	Overall	Low Need	447	8.72	25.50	38.93	26.85
6	Overall	Charter Schools	18	–	22.22	61.11	16.67
6	Overall	0 Years LEP	1274	38.70	31.16	23.78	6.36
6	Overall	1 Years LEP	2053	42.18	31.42	20.51	5.89
6	Overall	2 Years LEP	1497	19.77	37.47	31.93	10.82
6	Overall	3 Years LEP	1095	12.15	32.60	40.09	15.16
6	Overall	4 Years LEP	920	7.17	28.15	49.02	15.65
6	Overall	5 Years LEP	833	4.80	24.25	49.70	21.25
6	Overall	6 Years LEP or More	3610	4.90	25.48	54.13	15.48
6	Overall	Arabic	280	29.64	36.43	26.79	7.14
6	Overall	Bengali	188	12.23	34.04	41.49	12.23
6	Overall	Chinese	214	26.64	28.50	32.24	12.62
6	Overall	English	551	12.34	27.22	43.92	16.52
6	Overall	Spanish	7615	19.02	30.15	39.66	11.18
6	Overall	Other Language	2243	15.56	27.20	40.48	16.76
6	Overall	Students with Disabilities	2057	16.33	40.01	37.34	6.32

H.8: Grade 7**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
7	LS	Female	4963	8.70	16.46	37.88	36.95
7	LS	Male	5895	8.69	16.34	38.35	36.62
7	LS	American Indian or Alaska Native	18	11.11	27.78	38.89	22.22
7	LS	Black or African-American	656	5.64	17.38	45.88	31.10
7	LS	Asian or Pacific Islander	1781	9.77	20.72	35.99	33.52
7	LS	Hispanic or Latino	7573	9.31	15.54	38.36	36.79
7	LS	White	824	3.03	13.71	34.59	48.67
7	LS	Multi-Racial (not Hispanic origin)	6	16.67	33.33	33.33	16.67
7	LS	NYC	7916	8.98	16.14	38.73	36.14
7	LS	Big 4 Cities	674	8.46	16.47	40.06	35.01
7	LS	High Need Urban/Suburban	949	10.64	18.55	33.19	37.62
7	LS	High Need Rural	80	8.75	11.25	37.50	42.50
7	LS	Average Need	845	5.56	17.87	36.21	40.36
7	LS	Low Need	379	5.54	13.98	39.05	41.42
7	LS	Charter Schools	13	–	15.38	46.15	38.46
7	LS	0 Years LEP	1286	22.94	23.09	31.10	22.86
7	LS	1 Years LEP	2154	21.87	33.43	31.48	13.23
7	LS	2 Years LEP	1557	7.06	25.31	39.31	28.32
7	LS	3 Years LEP	1108	2.08	13.81	45.76	38.36
7	LS	4 Years LEP	953	0.94	8.39	42.92	47.74
7	LS	5 Years LEP	759	0.40	4.87	40.45	54.28
7	LS	6 Years LEP or More	2910	0.65	2.82	40.62	55.91
7	LS	Arabic	273	9.52	21.61	35.90	32.97
7	LS	Bengali	178	8.43	14.61	34.83	42.13
7	LS	Chinese	198	15.66	29.80	33.33	21.21
7	LS	English	513	3.90	12.48	34.50	49.12
7	LS	Spanish	7096	9.37	15.90	38.94	35.79
7	LS	Other Language	2276	6.94	17.49	37.52	38.05
7	LS	Students with Disabilities	1641	2.68	9.93	51.68	35.71

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
7	RW	Female	4963	25.07	33.31	31.03	10.60
7	RW	Male	5895	29.43	35.18	26.99	8.40
7	RW	American Indian or Alaska Native	18	44.44	33.33	22.22	–
7	RW	Black or African-American	656	29.73	35.82	28.81	5.64
7	RW	Asian or Pacific Islander	1781	26.90	32.73	29.37	11.01
7	RW	Hispanic or Latino	7573	28.28	34.70	28.43	8.58
7	RW	White	824	18.33	33.13	31.80	16.75
7	RW	Multi-Racial (not Hispanic origin)	6	66.67	33.33	–	–
7	RW	NYC	7916	27.65	35.78	27.72	8.86
7	RW	Big 4 Cities	674	31.75	33.83	27.00	7.42
7	RW	High Need Urban/Suburban	949	31.40	29.29	31.72	7.59
7	RW	High Need Rural	80	23.75	33.75	31.25	11.25
7	RW	Average Need	845	22.72	31.24	32.78	13.25
7	RW	Low Need	379	17.15	24.01	39.58	19.26
7	RW	Charter Schools	13	15.38	53.85	7.69	23.08
7	RW	0 Years LEP	1286	45.72	31.73	17.03	5.52
7	RW	1 Years LEP	2154	53.06	27.25	14.67	5.01
7	RW	2 Years LEP	1557	33.53	34.36	22.93	9.18
7	RW	3 Years LEP	1108	22.47	34.03	31.95	11.55
7	RW	4 Years LEP	953	13.85	35.26	38.82	12.07
7	RW	5 Years LEP	759	10.54	35.31	38.47	15.68
7	RW	6 Years LEP or More	2910	7.97	40.38	40.69	10.96
7	RW	Arabic	273	40.29	33.70	20.51	5.49
7	RW	Bengali	178	21.35	33.71	29.78	15.17
7	RW	Chinese	198	41.41	26.77	21.21	10.61
7	RW	English	513	17.35	31.97	38.21	12.48
7	RW	Spanish	7096	28.83	35.09	27.80	8.27
7	RW	Other Language	2276	23.42	33.92	30.76	11.91
7	RW	Students with Disabilities	1641	24.56	48.45	23.40	3.60

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
7	Overall	Female	4963	25.41	33.55	32.12	8.93
7	Overall	Male	5895	29.96	34.93	27.94	7.18
7	Overall	American Indian or Alaska Native	18	44.44	33.33	22.22	–
7	Overall	Black or African-American	656	30.64	35.37	29.27	4.73
7	Overall	Asian or Pacific Islander	1781	27.18	33.41	30.21	9.21
7	Overall	Hispanic or Latino	7573	28.76	34.49	29.49	7.26
7	Overall	White	824	18.45	33.62	33.25	14.68
7	Overall	Multi-Racial (not Hispanic origin)	6	66.67	33.33	–	–
7	Overall	NYC	7916	28.21	35.62	28.71	7.45
7	Overall	Big 4 Cities	674	31.90	33.68	28.34	6.08
7	Overall	High Need Urban/Suburban	949	31.51	29.50	32.24	6.74
7	Overall	High Need Rural	80	23.75	35.00	31.25	10.00
7	Overall	Average Need	845	22.84	31.72	33.73	11.72
7	Overall	Low Need	379	17.41	24.80	41.69	16.09
7	Overall	Charter Schools	13	15.38	53.85	15.38	15.38
7	Overall	0 Years LEP	1286	46.81	31.18	17.96	4.04
7	Overall	1 Years LEP	2154	53.67	27.34	14.76	4.22
7	Overall	2 Years LEP	1557	33.98	34.55	23.76	7.71
7	Overall	3 Years LEP	1108	22.65	33.94	33.12	10.29
7	Overall	4 Years LEP	953	13.96	35.47	40.19	10.39
7	Overall	5 Years LEP	759	10.67	35.57	39.66	14.10
7	Overall	6 Years LEP or More	2910	8.32	40.24	42.27	9.18
7	Overall	Arabic	273	40.29	34.07	21.25	4.40
7	Overall	Bengali	178	21.91	33.71	32.58	11.80
7	Overall	Chinese	198	41.41	28.79	22.22	7.58
7	Overall	English	513	17.54	31.97	39.38	11.11
7	Overall	Spanish	7096	29.34	34.86	28.86	6.93
7	Overall	Other Language	2276	23.81	34.18	31.63	10.37
7	Overall	Students with Disabilities	1641	24.86	48.20	24.31	2.62

H.9: Grade 8**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
8	LS	Female	5267	9.70	18.64	29.41	42.24
8	LS	Male	6264	9.26	18.68	29.36	42.70
8	LS	American Indian or Alaska Native	10	10.00	20.00	10.00	60.00
8	LS	Black or African-American	722	4.57	16.76	32.55	46.12
8	LS	Asian or Pacific Islander	1873	8.65	23.60	30.11	37.64
8	LS	Hispanic or Latino	8055	10.70	18.13	29.45	41.73
8	LS	White	859	3.73	14.09	24.68	57.51
8	LS	Multi-Racial (not Hispanic origin)	12	8.33	50.00	33.33	8.33
8	LS	NYC	8751	9.90	18.90	29.49	41.71
8	LS	Big 4 Cities	563	8.53	15.28	31.08	45.12
8	LS	High Need Urban/Suburban	944	10.06	19.81	27.12	43.01
8	LS	High Need Rural	66	9.09	18.18	28.79	43.94
8	LS	Average Need	832	6.37	17.67	30.41	45.55
8	LS	Low Need	365	6.30	16.44	27.95	49.32
8	LS	Charter Schools	10	–	60.00	20.00	20.00
8	LS	0 Years LEP	1221	25.23	22.28	23.01	29.48
8	LS	1 Years LEP	2144	23.13	37.55	22.76	16.56
8	LS	2 Years LEP	1750	9.60	30.97	32.97	26.46
8	LS	3 Years LEP	1247	4.17	19.09	36.33	40.42
8	LS	4 Years LEP	1046	1.82	11.38	33.65	53.15
8	LS	5 Years LEP	862	0.70	6.50	32.13	60.67
8	LS	6 Years LEP or More	3136	1.05	3.06	29.59	66.29
8	LS	Arabic	268	8.58	30.60	29.10	31.72
8	LS	Bengali	192	3.13	21.35	30.73	44.79
8	LS	Chinese	183	11.48	24.59	33.33	30.60
8	LS	English	487	4.11	13.35	29.98	52.57
8	LS	Spanish	7579	11.07	18.42	29.25	41.26
8	LS	Other Language	2476	6.58	18.54	29.08	45.80
8	LS	Students with Disabilities	1464	3.89	11.13	37.91	47.06

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
8	RW	Female	5267	25.06	33.17	25.92	15.85
8	RW	Male	6264	27.09	35.03	24.76	13.12
8	RW	American Indian or Alaska Native	10	20.00	30.00	30.00	20.00
8	RW	Black or African-American	722	22.30	38.50	25.76	13.43
8	RW	Asian or Pacific Islander	1873	25.79	31.13	25.36	17.73
8	RW	Hispanic or Latino	8055	27.78	35.01	24.67	12.54
8	RW	White	859	15.02	29.10	30.73	25.15
8	RW	Multi-Racial (not Hispanic origin)	12	33.33	58.33	8.33	–
8	RW	NYC	8751	27.06	34.89	25.16	12.89
8	RW	Big 4 Cities	563	25.93	31.44	25.04	17.58
8	RW	High Need Urban/Suburban	944	27.75	33.26	25.21	13.77
8	RW	High Need Rural	66	30.30	36.36	19.70	13.64
8	RW	Average Need	832	17.55	33.53	26.68	22.24
8	RW	Low Need	365	19.45	24.93	26.85	28.77
8	RW	Charter Schools	10	40.00	30.00	20.00	10.00
8	RW	0 Years LEP	1221	44.88	29.81	16.87	8.44
8	RW	1 Years LEP	2144	50.98	30.22	12.50	6.30
8	RW	2 Years LEP	1750	35.77	35.26	17.03	11.94
8	RW	3 Years LEP	1247	22.69	37.13	26.54	13.63
8	RW	4 Years LEP	1046	13.67	38.15	28.20	19.98
8	RW	5 Years LEP	862	9.51	35.27	31.67	23.55
8	RW	6 Years LEP or More	3136	6.63	35.36	38.68	19.32
8	RW	Arabic	268	32.46	38.43	19.40	9.70
8	RW	Bengali	192	19.27	37.50	23.96	19.27
8	RW	Chinese	183	28.42	32.24	24.04	15.30
8	RW	English	487	19.10	33.26	25.67	21.97
8	RW	Spanish	7579	28.28	35.03	24.77	11.93
8	RW	Other Language	2476	21.53	31.34	27.83	19.31
8	RW	Students with Disabilities	1464	24.66	45.42	22.95	6.97

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
8	Overall	Female	5267	25.65	33.17	26.98	14.20
8	Overall	Male	6264	28.02	34.74	25.67	11.57
8	Overall	American Indian or Alaska Native	10	20.00	30.00	30.00	20.00
8	Overall	Black or African-American	722	22.99	38.09	26.73	12.19
8	Overall	Asian or Pacific Islander	1873	26.32	31.29	26.80	15.59
8	Overall	Hispanic or Latino	8055	28.63	34.81	25.50	11.06
8	Overall	White	859	15.72	28.87	32.13	23.28
8	Overall	Multi-Racial (not Hispanic origin)	12	33.33	58.33	8.33	–
8	Overall	NYC	8751	28.00	34.66	25.93	11.42
8	Overall	Big 4 Cities	563	26.11	31.26	28.06	14.56
8	Overall	High Need Urban/Suburban	944	28.18	33.05	25.53	13.24
8	Overall	High Need Rural	66	30.30	37.88	18.18	13.64
8	Overall	Average Need	832	17.67	34.01	28.73	19.59
8	Overall	Low Need	365	19.73	24.93	29.59	25.75
8	Overall	Charter Schools	10	40.00	30.00	20.00	10.00
8	Overall	0 Years LEP	1221	47.26	28.26	17.04	7.45
8	Overall	1 Years LEP	2144	52.10	29.85	12.50	5.55
8	Overall	2 Years LEP	1750	36.17	35.49	17.83	10.51
8	Overall	3 Years LEP	1247	22.85	37.13	27.99	12.03
8	Overall	4 Years LEP	1046	13.96	38.53	29.73	17.78
8	Overall	5 Years LEP	862	9.74	35.38	33.29	21.58
8	Overall	6 Years LEP or More	3136	7.30	35.40	40.21	17.09
8	Overall	Arabic	268	33.96	38.06	20.15	7.84
8	Overall	Bengali	192	19.79	36.98	25.00	18.23
8	Overall	Chinese	183	29.51	32.79	25.68	12.02
8	Overall	English	487	19.30	33.47	26.90	20.33
8	Overall	Spanish	7579	29.17	34.81	25.48	10.54
8	Overall	Other Language	2476	22.01	31.34	29.24	17.41
8	Overall	Students with Disabilities	1464	25.61	45.15	23.02	6.22

H.10: Grade 9**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
9	LS	Female	5866	16.37	30.31	25.62	27.70
9	LS	Male	7131	17.37	30.85	23.99	27.78
9	LS	American Indian or Alaska Native	25	20.00	40.00	12.00	28.00
9	LS	Black or African-American	1108	14.35	33.30	25.54	26.81
9	LS	Asian or Pacific Islander	2646	16.70	36.21	25.28	21.81
9	LS	Hispanic or Latino	8103	18.36	28.80	24.16	28.67
9	LS	White	1057	8.33	26.87	27.72	37.09
9	LS	Multi-Racial (not Hispanic origin)	58	29.31	39.66	13.79	17.24
9	LS	NYC	9548	17.57	31.16	23.82	27.45
9	LS	Big 4 Cities	583	9.95	24.36	29.33	36.36
9	LS	High Need Urban/Suburban	1107	21.77	29.81	24.57	23.85
9	LS	High Need Rural	77	7.79	28.57	38.96	24.68
9	LS	Average Need	1125	14.13	30.58	27.11	28.18
9	LS	Low Need	524	10.69	30.92	28.24	30.15
9	LS	Charter Schools	7	–	28.57	14.29	57.14
9	LS	0 Years LEP	1547	34.26	32.77	16.03	16.94
9	LS	1 Years LEP	3500	30.20	44.43	17.03	8.34
9	LS	2 Years LEP	1957	15.28	42.00	24.32	18.40
9	LS	3 Years LEP	1350	8.52	35.78	30.52	25.19
9	LS	4 Years LEP	1047	4.68	23.59	38.11	33.62
9	LS	5 Years LEP	825	3.15	14.67	33.21	48.97
9	LS	6 Years LEP or More	2574	3.15	7.23	29.88	59.75
9	LS	Arabic	340	15.29	39.41	25.59	19.71
9	LS	Bengali	270	14.44	37.04	27.04	21.48
9	LS	Chinese	436	27.75	43.35	18.58	10.32
9	LS	English	637	11.15	21.04	27.16	40.66
9	LS	Spanish	7574	18.63	29.01	24.12	28.24
9	LS	Other Language	3326	13.41	33.19	26.28	27.12
9	LS	Students with Disabilities	196	9.18	17.35	36.22	37.24

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
9	RW	Female	5866	14.92	44.27	23.78	17.03
9	RW	Male	7131	19.66	43.74	20.99	15.61
9	RW	American Indian or Alaska Native	25	20.00	36.00	24.00	20.00
9	RW	Black or African-American	1108	16.61	44.95	23.83	14.62
9	RW	Asian or Pacific Islander	2646	16.48	43.24	20.82	19.46
9	RW	Hispanic or Latino	8103	18.62	45.27	21.91	14.20
9	RW	White	1057	12.20	34.91	27.06	25.83
9	RW	Multi-Racial (not Hispanic origin)	58	24.14	48.28	17.24	10.34
9	RW	NYC	9548	17.33	44.65	22.18	15.84
9	RW	Big 4 Cities	583	20.41	40.31	22.47	16.81
9	RW	High Need Urban/Suburban	1107	22.76	44.26	20.33	12.65
9	RW	High Need Rural	77	19.48	36.36	33.77	10.39
9	RW	Average Need	1125	15.29	43.02	23.02	18.67
9	RW	Low Need	524	12.02	38.74	23.28	25.95
9	RW	Charter Schools	7	–	42.86	–	57.14
9	RW	0 Years LEP	1547	30.83	43.89	14.61	10.67
9	RW	1 Years LEP	3500	29.09	50.77	12.54	7.60
9	RW	2 Years LEP	1957	19.37	48.65	18.45	13.54
9	RW	3 Years LEP	1350	12.15	50.00	22.07	15.78
9	RW	4 Years LEP	1047	8.12	44.13	28.56	19.20
9	RW	5 Years LEP	825	5.94	32.48	33.58	28.00
9	RW	6 Years LEP or More	2574	2.53	31.97	36.95	28.55
9	RW	Arabic	340	26.18	45.88	17.35	10.59
9	RW	Bengali	270	17.04	46.30	23.33	13.33
9	RW	Chinese	436	24.77	47.02	13.99	14.22
9	RW	English	637	14.13	36.42	26.37	23.08
9	RW	Spanish	7574	18.55	45.71	21.51	14.23
9	RW	Other Language	3326	13.89	41.52	24.53	20.05
9	RW	Students with Disabilities	196	21.43	39.29	29.59	9.69

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
9	Overall	Female	5866	21.48	41.44	24.53	12.55
9	Overall	Male	7131	25.55	41.37	21.12	11.96
9	Overall	American Indian or Alaska Native	25	24.00	44.00	16.00	16.00
9	Overall	Black or African-American	1108	23.19	42.69	23.19	10.92
9	Overall	Asian or Pacific Islander	2646	22.56	43.01	22.03	12.40
9	Overall	Hispanic or Latino	8103	25.24	41.27	22.21	11.28
9	Overall	White	1057	14.85	36.99	27.63	20.53
9	Overall	Multi-Racial (not Hispanic origin)	58	34.48	41.38	15.52	8.62
9	Overall	NYC	9548	24.19	41.79	22.14	11.88
9	Overall	Big 4 Cities	583	21.96	40.14	23.67	14.24
9	Overall	High Need Urban/Suburban	1107	29.18	40.20	20.78	9.85
9	Overall	High Need Rural	77	22.08	37.66	32.47	7.79
9	Overall	Average Need	1125	19.91	40.98	25.16	13.96
9	Overall	Low Need	524	14.89	39.89	27.48	17.75
9	Overall	Charter Schools	7	–	42.86	14.29	42.86
9	Overall	0 Years LEP	1547	42.15	36.20	13.51	8.14
9	Overall	1 Years LEP	3500	39.14	44.83	11.46	4.57
9	Overall	2 Years LEP	1957	24.99	47.37	18.29	9.35
9	Overall	3 Years LEP	1350	15.93	50.00	22.44	11.63
9	Overall	4 Years LEP	1047	9.93	45.85	29.61	14.61
9	Overall	5 Years LEP	825	7.76	33.70	36.48	22.06
9	Overall	6 Years LEP or More	2574	5.13	32.09	39.51	23.27
9	Overall	Arabic	340	29.41	45.29	17.35	7.94
9	Overall	Bengali	270	21.85	46.67	21.48	10.00
9	Overall	Chinese	436	32.80	46.56	13.53	7.11
9	Overall	English	637	17.43	35.48	27.63	19.47
9	Overall	Spanish	7574	25.46	41.54	21.79	11.22
9	Overall	Other Language	3326	19.60	41.13	25.29	13.98
9	Overall	Students with Disabilities	196	23.98	37.24	30.10	8.67

H.11: Grade 10**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
10	LS	Female	5509	11.18	38.34	25.63	24.85
10	LS	Male	5954	10.70	36.19	26.76	26.35
10	LS	American Indian or Alaska Native	19	5.26	31.58	42.11	21.05
10	LS	Black or African-American	1019	7.16	37.00	28.26	27.58
10	LS	Asian or Pacific Islander	2942	11.83	44.80	25.32	18.05
10	LS	Hispanic or Latino	6626	11.91	34.52	26.12	27.45
10	LS	White	817	4.04	32.44	27.17	36.35
10	LS	Multi-Racial (not Hispanic origin)	40	22.50	35.00	27.50	15.00
10	LS	NYC	8751	11.59	37.65	25.64	25.12
10	LS	Big 4 Cities	407	9.58	38.57	25.06	26.78
10	LS	High Need Urban/Suburban	933	11.68	37.41	25.72	25.19
10	LS	High Need Rural	55	9.09	29.09	32.73	29.09
10	LS	Average Need	831	7.34	35.02	28.40	29.24
10	LS	Low Need	474	4.85	32.49	33.97	28.69
10	LS	Charter Schools	12	16.67	41.67	33.33	8.33
10	LS	0 Years LEP	1239	25.91	38.98	18.32	16.79
10	LS	1 Years LEP	2625	18.32	51.81	18.97	10.90
10	LS	2 Years LEP	2265	9.67	48.26	25.78	16.29
10	LS	3 Years LEP	1342	5.81	40.01	30.63	23.55
10	LS	4 Years LEP	1011	5.54	30.17	33.63	30.66
10	LS	5 Years LEP	767	2.09	24.25	32.33	41.33
10	LS	6 Years LEP or More	2067	3.10	12.24	32.12	52.54
10	LS	Arabic	205	5.37	37.56	33.17	23.90
10	LS	Bengali	266	7.52	40.98	27.82	23.68
10	LS	Chinese	520	14.42	54.04	21.54	10.00
10	LS	English	519	6.74	32.95	31.60	28.71
10	LS	Spanish	6206	12.33	34.22	25.93	27.52
10	LS	Other Language	3414	9.58	40.69	26.04	23.70
10	LS	Students with Disabilities	135	7.41	28.15	33.33	31.11

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
10	RW	Female	5509	10.47	44.60	28.03	16.90
10	RW	Male	5954	12.90	43.85	27.54	15.70
10	RW	American Indian or Alaska Native	19	5.26	52.63	26.32	15.79
10	RW	Black or African-American	1019	9.03	46.71	27.77	16.49
10	RW	Asian or Pacific Islander	2942	12.17	42.05	28.69	17.10
10	RW	Hispanic or Latino	6626	12.50	45.83	27.17	14.50
10	RW	White	817	7.10	35.74	29.38	27.78
10	RW	Multi-Racial (not Hispanic origin)	40	20.00	40.00	30.00	10.00
10	RW	NYC	8751	11.51	44.39	28.43	15.67
10	RW	Big 4 Cities	407	20.88	45.70	20.64	12.78
10	RW	High Need Urban/Suburban	933	14.04	46.30	24.76	14.90
10	RW	High Need Rural	55	12.73	40.00	30.91	16.36
10	RW	Average Need	831	9.27	42.60	27.20	20.94
10	RW	Low Need	474	7.17	39.03	28.90	24.89
10	RW	Charter Schools	12	33.33	33.33	8.33	25.00
10	RW	0 Years LEP	1239	21.07	46.81	19.85	12.27
10	RW	1 Years LEP	2625	19.01	51.01	18.90	11.09
10	RW	2 Years LEP	2265	11.48	51.57	23.18	13.77
10	RW	3 Years LEP	1342	9.24	46.42	29.81	14.53
10	RW	4 Years LEP	1011	7.62	42.24	32.25	17.90
10	RW	5 Years LEP	767	5.35	37.55	34.94	22.16
10	RW	6 Years LEP or More	2067	2.90	28.45	42.57	26.08
10	RW	Arabic	205	9.76	50.73	29.76	9.76
10	RW	Bengali	266	9.40	45.49	28.57	16.54
10	RW	Chinese	520	11.15	47.31	26.15	15.38
10	RW	English	519	13.10	41.23	26.40	19.27
10	RW	Spanish	6206	12.58	45.75	27.31	14.36
10	RW	Other Language	3414	10.37	41.27	29.12	19.24
10	RW	Students with Disabilities	135	23.70	39.26	26.67	10.37

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
10	Overall	Female	5509	15.94	46.83	26.39	10.84
10	Overall	Male	5954	17.08	46.24	26.18	10.50
10	Overall	American Indian or Alaska Native	19	10.53	47.37	36.84	5.26
10	Overall	Black or African-American	1019	12.66	49.26	26.30	11.78
10	Overall	Asian or Pacific Islander	2942	17.17	49.25	24.78	8.80
10	Overall	Hispanic or Latino	6626	17.82	45.40	26.43	10.35
10	Overall	White	817	8.32	42.47	30.48	18.73
10	Overall	Multi-Racial (not Hispanic origin)	40	25.00	45.00	22.50	7.50
10	Overall	NYC	8751	16.99	46.98	26.10	9.93
10	Overall	Big 4 Cities	407	22.85	45.95	21.13	10.07
10	Overall	High Need Urban/Suburban	933	18.11	46.20	24.22	11.47
10	Overall	High Need Rural	55	12.73	47.27	29.09	10.91
10	Overall	Average Need	831	11.55	44.28	28.76	15.40
10	Overall	Low Need	474	8.02	43.67	33.54	14.77
10	Overall	Charter Schools	12	41.67	25.00	25.00	8.33
10	Overall	0 Years LEP	1239	30.99	44.07	17.76	7.18
10	Overall	1 Years LEP	2625	26.10	52.84	15.31	5.75
10	Overall	2 Years LEP	2265	16.11	55.32	20.13	8.43
10	Overall	3 Years LEP	1342	12.15	50.37	28.17	9.31
10	Overall	4 Years LEP	1011	10.68	46.69	30.56	12.07
10	Overall	5 Years LEP	767	6.65	40.81	36.11	16.43
10	Overall	6 Years LEP or More	2067	5.37	30.38	44.90	19.35
10	Overall	Arabic	205	11.22	55.12	28.29	5.37
10	Overall	Bengali	266	13.53	49.25	28.20	9.02
10	Overall	Chinese	520	18.85	56.92	18.27	5.96
10	Overall	English	519	15.61	42.97	27.94	13.49
10	Overall	Spanish	6206	18.18	45.04	26.62	10.17
10	Overall	Other Language	3414	14.26	47.69	26.39	11.66
10	Overall	Students with Disabilities	135	24.44	39.26	28.15	8.15

H.12: Grade 11**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
11	LS	Female	3476	7.39	37.14	29.46	26.01
11	LS	Male	3378	6.25	36.77	30.67	26.32
11	LS	American Indian or Alaska Native	14	7.14	14.29	57.14	21.43
11	LS	Black or African-American	617	4.05	35.33	34.36	26.26
11	LS	Asian or Pacific Islander	1870	5.03	45.45	28.56	20.96
11	LS	Hispanic or Latino	3809	8.64	34.31	29.56	27.49
11	LS	White	509	2.55	27.90	33.60	35.95
11	LS	Multi-Racial (not Hispanic origin)	35	17.14	40.00	25.71	17.14
11	LS	NYC	4792	7.32	37.00	29.70	25.98
11	LS	Big 4 Cities	226	5.31	43.36	27.88	23.45
11	LS	High Need Urban/Suburban	634	7.57	38.49	29.97	23.97
11	LS	High Need Rural	40	5.00	40.00	27.50	27.50
11	LS	Average Need	747	4.95	34.00	29.72	31.33
11	LS	Low Need	411	4.38	35.52	36.25	23.84
11	LS	Charter Schools	4	–	50.00	50.00	–
11	LS	0 Years LEP	536	13.25	37.87	28.36	20.52
11	LS	1 Years LEP	1207	14.25	47.39	23.53	14.83
11	LS	2 Years LEP	1361	6.25	46.73	29.10	17.93
11	LS	3 Years LEP	1128	5.32	41.76	30.05	22.87
11	LS	4 Years LEP	719	2.09	35.61	32.27	30.04
11	LS	5 Years LEP	532	3.38	26.50	34.21	35.90
11	LS	6 Years LEP or More	1262	2.77	17.75	34.71	44.77
11	LS	Arabic	130	2.31	38.46	33.08	26.15
11	LS	Bengali	155	3.87	44.52	28.39	23.23
11	LS	Chinese	338	5.33	52.07	27.22	15.38
11	LS	English	391	4.35	28.90	35.04	31.71
11	LS	Spanish	3474	8.69	34.46	29.82	27.03
11	LS	Other Language	2104	4.80	39.69	29.80	25.71
11	LS	Students with Disabilities	74	2.70	28.38	35.14	33.78

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
11	RW	Female	3476	6.33	49.05	25.63	18.99
11	RW	Male	3378	6.69	49.94	26.05	17.32
11	RW	American Indian or Alaska Native	14	–	57.14	42.86	–
11	RW	Black or African-American	617	6.65	53.48	22.69	17.18
11	RW	Asian or Pacific Islander	1870	5.78	44.76	27.59	21.87
11	RW	Hispanic or Latino	3809	7.30	52.35	24.68	15.67
11	RW	White	509	2.95	40.86	30.84	25.34
11	RW	Multi-Racial (not Hispanic origin)	35	11.43	42.86	34.29	11.43
11	RW	NYC	4792	6.59	49.08	26.04	18.28
11	RW	Big 4 Cities	226	14.60	59.29	16.81	9.29
11	RW	High Need Urban/Suburban	634	5.84	55.36	22.71	16.09
11	RW	High Need Rural	40	10.00	57.50	20.00	12.50
11	RW	Average Need	747	4.82	46.99	27.71	20.48
11	RW	Low Need	411	4.87	43.80	30.17	21.17
11	RW	Charter Schools	4	–	25.00	50.00	25.00
11	RW	0 Years LEP	536	10.82	54.10	21.64	13.43
11	RW	1 Years LEP	1207	11.43	53.77	22.62	12.18
11	RW	2 Years LEP	1361	7.27	51.29	23.73	17.71
11	RW	3 Years LEP	1128	5.05	53.63	24.29	17.02
11	RW	4 Years LEP	719	4.59	49.51	26.29	19.61
11	RW	5 Years LEP	532	2.44	43.98	28.38	25.19
11	RW	6 Years LEP or More	1262	3.17	40.02	33.20	23.61
11	RW	Arabic	130	3.08	53.08	26.92	16.92
11	RW	Bengali	155	7.74	52.26	23.23	16.77
11	RW	Chinese	338	3.85	43.20	31.07	21.89
11	RW	English	391	8.70	47.57	25.06	18.67
11	RW	Spanish	3474	7.25	52.39	24.78	15.57
11	RW	Other Language	2104	5.51	45.82	27.14	21.53
11	RW	Students with Disabilities	74	8.11	54.05	28.38	9.46

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
11	Overall	Female	3476	10.59	52.50	25.98	10.93
11	Overall	Male	3378	10.12	53.64	26.02	10.21
11	Overall	American Indian or Alaska Native	14	7.14	50.00	42.86	–
11	Overall	Black or African-American	617	8.59	56.56	23.99	10.86
11	Overall	Asian or Pacific Islander	1870	8.77	54.76	26.26	10.21
11	Overall	Hispanic or Latino	3809	12.05	52.90	25.18	9.87
11	Overall	White	509	4.91	44.20	33.20	17.68
11	Overall	Multi-Racial (not Hispanic origin)	35	22.86	48.57	25.71	2.86
11	Overall	NYC	4792	10.93	52.80	25.81	10.45
11	Overall	Big 4 Cities	226	16.37	61.06	15.49	7.08
11	Overall	High Need Urban/Suburban	634	10.41	56.78	23.66	9.15
11	Overall	High Need Rural	40	12.50	55.00	25.00	7.50
11	Overall	Average Need	747	6.96	50.74	29.05	13.25
11	Overall	Low Need	411	6.33	50.12	31.87	11.68
11	Overall	Charter Schools	4	–	50.00	50.00	–
11	Overall	0 Years LEP	536	17.16	53.73	21.27	7.84
11	Overall	1 Years LEP	1207	18.64	55.01	20.30	6.05
11	Overall	2 Years LEP	1361	10.87	57.60	23.59	7.94
11	Overall	3 Years LEP	1128	8.51	57.80	24.20	9.49
11	Overall	4 Years LEP	719	5.98	55.22	26.15	12.66
11	Overall	5 Years LEP	532	5.26	47.93	31.39	15.41
11	Overall	6 Years LEP or More	1262	5.23	42.87	35.50	16.40
11	Overall	Arabic	130	4.62	59.23	23.85	12.31
11	Overall	Bengali	155	10.97	56.13	24.52	8.39
11	Overall	Chinese	338	7.69	58.88	23.67	9.76
11	Overall	English	391	9.72	50.64	26.34	13.30
11	Overall	Spanish	3474	12.12	52.94	25.16	9.79
11	Overall	Other Language	2104	8.37	52.61	27.42	11.60
11	Overall	Students with Disabilities	74	9.46	59.46	22.97	8.11

H.13: Grade 12**Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
12	LS	Female	1691	6.15	42.11	22.41	29.33
12	LS	Male	1578	5.96	37.39	25.22	31.43
12	LS	American Indian or Alaska Native	6	–	66.67	16.67	16.67
12	LS	Black or African-American	291	5.50	37.46	27.49	29.55
12	LS	Asian or Pacific Islander	888	5.63	48.99	18.81	26.58
12	LS	Hispanic or Latino	1807	6.70	36.86	25.57	30.88
12	LS	White	264	3.79	30.30	24.24	41.67
12	LS	Multi-Racial (not Hispanic origin)	13	7.69	61.54	23.08	7.69
12	LS	NYC	2233	7.39	43.22	21.81	27.59
12	LS	Big 4 Cities	134	7.46	38.06	27.61	26.87
12	LS	High Need Urban/Suburban	270	2.59	29.63	30.74	37.04
12	LS	High Need Rural	26	3.85	30.77	23.08	42.31
12	LS	Average Need	395	2.53	33.67	26.58	37.22
12	LS	Low Need	210	2.38	30.95	28.10	38.57
12	LS	Charter Schools	1	–	–	–	100.00
12	LS	0 Years LEP	392	6.63	31.89	22.19	39.29
12	LS	1 Years LEP	302	9.27	40.40	21.19	29.14
12	LS	2 Years LEP	612	8.01	50.00	21.24	20.75
12	LS	3 Years LEP	510	5.29	45.29	23.73	25.69
12	LS	4 Years LEP	436	5.05	44.04	21.79	29.13
12	LS	5 Years LEP	310	5.16	34.52	27.10	33.23
12	LS	6 Years LEP or More	657	4.57	31.05	27.40	36.99
12	LS	Arabic	49	2.04	36.73	28.57	32.65
12	LS	Bengali	60	8.33	51.67	18.33	21.67
12	LS	Chinese	140	5.71	52.14	19.29	22.86
12	LS	English	250	2.40	30.80	27.60	39.20
12	LS	Spanish	1596	7.33	37.72	25.50	29.45
12	LS	Other Language	1048	5.63	43.61	21.37	29.39
12	LS	Students with Disabilities	72	19.44	43.06	16.67	20.83

Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
12	RW	Female	1691	5.03	49.26	24.66	21.05
12	RW	Male	1578	6.21	45.75	28.77	19.26
12	RW	American Indian or Alaska Native	6	–	50.00	16.67	33.33
12	RW	Black or African-American	291	6.87	47.77	24.74	20.62
12	RW	Asian or Pacific Islander	888	3.38	45.72	27.03	23.87
12	RW	Hispanic or Latino	1807	6.75	49.92	25.90	17.43
12	RW	White	264	3.79	37.12	32.58	26.52
12	RW	Multi-Racial (not Hispanic origin)	13	7.69	53.85	30.77	7.69
12	RW	NYC	2233	6.05	51.19	25.88	16.88
12	RW	Big 4 Cities	134	12.69	50.75	28.36	8.21
12	RW	High Need Urban/Suburban	270	4.07	43.33	27.41	25.19
12	RW	High Need Rural	26	11.54	42.31	30.77	15.38
12	RW	Average Need	395	2.53	37.72	30.13	29.62
12	RW	Low Need	210	3.33	31.90	25.24	39.52
12	RW	Charter Schools	1	–	–	100.00	–
12	RW	0 Years LEP	392	4.08	36.22	27.04	32.65
12	RW	1 Years LEP	302	7.62	37.42	27.81	27.15
12	RW	2 Years LEP	612	7.03	51.31	22.88	18.79
12	RW	3 Years LEP	510	3.14	48.04	29.61	19.22
12	RW	4 Years LEP	436	7.11	51.83	24.08	16.97
12	RW	5 Years LEP	310	2.90	49.03	29.03	19.03
12	RW	6 Years LEP or More	657	6.70	51.60	27.85	13.85
12	RW	Arabic	49	4.08	51.02	28.57	16.33
12	RW	Bengali	60	8.33	50.00	26.67	15.00
12	RW	Chinese	140	0.71	43.57	34.29	21.43
12	RW	English	250	6.40	40.80	29.20	23.60
12	RW	Spanish	1596	6.89	51.44	25.38	16.29
12	RW	Other Language	1048	4.39	44.94	26.72	23.95
12	RW	Students with Disabilities	72	34.72	44.44	12.50	8.33

Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
12	Overall	Female	1691	9.28	54.76	23.00	12.95
12	Overall	Male	1578	10.01	52.15	25.92	11.91
12	Overall	American Indian or Alaska Native	6	–	66.67	16.67	16.67
12	Overall	Black or African-American	291	9.62	51.55	26.80	12.03
12	Overall	Asian or Pacific Islander	888	7.66	57.77	21.62	12.95
12	Overall	Hispanic or Latino	1807	11.12	52.85	24.96	11.07
12	Overall	White	264	6.06	45.08	28.03	20.83
12	Overall	Multi-Racial (not Hispanic origin)	13	15.38	61.54	15.38	7.69
12	Overall	NYC	2233	11.51	56.20	22.12	10.17
12	Overall	Big 4 Cities	134	14.93	54.48	25.37	5.22
12	Overall	High Need Urban/Suburban	270	4.81	48.15	30.37	16.67
12	Overall	High Need Rural	26	11.54	46.15	30.77	11.54
12	Overall	Average Need	395	3.54	47.34	30.13	18.99
12	Overall	Low Need	210	3.81	43.81	28.57	23.81
12	Overall	Charter Schools	1	–	–	100.00	–
12	Overall	0 Years LEP	392	9.18	43.37	26.79	20.66
12	Overall	1 Years LEP	302	12.25	49.01	23.18	15.56
12	Overall	2 Years LEP	612	11.44	58.33	19.61	10.62
12	Overall	3 Years LEP	510	7.84	55.88	25.29	10.98
12	Overall	4 Years LEP	436	9.40	56.42	22.25	11.93
12	Overall	5 Years LEP	310	7.42	54.19	27.42	10.97
12	Overall	6 Years LEP or More	657	10.20	52.82	27.25	9.74
12	Overall	Arabic	49	4.08	61.22	20.41	14.29
12	Overall	Bengali	60	13.33	51.67	25.00	10.00
12	Overall	Chinese	140	5.71	59.29	23.57	11.43
12	Overall	English	250	6.80	48.40	30.00	14.80
12	Overall	Spanish	1596	11.84	53.51	24.25	10.40
12	Overall	Other Language	1048	8.40	54.20	23.38	14.03
12	Overall	Students with Disabilities	72	40.28	40.28	15.28	4.17

APPENDIX I: EXIT RATE BY SUBGROUP

I.1: Kindergarten

Grade	Group	N-count	Percent
K	Female	587	5.09
K	Male	482	3.80
K	Black or African-American	28	3.26
K	Asian or Pacific Islander	425	9.05
K	Hispanic or Latino	465	2.86
K	White	149	6.40
K	Multi-Racial (not Hispanic origin)	2	5.71
K	NYC	870	5.54
K	Big 4 Cities	16	1.36
K	High Need Urban/Suburban	46	1.57
K	Average Need	82	3.07
K	Low Need	45	3.16
K	Charter Schools	7	5.15
K	0 Years LEP	60	2.30
K	1 Years LEP	1002	4.72
K	Arabic	28	5.61
K	Bengali	60	9.39
K	Chinese	96	11.39
K	English	48	2.55
K	Spanish	411	2.83
K	Other Language	398	8.18
K	Students with Disabilities	23	1.12

I.2: Grade 1

Grade	Group	N-count	Percent
1	Female	1742	14.37
1	Male	1523	11.08
1	American Indian or Alaska Native	5	15.63
1	Black or African-American	109	10.06
1	Asian or Pacific Islander	1083	20.53
1	Hispanic or Latino	1605	9.39
1	White	461	19.56
1	Multi-Racial (not Hispanic origin)	2	8.00
1	NYC	2177	12.87
1	Big 4 Cities	96	7.55
1	High Need Urban/Suburban	296	9.41
1	High Need Rural	20	10.47
1	Average Need	384	13.87
1	Low Need	273	18.65
1	Charter Schools	16	15.84
1	0 Years LEP	125	7.42
1	1 Years LEP	510	8.24
1	2 Years LEP	2578	14.63
1	Arabic	67	11.59
1	Bengali	111	16.79
1	Chinese	180	23.81
1	English	199	11.36
1	Spanish	1495	9.40
1	Other Language	1049	20.01
1	Students with Disabilities	135	5.03

I.3: Grade 2

Grade	Group	N-count	Percent
2	Female	1245	12.41
2	Male	1269	10.78
2	American Indian or Alaska Native	3	9.68
2	Black or African-American	81	8.10
2	Asian or Pacific Islander	789	20.88
2	Hispanic or Latino	1333	8.81
2	White	307	16.45
2	Multi-Racial (not Hispanic origin)	1	11.11
2	NYC	1569	10.69
2	Big 4 Cities	69	5.83
2	High Need Urban/Suburban	283	10.93
2	High Need Rural	18	15.38
2	Average Need	332	16.04
2	Low Need	229	21.26
2	Charter Schools	10	11.36
2	0 Years LEP	70	5.46
2	1 Years LEP	238	8.14
2	2 Years LEP	523	12.61
2	3 Years LEP	1634	12.36
2	Arabic	33	6.93
2	Bengali	96	18.90
2	Chinese	109	23.34
2	English	163	11.26
2	Spanish	1205	8.70
2	Other Language	767	17.78
2	Students with Disabilities	105	3.36

I.4: Grade 3

Grade	Group	N-count	Percent
3	Female	1746	19.85
3	Male	1821	17.38
3	American Indian or Alaska Native	4	11.11
3	Black or African-American	125	13.48
3	Asian or Pacific Islander	823	27.31
3	Hispanic or Latino	2195	15.98
3	White	419	26.95
3	Multi-Racial (not Hispanic origin)	1	20.00
3	NYC	2368	17.68
3	Big 4 Cities	107	10.70
3	High Need Urban/Suburban	428	19.47
3	High Need Rural	21	19.63
3	Average Need	396	23.76
3	Low Need	231	27.60
3	Charter Schools	12	20.34
3	0 Years LEP	117	9.41
3	1 Years LEP	217	9.23
3	2 Years LEP	438	20.12
3	3 Years LEP	626	20.74
3	4 Years LEP	2123	20.63
3	Arabic	76	15.90
3	Bengali	113	30.62
3	Chinese	104	30.95
3	English	194	17.15
3	Spanish	2010	15.85
3	Other Language	911	24.95
3	Students with Disabilities	225	6.79

I.5: Grade 4

Grade	Group	N-count	Percent
4	Female	1728	23.55
4	Male	1758	20.12
4	American Indian or Alaska Native	4	17.39
4	Black or African-American	125	14.55
4	Asian or Pacific Islander	702	28.09
4	Hispanic or Latino	2250	19.73
4	White	405	31.47
4	NYC	2459	21.30
4	Big 4 Cities	109	12.99
4	High Need Urban/Suburban	363	22.13
4	High Need Rural	16	17.39
4	Average Need	316	25.40
4	Low Need	212	31.88
4	Charter Schools	11	22.92
4	0 Years LEP	125	10.30
4	1 Years LEP	243	10.76
4	2 Years LEP	381	22.16
4	3 Years LEP	430	27.35
4	4 Years LEP	574	26.06
4	5 Years LEP	1698	24.35
4	Arabic	67	18.98
4	Bengali	78	28.06
4	Chinese	91	29.84
4	English	220	25.00
4	Spanish	2044	19.33
4	Other Language	861	26.94
4	Students with Disabilities	317	10.05

I.6: Grade 5

Grade	Test	Group	N-count	Percent
5	LS	Female	975	16.27
5	LS	Male	960	13.42
5	LS	American Indian or Alaska Native	1	6.67
5	LS	Black or African-American	85	10.51
5	LS	Asian or Pacific Islander	385	19.14
5	LS	Hispanic or Latino	1228	13.23
5	LS	White	236	23.05
5	LS	NYC	1251	13.65
5	LS	Big 4 Cities	103	11.60
5	LS	High Need Urban/Suburban	200	14.84
5	LS	High Need Rural	11	15.49
5	LS	Average Need	240	21.68
5	LS	Low Need	124	23.22
5	LS	Charter Schools	5	27.78
5	LS	0 Years LEP	93	8.45
5	LS	1 Years LEP	163	7.37
5	LS	2 Years LEP	232	14.38
5	LS	3 Years LEP	226	18.43
5	LS	4 Years LEP	232	20.60
5	LS	5 Years LEP	318	20.28
5	LS	6 Years LEP or More	647	15.63
5	LS	Arabic	28	9.24
5	LS	Bengali	42	20.79
5	LS	Chinese	28	14.07
5	LS	English	132	18.94
5	LS	Spanish	1118	12.88
5	LS	Other Language	516	19.28
5	LS	Students with Disabilities	165	6.07

I.7: Grade 6

Grade	Group	N-count	Percent
6	Female	762	14.66
6	Male	685	10.95
6	American Indian or Alaska Native	3	13.04
6	Black or African-American	62	9.00
6	Asian or Pacific Islander	252	14.57
6	Hispanic or Latino	940	11.56
6	White	188	21.61
6	Multi-Racial (not Hispanic origin)	2	22.22
6	NYC	864	10.62
6	Big 4 Cities	99	13.73
6	High Need Urban/Suburban	179	15.39
6	High Need Rural	13	15.12
6	Average Need	169	19.23
6	Low Need	120	26.85
6	Charter Schools	3	16.67
6	0 Years LEP	81	6.36
6	1 Years LEP	121	5.89
6	2 Years LEP	162	10.82
6	3 Years LEP	166	15.16
6	4 Years LEP	144	15.65
6	5 Years LEP	177	21.25
6	6 Years LEP or More	559	15.48
6	Arabic	20	7.14
6	Bengali	23	12.23
6	Chinese	27	12.62
6	English	91	16.52
6	Spanish	851	11.18
6	Other Language	376	16.76
6	Students with Disabilities	130	6.32

I.8: Grade 7

Grade	Group	N-count	Percent
7	Female	443	8.93
7	Male	423	7.18
7	Black or African-American	31	4.73
7	Asian or Pacific Islander	164	9.21
7	Hispanic or Latino	550	7.26
7	White	121	14.68
7	NYC	590	7.45
7	Big 4 Cities	41	6.08
7	High Need Urban/Suburban	64	6.74
7	High Need Rural	8	10.00
7	Average Need	99	11.72
7	Low Need	61	16.09
7	Charter Schools	2	15.38
7	0 Years LEP	52	4.04
7	1 Years LEP	91	4.22
7	2 Years LEP	120	7.71
7	3 Years LEP	114	10.29
7	4 Years LEP	99	10.39
7	5 Years LEP	107	14.10
7	6 Years LEP or More	267	9.18
7	Arabic	12	4.40
7	Bengali	21	11.80
7	Chinese	15	7.58
7	English	57	11.11
7	Spanish	492	6.93
7	Other Language	236	10.37
7	Students with Disabilities	43	2.62

I.9: Grade 8

Grade	Group	N-count	Percent
8	Female	748	14.20
8	Male	725	11.57
8	American Indian or Alaska Native	2	20.00
8	Black or African-American	88	12.19
8	Asian or Pacific Islander	292	15.59
8	Hispanic or Latino	891	11.06
8	White	200	23.28
8	NYC	999	11.42
8	Big 4 Cities	82	14.56
8	High Need Urban/Suburban	125	13.24
8	High Need Rural	9	13.64
8	Average Need	163	19.59
8	Low Need	94	25.75
8	Charter Schools	1	10.00
8	0 Years LEP	91	7.45
8	1 Years LEP	119	5.55
8	2 Years LEP	184	10.51
8	3 Years LEP	150	12.03
8	4 Years LEP	186	17.78
8	5 Years LEP	186	21.58
8	6 Years LEP or More	536	17.09
8	Arabic	21	7.84
8	Bengali	35	18.23
8	Chinese	22	12.02
8	English	99	20.33
8	Spanish	799	10.54
8	Other Language	431	17.41
8	Students with Disabilities	91	6.22

I.10: Grade 9

Grade	Group	N-count	Percent
9	Female	736	12.55
9	Male	853	11.96
9	American Indian or Alaska Native	4	16.00
9	Black or African-American	121	10.92
9	Asian or Pacific Islander	328	12.40
9	Hispanic or Latino	914	11.28
9	White	217	20.53
9	Multi-Racial (not Hispanic origin)	5	8.62
9	NYC	1134	11.88
9	Big 4 Cities	83	14.24
9	High Need Urban/Suburban	109	9.85
9	High Need Rural	6	7.79
9	Average Need	157	13.96
9	Low Need	93	17.75
9	Charter Schools	3	42.86
9	0 Years LEP	126	8.14
9	1 Years LEP	160	4.57
9	2 Years LEP	183	9.35
9	3 Years LEP	157	11.63
9	4 Years LEP	153	14.61
9	5 Years LEP	182	22.06
9	6 Years LEP or More	599	23.27
9	Arabic	27	7.94
9	Bengali	27	10.00
9	Chinese	31	7.11
9	English	124	19.47
9	Spanish	850	11.22
9	Other Language	465	13.98
9	Students with Disabilities	17	8.67

I.11: Grade 10

Grade	Group	N-count	Percent
10	Female	597	10.84
10	Male	625	10.50
10	American Indian or Alaska Native	1	5.26
10	Black or African-American	120	11.78
10	Asian or Pacific Islander	259	8.80
10	Hispanic or Latino	686	10.35
10	White	153	18.73
10	Multi-Racial (not Hispanic origin)	3	7.50
10	NYC	869	9.93
10	Big 4 Cities	41	10.07
10	High Need Urban/Suburban	107	11.47
10	High Need Rural	6	10.91
10	Average Need	128	15.40
10	Low Need	70	14.77
10	Charter Schools	1	8.33
10	0 Years LEP	89	7.18
10	1 Years LEP	151	5.75
10	2 Years LEP	191	8.43
10	3 Years LEP	125	9.31
10	4 Years LEP	122	12.07
10	5 Years LEP	126	16.43
10	6 Years LEP or More	400	19.35
10	Arabic	11	5.37
10	Bengali	24	9.02
10	Chinese	31	5.96
10	English	70	13.49
10	Spanish	631	10.17
10	Other Language	398	11.66
10	Students with Disabilities	11	8.15

I.12: Grade 11

Grade	Group	N-count	Percent
11	Female	380	10.93
11	Male	345	10.21
11	Black or African-American	67	10.86
11	Asian or Pacific Islander	191	10.21
11	Hispanic or Latino	376	9.87
11	White	90	17.68
11	Multi-Racial (not Hispanic origin)	1	2.86
11	NYC	501	10.45
11	Big 4 Cities	16	7.08
11	High Need Urban/Suburban	58	9.15
11	High Need Rural	3	7.50
11	Average Need	99	13.25
11	Low Need	48	11.68
11	0 Years LEP	42	7.84
11	1 Years LEP	73	6.05
11	2 Years LEP	108	7.94
11	3 Years LEP	107	9.49
11	4 Years LEP	91	12.66
11	5 Years LEP	82	15.41
11	6 Years LEP or More	207	16.40
11	Arabic	16	12.31
11	Bengali	13	8.39
11	Chinese	33	9.76
11	English	52	13.30
11	Spanish	340	9.79
11	Other Language	244	11.60
11	Students with Disabilities	6	8.11

I.13: Grade 12

Grade	Group	N-count	Percent
12	Female	219	12.95
12	Male	188	11.91
12	American Indian or Alaska Native	1	16.67
12	Black or African-American	35	12.03
12	Asian or Pacific Islander	115	12.95
12	Hispanic or Latino	200	11.07
12	White	55	20.83
12	Multi-Racial (not Hispanic origin)	1	7.69
12	NYC	227	10.17
12	Big 4 Cities	7	5.22
12	High Need Urban/Suburban	45	16.67
12	High Need Rural	3	11.54
12	Average Need	75	18.99
12	Low Need	50	23.81
12	0 Years LEP	81	20.66
12	1 Years LEP	47	15.56
12	2 Years LEP	65	10.62
12	3 Years LEP	56	10.98
12	4 Years LEP	52	11.93
12	5 Years LEP	34	10.97
12	6 Years LEP or More	64	9.74
12	Arabic	7	14.29
12	Bengali	6	10.00
12	Chinese	16	11.43
12	English	37	14.80
12	Spanish	166	10.40
12	Other Language	147	14.03
12	Students with Disabilities	3	4.17

APPENDIX J: DIF CATEGORIES FOR OPERATIONAL ITEMS**Table J.1 Results of DIF Analyses**

Grade Span	Modality	Item Seq. Number	DIF– Male/Female	DIF– Hispanic/Asian	DIF– Hispanic/Black	DIF– Hispanic/White
K-1	Listen/Speak	3	A	>B	A	A
		4	A	>B	A	A
		5	A	>C	A	A
		6	A	>C	>C	A
		7	A	>B	A	A
		8	>B	>C	>B	A
		10	A	A	A	<B
		11	A	>B	A	A
		14	A	A	A	<B
		18	A	>C	A	A
		20	A	>B	>B	A
		22	A	>B	A	A
		31	AA	>BB	AA	AA
		33	AA	>BB	AA	AA
		35	AA	>BB	AA	AA
		37	AA	>BB	AA	AA
		38	AA	>BB	AA	AA
40	AA	>BB	AA	AA		
K-1	Read/Write	8	A	<B	A	A
		15	<B	A	A	A
		19	A	<B	A	A
		22	AA	AA	>BB	AA
		23	AA	AA	>CC	AA
		25	AA	<BB	<BB	<BB
		26	AA	<CC	AA	AA
		27	AA	<BB	AA	AA
		28	AA	<BB	AA	AA
		29	AA	<BB	AA	AA
		30	AA	<BB	AA	AA

Table J.1 Results of DIF Analyses (Continued)

Grade Span	Modality	Item Seq. Number	DIF– Male/Female	DIF– Hispanic/Asian	DIF– Hispanic/Black	DIF– Hispanic/White
2–4	Listen/Speak	3	A	>C	A	A
		5	A	A	<C	<B
		23	A	>B	A	A
		28	AA	>BB	AA	AA
		29	AA	>BB	AA	AA
2–4	Read/Write	3	A	<B	A	A
		36	A	<B	A	A
5–6	Listen/Speak	1	A	A	A	<B
		2	A	A	A	<B
		3	A	A	<B	A
		12	A	>B	A	A
		28	AA	>CC	AA	AA
		30	AA	>CC	AA	AA
		31	AA	>CC	AA	AA
		33	AA	>BB	AA	AA
		34	AA	>BB	AA	AA
		35	AA	>BB	AA	AA
5–6	Read/Write	2	A	>C	>C	>B
		4	A	<B	A	A
		6	A	>B	A	>B
		31	A	<C	A	<B
		35	A	>C	A	A
7–8	Listen/Speak	2	>B	>C	A	A
		3	A	>C	A	>B
		14	A	<B	A	A
		27	AA	AA	<BB	AA
		35	AA	>BB	AA	AA
7–8	Read/Write	2	>B	A	A	A
		4	A	<B	A	A
		12	A	A	>B	A
		17	A	>B	>B	>B
		19	A	>C	A	>B
		21	A	<B	A	A
		23	A	<B	A	<B
		27	A	<B	A	A
		28	B	A	>B
		29	A	A	>C	>C
		31	A	A	<B	A
		32	A	A	A	>B
		36	A	A	<C	A
		37	A	<C	<C	<C
		38	A	>C	A	A
		39	A	<B	A	<B
41	A	A	<B	A		
42	A	A	>B	A		
46	<BB	AA	AA	AA		

Table J.1 Results of DIF Analyses (Continued)

Grade Span	Modality	Item Seq. Number	DIF– Male/Female	DIF– Hispanic/Asian	DIF– Hispanic/Black	DIF– Hispanic/White
9–12	Listen/Speak	2	A	<B	<C	<C
		3	A	>B	A	A
		4	A	<B	A	A
		5	A	A	A	<B
		14	A	<B	A	A
		17	A	>C	>C	>B
		18	A	>B	A	A
		22	A	A	<B	A
		26	AA	>BB	AA	AA
		28	AA	>CC	AA	AA
		31	AA	>BB	AA	AA
		33	AA	>BB	AA	AA
		35	AA	>BB	AA	AA
		38	AA	>BB	AA	AA
9–12	Read/Write	1	A	A	>B	A
		3	A	<B	A	A
		4	A	>B	A	A
		18	A	<B	A	A
		19	A	>B	>C	>B
		25	A	>C	>C	>C
		26	A	>B	A	A
		31	A	<C	<B	<B
		35	A	A	A	>B
		36	A	<B	A	A
		37	A	<B	A	<B
		38	A	<B	<C	A
		40	A	<B	A	A
		41	A	A	<B	A
42	A	<B	<B	A		
43	AA	>BB	AA	AA		
46	<BB	AA	AA	AA		

Note: > indicates in favor of Males or Hispanics while < indicates in favor of the focal groups.

APPENDIX K: REFERENCES

- Altman, D. G. (1991). *Practical Statistics for Medical Research*. Chapman and Hall/CRC. 404.
- Andrich, A., & Luo, G. (2004). *Modern Measurement and Analysis in Social Science*. Perth, Western Australia: Murdoch University.
- Angoff, W. H. (1971). "Scales, norms and equivalent scores." In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.). Washington, D.C.: American Council on Education. 508–600.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Psychological Association.
- Browne, M. W. (1982). "Covariance structures." In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis*. Cambridge: Cambridge University Press. 72–141.
- Browne, M. W. & Cudeck, R. (1993). "Alternative ways of assessing model fit." In K.A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage Publication. 136–162.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16: 297–334.
- Steinley, D. & Wood, P. (2000). *ICC.SAS—a Program to Calculate Intra-class Correlations and Confidence Intervals*. Columbia, MO: University of Missouri.
- Embretson, S. & Reise, S. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Haertel, E. H. (1996). "Estimating the Decision Consistency from a Single Administration of a Performance Assessment Battery." *A Report on the National Board of Professional Teaching Standards McGEN Assessment*. Palo Alto, CA: Stanford University.
- Hanson, B. A. & Beguin, A. A. (1999). *Obtaining a common scale for IRT item parameters using separate versus concurrent estimation in the common item nonequivalent groups equating design*. Paper presented at the meeting of the National Council on Measurement in Education in Montreal.
- Hanson, B. A. (1995). *USmooth: A Program for Smoothing Univariate Test Score Distribution (Version 1.5)*. Iowa City, IA: American College Testing.
- Hanson, B. A. (1991). *Method of moments estimates for the four-parameter beta compound binomial model and the calculation of classification consistency indexes*. *ACT Research Report 91–5*. Iowa City, IA: American College Testing.
- Hu, M. & Bentler, P. M. (1999). "Cut off Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling* 6: 1–55.

Huynh, H. (2000). Robust linking for PACT assessments. *South Carolina PACT Technical Report*.

Jöreskog, K. G. & Sörbom, D. (1995a). *Liseral 8 User's Reference Guide*. Chicago: Scientific Software, Inc.

Jöreskog, K. G. & Sörbom, D. (2004). *Liseral 8.7 for Windows/Computer Software*. Lincolnwood, IL: Scientific Software International, Inc.

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

Kolen, M. J. & Brennan R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices* (2nd ed.). New York: Springer-Verlag.

Linacre, J. M. (2003). *A User's Guide to WINSTEPS: MINISTEP: Rasch-Model Computer Program*. Chicago: MESA Press.

Linacre, J. M. & Wright, B. D. (2000). *A User's Guide to WINSTEPS: Rasch-Model Computer Program*. Chicago: MESA Press

Livingston, S. A. & Lewis, C. (1995). Estimating the Consistency and Accuracy of Classifications Based on Test Scores. *Journal of Educational Measurement* 32: 179–1987.

Lord, F. M. & Wingersky, M. S. (1984). Comparison of IRT True-Score and Equipercentile Observed-Score "Equatings". *Applied Psychological Measurement* 8: 452–461.

MacCallum, R. C., Roznowski, M. & Necowitz, L. B. (1992). Model Modifications in Covariance Structure Analysis: The Problem of Capitalization on Chance. *Psychological Bulletin* 111: 490–504.

MacCallum, R. C., Browne, M. & Sugawara, H. M. (1996). Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods* 2(1): 130–149.

Masters, G. N. (1982). "A Rasch Model for Partial Credit Scoring". *Psychometrika* 47: 149–174.

McLeod, L. D., Swygert, K. A. & Thissen, D. (2001). Factor Analysis for Items Scored in Two Categories. In D. Thissen and H. Wainer (Eds.), *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum Associates. 189–216.

Mitzel, H.C., Lewis, D.M. & Green, D.R. (2001). The Bookmark Procedure. In G.J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates. 249–282.

Nitko, A. J. (2004). *Educational Assessment of Students* (4th Ed.). Upper Saddle River, NJ: Pearson Education Inc.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin* 86(2): 420–428.

Swygert, K. A., McLeod, L. D. & Thissen, D. (2001). Factor Analysis for Items or Testlets Scored in More than Two Categories. In D. Thissen and H. Wainer (Eds.), *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum Associates. 217–250.

Taherbhai, H. & Seo, D. (2007). Comparing Concurrent vs. Fixed Parameter Equating with Common Items, Using the Rasch Dichotomous and the Partial Credit Models in a Mixed-Item Format Test. *Journal of Applied Measurement* 8(1).

Thissen, D. & Steinberg, L. (1983). *A Response Model for Multiple Choice Items*. *Psychometric Technical Report 1*. Chicago: National Opinion Research Center.

Young, M. J. & Yoon, B. (1998, April). Estimating the Consistency and Accuracy of Classifications in a Standards-Referenced Assessment. *CSE Technical Report 475*. Los Angeles: University of California, Center for the Study of Evaluation, Standards, and Student Testing.