



NEW YORK STATE ENGLISH AS A SECOND LANGUAGE ACHIEVEMENT TEST (NYSESLAT)

2011 OPERATIONAL TEST TECHNICAL REPORT

SUBMITTED BY:

QUESTAR ASSESSMENT, INC.

TO: THE NEW YORK STATE

EDUCATION DEPARTMENT

DECEMBER 2011

REVISED OCTOBER 2014



5550 Upper 147th Street West
Apple Valley, MN 55124
(952) 997-2700

www.questarAI.com

TABLE OF CONTENTS

OVERVIEW	1
<i>Chapter 1: Introduction</i>	1
<i>Chapter 2: Test Design and Development</i>	1
<i>Chapter 3: Scoring</i>	1
<i>Chapter 4: Classical Item-Level Statistics</i>	1
<i>Chapter 5: Reliability</i>	1
<i>Chapter 6: Validity</i>	2
<i>Chapter 7: Calibration, Equating, and Scaling</i>	2
<i>Chapter 8: Standard Setting</i>	2
<i>Chapter 9: Summary of Operational Test Results</i>	2
CHAPTER 1: INTRODUCTION	3
<i>1.1 Background</i>	3
<i>1.2 Rationale and Purpose</i>	3
<i>1.3 Recommended Test Use</i>	4
<i>1.4 Test Accommodations</i>	4
CHAPTER 2: TEST DESIGN AND DEVELOPMENT.....	6
<i>2.1 Overview</i>	6
<i>2.2 Test Blueprint and Specifications by Modality and Grade Span</i>	6
<i>2.3 Test Mapping and Specifications by New York State Learning Standards for English as a Second Language</i>	10
<i>2.4 Item Development and Review Processes</i>	13
<i>2.5 Field Test</i>	13
<i>2.6 Test Construction</i>	14
CHAPTER 3: SCORING	18
<i>3.1 Scoring Audit: 10% of the Operational Writing Responses</i>	18
<i>3.2 Scoring the Speaking Field-Test Responses</i>	18
<i>3.3 NYSESLAT Writing Range-finding</i>	19
<i>3.4 Scoring the Writing Field-Test Responses</i>	19
<i>3.5 Reader Reliability</i>	19
CHAPTER 4: CLASSICAL ITEM-LEVEL STATISTICS.....	20
<i>4.1 Item-Level Descriptive Statistics</i>	20
<i>4.2 Differential Item Functioning</i>	22
CHAPTER 5: RELIABILITY	27
<i>5.1 Internal Consistency Reliability</i>	27
<i>5.2 Classical SEM (Based on Classical Test Theory)</i>	27
<i>5.3 Standard Error of the Mean (SEM_n)</i>	28
<i>5.4 Conditional SEM (Based on Item Response Theory)</i>	29
<i>5.5 Summary of Descriptive and Reliability Statistics</i>	30
<i>5.6 Inter-Rater Reliability</i>	35
<i>5.7 Reliability of Classification Decision at Proficient Cut</i>	40
CHAPTER 6: VALIDITY	44
<i>6.1 Content Validity</i>	44
<i>6.2 Internal Structure</i>	45
<i>6.3 External Structure</i>	50
<i>6.3.1 Relationship with the New York English Language Arts Test, Grades 3–8</i>	51
<i>6.3.2 Relationship with the New York State Regents Comprehensive Examination in English</i>	58

CHAPTER 7: CALIBRATION, EQUATING, AND SCALING.....	63
7.1 <i>Item Response Model and Rational for Use</i>	63
7.2 <i>Evidence of Model Fit</i>	64
7.3 <i>Description of Calibration Sample, Process</i>	65
7.4 <i>Equating of the NYSESLAT</i>	65
7.5 <i>Scaling of the NYSESLAT</i>	80
7.6 <i>Rasch Information</i>	81
CHAPTER 8: STANDARD SETTING.....	83
8.1 <i>Introduction</i>	83
8.2 <i>Standard Setting Model</i>	83
8.3 <i>Committees of Panelists</i>	83
8.4 <i>Performance Levels and Cut Scores</i>	84
8.5 <i>Standard-Setting Process</i>	84
8.6 <i>Summary Statistics for the Three Rounds of Ratings</i>	86
8.7 <i>Post-Standard-Setting Analyses</i>	86
8.8 <i>Final Performance-Level Cut Points</i>	86
CHAPTER 9: SUMMARY OF OPERATIONAL TEST RESULTS	88
APPENDIX A: ITEM MAP FOR 2011 NYSESLAT	94
APPENDIX B: ITEM LEVEL STATISTICS BY GRADE SPAN	139
APPENDIX C: RAW SCORE TO SCALE SCORE CONVERSION TABLES	149
APPENDIX D: IRT STATISTICS	165
APPENDIX E: ITEM INFORMATION AT DIFFERENT CUT POINTS BY GRADES	175
APPENDIX F: CONFIRMATORY FACTOR ANALYSIS FOR THE 2009 NYSESLAT	201
APPENDIX G: SCALE SCORE SUMMARY BY SUBGROUPS.....	210
APPENDIX H: PROFICIENCY PERCENTAGES	236
APPENDIX I: EXIT RATE BY SUBGROUP	275
APPENDIX J: DIF CATEGORIES FOR OPERATIONAL ITEMS	288
APPENDIX K: REFERENCES	290

LIST OF TABLES

Table 1—2011 NYSESLAT Operational Test Blueprint	7
Table 2—2011 Test Specifications by Modality and Grade Span	9
Table 3—Maximum Number of Points by Modality and Grade Span	9
Table 4—Test Specifications for New York State ESL Learning Standards by Grade Span	10
Table 5—Item Mapping by New York State ESL Learning Standards and Performance Indicators	11
Table 6—Four-Year Sampling Matrix for the NYSESLAT Field Test	14
Table 7—N-Count by Grade Band	21
Table 8—Summary of Classical Item Difficulty and Item Discrimination Indices for Each Grade Span.....	21
Table 9—DIF Classification for Constructed-Response Items.....	24
Table 10—DIF Classification Labels for Multiple-Choice Items.....	25
Table 11—Results of DIF Analyses for 2009 Operational Test Items	26
Table 12—Descriptive Statistics and Reliability by Grade Span and Modality	31
Table 13—Descriptive Statistics and Reliability by Grade and Modality	32
Table 14—Audit Sample	36
Table 15—Rater Agreement for Pre-Writing and Writing Prompts	38
Table 16—Percentages of the Score Difference Between Raters.....	39
Table 17—Comparison Between Local and Audit Raters.....	40
Table 18—Classification Accuracy and Consistency by Grade	43
Table 19—Intercorrelation Among the Modalities by Grade Span	46
Table 20 - Intercorrelation Among the Modalities by Grade.....	46
Table 21—Global Fit Indices for the One- and Two-Factor Models (Grades K–1).....	50
Table 22—Global Fit Indices for the One- and Two-Factor Models (Grades 2–4).....	50
Table 23—Global Fit Indices for the One- and Two-Factor Models (Grades 5–6).....	50
Table 24—Global Fit Indices for the One- and Two-Factor Models (Grades 7–8).....	50
Table 25—Global Fit Indices for the One- and Two-Factor Models (Grades 9–12).....	50
Table 26—Descriptive Statistics of the NYSESLAT Raw Scores of All Modalities and Their Correlation with the ELA Test Total Scale Scores.....	53
Table 27—Descriptive Statistics of the New York ELA Test and NYSESLAT Combined Scale Scores of Listening/Speaking and Reading/Writing.....	54
Table 28—Grade 3: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level.....	57
Table 29—Grade 4: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level.....	57
Table 30—Grade 5: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level.....	57
Table 31—Grade 6: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level.....	57
Table 32—Grade 7: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level.....	58
Table 33—Grade 8: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level.....	58
Table 34—Descriptive Statistics of the NYSESLAT Raw Scores of Listening, Speaking, Reading, and Writing Modalities and Their Correlations With the Regents Comp Exam in English Total Scale Scores.....	59
Table 35—Descriptive Statistics of the NYSESLAT Scale Scores of Listening/Speaking, Reading/Writing Modalities, and Total Scale Scores and Their Correlations With the Regents Comp Exam in English Total Scale Scores	59
Table 36—Grade 10: Percentage of ELL Students Passing/Failing the Regents Comp Exam in English at Each NYSESLAT Proficiency Level	61
Table 37—Grade 11: Percentage of ELL Students Passing/Failing the Regents Comp Exam in English at Each NYSESLAT Proficiency Level	62
Table 38—Grade 12: Percentage of ELL Students Passing/Failing the Regents Comp Exam in English at Each NYSESLAT Proficiency Level	62

Table 39 - Common Items as Anchor Items by Modality Combination and Grade Span	68
Table 40—Average Rasch Difficulty by Grade Span and Modality	82
Table 41—Panel Composition for Standard-Setting Committees	84
Table 42—Final Performance-Level Cut Points for Listening and Speaking	87
Table 43—Final Performance-Level Cut Points for Reading and Writing.....	87
Table 44—Raw Score Summary by Grade.....	89
Table 45—Raw Score Summary by Grade Span.....	89
Table 46—Scale Score Summary by Grade	90
Table 47— Scale Score Summary by Grade Span	90
Table 48—Percentage of Students in Each Proficiency Level by Grade.....	91
Table 49—Percentage of Students in Each Proficiency Level by Grade Span.....	92
Table 50—Exit Rate by Grade for 2008, 2009, 2010, and 2011 Administrations.....	92
Table 51—Exit Rate by Grade Span for 2008, 2009, 2010, and 2011 Administrations	93
Table 52 - Maximum Number of Points by Modality and Grade Span for the 2009 NYSESLAT	201
Table 53—Summary Statistics	204
Table 54—Factor Loading Estimates for the One- and Two-Factor Models	204
Table 55—Global Fit Indices for the One- and Two-Factor Models.....	204
Table 56—Summary Statistics	206
Table 57—Factor Loading Estimates for the One- and Two-Factor Models	206
Table 58—Global Fit Indices for the One- and Two-Factor Models.....	206
Table 59—Summary Statistics	206
Table 60—Factor Loading Estimates for the One- and Two-Factor Models	207
Table 61—Global Fit Indices for the One- and Two-Factor Models.....	207
Table 62—Summary Statistics	207
Table 63—Factor Loading Estimates for the One- and Two-Factor Models	207
Table 64 - Global Fit Indices for the One- and Two-Factor Models	207
Table 65—Summary Statistics	208
Table 66—Factor Loading Estimates for the One- and Two-Factor Models	208
Table 67—Global Fit Indices for the One- and Two-Factor Models.....	208

LIST OF FIGURES

Figure 1—Classification Accuracy.....41

Figure 2—Classification Consistency41

Figure 3—2011 TCC L/S Grade Span 1–5.....69

Figure 4—2011 TCC R/W Grade Span 1–5.....70

Figure 5—2010–11 TCC L/S Grade Span 1.....71

Figure 6—2010–11 TCC L/S Grade Span 2.....71

Figure 7—2010–11 TCC L/S Grade Span 3.....72

Figure 8—2010–11 TCC L/S Grade Span 4.....73

Figure 9—2010–11 TCC L/S Grade Span 5.....73

Figure 10—2010–11 SE Curves L/S Grade Span 1.....73

Figure 11—2010–11 SE Curves L/S Grade Span 2.....74

Figure 12—2010–11 SE Curves L/S Grade Span 3.....74

Figure 13—2010–11 SE Curves L/S Grade Span 4.....75

Figure 14— 2010–11 SE Curves L/S Grade Span 5.....75

Figure 15—2010–11 TCC R/W Grade Span 176

Figure 16—2010–11 TCC R/W Grade Span 276

Figure 17—2010–11 TCC R/W Grade Span 377

Figure 18—2010–11 TCC R/W Grade Span 477

Figure 19— 2010–11 TCC R/W Grade Span 578

Figure 20—2010–11 SE Curves R/W Grade Span 178

Figure 21—2010–11 SE Curves R/W Grade Span 2.....79

Figure 22—2010–11 SE Curves R/W Grade Span 3.....79

Figure 23—2010–11 SE Curves R/W Grade Span 4.....80

Figure 24—2010–11 SE Curves R/W Grade Span 5.....80

Figure 25—Path Diagram for the One-Factor Model.....203

Figure 26—Path Diagram for the Two-Factor Model204

OVERVIEW

The 2011 New York State English as a Second Language Achievement Test (NYSESLAT) Technical Report has nine chapters. The chapter headings are: Introduction; Test Design and Development; Scoring; Classical Item-Level Statistics; Reliability; Validity; Calibration, Equating and Scaling; Standard Setting; and Summary of Operational Test Results.

Chapter 1: Introduction

Chapter 1 contains background information on the NYSESLAT. This includes the rationale, purpose, and recommended use of the assessment. In addition, it describes the available test accommodations with detailed information on Large Type and Braille versions of the test.

Chapter 2: Test Design and Development

Chapter 2 details the development of the assessment and the process for its creation. The chapter includes the test specifications, a description of the item development and review processes, field-testing of the items, and test construction. In 2011, a reprint of the 2009 form was administered.

Chapter 3: Scoring

Chapter 3 provides a description of the scoring process. It includes a description of the range-finding meeting in Albany, New York, in 2011. There is also information about the audit-scoring process conducted on a ten-percent sample followed by a description of the rater training and inter-rater reliability of those raters.

Chapter 4: Classical Item-Level Statistics

Chapter 4 presents the item-level descriptive statistics based on Classical Test Theory (CTT) for the administration.

Chapter 5: Reliability

Chapter 5 relays reliability information of the assessment. This includes internal consistency reliability, classical Standard Error of Measurement (SEM), conditional SEM based on IRT, and inter-rater reliability. Moreover, the chapter 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 the classification decision at the proficient cut.

Chapter 6: Validity

Chapter 6 describes the validity studies that have been conducted for the assessment. The chapter also includes collected evidence of validity based upon the content of the test, its internal structure, and relationships to other variables.

Chapter 7: Calibration, Equating, and Scaling

Chapter 7 explains the Rasch and Partial Credit Models used in the NYSESLAT. The chapter also contains sample-item characteristic curves for dichotomous and polytomous items. Following these, the chapter describes the process of the calibration, equating, and scaling for the 2011 administration of the NYSESLAT. Chapter 7 also explains the rationale for the use of the Rasch IRT model. Finally, there are also summaries of the IRT model fit statistics and the average Rasch difficulty of the subtests.

Chapter 8: Standard Setting

Chapter 8 is a reiteration of the standard-setting process followed by the previous vendor. This process was done in order to establish the performance level cuts. The chapter 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 as reported in previous technical reports.

Chapter 9: Summary of Operational Test Results

Chapter 9 has the raw score summary, scale score summary, percentage of students in each performance category, and exit rate for the 2007, 2008, 2009, 2010, and 2011 administrations of the NYSESLAT.

CHAPTER 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 English language learner (ELL) 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 ELL students using the New York State English as a Second Language Achievement Test (NYSESLAT).

NCLB requires that the annual assessment of ELL students be based on specific student achievement objectives. 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* to meet the requirements of NCLB. New York’s learning standards and performance indicators (achievement objectives) for English as a Second Language (ESL) students 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: PreK–1, 2–4, 5–8, and 9–12. To meet Federal and State requirements regarding the assessment of ELL 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.

1.2 Rationale and Purpose

The New York State Board of Regents has established learning standards for all ELL students attending New York State schools. Based on NCLB mandates, all ELL students in Kindergarten through Grade 12 are assessed every year to measure English language proficiency in speaking,

listening, reading, and writing. In addition, each student's annual progress toward proficiency must be tracked. In compliance with this directive, NYSED developed an annual test that measures student progress toward meeting the State's ESL learning standards, NYSESLAT. The NYSESLAT helps schools determine which instructional standards teachers must devote time to in order to ensure their ELL students fully acquire the language proficiency that will prepare them for success in the classroom. The purpose of the NYSESLAT is to measure annual student improvement in achieving English language proficiency in order for students to ultimately exit LEP/ELL status and move into English mainstream programs.

1.3 Recommended Test Use

The NYSESLAT is used to make decisions for language instructional programs and accountability determinations. First, the test is used to measure the level of English proficiency of all ELL students in listening, speaking, reading, and writing. The English proficiency level, as measured by the NYSESLAT, is used by school districts to decide the type and amount of instructional services to which the students are entitled. Second, the NYSESLAT is used to measure students' annual progress toward learning English and attaining English language proficiency; it determines whether school districts and the State meet the required progress and attainment targets, as defined in the New York State Title III accountability system. Third, the NYSESLAT is used as the sole criteria for exiting ELL students who score at the proficient level from the ESL/Bilingual programs.

1.4 Test Accommodations

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. Adherence to these guidelines ensures that the assessments are accessible and valid for the widest range of students, including students with disabilities. Applying Universal Test Design during the development process helps eliminate the need to address after-the-fact accommodations, and Universal Test Design provides a better assessment for all students. Checklists are used to review every item to ensure that each is 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, Questar Assessment, Inc. utilizes in-house content and fairness experts to ensure that the forms are pulled with concepts of Universal Design in mind. Questar stringently reviews forms for special populations—such as visually or hearing-impaired students—to ensure that items are fair, reliable, and accessible to all.

Large Type

Questar Assessment, Inc. applied large-type product specifications that ease the test-taking experience for visually impaired children who require large type for each grade band. 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.

Braille

Questar Assessment, Inc. produces a Braille version of the NYSESLAT for visually impaired children using certified and experienced transcribers who can work with the multiple codes, rules, and guidelines of Braille. Questar produces Braille forms for each NYSESLAT subtest and grade span. For the K–1 grade band, a checklist is provided rather than a Braille test.

If a test question is difficult to Braille, Questar and content specialists determine if there are other ways that the construct could be worded or measured. To adapt some items, pictures are described. Care is taken not to convey the correct answer, but to give a description that would enable the student to ascertain the correct answer.

CHAPTER 2: TEST DESIGN AND DEVELOPMENT

2.1 Overview

In order 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 took several steps. First, NYSED developed the *Learning Standards for English as a Second Language*, which is a comprehensive set of ESL learning standards and performance indicators in the domains of Speaking, Listening, Reading, and Writing. The NYSED then requested test development, research, and scoring based on these standards. A determination was made to have the NYSESLAT be an assessment for five grade spans (K–1, 2–4, 5–6, 7–8, 9–12). The five tests are vertically scaled and each consists of four domains (Listening, Speaking, Reading, and Writing). Following the mandate of the legislation, the tests are designed to assess the English language skills of ELL students in Grades K–12 and to capture their progress toward achieving full English language proficiency. All of the tests were 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 consistent with applicable federal and state testing requirements.

2.2 Test Blueprint and Specifications by Modality and Grade Span

The NYSESLAT is divided into basic modalities—Speaking, Listening, Reading, and Writing—for each of the five grade bands. The tests have multiple-choice, constructed-response, short-response, and extended-response items. The number of items varies by grade span: K–1 has 70 items, 2–4 has 80 items, and the remaining grade spans (5–6, 7–8, and 9–12) have 87 items each.

Across all grade spans the Speaking modality has 16 constructed-response items. Listening and Reading modalities are assessed solely by the use of multiple-choice items. There are between 24 and 25 items in Listening depending on the grade span. In Reading, the number of items ranges from 15–27 for the different grade spans. Finally, the Writing modality contains between 15–19 items for the different grade spans and consists of the following parts:

- Multiple-choice section that assesses ELL students’ 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 response to graphic-based prompt (Grades K–12)

The operational test blueprint for the 2011 administration (a reprint of the 2009 assessment) of the NYSESLAT is presented in Table 1. Test specifications by modality and grade span are shown in Table 2. Table 3 provides the maximum number of points by modality and grade span. Table 4 provides a breakdown of the number of items based on New York State’s ESL learning

standards and grade spans. The 2011 reprint of the 2009 test design consists completely of items developed specifically for the NYSESLAT.

Table 1—2011 NYSESLAT Operational Test Blueprint

NYSESLAT 2011 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 Listening)		5	10	10	10	
	Number of Items	24	24	25	25	25	
Number of Points		24	24	25	25	25	
Reading	Word Reading	11					All items in this strand are multiple-choice items with 0-1 score points.
	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	
Number of Points		15	24	27	27	27	

Table 1—2011 NYSESLAT Operational Test Blueprint (continued)

NYSESLAT 2011 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		8	13	13	13	
Writing	Developmental Writing	8					Level K-1 has 8 CR items with 0-2 score points.
	Pre-Writing		3	3	3	3	Levels 2-4, 5-6, 7-8, and 9-12 each 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	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	5	5	5	5	5	Each of the five levels has 5 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	

Table 2—2011 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	CR/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 ER	80
5-6	16	25	27/5	15	3	1 ER	87
7-8	16	25	27/5	15	3	1 ER	87
9-12	16	25	27/5	15	3	1 ER	87

Table 3—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	CR/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 4—Test Specifications for New York State ESL Learning Standards by Grade Span

Grade Span	Standard*	Number of Items	Percentage of Total Items	Number of Score Points	Percentage of Total Points
K-1	1	37	53	46	47
	2	6	09	11	11
	3	11	16	15	15
	4	8	11	14	14
	5	8	11	11	11
	Total Test		70	100	97
2-4	1	45	56	50	48
	2	12	15	15	14
	3	12	15	22	21
	4	9	11	14	13
	5	2	03	3	03
	Total Test		80	100	104
5-6	1	46	53	52	47
	2	14	16	18	16
	3	9	10	16	14
	4	15	17	21	19
	5	3	03	4	04
	Total Test		87	100	111
7-8	1	45	52	48	43
	2	10	11	13	12
	3	13	15	21	19
	4	14	16	22	20
	5	5	06	7	06
	Total Test		87	100	111
9-12	1	34	39	36	32
	2	12	14	15	14
	3	14	16	19	17
	4	17	20	26	23
	5	10	11	15	14
	Total Test		87	100	111

*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 Mapping and Specifications by New York State Learning Standards for English as a Second Language

Table 5 provides an item-mapping summary, and Appendix A.1 provides, in detail, the item mapping for the 2011 NYSESLAT (2009 reprint) by New York State Learning Standards for each item within each grade span and modality. Item specifications are presented in Appendix A.2.

Table 5—Item Mapping by New York State ESL Learning Standards and Performance Indicators

Grade Span	Subtest	Standards	Performance Indicators	Items in Subtest	Percentage of Subtest Score
K-1	Speaking	Standard 2	PI 5, 7, 8	25, 26, 28, 30	25
		Standard 3	PI 1, 9	27, 29, 33	19
		Standard 4	PI 2, 3, 5	32, 34–38, 40	44
		Standard 5	PI 3, 4	31, 39	13
	Listening	Standard 1	PI 1, 3, 4, 5, 12, 16	1–15, 18, 22–24	79
		Standard 2	PI 2	16	4
		Standard 3	PI 9	20	4
		Standard 4	PI 1	19	4
		Standard 5	PI 2	17, 21	8
	Reading	Standard 1	PI 1, 3, 4, 5, 12, 16	1-13	87
		Standard 2	PI 2	14	7
		Standard 3	PI 5	15	7
	Writing Conventions	Standard 1	PI 12	16-21	40
Standard 1		PI 8, 11, 12	22–29	53	
	Standard 3	PI 1	31	7	
2-4	Speaking	Standard 1	PI 5, 6, 9	25–27, 29, 31	31
		Standard 2	PI 8	30	6
		Standard 3	PI 1, 3	28, 32, 39	19
		Standard 4	PI 2, 3, 5, 8, 9,	34–38, 40	38
		Standard 5	PI 4	33	6
	Listening	Standard 1	PI 1, 3, 4, 5, 6, 10, 16	1–10, 13–16, 18, 21	67
		Standard 2	PI 2	17, 19	8
		Standard 4	PI 5, 7	12, 20, 22, 23, 24	21
		Standard 5	PI 3	11	4
		Standard 1	PI 1, 3, 4, 6, 16	1–7, 12, 21–23	46
	Reading	Standard 2	PI 2, 5, 12	8–11, 16–20	38
		Standard 3	PI 3, 5	13, 24	8
		Standard 5	PI 3, 6	14, 15	8
		Standard 1	PI 1, 6, 12	25–35	69
	Writing Conventions	Standard 3	PI 1	36	6
		Standard 1	PI 6	37	6
		Standard 3	PI 1, 6	38–40	19
5-6	Speaking	Standard 1	PI 5, 9	30, 31	13
		Standard 4	PI 3, 5	26–28, 32–37, 39, 41	75
		Standard 5	PI 3	29, 38	13
	Listening	Standard 1	PI 1, 3, 4, 6, 10, 16	1, 2, 4, 5, 6, 8–12, 14, 15 18, 20–23	68
		Standard 3	PI 5	3	4
		Standard 4	PI 7, 8	7, 13, 16, 17, 19, 24, 25	28

Table 5—Item Mapping by New York State ESL Learning Standards and Performance Indicators (continued)

Grade Span	Subtest	Standards	Performance Indicators	Items in Subtest	Percentage of Subtest Score	
5-6	Reading	Standard 1	PI 1, 3, 4, 6, 10, 16	1-5, 7, 13, 14, 16, 19-23	52	
		Standard 2	PI 2, 4, 5, 12	9, 10, 11, 12, 17, 24-27	33	
		Standard 3	PI 9	6, 8, 15,	11	
		Standard 5	PI 6	18	4	
		Standard 1	PI 1, 12	28-42	79	
	Writing Conventions	Standard 3	PI 1, 3, 5, 9	43-46	21	
7-8	Speaking	Standard 1	PI 5, 6, 8	32, 34, 36, 38, 40	31	
		Standard 2	PI 8	31	6	
		Standard 3	PI 5	33	6	
		Standard 4	PI 2, 3, 5, 8, 9	26-30, 37, 39, 41	50	
		Standard 5	PI 4	35	6	
	Listening	Standard 1	PI 1, 2, 3, 6, 9, 12	1, 3-9, 11, 14-23, 25	80	
		Standard 2	PI 4	24	4	
		Standard 3	PI 2, 9	2, 10, 12	12	
		Standard 5	PI 7	13	4	
	Reading	Standard 1	PI 1, 3, 4, 6, 7, 9, 16	1,2, 3, 8-13, 25	37	
		Standard 2	PI 1, 2, 4, 5, 7, 12	4-7, 14-21	44	
		Standard 3	PI 2, 3, 5, 9	22-24, 26, 27	19	
	Writing Conventions	Standard 1	PI 1, 12	28-42	79	
		Standard 3	PI 1, 9	43-46	21	
	9-12	Speaking	Standard 1	PI 6	32, 34, 35	19
Standard 2			PI 8	31	6	
Standard 3			PI 1, 9	29, 30	13	
Standard 4			PI 2, 3, 5, 8	26-28, 33, 36-41	63	
Listening		Standard 1	PI 1, 2, 3, 9, 16	1-5, 11, 14, 16, 17, 20-22, 24	52	
		Standard 3	PI 5, 9	10, 15, 23, 25	16	
		Standard 4	PI 7	6, 7, 18, 19	16	
		Standard 5	PI 1, 2	8, 9, 12, 13	16	
		Standard 1	PI 1, 3, 4, 6, 16	1-7, 13, 14, 16, 17	41	
Reading		Standard 2	PI 1, 2, 4, 5	18, 19, 21, 22, 24-27	30	
		Standard 3	PI 3, 5, 9	9-11, 15, 23	19	
		Standard 4	PI 1, 3, 5	8, 12, 20	11	
		Writing Conventions	Standard 1	PI 6, 12, 15	28-30, 32-36, 38-42	68
			Standard 5	PI 1	31, 37	11
		Writing	Standard 3	PI 1, 6	45, 46	11
		Standard 5	PI 3	43, 44	11	

2.4 Item Development and Review Processes

All NYSESLAT items are written by educators of ELL students, which includes a number of New York State teachers. Item writers were given an Item Writers' Training Guide to assist them in developing the items. In addition, content specialists were available via phone and/or e-mail to answer any questions the item writers had. Subsequently, assessment specialists reviewed the passages and items sent in by the item writers and assessed the items with respect to item specifications. In addition, the assessment specialists worked to ensure the following:

- Absence of bias and sensitive topics in passages
- Item soundness. Item soundness can 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
- Items matched to the intended New York State Learning Standards for English as a Second Language

A committee of New York State ESL teachers was charged with reviewing the reading passages as a next step in the development process. Next, each test question was reviewed by grade-level ESL educators from New York State. Passages and questions are only accepted to continue in the development process if they are judged to be of acceptable quality and if they are fair to students from the various ethnic/cultural backgrounds of those who presently live in New York State. Following the committee reviews, a "Final Eyes" review was conducted by a separate group of New York State teachers to ensure items were error-free and appropriate. Once items were field tested, a group of New York State teachers were invited to conduct a range-finding session. During range-finding, the teachers reviewed potential anchor papers and made selections for practice sets in order to aid teachers in the local scoring of the NYSESLAT.

To facilitate equating, a selection of items is identified as an anchor set. These anchor items serve the purpose of providing statistics that help equate tests from one year to the next. This process allows the scaled scores from one year to be comparable to those of previous or subsequent years. Anchor item selection criteria are described in Section 7.4.

2.5 Field Test

For the field test, the test was constructed, NYSED reviewed and approved the field test forms, and then a stand-alone field test was administered to a representative sample of limited English proficient students around New York State. In this administration, there were two field test forms for each grade span. Each form was additionally split into the four modalities (Listening, Speaking, Reading, and Writing). Each modality's form was then administered to a representative sample of New York State schools. All field tests are conducted in New York State schools with current ELL populations. In order to accommodate the field test needs, school

districts have been divided into stratified random samples using the Need/Resource Capacity Index as the strata. The field test assignments were originally designed to rotate among the field test samples each year (see Table 6). For example, districts in Sample 2 (S2) were assigned to administer a field test in Listening in 2007. In 2008, the same Sample 2 districts were assigned to administer a field test in Writing.

Table 6—Four-Year Sampling Matrix for the NYSESLAT Field Test

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

The 2011 NYSESLAT test administration included a field test administration. The data from this field-test administration is used for forms construction of the 2012 NYSESLAT operational test. The 2011 field-testing sampling plan and statistics presented to NYSED are discussed in the *2011 NYSESLAT Field Test Technical Report*.

2.6 Test Construction

There was no test construction for the 2011 form because it was a repeat of the 2009 form. This section serves as part of the validity arguments and evidence that can be made to address the technical quality of the 2009 and 2011 forms. What follows is a reiteration of the procedures that the previous vendor used to construct the 2009 form.

Items selected for the 2009 NYSESLAT represented a complete range of difficulty at all grade levels from K–12. Items ranged from easier ones with high p -values that were primarily aimed at measuring the skills of students with very limited ability in English to items with fairly low p -values that were aimed at measuring the skills of students with advanced ability in English. The number of both multiple-choice and constructed-response items increased at each proficiency level, meeting the requirements of the NYSED.

The selection criteria for choosing items for the 2009 NYSESLAT operational forms included the following: 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 is 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 the ability of students to read a mixture of nouns, verbs, noun phrases, verb phrases, adjectives and adjective phrases sometimes, and adverbs occasionally. The Reading Comprehension items in the second section of the Grades 2–12 Reading subtest tested the ability of students 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 word from context; to infer the author’s purpose; and to understand why information was organized or presented in a particular way in a passage. Similar specific content criteria were applied for the other subtests. As explained above, items ranged

from easier ones with high p -values to items with low p -values. The 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 was 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 4.2, were performed to identify potential items that needed to be scrutinized for item bias. None of the items that showed statistical significant DIF were considered by content specialists to be biased based on either gender or ethnicity.

The process for constructing the 2009 NYSESLAT operational forms was as follows: As soon as clean statistics were available from the spring 2008 Field Test, ESL assessment specialists reviewed all items from the field test and pulled operational forms based on the above criteria. In addition to the spring 2008 Field Test items, items were pulled from the 2008 Operational Test to be used as anchor items in the 2009 Operational Test. The anchor items were necessary for placing the 2009 Operational Items on the NYSESLAT established scale and constituted, roughly, 25% of the items in each subtest. The previously indicated selection criteria for the non-anchor items (unique items) apply to anchor items as well. A draft of the operational forms was produced and then presented to NYSED. Next, a joint team representing the vendor and NYSED reviewed the forms, following all item-selection criteria, and made various changes to the draft forms. Final forms were then reviewed by 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 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:
 1. Check item difficulty
 2. Check the range of item difficulty: Items should be flagged if p -value is < 0.30 or > 0.95
 3. Check the point-biserial range: Try to avoid items with pt. biserial < 0.30
 4. Check the omit rate: Watch for items with an omit rate $> 5\%$
 5. Avoid items with a DIF bias flag. If it is necessary to select an item with a DIF 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 using the following statistical targets:
 1. Average p -value comparison between the spring 2009 form and spring 2008 form
The average p -values should be similar to one another
 2. Average p -value comparison between the anchor item set for spring 2009 and the entire spring 2008 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 are 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, advanced, and proficient 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 identify accurately 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, assessment specialists, editorial staff, 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 evaluated the coherence of a passage, the use 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. The sum of these analyses and evaluations determines the appropriateness of the language of a passage.

There is a gradual increase in difficulty from passage to passage at every grade span. 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. In order to assess age- and grade-appropriate materials for ELL items and stimuli for the oral-language subtests, the *Educational Developmental Laboratory (EDL) Core Vocabularies in Reading, Mathematics, Science, and Social Studies*, published by Steck-Vaughn, is used. Furthermore, 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 ELL

students have been investigated by NYSED and its vendors. These 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 ELL students in cognitive laboratories. This setting allows for careful observation and recording of student responses and student reactions to items. Outcomes of this procedure led to the following important design decisions:

- Item types
- Number of items
- Length of pauses between items
- Use of recorded stimuli
- Recording of students' 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 reflects current spoken language as much as possible, Speaking and Listening scripts are submitted to a read-aloud proofing process with ELL 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.

CHAPTER 3: SCORING

This chapter details the process for scoring of the field test, the operational test, and the audit. All scoring of 10% of the operational writing constructed responses (the audit) and the constructed-response field-test items for speaking and writing was conducted at Questar’s scoring center in Apple Valley. Training and monitoring was conducted by experienced Questar scoring directors and team leaders. All Questar readers have a minimum of a B.A./B.S. degree, and all NYSESLAT readers had experience scoring other ELL assessments. The following sections describe our scoring processes.

3.1 Scoring Audit: 10% of the Operational Writing Responses

The operational file for the spring 2011 NYSESLAT came from the New York State Data Repository after scoring by New York teachers was completed. The file was verified by Questar’s Information Services and Technology Group to ensure data accuracy based on the description values in the file layout.

One scoring director and four experienced team leaders trained thirty-nine readers to score the writing audit. Training consisted of reviewing the 2010-2011 Scoring Manuals and associated training sets and was accomplished in four days. All readers were trained to score the items in grade band K–1. Readers were then divided into four groups, each group scoring a different grade band (2–4, 5–6, 7–8, 9–12).

Training began on June 15, 2011, and scoring was completed by June 30, 2011. Each team leader was responsible for 9–10 readers. Duties included:

- Reading behind each member of the team to ensure accuracy; reviewing daily statistics to ensure scoring consistency; and planning and implementing retraining activities as necessary
- A 10% second reading of all responses in the 10% audit sample was performed as a reliability check.

3.2 Scoring the Speaking Field-Test Responses

One scoring director and seven team leaders trained sixty-two readers to score the speaking responses. Readers were trained using the 2010–2011 Operational Scoring Guide and the Exemplar CD for Speaking. The responses were scored from June 2–23, 2011. During scoring, readers identified possible exemplars to be used in future scoring guides. Team leaders reviewed the identified responses and selected five responses for each score point for each field-test item to upload as wav files for future use. From those five sample responses, we identified the best two in each score point. Transcripts and annotations for these items were compiled and saved. When the 2011–2012 operational forms for speaking are constructed, these responses will be used to develop the new scoring guides and CDs for teacher use.

3.3 NYSESLAT Writing Range-finding

Range-finding meetings were convened in Albany, New York, on July 11–14, 2011, prior to scoring the field-test items. One committee per grade band scored responses to be used to develop scoring guides and training sets. One day, July 15th, was set aside for Questar and NYSED to resolve scoring discrepancies. Upon NYSED approval of scored responses, Questar scoring directors developed scoring guides and training sets for all field-test responses. These materials were used to train readers in order to score the writing field test. When the 2011–2012 test forms are constructed, the field-test training materials will be used to build the 2011–2012 scoring materials for teachers.

3.4 Scoring the Writing Field-Test Responses

Two scoring directors and four team leaders trained thirty-five readers to score the writing field-test responses. The training materials developed after range-finding were used to train the readers. Scoring began on August 1 and was completed by August 12. The monitoring procedures used in the audit scoring were employed for field-test scoring.

3.5 Reader Reliability

All readers were trained to score using the same 2010–2011 materials that were used to train New York State teachers. Questar employed experienced readers who had proven themselves on other ELL programs. Our stringent monitoring procedures were applied during all scoring of NYSESLAT responses. The statistics of rater reliability are presented in Section 5.6.

CHAPTER 4: CLASSICAL ITEM-LEVEL STATISTICS

4.1 Item-Level Descriptive Statistics

Summary statistics for the raw scores of all items administered on the spring 2011 operational NYSESLAT forms are contained in this chapter. The data file used for this analysis was the 100% student file compiled by the NYSED data team and provided to Questar Assessment, Inc. The summary statistics are based in Classical Test Theory and include information such as the p -value and point-biserial correlation among other indices detailed below.

A p -value is defined as the proportion of students to answer an item correctly for the multiple-choice items. Higher p -values mean that items are easier, while lower p -value mean that items are difficult. For the constructed-response items, the p -value is reported as a proportion 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. Typically, this is interpreted to measure an item's ability to discriminate between low-ability and high-ability students. An item with a high point-biserial correlation is better at discriminating between low-ability and high-ability students than an item with a low point-biserial correlation.

Item-level statistics for the operational 2011 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 items (for multiple choice this is the percentage of examinees that answered the item correctly, for constructed response it is the proportion of the maximum points possible)
- Point biserial (index of discrimination between high and low scoring students)

Items that are too easy or too difficult are flagged based on their p -values. Items like these provide inadequate information. The inclusion of items outside of the pre-defined difficulty range is typically beneficial to the measurement process. However, since the NYSESLAT is a standard-referenced examination, content experts may determine that the inclusion of specific items is necessary. There is no consensus as to what is an acceptable p -value, but the policy guiding item review and form building has been to choose a p -value ≤ 0.90 and ≥ 0.30 .

As previously explained, the point biserial is an index that is used to assess the statistical suitability of an item for inclusion in an examination. As a correlation of the item to the total test, typically a larger point biserial is viewed as desirable for item inclusion. For form building, items are selected using the point biserial in two ways: 1) strength of association with the underlying trait—unidimensionality (high correlation) and 2) degree of local independence of the item (low correlation). A very high correlation, in fact, may mean the item is actually redundant or assesses essentially the same aspect of the construct. Too low a point biserial value may demonstrate a looser relationship between the underlying trait and the item than is acceptable.

For the NYSESLAT, the point biserial correlation for items is constrained to an upper and lower limit. The upper limit of the point biserial is fixed at 0.80, while the lower limit is fixed at 0.25. On occasion, there is an occurrence of an upper limit and lower limit violation, the exclusion of any items with a negative biserial correlation is stringently observed, since this may indicate a wrong key during the scoring process or that an item has an inappropriate or overly attractive distractor or distractors. Items with negative point biserial correlations are recommended for exclusion from the form-building process. The upper and lower bounds should be viewed as subjective guidelines or flags for further investigation. On occasion, upon investigation of the response pattern and distractors, an item may need to be rekeyed, or it may be determined that, based on content needs, the item needs to be included and is correctly keyed.

The total population had 208,685 records. After removing duplicates and off-level records (N=128), the total N=208,596. Invalid raw scores were also removed from the data for the final sample sizes used in Table 7. Table 7 also details the removal of invalid records for both Listening/Speaking and Reading/Writing. With the exception of some high *p*-values and point-biserial values, all the items fall well within the pre-set level of acceptance, both in terms of the *p*-value and the point biserial. These results can be reviewed in Appendix B. The mean *p*-values and point-biserial indices are presented in Table 8. The mean *p*-values and point-biserial correlations for each grade span in the combined modalities of Listening/Speaking and Reading/Writing are between the upper and lower bounds for an acceptable range of these values.

Table 7—N-Count by Grade Band

	Listening/Speaking					Reading/Writing				
	Initial N	Dup Removed	Off-Level Removed	Invalid RS	Final N	Initial N	Dup Removed	Off-Level Removed	Invalid RS	Final N
K-1	51,129	24	5	5	51,100	51,129	24	5	5	51,100
2-4	59,486	13	22	22	59,451	59,486	13	22	22	59,451
5-6	29,508	1	6	6	29,501	29,508	1	6	6	29,501
7-8	25,306	6	6	6	25,294	25,306	6	6	6	25,294
9-12	43,256	6	0	0	43,250	43,256	6	0	0	43,250

Table 8—Summary of Classical Item Difficulty and Item Discrimination Indices for Each Grade Span

Grade Span	Modality	N-Count	Item Difficulty	Item Discrimination
------------	----------	---------	-----------------	---------------------

			Mean	SD	Mean	SD
K-1	Listening	51,100	0.86	0.11	0.37	0.06
	Speaking	51,100	0.79	0.05	0.74	0.03
	Reading	51,100	0.76	0.09	0.50	0.11
	Writing	51,100	0.80	0.15	0.55	0.18
	Listening and Speaking	51,100	0.84	0.10	0.52	0.19
	Reading and Writing	51,100	0.78	0.12	0.53	0.15
2-4	Listening	59,451	0.81	0.11	0.40	0.06
	Speaking	59,451	0.84	0.03	0.75	0.03
	Reading	59,451	0.73	0.18	0.45	0.09
	Writing	59,451	0.73	0.12	0.54	0.10
	Listening and Speaking	59,451	0.82	0.09	0.54	0.18
	Reading and Writing	59,451	0.73	0.15	0.49	0.10
5-6	Listening	29,501	0.77	0.12	0.38	0.10
	Speaking	29,501	0.83	0.04	0.79	0.03
	Reading	29,501	0.72	0.12	0.49	0.11
	Writing	29,501	0.72	0.14	0.51	0.14
	Listening and Speaking	29,501	0.79	0.10	0.54	0.22
	Reading and Writing	29,501	0.72	0.13	0.50	0.12
7-8	Listening	25,294	0.70	0.15	0.38	0.07
	Speaking	25,294	0.79	0.03	0.81	0.02
	Reading	25,294	0.65	0.13	0.46	0.07
	Writing	25,294	0.69	0.13	0.49	0.13
	Listening and Speaking	25,294	0.73	0.12	0.55	0.22
	Reading and Writing	25,294	0.66	0.13	0.47	0.10
9-12	Listening	43,250	0.69	0.12	0.41	0.08
	Speaking	43,250	0.79	0.03	0.76	0.03
	Reading	43,250	0.62	0.16	0.40	0.08
	Writing	43,250	0.71	0.14	0.48	0.13
	Listening and Speaking	43,250	0.73	0.11	0.55	0.18
	Reading and Writing	43,250	0.66	0.16	0.43	0.11

Note: The data file used in the table above is the final research file. The total n-counts for Grades K-12 are 208,596 after cleaning. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

4.2 Differential Item Functioning

Differential Item Functioning (DIF) analyses are statistical procedures used to flag items for potential bias. DIF analyses are not proof of bias, but an indication that items need to be further scrutinized. In general, DIF statistics are used to compute the probability that one demographic group is more likely to correctly answer an item than another group or subgroup after having controlled for groups differences in ability. Once DIF analyses are performed, the results are useful in reviewing items and tests for potential bias in items. To reiterate: the presence of DIF does not necessarily indicate that an item is biased. Therefore, flagged items should be evaluated substantively for bias indication. Since the 2011 test is a repeated form, the results and text in this section are reprinted from the 2009 studies.

The NYSESLAT has constructed-response items that are polytomously scored. Therefore, the Mantel-Haenszel odds ratio α should not be used as a DIF index for all the items in the assessment. However, a generalization of the Mantel-Haenszel (1959) procedure for ordered categories, called the *Mantel Statistic* (Mantel, 1963), can be used for the assessment of DIF in examinations which contain both polytomous and dichotomous items. The Mantel Chi-square involves comparing the mean for two groups, conditioned upon a matching variable (in this case the total test score). The statistic has one degree of freedom under the null hypothesis of no relationship between group membership and correct or incorrect responses. 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 as follows:

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

The Mantel-Haenszel and the Mantel statistic, however, offer a significance test of the presence of DIF, but not an indication of the directionality of detected DIF. In other words, the significance does not indicate whether an item favors the reference or the focal group. The statistic is less likely to indicate 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

$$m_{Fk} = \frac{1}{n_{F+k}(\sum_t y_{tFtk})}, \quad (6)$$

and

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 9 below. DIF is categorized as “no DIF” (AA), “moderate DIF” (BB), or “large DIF” (CC).

Table 9—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	Moderate <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) and the Mantel-Haenszel odds ratio are transformed to what ETS has named the Delta Scale (D). To calculate delta, the odds ratio calculated as the odds of a correct response or P/Q (i.e., $P/[1-P]$). The odds ratio is then calculated as the odds of a correct response for the reference group divided by the odds of a correct response for the focal group. Thus, the odds ratio is as follows:

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

The null hypothesis for DIF is that the odds of correctly answering the item are equal for the two groups (i.e., the odds ratio is equal to 1):

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

Customarily, the odds ratio is then log-transformed and made symmetrical around zero with a range being in the interval $-\infty$ to $+\infty$. It is then called the log odds ratio, as follows:

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

The natural logarithm transformation of the odds ratio is symmetrical around zero. A zero value means that either group has equal odds of correctly answering the question. A positive DIF measure means that reference group is favored, while a negative value indicates DIF in favor of the focal group. β_{M-H} can also be transformed in a linear fashion in order to convert to an interval scale (Camilli & Shepard, 1994). This is the process which ETS used to transform β_{M-H} to their Delta Scale metric (D), via:

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

On this scale, the quantity that D represents is the relative difference in the difficulty of the items for the two groups whose performance is being compared. Table 10 depicts the ETS DIF classifications for multiple-choice items based on the M-H χ^2 and the item Delta Scale difference value (D).

Table 10—DIF Classification Labels for Multiple-Choice Items

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

NYSESLAT DIF analyses were completed on both gender and ethnic groups. For the gender DIF analysis, female students were considered the focal group while male students were considered the reference group. Three ethnic DIF analyses were conducted. Asian, Hispanic, and White students were considered the focal group in each analysis and the remaining ethnic groups were considered the reference group. Because of low n-counts, DIF analysis was not conducted using other ethnic groups as the focal group (e.g., African-American students). The same focal and reference groups were used in the DIF analysis of field test items, provided that enough n-counts were available for the comparison.

Appendix J (a reprint from the 2009 technical report) of this report provides the same information for the operational items. The > sign next to the DIF category indicates that the item

is in favor of the reference group, while the < sign indicates that the item is in favor of the focal group.

Summary of DIF Analyses

Table 11 contains the summary of DIF analysis of operational test items, including n-counts for each DIF analysis group and the number of items flagged by both SMD and/or Mantel for constructed response items and the ETS Delta Scale metric (D) and M-H criteria for multiple-choice items (see details in Table 9 and Table 10). 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 11 were identified as “biased” by the content specialists.

Table 11—Results of DIF Analyses for 2009 Operational Test Items

Grade Span	Modality	N-Counts	Total Number of DIF Items	Total Number of Moderate DIF Items	Total Number of Large DIF Items
K-1	Listening/Speaking	51	3	2	1
	Reading/Writing	51	1	0	0
2-4	Listening/Speaking	62	8	1	0
	Reading/Writing	60	4	2	0
5-6	Listening/Speaking	28	8	3	0
	Reading/Writing	27	4	3	0
7-8	Listening/Speaking	24	3	0	2
	Reading/Writing	23	11	5	2
9-12	Listening/Speaking	39	7	1	2
	Reading/Writing	38	13	6	6

Note: Detailed results can be found in Appendix J.

CHAPTER 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. There are several methods that can be used to estimate internal consistency. One method is called the “split-half” approach. This method splits all test questions into two groups and then correlates student scores on the two half-tests. The advantage of this method is to avoid any implications of changes in the individual students by administering only a single test. The theory is that if scores have high correlations on the two half-tests, it can be concluded that the test questions complement one another, function well as a group, and measure the same construct. In addition, a positive (high correlation) result would suggest that measurement error is lesser.

The problem with the split-half method is that the decision about which questions belong in which half of the test can have a rather large impact on the resulting correlation. Therefore, Questar, like many other vendors, prefers to use Cronbach’s coefficient alpha statistic (Cronbach, 1951) to sidestep this problem. Rather than elect specific items to put into two halves, coefficient alpha is the average split-half correlation based on all possible divisions of a test into two parts. Coefficient alpha has the additional advantage in that it can be used to estimate the internal consistency of both dichotomous and polytomous 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 is found in Tables 12 and 13 in Section 5.5. Reliability values of close to or more than 0.90 are considered high, and the test is considered a reliable test.

5.2 Classical SEM (Based on Classical Test Theory)

Because there is no such thing as perfect measurement of ability, it is important to analyze the amount of measurement errors on an assessment. Psychometricians sometimes think about reliability of an assessment as the ability of a test to consistently produce the same results. In essence, if a student repeatedly took the same assessment and theoretically was unable to remember its contents, that student should earn roughly the same score. Obviously, this cannot

be done in reality. However, the spread or standard deviation of these hypothetical scores is what is known as the standard error of measurement (SEM).

The SEM is another measure of the reliability of an assessment. It provides an estimate of the amount of the error that exists in an individual's observed test score. A given individual's observed total score, from this perspective, is thought of as an estimate of the student's true score. The SEM is inversely related to the reliability of a test, since the greater the reliability, the lower the standard error of measurement will be. Hence, there is more accuracy or precision in an observed test score when the SEM is small. Measurement error is commonly expressed in terms of standard deviations. That is to say, the SEM can be thought of as the standard deviation of the distribution of measurement error. The SEM is calculated using 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 12 and 13 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)

Yet another way to conceptualize reliability is to examine the standard error of the mean. This statistic is an estimate of the sampling error of the sample mean as an estimate of the population mean. Since a given administration can be thought of as all possible students that could take the NYSESLAT, a single administration can similarly be thought of as a sample of all possible tests takers. It is defined as follows:

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

where $SEMn$ is the standard error of the mean,
 σ is the standard deviation of the population, and
 n is the number of responses in each sample.

The *SEMn*'s are presented in Tables 12 and 13 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)

In classical testing theory, the SEM is the same for all students. However, in item-response theory (IRT), the SEM will not be the same for each and every student. For example, if a student gets either a few or a large number of items correct (extreme score), the standard error is greater than if the student gets a moderate number of items correct. In effect, the SEM depends on the total score (Andrich & Luo, 2004). Under the Rasch model, the SEM for each student 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 student answers an item correctly. It is defined as follows:

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

where β_v is student v 's ability, and δ_i is the difficulty of the item.

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

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

It should be noted that the standard error of measurement is smallest when the probability of passing an item is close to the probability of failing the item. To expand on this, if an item's difficulty is closely matched to the ability of many students in the sample, the standard error is small (Embretson & Reise, 2000).

One reason for the popularity of IRT over classical test theory in recent years is its ability to differentiate between students and item measurements. In other words, in classical test theory item measurements for all examinees and students themselves are assumed to be equivalent in a given test. IRT models varied student and item performances; this is taken into account when estimating the reliability of the measurement (Crocker & Algina, 1986). Equation 13 and its results can fluctuate in terms of the standard errors of students because they are a function of the n-counts. As such, the standard errors for each of the ability score estimates are the smallest in the middle of the score distribution (where most examinees perform) and greatest for estimates in the extreme. Hence, fewer and fewer students in the extremes lead to less precise estimates of ability. For this reason, IRT estimates individual standard errors at score points, known as

conditional standard errors of measurement. Essentially, the standard error is conditioned on theta (student ability). This is preferred to classical SEMs, since the conditional standard error of measurement (CSEM) allows for a more precise differentiation of students at different levels of performance.

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

5.5 Summary of Descriptive and Reliability Statistics

Table 12 and Table 13 provide descriptive statistics for the raw score and reliabilities by grade span and grade level. The tables contain the following:

- 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 (SEM_n)
- Cronbach's alpha internal consistency reliability
- Standard error of measurement (SEM)

In analyzing the information provided in Tables 12 and 13, 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 12, 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 SEM_n values of Listening/Speaking and Reading/Writing across the five grade spans ranged from 0.01 to 0.02, which is very small and thus acceptable. The SEM values of Listening/Speaking and Reading/Writing across the five grade spans ranged from 2.36 to 3.15, which are also acceptable. Of course, if an item has, for example, a SEM_n greater than the standard deviation of the mean, then the relevancy of the information would be questioned (which might be the result of an outlier or outliers).

In Table 13, most test reliability values of Listening/Speaking and Reading/Writing across the thirteen grades are above 0.86. The SEM_n and SEM values are small and within acceptable ranges at each grade level.

Table 12—Descriptive Statistics and Reliability by Grade Span and Modality

Grade	Test	Number of Items	Max		RS Mean	SD	Mean		Reliability	SEM
			Points	N-Count			P-value	SEMn		
K-1	Listening	24	24	51,100	20.64	3.53	0.86	0.02	0.80	1.57
	Speaking	16	34	51,100	26.64	7.97	0.79	0.01	0.95	1.81
	Reading	15	15	51,100	11.24	3.51	0.76	0.02	0.83	1.46
	Writing	15	24	51,100	18.65	5.00	0.80	0.04	0.87	1.79
	Listening and Speaking	40	58	51,100	47.27	10.67	0.84	0.02	0.94	2.70
	Reading and Writing	30	39	51,100	29.89	8.04	0.78	0.02	0.91	2.36
2-4	Listening	24	24	59,451	19.31	4.11	0.81	0.02	0.82	1.76
	Speaking	16	34	59,451	28.39	7.44	0.84	0.01	0.95	1.60
	Reading	24	24	59,451	17.37	4.63	0.73	0.04	0.84	1.87
	Writing	16	22	59,451	15.29	4.94	0.73	0.03	0.86	1.85
	Listening and Speaking	40	58	59,451	47.70	10.66	0.82	0.01	0.94	2.62
	Reading and Writing	40	46	59,451	32.66	9.07	0.73	0.02	0.91	2.68
5-6	Listening	25	25	29,501	19.12	4.40	0.77	0.02	0.81	1.90
	Speaking	16	34	29,501	27.97	8.33	0.83	0.01	0.96	1.59
	Reading	27	27	29,501	19.38	5.99	0.72	0.02	0.89	2.00
	Writing	19	25	29,501	17.46	5.45	0.72	0.03	0.86	2.03
	Listening and Speaking	41	59	29,501	47.10	11.78	0.79	0.02	0.94	2.79
	Reading and Writing	46	52	29,501	36.84	10.84	0.72	0.02	0.93	2.88
7-8	Listening	25	25	25,294	17.35	4.94	0.70	0.03	0.83	2.02
	Speaking	16	34	25,294	26.74	9.42	0.79	0.01	0.97	1.64
	Reading	27	27	25,294	17.46	6.10	0.65	0.03	0.88	2.14
	Writing	19	25	25,294	16.73	5.69	0.69	0.03	0.85	2.17
	Listening and Speaking	41	59	25,294	44.09	13.22	0.73	0.02	0.95	2.99
	Reading and Writing	46	52	25,294	34.19	11.00	0.66	0.02	0.92	3.09
9-12	Listening	25	25	43,250	17.04	5.44	0.69	0.02	0.86	2.07
	Speaking	16	34	43,250	26.88	8.49	0.79	0.01	0.96	1.69
	Reading	27	27	43,250	16.69	5.42	0.62	0.03	0.83	2.24
	Writing	19	25	43,250	17.50	5.51	0.71	0.03	0.85	2.14
	Listening and Speaking	41	59	43,250	43.91	12.77	0.73	0.02	0.95	2.96
	Reading and Writing	46	52	43,250	34.19	10.07	0.66	0.02	0.90	3.15

Note: The data file used in the table above is the final research file. The total n-counts for Grades K-12 are 208,596. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

Table 13—Descriptive Statistics and Reliability by Grade and Modality

Grade	Test	Number of Items	Max Points	N-Count	RS Mean	SD	Mean P-value	SEMn	Reliability	SEM
K	Listening	24	24	25,432	19.57	3.77	0.82	0.03	0.93	1.77
	Speaking	16	34	25,432	24.67	8.35	0.74	0.02	0.89	1.95
	Reading	15	15	25,432	9.51	3.37	0.64	0.03	0.79	1.69
	Writing	15	24	25,432	16.45	4.99	0.72	0.05	0.95	1.91
	Listening and Speaking	40	58	25,432	44.24	11.12	0.79	0.02	0.81	2.94
	Reading and Writing	30	39	25,432	25.96	7.72	0.68	0.03	0.85	2.60
1	Listening	24	24	25,668	21.69	2.92	0.91	0.02	0.93	1.33
	Speaking	16	34	25,668	28.59	7.06	0.85	0.01	0.90	1.64
	Reading	15	15	25,668	12.96	2.70	0.87	0.02	0.78	1.17
	Writing	15	24	25,668	20.84	3.94	0.88	0.02	0.95	1.51
	Listening and Speaking	40	58	25,668	50.28	9.27	0.88	0.01	0.81	2.38
	Reading and Writing	30	39	25,668	33.79	6.24	0.88	0.01	0.84	1.95
2	Listening	24	24	21,013	17.89	4.13	0.75	0.03	0.93	1.95
	Speaking	16	34	21,013	27.63	7.47	0.82	0.01	0.90	1.70
	Reading	24	24	21,013	15.54	4.58	0.65	0.04	0.81	2.00
	Writing	16	22	21,013	13.46	4.93	0.64	0.04	0.95	1.97
	Listening and Speaking	40	58	21,013	45.52	10.59	0.78	0.02	0.82	2.82
	Reading and Writing	40	46	21,013	29.00	8.93	0.65	0.03	0.85	2.84
3	Listening	24	24	20,140	19.68	3.89	0.82	0.02	0.94	1.71
	Speaking	16	34	20,140	28.62	7.32	0.85	0.01	0.91	1.57
	Reading	24	24	20,140	17.81	4.35	0.74	0.04	0.83	1.83
	Writing	16	22	20,140	15.76	4.65	0.75	0.03	0.96	1.80
	Listening and Speaking	40	58	20,140	48.30	10.39	0.83	0.01	0.84	2.57
	Reading and Writing	40	46	20,140	33.58	8.50	0.75	0.02	0.85	2.61
4	Listening	24	24	18,298	20.52	3.82	0.86	0.02	0.95	1.56
	Speaking	16	34	18,298	29.00	7.47	0.86	0.01	0.91	1.49
	Reading	24	24	18,298	18.98	4.24	0.79	0.03	0.93	1.71
	Writing	16	22	18,298	16.86	4.56	0.80	0.03	0.89	1.77
	Listening and Speaking	40	58	18,298	49.52	10.60	0.86	0.01	0.79	2.41
	Reading and Writing	40	46	18,298	35.84	8.34	0.79	0.02	0.95	2.49

Note: The data file used in the table above is the final research file. The total n-counts for Grades K–12 are 208,596. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

Table 13—Descriptive Statistics and Reliability by Grade and Modality (continued)

Grade	Test	Number of Items	Max Points	N-Count	RS Mean	SD	Mean P-value	SEMn	Reliability	SEM
5	Listening	25	25	15,833	19.15	4.35	0.77	0.02	0.81	1.90
	Speaking	16	34	15,833	28.22	8.05	0.84	0.01	0.96	1.58
	Reading	27	27	15,833	19.44	5.96	0.72	0.02	0.89	2.00
	Writing	19	25	15,833	17.50	5.36	0.72	0.03	0.86	2.02
	Listening and Speaking	41	59	15,833	47.37	11.46	0.79	0.02	0.94	2.76
	Reading and Writing	46	52	15,833	36.93	10.73	0.72	0.02	0.93	2.87
6	Listening	25	25	13,668	19.09	4.47	0.77	0.02	0.82	1.91
	Speaking	16	34	13,668	27.69	8.63	0.82	0.01	0.97	1.60
	Reading	27	27	13,668	19.31	6.03	0.72	0.02	0.89	2.00
	Writing	19	25	13,668	17.42	5.56	0.72	0.03	0.87	2.03
	Listening and Speaking	41	59	13,668	46.78	12.14	0.79	0.02	0.95	2.81
	Reading and Writing	46	52	13,668	36.73	10.97	0.72	0.02	0.93	2.88
7	Listening	25	25	12,632	17.10	4.84	0.69	0.03	0.82	2.04
	Speaking	16	34	12,632	26.85	9.33	0.79	0.01	0.97	1.64
	Reading	27	27	12,632	17.03	6.07	0.63	0.03	0.87	2.16
	Writing	19	25	12,632	16.56	5.65	0.68	0.03	0.85	2.16
	Listening and Speaking	41	59	12,632	43.95	13.05	0.73	0.02	0.95	3.00
	Reading and Writing	46	52	12,632	33.59	10.92	0.65	0.02	0.92	3.10
8	Listening	25	25	12,662	17.60	5.02	0.71	0.03	0.84	2.00
	Speaking	16	34	12,662	26.62	9.51	0.79	0.01	0.97	1.65
	Reading	27	27	12,662	17.90	6.09	0.67	0.02	0.88	2.11
	Writing	19	25	12,662	16.89	5.72	0.69	0.03	0.86	2.17
	Listening and Speaking	41	59	12,662	44.22	13.38	0.74	0.02	0.95	2.98
	Reading and Writing	46	52	12,662	34.79	11.04	0.68	0.02	0.92	3.07
9	Listening	25	25	14,640	15.91	5.81	0.64	0.02	0.87	2.12
	Speaking	16	34	14,640	25.01	9.99	0.74	0.01	0.97	1.75
	Reading	27	27	14,640	15.25	5.53	0.57	0.03	0.83	2.30
	Writing	19	25	14,640	16.10	6.13	0.66	0.03	0.87	2.22
	Listening and Speaking	41	59	14,640	40.92	14.66	0.68	0.02	0.95	3.13
	Reading and Writing	46	52	14,640	31.35	10.77	0.61	0.02	0.91	3.26

Note: The data file used in the table above is the final research file. The total n-counts for Grades K–12 are 208,596. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

Table 13—Descriptive Statistics and Reliability by Grade and Modality (continued)

Grade	Test	Number of Items	Max Points	N-Count	RS Mean	SD	Mean		Reliability	SEM
							P-value	SEMn		
10	Listening	25	25	13,276	16.60	5.38	0.67	0.03	0.85	2.10
	Speaking	16	34	13,276	26.47	8.33	0.78	0.01	0.96	1.74
	Reading	27	27	13,276	16.44	5.30	0.61	0.03	0.82	2.25
	Writing	19	25	13,276	17.26	5.40	0.70	0.03	0.84	2.13
	Listening and Speaking	41	59	13,276	43.07	12.52	0.71	0.02	0.94	3.00
	Reading and Writing	46	52	13,276	33.70	9.83	0.65	0.02	0.90	3.14
11	Listening	25	25	8,872	17.94	4.97	0.72	0.02	0.83	2.02
	Speaking	16	34	8,872	28.34	6.82	0.84	0.01	0.94	1.63
	Reading	27	27	8,872	17.86	5.10	0.67	0.03	0.82	2.19
	Writing	19	25	8,872	18.64	4.68	0.75	0.03	0.81	2.03
	Listening and Speaking	41	59	8,872	46.28	10.55	0.77	0.02	0.93	2.80
	Reading and Writing	46	52	8,872	36.50	8.90	0.70	0.02	0.88	3.03
12	Listening	25	25	6,462	19.24	4.36	0.77	0.02	0.80	1.95
	Speaking	16	34	6,462	29.92	5.44	0.88	0.01	0.92	1.50
	Reading	27	27	6,462	18.85	4.79	0.70	0.03	0.80	2.17
	Writing	19	25	6,462	19.60	4.17	0.79	0.03	0.78	1.97
	Listening and Speaking	41	59	6,462	49.16	8.52	0.82	0.02	0.90	2.64
	Reading and Writing	46	52	6,462	38.44	8.01	0.74	0.02	0.86	2.98

Note: The data file used in the table above is the final research file. The total n-counts for Grades K–12 are 208,596. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

5.6 Inter-Rater Reliability

Another source of measurement error can occur in hand-scoring. Inter-rater reliability investigates the extent to which examinees would receive the same score if they were scored either again by the same rater or by different raters. Inter-rater reliability is commonly evaluated in two ways. First, the correlation of two raters scoring each students' papers can be used as an estimate of this type of reliability. Using this approach, the consistency of ratings (e.g., a higher correlation) is the measure of reliability. Another type of evidence for inter-rater reliability is to calculate the percentage of agreement between raters. In a scenario of perfect agreement between raters, the result would be a 100% match in their assignment of scores. The lower bound to agreement rates (i.e. raters never agree with each other) is 0% agreement.

NYSED has chosen to carry out a yearly audit of the NYSESLAT for two reasons. The audit seeks to ensure that local teachers apply the same rigorous scoring standards as intended by NYSED. In addition, this audit also provides evidence of inter-rater reliability. To achieve this, the Department requires Questar to rescore 10% of all test papers after each test administration. This 10% audit sample within each grade span is selected as a stratified random sample based on the Need/Resource Category (NRC). The NRC for the State schools is divided 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 and were 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 2011 administration was used to select schools for the sampling and analysis, with the goal of acquiring at least 95%, but no more than 105%, of the target count for each grade span. These schools returned their booklets to Questar for rescoring. The following sections (or tables) indicate the procedures that Questar conducted to ensure reliability and accurate scoring of the items.

To perform the audit, scoring personnel and team leaders are usually the same people who have participated in the range-finding process and are familiar with the New York State Test. All training is done using the same scoring materials as those used by the New York State teachers for scoring the operational test. Readers must qualify for scoring the audit by scoring sample papers prior to actual audit scoring. Following this, readers are rechecked by team leaders to maintain their accuracy.

Table 14 provides the audit sample information regarding the number of students obtained for the rater analyses after merging the auditing file and final research file. The audit data and the research file were merged by student ID (SID). Questar selected the stratified random sample to achieve the desired percentage (10%) of the full population.

Table 14—Audit Sample

Grade Span	Number of Students in Audit Sample after Merging	Number of Students in Population by Grade Span	Percentage of Total Population
K–1	5,156	51,100	10
2–4	6,019	59,451	10
5–6	3,009	29,500	10
7–8	2,566	25,294	10
9–12	4,192	43,249	10
Total	20,942	208,594	10

Inter-Rater Agreement

Table 15 provides the rater agreement for the Pre-Writing and Writing constructed-response items between local raters and Questar raters. The Questar 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 item. For Grades K–1, items 1–9 were scored 2. For Grades 2–12, items 1–3 were scored 0, 1, or 2, and item 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 Audit Raters
- **Adjacent:** +/-1 Score Point Difference between Local and Audit Raters
- **Non-Adjacent:** +/-2 Score Point Difference between Local and Audit Raters

In Table 15, for the 2-point items, the percent of exact rating ranged from 65.30 to 96.39 across all the grade spans. The percent of adjacent ratings for 2-point items ranged from 3.06 to 32.14. The total percent of the exact and adjacent ratings ranged from 97.44 to 99.92. 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 for 4-point items ranged from 47.09 to 57.08 across the four grade spans. In addition, the percent of adjacent ratings for 4-point items ranged from 38.97 to 44.47. The total percent of the exact and adjacent ratings for the four-point items ranged from 91.56 to 96.05.

In addition to agreement ratings, Table 15 also includes the intraclass correlations between the local raters and Questar raters, the weighted Kappa coefficient, the asymptotic standard error (ASE) of the weighted Kappa coefficient, and the lower and upper 95 percent confidence limits of the Kappa coefficient.

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 intraclass correlations in Table 15 were calculated by using the random and fixed effects model introduced by Shrout and Fleiss (1979) to measure the agreement of the local and Questar 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 0.61 to 0.95 across all the grade spans. The lowest correlations occurred at the 5–6 grade span.

Kappa Coefficient

The Kappa coefficient is an index measuring the rater agreement. The value of Kappa is always less than 1. A value of 1 signifies perfect agreement while values less than 1 indicate less than perfect agreement. In extremely rare situations, a weighted Kappa value can be negative. This is a sign that the two observers agreed less than would be expected just by chance. It is also rare that there is perfect agreement. There is not complete agreement on the interpretation as to what degree of agreement can be considered an acceptable level of agreement. One guideline to the interpretation of the weighted Kappa coefficient is as follows (Altman, D.G. *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 15, the values of the weighted Kappa across grade spans indicate that the rater agreement ranges within moderate agreement to good agreement, using the interpretation rule above. The asymptotic standard error (ASE) of the Weighted Kappa coefficient and its lower and upper 95 percent confidence limits indicate that the errors of the weighted Kappa coefficient are very small and within a limited range.

Table 15—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	5,156	83.55	16.35	0.10	0.68	0.65	0.01	0.63	0.67
	2	2	5,156	82.33	17.59	0.08	0.93	0.67	0.01	0.65	0.69
	3	2	5,156	96.39	3.06	0.54	0.85	0.78	0.02	0.75	0.82
	4	2	5,156	94.98	4.29	0.74	0.85	0.77	0.02	0.74	0.80
	5	2	5,156	92.73	6.86	0.41	0.95	0.88	0.01	0.87	0.89
	6	2	5,156	89.24	10.16	0.60	0.89	0.84	0.01	0.83	0.85
	7	2	5,156	78.78	21.06	0.16	0.92	0.75	0.01	0.74	0.77
	8	2	5,156	76.67	23.02	0.31	0.81	0.73	0.01	0.71	0.74
	9	2	5,156	75.70	23.97	0.33	0.80	0.71	0.01	0.70	0.73
2-4	1	2	6,018	91.58	7.85	0.58	0.79	0.73	0.01	0.71	0.76
	2	2	6,018	85.04	14.04	0.91	0.81	0.75	0.01	0.73	0.76
	3	2	6,018	79.88	19.51	0.61	0.78	0.72	0.01	0.70	0.73
	4	4	6,018	57.08	38.97	3.95	0.71	0.56	0.01	0.55	0.58
5-6	1	2	3,009	88.33	11.10	0.56	0.77	0.69	0.02	0.66	0.73
	2	2	3,009	86.37	13.09	0.53	0.83	0.76	0.01	0.73	0.78
	3	2	3,009	65.30	32.14	2.56	0.61	0.53	0.01	0.50	0.55
	4	4	3,009	50.28	42.91	6.81	0.65	0.49	0.01	0.47	0.51
7-8	1	2	2,566	76.93	21.09	1.99	0.76	0.69	0.01	0.66	0.71
	2	2	2,566	92.32	6.63	1.05	0.87	0.84	0.01	0.82	0.86
	3	2	2,566	76.46	21.83	1.71	0.74	0.66	0.01	0.64	0.69
	4	4	2,566	55.69	39.24	5.07	0.75	0.60	0.01	0.58	0.62
9-12	1	2	4,190	85.97	12.70	1.34	0.78	0.71	0.01	0.69	0.73
	2	2	4,190	84.20	14.27	1.53	0.77	0.70	0.01	0.68	0.72
	3	2	4,190	70.69	26.85	2.46	0.63	0.56	0.01	0.53	0.58
	4	4	4,190	47.09	44.47	8.45	0.68	0.51	0.01	0.49	0.53

Note: The statistics in the table are based on the merged file between audit and Questar's final research file. The audit data and research file were merged by SIDs.

Table 16 provides the frequency distribution of the score point differences between the local raters' and the audit raters' scores by grade span. For the 2-point items, the percentage of 0 differences ranged from 65.30 to 96.39. For the 4-point items, the percentage of 0 differences ranged from 47.09 to 57.08. The above evidence indicates that the rater-agreement percentage is higher for the 2-point item than for the 4-point item.

Table 16—Percentages of the Score Difference Between Raters

Grade Span	Item	Percentage of Point Difference								
		Score Difference (Local score minus Questar score)								
		-4	-3	-2	-1	0	1	2	3	4
K-1	1			0.02	6.57	83.55	9.78	0.08		
	2			0.06	4.23	82.33	13.36	0.02		
	3			0.14	0.64	96.39	2.42	0.41		
	4			0.31	0.60	94.98	3.69	0.43		
	5			0.16	1.57	92.73	5.29	0.25		
	6			0.17	1.51	89.24	8.65	0.43		
	7			0.04	8.90	78.78	12.16	0.12		
	8			0.06	5.33	76.67	17.69	0.25		
	9			0.10	9.50	75.70	14.47	0.23		
2-4	1			0.45	5.32	91.58	2.53	0.13		
	2			0.60	8.91	85.04	5.13	0.32		
	3			0.28	10.82	79.88	8.69	0.33		
	4	0.03	0.17	2.36	21.27	57.08	17.70	1.38	0.02	
5-6	1			0.20	5.08	88.33	6.02	0.37		
	2			0.23	5.08	86.37	8.01	0.30		
	3			0.50	7.91	65.30	24.23	2.06		
	4			0.76	11.57	50.28	31.34	5.72	0.30	0.03
7-8	1			1.40	10.41	76.93	10.68	0.58		
	2			0.70	3.86	92.32	2.77	0.35		
	3			1.17	8.38	76.46	13.45	0.55		
	4	0.04	0.23	1.99	18.12	55.69	21.12	2.61	0.16	0.04
9-12	1			0.53	7.88	85.97	4.82	0.81		
	2			0.76	8.33	84.20	5.94	0.76		
	3			1.03	11.12	70.69	15.73	1.43		
	4		0.17	1.41	13.63	47.09	30.84	6.35	0.50	0.02

Table 17 provides the mean and standard deviation of each item for: 1) the local raters, 2) the Questar raters, and 3) the difference between local and Questar raters. The first pair of columns contains the mean and standard deviation of the local raters. The next pair of columns are the mean and standard deviation of the Questar raters. The final pair of columns contains the mean and standard deviation of the differences between the local and Questar raters. The largest mean difference (0.33) occurred at grade span 5-6 (item 4). The SD ratio ranged from 0.89 to 1.18, with the largest SD ratio occurring in grade span 5-6 (item 4). For all items, the mean difference is close to 0, and SD ratio is close to 1, which is a good indication of rater agreement.

Table 17—Comparison Between Local and Audit Raters

Grade Span	Item	Local		Questar		Differences	
		Mean	SD	Mean	SD	Mean	SD Ratio
K-1	1	1.67	0.50	1.64	0.51	0.03	0.98
	2	1.56	0.53	1.47	0.53	0.09	1.00
	3	1.91	0.39	1.89	0.41	0.02	0.95
	4	1.88	0.46	1.85	0.49	0.03	0.94
	5	1.61	0.69	1.57	0.73	0.04	0.95
	6	1.53	0.70	1.45	0.79	0.08	0.89
	7	1.21	0.81	1.18	0.79	0.03	1.03
	8	1.18	0.80	1.06	0.78	0.12	1.03
	9	1.00	0.79	0.95	0.78	0.05	1.01
2-4	1	1.79	0.51	1.82	0.47	-0.03	1.09
	2	1.55	0.69	1.59	0.67	-0.04	1.03
	3	1.38	0.70	1.40	0.71	-0.02	0.99
	4	1.99	1.04	2.05	0.96	-0.06	1.08
5-6	1	1.77	0.53	1.76	0.54	0.01	0.98
	2	1.64	0.66	1.60	0.67	0.04	0.99
	3	1.35	0.73	1.16	0.73	0.19	1.00
	4	2.17	1.07	1.86	0.91	0.31	1.18
7-8	1	1.37	0.79	1.38	0.77	-0.01	1.03
	2	1.66	0.66	1.67	0.64	-0.01	1.03
	3	1.44	0.75	1.40	0.73	0.04	1.03
	4	2.08	1.18	2.04	1.06	0.04	1.11
9-12	1	1.66	0.64	1.69	0.63	-0.03	1.02
	2	1.61	0.66	1.64	0.66	-0.03	1.00
	3	1.45	0.70	1.39	0.70	0.06	1.00
	4	2.44	1.18	2.16	1.07	0.28	1.10

Summary

Inter-rater reliability statistics and evidence collected and presented here demonstrate that the local raters and Questar raters in general have a high degree of agreement on the NYSESLAT scores. There are cases of a few items of lesser agreement (e.g., item 4 in grade span 5–6). In the cases of greater disagreement, the local raters generally gave higher scores than those given by Questar raters.

5.7 Reliability of Classification Decision at Proficient Cut

The NYSESLAT scale scores are used to categorize student performance into one of four proficiency levels (see Chapter 8 for details on standard setting). Although it is extremely important to know that student scores are highly reliable in any examination, it is perhaps of even greater import to evaluate the reliability of the decisions based on these scores. The previous vendor assessed the reliability of classification decisions via an analysis of the probabilities of correct and consistent classification of student performance. Livingston and Lewis (1995) developed procedures to do this and these procedures were applied to derive measures of the accuracy and consistency of the classifications for the NYSESLAT. The brief descriptions of the procedures used and results obtained are presented in this section.

The accuracy of decisions is defined as 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 those that would have been made if the students had taken a parallel form of the NYSESLAT, which would be defined as equal in difficulty and covering the same content as the form they actually took. These notions are graphically displayed in Figures 1 and 2.

		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 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 2—Classification Consistency

In Figure 1, accurate classifications occur when the decision made based on the average of all-theoretical forms average agrees with the decision made on the form actually taken. Misclassifications are defined as the case where a student who achieves a score of “Does Not Achieve Proficient Status” for his or her all-theoretical forms average is classified incorrectly as “Achieves Proficient Status”. Consistent classification occurs (Figure 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.

Decision consistency and accuracy techniques are outlined and implemented by Hanson (1991), Haertel (1996), Livingston and Lewis (1995), and Young and Yoon (1998). The previous vendor used the BB-Class software developed by Hanson (1995) to conduct these 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. Since the examination was not changed from the 2009 edition, the decision consistency and accuracy numbers also remain the same and, therefore, this section is repeated from the previous technical report.

Table 18 contains the proportions of False Positive and False Negative classifications. The sum of values of Accuracy, False Positive, and False Negative should be equal to 1.00. However, due to rounding, the table values may not always equal 1.00. False Positive and False Negative classifications refer to the mismatch between student true scores and observed scores. The proportion of student scores misclassified into the category of Achieves Proficient Status are labeled as False Positive. The proportion of student scores misclassified into the category Does Not Achieve Proficient Status when student scores meet proficient status is contained in the False Negative column.

Table 18 presents the results of the decision accuracy and consistency for the Achieves Proficient Status cut scores by the paired modalities of Listening/Speaking and Reading/Writing. The table has the following information:

- 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, 89% of the classifications for Listening/Speaking at the proficiency level were accurate, while 6% percent of the students were classified as proficient when they were not. Five percent of students were classified as not proficient when they were proficient. For Listening/Speaking combination proficiency-level classification, the accuracy of the classification ranged from 75% to 92% across all grade levels. For Reading/Writing combination proficiency-level classification, the accuracy ranged from 79% to 95% 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 based on an alternate form. For example, for K students, the chance that students were classified the same was 81% for Listening/Speaking. For the Listening/Speaking combination proficiency-level classifications, the consistency ranged from 72% to 89% across all the grade levels. For the Reading/Writing combination proficiency-level classifications, the consistency ranged from 77% to 93% across all the grade levels. In most cases, 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 lower than decision accuracy. This is understandable because 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 3% to 10% for Listening/Speaking and 3% to 21% for Reading/Writing across all the grade levels. The false negative ranged from 3% to 16% for Listening/Speaking and 0% to 10% for Reading/Writing across all the grade levels. Relatively speaking, more attention should be paid to the high false positive values because it is less acceptable to see a high percentages of students who were not proficient classified as proficient and exit the program. The NYSESLAT shows

much higher percentages of accuracy and consistency in comparison to the false positive and negative percentages.

Table 18—Classification Accuracy and Consistency by Grade

Grade	Test	Accuracy	False Positives	False Negatives	Consistency
K	Listening and Speaking	0.89	0.06	0.05	0.85
	Reading and Writing	0.95	0.03	0.01	0.93
1	Listening and Speaking	0.77	0.07	0.16	0.73
	Reading and Writing	0.85	0.05	0.10	0.82
2	Listening and Speaking	0.90	0.04	0.06	0.86
	Reading and Writing	0.92	0.05	0.02	0.89
3	Listening and Speaking	0.90	0.05	0.05	0.86
	Reading and Writing	0.87	0.08	0.06	0.83
4	Listening and Speaking	0.87	0.03	0.10	0.81
	Reading and Writing	0.79	0.21	0.00	0.77
5	Listening and Speaking	0.85	0.06	0.09	0.78
	Reading and Writing	0.87	0.05	0.07	0.82
6	Listening and Speaking	0.75	0.10	0.14	0.72
	Reading and Writing	0.89	0.06	0.05	0.84
7	Listening and Speaking	0.89	0.04	0.07	0.85
	Reading and Writing	0.92	0.04	0.04	0.88
8	Listening and Speaking	0.91	0.04	0.05	0.87
	Reading and Writing	0.90	0.04	0.05	0.86
9	Listening and Speaking	0.92	0.03	0.04	0.89
	Reading and Writing	0.93	0.04	0.03	0.90
10	Listening and Speaking	0.92	0.05	0.03	0.88
	Reading and Writing	0.91	0.06	0.02	0.88
11	Listening and Speaking	0.88	0.06	0.06	0.83
	Reading and Writing	0.89	0.07	0.03	0.86
12	Listening and Speaking	0.88	0.06	0.06	0.83
	Reading and Writing	0.89	0.07	0.03	0.85

Note: The data file used in the 2009 analysis from above was conducted upon the final research file and was not redone for the 2011 administration since the item parameters were maintained. The total n-count for Grades K–12 in that data file was 218,077. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score=0 or >900.) Cases with missing grades are also removed.

CHAPTER 6: VALIDITY

Assessments constructed by Questar, as well as other vendors, support the standards as set forth in the Standards for Educational and Psychological Testing. Specifically, the standards related to validity are supported in several manners. The judgments made in this technical report regarding test validity are based upon the following:¹

- 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

Test content as evidence of validity is evidenced by the relationship that test material and items have in that they should represent an appropriate sampling of the knowledge, skills, and understanding of the construct or domain being assessed. The definition of those skills, knowledge, and understanding comes from *The New York State Learning Standards for English as a Second Language*. As such, the Standards define what is expected of ELL students in New York State. To ensure the content validity of the assessment, multiple steps are taken to align the NYSESLAT items and test forms to these standards.

Psychometricians and ELL assessment specialists, with input from NYSED assessment and content specialists, developed a test blueprint (see Chapter 2 for further details) that includes a range of item types that can measure the appropriate (in terms of which standards and performance indicators lend themselves to assessment on a standardized test) standards and performance indicators from the *New York State Learning Standards for English as a Second Language*. This determination was made by ELL assessment specialists, working with committees of New York State ESL teachers in order to identify which performance indicators are suited to classroom use only and consequently would not become part of the assessment.

The item writers for NYSESLAT were given item-writing assignments and then instructed to use NYSESLAT item type specifications and the *New York State Learning Standards for English as a Second Language* to write their items. Specifically, they were asked to use these materials to guide their writing and then identify which standard and performance indicator was a best match

¹ The page number in the parenthesis is the page number in the Standards for Educational and Psychological Testing, 1999.

to what the items were intended to measure. Items were then reviewed by committees of New York State teachers as part of the review process. This included discussion and agreement upon the standards and performance indicators assigned to each item.

Performance indicators and standards were part of several criteria (including item statistics, content diversity, gender and ethnic balance of item subjects, etc.) that were taken into consideration by ELL assessment specialists and NYSED assessment and content specialists during forms construction to ensure that forms reflected broad coverage of the *Learning Standards*. Following forms construction, assessment specialists developed item maps for each form to show this coverage. ELL assessment specialists 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*. This item mapping is detailed both in Table 19 and in Appendices A.1–A.2 as evidence for the alignment to the *New York State Learning Standards for English as a Second Language*.

6.2 Internal Structure

A coherent assessment is not a random collection of assessment tasks or test questions. The tasks selected for inclusion in an assessment are intended to contribute positively to the total result. One way of providing evidence for this is to assess the interrelationship of the tasks for the test. This is referred to as the internal structure of the assessment. The following questions are often posed to investigate the internal structure of education assessments (Nitko, 2004):

- 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?

Correlations based on raw scores of the four modalities are calculated to investigate the answers to these questions. The intercorrelation of the four modalities (by grade span and grade) is found in Tables 19 and 20. The evidence of internal structure of the 2011 NYSESLAT can also be illustrated by examining the point-biserial correlation coefficient and fit statistics. Appendices B.1–B.5 and D.1–D.5 provide these statistics for the 2011 NYSESLAT administration.

Table 19—Intercorrelation Among the Modalities by Grade Span

Grade Span	Test	Correlation Coefficient			
		Listening	Speaking	Reading	Writing
K-1	Listening	1.00			
	Speaking	0.67	1.00		
	Reading	0.60	0.50	1.00	
	Writing	0.64	0.57	0.78	1.00
2-4	Listening	1.00			
	Speaking	0.68	1.00		
	Reading	0.70	0.57	1.00	
	Writing	0.72	0.66	0.80	1.00
5-6	Listening	1.00			
	Speaking	0.68	1.00		
	Reading	0.72	0.63	1.00	
	Writing	0.74	0.74	0.79	1.00
7-8	Listening	1.00			
	Speaking	0.66	1.00		
	Reading	0.72	0.56	1.00	
	Writing	0.73	0.74	0.74	1.00
9-12	Listening	1.00			
	Speaking	0.66	1.00		
	Reading	0.68	0.54	1.00	
	Writing	0.73	0.73	0.70	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 208,596. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900). Additionally, cases with missing grade are also removed.

Table 20 - Intercorrelation Among the Modalities by Grade

Grade	Test	Correlation Coefficient			
		Listening	Speaking	Reading	Writing
K	Listening	1.00			
	Speaking	0.63	1.00		
	Reading	0.51	0.39	1.00	
	Writing	0.57	0.49	0.70	1.00
1	Listening	1.00			
	Speaking	0.67	1.00		
	Reading	0.59	0.54	1.00	
	Writing	0.61	0.59	0.76	1.00
2	Listening	1.00			
	Speaking	0.64	1.00		
	Reading	0.64	0.53	1.00	
	Writing	0.66	0.61	0.76	1.00

Table 20—Intercorrelation Among the Modalities by Grade (continued)

Grade	Test	Correlation		Coefficient	
		Listening	Speaking	Reading	Writing
3	Listening	1.00			
	Speaking	0.69	1.00		
	Reading	0.67	0.58	1.00	
	Writing	0.71	0.68	0.78	1.00
4	Listening	1.00			
	Speaking	0.74	1.00		
	Reading	0.70	0.63	1.00	
	Writing	0.74	0.73	0.79	1.00
5	Listening	1.00			
	Speaking	0.68	1.00		
	Reading	0.72	0.62	1.00	
	Writing	0.74	0.74	0.80	1.00
6	Listening	1.00			
	Speaking	0.69	1.00		
	Reading	0.72	0.63	1.00	
	Writing	0.74	0.75	0.79	1.00
7	Listening	1.00			
	Speaking	0.66	1.00		
	Reading	0.71	0.56	1.00	
	Writing	0.73	0.74	0.74	1.00
8	Listening	1.00			
	Speaking	0.66	1.00		
	Reading	0.72	0.58	1.00	
	Writing	0.74	0.75	0.74	1.00
9	Listening	1.00			
	Speaking	0.70	1.00		
	Reading	0.70	0.57	1.00	
	Writing	0.76	0.77	0.71	1.00
10	Listening	1.00			
	Speaking	0.65	1.00		
	Reading	0.68	0.53	1.00	
	Writing	0.71	0.70	0.69	1.00
11	Listening	1.00			
	Speaking	0.59	1.00		
	Reading	0.64	0.48	1.00	
	Writing	0.66	0.64	0.66	1.00
12	Listening	1.00			
	Speaking	0.51	1.00		
	Reading	0.60	0.39	1.00	
	Writing	0.59	0.54	0.60	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 208,596. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900). Additionally, cases with missing grade are also removed.

Observations of language proficiency assessment subtests in Table 19:

- Listening and Speaking are moderately correlated for grade spans ranging from 0.66–0.68.
- Listening and Reading are moderately correlated for grade spans ranging from 0.60–0.72.
- Listening and Writing are moderately correlated for grade spans ranging from 0.64–0.74.
- Speaking and Reading are moderately correlated for grade spans ranging from 0.54–0.63.
- Speaking and Writing are moderately correlated for grade spans ranging from 0.57–0.74.
- Reading and Writing are moderately correlated for grade spans ranging from 0.70–0.80.
- The language domain pairs of Listening and Speaking and Reading and Writing are moderately to highly correlated, while Speaking and Reading are not as highly correlated.

The results of these internal correlations follow what is theoretically expected. The Reading and Writing subtests have the highest correlation range (from 0.78 to 0.80) across the five grade spans. The Speaking and Reading subtests have the lowest correlation range (from 0.50 to 0.63).

The following bullet lists contain some observations regarding the NYSESLAT subtests and the results found in Table 20:

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 20, for Kindergarten students, Speaking and Listening have a correlation of 0.63, and Reading and Writing have a correlation of 0.70. Speaking and Reading have a low correlation of 0.39, Listening and Reading have a low correlation of 0.51, Speaking and Writing have a correlation of 0.49, and Listening and Writing have a correlation of 0.57.

Grades 1–8

- A trend of increase in the correlation (from 0.54 to 0.71) between Listening and Writing is observed. The previous vendor found an exception in Grade 2, but this did not occur in the 2011 data.
- A possible explanation for this trend is that students in this age span are required to expand the use of and development of their Writing skills. At the same time, there are limited increases in the demands placed upon their Listening skills.

Grades 9–12

- A steady decrease in the correlation (from 0.76 to 0.59) between Listening and Writing is observed.
- One plausible explanation is that high school content contains an increased focus on the use of writing skills, especially an increased focus on academic content. Listening skills may also increase, but not to the same degree as the Writing demands require.
- Similar arguments may be made for the correlation behavior between Speaking and Writing in Grades 1–12. There is a trend of increase in the correlation (from 0.49 to 0.75) for Grades 1–8, and there is a steady decrease in the correlation from (0.77 to 0.54) for Grades 9–12.

Again, the internal correlations, and thus the internal structure of the test, for the most part follow expectations. The Reading and Writing subtests have the highest correlation range (from 0.60 to 0.80) across the thirteen grades. Speaking and Reading have the lowest correlation range (from 0.39 to 0.63).

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

A confirmatory factor analysis was not conducted in 2011 since the assessment was a reprint of the 2009 assessment. However, the validity of those inferences regarding scoring structure (i.e., Speaking and Listening as one score and Reading and Writing as the other) and a confirmatory factor analysis undertaken with the 2009 full population data remain relevant evidence of the unidimensional nature of model (i.e., the total test providing a single score). What follows is a reprint of the 2009 Confirmatory Factor study conducted by the previous vendor.

The 2009 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 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 21–25 below; Appendix F provides a reprint of the 2009 CFA analyses in their entirety.)

Table 21—Global Fit Indices for the One- and Two-Factor Models (Grades K–1)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	7963.838	535.992
GFI	0.930	0.995
AGFI	0.652	0.975
RMR	0.057	0.025
RMSEA	0.274	0.071

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.

Table 22—Global Fit Indices for the One- and Two-Factor Models (Grades 2–4)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	5855.534	2107.898
GFI	0.957	0.984
AGFI	0.783	0.920
RMR	0.036	0.034
RMSEA	0.213	0.128

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.

Table 23—Global Fit Indices for the One- and Two-Factor Models (Grades 5–6)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	1572.171	2500.706
GFI	0.974	0.959
AGFI	0.870	0.797
RMR	0.023	0.093
RMSEA	0.163	0.206

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.

Table 24—Global Fit Indices for the One- and Two-Factor Models (Grades 7–8)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	1661.807	2243.666
GFI	0.968	0.958
AGFI	0.842	0.789
RMR	0.025	0.062
RMSEA	0.181	0.210

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.

Table 25—Global Fit Indices for the One- and Two-Factor Models (Grades 9–12)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	2325.567	2606.121
GFI	0.972	0.969
AGFI	0.862	0.846
RMR	0.026	0.042
RMSEA	0.168	0.178

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 as one progressed from the lower grade spans to the higher ones.

6.3 External Structure

An additional source of validity evidence that can be assessed is sometimes called external criterion validity. To provide evidence of this, the external structure of the NYSESLAT with an external criterion was assessed. In particular, the data from two other tests were used. For ELL students in Grades 3–8, data from those who took both the NYSESLAT in May 2011, and the New York State English Language Assessment (ELA) Test in January 2011 were examined. In New York State, all ELL students, with the exception of first year ESL students, must take the ELA test. For ELL students in Grades 10–12, data from those who took both the NYSESLAT in May 2011 and the New York State Regents Examination in Comprehensive English (Regents English exam) in June 2011 were analyzed. The Regents English exam is an end-of-course exam that all high school students, including ELL students, are required to pass as a high school graduation requirement. Students may take the exam in any grade, but most students take the exam in Grade 11.

Since the NYSESLAT is intended to be a measure of annual student progress in the achievement of English language proficiency, it should be expected that ELL students who score proficient on the NYSESLAT will be able to perform well in mainstream classrooms as well. Therefore, they should in theory have a similar chance to pass the Regents Examinations as native English speakers who are required to take the state examinations. Hence, there should be a positive relationship between the NYSESLAT and the ELA Test/Regents English exam wherein those who perform well on the NYSESLAT are generally expected to perform well on the ELA test or the Regents English examinations.

6.3.1 Relationship with the New York English Language Arts Test, Grades 3–8

Correlations and Other Descriptive Statistics

Table 26 gives the sample size, minimum and maximum observed scores, means, standard deviations, and correlation between the New York State ELA Test and the NYSESLAT raw score total of Reading, Writing, and Listening for each grade from Grade 3 to Grade 8. Following this, Table 27 contains the sample size, minimum and maximum observed scores, means, standard deviations, and correlation between the ELA Test and the NYSESLAT combined scale scores of Listening and Speaking, and Reading and Writing, for each grade from Grade 3 to Grade 8.

The correlations between the ELA Test and the NYSESLAT raw score total of the three modalities (Reading, Writing, and Listening) ranged from 0.63 to 0.75 across the six grades. The correlations are positive between the two tests, which is logical since the modalities tested (Reading, Writing, and Listening) are the same as those in the ELA Test. The correlations between New York ELA test and the NYSESLAT combined scale scores of the four modalities ranged from 0.60 to 0.69 in grades 3–8. The correlations in Table 27 are positive, but slightly lower than the correlations in Table 26, which can be explained by the fact that the NYSESLAT total combined scale scores include the Speaking modality, while the ELA Test does not include a Speaking portion.

In sum, one notes a positive relationship between the NYSESLAT and the ELA Tests. This provides evidence of the external structure check and demonstrates a positive relationship

(moderate) between the two examinations (see Table 26 and Table 27). One can reasonably conclude that higher scores on the NYSESLAT are associated with higher scores on the ELA Test for each grade.

Table 26—Descriptive Statistics of the NYSESLAT Raw Scores of All Modalities and Their Correlation with the ELA Test Total Scale Scores

Grade	N	Test	Max Points	Mean	SD	Sample Min.	Sample Max.	Correlation with ELA
03	17,938	Listening	24	20.33	3.08	0	24	0.58
03	17,938	Speaking	34	30.21	4.65	0	34	0.43
03	17,938	Reading	24	18.39	3.84	0	24	0.69
03	17,938	Writing	22	16.52	3.93	0	22	0.69
03	17,938	LI/RD/WR	70	55.23	9.50	0	70	0.75
04	16,067	Listening	24	21.28	2.78	0	24	0.56
04	16,067	Speaking	34	30.80	4.40	0	34	0.44
04	16,067	Reading	24	19.66	3.57	0	24	0.66
04	16,067	Writing	22	17.71	3.65	0	22	0.65
04	16,067	LI/RD/WR	70	58.65	8.75	0	70	0.72
05	13,667	Listening	25	19.96	3.49	0	25	0.58
05	13,667	Speaking	34	30.26	4.99	0	34	0.46
05	13,667	Reading	27	20.43	5.28	0	27	0.69
05	13,667	Writing	25	18.57	4.35	0	25	0.67
05	13,667	LI/RD/WR	77	58.97	11.65	0	77	0.73
06	11,563	Listening	25	19.97	3.64	0	25	0.52
06	11,563	Speaking	34	30.00	5.55	0	34	0.43
06	11,563	Reading	27	20.38	5.36	0	27	0.62
06	11,563	Writing	25	18.62	4.56	0	25	0.61
06	11,563	LI/RD/WR	77	58.98	12.04	0	77	0.66
07	10,427	Listening	25	18.04	4.24	0	25	0.53
07	10,427	Speaking	34	29.29	6.74	0	34	0.43
07	10,427	Reading	27	17.95	5.73	0	27	0.58
07	10,427	Writing	25	17.77	4.80	0	25	0.56
07	10,427	LI/RD/WR	77	53.76	13.07	0	77	0.63
08	10,361	Listening	25	18.56	4.41	0	25	0.55
08	10,361	Speaking	34	29.11	6.99	0	34	0.45
08	10,361	Reading	27	18.84	5.71	0	27	0.61
08	10,361	Writing	25	18.07	4.92	0	25	0.58
08	10,361	LI/RD/WR	77	55.48	13.37	0	77	0.66
2-4	34,005	Listening	24	20.78	2.98	0	24	0.55
2-4	34,005	Speaking	34	30.49	4.54	0	34	0.43
2-4	34,005	Reading	24	18.99	3.77	0	24	0.65
2-4	34,005	Writing	22	17.08	3.85	0	22	0.65
2-4	34,005	LI/RD/WR	70	56.85	9.31	0	70	0.71
5-6	25,230	Listening	25	19.97	3.56	0	25	0.54
5-6	25,230	Speaking	34	30.14	5.26	0	34	0.44
5-6	25,230	Reading	27	20.41	5.32	0	27	0.64
5-6	25,230	Writing	25	18.59	4.45	0	25	0.62
5-6	25,230	LI/RD/WR	77	58.97	11.83	0	77	0.68
7-8	20,788	Listening	25	18.30	4.34	0	25	0.50
7-8	20,788	Speaking	34	29.20	6.87	0	34	0.43
7-8	20,788	Reading	27	18.40	5.74	0	27	0.55
7-8	20,788	Writing	25	17.92	4.86	0	25	0.54
7-8	20,788	LI/RD/WR	77	54.62	13.25	0	77	0.60

Table 27—Descriptive Statistics of the New York State ELA Test and NYSESLAT Combined Scale Scores of Listening/Speaking and Reading/Writing

Grade	N	Test	Mean	SD	Sample Min.	Sample Max.	Correlation
3	17,938	NYSESLAT Scale Score	1337	77.34	762	1,611	0.68
		NYS ELA Scale Score	647	22.71	475	714	
4	16,067	NYSESLAT Scale Score	1370	82.57	972	1,611	0.65
		NYS ELA Scale Score	647	32.34	430	745	
5	13,667	NYSESLAT Scale Score	1375	74.99	828	1,657	0.69
		NYS ELA Scale Score	649	20.91	495	710	
6	11,563	NYSESLAT Scale Score	1376	78.59	828	1,657	0.63
		NYS ELA Scale Score	641	17.66	480	694	
7	10,427	NYSESLAT Scale Score	1378	78.08	861	1,683	0.60
		NYS ELA Scale Score	639	23.01	470	707	
8	10,361	NYSESLAT Scale Score	1388	84.14	1026	1,683	0.63
		NYS ELA Scale Score	627	21.82	430	695	

Proficiency Level

As an additional way to demonstrate the positive relationship of student success on both exams, a crosstabs of proficiency level of the ELA Test and NYSESLAT was performed. Tables 28–33 contain the percentage distributions of the ELA proficiency level by the NYSESLAT proficiency level at each grade level.

Table 28 shows that for those Grade 3 students who were classified as beginning students on the NYSESLAT, nearly 100% 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% 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, 81% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, while 19% of them were classified as proficient on the ELA Test. For those students who were classified as proficient students on the NYSESLAT, 44% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 56% of them were classified as proficient on the ELA Test.

In Table 29, for those Grade 4 students who were classified as beginning students on the NYSESLAT, nearly 100% 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% 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, 84% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 16% of them were classified as proficient on the ELA Test. For those students who were classified as proficient students on the NYSESLAT, 55% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 45% of them were classified as proficient on the ELA Test.

Table 30 shows that for those Grade 5 students who were classified as beginning students on the NYSESLAT, 99% 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% 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, 93% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 7% of them were classified as proficient on the ELA Test. For those students who were classified as proficient students on the NYSESLAT, 68% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 32% of them were classified as proficient on the ELA Test.

For Grade 6 students, the results are summarized in Table 31. This table shows that for those Grade 6 students who were classified as beginning students on the NYSESLAT, 100% 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, 100% 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, 96% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 4% of them were classified as proficient on the ELA Test. For those students who were classified as proficient students on the NYSESLAT, 82% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 18% of them were classified as proficient on the ELA Test.

In the next table (Table 32), the classifications for Grade 7 students are displayed. Table 32 shows that for those Grade 7 students who were classified as beginning students on the NYSESLAT, 100% 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, 100% 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, 98% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 2% of them were classified as proficient on the ELA Test. For those students who were classified as proficient students on the NYSESLAT, 85% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 15% of them were classified as proficient on the ELA Test.

Table 33 shows that for those Grade 8 students who were classified as beginning students on the NYSESLAT, 100% 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, 100% 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, 98% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 2% of them were classified as proficient on the ELA Test. For those students who were classified as proficient

students on the NYSESLAT, 85% of them were classified as not proficient (Level 1 and Level 2) on the ELA Test, and 15% of them were classified as proficient on the ELA Test.

Between 97%–100% of students who were classified as beginning and intermediate on the NYSESLAT were also classified as not proficient (Level 1 and Level 2) by the ELA cut scores across the six grade levels. About 81%–98% of students who were classified as advanced on the NYSESLAT were also classified as not proficient (Level 1 and Level 2) by the ELA cut scores across the six grade levels. About 2%–19% of students who were classified as advanced on the NYSESLAT were classified as proficient by the ELA cut scores across the six grade levels. The higher the grade level, the lower the percentage. Approximately 32%–85% of students who were classified as proficient on the NYSESLAT were classified as not proficient by the ELA cut scores across the six grade levels. About 15%–68% of students who were classified as proficient on the NYSESLAT were also classified as proficient by the ELA cut scores across the six grade levels. The higher the grade level, the lower the percentage.

Table 28—Grade 3: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=1,174)	Intermediate (N=3,839)	Advanced (N=9,493)	Proficient (N=3,432)
Level 1 on ELA	98.30%	75.88%	19.65%	2.65%
Level 2 on ELA	1.45%	22.19%	60.98%	41.17%
Level 3 on ELA	0.26%	1.93%	19.25%	55.59%
Level 4 on ELA	0.00%	0.00%	0.13%	0.58%
Total	100%	100%	100%	100%

Table 29—Grade 4: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=863)	Intermediate (N=3,456)	Advanced (N=7,534)	Proficient (N=4,214)
Level 1 on ELA	95.37%	63.72%	13.60%	1.95%
Level 2 on ELA	4.40%	34.23%	70.65%	53.37%
Level 3 on ELA	0.23%	2.03%	15.72%	44.42%
Level 4 on ELA	0.00%	0.03%	0.03%	0.26%
Total	100%	100%	100%	100%

Table 30—Grade 5: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=985)	Intermediate (N=2,385)	Advanced (N=6,016)	Proficient (N=4,281)
Level 1 on ELA	96.95%	80.80%	30.87%	5.23%
Level 2 on ELA	2.44%	18.36%	62.02%	62.72%
Level 3 on ELA	0.61%	0.84%	7.10%	31.60%
Level 4 on ELA	0.00%	0.00%	0.02%	0.44%
Total	100%	100%	100%	100%

Table 31—Grade 6: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=1,025)	Intermediate (N=2,550)	Advanced (N=5,385)	Proficient (N=2,603)
Level 1 on ELA	96.78%	87.37%	44.22%	14.37%
Level 2 on ELA	2.93%	12.35%	51.42%	67.35%
Level 3 on ELA	0.29%	0.27%	4.35%	18.17%
Level 4 on ELA	0.00%	0.00%	0.02%	0.12%
Total	100%	100%	100%	100%

Table 32— Grade 7: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=1,576)	Intermediate (N=3,576)	Advanced (N=3,461)	Proficient (N=1,814)
Level 1 on ELA	96.78%	87.37%	44.22%	14.37%
Level 2 on ELA	2.93%	12.35%	51.42%	67.35%
Level 3 on ELA	0.29%	0.27%	4.35%	18.17%
Level 4 on ELA	0.00%	0.00%	0.02%	0.12%
Total	100%	100%	100%	100%

Table 33—Grade 8: Percentage of English Language Learners Scoring at Each ELA Level at NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=1,589)	Intermediate (N=3,696)	Advanced (N=2,651)	Proficient (N=2,425)
Level 1 on ELA	95.05%	69.35%	25.60%	6.78%
Level 2 on ELA	4.89%	30.31%	72.49%	78.22%
Level 3 on ELA	0.06%	0.34%	1.91%	14.83%
Level 4 on ELA	0.00%	0.00%	0.00%	0.17%
Total	100%	100%	100%	100%

Summary

These analyses and resulting correlations demonstrate that there is a positive relationship between the NYSESLAT and the ELA Test (ranging from 0.60 to 0.69). Positive correlations like these are evidence of the external validity of an assessment. In addition, the positive correlations also show that higher scores on the NYSESLAT are associated with higher scores on the ELA Test. Since the ELA Test and the NYSESLAT are developed for different populations of students, these relationships will also be less than complete correlation. Hence, it is of no surprise that there is only a moderate, rather than a high, positive correlation between the two tests. Similarly, the analyses also confirm the validity of inferences made from the NYSESLAT cut scores since about 86% to 98% (across six grade levels) of those who scored at the proficiency level on the NYSESLAT were also classified as Level 2 or Level 3 (proficient) by the ELA Test. Moreover, 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.

6.3.2 Relationship with the New York State Regents Comprehensive Examination in English

Correlations and Other Descriptive Statistics

Table 34 contains the sample size, maximum score points, minimum and maximum observed scores, means, standard deviations, and correlation between the New York State Regents Comprehensive Examination in English (Regents Comp Exam in English) and each of the NYSESLAT modalities raw scores, and the total raw scores of Reading, Writing, and Listening for Grades 10 through 12. Table 35 presents the sample size, minimum and maximum observed

scores, means, standard deviations, and correlation between the Regents Comp Exam in English and the combined scale scores of Listening/Speaking and Reading/Writing modalities, and total scale scores of the four modalities (Listening/Speaking/ Reading/Writing) for Grades 10 through 12.

Table 34—Descriptive Statistics of the NYSESLAT Raw Scores of Listening, Speaking, Reading, and Writing Modalities and Their Correlations With the Regents Comp Exam in English Total Scale Scores

Grade	N	Test	Max Points	Mean	SD	Sample Min.	Sample Max.	Correlation with Regents English
10	3,452	Listening	25	17.93	4.84	1	25	0.49
10	3,452	Speaking	34	28.83	6.65	0	34	0.43
10	3,452	Reading	27	17.56	5.05	0	27	0.56
10	3,452	Writing	25	18.71	4.58	0	25	0.56
10	3,452	LI/RD/WR Total	77	54.20	12.60	12	76	0.61
11	9,829	Listening	25	18.45	4.52	0	25	0.45
11	9,829	Speaking	34	29.25	5.65	0	34	0.38
11	9,829	Reading	27	18.18	4.82	0	27	0.52
11	9,829	Writing	25	19.19	4.06	0	25	0.53
11	9,829	LI/RD/WR Total	77	55.83	11.52	0	77	0.58
12	7,363	Listening	25	18.42	4.44	0	25	0.35
12	7,363	Speaking	34	29.10	5.68	0	34	0.29
12	7,363	Reading	27	17.98	4.68	0	27	0.41
12	7,363	Writing	25	18.84	4.15	0	25	0.41
12	7,363	LI/RD/WR Total	77	55.24	11.17	0	77	0.47
10–12	20,644	Listening	25	18.35	4.55	0	25	0.43
10–12	20,644	Speaking	34	29.13	5.84	0	34	0.36
10–12	20,644	Reading	27	18.01	4.81	0	27	0.49
10–12	20,644	Writing	25	18.98	4.19	0	25	0.49
10–12	20,644	LI/RD/WR Total	77	55.35	11.60	0	77	0.55

Table 35—Descriptive Statistics of the NYSESLAT Scale Scores of Listening/Speaking, Reading/Writing Modalities, and Total Scale Scores and Their Correlations With the Regents Comp Exam in English Total Scale Scores

Grade	N	Test	Mean	SD	Sample Min.	Sample Max.	Correlation with Regents English
		Regents English	50.17	21.48	0	98	–
10	2347	NYSESLAT Listening and Speaking	723.97	48.28	588	847	0.45
		NYSESLAT Reading and Writing	703.56	34.00	589	862	0.60
		Total NYSESLAT	1428.00	74.56	120	170	0.56

Table 35—Descriptive Statistics of the NYSESLAT Scale Scores of Listening/Speaking, Reading/Writing Modalities, and Total Scale Scores and Their Correlations With the Regents Comp Exam in English Total Scale Scores (continued)

Grade	N	Test	Mean	SD	Sample Min.	Sample Max.	Correlation with Regents English
10	3,452	Regents English	50.69	19.37	0	100	-
		NYSESLAT Listening and Speaking	46.75	10.35	2	59	0.51
		NYSESLAT Reading and Writing	36.27	8.69	0	52	0.62
		Total NYSESLAT	1413.00	79.58	1119	1662	0.58
11	9,829	Regents English	55.70	17.98	0	100	-
		NYSESLAT Listening and Speaking	47.70	8.85	0	59	0.47
		NYSESLAT Reading and Writing	37.37	7.98	0	52	0.58
		Total NYSESLAT	1420.00	73.66	891	1711	0.55
12	7,363	Regents English	57.25	16.75	0	98	-
		NYSESLAT Listening and Speaking	47.52	8.71	0	59	0.37
		NYSESLAT Reading and Writing	36.82	7.81	0	52	0.47
		Total NYSESLAT	1415.00	71.63	891	1711	0.42
10–12	20,644	Regents English	55.41	17.93	0	100	-
		NYSESLAT Listening and Speaking	47.48	9.07	0	59	0.45
		NYSESLAT Reading and Writing	36.99	8.06	0	52	0.55
		Total NYSESLAT	1417.00	74.03	891	1711	0.51

The correlation coefficients for the Regents Comp Exam in English scale scores and the NYSESLAT raw score total of the three modalities (Reading, Writing, and Listening) ranged from 0.47 to 0.61 across the three grade levels. Within a grade, the combined modalities have higher correlations between the scale scores of the Regents Comp Exam in English than any individual modality of the NYSESLAT raw scores. Additionally, it is logical that the correlations between the scale scores of the Regents Comp Exam in English and the NYSESLAT combined raw scores of Reading/Writing are higher than the correlations between the scale scores of the Regents Comp Exam in English and the NYSESLAT combined raw scores of Listening/Speaking since there is no speaking component on the Regents Comp Exam in English.

The correlations between the scale scores of the Regents Comp Exam in English and the NYSESLAT total scale scores (all four modalities combined) ranged from 0.42 to 0.58 across the three grade levels. All correlations in Table 35 are positive, but slightly lower than the correlations in Table 34 due to the NYSESLAT total combined scale scores including Speaking, which is not a portion of the Regents Comp Exam in English. For this same reason, the correlations between the scale scores of the Regents Comp Exam in English and the NYSESLAT

combined scale scores of Reading/Writing are higher than the correlations between the scale scores of the Regents Comp Exam in English and the NYSESLAT combined scale scores of Listening/Speaking.

In sum, the positive relationship (moderate) between the two examinations for Grades 10–12 demonstrated by the correlation analyses provides evidence of the external structure of the NYSESLAT.

Proficiency Level Classification

The classification percentages on the NYSESLAT (by proficiency level) were also compared to passing and failing on the Regents Comp Exam in English. The results are presented in Tables 36–38.

Table 36 shows that 64% of the Grade 10 students who scored proficient on the NYSESLAT passed the Regents Comp Exam in English. The percentage of students who passed the Regents English exam decreased to 41% for students at the advanced level, and 18% for students at the intermediate and beginning levels.

Table 36—Grade 10: Percentage of ELL Students Passing/Failing the Regents Comp Exam in English at Each NYSESLAT Proficiency Level

Regents English Exam	NYSESLAT Proficiency Levels			
	Beginning (N=327)	Intermediate (N=1,604)	Advanced (N=977)	Proficient (N=544)
Pass	1.22%	17.08%	41.35%	63.79%
Fail	98.78%	82.92%	58.65%	36.21%
Total	100%	100%	100%	100%

Table 37 shows that of the Grade 11 students who were classified as proficient on the NYSESLAT, 73% passed the Regents Comp Exam in English. For those students who were classified as advanced students on the NYSESLAT, 56% passed. For those students who were classified at the beginning and intermediate levels on the NYSESLAT, the percentage of students passing the Regents Comp Exam in English were 5% and 28% respectively.

Table 37—Grade 11: Percentage of ELL Students Passing/Failing the Regents Comp Exam in English at Each NYSESLAT Proficiency Level

NYSESLAT Proficiency Levels				
	Beginning	Intermediate	Advanced	Proficient
Regents English Exam	(N=720)	(N=5,312)	(N=2,663)	(N=1,134)
Pass	4.86%	28.18%	56.33%	72.57%
Fail	95.14%	71.82%	43.67%	27.43%
Total	100%	100%	100%	100%

Table 38 shows that of the Grade 12 students who were classified as proficient on the NYSESLAT, 68% passed the Regents Comp Exam in English. For those students who were classified as advanced students on the NYSESLAT, 59% passed. For those students who were classified at the beginning and intermediate levels on the NYSESLAT, the percentage of students passing the Regents Comp Exam in English were 10% and 36% respectively.

Table 38—Grade 12: Percentage of ELL Students Passing/Failing the Regents Comp Exam in English at Each NYSESLAT Proficiency Level

NYSESLAT Proficiency Levels				
	Beginning	Intermediate	Advanced	Proficient
Regents English Exam	(N=553)	(N=4,738)	(N=1,384)	(N=688)
Pass	10.49%	36.26%	59.10%	68.46%
Fail	89.51%	63.74%	40.90%	31.54%
Total	100%	100%	100%	100%

Summary

The correlation analyses detailed above provide additional evidence of a moderately positive relationship between the NYSESLAT and the Regents Comp Exam in English. Furthermore, the evidence continues to suggest, as other analyses do, that the higher NYSESLAT scores are associated with higher scores on the Regents Comp Exam in English.

In addition to the evidence of a positive correlation between the two assessments, these analyses also confirm the validity of inferences made based upon the NYSESLAT cut scores. That is to say, ELL students who were classified as proficient on the NYSESLAT have a higher degree of passing scores on the Regents Comp Exam in English as opposed to ELL students who scored below the proficient level on the NYSESLAT.

CHAPTER 7: CALIBRATION, EQUATING, AND SCALING

This chapter describes the IRT models and the processes that were used to calibrate, equate, and scale the NYSESLAT. As 2011 was a reprint of the 2009 administration, much of this chapter is a reprint of the previous contractor's work. Questar conducted an anchor stability study to verify the usage of the 2009 scoring tables and psychometric characteristics.

The Rasch Model (Rasch, 1960) and the Partial Credit Model (PCM) (Masters, 1982) were used to calibrate, equate, and scale the NYSESLAT. The Rasch Model was used for dichotomous items, and the PCM for polytomous items. 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 previous vendor used *Winsteps* Version 3.27 to calibrate and scale the NYSESLAT (Linacre & Wright, 2000). Questar performed its anchor stability checks using Version 3.68.1 and will use this version moving forward (Linacre 2009).

7.1 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 for dichotomous items and the PCM 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 Stanford English Language Proficiency (SELP) vertical scale that was developed using the Rasch Model. By using SELP items with known Rasch item difficulties, the vertical scale for the NYSESLAT could be created in a timely fashion.

Second, the sample size requirements for calibration, scaling, and equating under the Rasch Model and PCM are significantly smaller than they are for other IRT models. For example, the Rasch Model requires 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 the previous vendor to recommend that the Rasch model be adopted as the IRT methodology for the NYSESLAT.

7.2 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).

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 fit the IRT model perfectly. Items that degraded or distorted the measurement system were 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 the 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. These numbers are updated based upon the 2011 anchor stability results.

7.3 Description of Calibration Sample, Process

Calibration Sample

An equating sample (near 100% of the population) that was used for the equating analyses represented the ELL population in all six need/resource categories: NYC, Large City, and High Need Urban/Suburban, Rural, Average Need, and Low Need districts.

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 tested for the 2009 administration were calibrated to the 2008 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 2008 scale values.
- The results of this calibration were used as the operational item parameters to create the preliminary scales for the NYSESLAT spring 2009 administration. Item estimates of the common items from the 2008 administration were used as anchors to place the 2009 scale to the established 2008 vertical scale. The final reporting scale was derived from the vertical scale using a linear transformation.

7.4 Equating of the NYSESLAT

Equating Method and Rationale

In 2009, the 2008 NYSESLAT vertical scale was used to establish the 2009 NYSESLAT scale. The common-item, non-equivalent groups design was used for 2009 and 2008 equating because it is one of the most practical of the various designs used for equating. For example, this design does not 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). This makes the common-item non-equivalent groups design a very popular choice for large-scale assessments.

The common-item, non-equivalent groups 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 observation 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). 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 2009 NYSESLAT was successfully equated to the 2008 NYSESLAT while maintaining the established 2009 common vertical scale, which was based on the 2006 common vertical scale and the 2005 SELP scale.

In establishing the 2009 vertical scale, the data file used for the calibration and equating of the NYSESLAT had 190,178 cases, which was about 86% of the final operational research file (total $N=220,332$) received from New York SED. The calibration/equating data was representative of the population covering New York City and the Large City school districts (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 2009 operational form were calibrated, and the newly administered items were then located on the 2008 scale. Once the scale locations of the 2009 NYSESLAT were known, IRT true score equating was used to relate the raw scores on the 2009 NYSESLAT to the 2008 scale. In this process, the true score on the NYSESLAT with a given level of examinee ability was considered an equivalent true score on the 2008 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, the pre-existing 2008 vertical scale was used to create the NYSESLAT vertical scale of 2009. For the 2009 administration, the common items between 2008 and 2009 were the anchor items on the 2009 NYSESLAT and were, as explained earlier, fixed to the parameter values from the pre-existing vertical scale of 2008. That is, the common items between 2008 and 2009 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.

Anchor Item Evaluation

There are some general guidelines 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 locations as they do in the test that is being equated.
- Anchor items are expected to have strong statistical properties (e.g., point-biserial > 0.30) and span the range of difficulty including easy, medium, and difficult items. The mean difficulty (p-value/Rasch difficulty values) of a new form's anchor items should be very similar to the mean difficulty of the previous operational test.

Stability of Anchor Items

To assess the psychometric stability of the 2011 NYSESLAT reprint of the 2009 assessment, Questar psychometricians conducted an anchor stability study. The aim of this study was to determine if there had been a significant shift in the stability of the item parameters, and therefore, the feasibility of reusing the 2009 scoring tables.

The anchor stability check was conducted for every grade band on both the Listening/Speaking and Reading/Writing portions of the NYSESLAT assessment. Drs. Vansickle and Mueller created Winsteps control files, using Winsteps 3.68.1, to analyze both the 2009 data and the 2011 data files using anchored data runs. The 2009 b-parameters and associated step values for the indicated linking items were used to anchor the data sets. Additionally, a free run of the 2009 data files was compared to the 2009 anchored run to ensure that the use of an updated version of Winsteps would not have an impact on the anchored values. Previously, Winsteps 3.27 was used to calibrate these items. Comparisons utilized a Robust Z calculation, as well as review of displacement values, and theta values. Results indicated that there were no discernable differences in any grade band in either the Listening/Speaking or Reading/Writing portions of the data². In every instance of the study the test characteristic curve of 2011 overlays 2009 which is evidence of an extremely stable test.

For the 2011 NYSESLAT, the displacement value of 0.30 for an item (used in many large-scale assessments) was used in order to identify potential outliers. This criterion was used consistently across all the grade spans. For all grade spans, the overall percentage of anchor items between 2008 and 2009 assessment was around 25% of the total items. Following anchor evaluation, the general rule is that the anchor set should be no smaller than 20% of the test (e.g., Huynh, 2000). See Table 39 for the number of items excluded for each modality combination by grade span.

² Figures showing a comparison of the theta values between 2009 and 2011 are available in the technical write-up provided to SED by Questar (titled "NYSESLAT Anchor Stability Check: Operational 2011 Assessment").

Items which are excluded from being anchor items in the NYSESLAT have their parameters re-estimated, and the new estimate becomes the adjusted item parameter in the item bank for future use.

Table 39 - Common Items as Anchor Items by Modality Combination and Grade Span

Grade Span	Modality Combination	Total Number of Items	Common Items (2010 and 2011)	Percentage of Total Items	Anchor Items (2009 and 2010)	Percentage of Total Items
K-1	Listening/Speaking	40	10	25.0%	10	25.0%
	Reading/Writing	30	8	26.7%	8	26.7%
2-4	Listening/Speaking	40	10	25.0%	10	25.0%
	Reading/Writing	40	10	25.0%	10	25.0%
5-6	Listening/Speaking	41	10	24.4%	10	24.4%
	Reading/Writing	46	11	23.9%	11	23.9%
7-8	Listening/Speaking	41	11	26.8%	11	26.8%
	Reading/Writing	45	11	24.4%	11	24.4%
9-12	Listening/Speaking	41	11	26.8%	11	26.8%
	Reading/Writing	46	11	23.9%	11	23.9%

Note: Common items are the same items appearing in both the 2009 and 2010 operational tests.

Test Characteristic Curve and Standard Error Curves Comparison

In order to ascertain the maintenance of the scale across years, comparisons between the 2010, and 2011 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 thetas, or students' abilities, with the y -axis representing the expected score based on the number or proportion of items correctly solved by students. 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 thetas, or students' abilities, with the y -axis representing the SEM.

The 2011 TCC plots for all grade spans (see Figures 3 & 4) indicate that the TCC shifts to the right with each progression of grade span across the Listening/Speaking and Reading/Writing vertical scales with the exception of grade span 5-6 and grade span 7-8 in Reading/Writing. Overall, the TCCs of 2010 and 2011 align well.

The differences in terms of the SE curves between the 2010 and 2011 tests for both the Listening/Speaking and Reading/Writing are minimal across all the grade spans. The noticeable differences occur only at the tail portion of the SE curves, which is a typical result.

Listening and Speaking Test Vertical Scale

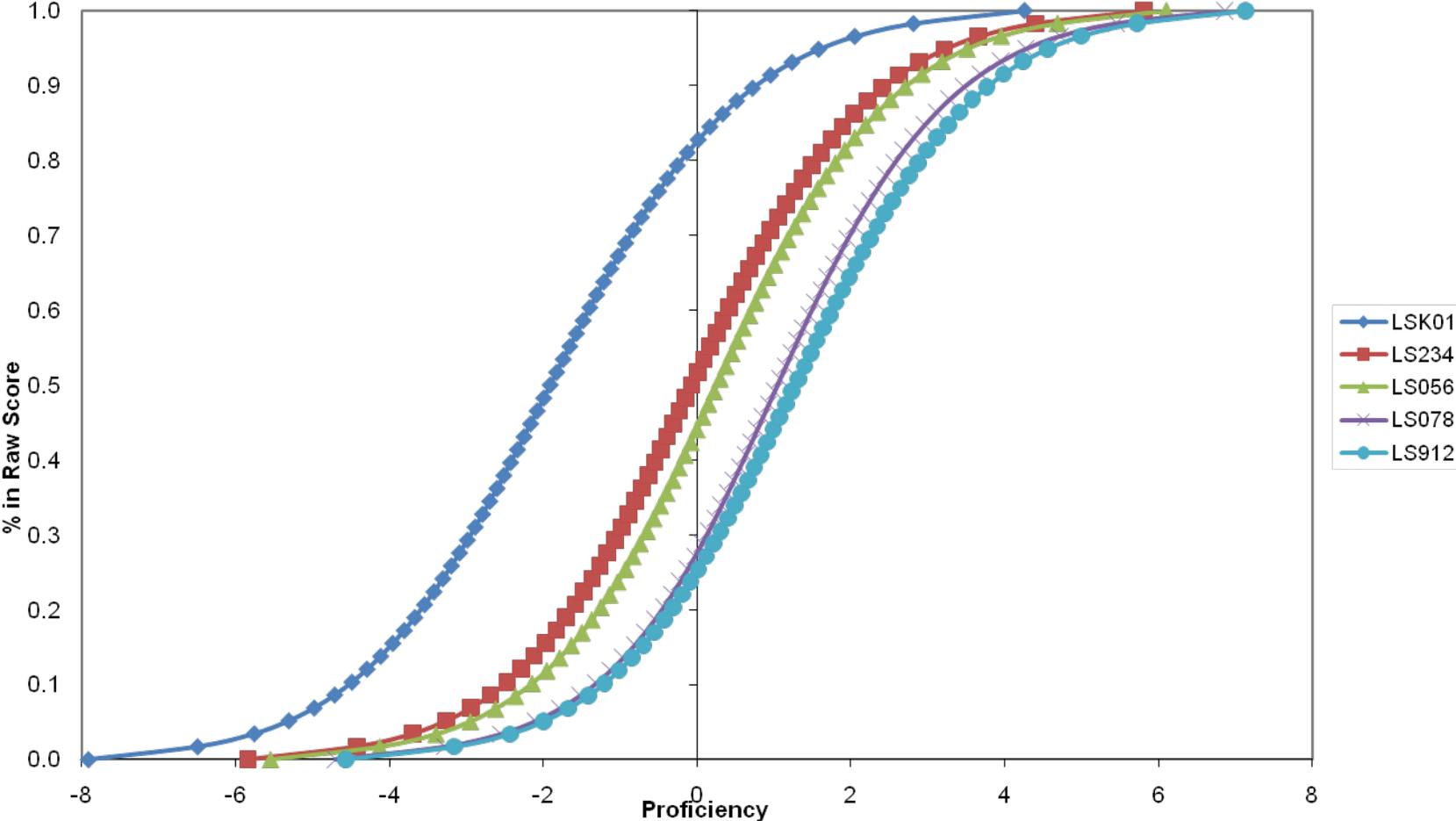


Figure 3—2011 TCC L/S Grade Span 1–5

Reading and Writing Test Vertical Scale

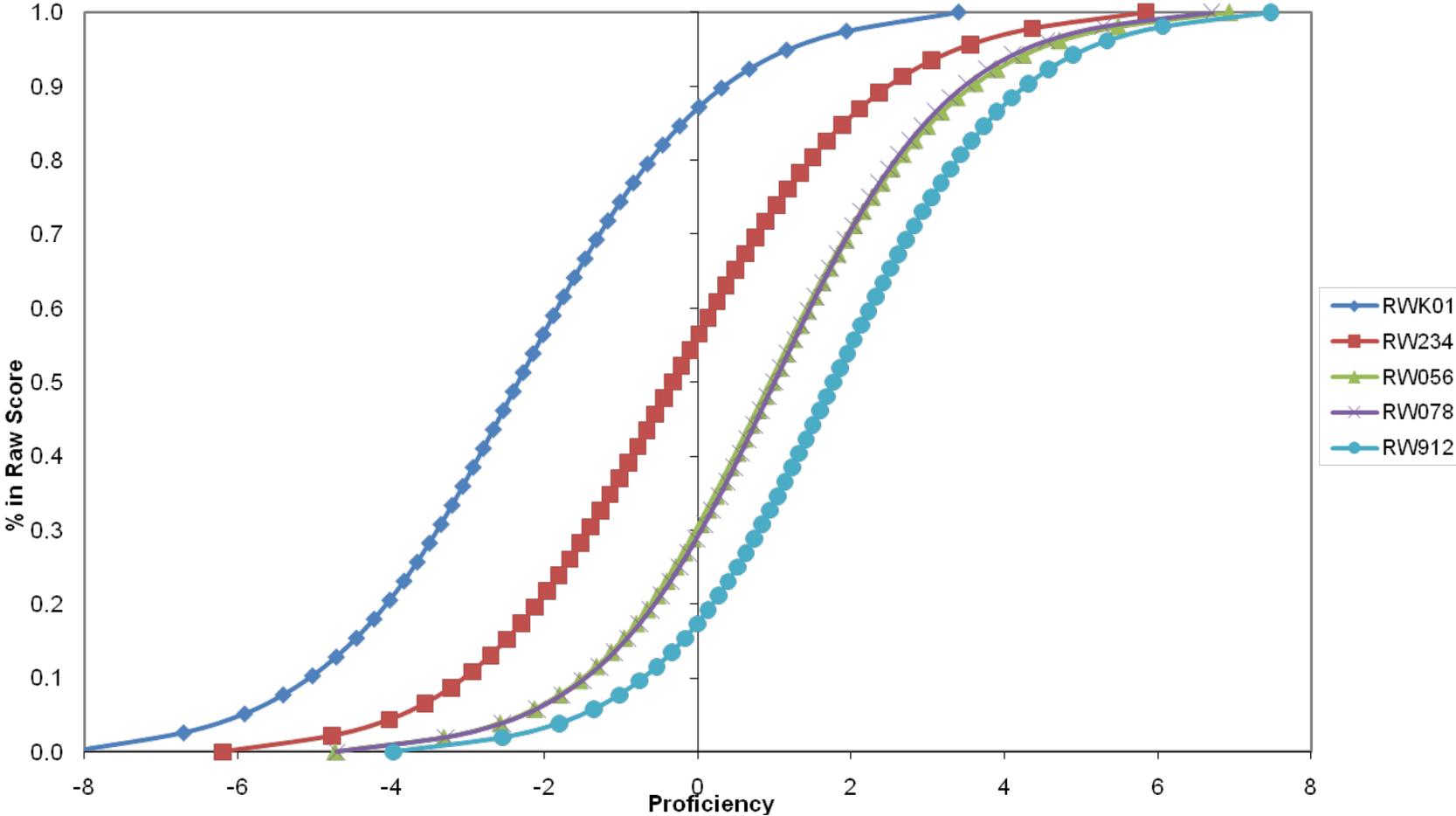


Figure 4—2011 TCC R/W Grade Span 1–5

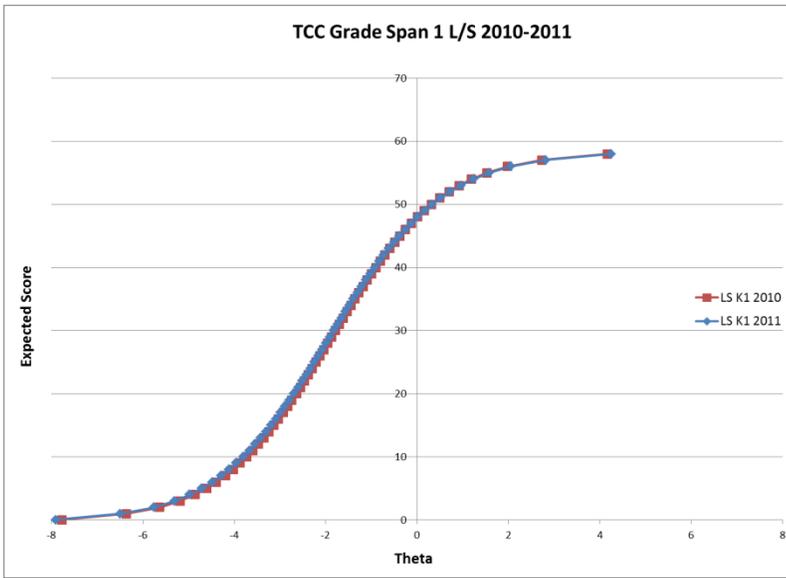


Figure 5—2010–11 TCC L/S Grade Span 1

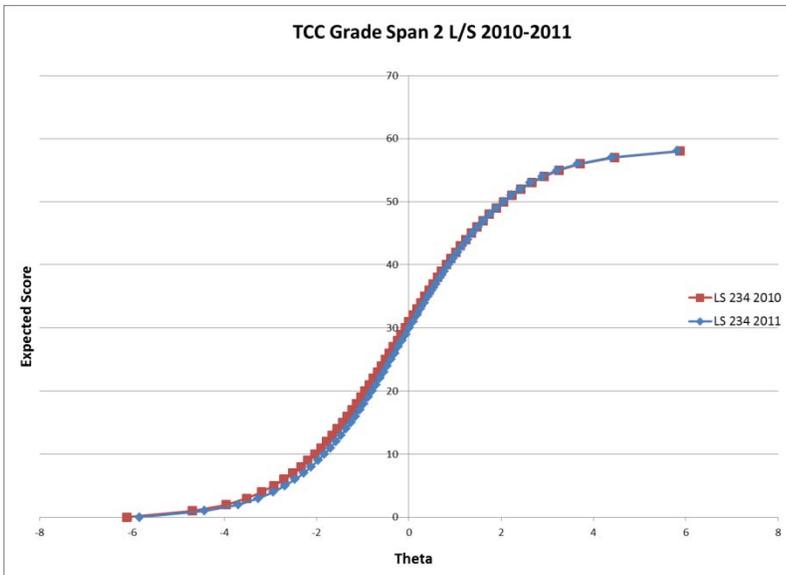


Figure 6—2010–11 TCC L/S Grade Span 2

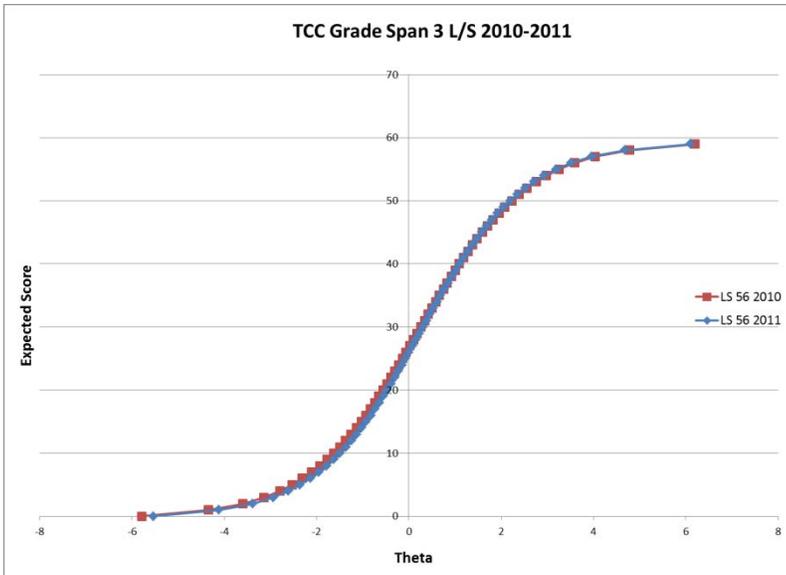


Figure 7—2010–11 TCC L/S Grade Span 3

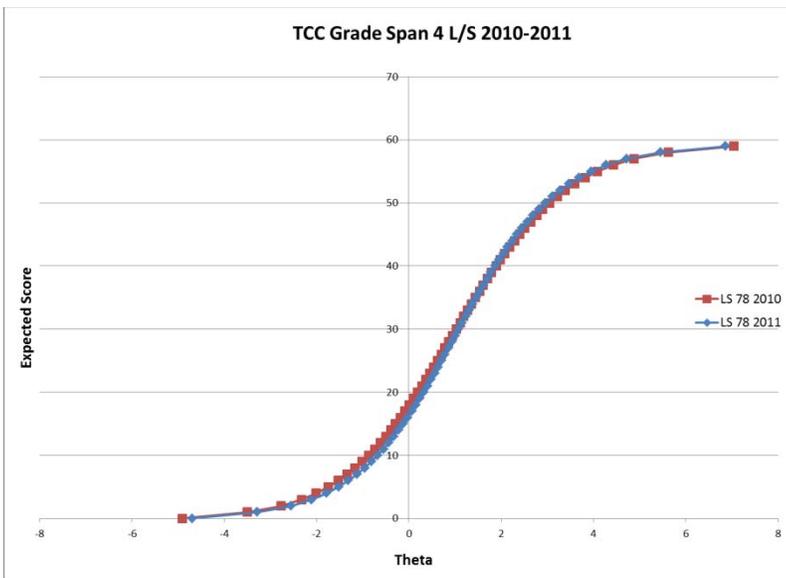


Figure 8—2010–11 TCC L/S Grade Span 4

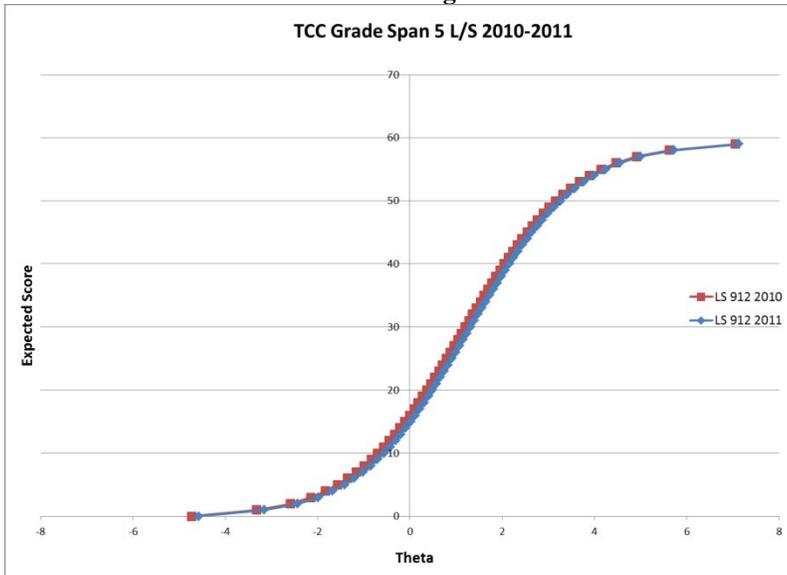


Figure 9—2010–11 TCC L/S Grade Span 5

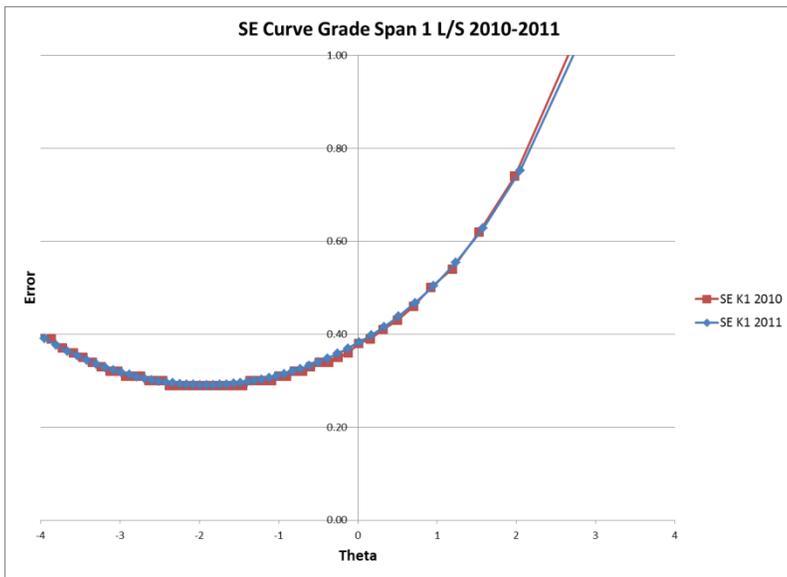


Figure 10—2010–11 SE Curves L/S Grade Span 1

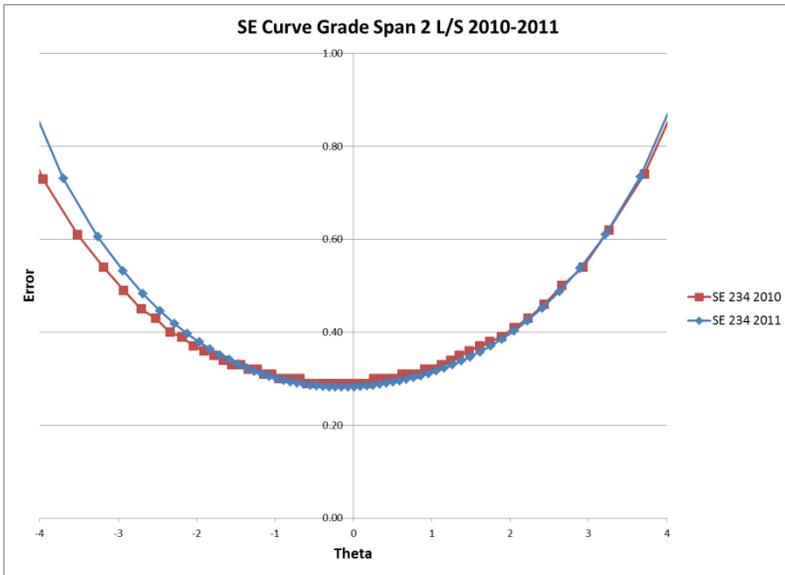


Figure 11—2010–11 SE Curves L/S Grade Span 2

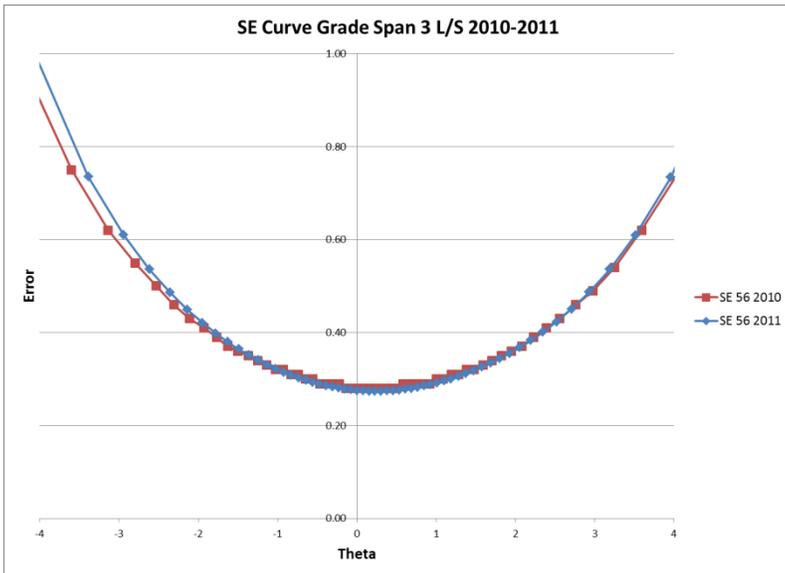


Figure 12—2010–11 SE Curves L/S Grade Span 3

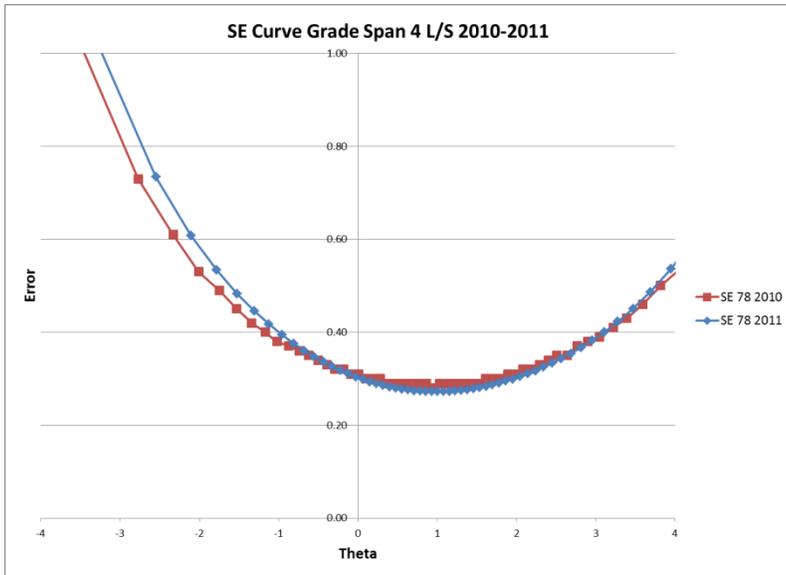


Figure 13—2010–11 SE Curves L/S Grade Span 4

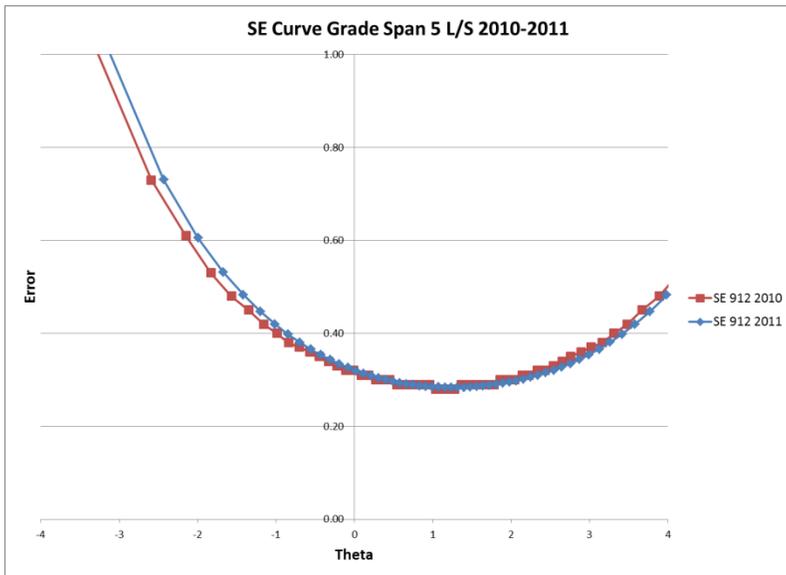


Figure 14— 2010–11 SE Curves L/S Grade Span 5

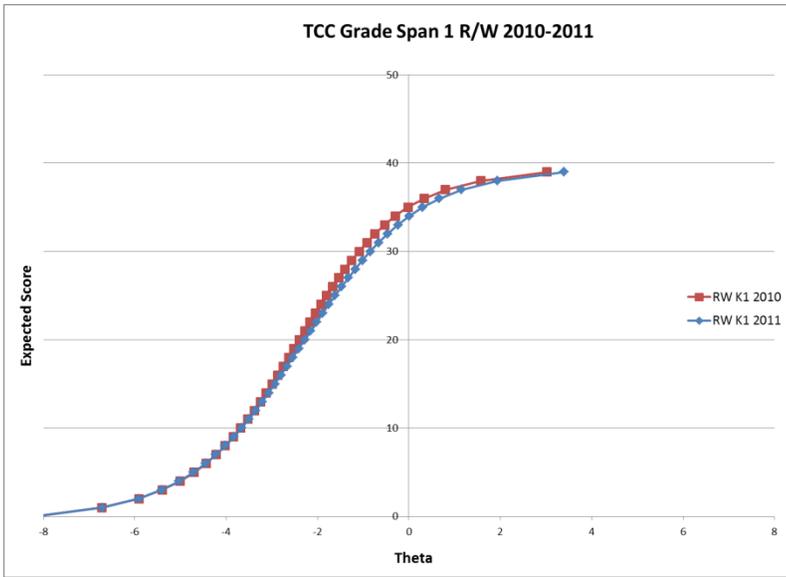


Figure 15—2010–11 TCC R/W Grade Span 1

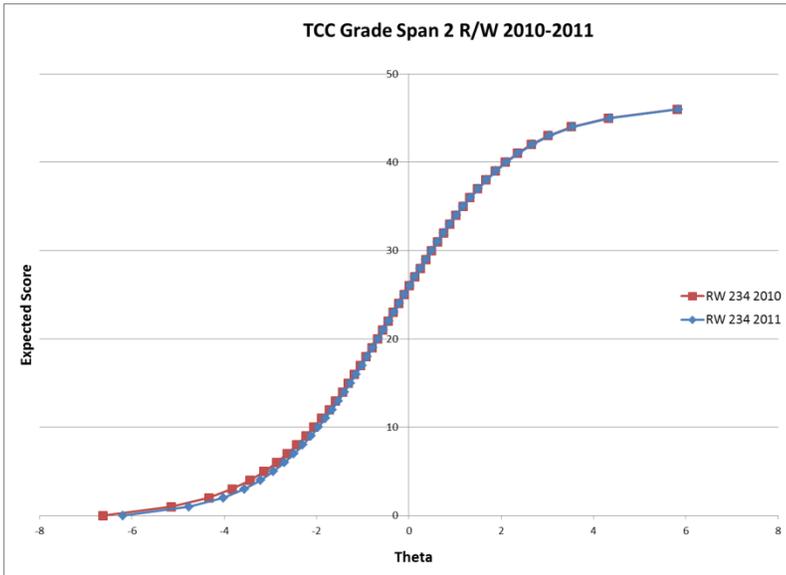


Figure 16—2010–11 TCC R/W Grade Span 2

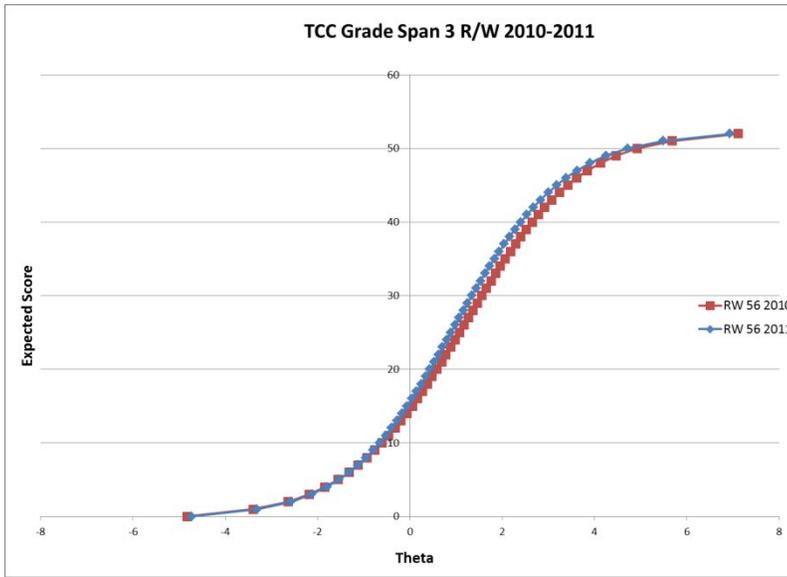


Figure 17—2010–11 TCC R/W Grade Span 3

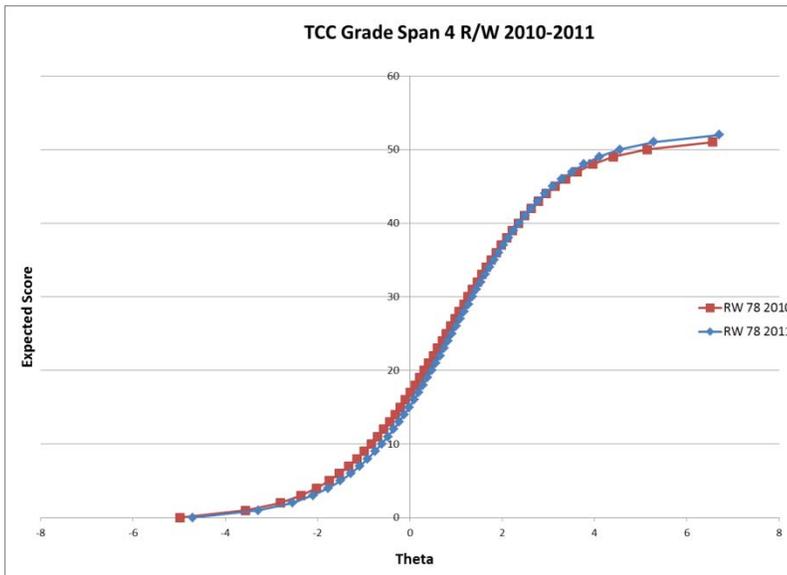


Figure 18—2010–11 TCC R/W Grade Span 4

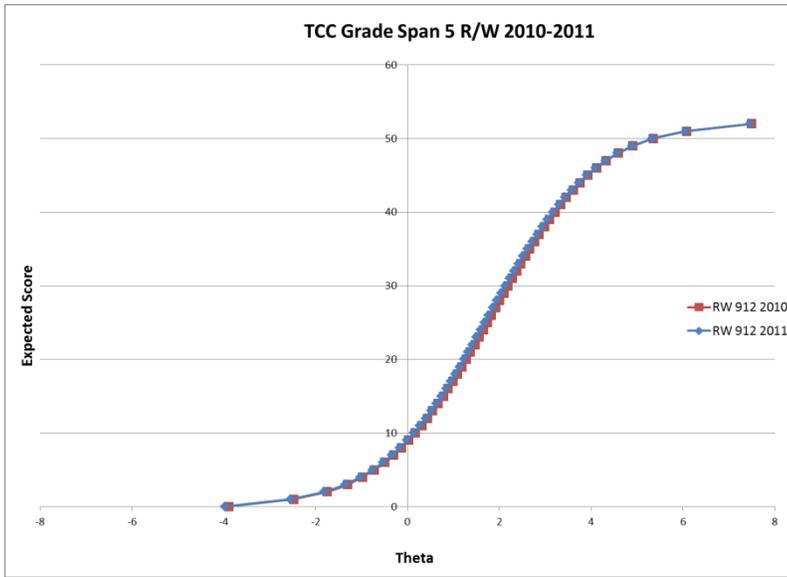


Figure 19— 2010–11 TCC R/W Grade Span 5

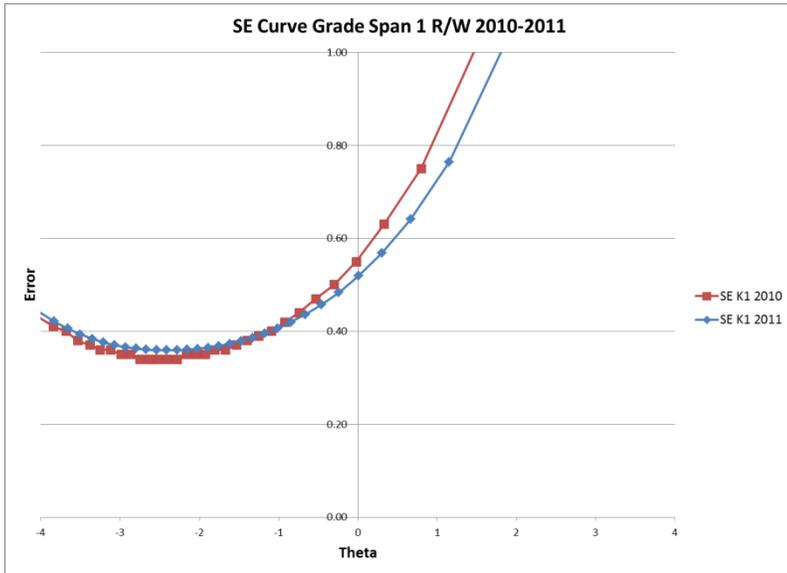


Figure 20—2010–11 SE Curves R/W Grade Span 1

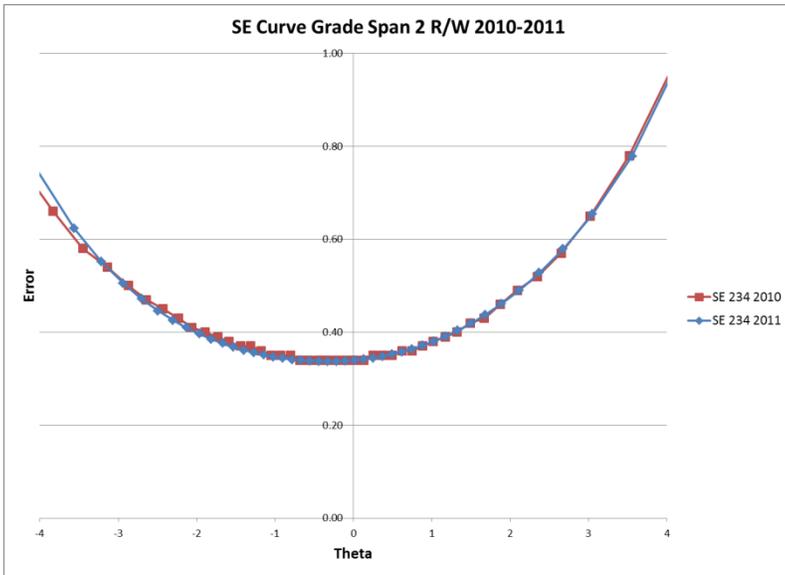


Figure 21—2010–11 SE Curves R/W Grade Span 2

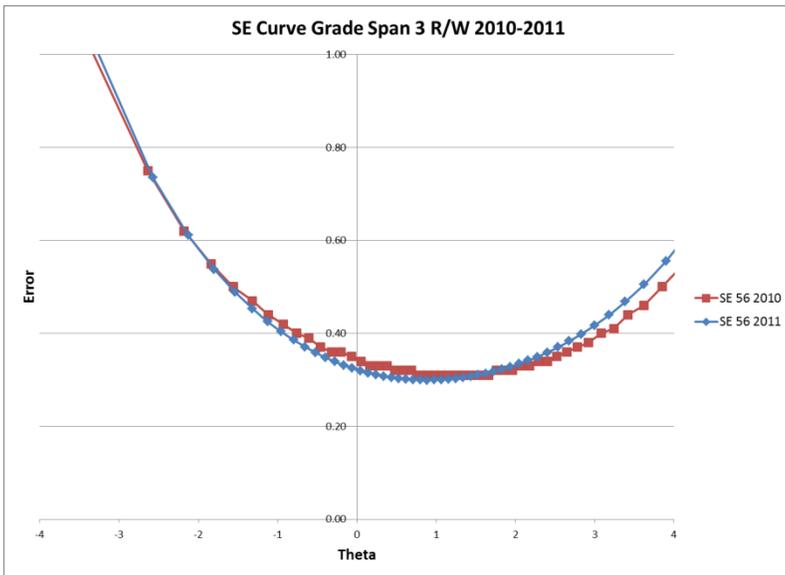


Figure 22—2010–11 SE Curves R/W Grade Span 3

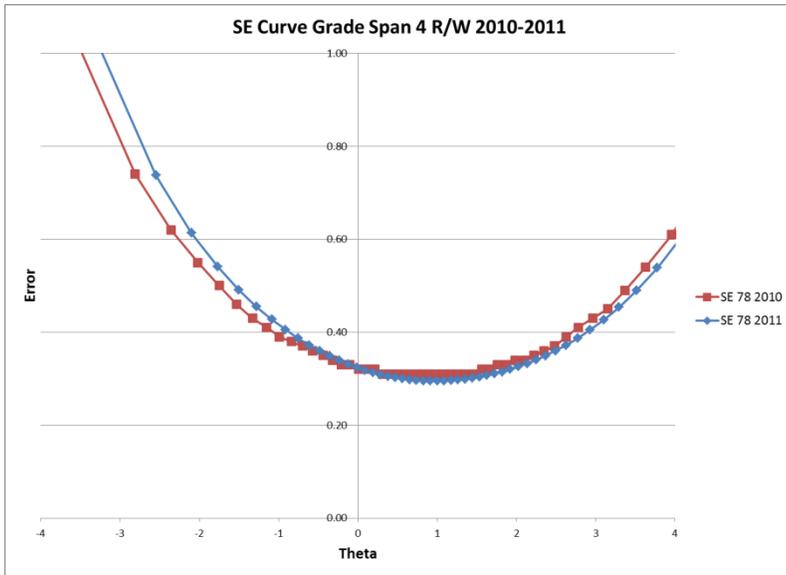


Figure 23—2010–11 SE Curves R/W Grade Span 4

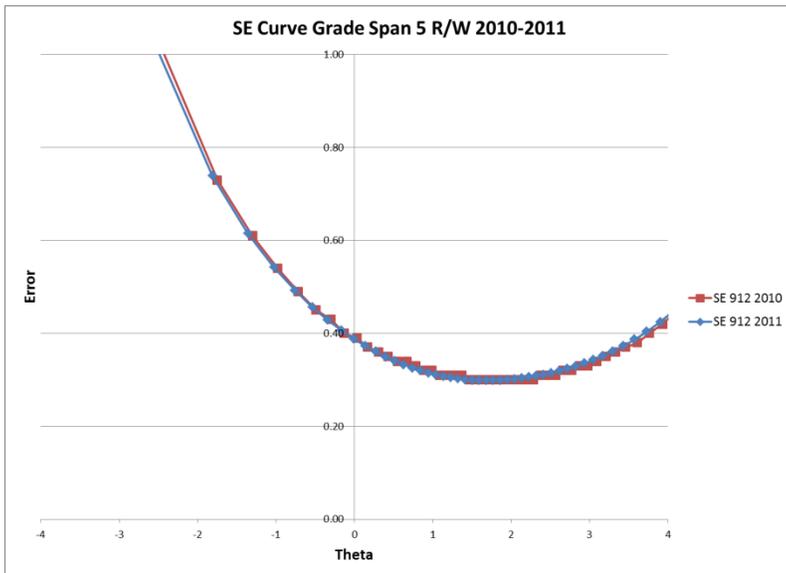


Figure 24—2010–11 SE Curves R/W Grade Span 5

7.5 Scaling of the NYSESLAT

As explained earlier, the NYSESLAT vertical scale was created based on the pre-existing 2008 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:

$$\text{Scale Score} = 35 * (\text{theta}) + 600 \quad (17)$$

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. For 2011, the result of the anchor stability analyses was a determination that the 2009 scoring tables could be applied for the 2011 tests. Appendices C.1–C.5 provide the raw-to-scale score conversion tables for the Speaking/Listening and Reading/Writing combinations by grade span.

7.6 Rasch Information

Appendix D contains the results of the operational items for the NYSESLAT (this Appendix is reprinted from 2009 as well since the 2011 test is a reprint of 2009). 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

Table 40 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. Appendix E contains the 2009 results of this analysis since the information was not updated for the 2011 assessment because the item parameters were maintained from the 2009 assessment.

Table 40—Average Rasch Difficulty by Grade Span and Modality

	Modality	Number of Items	Max Points	Average Rasch Difficulty
K-1	Listening	24	24	-2.431
	Speaking	16	34	-1.573
	Reading	15	15	-2.135
	Writing	15	24	-2.693
	Listening and Speaking	40	58	-2.088
	Reading and Writing	30	39	-2.414
	Total	70	97	-2.228
2-4	Listening	24	24	0.035
	Speaking	16	34	-0.165
	Reading	24	24	-0.462
	Writing	16	22	-0.337
	Listening and Speaking	40	58	-0.045
	Reading and Writing	40	46	-0.412
	Total	80	104	-0.228
5-6	Listening	25	25	0.513
	Speaking	16	34	0.055
	Reading	27	27	0.943
	Writing	19	25	0.938
	Listening and Speaking	41	59	0.334
	Reading and Writing	46	52	0.941
	Total	87	111	0.655
7-8	Listening	25	25	1.387
	Speaking	16	34	0.808
	Reading	27	27	1.068
	Writing	19	25	0.814
	Listening and Speaking	41	59	1.161
	Reading and Writing	46	52	0.963
	Total	87	111	1.056
9-12	Listening	25	25	1.735
	Speaking	16	34	0.919
	Reading	27	27	1.976
	Writing	19	25	1.467
	Listening and Speaking	41	59	1.416
	Reading and Writing	46	52	1.765
	Total	87	111	1.601

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

CHAPTER 8: STANDARD SETTING

8.1 Introduction

The previous vendor for the NYSESLAT carried out the standard setting study. This chapter is not modified to maintain consistency with their representation of the standard setting activities.

The contractor for the NYSESLAT 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 the contractor to facilitate the technical part of the standard setting. In addition, a content specialist from the vendor 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 as the booklet progresses. 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 41, 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, ELA, 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 represent all regions of the state of New York. Other educational stakeholders, such as administrators, curriculum specialists,

NYSESED members, and members of the professional community were also included, as per recommendations. These additional members tend to provide valuable insights from their area of expertise and help strengthen the consequential validity argument during panel discussions.

Table 41—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, and they respond 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 NYSED, the vendor, 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 the 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 panelists 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 session was held. The focus was on the characteristics and interrelationships of 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 were 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 among 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, the vendor 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 help them 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 2009 administration of the NYSESLAT for the Listening/Speaking and Reading/Writing combinations in raw score points, scale score, and theta metric (from standard setting) are presented in Table 42 and Table 43. There are three cut points that correspond to four performance levels. Any score below the Intermediate cut point is the Beginning performance level. Questar applied the same scoring tables and cut points for the 2011 administration as reported in the 2009 technical report.

Table 42—Final Performance-Level Cut Points for Listening and Speaking

Grade	Listening & Speaking								
	Raw Score Cut, 2009			Scale Score			Theta		
	I	A	P	I	A	P	I	A	P
K	22	42	51	511	573	616	-2.55	-0.76	0.47
1	25	44	55	520	582	649	-2.28	-0.52	1.39
2	13	32	48	543	602	661	-1.63	0.06	1.73
3	17	35	51	555	611	675	-1.27	0.31	2.14
4	20	38	52	567	621	684	-0.95	0.59	2.39
5	20	38	52	579	631	686	-0.59	0.89	2.46
6	23	41	54	587	639	696	-0.37	1.11	2.74
7	17	35	50	597	651	703	-0.08	1.45	2.95
8	19	38	50	605	661	703	0.15	1.74	2.95
9	22	41	50	623	678	715	0.66	2.23	3.28
10	24	43	51	628	686	721	0.80	2.44	3.46
11	26	45	53	635	694	728	1.00	2.68	3.66
12	28	48	53	640	704	728	1.14	2.96	3.66

Table 43—Final Performance-Level Cut Points for Reading and Writing

Grade	Reading & Writing								
	Raw Score Cut, 2009			Scale Score			Theta		
	I	A	P	I	A	P	I	A	P
K	25	32	35	537	582	609	-1.79	-0.52	0.25
1	29	35	38	559	609	655	-1.17	0.25	1.58
2	20	32	39	571	623	663	-0.82	0.64	1.81
3	22	33	41	580	630	679	-0.57	0.86	2.25
4	24	36	42	589	642	691	-0.33	1.19	2.61
5	23	35	44	620	660	704	0.56	1.73	2.98
6	24	37	45	623	668	710	0.66	1.94	3.15
7	25	38	45	629	674	710	0.83	2.13	3.15
8	26	40	45	632	682	710	0.92	2.34	3.15
9	22	38	44	643	698	726	1.21	2.80	3.60
10	23	39	45	646	702	732	1.30	2.91	3.78
11	24	41	46	649	711	739	1.39	3.16	3.97
12	24	42	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, and P = Proficient.

CHAPTER 9: SUMMARY OF OPERATIONAL TEST RESULTS

Chapter 9 contains raw score and scale score summaries for each of the reporting scales (Listening/Speaking and Reading/Writing). Table 44 details raw scores by grade. Table 45 contains the raw score summary by grade span. Subsequently, Table 46 includes the sample size (n-count), the mean, median, interquartile range (IQR), range, and the standard deviation (SD) of the scale score. Table 47 presents the summary of the scale scores by grade span. Tables 48 presents the percentage of students in each of the proficiency levels by grade. Table 49 presents the percentage of students in each of the proficiency levels by grade span. The overall proficiency level is defined as the lower of the two proficiency levels that a student achieves (e.g., Listening/Speaking level=3, Reading/Writing level=4, Overall level=3). Table 50 presents the exit rate by grade for 2008, 2009, 2010, and 2011 administrations. Table 51 presents the exit rate by grade span for 2008, 2009, 2010, and 2011 administrations. The two tables include n-counts and percentages of ELL students obtaining exit status after scoring at the Proficient Level (level 4) on both the Listening/Speaking and Reading/Writing modality combinations. Similar analyses were done according to the following subgroups (See Appendices G, H, and I):

- Gender
- Ethnicity (American Indian or Alaska Native, Black or African American, Asian, Native Hawaiian/other Pacific Islander, Hispanic or Latino, White, and Multi-Racial [not of Hispanic origin])
- Needs/Resource Group (NYC, Large City, 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 as an ELL
- Students with disabilities (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 44—Raw Score Summary by Grade

Grade	Test	N	Mean	Median	IQR	Range	SD
K	Listening and Speaking	25,432	44.24	47	12	0~58	11.12
	Reading and Writing	25,432	25.96	27	11	0~39	7.72
1	Listening and Speaking	25,668	50.28	53	8	0~58	9.27
	Reading and Writing	25,668	33.79	36	6	0~39	6.24
2	Listening and Speaking	21,013	45.52	49	11	0~58	10.59
	Reading and Writing	21,013	29.00	30	13	0~46	8.93
3	Listening and Speaking	20,140	48.30	52	9	0~58	10.39
	Reading and Writing	20,140	33.58	36	10	0~46	8.50
4	Listening and Speaking	18,298	49.52	53	8	0~58	10.60
	Reading and Writing	18,298	35.84	39	9	0~46	8.34
5	Listening and Speaking	15,833	47.37	51	10	0~59	11.46
	Reading and Writing	15,833	36.93	40	14	0~52	10.73
6	Listening and Speaking	13,667	46.78	51	11	0~59	12.14
	Reading and Writing	13,667	36.73	40	15	0~52	10.97
7	Listening and Speaking	12,632	43.95	49	14	0~59	13.05
	Reading and Writing	12,632	33.59	36	16	0~52	10.92
8	Listening and Speaking	12,662	44.22	50	16	0~59	13.38
	Reading and Writing	12,662	34.79	37	17	0~52	11.04
9	Listening and Speaking	14,640	40.92	46	22	0~59	14.66
	Reading and Writing	14,640	31.35	33	17	0~52	10.77
10	Listening and Speaking	13,275	43.07	47	17	0~59	12.52
	Reading and Writing	13,275	33.70	35	15	0~52	9.83
11	Listening and Speaking	8,872	46.28	49	13	0~59	10.55
	Reading and Writing	8,872	36.50	38	12	0~52	8.90
12	Listening and Speaking	6,462	49.16	52	10	0~59	8.52
	Reading and Writing	6,462	38.44	40	10	0~52	8.01

Table 45—Raw Score Summary by Grade Span

Grade	Test	N	Mean	Median	IQR	Range	SD
Band K-1	Listening and Speaking	51,100	47.27	51	11	0~58	10.67
	Reading and Writing	51,100	29.89	32	12	0~39	8.04
Band 2 -4	Listening and Speaking	59,451	47.70	51	10	0~58	10.66
	Reading and Writing	59,451	32.66	35	13	0~46	9.07
Band 5-6	Listening and Speaking	29,500	47.10	51	10	0~59	11.78
	Reading and Writing	29,500	36.84	40	14	0~52	10.84
Band 7-8	Listening and Speaking	25,294	44.09	49	16	0~59	13.22
	Reading and Writing	25,294	34.19	37	17	0~52	11.00
Band 9-12	Listening and Speaking	43,249	43.91	48	17	0~59	12.77
	Reading and Writing	43,249	34.19	36	15	0~52	10.07

Note: The data file used in Tables 44 and 45, above, is the final research file. The total n-counts for Grades K–12 are 208,596. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

Table 46—Scale Score Summary by Grade

Grade	Test	N	Mean	Median	IQR	Range	SD
K	Listening and Speaking	25,432	596.15	596	59	322~749	54.39
	Reading and Writing	25,432	553.93	551	60	315~718	51.90
1	Listening and Speaking	25,668	637.32	634	71	322~749	59.92
	Reading and Writing	25,668	624.63	622	85	315~718	64.03
2	Listening and Speaking	21,013	661.76	667	56	388~805	49.37
	Reading and Writing	21,013	615.51	616	60	374~806	46.09
3	Listening and Speaking	20,140	680.10	685	61	388~805	54.21
	Reading and Writing	20,140	640.87	646	58	374~806	47.83
4	Listening and Speaking	18,298	692.28	693	68	388~805	59.39
	Reading and Writing	18,298	657.05	666	64	374~806	51.95
5	Listening and Speaking	15,833	677.68	682	58	402~814	50.41
	Reading and Writing	15,833	678.00	683	63	426~843	48.79
6	Listening and Speaking	13,667	676.21	682	61	402~814	52.79
	Reading and Writing	13,667	677.62	683	66	426~843	50.55
7	Listening and Speaking	12,632	692.49	702	61	429~845	53.30
	Reading and Writing	12,632	664.82	668	59	432~838	45.12
8	Listening and Speaking	12,662	695.85	707	72	429~845	56.76
	Reading and Writing	12,662	670.63	672	66	432~838	47.40
9	Listening and Speaking	14,640	691.05	698	84	436~851	61.56
	Reading and Writing	14,640	679.52	682	61	455~960	44.72
10	Listening and Speaking	13,275	698.26	702	70	436~851	53.86
	Reading and Writing	13,275	688.74	689	56	455~860	41.37
11	Listening and Speaking	8,872	711.84	711	61	436~851	50.81
	Reading and Writing	8,872	700.16	701	48	455~860	40.22
12	Listening and Speaking	6,462	724.89	727	56	436~851	47.34
	Reading and Writing	6,462	707.92	709	44	455~860	37.94

Table 47— Scale Score Summary by Grade Span

Grade	Test	N	Mean	Median	IQR	Range	SD
Span K-1	Listening and Speaking	51,100	616.83	618	73	322~749	60.83
	Reading and Writing	51,100	589.44	582	99	315~718	68.19
Span 2 -4	Listening and Speaking	59,451	677.37	679	65	388~805	55.67
	Reading and Writing	59,451	636.88	641	71	374~806	51.46
Span 5-6	Listening and Speaking	29,500	677.00	682	58	402~814	51.53
	Reading and Writing	29,500	677.83	683	63	426~843	49.61
Span 7-8	Listening and Speaking	25,294	694.17	702	72	429~845	55.08
	Reading and Writing	25,294	667.73	672	65	432~838	46.37
Span 9-12	Listening and Speaking	43,249	702.58	706	73	436~851	56.39
	Reading and Writing	43,249	690.83	693	56	455~860	43.07

Note: 1. Generally speaking, the mean for each grade/grade span should increase from one grade/grade span to the next higher grade/grade span in a similar manner, as shown in Table 42 and Table 43 of this report, which depicts increases across the grade levels. However, due to artifacts of the population whereby some grades/grade spans may have a greater percentage of higher scoring students than the next higher grade/grade span, the mean for the lower grade/grade span can be higher than the next higher grades/grade spans.

2. The statistics in Tables 46 and 47 are based on the final research file. The total n-counts for grade K–12 are 208,596. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

Table 48—Percentage of Students in Each Proficiency Level by Grade

Grade	Test	Proficiency Levels (Percent)			
		Beginning	Intermediate	Advanced	Proficient
K	Listening And Speaking	5.97	23.82	35.95	34.26
	Reading and Writing	39.71	33.49	12.78	14.03
	Overall	40.36	35.36	14.76	9.52
1	Listening And Speaking	3.81	10.13	45.64	40.42
	Reading and Writing	15.60	22.38	27.54	34.48
	Overall	16.18	24.04	37.59	22.20
2	Listening And Speaking	2.56	7.10	35.04	55.29
	Reading and Writing	16.75	37.52	31.81	13.92
	Overall	16.88	37.49	32.51	13.11
3	Listening And Speaking	3.16	6.08	33.25	57.51
	Reading and Writing	11.34	22.00	47.17	19.50
	Overall	11.57	22.00	48.98	17.45
4	Listening And Speaking	4.05	6.72	26.84	62.39
	Reading and Writing	10.70	22.37	40.97	25.96
	Overall	11.02	22.34	43.05	23.59
5	Listening And Speaking	5.32	9.16	36.23	49.28
	Reading and Writing	13.45	18.16	34.07	34.31
	Overall	13.81	18.42	40.07	27.70
6	Listening And Speaking	7.35	12.97	44.25	35.43
	Reading and Writing	15.81	22.18	32.55	29.46
	Overall	16.49	22.59	41.35	19.57
7	Listening And Speaking	6.23	14.39	31.36	48.01
	Reading and Writing	22.70	33.04	27.53	16.74
	Overall	23.11	32.96	29.01	14.93
8	Listening And Speaking	7.82	16.72	25.44	50.02
	Reading and Writing	22.58	34.43	21.12	21.88
	Overall	23.16	34.37	22.43	20.04
9	Listening And Speaking	13.94	25.49	21.11	39.45
	Reading and Writing	21.84	43.87	20.51	13.78
	Overall	24.45	42.23	20.84	12.48
10	Listening And Speaking	9.30	29.87	24.18	36.65
	Reading and Writing	15.29	47.43	23.16	14.12
	Overall	17.69	47.34	22.99	11.97
11	Listening And Speaking	5.28	28.19	31.74	34.79
	Reading and Writing	9.31	51.65	23.86	15.18
	Overall	11.07	52.21	24.91	11.81
12	Listening And Speaking	2.75	28.86	23.99	44.40
	Reading and Writing	5.32	52.32	23.40	18.96
	Overall	6.34	55.08	23.58	15.00

Table 49—Percentage of Students in Each Proficiency Level by Grade Span

Grade	Test	Proficiency Levels (Percent)			
		Beginning	Intermediate	Advanced	Proficient
Span K-1	Listening And Speaking	4.89	16.94	40.82	37.35
	Reading and Writing	27.6	27.91	20.19	24.3
	Overall	28.21	29.67	26.23	15.89
Span 2-4	Listening And Speaking	3.22	6.64	31.91	58.23
	Reading and Writing	13.05	27.6	39.83	19.52
	Overall	13.28	27.58	41.33	17.81
Span 5-6	Listening And Speaking	6.26	10.93	39.95	42.86
	Reading and Writing	14.55	20.03	33.36	32.06
	Overall	15.05	20.36	40.66	23.93
Span 7-8	Listening And Speaking	7.03	15.56	28.4	49.02
	Reading and Writing	22.64	33.73	24.32	19.31
	Overall	23.13	33.66	25.71	17.49
Span 9-12	Listening And Speaking	9.07	27.89	24.66	38.38
	Reading and Writing	14.79	47.82	22.44	14.95
	Overall	16.93	47.77	22.75	12.56

Note: The statistics in Tables 48 and 49 are based on the final research file. The total n-counts for Grades K–12 are 208,596. The statistics presented in the table are based on the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

Table 50—Exit Rate by Grade for 2008, 2009, 2010, and 2011 Administrations

Grade	2008		2009		2010		2011	
	N	Percent	N	Percent	N	Percent	N	Percent
K	1,191	4.94	1,663	7.04	2,515	9.89	2,422	9.52
1	4,132	16.05	5,148	19.02	6,059	22.17	5,698	22.20
2	3,256	14.12	2,875	12.8	3,071	13.92	2,755	13.11
3	3,708	19.12	3,455	16.75	3,982	19.05	3,515	17.45
4	3,314	19.88	3,386	19.49	4,446	23.24	4,316	23.59
5	2,442	17.71	3,489	23.64	2,624	17	4,385	27.70
6	1,890	15.52	2,172	17.45	1,545	11.79	2,674	19.57
7	1,192	10.1	1,873	16.29	1,675	13.4	1,886	14.93
8	1,608	14.57	2,607	22.71	2,104	17.09	2,538	20.04
9	1,612	12.29	1,896	15.46	1,470	11.04	1,827	12.48
10	1,142	9.68	1,644	13.83	1,407	11.32	1,589	11.97
11	897	11.6	1,032	12.8	1,187	13.26	1,048	11.81
12	726	14.51	836	14.78	917	15.25	969	15.00
Total	27,110	13.85	32,076	16.10	33,002	15.8	35,622	16.80

Table 51—Exit Rate by Grade Span for 2008, 2009, 2010, and 2011 Administrations

Grade	2008		2009		2010		2011	
	N	Percent	N	Percent	N	Percent	N	Percent
Span K-1	5,323	10.68	6,811	13.44	8,574	16.25	8,120	15.89
Span 2-4	10,278	17.38	9,716	16.07	11,499	18.52	10,586	17.81
Span 5-6	4,332	16.68	5,661	20.81	4,169	14.61	7,059	23.93
Span 7-8	2,800	12.26	4,480	19.5	3,779	15.23	4,424	17.49
Span 9-12	4,377	11.63	5,408	14.28	4,981	12.24	5,433	12.56
Total	27,110	13.85	32,076	16.10	33,002	15.80	35,622	17.00

APPENDIX A: ITEM MAP FOR 2011 NYSESLAT

A. 1: 2011 NYSESLAT Operational Test Information Grades K–1

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
1	SS	2	2	Make predictions and inferences, and discuss the meaning of literary works to understand text presented orally and in written form	Sentence Completion
2	SS	2	2	Present personal responses to published literature through words or pictures, referring to features of the text	Sentence Completion
3	SS	2	3	Form and express responses to ideas through reading, listening, viewing, discussing, and writing	Sentence Completion
4	SS	2	2	Create personal stories, using appropriate vocabulary and elements of the literature students have read or heard	Sentence Completion
5	SS	2	3	Apply learning strategies to explore a variety of materials	Sentence Completion
6	SE	4	2	Create personal stories, using appropriate vocabulary and elements of the literature students have read or heard	Storytelling
7	SS	2	5	Interpret and demonstrate knowledge of nonverbal communication, and understand the contexts in which they are used appropriately	Picture Description
8	SS	2	4	Explain actions, choices, and decisions in social and classroom situations	Picture Description
9	SS	2	3	Form and express responses to ideas through reading, listening, viewing, discussing, and writing	Picture Description
10	SS	2	4	Explain actions, choices, and decisions in social and classroom situations	Picture Description
11	SS	2	4	Describe, read about, participate in, or recommend a favorite activity, book, song, or other interest to various audiences	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	5	Share cross-cultural experiences and ideas with others	Social Interaction

A.1: 2011 NYSESLAT Operational Test Information Grades K–1 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
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	Become familiar with some conventions of English	Word/Sentence Comprehension
2	MC	1	1	Become familiar with some conventions of English	Word/Sentence Comprehension
3	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful	Word/Sentence Comprehension
4	MC	1	1	Compare, contrast, and categorize to gain a deeper understanding of information and objects	Word/Sentence Comprehension
5	MC	1	1	Select information appropriate to the purpose of the investigation	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	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful	Word/Sentence Comprehension
8	MC	1	1	Select information appropriate to the purpose of the investigation	Word/Sentence Comprehension
9	MC	1	1	Select information appropriate to the purpose of the investigation	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	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful	Word/Sentence Comprehension
12	MC	1	1	Compare, contrast, and categorize to gain a deeper understanding of information and objects	Word/Sentence Comprehension
13	MC	1	1	Compare, contrast, and categorize to gain a deeper understanding of information and objects	Word/Sentence Comprehension
14	MC	1	1	Compare, contrast, and categorize to gain a deeper understanding of information and objects	Word/Sentence Comprehension
15	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful	Comprehension of Conversational Language
16	MC	1	2	Use basic reading and listening strategies to make literary text comprehensible and meaningful	Comprehension of Conversational Language

A.1: 2011 NYSESLAT Operational Test Information Grades K–1 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
17	MC	1	5	Demonstrate familiarity with some U.S. cultural referents at the local and national levels	Comprehension of Conversational Language
18	MC	1	1	Compare, contrast, and categorize to gain a deeper understanding of information and objects	Comprehension of Conversational Language
19	MC	1	4	Use a variety of oral, print, and electronic forms for social communication	Comprehension of Conversational Language
20	MC	1	3	Apply learning strategies to explore a variety of materials	Comprehension of Conversational Language
21	MC	1	5	Demonstrate familiarity with some U.S. cultural referents at the local and national levels	Comprehension of Conversational Language
22	MC	1	1	Formulate, ask, and respond to questions to obtain	Comprehension of Conversational Language
23	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful	Comprehension of Conversational Language
24	MC	1	1	Select information appropriate to the purpose of the investigation	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	Select information appropriate to the purpose of the investigation	Word Reading
5	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful	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	Formulate, ask, and respond to questions to obtain	Word Reading
8	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful	Word Reading

A.1: 2011 NYSESLAT Operational Test Information Grades K–1 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
9	MC	1	1	Identify and use basic reading and listening strategies to make text comprehensible and meaningful	Word Reading
10	MC	1	1	Select information appropriate to the purpose of the investigation	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	Become familiar with some conventions of English	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	3	Recognize how structural features affect readers' and listeners' understanding and appreciation of text	Short Passage with Question
Writing					
1	MC	1	1	Become familiar with some conventions of English	Phonemic Understanding
2	MC	1	1	Become familiar with some conventions of English	Phonemic Understanding
3	MC	1	1	Become familiar with some conventions of English	Phonemic Understanding
4	MC	1	1	Become familiar with some conventions of English	Phonemic Understanding
5	MC	1	1	Become familiar with some conventions of English	Phonemic Understanding
6	MC	1	1	Become familiar with some conventions of English	Phonemic Understanding
7	CR	2	1	Present ideas clearly in written form	Developmental Writing
8	CR	2	1	Present ideas clearly in written form	Developmental Writing
9	CR	2	1	Express and develop ideas and understanding, using some elements of the "writing process." Such as word mapping, brainstorming, drawing, and writing letters, words, and simple sentences	Developmental Writing
10	CR	2	1	Express and develop ideas and understanding, using some elements of the "writing process." Such as word mapping, brainstorming, drawing, and writing letters, words, and simple sentences	Developmental Writing
11	CR	2	1	Become familiar with some conventions of English	Developmental Writing

A.1: 2011 NYSESLAT Operational Test Information Grades K–1 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
12	CR	2	1	Become familiar with some conventions of English	Developmental Writing
13	CR	2	1	Become familiar with some conventions of English	Developmental Writing
14	CR	2	1	Become familiar with some conventions of English	Developmental Writing
15	ER	4	3	Form and express responses to ideas through reading, listening, viewing, discussing, and writing	Extended Response

A.1: 2011 NYSESLAT Operational Test Information Grades 2–4

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
1	SS	2	1	Make and support inferences about information and ideas with reference to features in oral and written text	Sentence Completion
2	SS	2	1	Make and support inferences about information and ideas with reference to features in oral and written text	Sentence Completion
3	SS	2	1	Convey information, using a variety of organizational patterns and structures	Sentence Completion
4	SS	2	3	Recognize personal point of view in self and others in discussing, interpreting, and evaluating information	Sentence Completion
5	SS	2	1	Formulate, ask, and respond to questions to obtain, clarify, and extend information and meaning	Sentence Completion
6	SE	6	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
7	SS	2	1	Convey information, using a variety of organizational patterns and structures	Picture Description
8	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
9	SS	2	5	Interpret and demonstrate knowledge of nonverbal and oral communication features, and understand the contexts in which they are used appropriately	Picture Description
10	SS	2	4	Explain actions, choices, and decisions in social and academic situations.	Picture Description
11	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes	Picture Description
12	SS	2	4	Use appropriate vocabulary, expressions, language, routines, and interaction styles for various audiences and formal and informal social or school situations	Social Interaction
13	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes	Social Interaction

A.1: 2011 NYSESLAT Operational Test Information Grades 2–4 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
14	SS	2	4	Negotiate and manage interactions to accomplish social and classroom tasks	Social Interaction
15	SS	2	3	Recognize personal point of view in self and others in discussing, interpreting, and evaluating information	Social Interaction
16	SS	2	4	Describe, read about, participate in, or recommend a favorite activity, book, song, or other interest to various audiences	Social Interaction
Listening					
1	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Word/Sentence Comprehension
2	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Word/Sentence Comprehension
3	MC	1	1	Select information appropriate to the purpose of the investigation, and relate ideas from one written or spoken source to another	Word/Sentence Comprehension
4	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful	Word/Sentence Comprehension
5	MC	1	1	Compare, contrast, and categorize, to gain a deeper understanding of information and objects	Word/Sentence Comprehension
6	MC	1	1	Compare, contrast, and categorize, to gain a deeper understanding of information and objects	Word/Sentence Comprehension
7	MC	1	1	Select information appropriate to the purpose of the investigation, and relate ideas from one written or spoken source to another	Word/Sentence Comprehension
8	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Word/Sentence Comprehension
9	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful	Word/Sentence Comprehension
10	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Word/Sentence Comprehension
11	MC	1	5	Recognize and share cross-cultural experiences, and ideas, and connect with those of others	Comprehension of Conversational Language

A.1: 2011 NYSESLAT Operational Test Information Grades 2–4 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
12	MC	1	4	Follow oral and written directions to participate in classroom and social activities	Comprehension of Conversational Language
13	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	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	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension of Conversational Language
16	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension of Conversational Language
17	MC	1	2	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension of Conversational Language
18	MC	1	1	Formulate, ask, and respond to questions to obtain, clarify, and extend information and meaning	Comprehension of Conversational Language
19	MC	1	2	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension of Conversational Language
20	MC	1	4	Follow oral and written directions to participate in classroom and social activities	Task-based
21	MC	1	1	Distinguish between fact and opinion, and relevant and irrelevant information	Task-based
22	MC	1	4	Follow oral and written directions to participate in classroom and social activities	Task-based
23	MC	1	4	Follow oral and written directions to participate in classroom and social activities	Task-based
24	MC	1	4	Explain actions, choices, and decisions in social and academic situations	Task-based
Reading					
1	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful.	Word/Sentence Reading
2	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful.	Word/Sentence Reading
3	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text.	Word/Sentence Reading

A.1: 2011 NYSESLAT Operational Test Information Grades 2–4 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
4	MC	1	1	Compare, contrast, and categorize, to gain a deeper understanding of information and objects	Word/Sentence Reading
5	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Word/Sentence Reading
6	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful	Word/Sentence Reading
7	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Word/Sentence Reading
8	MC	1	2	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension
9	MC	1	2	Make predictions, inferences, and deductions, and discuss the meaning of literary works with some attention to meaning beyond the literal level, to understand and interpret text presented orally and in written form	Comprehension
10	MC	1	2	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension
11	MC	1	2	Make predictions, inferences, and deductions, and discuss the meaning of literary works with some attention to meaning beyond the literal level, to understand and interpret text presented orally and in written form	Comprehension
12	MC	1	1	Apply learning strategies to acquire information and make oral and written texts comprehensible and meaningful	Comprehension
13	MC	1	3	Recognize personal point of view in self and others in discussing, interpreting, and evaluating information	Comprehension
14	MC	1	5	Recognize and share cross-cultural experiences, and ideas, and connect with those of others	Comprehension
15	MC	1	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	Comprehension
16	MC	1	2	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension

A.1: 2011 NYSESLAT Operational Test Information Grades 2–4 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
17	MC	1	2	Make predictions, inferences, and deductions, and discuss the meaning of literary works with some attention to meaning beyond the literal level, to understand and interpret text presented orally and in written form	Comprehension
18	MC	1	2	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension
19	MC	1	2	Apply learning strategies to comprehend and make inferences about literature and produce literary responses	Comprehension
20	MC	1	2	Make predictions, inferences, and deductions, and discuss the meaning of literary works with some attention to meaning beyond the literal level, to understand and interpret text presented orally and in written form	Comprehension
21	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Comprehension
22	MC	1	1	Select information appropriate to the purpose of the investigation, and relate ideas from one written or spoken source to another	Comprehension
23	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Comprehension
24	MC	1	3	Recognize and explain how structural features affect readers' and listeners' understanding and appreciation of text	Comprehension
Writing					
1	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Phonemic Understanding
2	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Phonemic Understanding
3	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English	Phonemic Understanding
4	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Phonemic Understanding
5	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English	Mechanics & Structure

A.1: 2011 NYSESLAT Operational Test Information Grades 2–4 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
6	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English	Mechanics & Structure
7	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English	Mechanics & Structure
8	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English	Mechanics & Structure
9	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English	Mechanics & Structure
10	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English	Mechanics & Structure
11	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English	Mechanics & Structure
12	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 as support	Mechanics & Structure
13	CR	2	1	Make and support inferences about information and ideas with reference to features in oral and written text	Pre-Writing
14	CR	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	Pre-Writing
15	CR	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	Pre-Writing
16	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)	Extended Response

A.1: 2011 NYSESLAT Operational Test Information Grades 5–6

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
1	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Sentence Completion
2	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes	Sentence Completion
3	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Sentence Completion
4	SS	2	5	Recognize and share cross-cultural experiences and ideas, and connect with those of others	Sentence Completion
5	SS	2	1	Formulate, ask, and respond to various question forms to obtain, clarify, and extend information and meaning	Sentence Completion
6	SE	4	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures	Storytelling
7	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Picture Description
8	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Picture Description
9	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Picture Description
10	SS	2	4	Explain actions, choices, and decisions in social and academic situations	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	5	Recognize and share cross-cultural experiences and ideas, and connect with those of others	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	Explain actions, choices, and decisions in social and academic situations	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.1: 2011 NYSESLAT Operational Test Information Grades 5–6 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
Listening					
1	MC	1	1	Compare, contrast, categorize, and synthesize to gain a deeper understanding of information and objects	Word/Sentence Comprehension
2	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Word/Sentence Comprehension
3	MC	1	3	Recognize, explain, and evaluate how structural features affect readers' and listeners' understanding and appreciation of text	Word/Sentence Comprehension
4	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
5	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
6	MC	1	1	Distinguish between fact and opinion, and relevant and irrelevant information, and exclude nonessential information in oral and written presentations	Comprehension of Conversational Language
7	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
8	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
9	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
10	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
11	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

A.1: 2011 NYSESLAT Operational Test Information Grades 5–6 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
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	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
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	1	Make and support inferences about information and ideas with reference to features in oral and written text	Comprehension of Conversational Language
16	MC	1	4	Follow oral and written directions to participate in classroom and social activities, and provide directions to peers in selected interactions	Task-based
17	MC	1	4	Follow oral and written directions to participate in classroom and social activities, and provide directions to peers in selected interactions	Task-based
18	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	Task-based
19	MC	1	4	Follow oral and written directions to participate in classroom and social activities, and provide directions to peers in selected interactions	Task-based
20	MC	1	1	Distinguish between fact and opinion, and relevant and irrelevant information, and exclude nonessential information in oral and written presentations	Task-based
21	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful	Task-based
22	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	Task-based
23	MC	1	1	Compare, contrast, categorize, and synthesize to gain a deeper understanding of information and objects	Task-based
24	MC	1	4	Negotiate and manage interactions to accomplish social and classroom tasks	Task-based

A.1: 2011 NYSESLAT Operational Test Information Grades 5–6 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
25	MC	1	4	Negotiate and manage interactions to accomplish social and classroom tasks	Task-based
Reading					
1	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Word/Sentence Reading
2	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Word/Sentence Reading
3	MC	1	1	Compare, contrast, categorize, and synthesize to gain a deeper understanding of information and objects	Word/Sentence Reading
4	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension
5	MC	1	1	Identify and use reading and listening 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	Select information appropriate to the purpose of the investigation, relate ideas from one written or spoken source to another, and exclude nonessential information	Comprehension
8	MC	1	3	Apply learning strategies to examine and interpret a variety of materials	Comprehension
9	MC	1	2	Identify and use reading and listening strategies to make literary text comprehensible and meaningful	Comprehension
10	MC	1	2	Apply learning strategies to comprehend and make inferences about literature and produce literary responses	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	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.1: 2011 NYSESLAT Operational Test Information Grades 5–6 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
13	MC	1	1	Distinguish between fact and opinion, and relevant and irrelevant information, and exclude nonessential information in oral and written presentations	Comprehension
14	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension
15	MC	1	3	Apply learning strategies to examine and interpret a variety of materials	Comprehension
16	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful	Comprehension
17	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
18	MC	1	5	Compare and contrast oral traditions, myths, folktales, and literature from the United States and international regions and cultures, including the students' own, and identify similarities, differences, and universal cultural themes	Comprehension
19	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
20	MC	1	1	Compare, contrast, categorize, and synthesize to gain a deeper understanding of information and objects	Comprehension
21	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
22	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Comprehension
23	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	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

A.1: 2011 NYSESLAT Operational Test Information Grades 5–6 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
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 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	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
Writing					
1	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Phonemic Understanding
2	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Phonemic Understanding
3	MC	1	1	Convey information and ideas through spoken and 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 spoken and 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 spoken and 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 spoken and 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 spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure

A.1: 2011 NYSESLAT Operational Test Information Grades 5–6 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
8	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
9	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
10	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
11	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
12	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
13	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
14	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
15	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
16	CR	2	3	Recognize, explain, and evaluate how structural features affect readers' and listeners' understanding and appreciation of text	Pre-Writing
17	CR	2	3	Apply learning strategies to examine and interpret a variety of materials	Pre-Writing
18	CR	2	3	Recognize and communicate personal and multiple points of view within and among groups, in discussing, interpreting, and evaluating information in texts and presentations	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	Extended Response

A.1: 2011 NYSESLAT Operational Test Information Grades 7–8

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
1	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Sentence Completion
2	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Sentence Completion
3	SS	2	4	Use appropriate vocabulary, expressions, language, routines, and interaction styles for various audiences and formal and informal social or school situations, noticing how intention is realized through language	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	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 styles appropriate to different audiences, and purpose	Storytelling
7	SS	2	1	Make and support inferences about information and ideas with reference to features in oral and written text	Social Interaction
8	SS	2	3	Recognize, explain, and evaluate how structural features affect readers' and listeners' understanding and appreciation of text	Social Interaction
9	SS	2	1	Formulate, ask, and respond to various question forms to obtain, clarify, and extend information and meaning	Social Interaction
10	SS	2	5	Interpret and demonstrate knowledge of nonverbal and oral communication features, and understand the contexts in which they are used appropriately	Social Interaction
11	SS	2	1	Formulate, ask, and respond to various question forms to obtain, clarify, and extend information and meaning	Social Interaction
12	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Picture Description
13	SS	2	1	Formulate, ask, and respond to various question forms to obtain, clarify, and extend information and meaning	Picture Description
14	SS	2	4	Request and provide information and assistance, orally or in writing, for personal, social, and academic purposes	Picture Description

A.1: 2011 NYSESLAT Operational Test Information Grades 7–8 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
15	SS	2	1	Select a focus, organization, and point of view for oral and written presentations, and justify this selection	Picture Description
16	SS	2	4	Listen attentively, take turns speaking, and build on others' ideas when engaged in pair, group, or full-class discussions on personal, social, community, and academic topics	Picture Description
Listening					
1	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Word/Sentence Comprehension
2	MC	1	3	Apply learning strategies to examine and interpret a variety of materials	Word/Sentence Comprehension
3	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Word/Sentence Comprehension
4	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension of Conversational Language
5	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension of Conversational Language
6	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Comprehension of Conversational Language
7	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension of Conversational Language
8	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension of Conversational Language
9	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension of Conversational Language
10	MC	1	3	Apply learning strategies to examine and interpret a variety of materials	Comprehension of Conversational Language
11	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

A.1: 2011 NYSESLAT Operational Test Information Grades 7–8 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
12	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 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	Comprehension of Conversational Language
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	Make and support inferences about information and ideas with reference to features in oral and written text	Comprehension of Conversational Language
16	MC	1	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures	Task-based
17	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	Task-based
18	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Task-based
19	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Task-based
20	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Task-based
21	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	Task-based
22	MC	1	1	Read, gather, view, listen to, organize, discuss, interpret, and analyze information related to academic content areas from various sources	Task-based
23	MC	1	1	Read, gather, view, listen to, organize, discuss, interpret, and analyze information related to academic content areas from various sources	Task-based

A.1: 2011 NYSESLAT Operational Test Information Grades 7–8 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
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	Task-based
25	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	Task-based
Reading					
1	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Word/Sentence Reading
2	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Word/Sentence Reading
3	MC	1	1	Compare, contrast, categorize, and synthesize to gain a deeper understanding of information and objects	Word/Sentence Reading
4	MC	1	2	Identify and use reading and listening strategies to make literary text comprehensible and meaningful	Comprehension
5	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
6	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
7	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
8	MC	1	1	Present information clearly in a variety of oral and written forms for different audiences and purposes related to all academic content areas	Comprehension
9	MC	1	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures	Comprehension

A.1: 2011 NYSESLAT Operational Test Information Grades 7–8 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
10	MC	1	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures	Comprehension
11	MC	1	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures	Comprehension
12	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful	Comprehension
13	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
14	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
15	MC	1	2	Compose and present personal and formal responses to and interpretations of published literary works and the work of peers, referring to details and features of text	Comprehension
16	MC	1	2	Identify and use reading and listening strategies to make literary text comprehensible and meaningful	Comprehension
17	MC	1	2	Read, listen to, view, write about, and discuss texts and performances from a wide range of authors, subjects, and genres	Comprehension
18	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
19	MC	1	2	Apply learning strategies to comprehend and make inferences about literature and produce literary responses	Comprehension
20	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.1: 2011 NYSESLAT Operational Test Information Grades 7–8 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
21	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
22	MC	1	3	Apply learning strategies to examine and interpret a variety of materials	Comprehension
23	MC	1	3	Recognize, explain, and evaluate how structural features affect readers' and listeners' understanding and appreciation of text	Comprehension
24	MC	1	3	Apply learning strategies to examine and interpret a variety of materials	Comprehension
25	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Comprehension
26	MC	1	3	Recognize and communicate personal and multiple points of view within and among groups, in discussing, interpreting, and evaluating information in texts and presentations	Comprehension
27	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
Writing					
1	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Phonemic Understanding
2	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Phonemic Understanding
3	MC	1	1	Convey information and ideas through spoken and 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 spoken and 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 spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure

A.1: 2011 NYSESLAT Operational Test Information Grades 7–8 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
6	MC	1	1	Convey information and ideas through spoken and 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 spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
8	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
9	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
10	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
11	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
12	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
13	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
14	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
15	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure

A.1: 2011 NYSESLAT Operational Test Information Grades 7–8 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
16	CR	2	3	Apply learning strategies to examine and interpret a variety of materials	Pre-Writing
17	CR	2	3	Apply learning strategies to examine and interpret a variety of materials	Pre-Writing
18	CR	2	3	Apply learning strategies to examine and interpret a variety of materials	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	Extended Response

A.1: 2011 NYSESLAT Operational Test Information Grades 9–12

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
Speaking					
1	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Sentence Completion
2	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Sentence Completion
3	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Sentence Completion
4	SS	2	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences; justify and explain the rationale for positions, using persuasive language, tone, evidence, and well-developed arguments	Sentence Completion
5	SS	2	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials	Sentence Completion
6	SE	4	2	Create stories, poems, sketches, songs, and plays, including those that reflect traditional and popular American culture, using typical features of a given genre; create an effective voice, using a variety of writing styles appropriate to different audiences	Storytelling
7	SS	2	1	Make and support inferences about information and ideas with reference to features in oral and written text	Picture Description
8	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Picture Description
9	SS	2	1	Make and support inferences about information and ideas with reference to features in oral and written text	Picture Description
10	SS	2	1	Make and support inferences about information and ideas with reference to features in oral and written text	Picture Description
11	SS	2	4	Explain actions, choices, and decisions in social and academic situations	Picture Description
12	SS	2	4	Explain actions, choices, and decisions in social and academic situations	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

A.1: 2011 NYSESLAT Operational Test Information Grades 9–12 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
16	SS	2	4	Describe, read about, participate in, or recommend a favorite activity, book, song, or other interest to various audiences	Social Interaction
Listening					
1	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful.	Word/Sentence Comprehension
2	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Word/Sentence Comprehension
3	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Word/Sentence Comprehension
4	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful	Comprehension of Conversational Language
5	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful	Comprehension of Conversational Language
6	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions	Comprehension of Conversational Language
7	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions	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 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
10	MC	1	3	Recognize, explain, evaluate, and analyze how structural features affect readers' and listeners' understanding and appreciation of text	Comprehension of Conversational Language
11	MC	1	1	Select information appropriate to the purpose of the investigation with suitable supporting material	Comprehension of Conversational Language

A.1: 2011 NYSESLAT Operational Test Information Grades 9–12 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
12	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 students' native community	Comprehension of Conversational Language
13	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 students' native community	Comprehension of Conversational Language
14	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful	Comprehension of Conversational Language
15	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials	Comprehension of Conversational Language
16	MC	1	1	Select information appropriate to the purpose of the investigation with suitable supporting material	Task-based
17	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Task-based
18	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions	Task-based
19	MC	1	4	Follow oral and written directions to participate in classroom activities, and provide directions to peers in selected interactions	Task-based
20	MC	1	1	Read, gather, view, listen to, organize, discuss, interpret, and analyze information related to academic content areas from various sources	Task-based
21	MC	1	1	Convey and organize information, using facts, details, illustrative examples, and a variety of patterns and structures.	Task-based
22	MC	1	1	Select information appropriate to the purpose of the investigation with suitable supporting material	Task-based
23	MC	1	3	Recognize, explain, evaluate, and analyze how structural features affect readers' and listeners' understanding and appreciation of text	Task-based

A.1: 2011 NYSESLAT Operational Test Information Grades 9–12 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
24	MC	1	1	Select information appropriate to the purpose of the investigation with suitable supporting material	Task-based
25	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials	Task-based
Reading					
1	MC	1	1	Compare, contrast, categorize, synthesize information and objects, and identify complexities and discrepancies in the information	Word/Sentence Reading
2	MC	1	1	Select information appropriate to the purpose of the investigation with suitable supporting material	Word/Sentence Reading
3	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful	Word/Sentence Reading
4	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension
5	MC	1	1	Apply learning strategies to acquire information and make texts comprehensible and meaningful	Comprehension
6	MC	1	1	Compare, contrast, categorize, synthesize information and objects, and identify complexities and discrepancies in the information	Comprehension
7	MC	1	1	Select information appropriate to the purpose of the investigation with suitable supporting material	Comprehension
8	MC	1	5	Compare and contrast oral traditions, myths, folktales, and literature from different national and international regions and cultures, including the students' own, identifying similarities and differences and universal cultural themes, and exploring how language and literature transmit culture	Comprehension
9	MC	1	3	Recognize and communicate personal and multiple points of view within and among groups, in discussing, interpreting, and evaluating information; make inferences about a writer or speaker's point of view	Comprehension
10	MC	1	3	Recognize and communicate personal and multiple points of view within and among groups, in discussing, interpreting, and evaluating information; make inferences about a writer or speaker's point of view	Comprehension

A.1: 2011 NYSESLAT Operational Test Information Grades 9–12 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
11	MC	1	3	Recognize, explain, evaluate, and analyze how structural features affect readers' and listeners' understanding and appreciation of text	Comprehension
12	MC	1	5	Recognize and share cross-cultural experiences and idea, and connect with those of others	Comprehension
13	MC	1	1	Compare, contrast, categorize, synthesize information and objects, and identify complexities and discrepancies in the information	Comprehension
14	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension
15	MC	1	3	Recognize and communicate personal and multiple points of view within and among groups, in discussing, interpreting, and evaluating information; make inferences about a writer or speaker's point of view	Comprehension
16	MC	1	1	Identify and use reading and listening strategies to make text comprehensible and meaningful	Comprehension
17	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Comprehension
18	MC	1	2	Apply reading and listening strategies to make literary text comprehensible and meaningful	Comprehension
19	MC	1	2	Read, listen to, view, write about, and discuss a variety of texts from a wide range of authors, subjects, genres, cultures, and historical periods	Comprehension
20	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
21	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
22	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

A.1: 2011 NYSESLAT Operational Test Information Grades 9–12 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
23	MC	1	3	Apply learning strategies to examine, interpret, analyze, synthesize, and evaluate a variety of materials	Comprehension
24	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
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	Apply self-monitoring and self-correcting strategies while reading, viewing, discussing, listening to, or producing literary texts and essays	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 spoken and 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 spoken and 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 spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
4	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	Mechanics & Structure

A.1: 2011 NYSESLAT Operational Test Information Grades 9–12 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
5	MC	1	1	Apply self-monitoring and self-correcting strategies for accurate language production and oral and written presentation, using established criteria for effective presentation of information	Mechanics & Structure
6	MC	1	1	Convey information and ideas through spoken and 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 spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
8	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
9	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
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, and identify and interpret how these patterns and norms are used	Mechanics & Structure
11	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
12	MC	1	1	Apply self-monitoring and self-correcting strategies for accurate language production and oral and written presentation, using established criteria for effective presentation of information	Mechanics & Structure
13	MC	1	1	Apply self-monitoring and self-correcting strategies for accurate language production and oral and written presentation, using established criteria for effective presentation of information	Mechanics & Structure

A.1: 2011 NYSESLAT Operational Test Information Grades 9–12 (continued)

Modality/ Position	Type	Score Point	Learning Standard	Performance Indicator	Cluster
14	MC	1	1	Convey information and ideas through spoken and written language, using conventions and features of American English appropriate to audience and purpose	Mechanics & Structure
15	MC	1	1	Make and support inferences about information and ideas with reference to features in oral and written text	Mechanics & Structure
16	CR	2	5	Recognize and share cross-cultural experiences and idea, and connect with those of others	Pre-Writing
17	CR	2	5	Recognize and share cross-cultural experiences and idea, and connect with those of others	Pre-Writing
18	CR	2	3	Develop and present clear interpretations, analyses, and evaluations of issues, ideas, texts, and experiences; justify and explain the rationale for positions, using persuasive language, tone, evidence, and well-developed arguments	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)	Extended Response

A. 1: 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.2: 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.2: 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?” and “What is the boy doing?” (Grades K–1); “Which word tells what is in the picture?” and “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.2: 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.2: 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.2: 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.2: 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.2: 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.2: 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.2: 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.2: 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.1a: Kindergarten–Grade 1 Listening and Speaking**

	Item Number	Item Format	Max Point	N-count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
LI	1	MC	1	50,542		4	93	2		1	0.94	0.36
	2	MC	1	50,517		4	92	3		1	0.93	0.36
	3	MC	1	50,501		95	2	3		1	0.95	0.39
	4	MC	1	50,612		4	94	1		0	0.95	0.35
	5	MC	1	50,594		2	95	3		0	0.96	0.38
	6	MC	1	50,562		1	1	97		0	0.98	0.38
	7	MC	1	50,632		2	94	4		0	0.95	0.36
	8	MC	1	50,613		16	82	1		0	0.83	0.41
	9	MC	1	50,586		88	7	5		0	0.89	0.42
	10	MC	1	50,658		1	2	96		0	0.96	0.33
	11	MC	1	50,587		20	76	3		0	0.77	0.32
	12	MC	1	50,601		7	88	5		0	0.88	0.33
	13	MC	1	50,616		8	2	89		0	0.89	0.53
	14	MC	1	50,557		81	4	14		0	0.82	0.34
	15	MC	1	50,561		3	95	1		0	0.96	0.40
	16	MC	1	50,513		3	3	93		1	0.94	0.45
	17	MC	1	50,524		6	88	6		1	0.89	0.35
	18	MC	1	50,453		89	5	5		1	0.90	0.31
	19	MC	1	50,548		9	2	89		1	0.89	0.32
	20	MC	1	50,524		84	5	10		1	0.85	0.41
	21	MC	1	50,541		9	23	67		1	0.68	0.49
	22	MC	1	50,477		12	32	56		1	0.56	0.31
	23	MC	1	50,548		11	77	11		1	0.78	0.28
	24	MC	1	50,457		6	64	29		1	0.64	0.34
SP	1	SS	2	50,791	9	30	61			0	0.76	0.71
	2	SS	2	50,791	6	21	73			0	0.84	0.74
	3	SS	2	50,791	10	30	60			0	0.75	0.72
	4	SS	2	50,791	9	27	64			0	0.78	0.68
	5	SS	2	50,791	7	24	69			0	0.81	0.74
	6	SE	4	50,791	6	9	27	33	25	0	0.65	0.82
	7	SS	2	50,791	6	30	63			0	0.78	0.76
	8	SS	2	50,791	6	38	57			0	0.76	0.72
	9	SS	2	50,791	6	24	70			0	0.82	0.75
	10	SS	2	50,791	6	36	59			0	0.76	0.74
	11	SS	2	50,791	7	34	58			0	0.76	0.75
	12	SS	2	50,791	9	14	78			0	0.84	0.74
	13	SS	2	50,791	7	17	76			0	0.84	0.73
	14	SS	2	50,791	6	11	83			0	0.89	0.73
	15	SS	2	50,791	8	19	73			0	0.83	0.75
	16	SS	2	50,791	11	18	71			0	0.80	0.73

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=208,596) after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

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

	Item Number	Item Format	Max Point	N-count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
RD	1	MC	1	50,418		7	88	4		1	0.89	0.47
	2	MC	1	50,587		84	5	10		0	0.85	0.33
	3	MC	1	50,560		9	6	85		0	0.85	0.57
	4	MC	1	50,410		15	13	71		1	0.72	0.56
	5	MC	1	50,559		21	75	3		0	0.76	0.53
	6	MC	1	50,473		11	8	80		1	0.81	0.64
	7	MC	1	50,377		9	84	6		1	0.85	0.56
	8	MC	1	50,468		67	24	9		1	0.68	0.37
	9	MC	1	50,469		16	9	74		1	0.75	0.65
	10	MC	1	50,371		13	70	15		1	0.71	0.58
	11	MC	1	50,309		9	28	62		1	0.62	0.60
	12	MC	1	50,315		16	74	9		1	0.75	0.44
	13	MC	1	50,280		11	22	67		1	0.67	0.49
	14	MC	1	50,099		25	61	13		1	0.62	0.34
	15	MC	1	50,093		10	80	8		1	0.82	0.39
WR	1	MC	1	50,543		3	2	95		1	0.95	0.44
	2	MC	1	50,548		90	7	3		1	0.90	0.41
	3	MC	1	50,518		11	83	5		1	0.84	0.45
	4	MC	1	50,494		6	86	8		1	0.86	0.37
	5	MC	1	50,470		88	4	7		1	0.89	0.35
	6	MC	1	50,498		92	2	5		1	0.93	0.39
	7	CR	2	50,783	2	30	68			0	0.83	0.47
	8	CR	2	50,783	2	42	56			0	0.77	0.48
	9	CR	2	50,783	4	2	93			0	0.94	0.48
	10	CR	2	50,783	6	1	93			0	0.93	0.47
	11	CR	2	50,783	14	15	71			0	0.79	0.74
	12	CR	2	50,783	15	22	63			0	0.74	0.79
	13	CR	2	50,783	26	29	45			0	0.60	0.82
	14	CR	2	50,783	25	33	42			0	0.58	0.83
	15	CR	2	50,783	33	37	30			0	0.48	0.80

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=208,596) after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

B.2a: 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	58,868		96	2	2		0	0.96	0.45	
	2	MC	1	58,814		7	12	81		0	0.81	0.33	
	3	MC	1	58,887		7	3	90		0	0.90	0.50	
	4	MC	1	58,890		89	8	3		0	0.89	0.31	
	5	MC	1	58,810		17	69	14		0	0.69	0.40	
	6	MC	1	58,915		94	3	3		0	0.94	0.45	
	7	MC	1	58,910		1	95	3		0	0.96	0.43	
	8	MC	1	58,896		73	20	7		0	0.73	0.36	
	9	MC	1	58,882		73	22	5		0	0.73	0.37	
	10	MC	1	58,825		3	38	59		0	0.59	0.32	
	11	MC	1	58,862		20	3	70	6		0	0.71	0.27
	12	MC	1	58,857		11	80	3	6		0	0.80	0.34
	13	MC	1	58,867		2	5	89	4		0	0.89	0.49
	14	MC	1	58,857		16	6	64	14		0	0.64	0.41
	15	MC	1	58,837		11	7	7	75		0	0.75	0.36
	16	MC	1	58,843		4	76	10	10		0	0.76	0.45
	17	MC	1	58,821		8	4	81	6		0	0.82	0.45
	18	MC	1	58,826		6	10	13	71		0	0.71	0.39
	19	MC	1	58,766		7	67	16	10		0	0.67	0.34
	20	MC	1	58,859		8	2	88	1		0	0.88	0.34
	21	MC	1	58,800		81	6	5	8		0	0.81	0.36
	22	MC	1	58,852		2	1	3	93		0	0.93	0.50
	23	MC	1	58,810		3	91	3	3		0	0.92	0.42
	24	MC	1	58,790		1	2	89	7		0	0.89	0.44
SP	1	SS	2	59,029	4	21	75			0	0.85	0.72	
	2	SS	2	59,029	6	21	72			0	0.83	0.76	
	3	SS	2	59,029	5	18	77			0	0.86	0.77	
	4	SS	2	59,029	6	20	74			0	0.84	0.75	
	5	SS	2	59,029	5	28	67			0	0.81	0.68	
	6	SE	4	59,029	4	4	17	38	38		0	0.76	0.81
	7	SS	2	59,029	4	20	76			0	0.86	0.76	
	8	SS	2	59,029	6	22	72			0	0.83	0.78	
	9	SS	2	59,029	5	22	73			0	0.84	0.77	
	10	SS	2	59,029	5	27	68			0	0.81	0.76	
	11	SS	2	59,029	6	26	68			0	0.81	0.74	
	12	SS	2	59,029	4	17	79			0	0.88	0.68	
	13	SS	2	59,029	6	13	81			0	0.87	0.74	
	14	SS	2	59,029	7	17	75			0	0.84	0.75	
	15	SS	2	59,029	7	12	81			0	0.87	0.76	
	16	SS	2	59,029	5	13	82			0	0.88	0.72	

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=208,596) after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

B.2b: 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	58,919	93	2	4			0	0.93	0.47
	2	MC	1	58,919	8	2	89			0	0.90	0.47
	3	MC	1	58,871	6	6	89			0	0.89	0.54
	4	MC	1	58,891	94	3	2			0	0.94	0.41
	5	MC	1	58,896	2	91	6			0	0.91	0.46
	6	MC	1	58,865	93	4	3			0	0.93	0.49
	7	MC	1	58,809	3	91	5			0	0.91	0.39
	8	MC	1	58,833	3	3	90	3		0	0.91	0.54
	9	MC	1	58,773	6	18	8	68		0	0.68	0.49
	10	MC	1	58,762	5	79	4	11		0	0.80	0.60
	11	MC	1	58,788	58	7	14	21		0	0.58	0.46
	12	MC	1	58,822	87	3	3	7		0	0.87	0.45
	13	MC	1	58,718	10	7	18	63		1	0.64	0.41
	14	MC	1	58,764	7	66	8	19		0	0.66	0.55
	15	MC	1	58,721	28	37	23	11		1	0.37	0.25
	16	MC	1	58,755	6	30	8	56		0	0.56	0.41
	17	MC	1	58,695	21	63	9	7		1	0.64	0.40
	18	MC	1	58,593	54	17	12	17		1	0.54	0.47
	19	MC	1	58,683	9	10	11	70		1	0.70	0.44
	20	MC	1	58,707	75	11	7	7		1	0.75	0.58
	21	MC	1	58,707	21	22	10	46		1	0.46	0.40
	22	MC	1	58,680	11	63	12	14		1	0.63	0.55
	23	MC	1	58,686	79	5	10	6		1	0.79	0.32
	24	MC	1	58,646	17	29	8	45		1	0.45	0.28
WR	1	MC	1	58,862	5	91	4			0	0.91	0.52
	2	MC	1	58,907	4	2	94			0	0.94	0.45
	3	MC	1	58,880	11	16	73			0	0.73	0.47
	4	MC	1	58,820	83	8	8			0	0.84	0.47
	5	MC	1	58,835	14	59	7	19		0	0.59	0.40
	6	MC	1	58,812	10	73	8	9		0	0.73	0.57
	7	MC	1	58,804	8	7	22	63		0	0.63	0.48
	8	MC	1	58,778	23	67	4	5		0	0.67	0.52
	9	MC	1	58,733	19	9	6	65		1	0.65	0.49
	10	MC	1	58,763	4	6	81	9		0	0.81	0.52
	11	MC	1	58,736	16	9	4	71		1	0.71	0.52
	12	MC	1	58,688	65	13	6	16		1	0.66	0.48
	13	CR	2	59,013	6	14	80			0	0.87	0.64
	14	CR	2	59,013	15	22	63			0	0.74	0.68
	15	CR	2	59,013	15	37	48			0	0.66	0.70
	16	ER	4	59,013	11	22	38	22	7	0	0.48	0.78

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=208,596) after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

B.3a: 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	29,172		2	4	94	0	0	0.94	0.52
	2	MC	1	29,180		15	73	12	0	0	0.73	0.46
	3	MC	1	29,162		72	24	4	0	0	0.72	0.31
	4	MC	1	29,192		2	4	3	91	0	0.91	0.49
	5	MC	1	29,178		4	4	87	4	0	0.88	0.54
	6	MC	1	29,197		3	15	80	2	0	0.80	0.28
	7	MC	1	29,179		12	73	10	4	0	0.73	0.32
	8	MC	1	29,191		88	4	3	5	0	0.88	0.48
	9	MC	1	29,196		6	86	5	3	0	0.86	0.51
	10	MC	1	29,188		4	6	8	82	0	0.82	0.47
	11	MC	1	29,184		7	5	14	74	0	0.75	0.33
	12	MC	1	29,186		78	7	6	8	0	0.79	0.47
	13	MC	1	29,164		6	12	69	12	0	0.70	0.39
	14	MC	1	29,180		2	3	3	91	0	0.92	0.45
	15	MC	1	29,154		64	12	15	8	0	0.64	0.43
	16	MC	1	29,170		85	5	5	5	0	0.85	0.52
	17	MC	1	29,172		13	55	26	5	0	0.55	0.27
	18	MC	1	29,173		15	17	61	6	0	0.62	0.23
	19	MC	1	29,174		3	4	7	85	0	0.85	0.32
	20	MC	1	29,177		1	1	12	84	0	0.85	0.34
	21	MC	1	29,163		9	70	10	10	0	0.70	0.24
	22	MC	1	29,171		1	15	82	2	0	0.82	0.21
	23	MC	1	29,168		29	49	13	9	0	0.49	0.29
	24	MC	1	29,165		82	8	8	2	0	0.82	0.37
	25	MC	1	29,135		13	56	15	16	0	0.56	0.34
SP	1	SS	2	29,240	5	21	74			0	0.84	0.76
	2	SS	2	29,240	5	16	78			0	0.86	0.80
	3	SS	2	29,240	8	21	71			0	0.82	0.78
	4	SS	2	29,240	8	19	73			0	0.82	0.79
	5	SS	2	29,240	9	26	65			0	0.78	0.76
	6	SE	4	29,240	5	5	19	35	36	0	0.73	0.84
	7	SS	2	29,240	5	19	76			0	0.85	0.78
	8	SS	2	29,240	7	21	72			0	0.82	0.81
	9	SS	2	29,240	8	23	69			0	0.80	0.81
	10	SS	2	29,240	6	22	72			0	0.83	0.78
	11	SS	2	29,240	5	16	79			0	0.87	0.80
	12	SS	2	29,240	10	8	82			0	0.86	0.79
	13	SS	2	29,240	9	21	70			0	0.80	0.72
	14	SS	2	29,240	6	9	86			0	0.90	0.77
	15	SS	2	29,240	9	12	79			0	0.85	0.81
	16	SS	2	29,240	12	16	72			0	0.80	0.80

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=208,596) after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score > 900).

B.3b: 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	29,199		2	94	4	0	0	0.94	0.41
	2	MC	1	29,185		7	6	87	0	0	0.88	0.37
	3	MC	1	29,175		7	10	83	0	0	0.83	0.39
	4	MC	1	29,190		2	2	93	2	0	0.94	0.39
	5	MC	1	29,180		63	23	8	6	0	0.63	0.47
	6	MC	1	29,173		3	13	4	80	0	0.80	0.37
	7	MC	1	29,179		3	82	7	7	0	0.82	0.58
	8	MC	1	29,169		6	7	8	79	0	0.79	0.62
	9	MC	1	29,163		25	22	42	10	0	0.42	0.31
	10	MC	1	29,167		7	5	81	7	0	0.81	0.63
	11	MC	1	29,155		16	67	5	11	0	0.67	0.50
	12	MC	1	29,152		75	4	5	15	0	0.75	0.48
	13	MC	1	29,152		15	14	19	53	0	0.53	0.42
	14	MC	1	29,126		80	8	3	8	0	0.80	0.55
	15	MC	1	29,166		9	22	8	60	0	0.60	0.37
	16	MC	1	29,152		9	6	69	15	0	0.69	0.45
	17	MC	1	29,144		70	12	8	9	0	0.70	0.58
	18	MC	1	29,157		8	6	76	10	0	0.76	0.64
	19	MC	1	29,147		10	6	79	5	0	0.79	0.54
	20	MC	1	29,147		6	13	14	66	0	0.67	0.56
	21	MC	1	29,136		54	10	7	28	0	0.54	0.41
	22	MC	1	29,137		12	62	15	11	0	0.62	0.54
	23	MC	1	29,133		9	10	71	9	0	0.72	0.60
	24	MC	1	29,122		12	13	62	12	0	0.62	0.28
	25	MC	1	29,136		8	76	5	10	0	0.76	0.70
	26	MC	1	29,121		8	68	11	12	0	0.69	0.56
	27	MC	1	29,127		7	19	6	68	0	0.68	0.44
WR	1	MC	1	29,179		18	77	5	0	0	0.77	0.27
	2	MC	1	29,176		7	88	5	0	0	0.88	0.46
	3	MC	1	29,185		11	3	83	3	0	0.83	0.43
	4	MC	1	29,182		80	5	11	3	0	0.81	0.57
	5	MC	1	29,177		6	8	56	29	0	0.56	0.41
	6	MC	1	29,173		6	82	6	5	0	0.82	0.59
	7	MC	1	29,158		18	9	53	20	0	0.53	0.40
	8	MC	1	29,177		88	4	3	4	0	0.89	0.57
	9	MC	1	29,182		2	12	6	80	0	0.80	0.50
	10	MC	1	29,182		79	10	4	7	0	0.79	0.54
	11	MC	1	29,160		24	9	19	48	0	0.49	0.38
	12	MC	1	29,164		88	2	2	7	0	0.88	0.51
	13	MC	1	29,151		11	66	17	6	0	0.67	0.44
	14	MC	1	29,136		6	24	60	9	0	0.60	0.38
	15	MC	1	29,139		24	53	6	17	0	0.53	0.35
	16	CR	2	29,239	7	14	79			0	0.86	0.68
	17	CR	2	29,239	12	18	70			0	0.79	0.74
	18	CR	2	29,239	18	35	47			0	0.64	0.68
	19	ER	4	29,239	11	16	36	27	9	0	0.52	0.78

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=208,596) after cleaning (with valid corresponding subtest raw score).

B.4a: 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	25,015		3	96	1		0	0.96	0.31
	2	MC	1	25,001		91	6	3		0	0.91	0.35
	3	MC	1	25,001		82	14	4		0	0.82	0.42
	4	MC	1	25,019		80	9	6	4	0	0.81	0.55
	5	MC	1	24,954		12	57	19	12	1	0.57	0.33
	6	MC	1	25,017		6	16	55	23	0	0.55	0.33
	7	MC	1	24,998		12	19	19	50	0	0.51	0.36
	8	MC	1	24,980		8	8	72	11	0	0.72	0.28
	9	MC	1	25,003		6	7	83	4	0	0.83	0.44
	10	MC	1	24,963		13	63	8	15	1	0.64	0.33
	11	MC	1	24,999		15	8	18	59	0	0.59	0.29
	12	MC	1	24,974		12	18	59	9	0	0.60	0.39
	13	MC	1	25,000		42	6	37	14	0	0.43	0.40
	14	MC	1	24,999		16	5	54	25	0	0.54	0.33
	15	MC	1	24,983		12	74	8	6	0	0.74	0.47
	16	MC	1	24,994		77	13	5	5	0	0.77	0.43
	17	MC	1	25,015		83	4	2	11	0	0.83	0.29
	18	MC	1	25,002		7	12	14	66	0	0.66	0.42
	19	MC	1	24,997		8	8	67	17	0	0.68	0.47
	20	MC	1	25,004		73	10	11	6	0	0.73	0.46
	21	MC	1	25,005		4	79	10	5	0	0.80	0.39
	22	MC	1	25,024		2	3	5	90	0	0.90	0.40
	23	MC	1	25,008		11	66	14	9	0	0.67	0.40
	24	MC	1	24,996		20	16	42	21	0	0.42	0.31
	25	MC	1	24,990		5	7	11	77	0	0.77	0.40
SP	1	SS	2	25,081	8	22	70			0	0.81	0.78
	2	SS	2	25,081	12	23	65			0	0.76	0.80
	3	SS	2	25,081	12	22	66			0	0.77	0.82
	4	SS	2	25,081	7	26	67			0	0.80	0.77
	5	SS	2	25,081	13	23	64			0	0.75	0.82
	6	SE	4	25,081	6	6	15	29	44	0	0.75	0.87
	7	SS	2	25,081	7	24	69			0	0.81	0.79
	8	SS	2	25,081	8	21	71			0	0.82	0.81
	9	SS	2	25,081	8	21	71			0	0.82	0.81
	10	SS	2	25,081	8	22	71			0	0.82	0.80
	11	SS	2	25,081	6	18	76			0	0.85	0.80
	12	SS	2	25,081	11	23	66			0	0.78	0.78
	13	SS	2	25,081	14	20	66			0	0.76	0.82
	14	SS	2	25,081	11	17	71			0	0.80	0.83
	15	SS	2	25,081	20	16	64			0	0.72	0.81
	16	SS	2	25,081	12	17	71			0	0.79	0.81

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=208,596) after cleaning (with valid corresponding subtest raw score).

B.4b: 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	25,042		4	12	84		0	0.84	0.41
	2	MC	1	25,033		90	8	2		0	0.90	0.45
	3	MC	1	25,017		4	3	92		0	0.92	0.36
	4	MC	1	25,030		23	9	60	7	0	0.60	0.54
	5	MC	1	25,010		13	72	10	5	0	0.73	0.49
	6	MC	1	25,032		4	5	6	85	0	0.85	0.47
	7	MC	1	25,031		20	7	60	13	0	0.60	0.49
	8	MC	1	25,012		16	67	12	5	0	0.67	0.41
	9	MC	1	25,019		13	67	7	13	0	0.67	0.58
	10	MC	1	24,995		13	9	67	11	0	0.67	0.49
	11	MC	1	24,998		18	65	11	7	0	0.65	0.47
	12	MC	1	25,020		4	80	9	8	0	0.80	0.45
	13	MC	1	24,987		7	7	9	76	0	0.77	0.55
	14	MC	1	24,972		25	15	12	48	0	0.48	0.32
	15	MC	1	25,022		9	66	17	8	0	0.66	0.51
	16	MC	1	25,030		20	11	7	61	0	0.61	0.40
	17	MC	1	25,016		62	8	17	13	0	0.62	0.47
	18	MC	1	25,005		57	14	16	13	0	0.57	0.49
	19	MC	1	25,009		11	55	26	7	0	0.55	0.38
	20	MC	1	25,004		15	51	26	8	0	0.51	0.48
	21	MC	1	24,999		21	9	26	44	0	0.44	0.39
	22	MC	1	24,971		57	21	11	11	0	0.57	0.54
	23	NS	1	24,973		23	40	25	12	0	0.40	0.34
	24	MC	1	24,962		68	12	13	7	1	0.69	0.54
	25	MC	1	24,975		13	23	10	54	0	0.54	0.43
	26	MC	1	24,973		13	10	9	67	0	0.67	0.53
	27	MC	1	24,981		13	13	54	20	0	0.54	0.45
WR	1	MC	1	24,988		88	8	3		0	0.88	0.30
	2	MC	1	24,996		6	88	5		0	0.88	0.40
	3	MC	1	25,016		40	32	19	7	0	0.40	0.30
	4	MC	1	25,024		16	4	60	20	0	0.60	0.46
	5	MC	1	25,016		4	6	10	80	0	0.80	0.54
	6	MC	1	25,014		3	80	3	13	0	0.80	0.40
	7	MC	1	24,986		5	7	23	65	0	0.65	0.49
	8	MC	1	25,002		9	11	3	77	0	0.77	0.53
	9	MC	1	24,996		53	22	11	14	0	0.53	0.33
	10	MC	1	25,011		24	67	5	3	0	0.67	0.48
	11	MC	1	24,990		8	5	26	61	0	0.61	0.48
	12	MC	1	24,998		10	81	4	4	0	0.81	0.56
	13	MC	1	24,990		65	5	15	15	0	0.65	0.45
	14	MC	1	24,988		22	4	66	7	0	0.67	0.40
	15	MC	1	24,982		2	17	21	59	0	0.60	0.43
	16	CR	2	25,084	21	25	54			0	0.66	0.64
	17	CR	2	25,084	11	14	75			0	0.82	0.62
	18	CR	2	25,084	17	26	56			0	0.70	0.66
	19	ER	4	25,084	13	16	33	26	12	0	0.52	0.77

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=208,596) after cleaning (with valid corresponding subtest raw score).

B.5a: 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	42,651		7	87	5	0	1	0.88	0.38
	2	MC	1	42,487		80	9	10	0	1	0.81	0.48
	3	MC	1	42,665		8	85	5	0	1	0.86	0.46
	4	MC	1	42,741		10	9	75	5	1	0.76	0.45
	5	MC	1	42,558		65	10	9	15	1	0.65	0.41
	6	MC	1	42,775		4	82	5	8	1	0.82	0.30
	7	MC	1	42,785		83	8	5	5	1	0.83	0.41
	8	MC	1	42,690		72	8	10	9	1	0.73	0.39
	9	MC	1	42,699		12	5	4	78	1	0.78	0.50
	10	MC	1	42,727		27	53	13	6	1	0.54	0.38
	11	MC	1	42,718		8	13	49	29	1	0.49	0.41
	12	MC	1	42,722		12	13	68	5	1	0.69	0.48
	13	MC	1	42,726		56	20	18	6	1	0.56	0.33
	14	MC	1	42,699		58	20	16	6	1	0.58	0.42
	15	MC	1	42,668		14	22	11	52	1	0.53	0.26
	16	MC	1	42,674		14	14	12	59	1	0.59	0.38
	17	MC	1	42,751		14	62	13	11	1	0.62	0.38
	18	MC	1	42,700		6	8	25	61	1	0.61	0.45
	19	MC	1	42,766		67	16	5	10	1	0.68	0.60
	20	MC	1	42,729		63	10	5	21	1	0.64	0.51
	21	MC	1	42,715		5	27	13	54	1	0.54	0.40
	22	MC	1	42,739		80	10	4	5	1	0.81	0.49
	23	MC	1	42,733		4	5	87	3	1	0.88	0.42
	24	MC	1	42,717		15	10	14	61	1	0.62	0.42
	25	MC	1	42,708		4	68	20	8	1	0.68	0.24
SP	1	SS	2	43,021	6	33	61			0	0.78	0.71
	2	SS	2	43,021	5	31	64			0	0.79	0.75
	3	SS	2	43,021	8	18	75			0	0.84	0.76
	4	SS	2	43,021	9	29	62			0	0.77	0.74
	5	SS	2	43,021	9	19	72			0	0.81	0.74
	6	SE	4	43,021	5	6	17	32	40	0	0.74	0.86
	7	SS	2	43,021	5	25	70			0	0.82	0.76
	8	SS	2	43,021	8	30	62			0	0.77	0.78
	9	SS	2	43,021	6	28	66			0	0.80	0.76
	10	SS	2	43,021	5	25	70			0	0.83	0.75
	11	SS	2	43,021	8	28	64			0	0.78	0.77
	12	SS	2	43,021	9	24	68			0	0.80	0.77
	13	SS	2	43,021	12	18	70			0	0.79	0.76
	14	SS	2	43,021	10	21	69			0	0.80	0.75
	15	SS	2	43,021	13	23	64			0	0.75	0.76
	16	SS	2	43,021	6	21	73			0	0.83	0.72

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=208,596) after cleaning (with valid corresponding subtest raw score).

B.5b: 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	42,790		11	86	3	0	1	0.87	0.42
	2	MC	1	42,722		6	5	89	0	1	0.90	0.35
	3	MC	1	42,789		84	13	3	0	1	0.84	0.30
	4	MC	1	42,866		92	4	2	2	0	0.92	0.36
	5	MC	1	42,845		15	63	8	13	0	0.64	0.39
	6	MC	1	42,578		21	11	51	16	1	0.52	0.29
	7	MC	1	42,847		6	84	3	7	0	0.84	0.46
	8	MC	1	42,820		13	52	8	26	1	0.53	0.39
	9	MC	1	42,735		68	8	12	11	1	0.69	0.43
	10	MC	1	42,797		17	7	31	45	1	0.45	0.35
	11	MC	1	42,716		20	9	25	46	1	0.46	0.41
	12	MC	1	42,765		64	15	11	9	1	0.64	0.55
	13	MC	1	42,767		8	10	3	77	1	0.78	0.31
	14	MC	1	42,740		15	8	6	71	1	0.71	0.47
	15	MC	1	42,758		10	14	71	5	1	0.71	0.35
	16	MC	1	42,789		19	65	13	1	1	0.66	0.49
	17	MC	1	42,723		6	21	61	10	1	0.62	0.35
	18	MC	1	42,732		6	63	15	16	1	0.63	0.47
	19	MC	1	42,726		18	48	18	15	1	0.49	0.47
	20	MC	1	42,655		58	8	15	19	1	0.58	0.49
	21	MC	1	42,685		19	21	51	8	1	0.51	0.46
	22	MC	1	42,501		22	14	52	12	1	0.52	0.37
	23	MC	1	42,564		17	36	18	28	1	0.36	0.28
	24	MC	1	42,594		42	23	20	14	1	0.43	0.47
	25	MC	1	42,567		20	22	9	48	1	0.48	0.39
	26	MC	1	42,497		19	10	65	5	1	0.66	0.44
	27	MC	1	42,468		27	38	16	17	1	0.39	0.21
WR	1	MC	1	42,738		6	87	6	0	1	0.88	0.38
	2	MC	1	42,756		3	3	93	0	1	0.93	0.36
	3	MC	1	42,758		9	69	13	8	1	0.69	0.51
	4	MC	1	42,798		5	84	5	5	1	0.84	0.50
	5	MC	1	42,791		19	61	16	4	1	0.61	0.42
	6	MC	1	42,709		21	12	19	48	1	0.48	0.43
	7	MC	1	42,771		88	5	4	3	1	0.88	0.41
	8	MC	1	42,775		12	3	79	5	1	0.80	0.45
	9	MC	1	42,751		4	16	59	20	1	0.60	0.39
	10	MC	1	42,787		7	79	9	4	1	0.80	0.51
	11	MC	1	42,781		83	3	10	4	1	0.83	0.51
	12	MC	1	42,761		48	7	5	39	1	0.48	0.29
	13	MC	1	42,717		10	17	23	50	1	0.50	0.42
	14	MC	1	42,701		18	9	61	11	1	0.62	0.28
	15	MC	1	42,703		6	9	67	17	1	0.67	0.57
	16	CR	2	43,017	11	15	75			0	0.82	0.64
	17	CR	2	43,017	12	18	70			0	0.79	0.66
	18	CR	2	43,017	14	32	54			0	0.70	0.64
	19	ER	4	43,017	10	13	28	30	19	0	0.59	0.77

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=208,596) after cleaning (with valid corresponding subtest raw score).

APPENDIX C: RAW SCORE TO SCALE SCORE CONVERSION TABLES

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

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-7.93	2.01	322	70
1	-6.51	1.02	372	36
2	-5.78	0.74	398	26
3	-5.33	0.61	413	21
4	-5.00	0.54	425	19
5	-4.74	0.49	434	17
6	-4.51	0.46	442	16
7	-4.32	0.43	449	15
8	-4.14	0.41	455	14
9	-3.98	0.39	461	14
10	-3.84	0.38	466	13
11	-3.70	0.36	470	13
12	-3.57	0.35	475	12
13	-3.45	0.34	479	12
14	-3.34	0.34	483	12
15	-3.23	0.33	487	12
16	-3.12	0.32	491	11
17	-3.02	0.32	494	11
18	-2.92	0.31	498	11
19	-2.82	0.31	501	11
20	-2.73	0.31	505	11
21	-2.63	0.30	508	11
22	-2.54	0.30	511	11
23	-2.45	0.30	514	10
24	-2.36	0.30	517	10
25	-2.28	0.30	520	10
26	-2.19	0.29	523	10
27	-2.10	0.29	526	10
28	-2.02	0.29	529	10
29	-1.93	0.29	532	10
30	-1.85	0.29	535	10
31	-1.76	0.29	538	10
32	-1.67	0.30	541	10
33	-1.59	0.30	544	10
34	-1.50	0.30	548	10
35	-1.41	0.30	551	11
36	-1.32	0.30	554	11
37	-1.22	0.31	557	11
38	-1.13	0.31	560	11
39	-1.03	0.31	564	11
40	-0.93	0.32	567	11
41	-0.83	0.32	571	11
42	-0.72	0.33	575	11

C.1a: Kindergarten–Grade 1 Listening and Speaking (continued)

Raw Score	Measure	SE	Scale Score	SE(SS)
43	-0.61	0.34	578	12
44	-0.50	0.34	583	12
45	-0.38	0.35	587	12
46	-0.25	0.36	591	13
47	-0.12	0.37	596	13
48	0.02	0.38	601	13
49	0.18	0.40	606	14
50	0.34	0.42	612	15
51	0.53	0.44	618	15
52	0.73	0.47	626	16
53	0.97	0.50	634	18
54	1.25	0.55	644	19
55	1.59	0.63	656	22
56	2.06	0.75	672	26
57	2.83	1.04	699	36
58	4.27	2.02	749	71

Note: The statistics in Tables C.1–C.5 were based on the equating sample. Invalid cases were removed after cleaning procedures.

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

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-8.15	2.03	315	71
1	-6.67	1.06	367	37
2	-5.87	0.78	395	27
3	-5.36	0.65	412	23
4	-4.99	0.57	425	20
5	-4.69	0.52	436	18
6	-4.45	0.48	444	17
7	-4.23	0.45	452	16
8	-4.05	0.42	458	15
9	-3.88	0.40	464	14
10	-3.72	0.39	470	14
11	-3.57	0.38	475	13
12	-3.43	0.37	480	13
13	-3.29	0.37	485	13
14	-3.16	0.36	490	13
15	-3.02	0.36	494	13
16	-2.89	0.36	499	13
17	-2.76	0.36	503	13
18	-2.63	0.36	508	13
19	-2.50	0.36	512	13
20	-2.37	0.36	517	13
21	-2.24	0.36	522	13
22	-2.10	0.37	526	13
23	-1.97	0.37	531	13
24	-1.83	0.37	536	13
25	-1.69	0.38	541	13
26	-1.54	0.38	546	13
27	-1.39	0.39	551	14
28	-1.24	0.40	557	14
29	-1.07	0.41	563	14
30	-0.90	0.42	569	15
31	-0.71	0.44	575	15
32	-0.51	0.46	582	16
33	-0.28	0.49	590	17
34	-0.03	0.52	599	18
35	0.27	0.57	610	20
36	0.64	0.64	622	23
37	1.13	0.77	640	27
38	1.92	1.05	667	37
39	3.38	2.03	718	71

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

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-6.06	2.01	388	70
1	-4.65	1.02	437	36
2	-3.92	0.73	463	26
3	-3.47	0.61	478	21
4	-3.15	0.54	490	19
5	-2.89	0.49	499	17
6	-2.67	0.45	507	16
7	-2.47	0.42	513	15
8	-2.30	0.40	519	14
9	-2.15	0.39	525	13
10	-2.00	0.37	530	13
11	-1.87	0.36	535	13
12	-1.75	0.35	539	12
13	-1.63	0.34	543	12
14	-1.51	0.33	547	12
15	-1.41	0.33	551	11
16	-1.30	0.32	554	11
17	-1.20	0.31	558	11
18	-1.10	0.31	561	11
19	-1.01	0.31	565	11
20	-0.92	0.30	568	11
21	-0.82	0.30	571	11
22	-0.73	0.30	574	10
23	-0.65	0.30	577	10
24	-0.56	0.29	580	10
25	-0.47	0.29	583	10
26	-0.39	0.29	586	10
27	-0.30	0.29	589	10
28	-0.22	0.29	592	10
29	-0.13	0.29	595	10
30	-0.05	0.29	598	10
31	0.04	0.29	601	10
32	0.13	0.29	604	10
33	0.21	0.29	607	10
34	0.30	0.30	610	10
35	0.39	0.30	614	10
36	0.48	0.30	617	11
37	0.57	0.30	620	11
38	0.66	0.31	623	11
39	0.76	0.31	626	11
40	0.85	0.31	630	11
41	0.95	0.32	633	11
42	1.06	0.32	637	11
43	1.16	0.33	641	11
44	1.27	0.34	645	12
45	1.39	0.34	649	12

C.2a: Grade 2–Grade 4 Listening and Speaking (continued)

Raw Score	Measure	SE	Scale Score	SE(SS)
46	1.51	0.35	653	12
47	1.64	0.36	657	13
48	1.77	0.37	662	13
49	1.92	0.39	667	14
50	2.07	0.41	673	14
51	2.25	0.43	679	15
52	2.44	0.45	685	16
53	2.66	0.49	693	17
54	2.93	0.54	702	19
55	3.26	0.61	714	21
56	3.70	0.74	730	26
57	4.44	1.02	755	36
58	5.85	2.01	805	70

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

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-6.46	2.01	374	70
1	-5.03	1.03	424	36
2	-4.27	0.75	450	26
3	-3.81	0.63	467	22
4	-3.46	0.56	479	20
5	-3.17	0.51	489	18
6	-2.92	0.48	498	17
7	-2.71	0.46	505	16
8	-2.51	0.44	512	15
9	-2.32	0.42	519	15
10	-2.15	0.41	525	14
11	-1.99	0.40	530	14
12	-1.84	0.39	536	14
13	-1.69	0.38	541	13
14	-1.55	0.37	546	13
15	-1.41	0.37	551	13
16	-1.27	0.36	555	13
17	-1.14	0.36	560	13
18	-1.02	0.36	564	12
19	-0.89	0.35	569	12
20	-0.77	0.35	573	12
21	-0.64	0.35	578	12
22	-0.52	0.35	582	12
23	-0.40	0.35	586	12
24	-0.28	0.35	590	12
25	-0.16	0.35	594	12
26	-0.04	0.35	599	12
27	0.08	0.35	603	12
28	0.21	0.35	607	12
29	0.33	0.36	612	12
30	0.46	0.36	616	13
31	0.59	0.36	621	13
32	0.73	0.37	625	13
33	0.87	0.38	630	13
34	1.01	0.39	635	13
35	1.16	0.40	641	14
36	1.32	0.41	646	14
37	1.50	0.42	652	15
38	1.68	0.44	659	15
39	1.89	0.46	666	16
40	2.12	0.49	674	17
41	2.38	0.53	683	19
42	2.68	0.58	694	20
43	3.07	0.66	707	23
44	3.58	0.78	725	27
45	4.39	1.06	754	37
46	5.89	2.04	806	71

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

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-5.66	2.01	402	70
1	-4.24	1.02	451	36
2	-3.51	0.73	477	26
3	-3.07	0.61	493	21
4	-2.74	0.54	504	19
5	-2.48	0.49	513	17
6	-2.26	0.45	521	16
7	-2.07	0.42	527	15
8	-1.91	0.40	533	14
9	-1.75	0.38	539	13
10	-1.61	0.37	544	13
11	-1.48	0.35	548	12
12	-1.36	0.34	552	12
13	-1.25	0.33	556	12
14	-1.14	0.33	560	11
15	-1.04	0.32	564	11
16	-0.94	0.31	567	11
17	-0.84	0.31	571	11
18	-0.75	0.30	574	11
19	-0.66	0.30	577	10
20	-0.57	0.29	580	10
21	-0.49	0.29	583	10
22	-0.40	0.29	586	10
23	-0.32	0.29	589	10
24	-0.24	0.28	592	10
25	-0.16	0.28	594	10
26	-0.08	0.28	597	10
27	0.00	0.28	600	10
28	0.08	0.28	603	10
29	0.15	0.28	605	10
30	0.23	0.28	608	10
31	0.31	0.28	611	10
32	0.39	0.28	614	10
33	0.47	0.28	616	10
34	0.55	0.28	619	10
35	0.63	0.29	622	10
36	0.71	0.29	625	10
37	0.80	0.29	628	10
38	0.88	0.29	631	10
39	0.97	0.30	634	10
40	1.06	0.30	637	11
41	1.15	0.31	640	11
42	1.25	0.31	644	11

C.3a: Grade 5–Grade 6 Listening and Speaking (continued)

Raw Score	Measure	SE	Scale Score	SE(SS)
43	1.34	0.32	647	11
44	1.45	0.32	651	11
45	1.55	0.33	654	12
46	1.67	0.34	658	12
47	1.78	0.35	662	12
48	1.91	0.36	667	13
49	2.04	0.37	671	13
50	2.19	0.39	677	14
51	2.34	0.40	682	14
52	2.52	0.43	688	15
53	2.71	0.45	695	16
54	2.93	0.49	703	17
55	3.19	0.54	712	19
56	3.52	0.61	723	21
57	3.97	0.74	739	26
58	4.71	1.02	765	36
59	6.12	2.01	814	70

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

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-4.98	2.01	426	70
1	-3.56	1.02	475	36
2	-2.82	0.74	501	26
3	-2.37	0.61	517	21
4	-2.04	0.54	529	19
5	-1.77	0.49	538	17
6	-1.55	0.46	546	16
7	-1.35	0.43	553	15
8	-1.17	0.41	559	14
9	-1.01	0.39	565	14
10	-0.86	0.38	570	13
11	-0.72	0.37	575	13
12	-0.59	0.36	579	12
13	-0.47	0.35	584	12
14	-0.35	0.34	588	12
15	-0.24	0.33	592	12
16	-0.13	0.33	596	12
17	-0.02	0.32	599	11
18	0.08	0.32	603	11
19	0.18	0.32	606	11
20	0.28	0.31	610	11
21	0.38	0.31	613	11
22	0.48	0.31	617	11
23	0.58	0.31	620	11
24	0.67	0.31	624	11
25	0.77	0.31	627	11
26	0.86	0.31	630	11
27	0.96	0.31	634	11
28	1.05	0.31	637	11
29	1.15	0.31	640	11
30	1.25	0.31	644	11
31	1.35	0.32	647	11
32	1.45	0.32	651	11
33	1.55	0.32	654	11
34	1.66	0.33	658	11
35	1.76	0.33	662	12
36	1.87	0.34	666	12
37	1.99	0.34	670	12
38	2.11	0.35	674	12
39	2.23	0.36	678	12
40	2.36	0.36	683	13
41	2.50	0.38	687	13
42	2.64	0.39	692	14
43	2.80	0.40	698	14
44	2.97	0.42	704	15
45	3.16	0.44	710	16
46	3.36	0.47	718	16

C.3b: Grade 5–Grade 6 Reading and Writing (continued)

Raw Score	Measure	SE	Scale Score	SE(SS)
47	3.60	0.51	726	18
48	3.89	0.56	736	20
49	4.24	0.63	748	22
50	4.71	0.76	765	26
51	5.49	1.04	792	36
52	6.94	2.02	843	71

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

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-4.91	2.01	428	70
1	-3.50	1.02	478	36
2	-2.77	0.73	503	26
3	-2.33	0.61	518	21
4	-2.01	0.53	530	19
5	-1.75	0.49	539	17
6	-1.53	0.45	546	16
7	-1.34	0.42	553	15
8	-1.17	0.40	559	14
9	-1.02	0.38	564	13
10	-0.87	0.37	569	13
11	-0.74	0.36	574	12
12	-0.62	0.35	578	12
13	-0.50	0.34	582	12
14	-0.39	0.33	586	12
15	-0.29	0.32	590	11
16	-0.18	0.32	594	11
17	-0.09	0.31	597	11
18	0.01	0.31	600	11
19	0.10	0.30	604	11
20	0.19	0.30	607	10
21	0.28	0.30	610	10
22	0.37	0.29	613	10
23	0.45	0.29	616	10
24	0.54	0.29	619	10
25	0.62	0.29	622	10
26	0.70	0.29	625	10
27	0.78	0.29	627	10
28	0.86	0.29	630	10
29	0.95	0.28	633	10
30	1.03	0.29	636	10
31	1.11	0.29	639	10
32	1.19	0.29	642	10
33	1.27	0.29	645	10
34	1.36	0.29	647	10
35	1.44	0.29	650	10
36	1.53	0.29	653	10
37	1.61	0.30	656	10
38	1.70	0.30	660	10
39	1.79	0.30	663	11
40	1.89	0.31	666	11
41	1.98	0.31	669	11
42	2.08	0.32	673	11
43	2.18	0.32	676	11
44	2.29	0.33	680	12
45	2.40	0.34	684	12

C.4a: Grade 7–Grade 8 Listening and Speaking (continued)

Raw Score	Measure	SE	Scale Score	SE(SS)
46	2.54	0.34	689	12
47	2.65	0.35	693	12
48	2.78	0.36	697	13
49	2.91	0.37	702	13
50	3.06	0.39	707	14
51	3.21	0.40	712	14
52	3.39	0.43	718	15
53	3.58	0.45	725	16
54	3.80	0.49	733	17
55	4.06	0.54	742	19
56	4.39	0.61	754	21
57	4.84	0.74	769	26
58	5.57	1.02	795	36
59	6.99	2.01	845	70

C.4b: Grade 7-Grade 8 Reading and Writing

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-4.81	2.01	432	70
1	-3.39	1.02	481	36
2	-2.66	0.74	507	26
3	-2.21	0.61	523	21
4	-1.88	0.54	534	19
5	-1.62	0.49	543	17
6	-1.40	0.46	551	16
7	-1.20	0.43	558	15
8	-1.02	0.41	564	14
9	-0.87	0.39	570	14
10	-0.72	0.38	575	13
11	-0.58	0.37	580	13
12	-0.45	0.36	584	12
13	-0.33	0.35	589	12
14	-0.21	0.34	593	12
15	-0.10	0.33	597	12
16	0.01	0.33	600	11
17	0.12	0.32	604	11
18	0.22	0.32	608	11
19	0.32	0.32	611	11
20	0.42	0.31	615	11
21	0.52	0.31	618	11
22	0.61	0.31	621	11
23	0.71	0.31	625	11
24	0.80	0.30	628	11
25	0.89	0.30	631	11
26	0.98	0.30	634	11
27	1.08	0.30	638	11
28	1.17	0.30	641	11
29	1.26	0.31	644	11
30	1.36	0.31	647	11
31	1.45	0.31	651	11
32	1.55	0.31	654	11
33	1.64	0.31	658	11
34	1.74	0.32	661	11
35	1.84	0.32	665	11
36	1.95	0.33	668	11
37	2.06	0.33	672	12
38	2.17	0.34	676	12
39	2.28	0.34	680	12
40	2.41	0.35	684	12
41	2.53	0.36	689	13
42	2.67	0.38	693	13
43	2.82	0.39	699	14
44	2.98	0.41	704	14
45	3.15	0.43	710	15
46	3.35	0.46	717	16
47	3.57	0.49	725	17
48	3.84	0.54	734	19
49	4.17	0.62	746	22
50	4.63	0.74	762	26
51	5.37	1.02	788	36
52	6.79	2.01	838	70

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

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-4.70	2.01	436	70
1	-3.28	1.02	485	36
2	-2.55	0.73	511	26
3	-2.10	0.61	526	21
4	-1.77	0.54	538	19
5	-1.51	0.49	547	17
6	-1.29	0.45	555	16
7	-1.10	0.43	562	15
8	-0.92	0.41	568	14
9	-0.77	0.39	573	14
10	-0.62	0.37	578	13
11	-0.49	0.36	583	13
12	-0.36	0.35	587	12
13	-0.24	0.34	592	12
14	-0.13	0.33	596	12
15	-0.02	0.33	599	11
16	0.09	0.32	603	11
17	0.19	0.32	607	11
18	0.29	0.31	610	11
19	0.38	0.31	613	11
20	0.48	0.30	617	11
21	0.57	0.30	620	11
22	0.66	0.30	623	10
23	0.74	0.30	626	10
24	0.83	0.29	629	10
25	0.92	0.29	632	10
26	1.00	0.29	635	10
27	1.09	0.29	638	10
28	1.17	0.29	641	10
29	1.25	0.29	644	10
30	1.34	0.29	647	10
31	1.42	0.29	650	10
32	1.50	0.29	653	10
33	1.59	0.29	656	10
34	1.67	0.29	659	10
35	1.76	0.29	661	10
36	1.84	0.29	664	10
37	1.93	0.30	668	10
38	2.02	0.30	671	10
39	2.11	0.30	674	11
40	2.20	0.30	677	11
41	2.29	0.31	680	11
42	2.39	0.31	684	11
43	2.49	0.32	687	11
44	2.59	0.32	691	11
45	2.69	0.33	694	11
46	2.80	0.34	698	12

C.5a: Grade 9–Grade 12 Listening and Speaking (continued)

Raw Score	Measure	SE	Scale Score	SE(SS)
47	2.92	0.34	702	12
48	3.04	0.35	706	12
49	3.17	0.37	711	13
50	3.31	0.38	716	13
51	3.46	0.40	721	14
52	3.63	0.42	727	15
53	3.82	0.45	734	16
54	4.03	0.48	741	17
55	4.29	0.53	750	19
56	4.61	0.60	761	21
57	5.04	0.73	777	25
58	5.77	1.02	802	36
59	7.18	2.01	851	70

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

Raw Score	Measure	SE	Scale Score	SE(SS)
0	-4.14	2.01	455	70
1	-2.73	1.02	505	36
2	-1.99	0.74	530	26
3	-1.54	0.61	546	21
4	-1.21	0.54	558	19
5	-0.95	0.49	567	17
6	-0.73	0.46	575	16
7	-0.53	0.43	581	15
8	-0.36	0.41	588	14
9	-0.20	0.39	593	14
10	-0.05	0.38	598	13
11	0.09	0.36	603	13
12	0.22	0.35	608	12
13	0.34	0.35	612	12
14	0.46	0.34	616	12
15	0.57	0.33	620	12
16	0.68	0.33	624	11
17	0.78	0.32	627	11
18	0.89	0.32	631	11
19	0.99	0.32	635	11
20	1.09	0.31	638	11
21	1.18	0.31	641	11
22	1.28	0.31	645	11
23	1.38	0.31	648	11
24	1.47	0.31	652	11
25	1.57	0.31	655	11
26	1.66	0.31	658	11
27	1.76	0.31	662	11
28	1.85	0.31	665	11
29	1.95	0.31	668	11
30	2.04	0.31	672	11
31	2.14	0.31	675	11
32	2.24	0.32	678	11
33	2.34	0.32	682	11
34	2.44	0.32	685	11
35	2.55	0.32	689	11
36	2.65	0.33	693	11
37	2.76	0.33	697	12
38	2.88	0.34	701	12
39	2.99	0.35	705	12
40	3.12	0.35	709	12
41	3.25	0.36	714	13
42	3.38	0.38	718	13
43	3.53	0.39	723	14
44	3.69	0.41	729	14
45	3.86	0.43	735	15
46	4.05	0.45	742	16
47	4.27	0.49	750	17
48	4.53	0.54	759	19
49	4.86	0.61	770	21
50	5.30	0.73	785	26
51	6.03	1.02	811	36
52	7.44	2.01	860	70

APPENDIX D: IRT STATISTICS

D.1a: Kindergarten–Grade 1 Listening and Speaking

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
LI	1	42,442	-3.13	0.02	1.06	1.27	--
	2	42,442	-2.79	0.02	1.09	1.21	--
	3	42,442	-3.58	0.03	1.00	1.13	--
	4	42,442	-3.14	0.02	1.06	1.25	--
	5	42,442	-3.46	0.02	0.91	1.19	--
	6	42,442	-4.47	0.04	0.94	0.94	--
	7	42,442	-3.46	0.02	1.04	1.33	--
	8	42,442	-1.83	0.01	1.09	1.09	--
	9	42,442	-2.54	0.02	1.03	0.95	--
	10	42,442	-3.76	0.03	1.03	1.08	--
	11	42,442	-1.09	0.01	1.28	1.68	--
	12	42,442	-2.25	0.02	1.18	1.79	--
	13	42,442	-2.43	0.02	0.95	0.80	--
	14	42,442	-1.60	0.01	1.14	1.09	--
	15	42,442	-3.81	0.03	0.97	0.89	--
	16	42,442	-3.31	0.02	0.97	0.79	--
	17	42,442	-1.98	0.02	0.97	1.00	--
	18	42,442	-2.41	0.02	1.16	1.44	--
	19	42,442	-1.98	0.02	1.00	1.11	--
	20	42,442	-1.73	0.01	1.09	1.08	--
	21	42,442	-0.54	0.01	1.07	1.13	--
	22	42,442	-0.06	0.01	1.30	1.55	--
	23	42,442	-1.18	0.01	1.29	1.62	--
	24	42,442	-0.26	0.01	1.26	1.49	--
SP	1	42,442	-1.38	0.01	0.91	0.85	--
	2	42,442	-1.87	0.01	0.78	0.69	--
	3	42,442	-1.17	0.01	1.02	1.09	--
	4	42,442	-1.51	0.01	0.99	1.02	--
	5	42,442	-1.76	0.01	0.84	0.79	--
	6	42,442	-0.63	0.01	0.76	0.77	--
	7	42,442	-1.76	0.01	0.80	0.76	--
	8	42,442	-1.62	0.01	0.83	0.80	--
	9	42,442	-1.96	0.01	0.80	0.72	--
	10	42,442	-1.74	0.01	0.81	0.76	--
	11	42,442	-1.55	0.01	0.80	0.76	--
	12	42,442	-1.82	0.01	0.87	0.80	--
	13	42,442	-1.99	0.01	0.89	0.81	--
	14	42,442	-2.33	0.01	0.84	0.73	--
	15	42,442	-1.84	0.01	0.84	0.75	--
	16	42,442	-1.42	0.01	0.89	0.78	--

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

D.1b: Kindergarten–Grade 1 Reading and Writing

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
RD	1	40,059	-3.36	0.02	0.93	1.03	--
	2	40,059	-2.96	0.02	1.19	1.61	--
	3	40,059	-3.05	0.02	0.84	0.66	--
	4	40,059	-1.86	0.01	0.99	0.93	--
	5	40,059	-2.21	0.01	1.04	1.29	--
	6	40,059	-2.50	0.01	0.80	0.57	--
	7	40,059	-2.97	0.02	0.88	0.70	--
	8	40,059	-1.45	0.01	1.31	1.50	--
	9	40,059	-2.05	0.01	0.83	0.67	--
	10	40,059	-1.78	0.01	0.96	0.84	--
	11	40,059	-1.18	0.01	0.94	0.91	--
	12	40,059	-1.99	0.01	1.14	1.04	--
	13	40,059	-1.52	0.01	1.12	1.19	--
	14	40,059	-1.17	0.01	1.43	1.68	--
	15	40,059	-2.55	0.01	1.16	1.33	--
WR	1	40,059	-4.50	0.02	0.86	0.56	--
	2	40,059	-3.59	0.02	1.03	0.97	--
	3	40,059	-2.80	0.01	1.05	1.04	--
	4	40,059	-3.07	0.02	1.14	2.19	--
	5	40,059	-3.40	0.02	1.03	1.37	--
	6	40,059	-4.06	0.02	0.97	1.07	--
	7	40,059	-3.48	0.01	1.29	1.43	--
	8	40,059	-3.01	0.01	1.33	1.44	--
	9	40,059	-3.62	0.01	0.86	1.36	--
	10	40,059	-3.38	0.01	1.07	2.69	--
	11	40,059	-2.24	0.01	0.80	0.80	--
	12	40,059	-1.93	0.01	0.74	0.66	--
	13	40,059	-0.97	0.01	0.65	0.60	--
	14	40,059	-0.90	0.01	0.62	0.59	--
	15	40,059	-0.17	0.01	0.67	0.64	--

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

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
LI	1	52,778	-2.01	0.03	0.95	0.63	--
	2	52,778	0.14	0.01	1.40	1.85	--
	3	52,778	-0.86	0.02	0.97	0.72	--
	4	52,778	-0.69	0.02	1.16	1.36	--
	5	52,778	0.96	0.01	1.07	1.10	--
	6	52,778	-1.50	0.02	0.96	0.70	--
	7	52,778	-1.81	0.02	0.95	0.81	--
	8	52,778	0.77	0.01	1.18	1.29	--
	9	52,778	1.22	0.01	1.17	1.24	--
	10	52,778	1.85	0.01	1.16	1.36	--
	11	52,778	1.14	0.01	1.25	1.45	--
	12	52,778	0.34	0.01	1.19	1.41	--
	13	52,778	-0.57	0.02	0.98	0.90	--
	14	52,778	1.52	0.01	1.06	1.13	--
	15	52,778	0.70	0.01	1.17	1.26	--
	16	52,778	0.83	0.01	1.08	1.14	--
	17	52,778	0.38	0.01	1.07	1.13	--
	18	52,778	1.15	0.01	1.11	1.17	--
	19	52,778	1.42	0.01	1.15	1.28	--
	20	52,778	-0.26	0.01	1.10	1.34	--
	21	52,778	0.07	0.01	1.24	1.41	--
	22	52,778	-1.34	0.02	0.94	0.91	--
	23	52,778	-0.81	0.02	1.06	1.20	--
	24	52,778	-0.42	0.02	1.08	1.30	--
SP	1	52,778	-0.63	0.01	0.87	0.81	--
	2	52,778	-0.15	0.01	0.79	0.71	--
	3	52,778	-0.52	0.01	0.78	0.70	--
	4	52,778	-0.13	0.01	0.83	0.77	--
	5	52,778	-0.18	0.01	0.93	0.90	--
	6	52,778	0.34	0.01	0.85	0.86	--
	7	52,778	-0.89	0.01	0.79	0.70	--
	8	52,778	0.03	0.01	0.71	0.62	--
	9	52,778	-0.29	0.01	0.79	0.72	--
	10	52,778	-0.17	0.01	0.79	0.72	--
	11	52,778	-0.05	0.01	0.85	0.81	--
	12	52,778	-0.92	0.01	0.93	0.82	--
	13	52,778	-0.40	0.01	0.87	0.74	--
	14	52,778	-0.20	0.01	0.90	0.84	--
	15	52,778	-0.15	0.01	0.83	0.73	--
	16	52,778	-0.64	0.01	0.90	0.82	--

D.2b: Grade 2–Grade 4 Reading and Writing

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
RD	1	52,121	-2.65	0.02	0.92	0.55	--
	2	52,121	-1.61	0.02	0.83	0.67	--
	3	52,121	-1.72	0.02	0.88	0.59	--
	4	52,121	-2.74	0.02	0.94	0.95	--
	5	52,121	-2.15	0.02	0.90	0.58	--
	6	52,121	-2.64	0.02	0.88	0.50	--
	7	52,121	-1.97	0.02	0.98	0.73	--
	8	52,121	-2.10	0.02	0.85	0.59	--
	9	52,121	0.10	0.01	1.01	0.99	--
	10	52,121	-0.82	0.01	0.86	0.73	--
	11	52,121	0.77	0.01	1.05	1.08	--
	12	52,121	-1.55	0.02	1.09	1.07	--
	13	52,121	0.39	0.01	1.10	1.14	--
	14	52,121	0.33	0.01	0.90	0.85	--
	15	52,121	1.94	0.01	1.25	1.57	--
	16	52,121	0.81	0.01	1.11	1.17	--
	17	52,121	0.42	0.01	1.12	1.20	--
	18	52,121	0.89	0.01	0.99	1.01	--
	19	52,121	0.02	0.01	1.07	1.01	--
	20	52,121	-0.34	0.01	0.90	0.78	--
	21	52,121	1.26	0.01	1.03	1.11	--
	22	52,121	0.47	0.01	0.93	0.88	--
	23	52,121	-0.76	0.01	1.21	1.59	--
	24	52,121	1.46	0.01	1.21	1.44	--
WR	1	52,121	-2.12	0.02	0.88	0.58	--
	2	52,121	-2.61	0.02	0.91	0.53	--
	3	52,121	-0.27	0.01	1.02	1.00	--
	4	52,121	-1.02	0.01	0.93	0.82	--
	5	52,121	0.61	0.01	1.11	1.17	--
	6	52,121	-0.29	0.01	0.91	0.83	--
	7	52,121	0.34	0.01	1.01	1.00	--
	8	52,121	0.31	0.01	0.93	0.90	--
	9	52,121	0.20	0.01	1.01	1.01	--
	10	52,121	-1.17	0.01	1.08	1.01	--
	11	52,121	-0.09	0.01	0.99	0.98	--
	12	52,121	0.33	0.01	1.02	1.02	--
	13	52,121	-2.03	0.01	0.98	0.95	--
	14	52,121	-0.42	0.01	1.05	1.04	--
	15	52,121	-0.04	0.01	0.95	0.94	--
	16	52,121	1.19	0.01	0.85	0.84	--

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

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
LI	1	24,463	-1.43	0.03	0.89	0.54	--
	2	24,463	1.04	0.02	0.99	0.98	--
	3	24,463	0.97	0.02	1.26	1.35	--
	4	24,463	-0.79	0.03	0.95	0.83	--
	5	24,463	-0.46	0.02	0.96	0.81	--
	6	24,463	0.35	0.02	1.28	1.36	--
	7	24,463	0.82	0.02	1.26	1.31	--
	8	24,463	-0.48	0.02	1.01	1.04	--
	9	24,463	-0.23	0.02	1.00	0.91	--
	10	24,463	0.30	0.02	1.05	1.01	--
	11	24,463	0.86	0.02	1.21	1.23	--
	12	24,463	0.70	0.02	1.08	1.03	--
	13	24,463	1.19	0.02	1.13	1.18	--
	14	24,463	-1.31	0.03	1.21	1.42	--
	15	24,463	1.52	0.02	1.08	1.15	--
	16	24,463	-0.09	0.02	0.99	0.91	--
	17	24,463	2.00	0.01	1.21	1.38	--
	18	24,463	1.91	0.01	1.33	1.57	--
	19	24,463	0.05	0.02	1.24	1.35	--
	20	24,463	0.02	0.02	1.05	1.20	--
	21	24,463	1.32	0.02	1.26	1.40	--
	22	24,463	0.37	0.02	1.35	1.65	--
	23	24,463	2.47	0.01	1.14	1.41	--
	24	24,463	0.43	0.02	1.20	1.28	--
	25	24,463	1.99	0.01	1.14	1.27	--
SP	1	24,463	-0.35	0.02	0.82	0.74	--
	2	24,463	-0.68	0.02	0.80	0.63	--
	3	24,463	0.15	0.01	0.78	0.73	--
	4	24,463	0.01	0.01	0.79	0.64	--
	5	24,463	0.38	0.01	0.82	0.78	--
	6	24,463	0.69	0.01	0.77	0.78	--
	7	24,463	-0.43	0.02	0.78	0.71	--
	8	24,463	0.05	0.01	0.71	0.61	--
	9	24,463	0.22	0.01	0.71	0.63	--
	10	24,463	-0.22	0.02	0.78	0.71	--
	11	24,463	-0.89	0.02	0.75	0.61	--
	12	24,463	-0.04	0.02	0.78	0.62	--
	13	24,463	0.25	0.01	0.95	0.88	--
	14	24,463	-0.88	0.02	0.89	0.67	--
	15	24,463	0.00	0.02	0.71	0.53	--
	16	24,463	0.45	0.01	0.74	0.65	--

D.3b: Grade 5–Grade 6 Reading and Writing

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
RD	1	23,627	-1.55	0.04	0.94	0.85	--
	2	23,627	-0.21	0.02	0.99	1.08	--
	3	23,627	0.26	0.02	1.07	1.04	--
	4	23,627	-1.33	0.03	0.98	1.17	--
	5	23,627	1.64	0.02	1.05	1.06	--
	6	23,627	0.53	0.02	1.12	1.29	--
	7	23,627	0.16	0.02	0.90	0.79	--
	8	23,627	0.28	0.02	0.83	0.68	--
	9	23,627	2.85	0.01	1.13	1.36	--
	10	23,627	0.30	0.02	0.82	0.61	--
	11	23,627	1.29	0.02	1.00	0.97	--
	12	23,627	0.78	0.02	1.02	1.02	--
	13	23,627	2.47	0.01	1.06	1.16	--
	14	23,627	0.76	0.02	0.76	0.62	--
	15	23,627	1.63	0.02	1.15	1.18	--
	16	23,627	1.16	0.02	1.04	1.10	--
	17	23,627	0.98	0.02	0.98	0.93	--
	18	23,627	0.57	0.02	0.80	0.64	--
	19	23,627	0.56	0.02	0.93	0.77	--
	20	23,627	1.33	0.02	0.89	0.81	--
	21	23,627	2.09	0.01	1.03	1.07	--
	22	23,627	1.66	0.02	0.94	0.89	--
	23	23,627	0.96	0.02	0.86	0.74	--
	24	23,627	1.66	0.02	1.28	1.40	--
	25	23,627	0.69	0.02	0.76	0.59	--
	26	23,627	1.25	0.02	0.93	0.89	--
	27	23,627	1.34	0.02	1.07	1.10	--
WR	1	23,627	0.56	0.02	1.22	1.31	--
	2	23,627	-0.44	0.02	0.93	0.72	--
	3	23,627	-0.05	0.02	1.02	1.30	--
	4	23,627	0.27	0.02	0.92	0.82	--
	5	23,627	1.73	0.02	1.11	1.15	--
	6	23,627	0.11	0.02	0.83	0.69	--
	7	23,627	2.24	0.01	1.08	1.15	--
	8	23,627	-0.72	0.03	0.86	0.73	--
	9	23,627	0.57	0.02	0.86	0.80	--
	10	23,627	0.33	0.02	0.94	0.99	--
	11	23,627	2.57	0.01	1.12	1.23	--
	12	23,627	-0.52	0.02	0.88	0.71	--
	13	23,627	1.36	0.02	1.11	1.17	--
	14	23,627	2.03	0.01	1.15	1.22	--
	15	23,627	2.14	0.01	1.14	1.22	--
	16	23,627	-0.57	0.02	0.89	0.78	--
	17	23,627	0.31	0.01	0.86	0.78	--
	18	23,627	1.27	0.01	0.99	0.99	--
	19	23,627	2.00	0.01	0.84	0.84	--

D.4a: Grade 7–Grade 8 Listening and Speaking

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
LI	1	21,224	-1.36	0.04	1.04	1.05	--
	2	21,224	-0.44	0.03	1.09	1.05	--
	3	21,224	0.35	0.02	1.13	1.12	--
	4	21,224	0.78	0.02	0.95	0.94	--
	5	21,224	2.42	0.02	1.25	1.39	--
	6	21,224	2.59	0.02	1.25	1.41	--
	7	21,224	2.83	0.02	1.17	1.35	--
	8	21,224	1.46	0.02	1.36	1.45	--
	9	21,224	0.52	0.02	1.10	1.06	--
	10	21,224	2.17	0.02	1.30	1.45	--
	11	21,224	2.33	0.02	1.32	1.49	--
	12	21,224	2.42	0.02	1.13	1.25	--
	13	21,224	3.42	0.02	1.05	1.33	--
	14	21,224	2.69	0.02	1.21	1.44	--
	15	21,224	1.23	0.02	1.11	1.13	--
	16	21,224	1.20	0.02	1.14	1.17	--
	17	21,224	0.62	0.02	1.26	1.39	--
	18	21,224	1.90	0.02	1.15	1.20	--
	19	21,224	1.79	0.02	1.08	1.10	--
	20	21,224	1.38	0.02	1.10	1.09	--
	21	21,224	0.90	0.02	1.14	1.16	--
	22	21,224	-0.22	0.03	1.06	1.04	--
	23	21,224	1.85	0.02	1.17	1.23	--
	24	21,224	3.32	0.02	1.15	1.50	--
	25	21,224	1.22	0.02	1.17	1.13	--
SP	1	21,224	0.42	0.02	0.77	0.66	--
	2	21,224	1.03	0.01	0.76	0.72	--
	3	21,224	0.86	0.01	0.74	0.68	--
	4	21,224	0.45	0.02	0.77	0.73	--
	5	21,224	1.14	0.01	0.71	0.65	--
	6	21,224	1.08	0.01	0.70	0.71	--
	7	21,224	0.35	0.02	0.73	0.67	--
	8	21,224	0.38	0.02	0.69	0.63	--
	9	21,224	0.41	0.02	0.72	0.62	--
	10	21,224	0.43	0.02	0.73	0.65	--
	11	21,224	-0.10	0.02	0.71	0.58	--
	12	21,224	0.85	0.01	0.77	0.75	--
	13	21,224	1.01	0.01	0.75	0.68	--
	14	21,224	0.82	0.01	0.66	0.55	--
	15	21,224	1.30	0.01	0.80	0.70	--
	16	21,224	0.94	0.01	0.73	0.61	--

D.4b: Grade 7–Grade 8 Reading and Writing

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
RD	1	20,224	-0.19	0.02	1.00	1.19	--
	2	20,224	-1.02	0.03	1.06	0.82	--
	3	20,224	-1.20	0.03	0.97	0.97	--
	4	20,224	1.53	0.02	0.93	0.91	--
	5	20,224	1.04	0.02	0.88	0.81	--
	6	20,224	-0.38	0.02	0.91	0.74	--
	7	20,224	1.65	0.02	0.95	0.95	--
	8	20,224	1.02	0.02	1.06	1.06	--
	9	20,224	1.04	0.02	0.88	0.79	--
	10	20,224	1.13	0.02	0.96	0.90	--
	11	20,224	1.19	0.02	1.01	1.00	--
	12	20,224	0.17	0.02	1.00	0.90	--
	13	20,224	0.45	0.02	0.89	0.77	--
	14	20,224	2.17	0.02	1.15	1.26	--
	15	20,224	1.12	0.02	0.95	0.93	--
	16	20,224	1.51	0.02	1.04	1.04	--
	17	20,224	1.40	0.02	0.99	0.95	--
	18	20,224	1.67	0.02	0.96	0.93	--
	19	20,224	1.78	0.02	1.13	1.16	--
	20	20,224	1.86	0.02	0.97	0.95	--
	21	20,224	2.37	0.02	1.04	1.12	--
	22	20,224	1.64	0.02	0.89	0.86	--
	23	20,224	2.50	0.02	1.08	1.20	--
	24	20,224	0.92	0.02	0.91	0.82	--
	25	20,224	1.83	0.02	1.04	1.09	--
	26	20,224	1.06	0.02	0.93	0.84	--
	27	20,224	1.86	0.02	0.99	0.99	--
WR	1	20,224	-0.51	0.02	1.04	1.21	--
	2	20,224	-0.82	0.03	1.11	0.99	--
	3	20,224	2.40	0.02	1.20	1.28	--
	4	20,224	1.41	0.02	1.04	1.05	--
	5	20,224	0.07	0.02	0.90	0.79	--
	6	20,224	0.00	0.02	1.02	1.14	--
	7	20,224	1.09	0.02	0.98	0.96	--
	8	20,224	0.14	0.02	0.95	0.91	--
	9	20,224	1.82	0.02	1.21	1.29	--
	10	20,224	1.03	0.02	1.01	1.07	--
	11	20,224	1.34	0.02	1.00	1.00	--
	12	20,224	-0.13	0.02	0.91	0.77	--
	13	20,224	1.18	0.02	1.01	0.98	--
	14	20,224	0.86	0.02	1.19	1.38	--
	15	20,224	1.49	0.02	1.09	1.15	--
	16	20,224	0.97	0.01	1.09	1.10	--
	17	20,224	-0.45	0.02	1.01	1.08	--
	18	20,224	0.76	0.01	1.01	1.04	--
	19	20,224	1.74	0.01	0.90	0.90	--

D.5a: Grade 9–Grade 12 Listening and Speaking

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
LI	1	35,590	0.30	0.02	1.11	1.00	--
	2	35,590	0.97	0.02	1.03	0.90	--
	3	35,590	0.23	0.02	1.12	0.95	--
	4	35,590	1.27	0.02	1.09	1.02	--
	5	35,590	2.06	0.01	1.18	1.28	--
	6	35,590	0.76	0.02	1.25	1.64	--
	7	35,590	0.93	0.02	1.11	1.27	--
	8	35,590	1.66	0.01	1.16	1.23	--
	9	35,590	1.15	0.02	0.99	0.90	--
	10	35,590	2.76	0.01	1.21	1.35	--
	11	35,590	2.93	0.01	1.09	1.24	--
	12	35,590	1.63	0.01	1.09	1.09	--
	13	35,590	2.51	0.01	1.24	1.42	--
	14	35,590	2.50	0.01	1.20	1.33	--
	15	35,590	2.90	0.01	1.37	1.71	--
	16	35,590	2.50	0.01	1.22	1.34	--
	17	35,590	2.34	0.01	1.22	1.37	--
	18	35,590	2.64	0.01	1.06	1.13	--
	19	35,590	1.51	0.01	0.96	0.85	--
	20	35,590	2.16	0.01	1.01	1.01	--
	21	35,590	2.67	0.01	1.18	1.33	--
	22	35,590	0.84	0.02	1.02	0.99	--
	23	35,590	0.25	0.02	1.05	1.10	--
	24	35,590	2.36	0.01	1.21	1.32	--
	25	35,590	2.07	0.01	1.46	1.86	--
SP	1	35,590	0.79	0.01	0.84	0.79	--
	2	35,590	0.65	0.01	0.77	0.70	--
	3	35,590	0.69	0.01	0.81	0.70	--
	4	35,590	0.95	0.01	0.88	0.78	--
	5	35,590	0.95	0.01	0.79	0.75	--
	6	35,590	1.35	0.01	0.69	0.70	--
	7	35,590	0.53	0.01	0.74	0.65	--
	8	35,590	0.90	0.01	0.75	0.69	--
	9	35,590	0.84	0.01	0.77	0.68	--
	10	35,590	0.42	0.01	0.75	0.65	--
	11	35,590	1.09	0.01	0.74	0.69	--
	12	35,590	1.04	0.01	0.73	0.65	--
	13	35,590	1.21	0.01	0.79	0.71	--
	14	35,590	0.86	0.01	0.92	0.83	--
	15	35,590	1.53	0.01	0.79	0.73	--
	16	35,590	0.63	0.01	0.82	0.72	--

D.5b: Grade 9–Grade 12 Reading and Writing

	Item	N-count	Rasch Difficulty	SE	MNSQ Infit	MNSQ Outfit	Misfit
RD	1	33,835	0.26	0.02	0.94	0.76	--
	2	33,835	0.21	0.02	0.84	0.67	--
	3	33,835	0.52	0.02	1.05	1.08	--
	4	33,835	-0.45	0.02	0.95	0.71	--
	5	33,835	2.00	0.01	1.04	1.01	--
	6	33,835	2.64	0.01	1.17	1.22	--
	7	33,835	0.48	0.02	0.92	0.83	--
	8	33,835	2.60	0.01	1.03	1.04	--
	9	33,835	1.67	0.01	0.99	0.98	--
	10	33,835	2.97	0.01	1.08	1.13	--
	11	33,835	2.97	0.01	1.05	1.07	--
	12	33,835	1.95	0.01	0.90	0.85	--
	13	33,835	0.92	0.02	1.17	1.32	--
	14	33,835	1.51	0.01	0.92	0.84	--
	15	33,835	1.55	0.01	1.08	1.16	--
	16	33,835	1.69	0.01	0.94	0.92	--
	17	33,835	2.10	0.01	1.08	1.09	--
	18	33,835	1.92	0.01	0.96	0.92	--
	19	33,835	2.71	0.01	0.93	0.92	--
	20	33,835	2.22	0.01	0.93	0.91	--
	21	33,835	2.59	0.01	0.97	0.96	--
	22	33,835	2.57	0.01	1.06	1.08	--
	23	33,835	3.46	0.01	1.12	1.26	--
	24	33,835	3.00	0.01	0.92	0.92	--
	25	33,835	2.75	0.01	1.04	1.05	--
	26	33,835	1.86	0.01	0.99	0.95	--
	27	33,835	3.32	0.01	1.22	1.36	--
WR	1	33,835	0.13	0.02	0.97	0.85	--
	2	33,835	-0.62	0.03	0.84	0.63	--
	3	33,835	1.65	0.01	0.91	0.84	--
	4	33,835	0.48	0.02	0.88	0.69	--
	5	33,835	1.93	0.01	1.05	1.06	--
	6	33,835	2.68	0.01	0.99	0.99	--
	7	33,835	0.14	0.02	0.95	0.84	--
	8	33,835	0.85	0.02	0.94	0.88	--
	9	33,835	2.19	0.01	1.06	1.07	--
	10	33,835	1.07	0.01	0.80	0.71	--
	11	33,835	0.57	0.02	0.88	0.78	--
	12	33,835	2.70	0.01	1.19	1.27	--
	13	33,835	2.59	0.01	1.05	1.07	--
	14	33,835	2.02	0.01	1.14	1.16	--
	15	33,835	1.72	0.01	0.86	0.80	--
	16	33,835	0.66	0.01	0.90	0.84	--
	17	33,835	0.93	0.01	0.89	0.87	--
	18	33,835	1.43	0.01	0.95	0.97	--
	19	33,835	2.08	0.01	0.92	0.93	--

APPENDIX E: ITEM INFORMATION AT DIFFERENT CUT POINTS BY GRADES

Note: The statistics in Tables E.1–E.5 were based on the equating sample. Invalid cases were removed after cleaning procedures.

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	-3.13	0.23	0.08	0.03
	2	-2.79	0.25	0.10	0.04
	3	-3.58	0.19	0.05	0.02
	4	-3.14	0.23	0.08	0.03
	5	-3.46	0.20	0.06	0.02
	6	-4.47	0.11	0.02	0.01
	7	-3.46	0.20	0.06	0.02
	8	-1.83	0.22	0.19	0.09
	9	-2.54	0.25	0.12	0.05
	10	-3.76	0.18	0.04	0.02
	11	-1.09	0.15	0.24	0.16
	12	-2.25	0.24	0.15	0.06
	13	-2.43	0.25	0.13	0.06
	14	-1.60	0.20	0.21	0.11
	15	-3.81	0.17	0.04	0.02
	16	-3.31	0.22	0.07	0.02
	17	-1.98	0.23	0.17	0.08
	18	-2.41	0.25	0.13	0.06
	19	-1.98	0.23	0.17	0.08
	20	-1.73	0.21	0.20	0.10
	21	-0.54	0.10	0.25	0.21
	22	-0.06	0.07	0.22	0.24
	23	-1.18	0.16	0.24	0.15
	24	-0.26	0.08	0.24	0.23
SP	1	-1.38	0.35	0.40	0.27
	2	-1.87	0.43	0.36	0.19
	3	-1.17	0.32	0.45	0.30
	4	-1.51	0.38	0.41	0.25
	5	-1.76	0.39	0.36	0.22
	6	-0.63	0.61	0.65	0.56
	7	-1.76	0.33	0.33	0.24
	8	-1.62	0.28	0.28	0.27
	9	-1.96	0.37	0.33	0.21
	10	-1.74	0.28	0.29	0.26
	11	-1.55	0.31	0.31	0.27
	12	-1.82	0.50	0.39	0.15
	13	-1.99	0.47	0.34	0.16
	14	-2.33	0.59	0.26	0.09
	15	-1.84	0.45	0.38	0.18
	16	-1.42	0.38	0.49	0.22

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	-3.36	0.13	0.05	0.03
	2	-2.96	0.17	0.07	0.04
	3	-3.05	0.16	0.07	0.03
	4	-1.86	0.25	0.16	0.09
	5	-2.21	0.23	0.13	0.07
	6	-2.50	0.21	0.11	0.06
	7	-2.97	0.17	0.07	0.04
	8	-1.45	0.25	0.20	0.13
	9	-2.05	0.24	0.14	0.08
	10	-1.78	0.25	0.17	0.10
	11	-1.18	0.23	0.22	0.15
	12	-1.99	0.24	0.15	0.09
	13	-1.52	0.25	0.20	0.12
	14	-1.17	0.23	0.22	0.15
	15	-2.55	0.21	0.10	0.05
WR	1	-4.50	0.05	0.02	0.01
	2	-3.59	0.11	0.04	0.02
	3	-2.80	0.19	0.08	0.04
	4	-3.07	0.16	0.07	0.03
	5	-3.40	0.13	0.05	0.02
	6	-4.06	0.08	0.03	0.01
	7	-3.48	0.27	0.20	0.12
	8	-3.01	0.22	0.26	0.20
	9	-3.62	0.11	0.02	0.01
	10	-3.38	0.14	0.02	0.00
	11	-2.24	0.58	0.21	0.09
	12	-1.93	0.54	0.31	0.16
	13	-0.97	0.44	0.48	0.35
	14	-0.90	0.41	0.46	0.36
	15	-0.17	0.30	0.40	0.39

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	-3.13	0.21	0.06	0.01
	2	-2.79	0.23	0.08	0.01
	3	-3.58	0.17	0.04	0.01
	4	-3.14	0.21	0.06	0.01
	5	-3.46	0.18	0.05	0.01
	6	-4.47	0.09	0.02	0.00
	7	-3.46	0.18	0.05	0.01
	8	-1.83	0.24	0.17	0.03
	9	-2.54	0.25	0.10	0.02
	10	-3.76	0.15	0.04	0.00
	11	-1.09	0.18	0.23	0.06
	12	-2.25	0.25	0.13	0.02
	13	-2.43	0.25	0.11	0.02
	14	-1.60	0.22	0.19	0.04
	15	-3.81	0.15	0.03	0.00
	16	-3.31	0.19	0.05	0.01
	17	-1.98	0.24	0.15	0.03
	18	-2.41	0.25	0.11	0.02
	19	-1.98	0.24	0.15	0.03
	20	-1.73	0.23	0.17	0.03
	21	-0.54	0.13	0.25	0.09
	22	-0.06	0.09	0.24	0.13
	23	-1.18	0.19	0.22	0.06
	24	-0.26	0.10	0.25	0.12
SP	1	-1.38	0.38	0.38	0.12
	2	-1.87	0.45	0.32	0.07
	3	-1.17	0.37	0.43	0.12
	4	-1.51	0.41	0.38	0.09
	5	-1.76	0.41	0.34	0.08
	6	-0.63	0.67	0.63	0.43
	7	-1.76	0.34	0.32	0.11
	8	-1.62	0.27	0.28	0.16
	9	-1.96	0.38	0.31	0.08
	10	-1.74	0.27	0.29	0.14
	11	-1.55	0.31	0.31	0.14
	12	-1.82	0.57	0.32	0.04
	13	-1.99	0.50	0.30	0.05
	14	-2.33	0.60	0.21	0.03
	15	-1.84	0.49	0.33	0.06
	16	-1.42	0.45	0.44	0.07

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	-3.36	0.08	0.03	0.01
	2	-2.96	0.11	0.04	0.01
	3	-3.05	0.11	0.03	0.01
	4	-1.86	0.21	0.09	0.02
	5	-2.21	0.18	0.07	0.02
	6	-2.50	0.16	0.06	0.01
	7	-2.97	0.11	0.04	0.01
	8	-1.45	0.24	0.13	0.03
	9	-2.05	0.20	0.08	0.02
	10	-1.78	0.22	0.10	0.02
	11	-1.18	0.25	0.15	0.04
	12	-1.99	0.20	0.09	0.02
	13	-1.52	0.24	0.12	0.03
	14	-1.17	0.25	0.15	0.04
	15	-2.55	0.15	0.05	0.01
WR	1	-4.50	0.03	0.01	0.00
	2	-3.59	0.07	0.02	0.00
	3	-2.80	0.13	0.04	0.01
	4	-3.07	0.11	0.03	0.01
	5	-3.40	0.08	0.02	0.00
	6	-4.06	0.05	0.01	0.00
	7	-3.48	0.25	0.12	0.03
	8	-3.01	0.26	0.20	0.06
	9	-3.62	0.04	0.01	0.00
	10	-3.38	0.05	0.00	0.00
	11	-2.24	0.37	0.09	0.02
	12	-1.93	0.44	0.16	0.03
	13	-0.97	0.50	0.35	0.10
	14	-0.90	0.47	0.36	0.12
	15	-0.17	0.37	0.39	0.23

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	-2.01	0.24	0.09	0.02
	2	0.14	0.12	0.25	0.14
	3	-0.86	0.22	0.20	0.06
	4	-0.69	0.20	0.21	0.07
	5	0.96	0.06	0.21	0.21
	6	-1.50	0.25	0.14	0.04
	7	-1.81	0.25	0.11	0.03
	8	0.77	0.08	0.23	0.20
	9	1.22	0.05	0.19	0.23
	10	1.85	0.03	0.13	0.25
	11	1.14	0.06	0.20	0.23
	12	0.34	0.11	0.25	0.16
	13	-0.57	0.19	0.22	0.08
	14	1.52	0.04	0.16	0.25
	15	0.70	0.08	0.23	0.19
	16	0.83	0.07	0.22	0.20
	17	0.38	0.10	0.25	0.16
	18	1.15	0.06	0.19	0.23
	19	1.42	0.04	0.17	0.24
	20	-0.26	0.16	0.24	0.10
	21	0.07	0.13	0.25	0.13
	22	-1.34	0.24	0.15	0.04
	23	-0.81	0.21	0.20	0.07
	24	-0.42	0.18	0.23	0.09
SP	1	-0.63	0.35	0.36	0.20
	2	-0.15	0.31	0.44	0.24
	3	-0.52	0.36	0.42	0.18
	4	-0.13	0.30	0.46	0.24
	5	-0.18	0.31	0.34	0.26
	6	0.34	0.62	0.67	0.52
	7	-0.89	0.32	0.31	0.19
	8	0.03	0.29	0.39	0.28
	9	-0.29	0.32	0.36	0.24
	10	-0.17	0.31	0.34	0.26
	11	-0.05	0.30	0.36	0.27
	12	-0.92	0.36	0.34	0.16
	13	-0.40	0.35	0.53	0.16
	14	-0.20	0.31	0.53	0.20
	15	-0.15	0.29	0.59	0.19
	16	-0.64	0.40	0.44	0.15

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	-2.65	0.11	0.03	0.01
	2	-1.61	0.21	0.08	0.03
	3	-1.72	0.20	0.07	0.03
	4	-2.74	0.11	0.03	0.01
	5	-2.15	0.16	0.05	0.02
	6	-2.64	0.12	0.03	0.01
	7	-1.97	0.18	0.06	0.02
	8	-2.10	0.17	0.05	0.02
	9	0.10	0.21	0.23	0.12
	10	-0.82	0.25	0.14	0.06
	11	0.77	0.15	0.25	0.19
	12	-1.55	0.22	0.08	0.03
	13	0.39	0.18	0.24	0.15
	14	0.33	0.19	0.24	0.14
	15	1.94	0.06	0.18	0.25
	16	0.81	0.14	0.25	0.19
	17	0.42	0.18	0.24	0.15
	18	0.89	0.13	0.25	0.20
	19	0.02	0.22	0.22	0.12
	20	-0.34	0.24	0.19	0.09
	21	1.26	0.10	0.23	0.23
	22	0.47	0.17	0.25	0.16
	23	-0.76	0.25	0.15	0.06
	24	1.46	0.09	0.22	0.24
WR	1	-2.12	0.16	0.05	0.02
	2	-2.61	0.12	0.03	0.01
	3	-0.27	0.24	0.20	0.09
	4	-1.02	0.25	0.13	0.05
	5	0.61	0.16	0.25	0.17
	6	-0.29	0.24	0.19	0.09
	7	0.34	0.19	0.24	0.14
	8	0.31	0.19	0.24	0.14
	9	0.20	0.20	0.23	0.13
	10	-1.17	0.24	0.11	0.04
	11	-0.09	0.22	0.21	0.11
	12	0.33	0.19	0.24	0.14
	13	-2.03	0.34	0.13	0.05
	14	-0.42	0.53	0.37	0.15
	15	-0.04	0.37	0.37	0.25
	16	1.19	0.47	0.52	0.55

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	-2.01	0.10	0.03	0.01
	2	0.14	0.19	0.07	0.02
	3	-0.86	0.18	0.07	0.02
	4	-0.69	0.09	0.03	0.01
	5	0.96	0.14	0.04	0.01
	6	-1.50	0.10	0.03	0.01
	7	-1.81	0.15	0.05	0.01
	8	0.77	0.14	0.05	0.01
	9	1.22	0.23	0.22	0.08
	10	1.85	0.24	0.13	0.04
	11	1.14	0.17	0.25	0.14
	12	0.34	0.19	0.08	0.02
	13	-0.57	0.20	0.24	0.11
	14	1.52	0.21	0.23	0.10
	15	0.70	0.07	0.19	0.24
	16	0.83	0.17	0.25	0.14
	17	0.38	0.20	0.24	0.11
	18	1.15	0.16	0.25	0.15
	19	1.42	0.23	0.21	0.08
	20	-0.26	0.25	0.18	0.06
	21	0.07	0.12	0.24	0.19
	22	-1.34	0.20	0.24	0.11
	23	-0.81	0.25	0.14	0.04
	24	-0.42	0.11	0.23	0.20
SP	1	-0.63	0.14	0.05	0.01
	2	-0.15	0.10	0.03	0.01
	3	-0.52	0.25	0.18	0.06
	4	-0.13	0.23	0.11	0.03
	5	-0.18	0.18	0.25	0.12
	6	0.34	0.25	0.18	0.06
	7	-0.89	0.21	0.23	0.10
	8	0.03	0.21	0.23	0.10
	9	-0.29	0.22	0.22	0.09
	10	-0.17	0.23	0.10	0.03
	11	-0.05	0.24	0.20	0.07
	12	-0.92	0.21	0.23	0.10
	13	-0.40	0.30	0.12	0.03
	14	-0.20	0.55	0.34	0.09
	15	-0.15	0.38	0.36	0.19
	16	-0.64	0.48	0.52	0.55

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	-2.65	0.10	0.03	0.01
	2	-1.61	0.19	0.07	0.02
	3	-1.72	0.18	0.07	0.02
	4	-2.74	0.09	0.03	0.01
	5	-2.15	0.14	0.04	0.01
	6	-2.64	0.10	0.03	0.01
	7	-1.97	0.15	0.05	0.01
	8	-2.10	0.14	0.05	0.01
	9	0.10	0.23	0.22	0.08
	10	-0.82	0.24	0.13	0.04
	11	0.77	0.17	0.25	0.14
	12	-1.55	0.19	0.08	0.02
	13	0.39	0.20	0.24	0.11
	14	0.33	0.21	0.23	0.10
	15	1.94	0.07	0.19	0.24
	16	0.81	0.17	0.25	0.14
	17	0.42	0.20	0.24	0.11
	18	0.89	0.16	0.25	0.15
	19	0.02	0.23	0.21	0.08
	20	-0.34	0.25	0.18	0.06
	21	1.26	0.12	0.24	0.19
	22	0.47	0.20	0.24	0.11
	23	-0.76	0.25	0.14	0.04
	24	1.46	0.11	0.23	0.20
WR	1	-2.12	0.14	0.05	0.01
	2	-2.61	0.10	0.03	0.01
	3	-0.27	0.25	0.18	0.06
	4	-1.02	0.23	0.11	0.03
	5	0.61	0.18	0.25	0.12
	6	-0.29	0.25	0.18	0.06
	7	0.34	0.21	0.23	0.10
	8	0.31	0.21	0.23	0.10
	9	0.20	0.22	0.22	0.09
	10	-1.17	0.23	0.10	0.03
	11	-0.09	0.24	0.20	0.07
	12	0.33	0.21	0.23	0.10
	13	-2.03	0.30	0.12	0.03
	14	-0.42	0.55	0.34	0.09
	15	-0.04	0.38	0.36	0.19
	16	1.19	0.48	0.52	0.55

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	-2.01	0.19	0.06	0.01
	2	0.14	0.19	0.23	0.08
	3	-0.86	0.25	0.15	0.03
	4	-0.69	0.25	0.16	0.04
	5	0.96	0.11	0.24	0.15
	6	-1.50	0.23	0.09	0.02
	7	-1.81	0.21	0.07	0.01
	8	0.77	0.13	0.25	0.13
	9	1.22	0.09	0.23	0.18
	10	1.85	0.06	0.18	0.23
	11	1.14	0.10	0.24	0.17
	12	0.34	0.17	0.24	0.10
	13	-0.57	0.24	0.18	0.04
	14	1.52	0.07	0.21	0.20
	15	0.70	0.14	0.25	0.13
	16	0.83	0.13	0.25	0.14
	17	0.38	0.17	0.25	0.10
	18	1.15	0.10	0.24	0.17
	19	1.42	0.08	0.22	0.19
	20	-0.26	0.22	0.20	0.06
	21	0.07	0.20	0.23	0.08
	22	-1.34	0.24	0.10	0.02
	23	-0.81	0.25	0.15	0.04
	24	-0.42	0.24	0.19	0.05
SP	1	-0.63	0.37	0.32	0.12
	2	-0.15	0.40	0.40	0.15
	3	-0.52	0.44	0.35	0.11
	4	-0.13	0.41	0.41	0.15
	5	-0.18	0.34	0.33	0.18
	6	0.34	0.77	0.60	0.44
	7	-0.89	0.31	0.30	0.12
	8	0.03	0.36	0.38	0.19
	9	-0.29	0.36	0.35	0.16
	10	-0.17	0.34	0.33	0.18
	11	-0.05	0.34	0.35	0.19
	12	-0.92	0.38	0.30	0.10
	13	-0.40	0.53	0.40	0.08
	14	-0.20	0.47	0.44	0.11
	15	-0.15	0.49	0.47	0.10
	16	-0.64	0.50	0.34	0.08

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	-2.65	0.08	0.02	0.00
	2	-1.61	0.17	0.05	0.01
	3	-1.72	0.15	0.04	0.01
	4	-2.74	0.07	0.02	0.00
	5	-2.15	0.12	0.03	0.01
	6	-2.64	0.08	0.02	0.00
	7	-1.97	0.13	0.03	0.01
	8	-2.10	0.12	0.03	0.01
	9	0.10	0.24	0.18	0.07
	10	-0.82	0.23	0.09	0.03
	11	0.77	0.19	0.23	0.11
	12	-1.55	0.17	0.05	0.01
	13	0.39	0.22	0.20	0.08
	14	0.33	0.23	0.20	0.08
	15	1.94	0.09	0.23	0.22
	16	0.81	0.19	0.23	0.11
	17	0.42	0.22	0.21	0.09
	18	0.89	0.18	0.24	0.12
	19	0.02	0.24	0.17	0.06
	20	-0.34	0.25	0.13	0.04
	21	1.26	0.15	0.25	0.16
	22	0.47	0.22	0.21	0.09
	23	-0.76	0.24	0.10	0.03
	24	1.46	0.13	0.25	0.18
WR	1	-2.12	0.12	0.03	0.01
	2	-2.61	0.08	0.02	0.00
	3	-0.27	0.25	0.14	0.05
	4	-1.02	0.22	0.08	0.02
	5	0.61	0.21	0.22	0.10
	6	-0.29	0.25	0.14	0.05
	7	0.34	0.23	0.20	0.08
	8	0.31	0.23	0.20	0.08
	9	0.20	0.24	0.19	0.07
	10	-1.17	0.21	0.07	0.02
	11	-0.09	0.25	0.16	0.06
	12	0.33	0.23	0.20	0.08
	13	-2.03	0.27	0.08	0.02
	14	-0.42	0.54	0.24	0.07
	15	-0.04	0.38	0.32	0.15
	16	1.19	0.50	0.53	0.55

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	-1.43	0.21	0.08	0.02
	2	1.04	0.14	0.25	0.15
	3	0.97	0.15	0.25	0.14
	4	-0.79	0.25	0.13	0.03
	5	-0.46	0.25	0.16	0.05
	6	0.35	0.20	0.23	0.09
	7	0.82	0.16	0.25	0.13
	8	-0.48	0.25	0.16	0.05
	9	-0.23	0.24	0.19	0.06
	10	0.30	0.21	0.23	0.09
	11	0.86	0.16	0.25	0.13
	12	0.70	0.17	0.25	0.12
	13	1.19	0.13	0.24	0.17
	14	-1.31	0.22	0.09	0.02
	15	1.52	0.10	0.23	0.20
	16	-0.09	0.24	0.20	0.06
	17	2.00	0.07	0.19	0.23
	18	1.91	0.07	0.19	0.23
	19	0.05	0.23	0.21	0.07
	20	0.02	0.23	0.21	0.07
	21	1.32	0.11	0.24	0.18
	22	0.37	0.20	0.23	0.09
	23	2.47	0.04	0.14	0.25
	24	0.43	0.20	0.24	0.10
	25	1.99	0.07	0.19	0.23
SP	1	-0.35	0.40	0.34	0.13
	2	-0.68	0.40	0.30	0.10
	3	0.15	0.44	0.44	0.16
	4	0.01	0.44	0.41	0.15
	5	0.38	0.37	0.41	0.21
	6	0.69	0.75	0.68	0.50
	7	-0.43	0.40	0.33	0.12
	8	0.05	0.43	0.40	0.16
	9	0.22	0.41	0.43	0.18
	10	-0.22	0.39	0.35	0.15
	11	-0.89	0.33	0.28	0.11
	12	-0.04	0.62	0.46	0.07
	13	0.25	0.42	0.45	0.17
	14	-0.88	0.56	0.23	0.05
	15	0.00	0.56	0.46	0.10
	16	0.45	0.41	0.58	0.16

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	-1.55	0.10	0.03	0.01
	2	-0.21	0.22	0.11	0.04
	3	0.26	0.24	0.15	0.06
	4	-1.33	0.11	0.04	0.01
	5	1.64	0.19	0.25	0.17
	6	0.53	0.25	0.17	0.07
	7	0.16	0.24	0.14	0.05
	8	0.28	0.24	0.15	0.06
	9	2.85	0.08	0.19	0.25
	10	0.30	0.25	0.15	0.06
	11	1.29	0.22	0.24	0.13
	12	0.78	0.25	0.20	0.09
	13	2.47	0.11	0.22	0.24
	14	0.76	0.25	0.20	0.09
	15	1.63	0.19	0.25	0.16
	16	1.16	0.23	0.23	0.12
	17	0.98	0.24	0.22	0.11
	18	0.57	0.25	0.18	0.08
	19	0.56	0.25	0.18	0.08
	20	1.33	0.22	0.24	0.14
	21	2.09	0.15	0.24	0.21
	22	1.66	0.19	0.25	0.17
	23	0.96	0.24	0.21	0.10
	24	1.66	0.19	0.25	0.17
	25	0.69	0.25	0.19	0.08
	26	1.25	0.22	0.23	0.13
	27	1.34	0.22	0.24	0.14
WR	1	0.56	0.25	0.18	0.08
	2	-0.44	0.19	0.09	0.03
	3	-0.05	0.23	0.12	0.04
	4	0.27	0.24	0.15	0.06
	5	1.73	0.18	0.25	0.17
	6	0.11	0.24	0.13	0.05
	7	2.24	0.13	0.24	0.22
	8	-0.72	0.17	0.07	0.02
	9	0.57	0.25	0.18	0.08
	10	0.33	0.25	0.16	0.06
	11	2.57	0.11	0.21	0.24
	12	-0.52	0.19	0.08	0.03
	13	1.36	0.22	0.24	0.14
	14	2.03	0.15	0.25	0.20
	15	2.14	0.14	0.24	0.21
	16	-0.57	0.37	0.16	0.06
	17	0.31	0.54	0.30	0.11
	18	1.27	0.41	0.42	0.27
	19	2.00	0.58	0.57	0.55

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	-1.43	0.19	0.07	0.01
	2	1.04	0.16	0.25	0.11
	3	0.97	0.17	0.25	0.11
	4	-0.79	0.24	0.11	0.02
	5	-0.46	0.25	0.14	0.03
	6	0.35	0.22	0.21	0.07
	7	0.82	0.18	0.24	0.10
	8	-0.48	0.25	0.14	0.03
	9	-0.23	0.25	0.16	0.04
	10	0.30	0.23	0.21	0.06
	11	0.86	0.18	0.24	0.10
	12	0.70	0.19	0.24	0.09
	13	1.19	0.15	0.25	0.13
	14	-1.31	0.20	0.07	0.01
	15	1.52	0.12	0.24	0.16
	16	-0.09	0.25	0.17	0.04
	17	2.00	0.08	0.21	0.20
	18	1.91	0.09	0.22	0.19
	19	0.05	0.24	0.19	0.05
	20	0.02	0.24	0.18	0.05
	21	1.32	0.14	0.25	0.14
	22	0.37	0.22	0.22	0.07
	23	2.47	0.05	0.17	0.24
	24	0.43	0.22	0.22	0.07
	25	1.99	0.08	0.21	0.20
SP	1	-0.35	0.40	0.31	0.10
	2	-0.68	0.39	0.26	0.07
	3	0.15	0.47	0.39	0.11
	4	0.01	0.47	0.36	0.10
	5	0.38	0.40	0.39	0.16
	6	0.69	0.77	0.65	0.43
	7	-0.43	0.40	0.29	0.09
	8	0.05	0.45	0.37	0.11
	9	0.22	0.44	0.39	0.13
	10	-0.22	0.40	0.32	0.11
	11	-0.89	0.33	0.25	0.08
	12	-0.04	0.70	0.35	0.04
	13	0.25	0.46	0.41	0.12
	14	-0.88	0.51	0.18	0.03
	15	0.00	0.62	0.37	0.06
	16	0.45	0.49	0.51	0.10

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	-1.55	0.09	0.03	0.01
	2	-0.21	0.21	0.09	0.03
	3	0.26	0.24	0.13	0.05
	4	-1.33	0.10	0.03	0.01
	5	1.64	0.20	0.24	0.15
	6	0.53	0.25	0.15	0.06
	7	0.16	0.23	0.12	0.05
	8	0.28	0.24	0.13	0.05
	9	2.85	0.09	0.21	0.24
	10	0.30	0.24	0.13	0.05
	11	1.29	0.23	0.22	0.12
	12	0.78	0.25	0.18	0.08
	13	2.47	0.12	0.24	0.22
	14	0.76	0.25	0.18	0.08
	15	1.63	0.20	0.24	0.15
	16	1.16	0.24	0.21	0.11
	17	0.98	0.24	0.20	0.09
	18	0.57	0.25	0.16	0.07
	19	0.56	0.25	0.16	0.06
	20	1.33	0.22	0.22	0.12
	21	2.09	0.16	0.25	0.19
	22	1.66	0.20	0.24	0.15
	23	0.96	0.24	0.19	0.09
	24	1.66	0.20	0.24	0.15
	25	0.69	0.25	0.17	0.07
	26	1.25	0.23	0.22	0.11
	27	1.34	0.22	0.23	0.12
WR	1	0.56	0.25	0.16	0.06
	2	-0.44	0.19	0.07	0.03
	3	-0.05	0.22	0.10	0.04
	4	0.27	0.24	0.13	0.05
	5	1.73	0.19	0.25	0.16
	6	0.11	0.23	0.12	0.04
	7	2.24	0.14	0.25	0.20
	8	-0.72	0.16	0.06	0.02
	9	0.57	0.25	0.16	0.07
	10	0.33	0.24	0.13	0.05
	11	2.57	0.11	0.23	0.23
	12	-0.52	0.18	0.07	0.02
	13	1.36	0.22	0.23	0.12
	14	2.03	0.16	0.25	0.18
	15	2.14	0.15	0.25	0.20
	16	-0.57	0.35	0.14	0.05
	17	0.31	0.53	0.25	0.09
	18	1.27	0.41	0.40	0.25
	19	2.00	0.59	0.56	0.55

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	-1.36	0.16	0.05	0.01
	2	-0.44	0.24	0.11	0.03
	3	0.35	0.24	0.18	0.06
	4	0.78	0.22	0.22	0.08
	5	2.42	0.07	0.20	0.23
	6	2.59	0.06	0.19	0.24
	7	2.83	0.05	0.17	0.25
	8	1.46	0.15	0.25	0.14
	9	0.52	0.23	0.20	0.07
	10	2.17	0.09	0.22	0.21
	11	2.33	0.08	0.21	0.22
	12	2.42	0.07	0.20	0.23
	13	3.42	0.03	0.11	0.24
	14	2.69	0.06	0.18	0.24
	15	1.23	0.17	0.25	0.12
	16	1.20	0.18	0.24	0.12
	17	0.62	0.23	0.21	0.07
	18	1.90	0.11	0.24	0.18
	19	1.79	0.12	0.24	0.17
	20	1.38	0.16	0.25	0.13
	21	0.90	0.20	0.23	0.09
	22	-0.22	0.25	0.13	0.04
	23	1.85	0.12	0.24	0.18
	24	3.32	0.03	0.12	0.25
	25	1.22	0.18	0.25	0.12
SP	1	0.42	0.41	0.36	0.15
	2	1.03	0.39	0.47	0.21
	3	0.86	0.43	0.48	0.17
	4	0.45	0.37	0.34	0.17
	5	1.14	0.37	0.50	0.22
	6	1.08	0.79	0.75	0.47
	7	0.35	0.36	0.33	0.16
	8	0.38	0.45	0.36	0.13
	9	0.41	0.41	0.36	0.15
	10	0.43	0.40	0.35	0.16
	11	-0.10	0.39	0.29	0.11
	12	0.85	0.42	0.45	0.18
	13	1.01	0.40	0.52	0.18
	14	0.82	0.45	0.49	0.15
	15	1.30	0.33	0.65	0.20
	16	0.94	0.43	0.52	0.16

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	-0.19	0.19	0.08	0.03
	2	-1.02	0.11	0.04	0.01
	3	-1.20	0.10	0.03	0.01
	4	1.53	0.23	0.23	0.14
	5	1.04	0.25	0.18	0.10
	6	-0.38	0.17	0.07	0.03
	7	1.65	0.22	0.23	0.15
	8	1.02	0.25	0.18	0.09
	9	1.04	0.25	0.18	0.10
	10	1.13	0.25	0.19	0.10
	11	1.19	0.24	0.20	0.11
	12	0.17	0.22	0.11	0.05
	13	0.45	0.24	0.13	0.06
	14	2.17	0.17	0.25	0.20
	15	1.12	0.25	0.19	0.10
	16	1.51	0.23	0.22	0.14
	17	1.40	0.23	0.22	0.13
	18	1.67	0.22	0.24	0.15
	19	1.78	0.21	0.24	0.16
	20	1.86	0.20	0.24	0.17
	21	2.37	0.15	0.25	0.22
	22	1.64	0.22	0.23	0.15
	23	2.50	0.14	0.24	0.23
	24	0.92	0.25	0.17	0.09
	25	1.83	0.20	0.24	0.17
	26	1.06	0.25	0.19	0.10
	27	1.86	0.20	0.24	0.17
WR	1	-0.51	0.16	0.06	0.02
	2	-0.82	0.13	0.05	0.02
	3	2.40	0.15	0.25	0.22
	4	1.41	0.23	0.22	0.13
	5	0.07	0.21	0.10	0.04
	6	0.00	0.21	0.09	0.04
	7	1.09	0.25	0.19	0.10
	8	0.14	0.22	0.10	0.04
	9	1.82	0.20	0.24	0.17
	10	1.03	0.25	0.18	0.10
	11	1.34	0.24	0.21	0.12
	12	-0.13	0.19	0.08	0.03
	13	1.18	0.24	0.20	0.11
	14	0.86	0.25	0.17	0.08
	15	1.49	0.23	0.22	0.13
	16	0.97	0.57	0.36	0.16
	17	-0.45	0.32	0.11	0.04
	18	0.76	0.51	0.31	0.15
	19	1.74	0.63	0.61	0.58

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	-1.36	0.15	0.04	0.01
	2	-0.44	0.23	0.09	0.03
	3	0.35	0.25	0.16	0.06
	4	0.78	0.23	0.20	0.08
	5	2.42	0.09	0.22	0.23
	6	2.59	0.08	0.21	0.24
	7	2.83	0.06	0.19	0.25
	8	1.46	0.17	0.24	0.14
	9	0.52	0.24	0.18	0.07
	10	2.17	0.11	0.24	0.21
	11	2.33	0.09	0.23	0.22
	12	2.42	0.09	0.22	0.23
	13	3.42	0.04	0.13	0.24
	14	2.69	0.07	0.20	0.24
	15	1.23	0.19	0.23	0.12
	16	1.20	0.20	0.23	0.12
	17	0.62	0.24	0.18	0.07
	18	1.90	0.13	0.25	0.18
	19	1.79	0.14	0.25	0.17
	20	1.38	0.18	0.24	0.13
	21	0.90	0.22	0.21	0.09
	22	-0.22	0.24	0.11	0.04
	23	1.85	0.13	0.25	0.18
	24	3.32	0.04	0.14	0.25
	25	1.22	0.19	0.23	0.12
SP	1	0.42	0.42	0.33	0.15
	2	1.03	0.42	0.44	0.21
	3	0.86	0.47	0.43	0.17
	4	0.45	0.37	0.32	0.17
	5	1.14	0.41	0.47	0.22
	6	1.08	0.82	0.73	0.47
	7	0.35	0.37	0.31	0.16
	8	0.38	0.46	0.32	0.13
	9	0.41	0.42	0.33	0.15
	10	0.43	0.40	0.33	0.16
	11	-0.10	0.39	0.26	0.11
	12	0.85	0.45	0.41	0.18
	13	1.01	0.45	0.47	0.18
	14	0.82	0.50	0.43	0.15
	15	1.30	0.39	0.60	0.20
	16	0.94	0.48	0.46	0.16

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	-0.19	0.18	0.06	0.03
	2	-1.02	0.10	0.03	0.01
	3	-1.20	0.09	0.03	0.01
	4	1.53	0.23	0.21	0.14
	5	1.04	0.25	0.16	0.10
	6	-0.38	0.16	0.05	0.03
	7	1.65	0.22	0.22	0.15
	8	1.02	0.25	0.16	0.09
	9	1.04	0.25	0.16	0.10
	10	1.13	0.25	0.17	0.10
	11	1.19	0.25	0.18	0.11
	12	0.17	0.21	0.09	0.05
	13	0.45	0.23	0.11	0.06
	14	2.17	0.18	0.25	0.20
	15	1.12	0.25	0.17	0.10
	16	1.51	0.23	0.21	0.14
	17	1.40	0.24	0.20	0.13
	18	1.67	0.22	0.22	0.15
	19	1.78	0.21	0.23	0.16
	20	1.86	0.21	0.23	0.17
	21	2.37	0.16	0.25	0.22
	22	1.64	0.22	0.22	0.15
	23	2.50	0.15	0.25	0.23
	24	0.92	0.25	0.15	0.09
	25	1.83	0.21	0.23	0.17
	26	1.06	0.25	0.16	0.10
	27	1.86	0.21	0.23	0.17
WR	1	-0.51	0.15	0.05	0.02
	2	-0.82	0.12	0.04	0.02
	3	2.40	0.16	0.25	0.22
	4	1.41	0.24	0.20	0.13
	5	0.07	0.20	0.08	0.04
	6	0.00	0.20	0.08	0.04
	7	1.09	0.25	0.17	0.10
	8	0.14	0.21	0.08	0.04
	9	1.82	0.21	0.23	0.17
	10	1.03	0.25	0.16	0.10
	11	1.34	0.24	0.19	0.12
	12	-0.13	0.19	0.07	0.03
	13	1.18	0.25	0.18	0.11
	14	0.86	0.25	0.14	0.08
	15	1.49	0.23	0.20	0.13
	16	0.97	0.57	0.30	0.16
	17	-0.45	0.30	0.09	0.04
	18	0.76	0.50	0.27	0.15
	19	1.74	0.63	0.61	0.58

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	0.30	0.24	0.11	0.04
	2	0.97	0.24	0.17	0.08
	3	0.23	0.24	0.10	0.04
	4	1.27	0.23	0.19	0.10
	5	2.06	0.16	0.25	0.17
	6	0.76	0.25	0.15	0.07
	7	0.93	0.25	0.16	0.08
	8	1.66	0.20	0.23	0.14
	9	1.15	0.24	0.18	0.09
	10	2.76	0.10	0.24	0.23
	11	2.93	0.08	0.23	0.24
	12	1.63	0.20	0.22	0.13
	13	2.51	0.12	0.25	0.21
	14	2.50	0.12	0.25	0.21
	15	2.90	0.09	0.23	0.24
	16	2.50	0.12	0.25	0.21
	17	2.34	0.13	0.25	0.20
	18	2.64	0.11	0.24	0.22
	19	1.51	0.21	0.22	0.12
	20	2.16	0.15	0.25	0.18
	21	2.67	0.10	0.24	0.23
	22	0.84	0.25	0.15	0.07
	23	0.25	0.24	0.10	0.04
	24	2.36	0.13	0.25	0.20
	25	2.07	0.16	0.25	0.17
SP	1	0.79	0.30	0.30	0.21
	2	0.65	0.32	0.29	0.19
	3	0.69	0.52	0.27	0.12
	4	0.95	0.38	0.32	0.20
	5	0.95	0.51	0.33	0.15
	6	1.35	0.81	0.64	0.47
	7	0.53	0.35	0.28	0.16
	8	0.90	0.35	0.31	0.20
	9	0.84	0.36	0.31	0.19
	10	0.42	0.33	0.27	0.16
	11	1.09	0.40	0.34	0.21
	12	1.04	0.45	0.34	0.18
	13	1.21	0.52	0.39	0.17
	14	0.86	0.45	0.31	0.16
	15	1.53	0.43	0.45	0.24
	16	0.63	0.45	0.28	0.14

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	0.26	0.19	0.06	0.03
	2	0.21	0.19	0.06	0.03
	3	0.52	0.22	0.08	0.04
	4	-0.45	0.13	0.03	0.02
	5	2.00	0.22	0.21	0.13
	6	2.64	0.16	0.25	0.19
	7	0.48	0.21	0.08	0.04
	8	2.60	0.17	0.25	0.19
	9	1.67	0.24	0.18	0.10
	10	2.97	0.13	0.25	0.22
	11	2.97	0.13	0.25	0.22
	12	1.95	0.22	0.20	0.13
	13	0.92	0.24	0.11	0.06
	14	1.51	0.25	0.16	0.09
	15	1.55	0.25	0.17	0.09
	16	1.69	0.24	0.18	0.10
	17	2.10	0.21	0.22	0.14
	18	1.92	0.23	0.20	0.12
	19	2.71	0.16	0.25	0.20
	20	2.22	0.20	0.22	0.15
	21	2.59	0.17	0.25	0.19
	22	2.57	0.17	0.24	0.19
	23	3.46	0.09	0.23	0.25
	24	3.00	0.13	0.25	0.22
	25	2.75	0.15	0.25	0.20
	26	1.86	0.23	0.20	0.12
	27	3.32	0.10	0.24	0.24
WR	1	0.13	0.18	0.06	0.03
	2	-0.62	0.11	0.03	0.01
	3	1.65	0.24	0.18	0.10
	4	0.48	0.21	0.08	0.04
	5	1.93	0.23	0.20	0.13
	6	2.68	0.16	0.25	0.20
	7	0.14	0.18	0.06	0.03
	8	0.85	0.24	0.10	0.05
	9	2.19	0.21	0.22	0.15
	10	1.07	0.25	0.12	0.06
	11	0.57	0.22	0.08	0.04
	12	2.70	0.16	0.25	0.20
	13	2.59	0.17	0.24	0.19
	14	2.02	0.22	0.21	0.13
	15	1.72	0.24	0.18	0.11
	16	0.66	0.52	0.14	0.06
	17	0.93	0.56	0.19	0.09
	18	1.43	0.47	0.31	0.18
	19	2.08	0.79	0.72	0.57

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	0.30	0.23	0.09	0.04
	2	0.97	0.25	0.15	0.07
	3	0.23	0.23	0.09	0.04
	4	1.27	0.24	0.18	0.09
	5	2.06	0.18	0.24	0.16
	6	0.76	0.25	0.13	0.06
	7	0.93	0.25	0.14	0.07
	8	1.66	0.21	0.21	0.12
	9	1.15	0.24	0.16	0.08
	10	2.76	0.11	0.25	0.22
	11	2.93	0.10	0.24	0.23
	12	1.63	0.21	0.21	0.12
	13	2.51	0.13	0.25	0.20
	14	2.50	0.13	0.25	0.20
	15	2.90	0.10	0.24	0.23
	16	2.50	0.13	0.25	0.20
	17	2.34	0.15	0.25	0.18
	18	2.64	0.12	0.25	0.21
	19	1.51	0.22	0.20	0.11
	20	2.16	0.17	0.24	0.17
	21	2.67	0.12	0.25	0.21
	22	0.84	0.25	0.14	0.06
	23	0.25	0.23	0.09	0.04
	24	2.36	0.15	0.25	0.19
	25	2.07	0.17	0.24	0.16
SP	1	0.79	0.30	0.29	0.19
	2	0.65	0.32	0.27	0.17
	3	0.69	0.52	0.24	0.10
	4	0.95	0.38	0.30	0.18
	5	0.95	0.52	0.29	0.13
	6	1.35	0.80	0.61	0.44
	7	0.53	0.35	0.26	0.15
	8	0.90	0.35	0.30	0.19
	9	0.84	0.36	0.29	0.17
	10	0.42	0.33	0.25	0.14
	11	1.09	0.40	0.32	0.19
	12	1.04	0.46	0.31	0.16
	13	1.21	0.55	0.34	0.15
	14	0.86	0.46	0.28	0.14
	15	1.53	0.46	0.41	0.21
	16	0.63	0.45	0.25	0.12

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	0.26	0.19	0.06	0.03
	2	0.21	0.18	0.05	0.02
	3	0.52	0.21	0.07	0.03
	4	-0.45	0.12	0.03	0.01
	5	2.00	0.23	0.20	0.12
	6	2.64	0.17	0.24	0.18
	7	0.48	0.21	0.07	0.03
	8	2.60	0.18	0.24	0.17
	9	1.67	0.24	0.17	0.09
	10	2.97	0.14	0.25	0.21
	11	2.97	0.14	0.25	0.21
	12	1.95	0.23	0.19	0.11
	13	0.92	0.24	0.10	0.05
	14	1.51	0.25	0.15	0.08
	15	1.55	0.25	0.15	0.08
	16	1.69	0.24	0.17	0.09
	17	2.10	0.22	0.21	0.13
	18	1.92	0.23	0.19	0.11
	19	2.71	0.17	0.24	0.18
	20	2.22	0.21	0.22	0.14
	21	2.59	0.18	0.24	0.17
	22	2.57	0.18	0.24	0.17
	23	3.46	0.10	0.24	0.24
	24	3.00	0.14	0.25	0.21
	25	2.75	0.16	0.25	0.19
	26	1.86	0.24	0.18	0.11
	27	3.32	0.11	0.24	0.23
WR	1	0.13	0.17	0.05	0.02
	2	-0.62	0.11	0.03	0.01
	3	1.65	0.25	0.16	0.09
	4	0.48	0.21	0.07	0.03
	5	1.93	0.23	0.19	0.11
	6	2.68	0.17	0.24	0.18
	7	0.14	0.17	0.05	0.02
	8	0.85	0.23	0.09	0.04
	9	2.19	0.21	0.21	0.13
	10	1.07	0.24	0.11	0.05
	11	0.57	0.21	0.07	0.03
	12	2.70	0.17	0.24	0.18
	13	2.59	0.18	0.24	0.17
	14	2.02	0.23	0.20	0.12
	15	1.72	0.24	0.17	0.09
	16	0.66	0.50	0.13	0.05
	17	0.93	0.55	0.17	0.07
	18	1.43	0.47	0.29	0.16
	19	2.08	0.79	0.70	0.53

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	0.30	0.22	0.08	0.03
	2	0.97	0.25	0.13	0.05
	3	0.23	0.22	0.07	0.03
	4	1.27	0.25	0.16	0.07
	5	2.06	0.19	0.23	0.13
	6	0.76	0.25	0.11	0.04
	7	0.93	0.25	0.12	0.05
	8	1.66	0.22	0.19	0.09
	9	1.15	0.25	0.14	0.06
	10	2.76	0.13	0.25	0.19
	11	2.93	0.11	0.25	0.21
	12	1.63	0.23	0.19	0.09
	13	2.51	0.15	0.25	0.17
	14	2.50	0.15	0.25	0.17
	15	2.90	0.11	0.25	0.20
	16	2.50	0.15	0.25	0.17
	17	2.34	0.16	0.24	0.15
	18	2.64	0.14	0.25	0.18
	19	1.51	0.23	0.18	0.08
	20	2.16	0.18	0.23	0.13
	21	2.67	0.13	0.25	0.18
	22	0.84	0.25	0.12	0.05
	23	0.25	0.22	0.07	0.03
	24	2.36	0.16	0.24	0.15
	25	2.07	0.19	0.23	0.13
SP	1	0.79	0.30	0.27	0.15
	2	0.65	0.32	0.26	0.13
	3	0.69	0.51	0.20	0.07
	4	0.95	0.38	0.28	0.14
	5	0.95	0.52	0.25	0.10
	6	1.35	0.79	0.58	0.36
	7	0.53	0.34	0.24	0.11
	8	0.90	0.35	0.27	0.15
	9	0.84	0.36	0.27	0.14
	10	0.42	0.33	0.23	0.11
	11	1.09	0.40	0.29	0.15
	12	1.04	0.46	0.28	0.12
	13	1.21	0.57	0.29	0.11
	14	0.86	0.46	0.25	0.11
	15	1.53	0.49	0.37	0.16
	16	0.63	0.44	0.22	0.09

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	0.26	0.18	0.05	0.02
	2	0.21	0.17	0.04	0.02
	3	0.52	0.20	0.06	0.03
	4	-0.45	0.11	0.02	0.01
	5	2.00	0.23	0.17	0.10
	6	2.64	0.18	0.23	0.16
	7	0.48	0.20	0.06	0.03
	8	2.60	0.19	0.23	0.15
	9	1.67	0.25	0.14	0.08
	10	2.97	0.15	0.25	0.19
	11	2.97	0.15	0.25	0.19
	12	1.95	0.24	0.17	0.10
	13	0.92	0.23	0.08	0.04
	14	1.51	0.25	0.13	0.07
	15	1.55	0.25	0.13	0.07
	16	1.69	0.25	0.14	0.08
	17	2.10	0.23	0.18	0.11
	18	1.92	0.24	0.17	0.09
	19	2.71	0.17	0.23	0.16
	20	2.22	0.22	0.19	0.12
	21	2.59	0.19	0.23	0.15
	22	2.57	0.19	0.22	0.15
	23	3.46	0.11	0.25	0.23
	24	3.00	0.15	0.25	0.19
	25	2.75	0.17	0.24	0.17
	26	1.86	0.24	0.16	0.09
	27	3.32	0.12	0.25	0.22
WR	1	0.13	0.16	0.04	0.02
	2	-0.62	0.10	0.02	0.01
	3	1.65	0.25	0.14	0.08
	4	0.48	0.20	0.06	0.03
	5	1.93	0.24	0.17	0.10
	6	2.68	0.18	0.23	0.16
	7	0.14	0.17	0.04	0.02
	8	0.85	0.23	0.08	0.04
	9	2.19	0.22	0.19	0.12
	10	1.07	0.24	0.09	0.05
	11	0.57	0.21	0.06	0.03
	12	2.70	0.18	0.23	0.16
	13	2.59	0.19	0.22	0.15
	14	2.02	0.23	0.18	0.10
	15	1.72	0.25	0.15	0.08
	16	0.66	0.47	0.10	0.04
	17	0.93	0.53	0.14	0.06
	18	1.43	0.47	0.25	0.14
	19	2.08	0.80	0.66	0.48

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	0.30	0.21	0.06	0.03
	2	0.97	0.25	0.10	0.05
	3	0.23	0.20	0.05	0.03
	4	1.27	0.25	0.12	0.07
	5	2.06	0.21	0.20	0.13
	6	0.76	0.24	0.08	0.04
	7	0.93	0.25	0.10	0.05
	8	1.66	0.24	0.16	0.09
	9	1.15	0.25	0.11	0.06
	10	2.76	0.14	0.25	0.19
	11	2.93	0.13	0.25	0.21
	12	1.63	0.24	0.16	0.09
	13	2.51	0.16	0.23	0.17
	14	2.50	0.17	0.23	0.17
	15	2.90	0.13	0.25	0.20
	16	2.50	0.17	0.23	0.17
	17	2.34	0.18	0.22	0.15
	18	2.64	0.15	0.24	0.18
	19	1.51	0.24	0.15	0.08
	20	2.16	0.20	0.21	0.13
	21	2.67	0.15	0.24	0.18
	22	0.84	0.24	0.09	0.05
	23	0.25	0.20	0.05	0.03
	24	2.36	0.18	0.22	0.15
	25	2.07	0.21	0.20	0.13
SP	1	0.79	0.31	0.24	0.15
	2	0.65	0.32	0.22	0.13
	3	0.69	0.49	0.15	0.07
	4	0.95	0.38	0.23	0.14
	5	0.95	0.51	0.19	0.10
	6	1.35	0.77	0.52	0.36
	7	0.53	0.34	0.19	0.11
	8	0.90	0.35	0.24	0.15
	9	0.84	0.36	0.22	0.14
	10	0.42	0.33	0.19	0.11
	11	1.09	0.40	0.25	0.15
	12	1.04	0.46	0.22	0.12
	13	1.21	0.57	0.22	0.11
	14	0.86	0.45	0.20	0.11
	15	1.53	0.52	0.29	0.16
	16	0.63	0.43	0.17	0.09

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	0.26	0.18	0.04	0.02
	2	0.21	0.17	0.04	0.02
	3	0.52	0.20	0.05	0.03
	4	-0.45	0.11	0.02	0.01
	5	2.00	0.23	0.16	0.10
	6	2.64	0.18	0.22	0.16
	7	0.48	0.20	0.05	0.03
	8	2.60	0.19	0.22	0.15
	9	1.67	0.25	0.13	0.08
	10	2.97	0.15	0.24	0.19
	11	2.97	0.15	0.24	0.19
	12	1.95	0.24	0.16	0.10
	13	0.92	0.23	0.07	0.04
	14	1.51	0.25	0.12	0.07
	15	1.55	0.25	0.12	0.07
	16	1.69	0.25	0.13	0.08
	17	2.10	0.23	0.17	0.11
	18	1.92	0.24	0.15	0.09
	19	2.71	0.17	0.22	0.16
	20	2.22	0.22	0.18	0.12
	21	2.59	0.19	0.21	0.15
	22	2.57	0.19	0.21	0.15
	23	3.46	0.11	0.25	0.23
	24	3.00	0.15	0.24	0.19
	25	2.75	0.17	0.23	0.17
	26	1.86	0.24	0.15	0.09
	27	3.32	0.12	0.25	0.22
WR	1	0.13	0.16	0.04	0.02
	2	-0.62	0.10	0.02	0.01
	3	1.65	0.25	0.13	0.08
	4	0.48	0.20	0.05	0.03
	5	1.93	0.24	0.15	0.10
	6	2.68	0.18	0.22	0.16
	7	0.14	0.17	0.04	0.02
	8	0.85	0.23	0.07	0.04
	9	2.19	0.22	0.18	0.12
	10	1.07	0.24	0.08	0.05
	11	0.57	0.21	0.05	0.03
	12	2.70	0.18	0.22	0.16
	13	2.59	0.19	0.21	0.15
	14	2.02	0.23	0.16	0.10
	15	1.72	0.25	0.13	0.08
	16	0.66	0.47	0.09	0.04
	17	0.93	0.53	0.12	0.06
	18	1.43	0.47	0.23	0.14
	19	2.08	0.80	0.63	0.48

APPENDIX F: CONFIRMATORY FACTOR ANALYSIS FOR THE 2009 NYSESLAT

New York State regulations, under Federal Title III of the *No Child Left Behind* (NCLB) Act of 2001, require annual assessment of limited English proficient students using the 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 ELL students from Kindergarten through Grade 12.

The maximum number of points by modality and grade span for the 2009 administration is depicted in the following table:

Table 52 - Maximum Number of Points by Modality and Grade Span for the 2009 NYSESLAT

Maximum Number of Points in NYSESLAT Subtests							
Grade Span	Speaking	Listening	Reading	Writing			Total Number of Points
				Writing Conventions	Pre-Writing	Writing Prompt	
	CR	MC	MC	MC	SR	CR/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: a scale that is a combination of Speaking and Listening modalities and another that is a combination of the Reading and Writing modalities. In 2009, the New York State Education Department (NYSED) expressed 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 loadings based on the different combinations 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 preconceived 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 (English Language Proficiency Assessment)
- A two-factor model that tested 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 2009 data set, and 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 stated by Jöreskog and Sörbom (1995a), 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 polychoric correlation coefficients. With this in mind, polychoric 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 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 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 average 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(1996) 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 the one- and the two-factor models are shown in Figures 25 and 26:

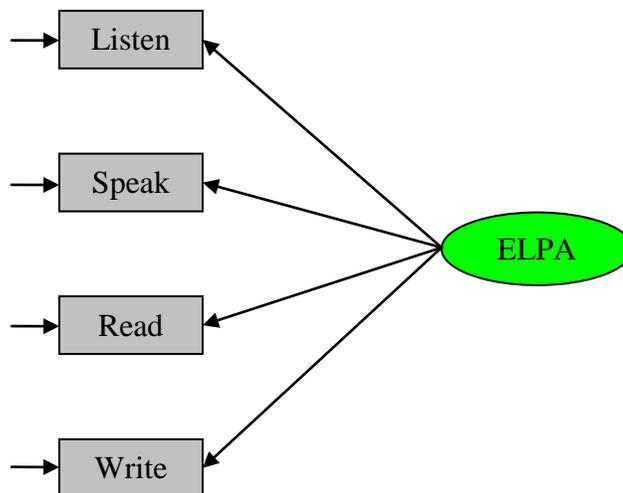


Figure 25—Path Diagram for the One-Factor Model

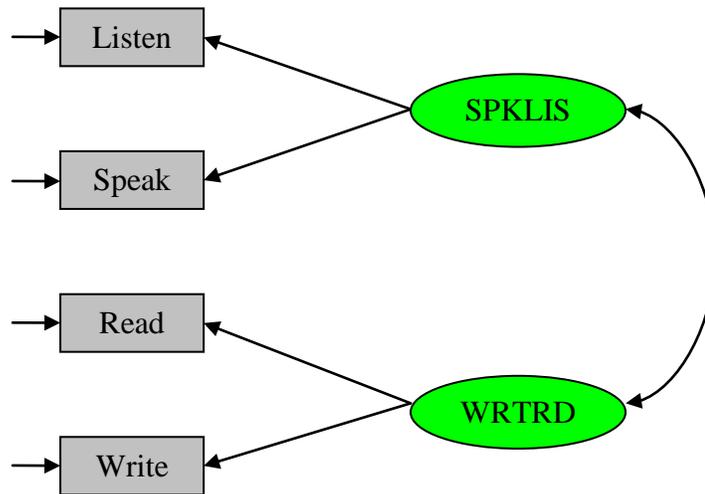


Figure 26—Path Diagram for the Two-Factor Model

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 Lisrel, version 8.80, computer program. The latent variable is standardized.

F.3: Results and Comments

Grade Span K–1

The summary statistics are provided in Table 53. Factor loadings for the one- and the two- factor models are provided in Table 54 with the fit indices for the two models displayed in Table 55.

Table 53—Summary Statistics

Variable	N-count	Mean	SD	Skewness	Kurtosis	Minimum	Maximum
Listen	53,211	20.37	3.82	-1.88	4.24	0.00	24.00
Speak	53,211	26.06	8.17	-1.51	1.92	0.00	34.00
Read	53,211	11.00	3.62	-0.71	-0.43	0.00	18.00
Write	53,211	18.12	5.23	-1.00	0.51	0.00	24.00

Table 54—Factor Loading Estimates for the One- and Two-Factor Models

Modalities	One-Factor Model	Two-Factor Model
	Λ	Λ_1 Λ_2
Listen	0.74	0.84 –
Speak	0.66	0.76 –
Read	0.84	– 0.85
Write	0.89	– 0.91
		Correlation 0.81

Table 55—Global Fit Indices for the One- and Two-Factor Models

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	7963.838	535.992
GFI	0.930	0.995
AGFI	0.652	0.975
RMR	0.057	0.025
RMSEA	0.274	0.071

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.

For the one-factor model, the values of GFI and AGFI are 0.933 and 0.663, respectively, and the RMR = 0.068 and the RMSEA = 0.269 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 Speaking and Listening on one factor (SPKLIS), and the Writing Conventions and Writing Constructed-Response and Reading on the second factor (WRTRD) was estimated. The results indicate a great improvement in fit of the model to the data. The rather weak loading of the Listening modality (0.68) is considerably improved with the two-factor model to 0.82. Although the p-value of the Chi-square is still in the range of significance at the 0.05 level of rejection ($\chi^2 = 512.309$, $p = 0.00$), the Chi-square itself is reduced tremendously. The RMR and RMSEA values have decreased a great deal to 0.028 and 0.070 respectively, and the GFI as well as the AGFI is approaching the peak of 1.000.

Grade Span 2–4

The summary statistics are provided in Table 56. Factor loadings for the one- and the two-factor models are provided in Table 57 with the fit indices for the two models displayed in Table 58.

Table 56—Summary Statistics

Variable	N-count	Mean	SD	Skewness	Kurtosis	Minimum	Maximum
Listen	64,510	19.16	4.18	-1.42	2.19	0.00	24.00
Speak	64,510	28.14	7.35	-2.03	4.15	0.00	34.00
Read	64,510	17.18	4.68	-0.88	0.29	0.00	26.00
Write	64,510	15.10	4.84	-0.90	0.13	0.00	24.00

Table 57—Factor Loading Estimates for the One- and Two-Factor Models

Modalities	One-Factor Model	Two-Factor Model	
	Λ	Λ_1	Λ_2
Listen	0.81	0.89	–
Speak	0.71	0.75	–
Read	0.85	–	0.84
Write	0.89	–	0.91
		Correlation	0.88

Table 58—Global Fit Indices for the One- and Two-Factor Models

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	5855.534	2107.898
GFI	0.957	0.984
AGFI	0.783	0.920
RMR	0.036	0.034
RMSEA	0.213	0.128

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.

The fit of the model to the data is very poor with the one-factor model. The Chi-square value is extremely large ($\chi^2 = 4309.044$ with 5 *df*, $p = 0.00$). The value of GFI is 0.966 with a poor value for the AGFI, i.e., 0.832. The RMR is acceptable at 0.046, while the RMSEA is large at 0.186. A summary of the fit indices suggest that an alternate model is justifiable.

With the two-factor model, the Chi-square reduces to 3263.943 (with 4 *df*, $p = 0.00$) and values for GFI and AGFI increased to 0.974 and 0.872 respectively. The increase in the GFI and AGFI to an the acceptable value suggest that the two-factor model is a substantial improvement over the one-factor model and is an acceptable model. However, the RMR and RMSEA decreased slightly. Although the two-factor model shows improvement over the one-factor model, the values of RMR = 0.085 and RMSEA = 0.162 indicate less than desirable fit.

Grade Span 5–6

The summary statistics are provided in Table 59. Factor loadings for the one- and the two-factor models are provided in Table 60 with the fit indices for the two models displayed in Table 61.

Table 59—Summary Statistics

Variable	N-count	Mean	SD	Skewness	Kurtosis	Minimum	Maximum
Listen	29,554	18.92	4.65	-1.33	1.44	0.00	25.00
Speak	29,554	27.59	8.34	-1.83	2.66	0.00	34.00
Read	29,554	18.98	6.02	-0.75	-0.44	0.00	27.00
Write	29,554	17.22	5.36	-1.05	0.40	0.00	25.00

Table 60—Factor Loading Estimates for the One- and Two-Factor Models

Modalities	One-Factor Model		Two-Factor Model	
	Λ	Λ_1	Λ_2	
Listen	0.83	0.88	–	
Speak	0.78	0.81	–	
Read	0.84	–	0.79	
Write	0.91	–	0.93	
		Correlation		0.93

Table 61—Global Fit Indices for the One- and Two-Factor Models

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	1572.171	2500.706
GFI	0.974	0.959
AGFI	0.870	0.797
RMR	0.023	0.093
RMSEA	0.163	0.206

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.

There was not any improvement in the 5–6 grade span for the two-factor model when compared to the one-factor model ($\chi^2 = 1820.544$ with 4 *df*, *p* = 0.00 for the two-factor model). The RMR and the RMSEA also show an increase. Similarly, there is only a small improvement for the GFI and the AGFI, with the GFI showing a value of 0.969.

Grade Span 7–8

The summary statistics are provided in Table 62. Factor loadings for the one- and the two-factor models are provided in Table 63 with the fit indices for the two models displayed in Table 64.

Table 62—Summary Statistics

Variable	N-count	Mean	SD	Skewness	Kurtosis	Minimum	Maximum
Listen	25,465	17.38	5.10	-0.74	-0.22	0.00	25.00
Speak	25,465	26.67	9.26	-1.45	1.14	0.00	34.00
Read	25,465	17.41	6.23	-0.35	-0.98	0.00	27.00
Write	25,465	16.76	5.74	-0.84	-0.20	0.00	25.00

Table 63—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	0.87	–	
Speak	0.79	0.80	–	
Read	0.82	–	0.78	
Write	0.91	–	0.91	
		Correlation		0.98

Table 64 - Global Fit Indices for the One- and Two-Factor Models

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	1661.807	2243.666
GFI	0.968	0.958
AGFI	0.842	0.789
RMR	0.025	0.062
RMSEA	0.181	0.210

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.

There was slight improvement in the 7–8 grade span for the two-factor model when compared to the one-factor model. The Chi-square sees an increase from the one-factor value of 1510.399 with 5 *df* to 998.103 with 4 degrees of freedom. Similarly, GFI, AGFI, RMR, and RMSEA all indicate slightly improved fit over their counterparts in the one-factor model.

Grade Span 9–12

The summary statistics are provided in Table 65. Factor loadings for the one- and the two-factor models are provided in Table 66 with the fit indices for the two models displayed in Table 67.

Table 65—Summary Statistics

Variable	N-count	Mean	SD	Skewness	Kurtosis	Minimum	Maximum
Listen	40,981	17.54	5.41	-0.73	-0.38	0.00	25.00
Speak	40,981	26.84	8.43	-1.45	1.46	0.00	34.00
Read	40,981	16.99	5.38	-0.26	-0.76	0.00	27.00
Write	40,981	17.73	5.44	-0.93	0.11	0.00	25.00

Table 66—Factor Loading Estimates for the One- and Two-Factor Models

Modalities	One-Factor Model		Two-Factor Model	
	Λ	Λ_1	Λ_1	Λ_2
Listen	0.85	0.84	–	–
Speak	0.79	0.79	–	–
Read	0.78	–	–	0.80
Write	0.88	–	–	0.89
Correlation				0.99

Table 67—Global Fit Indices for the One- and Two-Factor Models

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	2325.567	2606.121
GFI	0.972	0.969
AGFI	0.862	0.846
SRMR	0.026	0.042
RMSEA	0.168	0.178

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.

Test results for the 9–12 grade span showed slight improvement with the two-factor model over the one-factor model. The factor loadings increase with the use of the two-factor model in comparison to their loadings on the one-factor model. The Chi-square decreases ($\chi^2 = 2082.209$ with 5 *df* for the one-factor model and $\chi^2 = 1448.785$ with 4 *df* for the two-factor model). The other fit indices all indicated slightly poorer fit over the one-factor model.

F.4: Conclusion

Overall, the 2009 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), especially with lower grade spans. With higher grade spans, even though the one-factor model produced slightly better model fit than the two-factor model, the difference was not as noticeable as the difference observed with the lower grade span.

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 the 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 were very close to 1.0.

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

With respect to the analysis of the 5–6, 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 two-factor model does not show any improvement over the fit of the one-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 towards content and substantive issues in selecting items for the future.

APPENDIX G: SCALE SCORE SUMMARY BY SUBGROUPS**G.1a: Kindergarten Listening and Speaking**

Grade	Test	Group	N-count	Mean	Median	SD	IQR
K	LS	Female	12,086	599.25	601	54.45	67
K	LS	Male	13,346	593.34	596	54.18	62
K	LS	American Indian or Alaska Native	71	597.27	601	52.76	74
K	LS	Black or African American	1,038	599.02	596	52.03	55
K	LS	Asian	5,373	588.04	591	58.33	72
K	LS	Native Hawaiian/Other Pacific Islander	115	587.38	596	49.29	64
K	LS	Hispanic or Latino	16,415	597.23	596	51.72	59
K	LS	White	2,387	606.01	606	61.53	73
K	LS	Multiracial (not Hispanic origin)	33	603.00	606	58.27	51
K	LS	NYC	15,837	595.01	596	55.37	62
K	LS	Big 4 Cities	1,084	577.75	578	53.77	64
K	LS	High Need Urban/Suburban	3,382	596.36	596	54.15	59
K	LS	High Need Rural	195	596.12	596	49.85	55
K	LS	Average Need	3,042	603.85	601	50.88	59
K	LS	Low Need	1,307	599.34	601	51.73	51
K	LS	Charter Schools	575	612.58	606	42.52	47
K	LS	0 Years LEP	15,529	594.40	596	55.09	62
K	LS	1 Years LEP	9,408	599.10	601	53.29	55
K	LS	Arabic	657	586.29	587	61.75	75
K	LS	Bengali	692	604.42	606	54.64	59
K	LS	Chinese	1,414	586.07	587	53.64	64
K	LS	English	1,739	603.72	601	49.21	59
K	LS	Spanish	15,159	596.77	596	51.93	59
K	LS	Other Language	5,771	594.84	596	60.54	74
K	LS	Students with Disabilities	2,721	578.62	578	49.03	64

G.1b: Kindergarten Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
K	RW	Female	12,086	558.56	557	51.97	56
K	RW	Male	13,346	549.74	546	51.48	58
K	RW	American Indian or Alaska Native	71	553.77	551	43.07	60
K	RW	Black or African American	1,038	544.39	541	53.07	63
K	RW	Asian	5,373	570.40	569	59.60	68
K	RW	Native Hawaiian/Other Pacific Islander	115	558.20	546	61.21	78
K	RW	Hispanic or Latino	16,415	548.19	546	46.95	58
K	RW	White	2,387	560.25	557	55.92	64
K	RW	Multiracial (not Hispanic origin)	33	554.06	557	55.99	63
K	RW	NYC	15,837	557.99	551	54.03	68
K	RW	Big 4 Cities	1,084	529.27	526	46.45	48
K	RW	High Need Urban/Suburban	3,382	545.38	546	48.03	52
K	RW	High Need Rural	195	536.82	536	40.88	51
K	RW	Average Need	3,042	549.22	546	44.29	53
K	RW	Low Need	1,307	553.86	551	49.80	60
K	RW	Charter Schools	575	570.18	569	44.77	58
K	RW	0 Years LEP	15,529	557.28	551	53.57	68
K	RW	1 Years LEP	9,408	549.11	546	48.80	58
K	RW	Arabic	657	546.98	541	56.33	67
K	RW	Bengali	692	581.60	575	59.08	69
K	RW	Chinese	1,414	576.75	575	56.27	69
K	RW	English	1,739	549.19	546	46.70	53
K	RW	Spanish	15,159	548.44	546	47.19	58
K	RW	Other Language	5,771	561.66	557	58.65	64
K	RW	Students with Disabilities	2,721	533.39	531	48.55	60

G.2a: Grade 1 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
1	LS	Female	12,035	640.68	644	60.21	66
1	LS	Male	13,633	634.35	634	59.51	71
1	LS	American Indian or Alaska Native	48	635.40	639	61.07	71
1	LS	Black or African American	1,185	630.91	634	58.99	76
1	LS	Asian	5,464	634.42	634	63.58	76
1	LS	Native Hawaiian/Other Pacific Islander	86	620.48	615	59.55	78
1	LS	Hispanic or Latino	16,600	637.11	634	57.82	66
1	LS	White	2,254	650.07	656	64.54	87
1	LS	Multiracial (not Hispanic origin)	31	626.74	626	61.56	60
1	LS	NYC	16,162	635.56	634	60.44	71
1	LS	Big 4 Cities	1,247	617.10	618	60.02	69
1	LS	High Need Urban/Suburban	3,279	633.74	634	57.82	71
1	LS	High Need Rural	177	639.73	634	58.34	76
1	LS	Average Need	2,935	652.14	644	56.90	81
1	LS	Low Need	1,373	648.65	644	60.59	87
1	LS	Charter Schools	482	648.85	644	46.53	54
1	LS	0 Years LEP	14,363	634.10	634	60.79	71
1	LS	1 Years LEP	3,664	626.12	626	65.98	85
1	LS	2 Years LEP	7,149	649.40	644	52.43	60
1	LS	Arabic	702	623.37	626	69.17	89
1	LS	Bengali	715	638.95	644	63.75	66
1	LS	Chinese	1,230	636.68	634	60.74	71
1	LS	English	1,278	648.12	644	54.31	60
1	LS	Spanish	15,724	636.45	634	57.88	71
1	LS	Other Language	6,019	638.85	644	63.96	71
1	LS	Students with Disabilities	3,456	621.13	618	53.51	69

G.2b: Grade 1 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
1	RW	Female	12,035	631.59	640	63.00	77
1	RW	Male	13,633	618.47	622	64.31	92
1	RW	American Indian or Alaska Native	48	618.69	622	63.26	71
1	RW	Black or African American	1,185	611.56	610	67.00	98
1	RW	Asian	5,464	643.96	640	63.53	119
1	RW	Native Hawaiian/Other Pacific Islander	86	626.19	640	64.90	77
1	RW	Hispanic or Latino	16,600	618.26	622	62.26	92
1	RW	White	2,254	631.72	640	65.94	77
1	RW	Multiracial (not Hispanic origin)	31	612.87	610	69.32	98
1	RW	NYC	16,162	624.92	622	64.93	85
1	RW	Big 4 Cities	1,247	596.19	590	64.59	89
1	RW	High Need Urban/Suburban	3,279	620.21	622	62.07	92
1	RW	High Need Rural	177	619.85	610	58.10	85
1	RW	Average Need	2,935	632.18	622	59.31	77
1	RW	Low Need	1,373	637.51	640	62.80	77
1	RW	Charter Schools	482	638.20	640	53.05	68
1	RW	0 Years LEP	14,363	623.25	622	64.89	85
1	RW	1 Years LEP	3,664	613.28	610	68.02	98
1	RW	2 Years LEP	7,149	634.06	640	58.55	77
1	RW	Arabic	702	610.05	610	69.22	104
1	RW	Bengali	715	648.10	667	64.61	108
1	RW	Chinese	1,230	652.65	667	58.33	96
1	RW	English	1,278	629.72	622	58.74	77
1	RW	Spanish	15,724	617.65	610	62.44	92
1	RW	Other Language	6,019	634.96	640	65.79	128
1	RW	Students with Disabilities	3,456	595.18	590	62.40	89

G.3a: Grade 2 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
2	LS	Female	9,676	663.38	667	49.42	52
2	LS	Male	11,337	660.38	662	49.28	48
2	LS	American Indian or Alaska Native	34	663.44	664.5	60.48	72
2	LS	Black or African American	1,188	660.46	662	49.83	52
2	LS	Asian	4,060	660.80	667	55.22	60
2	LS	Native Hawaiian/Other Pacific Islander	61	644.41	649	64.84	92
2	LS	Hispanic or Latino	13,939	661.08	662	46.63	48
2	LS	White	1,709	671.14	673	53.80	57
2	LS	Multiracial (not Hispanic origin)	22	656.77	670	59.59	79
2	LS	NYC	13,426	658.17	662	49.31	52
2	LS	Big 4 Cities	1,171	649.55	653	52.88	59
2	LS	High Need Urban/Suburban	2,738	663.85	667	44.92	52
2	LS	High Need Rural	136	673.91	673	43.47	53
2	LS	Average Need	2,248	675.41	673	48.06	53
2	LS	Low Need	976	682.49	679	51.26	61
2	LS	Charter Schools	311	674.94	673	39.74	53
2	LS	0 Years LEP	11,761	657.24	662	50.01	52
2	LS	1 Years LEP	2,294	644.45	649	56.82	65
2	LS	2 Years LEP	2,263	670.07	667	42.92	48
2	LS	3 Years LEP	4,233	678.24	673	40.92	49
2	LS	Arabic	660	644.43	653	57.20	65
2	LS	Bengali	579	667.49	673	50.85	52
2	LS	Chinese	749	665.78	667	51.23	52
2	LS	English	843	677.63	673	48.03	53
2	LS	Spanish	13,341	660.39	662	46.45	48
2	LS	Other Language	4,841	663.81	667	54.52	56
2	LS	Students with Disabilities	3,595	650.50	653	40.36	47

G.3 b: Grade 2 Reading and Writing

Grade	Test	Group	N- count	Mean	Median	SD	IQR
2	RW	Female	9,676	618.60	621	45.78	56
2	RW	Male	11,337	612.87	616	46.19	64
2	RW	American Indian or Alaska Native	34	615.06	616	49.62	93
2	RW	Black or African American	1,188	607.99	607	47.50	63
2	RW	Asian	4,060	626.03	630	49.95	65
2	RW	Native Hawaiian/Other Pacific Islander	61	608.92	616	51.80	97
2	RW	Hispanic or Latino	13,939	612.05	616	43.55	55
2	RW	White	1,709	624.22	625	50.14	69
2	RW	Multiracial (not Hispanic origin)	22	613.36	621	46.56	70
2	RW	NYC	13,426	613.05	616	45.02	64
2	RW	Big 4 Cities	1,171	597.97	599	49.71	66
2	RW	High Need Urban/Suburban	2,738	616.99	621	43.87	56
2	RW	High Need Rural	136	618.26	618.5	45.42	62
2	RW	Average Need	2,248	626.19	630	46.33	53
2	RW	Low Need	976	637.20	641	51.47	54
2	RW	Charter Schools	311	628.29	630	34.30	39
2	RW	0 Years LEP	11,761	611.86	616	44.65	59
2	RW	1 Years LEP	2,294	604.08	607	52.50	77
2	RW	2 Years LEP	2,263	622.09	625	46.68	62
2	RW	3 Years LEP	4,233	628.34	630	42.43	49
2	RW	Arabic	660	599.17	599	49.81	75
2	RW	Bengali	579	629.36	635	47.65	67
2	RW	Chinese	749	632.97	635	46.73	52
2	RW	English	843	628.13	630	45.03	53
2	RW	Spanish	13,341	611.49	616	43.43	59
2	RW	Other Language	4,841	622.25	625	50.01	62
2	RW	Students with Disabilities	3,595	593.83	594	41.11	57

G.4a: Grade 3 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
3	LS	Female	9,172	682.30	685	54.59	57
3	LS	Male	10,968	678.27	679	53.83	61
3	LS	American Indian or Alaska Native	37	678.51	685	52.10	49
3	LS	Black or African American	1,169	683.17	685	53.50	61
3	LS	Asian	3,677	675.68	685	60.92	69
3	LS	Native Hawaiian/Other Pacific Islander	43	666.14	679	74.68	104
3	LS	Hispanic or Latino	13,725	680.49	685	51.75	57
3	LS	White	1,470	685.48	693	57.97	57
3	LS	Multiracial (not Hispanic origin)	19	688.37	702	47.38	41
3	LS	NYC	13,290	678.01	685	54.51	61
3	LS	Big 4 Cities	1,179	671.19	673	55.82	57
3	LS	High Need Urban/Suburban	2,480	683.44	685	50.81	57
3	LS	High Need Rural	117	677.73	685	62.37	73
3	LS	Average Need	1,977	688.59	693	54.08	52
3	LS	Low Need	848	692.25	693	52.05	52
3	LS	Charter Schools	243	693.65	693	49.93	57
3	LS	0 Years LEP	11,953	678.73	685	54.65	61
3	LS	1 Years LEP	2,062	650.20	657	62.98	83
3	LS	2 Years LEP	1,347	681.95	679	50.22	61
3	LS	3 Years LEP	1,089	693.65	693	43.63	47
3	LS	4 Years LEP	3,242	697.64	693	41.86	41
3	LS	Arabic	601	666.64	673	61.76	69
3	LS	Bengali	555	682.63	685	54.36	65
3	LS	Chinese	691	683.75	693	61.35	73
3	LS	English	758	689.12	693	51.90	52
3	LS	Spanish	13,158	680.02	685	51.49	57
3	LS	Other Language	4,377	679.73	685	59.62	65
3	LS	Students with Disabilities	4,112	674.32	673	42.84	53

G.4b: Grade 3 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
3	RW	Female	9,172	644.97	652	47.70	53
3	RW	Male	10,968	637.44	646	47.67	54
3	RW	American Indian or Alaska Native	37	643.81	652	42.77	41
3	RW	Black or African American	1,169	634.84	641	50.82	59
3	RW	Asian	3,677	646.27	652	53.24	67
3	RW	Native Hawaiian/Other Pacific Islander	43	628.30	635	55.55	96
3	RW	Hispanic or Latino	13,725	639.43	646	45.40	50
3	RW	White	1,470	645.68	652	51.42	67
3	RW	Multiracial (not Hispanic origin)	19	655.37	666	47.00	37
3	RW	NYC	13,290	639.69	646	47.51	62
3	RW	Big 4 Cities	1,179	627.56	635	52.00	72
3	RW	High Need Urban/Suburban	2,480	643.48	652	45.46	53
3	RW	High Need Rural	117	632.16	641	56.22	71
3	RW	Average Need	1,977	646.07	652	46.85	53
3	RW	Low Need	848	655.69	659	49.77	53
3	RW	Charter Schools	243	652.72	652	42.37	39
3	RW	0 Years LEP	11,953	639.86	646	46.71	58
3	RW	1 Years LEP	2,062	619.86	625	57.70	81
3	RW	2 Years LEP	1,347	642.33	646	51.75	62
3	RW	3 Years LEP	1,089	648.85	652	44.01	49
3	RW	4 Years LEP	3,242	654.47	659	39.27	39
3	RW	Arabic	601	626.91	635	57.35	76
3	RW	Bengali	555	651.22	659	47.24	58
3	RW	Chinese	691	654.44	659	51.48	53
3	RW	English	758	645.60	652	44.51	53
3	RW	Spanish	13,158	639.05	646	45.62	50
3	RW	Other Language	4,377	643.98	652	51.79	58
3	RW	Students with Disabilities	4,112	622.25	625	42.60	58

G.5a: Grade 4 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
4	LS	Female	8,416	693.78	702	59.66	63
4	LS	Male	9,882	691.01	693	59.12	68
4	LS	American Indian or Alaska Native	31	698.39	702	63.24	102
4	LS	Black or African American	1,128	692.40	693	58.47	73
4	LS	Asian	3,281	684.62	693	65.25	81
4	LS	Native Hawaiian/Other Pacific Islander	48	663.92	673	57.34	67.5
4	LS	Hispanic or Latino	12,444	693.70	693	57.14	63
4	LS	White	1,352	698.52	702	63.55	65.5
4	LS	Multiracial (not Hispanic origin)	14	702.14	693.5	62.54	93
4	LS	NYC	12,320	690.34	693	59.78	68
4	LS	Big 4 Cities	1,174	680.29	685	60.00	65
4	LS	High Need Urban/Suburban	2,169	696.11	702	56.58	63
4	LS	High Need Rural	109	705.95	702	52.58	88
4	LS	Average Need	1,673	702.52	702	59.89	57
4	LS	Low Need	697	706.97	702	56.03	76
4	LS	Charter Schools	152	698.28	697.5	40.57	57
4	LS	0 Years LEP	10,911	692.12	693	60.18	68
4	LS	1 Years LEP	1,989	652.11	653	64.05	79
4	LS	2 Years LEP	1,229	689.17	685	53.20	61
4	LS	3 Years LEP	592	707.53	702	48.22	51
4	LS	4 Years LEP	804	712.17	702	44.93	51
4	LS	5 Years LEP	2,400	714.93	714	42.95	45
4	LS	Arabic	614	671.95	679	67.91	81
4	LS	Bengali	508	697.31	702	57.17	68
4	LS	Chinese	543	682.81	685	65.04	85
4	LS	English	608	708.55	702	54.67	51
4	LS	Spanish	11,960	693.05	693	57.09	63
4	LS	Other Language	4,065	691.29	693	63.65	73
4	LS	Students with Disabilities	4,170	693.78	693	44.95	47

G.5b: Grade 4 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
4	RW	Female	8,416	661.16	666	50.82	59
4	RW	Male	9,882	653.54	659	52.63	58
4	RW	American Indian or Alaska Native	31	654.68	659	41.33	59
4	RW	Black or African American	1,128	647.92	652	55.34	62
4	RW	Asian	3,281	661.29	666	57.65	69
4	RW	Native Hawaiian/Other Pacific Islander	48	643.65	652	50.50	73.5
4	RW	Hispanic or Latino	12,444	656.11	666	49.46	53
4	RW	White	1,352	663.39	666	55.42	59
4	RW	Multiracial (not Hispanic origin)	14	670.71	678.5	46.22	42
4	RW	NYC	12,320	655.85	659	51.02	53
4	RW	Big 4 Cities	1,174	640.57	646	54.65	67
4	RW	High Need Urban/Suburban	2,169	660.88	666	50.90	59
4	RW	High Need Rural	109	655.99	666	53.64	42
4	RW	Average Need	1,673	663.74	666	53.49	53
4	RW	Low Need	697	675.95	674	53.71	61
4	RW	Charter Schools	152	666.99	674	49.04	48
4	RW	0 Years LEP	10,911	656.69	666	50.43	53
4	RW	1 Years LEP	1,989	629.71	630	60.58	88
4	RW	2 Years LEP	1,229	658.93	659	51.18	64
4	RW	3 Years LEP	592	663.09	666	51.79	59
4	RW	4 Years LEP	804	667.70	666	50.10	53
4	RW	5 Years LEP	2,400	674.79	674	43.11	42
4	RW	Arabic	614	636.55	646	59.59	80
4	RW	Bengali	508	669.55	674	49.54	66
4	RW	Chinese	543	661.74	674	56.21	69
4	RW	English	608	664.46	666	53.20	53
4	RW	Spanish	11,960	655.78	659	49.16	53
4	RW	Other Language	4,065	660.58	666	56.91	64
4	RW	Students with Disabilities	4,170	643.78	646	45.40	53

G.6a: Grade 5 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
5	LS	Female	7,002	677.56	682	50.59	54
5	LS	Male	8,831	677.77	682	50.28	58
5	LS	American Indian or Alaska Native	21	683.86	695	65.69	52
5	LS	Black or African American	1,063	675.27	682	45.93	49
5	LS	Asian	2,843	668.99	677	55.96	78
5	LS	Native Hawaiian/Other Pacific Islander	54	665.56	667	47.44	75
5	LS	Hispanic or Latino	10,717	679.94	688	48.83	54
5	LS	White	1,122	680.53	688	51.76	58
5	LS	Multiracial (not Hispanic origin)	13	703.69	703	51.00	41
5	LS	NYC	10,694	676.01	682	51.07	58
5	LS	Big 4 Cities	1,026	666.72	677	48.76	51
5	LS	High Need Urban/Suburban	1,802	681.50	688	50.68	50
5	LS	High Need Rural	119	685.20	688	50.49	50
5	LS	Average Need	1,335	687.01	688	46.60	45
5	LS	Low Need	568	688.58	695	48.42	50
5	LS	Charter Schools	280	687.03	688	35.59	32
5	LS	0 Years LEP	9,345	677.86	688	51.43	54
5	LS	1 Years LEP	1,918	646.39	651	53.94	71
5	LS	2 Years LEP	1,125	676.93	682	44.95	52
5	LS	3 Years LEP	532	688.99	688	41.53	45
5	LS	4 Years LEP	368	689.08	688	39.03	45
5	LS	5 Years LEP	610	697.35	695	40.32	35
5	LS	6 Years LEP or More	1,913	697.23	695	33.66	35
5	LS	Arabic	514	659.89	671	55.19	78
5	LS	Bengali	381	677.78	682	47.69	61
5	LS	Chinese	461	663.04	671	59.34	78
5	LS	English	478	688.58	695	47.75	41
5	LS	Spanish	10,350	679.56	688	48.87	54
5	LS	Other Language	3,649	675.25	682	52.38	65
5	LS	Students with Disabilities	3,899	683.23	682	36.27	41

G.6b: Grade 5 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
5	RW	Female	7,002	681.69	687	48.90	56
5	RW	Male	8,831	675.09	683	48.51	66
5	RW	American Indian or Alaska Native	21	676.48	687	54.31	63
5	RW	Black or African American	1,063	670.47	674	46.57	64
5	RW	Asian	2,843	677.06	683	56.20	81
5	RW	Native Hawaiian/Other Pacific Islander	54	669.74	668	49.68	76
5	RW	Hispanic or Latino	10,717	679.08	687	46.30	59
5	RW	White	1,122	677.41	683	52.82	66
5	RW	Multiracial (not Hispanic origin)	13	702.15	704	44.91	52
5	RW	NYC	10,694	676.56	683	48.67	66
5	RW	Big 4 Cities	1,026	660.98	670	49.51	71
5	RW	High Need Urban/Suburban	1,802	683.24	692	48.77	60
5	RW	High Need Rural	119	675.73	674	42.89	59
5	RW	Average Need	1,335	686.19	692	45.96	60
5	RW	Low Need	568	694.21	698	51.99	56
5	RW	Charter Schools	280	690.35	692	34.26	46
5	RW	0 Years LEP	9,345	678.17	683	47.96	59
5	RW	1 Years LEP	1,918	652.56	651	53.87	82
5	RW	2 Years LEP	1,125	678.15	683	49.75	66
5	RW	3 Years LEP	532	688.26	687	47.28	60
5	RW	4 Years LEP	368	685.77	687	45.87	52
5	RW	5 Years LEP	610	690.41	692	40.07	52
5	RW	6 Years LEP or More	1,913	694.45	698	38.99	44
5	RW	Arabic	514	654.21	660	55.82	92
5	RW	Bengali	381	683.63	687	52.26	82
5	RW	Chinese	461	673.79	674	57.01	91
5	RW	English	478	685.10	687	43.83	60
5	RW	Spanish	10,350	678.77	687	46.38	59
5	RW	Other Language	3,649	678.20	683	52.45	74
5	RW	Students with Disabilities	3,899	670.27	674	40.42	54

G.7a: Grade 6 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
6	LS	Female	5,885	675.50	682	52.25	61
6	LS	Male	7,783	676.74	682	53.19	61
6	LS	American Indian or Alaska Native	32	669.81	685	57.76	79.5
6	LS	Black or African American	1,008	675.24	682	48.09	52
6	LS	Asian	2,306	664.74	671	57.05	75
6	LS	Native Hawaiian/Other Pacific Islander	45	667.40	677	57.28	75
6	LS	Hispanic or Latino	9,335	679.00	688	51.50	54
6	LS	White	925	678.59	688	55.21	61
6	LS	Multiracial (not Hispanic origin)	17	659.12	658	47.83	51
6	LS	NYC	9,441	673.87	682	53.75	65
6	LS	Big 4 Cities	969	672.66	677	49.72	59
6	LS	High Need Urban/Suburban	1,458	681.31	688	50.96	50
6	LS	High Need Rural	95	691.51	695	50.75	65
6	LS	Average Need	1,067	683.53	688	51.18	50
6	LS	Low Need	424	688.00	695	48.59	56
6	LS	Charter Schools	211	694.89	688	35.92	35
6	LS	0 Years LEP	8,060	676.42	688	54.22	58
6	LS	1 Years LEP	1,847	646.00	647	54.08	71
6	LS	2 Years LEP	1,075	674.23	677	45.20	59
6	LS	3 Years LEP	420	686.08	688	39.07	54
6	LS	4 Years LEP	318	690.95	688	39.11	45
6	LS	5 Years LEP	310	696.82	695	40.60	35
6	LS	6 Years LEP or More	1,607	701.67	703	35.86	41
6	LS	Arabic	468	657.65	667	59.16	85.5
6	LS	Bengali	304	673.08	677	50.79	72
6	LS	Chinese	381	653.10	658	60.86	79
6	LS	English	351	686.54	688	44.09	45
6	LS	Spanish	9,033	678.76	688	51.68	58
6	LS	Other Language	3,131	673.58	682	53.53	72
6	LS	Students with Disabilities	3,461	687.87	688	37.24	45

G.7b: Grade 6 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
6	RW	Female	5,885	681.51	687	50.12	67
6	RW	Male	7,783	674.68	683	50.67	70
6	RW	American Indian or Alaska Native	32	682.78	692	49.79	50.5
6	RW	Black or African American	1,008	675.44	678	49.66	70
6	RW	Asian	2,306	674.06	678	56.35	88
6	RW	Native Hawaiian/Other Pacific Islander	45	667.60	678	48.22	80
6	RW	Hispanic or Latino	9,335	678.74	687	48.53	63
6	RW	White	925	678.03	683	55.70	78
6	RW	Multiracial (not Hispanic origin)	17	670.06	662	44.53	66
6	RW	NYC	9,441	675.26	683	50.24	70
6	RW	Big 4 Cities	969	669.25	674	47.43	67
6	RW	High Need Urban/Suburban	1,458	683.98	692	50.50	60
6	RW	High Need Rural	95	687.13	687	47.60	60
6	RW	Average Need	1,067	683.98	692	51.82	60
6	RW	Low Need	424	697.29	704	53.85	54
6	RW	Charter Schools	211	702.97	698	40.83	43
6	RW	0 Years LEP	8,060	677.33	683	49.58	66
6	RW	1 Years LEP	1,847	653.80	651	52.82	79
6	RW	2 Years LEP	1,075	678.47	683	51.45	78
6	RW	3 Years LEP	420	692.39	692	53.72	56
6	RW	4 Years LEP	318	691.63	695	43.62	48
6	RW	5 Years LEP	310	690.75	692	39.32	44
6	RW	6 Years LEP or More	1,607	696.70	698	42.00	40
6	RW	Arabic	468	655.84	662	55.70	85
6	RW	Bengali	304	679.66	683	57.13	81
6	RW	Chinese	381	668.39	670	56.89	77
6	RW	English	351	687.08	692	44.01	56
6	RW	Spanish	9,033	678.34	683	48.52	63
6	RW	Other Language	3,131	678.68	683	53.65	78
6	RW	Students with Disabilities	3,461	674.98	678	40.51	53

G.8a: Grade 7 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
7	LS	Female	5,643	692.64	702	53.55	64
7	LS	Male	6,989	692.37	702	53.11	61
7	LS	American Indian or Alaska Native	19	683.42	685	60.97	81
7	LS	Black or African American	931	692.80	697	46.42	61
7	LS	Asian	2,413	677.07	681	55.63	80
7	LS	Native Hawaiian/Other Pacific Islander	57	679.81	689	48.19	74
7	LS	Hispanic or Latino	8,304	696.41	707	52.37	59
7	LS	White	903	698.32	707	54.62	65
7	LS	Multiracial (not Hispanic origin)	5	696.40	693	65.11	36
7	LS	NYC	8,976	690.57	702	53.07	64
7	LS	Big 4 Cities	895	685.24	693	52.52	66
7	LS	High Need Urban/Suburban	1,197	694.08	707	56.19	61
7	LS	High Need Rural	79	703.87	702	52.26	64
7	LS	Average Need	887	702.91	712	51.70	52
7	LS	Low Need	458	709.69	712	50.16	57
7	LS	Charter Schools	133	722.27	718	36.70	35
7	LS	0 Years LEP	7,315	694.20	707	53.25	57
7	LS	1 Years LEP	2,101	663.37	661	54.19	78
7	LS	2 Years LEP	1,094	690.23	693	45.67	57
7	LS	3 Years LEP	408	706.81	707	46.72	48
7	LS	4 Years LEP	277	712.19	712	43.55	53
7	LS	5 Years LEP	245	719.97	718	36.63	45
7	LS	6 Years LEP or More	1,158	721.38	718	38.09	40
7	LS	Arabic	433	679.14	681	55.17	77
7	LS	Bengali	305	693.09	697	43.43	67
7	LS	Chinese	373	666.94	671	59.52	88
7	LS	English	314	706.59	712	48.96	44
7	LS	Spanish	8,071	695.99	707	52.56	54
7	LS	Other Language	3,136	686.89	693	53.75	73
7	LS	Students with Disabilities	3,127	707.87	707	37.37	36

G.8b: Grade 7 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
7	RW	Female	5,643	669.40	672	45.84	61
7	RW	Male	6,989	661.13	665	44.19	62
7	RW	American Indian or Alaska Native	19	659.74	672	49.54	85
7	RW	Black or African American	931	666.07	668	46.16	65
7	RW	Asian	2,413	659.39	658	49.52	72
7	RW	Native Hawaiian/Other Pacific Islander	57	663.18	661	40.88	55
7	RW	Hispanic or Latino	8,304	665.85	668	42.87	55
7	RW	White	903	668.76	672	50.70	70
7	RW	Multiracial (not Hispanic origin)	5	673.20	699	43.57	66
7	RW	NYC	8,976	663.83	665	44.05	62
7	RW	Big 4 Cities	895	651.93	654	47.91	63
7	RW	High Need Urban/Suburban	1,197	665.58	672	45.04	65
7	RW	High Need Rural	79	670.30	680	44.38	59
7	RW	Average Need	887	674.05	676	47.45	57
7	RW	Low Need	458	682.09	684	46.28	59
7	RW	Charter Schools	133	689.34	689	41.17	52
7	RW	0 Years LEP	7,315	665.62	668	43.72	59
7	RW	1 Years LEP	2,101	646.21	641	47.58	65
7	RW	2 Years LEP	1,094	665.77	665	45.90	68
7	RW	3 Years LEP	408	677.61	680	41.81	53
7	RW	4 Years LEP	277	680.58	680	43.50	52
7	RW	5 Years LEP	245	680.42	684	42.35	50
7	RW	6 Years LEP or More	1,158	681.37	684	38.09	43
7	RW	Arabic	433	649.34	647	51.05	73
7	RW	Bengali	305	670.99	668	39.94	58
7	RW	Chinese	373	654.90	651	53.23	82
7	RW	English	314	673.90	676	45.12	60
7	RW	Spanish	8,071	665.56	668	42.75	55
7	RW	Other Language	3,136	664.74	665	48.85	68
7	RW	Students with Disabilities	3,127	664.06	665	36.72	48

G.9a: Grade 8 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
8	LS	Female	5,576	695.19	702	56.79	72
8	LS	Male	7,086	696.37	707	56.72	72
8	LS	American Indian or Alaska Native	17	655.88	668	48.92	44
8	LS	Black or African American	991	696.03	702	51.86	65
8	LS	Asian	2,680	681.82	681	57.70	77
8	LS	Native Hawaiian/Other Pacific Islander	45	688.84	685	58.04	70
8	LS	Hispanic or Latino	8,042	699.88	712	56.37	62
8	LS	White	874	702.97	707	55.45	74
8	LS	Multiracial (not Hispanic origin)	13	680.08	681	57.06	98
8	LS	NYC	9,264	694.97	702	55.99	72
8	LS	Big 4 Cities	870	684.11	693	59.79	81
8	LS	High Need Urban/Suburban	1,176	696.49	707	58.71	75
8	LS	High Need Rural	64	710.44	712	53.11	61
8	LS	Average Need	816	706.02	712	58.49	71
8	LS	Low Need	391	710.03	712	52.06	61
8	LS	Charter Schools	81	731.00	733	45.23	47
8	LS	0 Years LEP	7,382	699.05	707	56.27	62
8	LS	1 Years LEP	2,161	663.57	664	55.04	70
8	LS	2 Years LEP	1,235	692.42	693	52.00	67
8	LS	3 Years LEP	442	705.49	707	50.07	59
8	LS	4 Years LEP	255	725.26	725	46.49	57
8	LS	5 Years LEP	232	729.22	725	39.54	47
8	LS	6 Years LEP or More	926	729.66	733	39.72	47
8	LS	Arabic	434	689.44	691	59.52	86
8	LS	Bengali	269	699.39	702	49.06	65
8	LS	Chinese	490	671.03	671	56.92	82
8	LS	English	270	708.91	715	53.83	64
8	LS	Spanish	7,843	699.50	707	56.41	62
8	LS	Other Language	3,356	690.45	693	56.51	73
8	LS	Students with Disabilities	2,724	713.00	712	39.31	40

G.9b: Grade 8 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
8	RW	Female	5,576	674.91	680	48.71	69
8	RW	Male	7,086	667.25	668	46.07	65
8	RW	American Indian or Alaska Native	17	653.00	638	40.24	66
8	RW	Black or African American	991	671.99	676	46.27	66
8	RW	Asian	2,680	666.38	668	50.42	68
8	RW	Native Hawaiian/Other Pacific Islander	45	673.09	672	46.08	63
8	RW	Hispanic or Latino	8,042	671.05	676	45.81	63
8	RW	White	874	678.36	680	52.15	69
8	RW	Multiracial (not Hispanic origin)	13	670.08	684	61.26	93
8	RW	NYC	9,264	670.18	672	45.82	66
8	RW	Big 4 Cities	870	654.16	654	52.43	72
8	RW	High Need Urban/Suburban	1,176	671.92	680	49.61	66
8	RW	High Need Rural	64	672.52	680	53.43	69
8	RW	Average Need	816	680.52	684	50.17	63
8	RW	Low Need	391	688.87	693	46.66	56
8	RW	Charter Schools	81	690.35	693	44.52	56
8	RW	0 Years LEP	7,382	672.65	676	45.40	63
8	RW	1 Years LEP	2,161	649.37	644	47.42	65
8	RW	2 Years LEP	1,235	669.20	668	47.72	70
8	RW	3 Years LEP	442	680.07	684	47.03	59
8	RW	4 Years LEP	255	689.57	693	48.99	67
8	RW	5 Years LEP	232	694.44	699	41.41	45
8	RW	6 Years LEP or More	926	690.47	693	44.93	49
8	RW	Arabic	434	661.74	661	52.90	74
8	RW	Bengali	269	681.37	680	44.73	59
8	RW	Chinese	490	658.99	658	48.47	68
8	RW	English	270	679.30	684	47.68	59
8	RW	Spanish	7,843	670.72	676	45.70	66
8	RW	Other Language	3,356	671.69	672	50.08	70
8	RW	Students with Disabilities	2,724	670.00	672	39.28	52

G.10a: Grade 9 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
9	LS	Female	6,369	689.74	694	59.90	84
9	LS	Male	8,271	692.06	698	62.80	84
9	LS	American Indian or Alaska Native	29	676.79	680	54.71	83
9	LS	Black or African American	1,409	692.99	698	55.09	66
9	LS	Asian	3,220	677.59	677	57.02	75
9	LS	Native Hawaiian/Other Pacific Islander	35	675.31	671	48.15	54
9	LS	Hispanic or Latino	8,910	694.29	706	63.78	91
9	LS	White	1,026	703.38	706	58.06	77
9	LS	Multiracial (not Hispanic origin)	11	699.91	694	53.12	70
9	LS	NYC	10,213	690.92	698	60.73	84
9	LS	Big 4 Cities	1,097	687.00	687	70.26	77
9	LS	High Need Urban/Suburban	1,571	689.40	694	62.77	87
9	LS	High Need Rural	63	707.13	716	48.50	70
9	LS	Average Need	1,077	690.33	694	62.04	87
9	LS	Low Need	474	698.17	702	55.07	66
9	LS	Charter Schools	142	726.33	727	37.37	44
9	LS	0 Years LEP	8,125	693.87	702	62.71	91
9	LS	1 Years LEP	3,078	664.24	664	54.10	69
9	LS	2 Years LEP	1,563	689.42	691	52.54	65
9	LS	3 Years LEP	516	712.73	716	49.72	61
9	LS	4 Years LEP	283	714.47	716	51.61	54
9	LS	5 Years LEP	227	728.20	727	52.09	44
9	LS	6 Years LEP or More	811	735.81	741	59.19	45
9	LS	Arabic	536	686.34	687	54.42	69.5
9	LS	Bengali	396	685.61	684	52.93	71
9	LS	Chinese	490	673.01	675.5	57.51	76
9	LS	English	403	694.44	698	59.23	78
9	LS	Spanish	8,648	694.57	706	63.71	88
9	LS	Other Language	4,167	686.65	687	58.57	77
9	LS	Students with Disabilities	2,212	722.65	727	47.90	48

G.10b: Grade 9 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
9	RW	Female	6,369	680.63	682	43.51	57
9	RW	Male	8,271	678.67	678	45.61	64
9	RW	American Indian or Alaska Native	29	674.17	678	36.51	60
9	RW	Black or African American	1,409	678.78	678	44.20	57
9	RW	Asian	3,220	676.99	675	44.59	60
9	RW	Native Hawaiian/Other Pacific Islander	35	672.91	665	43.36	48
9	RW	Hispanic or Latino	8,910	679.82	682	44.83	61
9	RW	White	1,026	686.42	685	44.44	63
9	RW	Multiracial (not Hispanic origin)	11	671.36	675	40.75	77
9	RW	NYC	10,213	681.20	682	42.15	61
9	RW	Big 4 Cities	1,097	665.01	665	59.82	66
9	RW	High Need Urban/Suburban	1,571	675.04	678	45.70	60
9	RW	High Need Rural	63	681.68	682	40.08	61
9	RW	Average Need	1,077	677.22	678	45.67	68
9	RW	Low Need	474	691.84	693	45.27	61
9	RW	Charter Schools	142	697.12	697	32.29	40
9	RW	0 Years LEP	8,125	682.51	685	42.56	62
9	RW	1 Years LEP	3,078	664.94	662	41.50	58
9	RW	2 Years LEP	1,563	677.49	678	44.98	57
9	RW	3 Years LEP	516	688.58	689	48.18	53
9	RW	4 Years LEP	283	689.01	693	46.70	53
9	RW	5 Years LEP	227	695.73	697	42.71	48
9	RW	6 Years LEP or More	811	695.95	705	57.52	41
9	RW	Arabic	536	667.69	665	42.18	55
9	RW	Bengali	396	678.16	678	40.47	57
9	RW	Chinese	490	679.47	678	44.39	61
9	RW	English	403	680.95	682	42.53	61
9	RW	Spanish	8,648	680.03	682	44.76	61
9	RW	Other Language	4,167	680.00	678	45.39	61
9	RW	Students with Disabilities	2,212	687.56	689	40.04	41

G.11a: Grade 10 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
10	LS	Female	6,099	697.20	698	52.26	73
10	LS	Male	7,177	699.15	702	55.16	70
10	LS	American Indian or Alaska Native	24	706.96	713.5	52.12	45
10	LS	Black or African American	1,206	707.79	706	44.10	54
10	LS	Asian	3,806	682.43	680	48.60	63
10	LS	Native Hawaiian/Other Pacific Islander	41	688.00	694	48.76	60
10	LS	Hispanic or Latino	7,379	703.53	706	56.46	73
10	LS	White	810	710.35	711	50.51	61
10	LS	Multiracial (not Hispanic origin)	10	719.30	716	57.92	54
10	LS	NYC	9,946	695.83	698	53.26	73
10	LS	Big 4 Cities	676	694.79	698	63.43	73
10	LS	High Need Urban/Suburban	1,151	703.38	706	54.67	70
10	LS	High Need Rural	66	714.88	716	46.74	59
10	LS	Average Need	872	710.45	716	52.69	61
10	LS	Low Need	489	710.39	711	46.18	57
10	LS	Charter Schools	75	736.29	741	39.18	50
10	LS	0 Years LEP	6,904	702.58	706	55.25	73
10	LS	1 Years LEP	3,172	678.58	677	47.70	56
10	LS	2 Years LEP	1,629	694.75	694	45.43	57
10	LS	3 Years LEP	461	708.20	706	42.51	54
10	LS	4 Years LEP	305	719.28	721	46.65	59
10	LS	5 Years LEP	193	727.99	727	58.45	59
10	LS	6 Years LEP or More	581	736.11	741	58.33	40
10	LS	Arabic	342	701.92	702	41.76	53
10	LS	Bengali	396	695.35	691	43.76	59
10	LS	Chinese	785	674.82	674	47.47	62
10	LS	English	318	713.37	721	52.82	66
10	LS	Spanish	7,208	703.20	706	56.42	73
10	LS	Other Language	4,227	693.02	694	50.43	66
10	LS	Students with Disabilities	1,416	725.52	727	47.52	48

G.11b: Grade 10 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
10	RW	Female	6,099	689.17	689	40.22	52
10	RW	Male	7,177	688.37	689	42.31	60
10	RW	American Indian or Alaska Native	24	691.29	685	40.36	44
10	RW	Black or African American	1,206	694.14	693	38.38	50
10	RW	Asian	3,806	684.82	685	40.42	51
10	RW	Native Hawaiian/Other Pacific Islander	41	688.54	693	38.83	51
10	RW	Hispanic or Latino	7,379	689.15	689	42.05	56
10	RW	White	810	695.22	697	42.10	55
10	RW	Multiracial (not Hispanic origin)	10	687.20	695	46.47	51
10	RW	NYC	9,946	688.26	689	39.35	52
10	RW	Big 4 Cities	676	675.97	678	52.24	53
10	RW	High Need Urban/Suburban	1,151	686.09	685	41.96	56
10	RW	High Need Rural	66	688.50	693	45.82	50
10	RW	Average Need	872	696.83	697	47.27	61
10	RW	Low Need	489	704.00	701	42.44	60
10	RW	Charter Schools	75	714.79	714	43.52	50
10	RW	0 Years LEP	6,904	691.81	693	40.04	53
10	RW	1 Years LEP	3,172	679.08	675	39.62	53
10	RW	2 Years LEP	1,629	687.50	685	40.63	56
10	RW	3 Years LEP	461	689.06	685	43.62	52
10	RW	4 Years LEP	305	698.31	697	44.73	57
10	RW	5 Years LEP	193	693.30	701	51.13	51
10	RW	6 Years LEP or More	581	702.23	709	48.49	44
10	RW	Arabic	342	684.58	685	37.15	51
10	RW	Bengali	396	686.09	689	33.48	47
10	RW	Chinese	785	683.99	682	41.07	54
10	RW	English	318	697.19	697	44.14	48
10	RW	Spanish	7,208	688.88	689	41.83	56
10	RW	Other Language	4,227	689.31	689	41.27	56
10	RW	Students with Disabilities	1,416	691.00	697	40.05	42

G.12a: Grade 11 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
11	LS	Female	4,306	709.97	711	50.31	61
11	LS	Male	4,566	713.60	716	51.21	57
11	LS	American Indian or Alaska Native	15	711.60	711	50.10	86
11	LS	Black or African American	836	716.28	716	44.59	52
11	LS	Asian	2,402	702.41	702	43.73	53
11	LS	Native Hawaiian/Other Pacific Islander	28	732.61	721	68.59	88
11	LS	Hispanic or Latino	4,992	714.13	716	53.92	66
11	LS	White	591	723.45	721	52.26	59
11	LS	Multiracial (not Hispanic origin)	8	717.63	704.5	43.74	49.5
11	LS	NYC	6,276	711.44	711	49.65	61
11	LS	Big 4 Cities	485	698.57	698	63.53	66
11	LS	High Need Urban/Suburban	882	714.83	716	48.24	54
11	LS	High Need Rural	41	726.71	727	35.18	48
11	LS	Average Need	707	718.11	721	49.64	63
11	LS	Low Need	418	712.13	711	56.21	61
11	LS	Charter Schools	62	731.13	741	52.45	55
11	LS	0 Years LEP	4,178	718.25	721	51.91	63
11	LS	1 Years LEP	1,760	691.35	691	47.05	60
11	LS	2 Years LEP	1,471	705.31	702	43.50	57
11	LS	3 Years LEP	463	710.38	711	42.86	47
11	LS	4 Years LEP	301	718.74	721	49.14	50
11	LS	5 Years LEP	222	728.00	727	43.66	44
11	LS	6 Years LEP or More	456	742.90	741	54.71	48
11	LS	Arabic	215	720.60	721	47.10	59
11	LS	Bengali	229	709.35	706	48.16	64
11	LS	Chinese	427	697.29	694	39.18	50
11	LS	English	214	722.67	721	50.01	52
11	LS	Spanish	4,886	714.24	716	53.93	66
11	LS	Other Language	2,901	708.69	706	46.65	54
11	LS	Students with Disabilities	851	730.35	734	47.61	44

G.12b: Grade 11 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
11	RW	Female	4,306	700.79	701	39.27	48
11	RW	Male	4,566	699.56	701	41.10	48
11	RW	American Indian or Alaska Native	15	702.27	689	47.63	55
11	RW	Black or African American	836	701.85	701	40.97	51
11	RW	Asian	2,402	700.68	701	39.44	48
11	RW	Native Hawaiian/Other Pacific Islander	28	702.86	707	37.10	38
11	RW	Hispanic or Latino	4,992	698.66	701	40.05	49.5
11	RW	White	591	708.20	709	42.65	57
11	RW	Multiracial (not Hispanic origin)	8	693.38	693	44.56	60
11	RW	NYC	6,276	701.40	701	38.07	48
11	RW	Big 4 Cities	485	681.80	682	53.43	59
11	RW	High Need Urban/Suburban	882	693.35	693	42.44	50
11	RW	High Need Rural	41	703.73	701	37.79	34
11	RW	Average Need	707	705.34	705	39.16	51
11	RW	Low Need	418	707.43	705	43.25	53
11	RW	Charter Schools	62	704.98	709	42.65	36
11	RW	0 Years LEP	4,178	703.73	705	38.43	51
11	RW	1 Years LEP	1,760	691.85	689	40.68	53
11	RW	2 Years LEP	1,471	698.34	697	39.06	51
11	RW	3 Years LEP	463	698.99	697	39.96	48
11	RW	4 Years LEP	301	697.29	701	44.83	55
11	RW	5 Years LEP	222	703.79	707	40.24	44
11	RW	6 Years LEP or More	456	707.46	714	47.80	46
11	RW	Arabic	215	699.20	701	34.73	48
11	RW	Bengali	229	698.35	701	35.46	43
11	RW	Chinese	427	703.38	701	39.27	54
11	RW	English	214	706.65	705	46.60	50
11	RW	Spanish	4,886	698.75	701	39.87	51
11	RW	Other Language	2,901	701.78	701	41.07	54
11	RW	Students with Disabilities	851	695.08	701	41.03	43

G.13a: Grade 12 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
12	LS	Female	3,246	723.25	727	45.72	56
12	LS	Male	3,216	726.54	727	48.88	52
12	LS	American Indian or Alaska Native	7	711.00	711	18.44	29
12	LS	Black or African American	626	727.37	727	46.02	48
12	LS	Asian	1,744	716.58	711	41.30	50
12	LS	Native Hawaiian/Other Pacific Islander	9	728.89	734	38.68	50
12	LS	Hispanic or Latino	3,672	727.08	727	49.09	63
12	LS	White	403	736.90	741	52.42	50
12	LS	Multiracial (not Hispanic origin)	1	851.00	851	-	0
12	LS	NYC	4,766	723.33	721	45.38	56
12	LS	Big 4 Cities	345	709.38	716	67.76	50
12	LS	High Need Urban/Suburban	517	730.98	727	44.71	55
12	LS	High Need Rural	38	733.47	734	65.95	50
12	LS	Average Need	488	736.88	734	42.32	50
12	LS	Low Need	293	735.86	734	51.82	50
12	LS	Charter Schools	15	742.20	741	32.10	50
12	LS	0 Years LEP	3,530	728.76	727	45.47	63
12	LS	1 Years LEP	847	704.79	702	45.71	57
12	LS	2 Years LEP	898	715.40	711	42.44	54
12	LS	3 Years LEP	253	723.52	721	45.43	52
12	LS	4 Years LEP	307	730.51	727	49.42	39
12	LS	5 Years LEP	199	738.97	741	43.32	50
12	LS	6 Years LEP or More	419	743.32	741	57.61	56
12	LS	Arabic	171	724.23	727	61.34	52
12	LS	Bengali	211	722.51	711	44.74	59
12	LS	Chinese	236	711.46	711	40.77	47
12	LS	English	131	736.53	741	48.52	45
12	LS	Spanish	3,589	726.98	727	48.96	63
12	LS	Other Language	2,124	722.43	721	43.59	56
12	LS	Students with Disabilities	680	732.14	734	45.24	55

G.13b: Grade 12 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
12	RW	Female	3,246	708.21	709	35.04	44
12	RW	Male	3,216	707.63	709	40.67	44
12	RW	American Indian or Alaska Native	7	702.29	714	28.56	50
12	RW	Black or African American	626	707.13	705	40.53	53
12	RW	Asian	1,744	709.58	709	36.11	40
12	RW	Native Hawaiian/Other Pacific Islander	9	716.33	714	36.84	57
12	RW	Hispanic or Latino	3,672	706.75	709	37.69	47
12	RW	White	403	712.51	714	43.30	50
12	RW	Multiracial (not Hispanic origin)	1	729.00	729	-	0
12	RW	NYC	4,766	708.20	709	35.50	44
12	RW	Big 4 Cities	345	687.54	697	58.84	46
12	RW	High Need Urban/Suburban	517	704.86	705	36.78	47
12	RW	High Need Rural	38	704.45	705	61.55	47
12	RW	Average Need	488	716.09	718	37.08	38
12	RW	Low Need	293	718.83	718	35.28	45
12	RW	Charter Schools	15	723.00	718	33.06	36
12	RW	0 Years LEP	3,530	709.57	709	35.52	50
12	RW	1 Years LEP	847	701.79	701	37.98	45
12	RW	2 Years LEP	898	708.71	709	37.00	47
12	RW	3 Years LEP	253	705.94	709	39.06	47
12	RW	4 Years LEP	307	709.87	714	43.57	42
12	RW	5 Years LEP	199	706.39	714	40.76	44
12	RW	6 Years LEP or More	419	705.34	714	49.97	36
12	RW	Arabic	171	700.26	705	52.28	47
12	RW	Bengali	211	701.29	701	34.12	48
12	RW	Chinese	236	711.58	709	36.69	48
12	RW	English	131	708.27	714	45.88	50
12	RW	Spanish	3,589	707.02	709	37.19	47
12	RW	Other Language	2,124	710.30	709	37.61	46
12	RW	Students with Disabilities	680	696.92	701	37.86	48

APPENDIX H: PROFICIENCY PERCENTAGES

H.1a: Kindergarten Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
K	LS	Female	12,086	5.63	21.96	36.03	36.39
K	LS	Male	13,346	6.29	25.51	35.88	32.32
K	LS	American Indian or Alaska Native	71	5.63	19.72	39.44	35.21
K	LS	Black or African American	1,038	4.72	22.93	36.51	35.84
K	LS	Asian	5,373	8.51	27.77	34.10	29.63
K	LS	Native Hawaiian/Other Pacific Islander	115	6.09	30.43	37.39	26.09
K	LS	Hispanic or Latino	16,415	5.16	23.23	37.24	34.36
K	LS	White	2,387	6.45	19.23	30.79	43.53
K	LS	Multiracial (not Hispanic origin)	33	3.03	21.21	36.36	39.39
K	LS	NYC	15,837	6.71	24.76	34.26	34.27
K	LS	Big 4 Cities	1,084	9.50	34.69	34.04	21.77
K	LS	High Need Urban/Suburban	3,382	5.35	23.57	37.32	33.77
K	LS	High Need Rural	195	4.62	22.56	42.56	30.26
K	LS	Average Need	3,042	3.62	18.97	39.05	38.36
K	LS	Low Need	1,307	3.75	20.05	43.69	32.52
K	LS	Charter Schools	575	0.87	14.09	41.57	43.48
K	LS	0 Years LEP	15,529	6.83	24.82	34.50	33.85
K	LS	1 Years LEP	9,408	4.53	22.28	38.23	34.96
K	LS	Arabic	657	10.35	27.85	30.44	31.35
K	LS	Bengali	692	4.62	19.94	35.55	39.88
K	LS	Chinese	1,414	7.57	32.32	34.58	25.53
K	LS	English	1,739	2.65	19.78	39.74	37.84
K	LS	Spanish	15,159	5.34	23.40	37.30	33.97
K	LS	Other Language	5,771	7.92	24.09	32.26	35.73
K	LS	Students with Disabilities	2,721	7.64	36.82	34.73	20.80

H.1b: Kindergarten Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
K	RW	Female	12,086	35.87	34.72	13.76	15.65
K	RW	Male	13,346	43.19	32.37	11.88	12.56
K	RW	American Indian or Alaska Native	71	40.85	30.99	16.90	11.27
K	RW	Black or African American	1,038	47.21	31.50	10.79	10.50
K	RW	Asian	5,373	28.62	31.01	15.43	24.94
K	RW	Native Hawaiian/Other Pacific Islander	115	42.61	24.35	13.91	19.13
K	RW	Hispanic or Latino	16,415	43.62	34.35	11.76	10.27
K	RW	White	2,387	34.35	34.44	14.54	16.67
K	RW	Multiracial (not Hispanic origin)	33	36.36	39.39	9.09	15.15
K	RW	NYC	15,837	37.51	32.03	13.77	16.69
K	RW	Big 4 Cities	1,084	62.64	25.74	6.73	4.89
K	RW	High Need Urban/Suburban	3,382	44.32	35.84	11.09	8.75
K	RW	High Need Rural	195	54.36	31.79	8.21	5.64
K	RW	Average Need	3,042	41.19	38.95	10.88	8.97
K	RW	Low Need	1,307	36.65	37.41	12.78	13.16
K	RW	Charter Schools	575	23.65	37.04	18.43	20.87
K	RW	0 Years LEP	15,529	37.91	32.05	13.68	16.36
K	RW	1 Years LEP	9,408	42.29	35.64	11.50	10.57
K	RW	Arabic	657	47.95	27.25	11.11	13.70
K	RW	Bengali	692	23.41	27.02	17.49	32.08
K	RW	Chinese	1,414	24.61	31.68	16.34	27.37
K	RW	English	1,739	41.06	37.26	11.56	10.12
K	RW	Spanish	15,159	43.66	33.99	11.93	10.41
K	RW	Other Language	5,771	33.63	32.94	14.11	19.32
K	RW	Students with Disabilities	2,721	57.70	26.98	8.31	7.02

H.1c: Kindergarten Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
K	Overall	Female	12,086	36.51	36.74	16.03	10.72
K	Overall	Male	13,346	43.84	34.11	13.61	8.44
K	Overall	American Indian or Alaska Native	71	40.85	32.39	19.72	7.04
K	Overall	Black or African American	1,038	47.59	32.18	11.56	8.67
K	Overall	Asian	5,373	29.96	35.77	19.77	14.50
K	Overall	Native Hawaiian/Other Pacific Islander	115	42.61	29.57	16.52	11.30
K	Overall	Hispanic or Latino	16,415	44.09	35.42	13.12	7.37
K	Overall	White	2,387	34.86	35.69	16.00	13.45
K	Overall	Multiracial (not Hispanic origin)	33	36.36	39.39	9.09	15.15
K	Overall	NYC	15,837	38.38	34.40	16.02	11.20
K	Overall	Big 4 Cities	1,084	63.28	26.85	7.01	2.86
K	Overall	High Need Urban/Suburban	3,382	44.65	36.78	12.39	6.18
K	Overall	High Need Rural	195	54.36	33.33	8.21	4.10
K	Overall	Average Need	3,042	41.42	39.48	12.00	7.10
K	Overall	Low Need	1,307	36.73	39.40	15.84	8.03
K	Overall	Charter Schools	575	24.00	38.96	23.30	13.74
K	Overall	0 Years LEP	15,529	38.80	34.38	15.96	10.86
K	Overall	1 Years LEP	9,408	42.56	36.82	13.00	7.62
K	Overall	Arabic	657	48.25	29.07	11.87	10.81
K	Overall	Bengali	692	23.99	30.92	23.41	21.68
K	Overall	Chinese	1,414	26.17	37.84	21.22	14.78
K	Overall	English	1,739	41.29	38.47	12.94	7.30
K	Overall	Spanish	15,159	44.16	35.07	13.34	7.43
K	Overall	Other Language	5,771	34.64	35.82	16.76	12.79
K	Overall	Students with Disabilities	2,721	58.14	28.81	8.93	4.12

H.2a: Grade 1 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
1	LS	Female	12,035	3.78	8.97	44.21	43.03
1	LS	Male	13,633	3.84	11.14	46.90	38.11
1	LS	American Indian or Alaska Native	48	2.08	12.50	47.92	37.50
1	LS	Black or African American	1,185	4.22	11.90	48.61	35.27
1	LS	Asian	5,464	4.90	12.13	43.03	39.93
1	LS	Native Hawaiian/Other Pacific Islander	86	1.16	24.42	46.51	27.91
1	LS	Hispanic or Latino	16,600	3.47	9.44	47.52	39.57
1	LS	White	2,254	3.59	8.78	36.34	51.29
1	LS	Multiracial (not Hispanic origin)	31	6.45	9.68	58.06	25.81
1	LS	NYC	16,162	4.38	10.74	44.67	40.21
1	LS	Big 4 Cities	1,247	6.98	15.64	50.44	26.94
1	LS	High Need Urban/Suburban	3,279	2.99	10.95	50.66	35.41
1	LS	High Need Rural	177	2.82	10.17	46.89	40.11
1	LS	Average Need	2,935	1.70	6.00	43.44	48.86
1	LS	Low Need	1,373	2.18	7.65	42.02	48.14
1	LS	Charter Schools	482	0.21	2.07	54.77	42.95
1	LS	0 Years LEP	14,363	4.77	10.88	44.96	39.40
1	LS	1 Years LEP	3,664	6.66	14.03	44.98	34.33
1	LS	2 Years LEP	7,149	0.55	6.73	47.03	45.70
1	LS	Arabic	702	7.55	16.24	43.16	33.05
1	LS	Bengali	715	3.78	10.49	41.82	43.92
1	LS	Chinese	1,230	3.82	11.22	44.72	40.24
1	LS	English	1,278	1.64	5.48	47.18	45.70
1	LS	Spanish	15,724	3.55	9.74	47.44	39.27
1	LS	Other Language	6,019	4.54	11.15	41.54	42.78
1	LS	Students with Disabilities	3,456	2.81	17.59	52.66	26.94

H.2b: Grade 1 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
1	RW	Female	12,035	13.14	20.01	27.93	38.92
1	RW	Male	13,633	17.77	24.47	27.20	30.57
1	RW	American Indian or Alaska Native	48	20.83	14.58	39.58	25.00
1	RW	Black or African American	1,185	22.03	24.98	23.80	29.20
1	RW	Asian	5,464	10.14	15.21	26.28	48.37
1	RW	Native Hawaiian/Other Pacific Islander	86	16.28	19.77	31.40	32.56
1	RW	Hispanic or Latino	16,600	17.17	24.98	28.19	29.67
1	RW	White	2,254	13.66	19.52	27.42	39.40
1	RW	Multiracial (not Hispanic origin)	31	22.58	22.58	25.81	29.03
1	RW	NYC	16,162	16.34	21.55	26.99	35.12
1	RW	Big 4 Cities	1,247	30.47	26.46	22.69	20.37
1	RW	High Need Urban/Suburban	3,279	15.92	24.49	29.00	30.59
1	RW	High Need Rural	177	13.56	27.12	32.77	26.55
1	RW	Average Need	2,935	9.30	23.78	29.30	37.61
1	RW	Low Need	1,373	9.69	20.25	27.02	43.04
1	RW	Charter Schools	482	6.02	20.54	37.14	36.31
1	RW	0 Years LEP	14,363	16.93	21.84	27.24	34.00
1	RW	1 Years LEP	3,664	22.16	23.47	24.56	29.80
1	RW	2 Years LEP	7,149	9.43	22.44	29.72	38.41
1	RW	Arabic	702	24.50	23.50	24.22	27.78
1	RW	Bengali	715	9.37	14.41	23.36	52.87
1	RW	Chinese	1,230	6.50	13.58	28.13	51.79
1	RW	English	1,278	10.95	23.40	30.28	35.37
1	RW	Spanish	15,724	17.64	24.87	28.08	29.41
1	RW	Other Language	6,019	12.83	18.26	26.30	42.62
1	RW	Students with Disabilities	3,456	28.70	28.82	24.59	17.88

H.2c: Grade 1 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
1	Overall	Female	12,035	13.84	21.70	38.97	25.48
1	Overall	Male	13,633	18.24	26.10	36.37	19.30
1	Overall	American Indian or Alaska Native	48	20.83	18.75	39.58	20.83
1	Overall	Black or African American	1,185	22.11	26.67	32.57	18.65
1	Overall	Asian	5,464	11.02	18.69	41.12	29.17
1	Overall	Native Hawaiian/Other Pacific Islander	86	16.28	22.09	48.84	12.79
1	Overall	Hispanic or Latino	16,600	17.70	26.08	37.08	19.13
1	Overall	White	2,254	14.11	20.76	34.96	30.17
1	Overall	Multiracial (not Hispanic origin)	31	22.58	25.81	32.26	19.35
1	Overall	NYC	16,162	17.00	23.56	37.22	22.22
1	Overall	Big 4 Cities	1,247	31.19	27.35	29.83	11.63
1	Overall	High Need Urban/Suburban	3,279	16.35	25.77	38.58	19.30
1	Overall	High Need Rural	177	14.12	29.38	42.37	14.12
1	Overall	Average Need	2,935	9.68	24.70	39.11	26.51
1	Overall	Low Need	1,373	10.12	21.49	38.24	30.15
1	Overall	Charter Schools	482	6.02	20.75	50.21	23.03
1	Overall	0 Years LEP	14,363	17.64	23.79	37.33	21.25
1	Overall	1 Years LEP	3,664	23.14	25.33	32.51	19.02
1	Overall	2 Years LEP	7,149	9.57	23.49	40.86	26.09
1	Overall	Arabic	702	25.21	24.64	31.05	19.09
1	Overall	Bengali	715	9.65	17.34	40.84	32.17
1	Overall	Chinese	1,230	7.56	17.40	44.80	30.24
1	Overall	English	1,278	11.11	23.87	40.69	24.33
1	Overall	Spanish	15,724	18.20	26.05	36.90	18.86
1	Overall	Other Language	6,019	13.46	20.90	37.63	28.01
1	Overall	Students with Disabilities	3,456	29.05	31.08	29.54	10.33

H.3a: Grade 2 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
2	LS	Female	9,676	2.47	6.61	33.93	56.99
2	LS	Male	11,337	2.64	7.52	36.00	53.85
2	LS	American Indian or Alaska Native	34	2.94	8.82	32.35	55.88
2	LS	Black or African American	1,188	2.02	8.42	35.94	53.62
2	LS	Asian	4,060	3.47	9.41	30.74	56.38
2	LS	Native Hawaiian/Other Pacific Islander	61	6.56	21.31	29.51	42.62
2	LS	Hispanic or Latino	13,939	2.35	6.32	36.97	54.36
2	LS	White	1,709	2.28	6.50	29.26	61.97
2	LS	Multiracial (not Hispanic origin)	22	4.55	9.09	31.82	54.55
2	LS	NYC	13,426	2.95	8.03	36.39	52.63
2	LS	Big 4 Cities	1,171	4.53	10.76	40.39	44.32
2	LS	High Need Urban/Suburban	2,738	1.68	5.41	36.49	56.43
2	LS	High Need Rural	136	0.74	3.68	31.62	63.97
2	LS	Average Need	2,248	1.25	4.27	27.40	67.08
2	LS	Low Need	976	1.43	3.48	24.18	70.90
2	LS	Charter Schools	311	0.00	1.61	35.05	63.34
2	LS	0 Years LEP	11,761	3.31	8.25	36.09	52.36
2	LS	1 Years LEP	2,294	5.84	13.69	38.93	41.54
2	LS	2 Years LEP	2,263	0.31	4.60	35.88	59.21
2	LS	3 Years LEP	4,233	0.07	2.22	29.86	67.85
2	LS	Arabic	660	5.91	14.24	37.27	42.58
2	LS	Bengali	579	1.73	8.12	29.71	60.45
2	LS	Chinese	749	1.87	8.14	31.91	58.08
2	LS	English	843	1.19	2.73	30.13	65.95
2	LS	Spanish	13,341	2.41	6.45	37.34	53.80
2	LS	Other Language	4,841	2.95	8.41	30.39	58.25
2	LS	Students with Disabilities	3,595	0.81	9.57	49.24	40.39

H.3b: Grade 2 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
2	RW	Female	9,676	14.75	36.96	32.80	15.49
2	RW	Male	11,337	18.46	38.00	30.96	12.58
2	RW	American Indian or Alaska Native	34	23.53	35.29	14.71	26.47
2	RW	Black or African American	1,188	21.55	39.81	27.95	10.69
2	RW	Asian	4,060	14.16	29.43	34.80	21.60
2	RW	Native Hawaiian/Other Pacific Islander	61	31.15	29.51	22.95	16.39
2	RW	Hispanic or Latino	13,939	17.23	40.44	31.48	10.85
2	RW	White	1,709	15.04	31.60	30.84	22.53
2	RW	Multiracial (not Hispanic origin)	22	18.18	40.91	22.73	18.18
2	RW	NYC	13,426	17.96	38.72	30.75	12.57
2	RW	Big 4 Cities	1,171	28.78	40.14	22.80	8.28
2	RW	High Need Urban/Suburban	2,738	14.68	37.95	34.00	13.37
2	RW	High Need Rural	136	8.82	44.12	33.09	13.97
2	RW	Average Need	2,248	11.52	32.78	36.21	19.48
2	RW	Low Need	976	8.61	26.02	37.81	27.56
2	RW	Charter Schools	311	4.50	39.23	41.16	15.11
2	RW	0 Years LEP	11,761	18.48	39.08	30.74	11.71
2	RW	1 Years LEP	2,294	27.07	35.48	25.11	12.34
2	RW	2 Years LEP	2,263	12.99	36.19	33.41	17.41
2	RW	3 Years LEP	4,233	8.69	34.85	37.14	19.32
2	RW	Arabic	660	30.91	37.12	21.06	10.91
2	RW	Bengali	579	12.61	27.46	34.54	25.39
2	RW	Chinese	749	9.08	28.04	39.25	23.63
2	RW	English	843	9.49	33.33	37.25	19.93
2	RW	Spanish	13,341	17.55	40.72	31.17	10.55
2	RW	Other Language	4,841	15.55	32.16	32.60	19.69
2	RW	Students with Disabilities	3,595	29.32	46.37	19.89	4.42

H.3c: Grade 2 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
2	Overall Female		9,676	14.88	36.92	33.61	14.59
2	Overall Male		11,337	18.59	37.98	31.58	11.85
2	Overall American Indian or Alaska Native		34	23.53	35.29	17.65	23.53
2	Overall Black or African American		1,188	21.80	39.56	28.70	9.93
2	Overall Asian		4,060	14.33	29.43	35.71	20.52
2	Overall Native Hawaiian/Other Pacific Islander		61	31.15	31.15	21.31	16.39
2	Overall Hispanic or Latino		13,939	17.33	40.42	32.12	10.13
2	Overall White		1,709	15.21	31.54	31.60	21.65
2	Overall Multiracial (not Hispanic origin)		22	18.18	40.91	22.73	18.18
2	Overall NYC		13,426	18.11	38.69	31.42	11.78
2	Overall Big 4 Cities		1,171	29.21	39.71	23.40	7.69
2	Overall High Need Urban/Suburban		2,738	14.72	37.91	34.84	12.53
2	Overall High Need Rural		136	8.82	44.85	32.35	13.97
2	Overall Average Need		2,248	11.61	32.78	36.74	18.86
2	Overall Low Need		976	8.61	26.23	39.14	26.02
2	Overall Charter Schools		311	4.50	39.23	42.44	13.83
2	Overall 0 Years LEP		11,761	18.65	39.04	31.37	10.95
2	Overall 1 Years LEP		2,294	27.38	35.44	25.76	11.42
2	Overall 2 Years LEP		2,263	13.04	36.19	34.33	16.44
2	Overall 3 Years LEP		4,233	8.69	34.85	37.99	18.47
2	Overall Arabic		660	31.21	36.82	21.82	10.15
2	Overall Bengali		579	12.61	27.81	35.41	24.18
2	Overall Chinese		749	9.08	28.17	40.32	22.43
2	Overall English		843	9.49	33.33	38.32	18.86
2	Overall Spanish		13,341	17.67	40.69	31.80	9.84
2	Overall Other Language		4,841	15.78	32.08	33.38	18.76
2	Overall Students with Disabilities		3,595	29.32	46.43	20.14	4.12

H.4a: Grade 3 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
3	LS	Female	9,172	3.21	5.70	31.68	59.41
3	LS	Male	10,968	3.13	6.39	34.56	55.92
3	LS	American Indian or Alaska Native	37	2.70	5.41	27.03	64.86
3	LS	Black or African American	1,169	1.88	6.59	32.76	58.77
3	LS	Asian	3,677	4.54	9.44	31.14	54.88
3	LS	Native Hawaiian/Other Pacific Islander	43	9.30	16.28	23.26	51.16
3	LS	Hispanic or Latino	13,725	2.90	5.06	34.50	57.54
3	LS	White	1,470	2.99	6.46	27.89	62.65
3	LS	Multiracial (not Hispanic origin)	19	5.26	5.26	21.05	68.42
3	LS	NYC	13,290	3.59	6.51	33.72	56.18
3	LS	Big 4 Cities	1,179	2.97	10.26	36.98	49.79
3	LS	High Need Urban/Suburban	2,480	2.22	4.48	35.04	58.27
3	LS	High Need Rural	117	3.42	8.55	32.48	55.56
3	LS	Average Need	1,977	2.58	4.15	28.38	64.90
3	LS	Low Need	848	1.42	3.89	27.59	67.10
3	LS	Charter Schools	243	1.23	0.82	29.63	68.31
3	LS	0 Years LEP	11,953	3.76	6.21	32.85	57.17
3	LS	1 Years LEP	2,062	7.86	17.26	38.31	36.57
3	LS	2 Years LEP	1,347	1.04	4.75	39.64	54.57
3	LS	3 Years LEP	1,089	0.28	2.02	30.95	66.76
3	LS	4 Years LEP	3,242	0.06	0.99	29.43	69.52
3	LS	Arabic	601	5.49	10.82	34.61	49.08
3	LS	Bengali	555	2.16	7.57	30.09	60.18
3	LS	Chinese	691	4.63	5.93	28.65	60.78
3	LS	English	758	1.72	3.96	28.89	65.44
3	LS	Spanish	13,158	2.92	5.15	34.79	57.14
3	LS	Other Language	4,377	3.72	8.41	30.32	57.55
3	LS	Students with Disabilities	4,112	0.56	4.94	47.67	46.84

H.4b: Grade 3 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
3	RW	Female	9,172	9.91	20.34	47.82	21.93
3	RW	Male	10,968	12.53	23.38	46.63	17.47
3	RW	American Indian or Alaska Native	37	13.51	13.51	54.05	18.92
3	RW	Black or African American	1,169	13.86	25.41	42.51	18.22
3	RW	Asian	3,677	12.54	18.93	41.37	27.17
3	RW	Native Hawaiian/Other Pacific Islander	43	25.58	16.28	37.21	20.93
3	RW	Hispanic or Latino	13,725	10.72	22.91	49.52	16.85
3	RW	White	1,470	11.56	19.12	43.40	25.92
3	RW	Multiracial (not Hispanic origin)	19	10.53	0.00	57.89	31.58
3	RW	NYC	13,290	11.90	22.61	46.76	18.73
3	RW	Big 4 Cities	1,179	18.32	25.11	42.49	14.08
3	RW	High Need Urban/Suburban	2,480	9.35	20.73	50.77	19.15
3	RW	High Need Rural	117	17.95	18.80	43.59	19.66
3	RW	Average Need	1,977	8.55	20.69	47.70	23.07
3	RW	Low Need	848	6.60	16.16	46.82	30.42
3	RW	Charter Schools	243	2.88	19.34	53.91	23.87
3	RW	0 Years LEP	11,953	11.55	22.28	47.94	18.23
3	RW	1 Years LEP	2,062	25.80	26.48	32.59	15.13
3	RW	2 Years LEP	1,347	11.66	23.01	42.54	22.79
3	RW	3 Years LEP	1,089	6.15	22.13	47.75	23.97
3	RW	4 Years LEP	3,242	3.61	17.55	54.29	24.55
3	RW	Arabic	601	20.97	23.13	38.60	17.30
3	RW	Bengali	555	8.65	18.20	43.06	30.09
3	RW	Chinese	691	9.70	15.05	41.68	33.57
3	RW	English	758	8.31	21.77	47.49	22.43
3	RW	Spanish	13,158	11.00	22.95	49.34	16.71
3	RW	Other Language	4,377	12.15	20.58	43.16	24.10
3	RW	Students with Disabilities	4,112	15.47	37.38	39.86	7.30

H.4c: Grade 3 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
3	Overall	Female	9,172	10.19	20.29	49.90	19.61
3	Overall	Male	10,968	12.73	23.42	48.20	15.65
3	Overall	American Indian or Alaska Native	37	13.51	13.51	54.05	18.92
3	Overall	Black or African American	1,169	14.03	25.32	44.05	16.60
3	Overall	Asian	3,677	13.05	18.93	43.65	24.37
3	Overall	Native Hawaiian/Other Pacific Islander	43	25.58	16.28	37.21	20.93
3	Overall	Hispanic or Latino	13,725	10.89	22.93	51.18	15.00
3	Overall	White	1,470	11.84	18.98	45.65	23.54
3	Overall	Multiracial (not Hispanic origin)	19	10.53	0.00	68.42	21.05
3	Overall	NYC	13,290	12.17	22.61	48.58	16.64
3	Overall	Big 4 Cities	1,179	18.74	24.77	44.19	12.30
3	Overall	High Need Urban/Suburban	2,480	9.40	20.89	52.54	17.18
3	Overall	High Need Rural	117	17.95	18.80	47.01	16.24
3	Overall	Average Need	1,977	8.85	20.59	48.81	21.75
3	Overall	Low Need	848	6.60	16.27	50.00	27.12
3	Overall	Charter Schools	243	2.88	19.75	55.56	21.81
3	Overall	0 Years LEP	11,953	11.85	22.25	49.59	16.31
3	Overall	1 Years LEP	2,062	26.29	26.48	34.43	12.80
3	Overall	2 Years LEP	1,347	11.80	23.01	45.29	19.90
3	Overall	3 Years LEP	1,089	6.15	22.22	49.86	21.76
3	Overall	4 Years LEP	3,242	3.64	17.61	56.29	22.46
3	Overall	Arabic	601	21.46	22.63	40.60	15.31
3	Overall	Bengali	555	9.01	18.74	45.05	27.21
3	Overall	Chinese	691	9.99	15.20	45.15	29.67
3	Overall	English	758	8.44	21.77	48.81	20.98
3	Overall	Spanish	13,158	11.15	22.99	51.03	14.84
3	Overall	Other Language	4,377	12.61	20.45	45.10	21.84
3	Overall	Students with Disabilities	4,112	15.52	37.52	40.42	6.54

H.5a: Grade 4 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
4	LS	Female	8,416	4.13	6.57	25.39	63.90
4	LS	Male	9,882	3.98	6.84	28.08	61.10
4	LS	American Indian or Alaska Native	31	3.23	12.90	22.58	61.29
4	LS	Black or African American	1,128	2.57	8.33	29.61	59.49
4	LS	Asian	3,281	5.43	10.45	28.86	55.26
4	LS	Native Hawaiian/Other Pacific Islander	48	6.25	16.67	37.50	39.58
4	LS	Hispanic or Latino	12,444	3.84	5.46	26.45	64.25
4	LS	White	1,352	3.85	7.40	22.86	65.90
4	LS	Multiracial (not Hispanic origin)	14	0.00	7.14	35.71	57.14
4	LS	NYC	12,320	4.48	7.01	27.07	61.44
4	LS	Big 4 Cities	1,174	4.51	10.82	31.60	53.07
4	LS	High Need Urban/Suburban	2,169	3.00	5.72	25.50	65.79
4	LS	High Need Rural	109	0.00	5.50	28.44	66.06
4	LS	Average Need	1,673	3.59	4.06	24.51	67.84
4	LS	Low Need	697	1.43	5.16	23.53	69.87
4	LS	Charter Schools	152	0.66	1.97	31.58	65.79
4	LS	0 Years LEP	10,911	4.83	6.33	25.09	63.74
4	LS	1 Years LEP	1,989	9.85	20.11	38.46	31.57
4	LS	2 Years LEP	1,229	0.98	7.40	37.67	53.95
4	LS	3 Years LEP	592	0.00	3.55	27.20	69.26
4	LS	4 Years LEP	804	0.12	0.75	24.88	74.25
4	LS	5 Years LEP	2,400	0.04	0.50	20.25	79.21
4	LS	Arabic	614	7.98	13.52	30.13	48.37
4	LS	Bengali	508	1.77	7.48	26.97	63.78
4	LS	Chinese	543	5.34	9.94	31.68	53.04
4	LS	English	608	1.81	3.62	23.36	71.22
4	LS	Spanish	11,960	3.90	5.50	26.84	63.75
4	LS	Other Language	4,065	4.33	9.20	26.22	60.25
4	LS	Students with Disabilities	4,170	0.38	3.81	36.12	59.69

H.5b: Grade 4 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
4	RW	Female	8,416	9.08	20.62	41.84	28.47
4	RW	Male	9,882	12.08	23.87	40.22	23.82
4	RW	American Indian or Alaska Native	31	6.45	25.81	41.94	25.81
4	RW	Black or African American	1,128	13.83	27.75	37.15	21.28
4	RW	Asian	3,281	12.59	19.63	34.53	33.25
4	RW	Native Hawaiian/Other Pacific Islander	48	16.67	27.08	37.50	18.75
4	RW	Hispanic or Latino	12,444	9.95	22.77	43.47	23.81
4	RW	White	1,352	10.36	20.78	36.83	32.03
4	RW	Multiracial (not Hispanic origin)	14	7.14	14.29	35.71	42.86
4	RW	NYC	12,320	11.06	23.04	40.97	24.93
4	RW	Big 4 Cities	1,174	18.06	29.13	34.33	18.48
4	RW	High Need Urban/Suburban	2,169	9.22	19.73	42.65	28.40
4	RW	High Need Rural	109	5.50	22.94	49.54	22.02
4	RW	Average Need	1,673	8.19	18.53	42.74	30.54
4	RW	Low Need	697	5.16	17.36	38.59	38.88
4	RW	Charter Schools	152	2.63	18.42	53.29	25.66
4	RW	0 Years LEP	10,911	10.68	22.06	42.27	24.99
4	RW	1 Years LEP	1,989	25.94	30.92	26.80	16.34
4	RW	2 Years LEP	1,229	9.11	25.96	37.43	27.50
4	RW	3 Years LEP	592	5.57	24.83	41.05	28.55
4	RW	4 Years LEP	804	4.98	21.02	41.79	32.21
4	RW	5 Years LEP	2,400	2.88	14.50	47.46	35.17
4	RW	Arabic	614	22.96	25.24	32.90	18.89
4	RW	Bengali	508	7.48	18.31	36.02	38.19
4	RW	Chinese	543	12.34	19.15	34.81	33.70
4	RW	English	608	7.73	19.41	42.93	29.93
4	RW	Spanish	11,960	10.03	23.02	43.51	23.44
4	RW	Other Language	4,065	11.46	21.43	35.84	31.27
4	RW	Students with Disabilities	4,170	10.60	35.28	40.62	13.50

H.5c: Grade 4 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
4	Overall	Female	8,416	9.45	20.65	44.13	25.77
4	Overall	Male	9,882	12.37	23.78	42.13	21.73
4	Overall	American Indian or Alaska Native	31	6.45	25.81	41.94	25.81
4	Overall	Black or African American	1,128	13.92	27.75	38.03	20.30
4	Overall	Asian	3,281	13.04	19.93	37.61	29.41
4	Overall	Native Hawaiian/Other Pacific Islander	48	16.67	27.08	41.67	14.58
4	Overall	Hispanic or Latino	12,444	10.25	22.62	45.40	21.72
4	Overall	White	1,352	10.72	20.93	38.83	29.51
4	Overall	Multiracial (not Hispanic origin)	14	7.14	14.29	42.86	35.71
4	Overall	NYC	12,320	11.45	22.95	42.84	22.76
4	Overall	Big 4 Cities	1,174	18.14	29.47	35.78	16.61
4	Overall	High Need Urban/Suburban	2,169	9.45	19.78	44.77	26.00
4	Overall	High Need Rural	109	5.50	22.94	51.38	20.18
4	Overall	Average Need	1,673	8.37	18.53	45.25	27.85
4	Overall	Low Need	697	5.45	17.50	43.90	33.14
4	Overall	Charter Schools	152	2.63	18.42	57.24	21.71
4	Overall	0 Years LEP	10,911	11.07	21.91	43.97	23.04
4	Overall	1 Years LEP	1,989	26.75	30.97	29.51	12.77
4	Overall	2 Years LEP	1,229	9.11	26.28	40.60	24.00
4	Overall	3 Years LEP	592	5.57	25.17	44.59	24.66
4	Overall	4 Years LEP	804	4.98	21.02	44.40	29.60
4	Overall	5 Years LEP	2,400	2.88	14.54	49.71	32.88
4	Overall	Arabic	614	23.29	25.08	34.69	16.94
4	Overall	Bengali	508	8.07	17.91	38.39	35.63
4	Overall	Chinese	543	12.52	19.34	38.49	29.65
4	Overall	English	608	7.73	19.57	44.57	28.13
4	Overall	Spanish	11,960	10.33	22.87	45.48	21.32
4	Overall	Other Language	4,065	11.86	21.75	38.13	28.27
4	Overall	Students with Disabilities	4,170	10.60	35.42	41.89	12.09

H.6a: Grade 5 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
5	LS	Female	7,002	5.26	9.47	35.86	49.41
5	LS	Male	8,831	5.38	8.91	36.53	49.18
5	LS	American Indian or Alaska Native	21	9.52	9.52	19.05	61.90
5	LS	Black or African American	1,063	3.95	10.72	39.79	45.53
5	LS	Asian	2,843	7.32	15.02	35.14	42.53
5	LS	Native Hawaiian/Other Pacific Islander	54	1.85	24.07	33.33	40.74
5	LS	Hispanic or Latino	10,717	5.05	7.19	36.56	51.20
5	LS	White	1,122	4.37	10.87	33.16	51.60
5	LS	Multiracial (not Hispanic origin)	13	0.00	7.69	23.08	69.23
5	LS	NYC	10,694	5.96	9.88	36.01	48.15
5	LS	Big 4 Cities	1,026	6.04	13.55	42.30	38.11
5	LS	High Need Urban/Suburban	1,802	4.33	7.16	35.41	53.11
5	LS	High Need Rural	119	4.20	5.04	38.66	52.10
5	LS	Average Need	1,335	3.07	5.62	34.68	56.63
5	LS	Low Need	568	2.82	6.51	32.22	58.45
5	LS	Charter Schools	280	1.43	2.50	42.14	53.93
5	LS	0 Years LEP	9,345	6.25	8.43	34.87	50.44
5	LS	1 Years LEP	1,918	11.63	24.24	39.83	24.30
5	LS	2 Years LEP	1,125	2.22	11.29	41.60	44.89
5	LS	3 Years LEP	532	0.19	5.26	41.54	53.01
5	LS	4 Years LEP	368	0.54	3.26	41.30	54.89
5	LS	5 Years LEP	610	0.49	1.80	31.64	66.07
5	LS	6 Years LEP or More	1,913	0.10	0.84	35.08	63.98
5	LS	Arabic	514	9.34	17.12	36.96	36.58
5	LS	Bengali	381	2.62	13.12	35.70	48.56
5	LS	Chinese	461	10.85	15.84	33.41	39.91
5	LS	English	478	2.51	5.02	35.15	57.32
5	LS	Spanish	10,350	5.12	7.32	36.70	50.86
5	LS	Other Language	3,649	5.29	12.52	35.38	46.81
5	LS	Students with Disabilities	3,899	0.90	4.23	47.60	47.27

H.6b: Grade 5 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
5	RW	Female	7,002	12.15	16.41	33.95	37.49
5	RW	Male	8,831	14.48	19.56	34.16	31.80
5	RW	American Indian or Alaska Native	21	14.29	14.29	28.57	42.86
5	RW	Black or African American	1,063	14.77	22.48	35.37	27.38
5	RW	Asian	2,843	18.01	18.19	26.38	37.43
5	RW	Native Hawaiian/Other Pacific Islander	54	18.52	29.63	18.52	33.33
5	RW	Hispanic or Latino	10,717	11.83	17.76	36.30	34.11
5	RW	White	1,122	15.95	17.65	31.82	34.58
5	RW	Multiracial (not Hispanic origin)	13	7.69	0.00	38.46	53.85
5	RW	NYC	10,694	14.40	18.68	33.32	33.60
5	RW	Big 4 Cities	1,026	20.08	23.88	35.87	20.18
5	RW	High Need Urban/Suburban	1,802	11.38	15.21	34.52	38.90
5	RW	High Need Rural	119	9.24	24.37	36.13	30.25
5	RW	Average Need	1,335	8.24	17.23	35.66	38.88
5	RW	Low Need	568	8.45	11.97	32.57	47.01
5	RW	Charter Schools	280	3.57	11.07	47.50	37.86
5	RW	0 Years LEP	9,345	13.47	17.34	34.65	34.54
5	RW	1 Years LEP	1,918	30.19	26.17	22.78	20.86
5	RW	2 Years LEP	1,125	13.78	19.73	31.47	35.02
5	RW	3 Years LEP	532	6.58	19.36	36.28	37.78
5	RW	4 Years LEP	368	5.43	21.47	38.32	34.78
5	RW	5 Years LEP	610	2.46	16.72	40.82	40.00
5	RW	6 Years LEP or More	1,913	3.29	12.75	40.36	43.60
5	RW	Arabic	514	30.35	19.65	27.24	22.76
5	RW	Bengali	381	12.60	19.16	27.30	40.94
5	RW	Chinese	461	20.39	20.82	24.73	34.06
5	RW	English	478	7.74	17.78	37.87	36.61
5	RW	Spanish	10,350	11.99	17.74	36.29	33.98
5	RW	Other Language	3,649	15.18	18.77	30.12	35.93
5	RW	Students with Disabilities	3,899	10.26	26.57	41.34	21.83

H.6c: Grade 5 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
5	Overall	Female	7,002	12.51	16.91	40.72	29.86
5	Overall	Male	8,831	14.85	19.62	39.55	25.98
5	Overall	American Indian or Alaska Native	21	14.29	14.29	38.10	33.33
5	Overall	Black or African American	1,063	15.15	22.39	41.49	20.98
5	Overall	Asian	2,843	18.47	19.10	32.92	29.51
5	Overall	Native Hawaiian/Other Pacific Islander	54	18.52	29.63	27.78	24.07
5	Overall	Hispanic or Latino	10,717	12.19	17.85	42.27	27.69
5	Overall	White	1,122	16.13	18.18	36.27	29.41
5	Overall	Multiracial (not Hispanic origin)	13	7.69	0.00	53.85	38.46
5	Overall	NYC	10,694	14.80	18.96	39.10	27.14
5	Overall	Big 4 Cities	1,026	20.18	24.07	40.06	15.69
5	Overall	High Need Urban/Suburban	1,802	11.82	15.26	41.12	31.80
5	Overall	High Need Rural	119	9.24	25.21	42.02	23.53
5	Overall	Average Need	1,335	8.46	17.53	42.92	31.09
5	Overall	Low Need	568	8.63	12.68	39.96	38.73
5	Overall	Charter Schools	280	3.93	10.71	55.71	29.64
5	Overall	0 Years LEP	9,345	13.86	17.49	40.49	28.16
5	Overall	1 Years LEP	1,918	30.97	26.96	27.79	14.29
5	Overall	2 Years LEP	1,125	13.87	20.36	38.22	27.56
5	Overall	3 Years LEP	532	6.77	19.92	42.67	30.64
5	Overall	4 Years LEP	368	5.71	21.74	43.75	28.80
5	Overall	5 Years LEP	610	2.62	16.89	45.90	34.59
5	Overall	6 Years LEP or More	1,913	3.35	12.81	47.99	35.86
5	Overall	Arabic	514	30.93	19.07	30.93	19.07
5	Overall	Bengali	381	12.86	19.42	34.65	33.07
5	Overall	Chinese	461	21.69	21.04	29.72	27.55
5	Overall	English	478	8.16	18.62	42.68	30.54
5	Overall	Spanish	10,350	12.33	17.84	42.36	27.48
5	Overall	Other Language	3,649	15.46	19.54	36.39	28.61
5	Overall	Students with Disabilities	3,899	10.31	26.78	45.52	17.39

H.7a: Grade 6 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
6	LS	Female	5,885	7.44	13.00	44.98	34.58
6	LS	Male	7,783	7.29	12.95	43.70	36.07
6	LS	American Indian or Alaska Native	32	12.50	15.63	34.38	37.50
6	LS	Black or African American	1,008	5.16	14.38	50.00	30.46
6	LS	Asian	2,306	10.32	20.16	40.76	28.75
6	LS	Native Hawaiian/Other Pacific Islander	45	11.11	22.22	37.78	28.89
6	LS	Hispanic or Latino	9,335	6.90	10.84	44.94	37.32
6	LS	White	925	6.59	14.27	40.32	38.81
6	LS	Multiracial (not Hispanic origin)	17	5.88	23.53	47.06	23.53
6	LS	NYC	9,441	8.34	13.70	43.80	34.17
6	LS	Big 4 Cities	969	5.88	15.79	48.30	30.03
6	LS	High Need Urban/Suburban	1,458	5.90	10.97	44.79	38.34
6	LS	High Need Rural	95	3.16	11.58	40.00	45.26
6	LS	Average Need	1,067	4.78	10.50	43.30	41.42
6	LS	Low Need	424	4.72	8.73	40.80	45.75
6	LS	Charter Schools	211	0.47	1.90	56.40	41.23
6	LS	0 Years LEP	8,060	8.61	11.28	43.57	36.54
6	LS	1 Years LEP	1,847	14.29	30.10	39.04	16.57
6	LS	2 Years LEP	1,075	2.88	18.79	49.30	29.02
6	LS	3 Years LEP	420	0.95	8.81	51.43	38.81
6	LS	4 Years LEP	318	0.94	7.23	52.83	38.99
6	LS	5 Years LEP	310	0.65	2.90	48.06	48.39
6	LS	6 Years LEP or More	1,607	0.25	2.12	46.11	51.52
6	LS	Arabic	468	12.61	20.51	41.67	25.21
6	LS	Bengali	304	5.92	17.11	47.04	29.93
6	LS	Chinese	381	14.96	24.93	37.27	22.83
6	LS	English	351	3.13	7.69	46.44	42.74
6	LS	Spanish	9,033	7.00	10.89	45.00	37.11
6	LS	Other Language	3,131	7.28	16.58	42.80	33.34
6	LS	Students with Disabilities	3,461	0.84	6.41	56.31	36.43

H.7b: Grade 6 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
6	RW	Female	5,885	14.36	20.99	32.20	32.46
6	RW	Male	7,783	16.91	23.10	32.80	27.19
6	RW	American Indian or Alaska Native	32	12.50	15.63	46.88	25.00
6	RW	Black or African American	1,008	15.48	26.69	31.55	26.29
6	RW	Asian	2,306	21.73	22.12	25.24	30.92
6	RW	Native Hawaiian/Other Pacific Islander	45	22.22	24.44	33.33	20.00
6	RW	Hispanic or Latino	9,335	14.25	21.71	34.86	29.18
6	RW	White	925	17.08	21.84	28.43	32.65
6	RW	Multiracial (not Hispanic origin)	17	11.76	52.94	5.88	29.41
6	RW	NYC	9,441	17.28	22.90	31.84	27.98
6	RW	Big 4 Cities	969	17.85	26.32	35.09	20.74
6	RW	High Need Urban/Suburban	1,458	12.89	18.86	33.06	35.19
6	RW	High Need Rural	95	9.47	20.00	37.89	32.63
6	RW	Average Need	1,067	10.87	21.46	34.68	32.99
6	RW	Low Need	424	8.49	15.33	30.19	45.99
6	RW	Charter Schools	211	2.37	13.27	40.76	43.60
6	RW	0 Years LEP	8,060	15.84	21.75	33.28	29.13
6	RW	1 Years LEP	1,847	31.84	29.56	21.66	16.95
6	RW	2 Years LEP	1,075	15.53	25.58	29.02	29.86
6	RW	3 Years LEP	420	7.38	21.43	34.05	37.14
6	RW	4 Years LEP	318	5.66	18.87	41.19	34.28
6	RW	5 Years LEP	310	4.19	21.29	40.97	33.55
6	RW	6 Years LEP or More	1,607	3.92	14.69	39.95	41.44
6	RW	Arabic	468	29.91	24.57	26.92	18.59
6	RW	Bengali	304	17.76	23.03	30.26	28.95
6	RW	Chinese	381	23.88	25.72	25.46	24.93
6	RW	English	351	8.55	19.66	37.32	34.47
6	RW	Spanish	9,033	14.39	21.96	34.82	28.83
6	RW	Other Language	3,131	17.44	22.26	27.37	32.93
6	RW	Students with Disabilities	3,461	9.48	30.45	39.82	20.25

H.7c: Grade 6 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
6	Overall	Female	5,885	15.26	21.39	42.28	21.07
6	Overall	Male	7,783	17.42	23.51	40.64	18.42
6	Overall	American Indian or Alaska Native	32	15.63	15.63	50.00	18.75
6	Overall	Black or African American	1,008	16.27	27.08	40.18	16.47
6	Overall	Asian	2,306	22.68	22.81	34.43	20.08
6	Overall	Native Hawaiian/Other Pacific Islander	45	24.44	22.22	37.78	15.56
6	Overall	Hispanic or Latino	9,335	14.87	22.09	43.74	19.30
6	Overall	White	925	17.30	22.16	36.00	24.54
6	Overall	Multiracial (not Hispanic origin)	17	17.65	47.06	17.65	17.65
6	Overall	NYC	9,441	18.07	23.32	40.29	18.31
6	Overall	Big 4 Cities	969	18.06	27.14	40.25	14.55
6	Overall	High Need Urban/Suburban	1,458	13.58	19.20	44.17	23.05
6	Overall	High Need Rural	95	9.47	20.00	43.16	27.37
6	Overall	Average Need	1,067	11.34	21.46	43.67	23.52
6	Overall	Low Need	424	8.73	16.04	42.92	32.31
6	Overall	Charter Schools	211	2.37	13.27	58.77	25.59
6	Overall	0 Years LEP	8,060	16.63	21.94	42.23	19.21
6	Overall	1 Years LEP	1,847	33.08	30.64	25.50	10.77
6	Overall	2 Years LEP	1,075	15.72	26.79	38.98	18.51
6	Overall	3 Years LEP	420	7.38	22.14	44.29	26.19
6	Overall	4 Years LEP	318	6.29	19.81	50.31	23.58
6	Overall	5 Years LEP	310	4.52	21.29	50.97	23.23
6	Overall	6 Years LEP or More	1,607	3.98	14.87	52.21	28.94
6	Overall	Arabic	468	30.34	24.36	32.69	12.61
6	Overall	Bengali	304	18.42	23.36	40.13	18.09
6	Overall	Chinese	381	25.98	27.03	30.71	16.27
6	Overall	English	351	8.83	20.51	43.59	27.07
6	Overall	Spanish	9,033	15.02	22.34	43.68	18.95
6	Overall	Other Language	3,131	18.17	22.71	37.05	22.07
6	Overall	Students with Disabilities	3,461	9.53	30.86	45.88	13.72

H.8a: Grade 7 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
7	LS	Female	5,643	6.42	14.25	30.53	48.80
7	LS	Male	6,989	6.08	14.51	32.04	47.37
7	LS	American Indian or Alaska Native	19	10.53	21.05	21.05	47.37
7	LS	Black or African American	931	3.01	14.07	40.82	42.11
7	LS	Asian	2,413	8.66	24.41	32.57	34.36
7	LS	Native Hawaiian/Other Pacific Islander	57	5.26	28.07	35.09	31.58
7	LS	Hispanic or Latino	8,304	6.01	11.63	30.09	52.26
7	LS	White	903	5.09	12.29	30.01	52.60
7	LS	Multiracial (not Hispanic origin)	5	0.00	20.00	40.00	40.00
7	LS	NYC	8,976	6.78	14.87	31.66	46.68
7	LS	Big 4 Cities	895	5.59	17.54	37.65	39.22
7	LS	High Need Urban/Suburban	1,197	5.76	15.46	28.65	50.13
7	LS	High Need Rural	79	3.80	8.86	37.97	49.37
7	LS	Average Need	887	4.74	9.58	29.43	56.26
7	LS	Low Need	458	2.84	9.39	26.20	61.57
7	LS	Charter Schools	133	0.00	3.01	19.55	77.44
7	LS	0 Years LEP	7,315	6.96	12.33	29.71	51.00
7	LS	1 Years LEP	2,101	11.42	30.65	34.65	23.27
7	LS	2 Years LEP	1,094	2.01	17.64	40.68	39.67
7	LS	3 Years LEP	408	0.98	8.09	35.78	55.15
7	LS	4 Years LEP	277	1.08	4.33	34.66	59.93
7	LS	5 Years LEP	245	0.00	2.86	31.02	66.12
7	LS	6 Years LEP or More	1,158	0.43	1.73	24.78	73.06
7	LS	Arabic	433	7.85	21.94	33.26	36.95
7	LS	Bengali	305	1.31	17.70	36.39	44.59
7	LS	Chinese	373	13.67	28.69	30.83	26.81
7	LS	English	314	2.87	8.60	31.21	57.32
7	LS	Spanish	8,071	6.11	11.77	30.03	52.09
7	LS	Other Language	3,136	6.25	18.65	34.12	40.98
7	LS	Students with Disabilities	3,127	1.12	4.09	37.45	57.34

H.8b: Grade 7 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
7	RW	Female	5,643	20.31	30.69	28.94	20.06
7	RW	Male	6,989	24.62	34.93	26.38	14.06
7	RW	American Indian or Alaska Native	19	31.58	21.05	36.84	10.53
7	RW	Black or African American	931	21.16	34.05	27.07	17.72
7	RW	Asian	2,413	30.54	29.51	22.88	17.07
7	RW	Native Hawaiian/Other Pacific Islander	57	19.30	45.61	24.56	10.53
7	RW	Hispanic or Latino	8,304	20.70	34.27	29.03	15.99
7	RW	White	903	21.71	29.57	26.47	22.26
7	RW	Multiracial (not Hispanic origin)	5	20.00	20.00	40.00	20.00
7	RW	NYC	8,976	23.22	33.77	27.17	15.84
7	RW	Big 4 Cities	895	32.51	34.75	20.22	12.51
7	RW	High Need Urban/Suburban	1,197	21.72	31.24	30.49	16.54
7	RW	High Need Rural	79	16.46	29.11	37.97	16.46
7	RW	Average Need	887	17.14	29.99	30.44	22.44
7	RW	Low Need	458	12.45	28.82	30.79	27.95
7	RW	Charter Schools	133	5.26	24.81	37.59	32.33
7	RW	0 Years LEP	7,315	21.35	33.55	28.78	16.32
7	RW	1 Years LEP	2,101	40.31	32.94	15.80	10.95
7	RW	2 Years LEP	1,094	23.58	32.72	25.14	18.56
7	RW	3 Years LEP	408	12.01	34.07	33.33	20.59
7	RW	4 Years LEP	277	11.19	29.60	32.85	26.35
7	RW	5 Years LEP	245	9.80	32.24	34.69	23.27
7	RW	6 Years LEP or More	1,158	7.25	31.17	38.34	23.23
7	RW	Arabic	433	36.95	29.79	19.40	13.86
7	RW	Bengali	305	18.03	35.08	29.18	17.70
7	RW	Chinese	373	35.12	28.69	19.57	16.62
7	RW	English	314	17.52	31.53	27.71	23.25
7	RW	Spanish	8,071	20.90	34.20	29.13	15.77
7	RW	Other Language	3,136	24.84	30.96	25.29	18.91
7	RW	Students with Disabilities	3,127	16.28	43.59	29.65	10.49

H.8c: Grade 7 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
7	Overall	Female	5,643	20.68	30.80	30.78	17.74
7	Overall	Male	6,989	25.07	34.70	27.57	12.66
7	Overall	American Indian or Alaska Native	19	31.58	21.05	36.84	10.53
7	Overall	Black or African American	931	21.48	33.83	29.54	15.15
7	Overall	Asian	2,413	30.75	29.84	24.99	14.42
7	Overall	Native Hawaiian/Other Pacific Islander	57	22.81	42.11	26.32	8.77
7	Overall	Hispanic or Latino	8,304	21.16	34.14	30.25	14.45
7	Overall	White	903	22.15	29.24	27.57	21.04
7	Overall	Multiracial (not Hispanic origin)	5	20.00	20.00	0.00	60.00
7	Overall	NYC	8,976	23.69	33.69	28.71	13.91
7	Overall	Big 4 Cities	895	32.63	34.86	21.45	11.06
7	Overall	High Need Urban/Suburban	1,197	22.39	30.83	31.24	15.54
7	Overall	High Need Rural	79	16.46	29.11	39.24	15.19
7	Overall	Average Need	887	17.14	30.21	31.79	20.86
7	Overall	Low Need	458	12.66	28.60	33.19	25.55
7	Overall	Charter Schools	133	5.26	24.81	41.35	28.57
7	Overall	0 Years LEP	7,315	21.86	33.41	30.42	14.31
7	Overall	1 Years LEP	2,101	40.79	32.94	16.80	9.47
7	Overall	2 Years LEP	1,094	23.67	32.72	27.33	16.27
7	Overall	3 Years LEP	408	12.75	34.07	33.58	19.61
7	Overall	4 Years LEP	277	11.19	29.60	35.38	23.83
7	Overall	5 Years LEP	245	9.80	32.24	36.33	21.63
7	Overall	6 Years LEP or More	1,158	7.34	31.17	39.12	22.37
7	Overall	Arabic	433	37.88	29.10	20.09	12.93
7	Overall	Bengali	305	19.02	35.08	29.51	16.39
7	Overall	Chinese	373	35.39	28.95	21.98	13.67
7	Overall	English	314	17.52	31.53	30.89	20.06
7	Overall	Spanish	8,071	21.37	34.06	30.31	14.26
7	Overall	Other Language	3,136	25.03	31.06	27.49	16.42
7	Overall	Students with Disabilities	3,127	16.34	43.59	30.73	9.34

H.9a: Grade 8 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
8	LS	Female	5,576	7.68	17.27	25.68	49.37
8	LS	Male	7,086	7.93	16.29	25.25	50.54
8	LS	American Indian or Alaska Native	17	17.65	23.53	41.18	17.65
8	LS	Black or African American	991	5.55	16.35	31.58	46.52
8	LS	Asian	2,680	10.37	24.93	28.43	36.27
8	LS	Native Hawaiian/Other Pacific Islander	45	6.67	20.00	37.78	35.56
8	LS	Hispanic or Latino	8,042	7.51	14.01	23.71	54.76
8	LS	White	874	5.26	16.36	24.14	54.23
8	LS	Multiracial (not Hispanic origin)	13	7.69	30.77	30.77	30.77
8	LS	NYC	9,264	7.92	16.81	26.07	49.20
8	LS	Big 4 Cities	870	10.80	21.03	26.32	41.84
8	LS	High Need Urban/Suburban	1,176	8.76	16.41	22.79	52.04
8	LS	High Need Rural	64	3.13	12.50	23.44	60.94
8	LS	Average Need	816	5.15	15.69	23.16	56.00
8	LS	Low Need	391	3.84	11.00	23.27	61.89
8	LS	Charter Schools	81	0.00	6.17	17.28	76.54
8	LS	0 Years LEP	7,382	8.05	13.83	23.83	54.29
8	LS	1 Years LEP	2,161	14.85	32.90	29.75	22.49
8	LS	2 Years LEP	1,235	4.62	22.02	33.28	40.08
8	LS	3 Years LEP	442	2.04	12.90	31.67	53.39
8	LS	4 Years LEP	255	0.00	10.20	18.82	70.98
8	LS	5 Years LEP	232	0.00	3.45	20.69	75.86
8	LS	6 Years LEP or More	926	0.43	1.84	18.03	79.70
8	LS	Arabic	434	9.68	21.66	26.04	42.63
8	LS	Bengali	269	2.60	17.84	34.94	44.61
8	LS	Chinese	490	14.49	27.55	29.80	28.16
8	LS	English	270	5.19	9.26	27.78	57.78
8	LS	Spanish	7,843	7.60	14.15	23.75	54.49
8	LS	Other Language	3,356	7.75	21.01	27.71	43.53
8	LS	Students with Disabilities	2,724	1.32	6.09	29.30	63.29

H.9b: Grade 8 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
8	RW	Female	5,576	20.01	32.94	21.83	25.22
8	RW	Male	7,086	24.60	35.59	20.56	19.25
8	RW	American Indian or Alaska Native	17	41.18	29.41	23.53	5.88
8	RW	Black or African American	991	21.39	34.11	21.90	22.60
8	RW	Asian	2,680	26.83	34.03	18.25	20.90
8	RW	Native Hawaiian/Other Pacific Islander	45	17.78	40.00	15.56	26.67
8	RW	Hispanic or Latino	8,042	21.55	34.92	22.10	21.44
8	RW	White	874	20.02	31.69	20.14	28.15
8	RW	Multiracial (not Hispanic origin)	13	38.46	7.69	30.77	23.08
8	RW	NYC	9,264	22.43	35.58	21.05	20.94
8	RW	Big 4 Cities	870	32.87	35.75	17.36	14.02
8	RW	High Need Urban/Suburban	1,176	23.13	30.87	21.34	24.66
8	RW	High Need Rural	64	15.63	39.06	17.19	28.13
8	RW	Average Need	816	18.87	28.80	23.28	29.04
8	RW	Low Need	391	13.04	27.37	25.32	34.27
8	RW	Charter Schools	81	9.88	27.16	27.16	35.80
8	RW	0 Years LEP	7,382	20.58	34.62	22.50	22.30
8	RW	1 Years LEP	2,161	39.57	36.51	12.54	11.38
8	RW	2 Years LEP	1,235	24.21	35.71	18.46	21.62
8	RW	3 Years LEP	442	14.48	35.07	22.17	28.28
8	RW	4 Years LEP	255	14.51	27.45	20.39	37.65
8	RW	5 Years LEP	232	5.17	31.47	29.31	34.05
8	RW	6 Years LEP or More	926	6.80	29.05	31.32	32.83
8	RW	Arabic	434	31.80	31.80	15.90	20.51
8	RW	Bengali	269	14.87	38.29	18.22	28.62
8	RW	Chinese	490	31.63	34.29	17.35	16.73
8	RW	English	270	15.19	32.96	25.19	26.67
8	RW	Spanish	7,843	21.70	35.15	21.97	21.18
8	RW	Other Language	3,356	23.33	32.90	20.26	23.51
8	RW	Students with Disabilities	2,724	16.19	44.57	24.05	15.20

H.9c: Grade 8 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
8	Overall	Female	5,576	20.66	33.05	23.26	23.03
8	Overall	Male	7,086	25.12	35.41	21.78	17.70
8	Overall	American Indian or Alaska Native	17	41.18	29.41	29.41	0.00
8	Overall	Black or African American	991	21.59	34.51	23.92	19.98
8	Overall	Asian	2,680	27.43	34.33	19.55	18.69
8	Overall	Native Hawaiian/Other Pacific Islander	45	17.78	40.00	17.78	24.44
8	Overall	Hispanic or Latino	8,042	22.20	34.68	23.35	19.77
8	Overall	White	874	20.37	31.69	20.94	27.00
8	Overall	Multiracial (not Hispanic origin)	13	38.46	7.69	38.46	15.38
8	Overall	NYC	9,264	23.12	35.47	22.43	18.98
8	Overall	Big 4 Cities	870	33.22	35.63	17.93	13.22
8	Overall	High Need Urban/Suburban	1,176	23.38	30.95	23.13	22.53
8	Overall	High Need Rural	64	15.63	39.06	20.31	25.00
8	Overall	Average Need	816	19.24	28.92	24.14	27.70
8	Overall	Low Need	391	13.04	27.88	25.58	33.50
8	Overall	Charter Schools	81	9.88	27.16	29.63	33.33
8	Overall	0 Years LEP	7,382	21.17	34.57	23.87	20.39
8	Overall	1 Years LEP	2,161	40.77	35.82	13.60	9.81
8	Overall	2 Years LEP	1,235	24.29	36.36	19.92	19.43
8	Overall	3 Years LEP	442	14.93	35.52	23.53	26.02
8	Overall	4 Years LEP	255	14.51	27.84	20.39	37.25
8	Overall	5 Years LEP	232	5.17	31.47	30.60	32.76
8	Overall	6 Years LEP or More	926	6.80	29.16	32.83	31.21
8	Overall	Arabic	434	32.03	32.26	15.67	20.05
8	Overall	Bengali	269	15.24	38.29	20.82	25.65
8	Overall	Chinese	490	32.04	34.90	17.76	15.31
8	Overall	English	270	15.19	33.33	26.67	24.81
8	Overall	Spanish	7,843	22.38	34.91	23.26	19.46
8	Overall	Other Language	3,356	23.81	33.08	21.84	21.28
8	Overall	Students with Disabilities	2,724	16.30	44.71	25.11	13.88

H.10a: Grade 9 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
9	LS	Female	6,369	13.53	26.27	21.93	38.26
9	LS	Male	8,271	14.25	24.89	20.48	40.37
9	LS	American Indian or Alaska Native	29	17.24	31.03	20.69	31.03
9	LS	Black or African American	1,409	9.65	26.83	26.61	36.91
9	LS	Asian	3,220	15.50	34.84	23.17	26.49
9	LS	Native Hawaiian/Other Pacific Islander	35	11.43	45.71	25.71	17.14
9	LS	Hispanic or Latino	8,910	14.79	21.83	19.18	44.20
9	LS	White	1,026	7.60	25.34	23.68	43.37
9	LS	Multiracial (not Hispanic origin)	11	9.09	18.18	27.27	45.45
9	LS	NYC	10,213	14.28	25.19	21.14	39.39
9	LS	Big 4 Cities	1,097	12.31	32.18	20.24	35.28
9	LS	High Need Urban/Suburban	1,571	15.21	25.97	19.35	39.47
9	LS	High Need Rural	63	6.35	20.63	22.22	50.79
9	LS	Average Need	1,077	14.48	25.26	22.01	38.25
9	LS	Low Need	474	9.70	21.94	24.68	43.67
9	LS	Charter Schools	142	1.41	6.34	26.06	66.20
9	LS	0 Years LEP	8,125	14.93	21.92	19.66	43.50
9	LS	1 Years LEP	3,078	21.02	39.80	22.12	17.06
9	LS	2 Years LEP	1,563	8.32	33.53	26.42	31.73
9	LS	3 Years LEP	516	2.33	20.16	27.33	50.19
9	LS	4 Years LEP	283	2.47	15.90	26.15	55.48
9	LS	5 Years LEP	227	1.76	7.49	21.15	69.60
9	LS	6 Years LEP or More	811	2.71	3.08	15.66	78.55
9	LS	Arabic	536	11.75	30.97	28.36	28.92
9	LS	Bengali	396	11.87	34.85	23.48	29.80
9	LS	Chinese	490	17.35	34.90	24.69	23.06
9	LS	English	403	12.66	24.81	23.08	39.45
9	LS	Spanish	8,648	14.66	21.58	19.25	44.51
9	LS	Other Language	4,167	12.65	30.98	23.21	33.17
9	LS	Students with Disabilities	2,212	2.49	8.91	25.45	63.16

H.10b: Grade 9 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
9	RW	Female	6,369	20.24	44.61	21.75	13.41
9	RW	Male	8,271	23.08	43.31	19.55	14.06
9	RW	American Indian or Alaska Native	29	27.59	44.83	17.24	10.34
9	RW	Black or African American	1,409	19.16	48.33	20.16	12.35
9	RW	Asian	3,220	24.13	45.09	17.24	13.54
9	RW	Native Hawaiian/Other Pacific Islander	35	20.00	60.00	8.57	11.43
9	RW	Hispanic or Latino	8,910	21.89	42.69	21.85	13.57
9	RW	White	1,026	17.84	43.57	20.08	18.52
9	RW	Multiracial (not Hispanic origin)	11	27.27	45.45	18.18	9.09
9	RW	NYC	10,213	20.53	44.47	20.94	14.05
9	RW	Big 4 Cities	1,097	31.91	42.11	15.59	10.39
9	RW	High Need Urban/Suburban	1,571	24.57	43.79	19.92	11.71
9	RW	High Need Rural	63	19.05	44.44	23.81	12.70
9	RW	Average Need	1,077	25.07	41.78	20.06	13.09
9	RW	Low Need	474	15.82	38.82	22.57	22.78
9	RW	Charter Schools	142	4.93	47.89	28.87	18.31
9	RW	0 Years LEP	8,125	20.21	42.95	22.01	14.83
9	RW	1 Years LEP	3,078	33.11	46.39	12.96	7.54
9	RW	2 Years LEP	1,563	22.14	46.71	18.43	12.73
9	RW	3 Years LEP	516	12.79	48.06	20.93	18.22
9	RW	4 Years LEP	283	13.43	41.70	29.68	15.19
9	RW	5 Years LEP	227	8.81	41.41	29.07	20.70
9	RW	6 Years LEP or More	811	6.91	36.74	32.43	23.92
9	RW	Arabic	536	29.85	49.63	12.31	8.21
9	RW	Bengali	396	21.46	48.23	18.18	12.12
9	RW	Chinese	490	22.24	45.31	17.96	14.49
9	RW	English	403	21.09	43.42	19.11	16.38
9	RW	Spanish	8,648	21.67	42.75	21.96	13.62
9	RW	Other Language	4,167	21.24	44.92	19.20	14.64
9	RW	Students with Disabilities	2,212	9.67	50.90	26.99	12.43

H.10c: Grade 9 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
9	Overall	Female	6,369	23.13	42.88	22.09	11.90
9	Overall	Male	8,271	25.46	41.74	19.88	12.92
9	Overall	American Indian or Alaska Native	29	27.59	44.83	17.24	10.34
9	Overall	Black or African American	1,409	20.72	48.26	20.01	11.00
9	Overall	Asian	3,220	27.27	44.01	17.61	11.12
9	Overall	Native Hawaiian/Other Pacific Islander	35	22.86	60.00	11.43	5.71
9	Overall	Hispanic or Latino	8,910	24.65	40.43	22.21	12.72
9	Overall	White	1,026	18.91	43.37	20.66	17.06
9	Overall	Multiracial (not Hispanic origin)	11	27.27	45.45	18.18	9.09
9	Overall	NYC	10,213	23.51	42.52	21.37	12.59
9	Overall	Big 4 Cities	1,097	32.73	42.48	14.59	10.21
9	Overall	High Need Urban/Suburban	1,571	27.24	41.50	20.24	11.01
9	Overall	High Need Rural	63	20.63	42.86	26.98	9.52
9	Overall	Average Need	1,077	27.02	40.67	19.78	12.53
9	Overall	Low Need	474	16.46	39.87	24.47	19.20
9	Overall	Charter Schools	142	5.63	47.18	30.99	16.20
9	Overall	0 Years LEP	8,125	23.18	40.90	22.40	13.53
9	Overall	1 Years LEP	3,078	37.10	43.99	12.74	6.17
9	Overall	2 Years LEP	1,563	22.97	46.71	19.26	11.07
9	Overall	3 Years LEP	516	12.79	48.06	22.29	16.86
9	Overall	4 Years LEP	283	13.43	42.40	29.68	14.49
9	Overall	5 Years LEP	227	8.81	41.85	29.52	19.82
9	Overall	6 Years LEP or More	811	7.40	36.50	32.80	23.30
9	Overall	Arabic	536	30.78	48.69	12.69	7.84
9	Overall	Bengali	396	23.23	48.48	17.93	10.35
9	Overall	Chinese	490	25.92	43.88	19.18	11.02
9	Overall	English	403	23.57	41.94	19.35	15.14
9	Overall	Spanish	8,648	24.43	40.47	22.33	12.77
9	Overall	Other Language	4,167	23.69	44.30	19.41	12.60
9	Overall	Students with Disabilities	2,212	10.04	50.72	27.49	11.75

H.11a: Grade 10 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
10	LS	Female	6,099	8.71	31.32	24.61	35.37
10	LS	Male	7,177	9.81	28.65	23.81	37.73
10	LS	American Indian or Alaska Native	24	8.33	16.67	29.17	45.83
10	LS	Black or African American	1,206	3.98	25.62	30.35	40.05
10	LS	Asian	3,806	11.80	40.91	25.04	22.25
10	LS	Native Hawaiian/Other Pacific Islander	41	9.76	34.15	34.15	21.95
10	LS	Hispanic or Latino	7,379	9.45	25.41	22.33	42.81
10	LS	White	810	4.32	25.31	27.04	43.33
10	LS	Multiracial (not Hispanic origin)	10	0.00	20.00	30.00	50.00
10	LS	NYC	9,946	10.08	31.32	23.70	34.90
10	LS	Big 4 Cities	676	10.65	31.36	23.96	34.02
10	LS	High Need Urban/Suburban	1,151	7.99	26.67	25.02	40.31
10	LS	High Need Rural	66	4.55	18.18	31.82	45.45
10	LS	Average Need	872	5.62	23.28	24.20	46.90
10	LS	Low Need	489	3.27	22.70	31.49	42.54
10	LS	Charter Schools	75	0.00	6.67	22.67	70.67
10	LS	0 Years LEP	6,904	9.66	25.61	22.64	42.09
10	LS	1 Years LEP	3,172	13.37	43.03	25.09	18.51
10	LS	2 Years LEP	1,629	6.08	36.10	28.85	28.97
10	LS	3 Years LEP	461	2.60	27.55	32.32	37.53
10	LS	4 Years LEP	305	2.62	17.05	25.90	54.43
10	LS	5 Years LEP	193	2.07	12.95	25.91	59.07
10	LS	6 Years LEP or More	581	2.58	5.51	16.35	75.56
10	LS	Arabic	342	2.92	31.58	31.87	33.63
10	LS	Bengali	396	5.81	38.64	25.76	29.80
10	LS	Chinese	785	14.78	43.69	23.69	17.83
10	LS	English	318	5.66	20.13	23.90	50.31
10	LS	Spanish	7,208	9.56	25.57	22.29	42.58
10	LS	Other Language	4,227	8.97	34.42	26.73	29.88
10	LS	Students with Disabilities	1,416	2.26	11.30	25.78	60.66

H.11b: Grade 10 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
10	RW	Female	6,099	13.87	49.27	23.40	13.46
10	RW	Male	7,177	16.50	45.87	22.95	14.69
10	RW	American Indian or Alaska Native	24	12.50	54.17	16.67	16.67
10	RW	Black or African American	1,206	9.54	50.41	24.30	15.75
10	RW	Asian	3,806	17.42	50.76	19.68	12.14
10	RW	Native Hawaiian/Other Pacific Islander	41	17.07	48.78	24.39	9.76
10	RW	Hispanic or Latino	7,379	15.52	45.51	24.47	14.50
10	RW	White	810	11.73	44.57	26.05	17.65
10	RW	Multiracial (not Hispanic origin)	10	20.00	50.00	10.00	20.00
10	RW	NYC	9,946	15.38	47.97	23.27	13.38
10	RW	Big 4 Cities	676	22.34	51.18	17.01	9.47
10	RW	High Need Urban/Suburban	1,151	16.68	48.31	22.68	12.34
10	RW	High Need Rural	66	12.12	43.94	31.82	12.12
10	RW	Average Need	872	12.04	41.28	25.92	20.76
10	RW	Low Need	489	8.18	42.74	23.52	25.56
10	RW	Charter Schools	75	4.00	34.67	29.33	32.00
10	RW	0 Years LEP	6,904	14.11	44.86	25.58	15.45
10	RW	1 Years LEP	3,172	20.90	52.81	16.96	9.33
10	RW	2 Years LEP	1,629	14.49	52.12	20.38	13.01
10	RW	3 Years LEP	461	13.45	51.84	19.09	15.62
10	RW	4 Years LEP	305	9.84	42.95	26.56	20.66
10	RW	5 Years LEP	193	8.29	44.56	27.98	19.17
10	RW	6 Years LEP or More	581	6.88	35.63	35.97	21.51
10	RW	Arabic	342	16.08	51.75	21.93	10.23
10	RW	Bengali	396	12.63	56.31	22.73	8.33
10	RW	Chinese	785	18.34	52.10	16.05	13.50
10	RW	English	318	8.49	47.17	28.62	15.72
10	RW	Spanish	7,208	15.75	45.49	24.31	14.46
10	RW	Other Language	4,227	14.64	48.71	22.24	14.41
10	RW	Students with Disabilities	1,416	11.16	48.38	29.10	11.37

H.11c: Grade 10 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
10	Overall	Female	6,099	16.30	49.35	22.97	11.38
10	Overall	Male	7,177	18.88	45.65	23.00	12.47
10	Overall	American Indian or Alaska Native	24	12.50	54.17	16.67	16.67
10	Overall	Black or African American	1,206	10.53	51.08	24.96	13.43
10	Overall	Asian	3,806	20.49	51.94	19.08	8.49
10	Overall	Native Hawaiian/Other Pacific Islander	41	19.51	48.78	24.39	7.32
10	Overall	Hispanic or Latino	7,379	18.02	44.48	24.38	13.12
10	Overall	White	810	12.22	46.05	26.05	15.68
10	Overall	Multiracial (not Hispanic origin)	10	20.00	50.00	10.00	20.00
10	Overall	NYC	9,946	18.19	47.90	22.80	11.11
10	Overall	Big 4 Cities	676	24.11	50.15	17.01	8.73
10	Overall	High Need Urban/Suburban	1,151	18.25	48.05	22.24	11.47
10	Overall	High Need Rural	66	13.64	42.42	33.33	10.61
10	Overall	Average Need	872	12.73	41.63	26.72	18.92
10	Overall	Low Need	489	8.79	43.56	27.20	20.45
10	Overall	Charter Schools	75	4.00	34.67	33.33	28.00
10	Overall	0 Years LEP	6,904	16.58	44.87	25.25	13.30
10	Overall	1 Years LEP	3,172	24.62	52.14	16.61	6.62
10	Overall	2 Years LEP	1,629	16.08	52.30	20.63	10.99
10	Overall	3 Years LEP	461	13.67	52.28	19.74	14.32
10	Overall	4 Years LEP	305	9.84	44.26	27.54	18.36
10	Overall	5 Years LEP	193	8.29	45.08	27.98	18.65
10	Overall	6 Years LEP or More	581	7.23	35.46	36.49	20.83
10	Overall	Arabic	342	16.67	52.34	22.51	8.48
10	Overall	Bengali	396	15.40	55.81	21.97	6.82
10	Overall	Chinese	785	22.29	53.38	16.05	8.28
10	Overall	English	318	10.38	46.86	28.30	14.47
10	Overall	Spanish	7,208	18.24	44.49	24.15	13.11
10	Overall	Other Language	4,227	16.75	49.94	22.03	11.28
10	Overall	Students with Disabilities	1,416	11.44	48.59	29.38	10.59

H.12a: Grade 11 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
11	LS	Female	4,306	5.74	29.31	31.03	33.93
11	LS	Male	4,566	4.84	27.14	32.41	35.61
11	LS	American Indian or Alaska Native	15	0.00	46.67	26.67	26.67
11	LS	Black or African American	836	2.27	26.56	35.05	36.12
11	LS	Asian	2,402	3.41	38.01	34.89	23.69
11	LS	Native Hawaiian/Other Pacific Islander	28	7.14	21.43	32.14	39.29
11	LS	Hispanic or Latino	4,992	7.01	24.40	29.63	38.96
11	LS	White	591	2.54	22.50	32.15	42.81
11	LS	Multiracial (not Hispanic origin)	8	0.00	25.00	37.50	37.50
11	LS	NYC	6,276	5.27	28.54	32.17	34.02
11	LS	Big 4 Cities	485	8.66	37.94	24.54	28.87
11	LS	High Need Urban/Suburban	882	5.22	24.72	33.45	36.62
11	LS	High Need Rural	41	2.44	14.63	43.90	39.02
11	LS	Average Need	707	3.54	24.05	31.82	40.59
11	LS	Low Need	418	5.26	28.71	30.38	35.65
11	LS	Charter Schools	62	1.61	17.74	20.97	59.68
11	LS	0 Years LEP	4,178	5.34	22.88	31.14	40.64
11	LS	1 Years LEP	1,760	8.47	44.09	29.20	18.24
11	LS	2 Years LEP	1,471	4.49	33.38	34.94	27.19
11	LS	3 Years LEP	463	2.38	29.16	38.23	30.24
11	LS	4 Years LEP	301	2.66	22.92	37.87	36.54
11	LS	5 Years LEP	222	0.90	14.86	37.84	46.40
11	LS	6 Years LEP or More	456	1.54	7.68	23.03	67.76
11	LS	Arabic	215	1.40	24.19	33.02	41.40
11	LS	Bengali	229	3.49	33.19	32.75	30.57
11	LS	Chinese	427	2.58	45.90	33.02	18.50
11	LS	English	214	3.74	17.29	37.85	41.12
11	LS	Spanish	4,886	6.94	24.62	29.47	38.97
11	LS	Other Language	2,901	3.41	32.30	34.75	29.54
11	LS	Students with Disabilities	851	2.23	14.10	32.55	51.12

H.12b: Grade 11 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
11	RW	Female	4,306	8.69	52.55	23.29	15.47
11	RW	Male	4,566	9.90	50.79	24.40	14.91
11	RW	American Indian or Alaska Native	15	13.33	53.33	13.33	20.00
11	RW	Black or African American	836	6.82	54.55	21.65	16.99
11	RW	Asian	2,402	8.20	53.41	23.27	15.11
11	RW	Native Hawaiian/Other Pacific Islander	28	7.14	53.57	17.86	21.43
11	RW	Hispanic or Latino	4,992	10.60	50.96	24.24	14.20
11	RW	White	591	6.43	45.85	26.90	20.81
11	RW	Multiracial (not Hispanic origin)	8	12.50	62.50	12.50	12.50
11	RW	NYC	6,276	8.49	51.77	24.71	15.03
11	RW	Big 4 Cities	485	21.03	52.37	16.91	9.69
11	RW	High Need Urban/Suburban	882	11.79	56.12	19.95	12.13
11	RW	High Need Rural	41	9.76	53.66	17.07	19.51
11	RW	Average Need	707	6.79	48.09	26.31	18.81
11	RW	Low Need	418	7.66	45.45	22.73	24.16
11	RW	Charter Schools	62	4.84	50.00	32.26	12.90
11	RW	0 Years LEP	4,178	8.35	48.54	26.78	16.32
11	RW	1 Years LEP	1,760	13.47	56.48	18.98	11.08
11	RW	2 Years LEP	1,471	8.50	56.90	19.78	14.82
11	RW	3 Years LEP	463	9.29	54.86	20.30	15.55
11	RW	4 Years LEP	301	10.96	48.84	24.58	15.61
11	RW	5 Years LEP	222	5.86	49.10	29.73	15.32
11	RW	6 Years LEP or More	456	5.04	43.86	29.82	21.27
11	RW	Arabic	215	6.98	55.81	23.26	13.95
11	RW	Bengali	229	8.73	61.14	20.96	9.17
11	RW	Chinese	427	6.32	51.99	26.23	15.46
11	RW	English	214	6.54	47.66	25.70	20.09
11	RW	Spanish	4,886	10.64	50.94	24.17	14.24
11	RW	Other Language	2,901	7.93	52.02	23.13	16.93
11	RW	Students with Disabilities	851	10.81	54.41	25.15	9.64

H.12c: Grade 11 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
11	Overall	Female	4,306	10.73	52.86	24.29	12.12
11	Overall	Male	4,566	11.39	51.60	25.49	11.52
11	Overall	American Indian or Alaska Native	15	13.33	53.33	13.33	20.00
11	Overall	Black or African American	836	7.54	55.14	23.92	13.40
11	Overall	Asian	2,402	9.49	56.58	24.23	9.70
11	Overall	Native Hawaiian/Other Pacific Islander	28	7.14	53.57	25.00	14.29
11	Overall	Hispanic or Latino	4,992	12.84	50.26	24.98	11.92
11	Overall	White	591	7.61	46.53	28.93	16.92
11	Overall	Multiracial (not Hispanic origin)	8	12.50	62.50	12.50	12.50
11	Overall	NYC	6,276	10.47	52.41	25.57	11.55
11	Overall	Big 4 Cities	485	21.86	53.40	16.49	8.25
11	Overall	High Need Urban/Suburban	882	13.27	55.67	20.52	10.54
11	Overall	High Need Rural	41	9.76	53.66	19.51	17.07
11	Overall	Average Need	707	7.92	49.22	27.86	14.99
11	Overall	Low Need	418	9.33	45.45	28.23	16.99
11	Overall	Charter Schools	62	4.84	51.61	33.87	9.68
11	Overall	0 Years LEP	4,178	10.12	48.88	27.84	13.16
11	Overall	1 Years LEP	1,760	15.91	57.73	19.38	6.99
11	Overall	2 Years LEP	1,471	10.81	57.10	20.73	11.35
11	Overall	3 Years LEP	463	9.94	56.16	23.33	10.58
11	Overall	4 Years LEP	301	11.63	49.83	26.58	11.96
11	Overall	5 Years LEP	222	5.86	49.55	31.53	13.06
11	Overall	6 Years LEP or More	456	5.04	44.08	30.48	20.39
11	Overall	Arabic	215	7.44	57.21	24.19	11.16
11	Overall	Bengali	229	9.61	60.26	21.40	8.73
11	Overall	Chinese	427	7.96	58.55	23.19	10.30
11	Overall	English	214	7.01	48.60	29.44	14.95
11	Overall	Spanish	4,886	12.91	50.20	24.89	11.99
11	Overall	Other Language	2,901	9.10	53.91	25.20	11.79
11	Overall	Students with Disabilities	851	11.05	54.88	25.62	8.46

H.13a: Grade 12 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
12	LS	Female	3,246	3.02	29.70	23.94	43.35
12	LS	Male	3,216	2.49	28.02	24.04	45.46
12	LS	American Indian or Alaska Native	7	0.00	28.57	57.14	14.29
12	LS	Black or African American	626	1.28	26.04	24.76	47.92
12	LS	Asian	1,744	2.06	37.84	27.52	32.57
12	LS	Native Hawaiian/Other Pacific Islander	9	0.00	22.22	22.22	55.56
12	LS	Hispanic or Latino	3,672	3.51	26.06	22.33	48.09
12	LS	White	403	1.24	20.10	22.08	56.58
12	LS	Multiracial (not Hispanic origin)	1	0.00	0.00	0.00	100.00
12	LS	NYC	4,766	2.71	31.28	23.63	42.38
12	LS	Big 4 Cities	345	7.25	32.17	25.80	34.78
12	LS	High Need Urban/Suburban	517	2.51	22.05	25.53	49.90
12	LS	High Need Rural	38	2.63	18.42	18.42	60.53
12	LS	Average Need	488	1.02	18.03	24.59	56.35
12	LS	Low Need	293	1.71	17.75	24.91	55.63
12	LS	Charter Schools	15	0.00	13.33	20.00	66.67
12	LS	0 Years LEP	3,530	2.18	26.69	23.09	48.05
12	LS	1 Years LEP	847	6.49	44.75	20.78	27.98
12	LS	2 Years LEP	898	2.56	37.42	28.17	31.85
12	LS	3 Years LEP	253	2.37	28.46	28.06	41.11
12	LS	4 Years LEP	307	1.63	19.54	30.62	48.21
12	LS	5 Years LEP	199	1.01	18.59	22.11	58.29
12	LS	6 Years LEP or More	419	2.15	9.31	22.67	65.87
12	LS	Arabic	171	2.92	25.15	22.81	49.12
12	LS	Bengali	211	0.47	38.86	22.27	38.39
12	LS	Chinese	236	3.39	39.83	28.39	28.39
12	LS	English	131	0.76	18.32	18.32	62.60
12	LS	Spanish	3,589	3.59	26.14	22.32	47.95
12	LS	Other Language	2,124	1.60	32.20	26.93	39.27
12	LS	Students with Disabilities	680	2.06	22.21	23.53	52.21

H.13b: Grade 12 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
12	RW	Female	3,246	4.74	53.02	23.69	18.55
12	RW	Male	3,216	5.91	51.62	23.10	19.37
12	RW	American Indian or Alaska Native	7	0.00	57.14	42.86	0.00
12	RW	Black or African American	626	4.79	54.31	19.33	21.57
12	RW	Asian	1,744	4.36	52.69	23.05	19.90
12	RW	Native Hawaiian/Other Pacific Islander	9	0.00	55.56	11.11	33.33
12	RW	Hispanic or Latino	3,672	6.02	52.48	24.05	17.46
12	RW	White	403	4.22	46.15	25.06	24.57
12	RW	Multiracial (not Hispanic origin)	1	0.00	0.00	100.00	0.00
12	RW	NYC	4,766	5.01	53.34	22.79	18.86
12	RW	Big 4 Cities	345	14.20	56.52	19.71	9.57
12	RW	High Need Urban/Suburban	517	6.58	54.16	23.21	16.05
12	RW	High Need Rural	38	7.89	50.00	23.68	18.42
12	RW	Average Need	488	2.66	44.26	29.10	23.98
12	RW	Low Need	293	1.71	42.32	28.67	27.30
12	RW	Charter Schools	15	6.67	33.33	20.00	40.00
12	RW	0 Years LEP	3,530	4.99	51.44	23.91	19.66
12	RW	1 Years LEP	847	7.32	57.73	18.77	16.17
12	RW	2 Years LEP	898	4.12	54.01	21.60	20.27
12	RW	3 Years LEP	253	6.32	49.80	26.09	17.79
12	RW	4 Years LEP	307	4.56	49.19	27.36	18.89
12	RW	5 Years LEP	199	5.03	50.75	28.14	16.08
12	RW	6 Years LEP or More	419	6.68	49.88	25.78	17.66
12	RW	Arabic	171	7.02	53.80	21.05	18.13
12	RW	Bengali	211	4.27	64.45	16.59	14.69
12	RW	Chinese	236	4.24	49.58	22.88	23.31
12	RW	English	131	3.05	51.15	23.66	22.14
12	RW	Spanish	3,589	5.99	52.35	24.05	17.61
12	RW	Other Language	2,124	4.43	51.32	23.21	21.05
12	RW	Students with Disabilities	680	11.03	55.59	22.35	11.03

H.13c: Grade 12 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
12	Overall	Female	3,246	5.85	55.67	24.31	14.17
12	Overall	Male	3,216	6.84	54.48	22.85	15.83
12	Overall	American Indian or Alaska Native	7	0.00	57.14	42.86	0.00
12	Overall	Black or African American	626	5.11	56.55	20.77	17.57
12	Overall	Asian	1,744	5.05	59.63	22.08	13.25
12	Overall	Native Hawaiian/Other Pacific Islander	9	0.00	55.56	11.11	33.33
12	Overall	Hispanic or Latino	3,672	7.41	53.51	24.48	14.60
12	Overall	White	403	4.47	47.39	26.05	22.08
12	Overall	Multiracial (not Hispanic origin)	1	0.00	0.00	100.0 0	0.00
12	Overall	NYC	4,766	6.23	56.36	22.81	14.60
12	Overall	Big 4 Cities	345	15.07	57.97	19.13	7.83
12	Overall	High Need Urban/Suburban	517	6.96	54.93	24.56	13.54
12	Overall	High Need Rural	38	7.89	52.63	23.68	15.79
12	Overall	Average Need	488	2.87	47.54	29.10	20.49
12	Overall	Low Need	293	2.39	45.05	30.38	22.18
12	Overall	Charter Schools	15	6.67	33.33	26.67	33.33
12	Overall	0 Years LEP	3,530	5.86	53.65	24.48	16.01
12	Overall	1 Years LEP	847	10.27	61.75	17.00	10.98
12	Overall	2 Years LEP	898	5.12	59.24	22.16	13.47
12	Overall	3 Years LEP	253	6.72	52.57	26.88	13.83
12	Overall	4 Years LEP	307	4.56	51.14	28.01	16.29
12	Overall	5 Years LEP	199	5.03	51.26	28.14	15.58
12	Overall	6 Years LEP or More	419	6.68	51.07	25.30	16.95
12	Overall	Arabic	171	7.60	54.39	22.22	15.79
12	Overall	Bengali	211	4.74	67.30	14.22	13.74
12	Overall	Chinese	236	5.93	56.78	23.31	13.98
12	Overall	English	131	3.05	52.67	24.43	19.85
12	Overall	Spanish	3,589	7.41	53.39	24.41	14.80
12	Overall	Other Language	2,124	4.85	56.73	23.21	15.21
12	Overall	Students with Disabilities	680	11.47	56.32	22.94	9.26

APPENDIX I: EXIT RATE BY SUBGROUP**I.1: Kindergarten**

Grade	Group	N-count	Percent
K	Female	1,296	10.72
K	Male	1,126	8.44
K	American Indian or Alaska Native	5	7.04
K	Black or African American	90	8.67
K	Asian	779	14.50
K	Native Hawaiian/Other Pacific Islander	13	11.30
K	Hispanic or Latino	1,210	7.37
K	White	321	13.45
K	Multiracial (not Hispanic origin)	5	15.15
K	NYC	1,774	11.20
K	Big 4 Cities	31	2.86
K	High Need Urban/Suburban	209	6.18
K	High Need Rural	8	4.10
K	Average Need	216	7.10
K	Low Need	105	8.03
K	Charter Schools	79	13.74
K	0 Years LEP	1,686	10.86
K	1 Years LEP	717	7.62
K	Arabic	71	10.81
K	Bengali	150	21.68
K	Chinese	209	14.78
K	English	127	7.30
K	Spanish	1,126	7.43
K	Other Language	738	12.79
K	Students with Disabilities	112	4.12

I.2: Grade 1

Grade	Group	N-count	Percent
1	Female	3,067	25.48
1	Male	2,631	19.30
1	American Indian or Alaska Native	10	20.83
1	Black or African American	221	18.65
1	Asian	1,594	29.17
1	Native Hawaiian/Other Pacific Islander	11	12.79
1	Hispanic or Latino	3,176	19.13
1	White	680	30.17
1	Multiracial (not Hispanic origin)	6	19.35
1	NYC	3,591	22.22
1	Big 4 Cities	145	11.63
1	High Need Urban/Suburban	633	19.30
1	High Need Rural	25	14.12
1	Average Need	778	26.51
1	Low Need	414	30.15
1	Charter Schools	111	23.03
1	0 Years LEP	3,052	21.25
1	1 Years LEP	697	19.02
1	2 Years LEP	1,865	26.09
1	Arabic	134	19.09
1	Bengali	230	32.17
1	Chinese	372	30.24
1	English	311	24.33
1	Spanish	2,966	18.86
1	Other Language	1,686	28.01
1	Students with Disabilities	357	10.33

I.3: Grade 2

Grade	Group	N-count	Percent
2	Female	1,412	14.59
2	Male	1,343	11.85
2	American Indian or Alaska Native	8	23.53
2	Black or African American	118	9.93
2	Asian	833	20.52
2	Native Hawaiian/Other Pacific Islander	10	16.39
2	Hispanic or Latino	1,412	10.13
2	White	370	21.65
2	Multiracial (not Hispanic origin)	4	18.18
2	NYC	1,582	11.78
2	Big 4 Cities	90	7.69
2	High Need Urban/Suburban	343	12.53
2	High Need Rural	19	13.97
2	Average Need	424	18.86
2	Low Need	254	26.02
2	Charter Schools	43	13.83
2	0 Years LEP	1,288	10.95
2	1 Years LEP	262	11.42
2	2 Years LEP	372	16.44
2	3 Years LEP	782	18.47
2	Arabic	67	10.15
2	Bengali	140	24.18
2	Chinese	168	22.43
2	English	159	18.86
2	Spanish	1,313	9.84
2	Other Language	908	18.76
2	Students with Disabilities	148	4.12

I.4: Grade 3

Grade	Group	N-count	Percent
3	Female	1,799	19.61
3	Male	1,716	15.65
3	American Indian or Alaska Native	7	18.92
3	Black or African American	194	16.60
3	Asian	896	24.37
3	Native Hawaiian/Other Pacific Islander	9	20.93
3	Hispanic or Latino	2,059	15.00
3	White	346	23.54
3	Multiracial (not Hispanic origin)	4	21.05
3	NYC	2,211	16.64
3	Big 4 Cities	145	12.30
3	High Need Urban/Suburban	426	17.18
3	High Need Rural	19	16.24
3	Average Need	430	21.75
3	Low Need	230	27.12
3	Charter Schools	53	21.81
3	0 Years LEP	1,950	16.31
3	1 Years LEP	264	12.80
3	2 Years LEP	268	19.90
3	3 Years LEP	237	21.76
3	4 Years LEP	728	22.46
3	Arabic	92	15.31
3	Bengali	151	27.21
3	Chinese	205	29.67
3	English	159	20.98
3	Spanish	1,953	14.84
3	Other Language	956	21.84
3	Students with Disabilities	269	6.54

I.5: Grade 4

Grade	Group	N-count	Percent
4	Female	2,169	25.77
4	Male	2,147	21.73
4	American Indian or Alaska Native	8	25.81
4	Black or African American	229	20.30
4	Asian	965	29.41
4	Native Hawaiian/Other Pacific Islander	7	14.58
4	Hispanic or Latino	2,703	21.72
4	White	399	29.51
4	Multiracial (not Hispanic origin)	5	35.71
4	NYC	2,804	22.76
4	Big 4 Cities	195	16.61
4	High Need Urban/Suburban	564	26.00
4	High Need Rural	22	20.18
4	Average Need	466	27.85
4	Low Need	231	33.14
4	Charter Schools	33	21.71
4	0 Years LEP	2,514	23.04
4	1 Years LEP	254	12.77
4	2 Years LEP	295	24.00
4	3 Years LEP	146	24.66
4	4 Years LEP	238	29.60
4	5 Years LEP	789	32.88
4	Arabic	104	16.94
4	Bengali	181	35.63
4	Chinese	161	29.65
4	English	171	28.13
4	Spanish	2,550	21.32
4	Other Language	1,149	28.27
4	Students with Disabilities	504	12.09

I.6: Grade 5

Grade	Group	N-count	Percent
5	Female	2,091	29.86
5	Male	2,294	25.98
5	American Indian or Alaska Native	7	33.33
5	Black or African American	223	20.98
5	Asian	839	29.51
5	Native Hawaiian/Other Pacific Islander	13	24.07
5	Hispanic or Latino	2,968	27.69
5	White	330	29.41
5	Multiracial (not Hispanic origin)	5	38.46
5	NYC	2,902	27.14
5	Big 4 Cities	161	15.69
5	High Need Urban/Suburban	573	31.80
5	High Need Rural	28	23.53
5	Average Need	415	31.09
5	Low Need	220	38.73
5	Charter Schools	83	29.64
5	0 Years LEP	2,632	28.16
5	1 Years LEP	274	14.29
5	2 Years LEP	310	27.56
5	3 Years LEP	163	30.64
5	4 Years LEP	106	28.80
5	5 Years LEP	211	34.59
5	6 Years LEP or More	686	35.86
5	Arabic	98	19.07
5	Bengali	126	33.07
5	Chinese	127	27.55
5	English	146	30.54
5	Spanish	2,844	27.48
5	Other Language	1,044	28.61
5	Students with Disabilities	678	17.39

I.7: Grade 6

Grade	Group	N-count	Percent
6	Female	1,240	21.07
6	Male	1,434	18.42
6	American Indian or Alaska Native	6	18.75
6	Black or African American	166	16.47
6	Asian	463	20.08
6	Native Hawaiian/Other Pacific Islander	7	15.56
6	Hispanic or Latino	1,802	19.30
6	White	227	24.54
6	Multiracial (not Hispanic origin)	3	17.65
6	NYC	1,729	18.31
6	Big 4 Cities	141	14.55
6	High Need Urban/Suburban	336	23.05
6	High Need Rural	26	27.37
6	Average Need	251	23.52
6	Low Need	137	32.31
6	Charter Schools	54	25.59
6	0 Years LEP	1,548	19.21
6	1 Years LEP	199	10.77
6	2 Years LEP	199	18.51
6	3 Years LEP	110	26.19
6	4 Years LEP	75	23.58
6	5 Years LEP	72	23.23
6	6 Years LEP or More	465	28.94
6	Arabic	59	12.61
6	Bengali	55	18.09
6	Chinese	62	16.27
6	English	95	27.07
6	Spanish	1,712	18.95
6	Other Language	691	22.07
6	Students with Disabilities	475	13.72

I.8: Grade 7

Grade	Group	N-count	Percent
7	Female	1,001	17.74
7	Male	885	12.66
7	American Indian or Alaska Native	2	10.53
7	Black or African American	141	15.15
7	Asian	348	14.42
7	Native Hawaiian/Other Pacific Islander	5	8.77
7	Hispanic or Latino	1,200	14.45
7	White	190	21.04
7	Multiracial (not Hispanic origin)	3	60.00
7	NYC	1,249	13.91
7	Big 4 Cities	99	11.06
7	High Need Urban/Suburban	186	15.54
7	High Need Rural	12	15.19
7	Average Need	185	20.86
7	Low Need	117	25.55
7	Charter Schools	38	28.57
7	0 Years LEP	1,047	14.31
7	1 Years LEP	199	9.47
7	2 Years LEP	178	16.27
7	3 Years LEP	80	19.61
7	4 Years LEP	66	23.83
7	5 Years LEP	53	21.63
7	6 Years LEP or More	259	22.37
7	Arabic	56	12.93
7	Bengali	50	16.39
7	Chinese	51	13.67
7	English	63	20.06
7	Spanish	1,151	14.26
7	Other Language	515	16.42
7	Students with Disabilities	292	9.34

I.9: Grade 8

Grade	Group	N-count	Percent
8	Female	1,284	23.03
8	Male	1,254	17.70
8	American Indian or Alaska Native	0	0.00
8	Black or African American	198	19.98
8	Asian	501	18.69
8	Native Hawaiian/Other Pacific Islander	11	24.44
8	Hispanic or Latino	1,590	19.77
8	White	236	27.00
8	Multiracial (not Hispanic origin)	2	15.38
8	NYC	1,758	18.98
8	Big 4 Cities	115	13.22
8	High Need Urban/Suburban	265	22.53
8	High Need Rural	16	25.00
8	Average Need	226	27.70
8	Low Need	131	33.50
8	Charter Schools	27	33.33
8	0 Years LEP	1,505	20.39
8	1 Years LEP	212	9.81
8	2 Years LEP	240	19.43
8	3 Years LEP	115	26.02
8	4 Years LEP	95	37.25
8	5 Years LEP	76	32.76
8	6 Years LEP or More	289	31.21
8	Arabic	87	20.05
8	Bengali	69	25.65
8	Chinese	75	15.31
8	English	67	24.81
8	Spanish	1,526	19.46
8	Other Language	714	21.28
8	Students with Disabilities	378	13.88

I.10: Grade 9

Grade	Group	N-count	Percent
9	Female	758	11.90
9	Male	1,069	12.92
9	American Indian or Alaska Native	3	10.34
9	Black or African American	155	11.00
9	Asian	358	11.12
9	Native Hawaiian/Other Pacific Islander	2	5.71
9	Hispanic or Latino	1,133	12.72
9	White	175	17.06
9	Multiracial (not Hispanic origin)	1	9.09
9	NYC	1,286	12.59
9	Big 4 Cities	112	10.21
9	High Need Urban/Suburban	173	11.01
9	High Need Rural	6	9.52
9	Average Need	135	12.53
9	Low Need	91	19.20
9	Charter Schools	23	16.20
9	0 Years LEP	1,099	13.53
9	1 Years LEP	190	6.17
9	2 Years LEP	173	11.07
9	3 Years LEP	87	16.86
9	4 Years LEP	41	14.49
9	5 Years LEP	45	19.82
9	6 Years LEP or More	189	23.30
9	Arabic	42	7.84
9	Bengali	41	10.35
9	Chinese	54	11.02
9	English	61	15.14
9	Spanish	1,104	12.77
9	Other Language	525	12.60
9	Students with Disabilities	260	11.75

I.11: Grade 10

Grade	Group	N-count	Percent
10	Female	694	11.38
10	Male	895	12.47
10	American Indian or Alaska Native	4	16.67
10	Black or African American	162	13.43
10	Asian	323	8.49
10	Native Hawaiian/Other Pacific Islander	3	7.32
10	Hispanic or Latino	968	13.12
10	White	127	15.68
10	Multiracial (not Hispanic origin)	2	20.00
10	NYC	1,105	11.11
10	Big 4 Cities	59	8.73
10	High Need Urban/Suburban	132	11.47
10	High Need Rural	7	10.61
10	Average Need	165	18.92
10	Low Need	100	20.45
10	Charter Schools	21	28.00
10	0 Years LEP	918	13.30
10	1 Years LEP	210	6.62
10	2 Years LEP	179	10.99
10	3 Years LEP	66	14.32
10	4 Years LEP	56	18.36
10	5 Years LEP	36	18.65
10	6 Years LEP or More	121	20.83
10	Arabic	29	8.48
10	Bengali	27	6.82
10	Chinese	65	8.28
10	English	46	14.47
10	Spanish	945	13.11
10	Other Language	477	11.28
10	Students with Disabilities	150	10.59

I.12: Grade 11

Grade	Group	N-count	Percent
11	Female	522	12.12
11	Male	526	11.52
11	American Indian or Alaska Native	3	20.00
11	Black or African American	112	13.40
11	Asian	233	9.70
11	Native Hawaiian/Other Pacific Islander	4	14.29
11	Hispanic or Latino	595	11.92
11	White	100	16.92
11	Multiracial (not Hispanic origin)	1	12.50
11	NYC	725	11.55
11	Big 4 Cities	40	8.25
11	High Need Urban/Suburban	93	10.54
11	High Need Rural	7	17.07
11	Average Need	106	14.99
11	Low Need	71	16.99
11	Charter Schools	6	9.68
11	0 Years LEP	550	13.16
11	1 Years LEP	123	6.99
11	2 Years LEP	167	11.35
11	3 Years LEP	49	10.58
11	4 Years LEP	36	11.96
11	5 Years LEP	29	13.06
11	6 Years LEP or More	93	20.39
11	Arabic	24	11.16
11	Bengali	20	8.73
11	Chinese	44	10.30
11	English	32	14.95
11	Spanish	586	11.99
11	Other Language	342	11.79
11	Students with Disabilities	72	8.46

I.13: Grade 12

Grade	Group	N-count	Percent
12	Female	460	14.17
12	Male	509	15.83
12	American Indian or Alaska Native	0	0.00
12	Black or African American	110	17.57
12	Asian	231	13.25
12	Native Hawaiian/Other Pacific Islander	3	33.33
12	Hispanic or Latino	536	14.60
12	White	89	22.08
12	Multiracial (not Hispanic origin)	0	0.00
12	NYC	696	14.60
12	Big 4 Cities	27	7.83
12	High Need Urban/Suburban	70	13.54
12	High Need Rural	6	15.79
12	Average Need	100	20.49
12	Low Need	65	22.18
12	Charter Schools	5	33.33
12	0 Years LEP	565	16.01
12	1 Years LEP	93	10.98
12	2 Years LEP	121	13.47
12	3 Years LEP	35	13.83
12	4 Years LEP	50	16.29
12	5 Years LEP	31	15.58
12	6 Years LEP or More	71	16.95
12	Arabic	27	15.79
12	Bengali	29	13.74
12	Chinese	33	13.98
12	English	26	19.85
12	Spanish	531	14.80
12	Other Language	323	15.21
12	Students with Disabilities	63	9.26

APPENDIX J: DIF CATEGORIES FOR OPERATIONAL ITEMS

J. 1 Results of DIF Analyses

Grade Span	Modality	Item Seq. Number	DIF Male/Female	DIF Asian/Other	DIF Hispanic/Other	DIF White/Other
K-1	Listen/Speaking	2	A	>B	A	A
		7	A	>B	A	A
		13	A	A	A	>B
K-1	Reading/Writing	21	A	A	B
2-4	Listen/Speaking	5	A	>B	<B	A
		11	A	>B	A	A
		20	A	B	A
		22	A	A	>B	A
		25	AA	<BB	AA	AA
		26	AA	<BB	AA	AA
		27	AA	<BB	AA	AA
		28	AA	<BB	AA	AA
2-4	Reading/Writing	6	A	B	A
		27	AA	>C	<BB	AA
		31	AA	AA	AA	>BB
		32	AA	AA	AA	>BB
5-6	Listen/Speaking	8	A	>BB	AA	AA
		12	A	>BB	AA	AA
		26	AA	<BB	AA	AA
		27	AA	<BB	AA	AA
		29	AA	<CC	AA	AA
		30	AA	<BB	AA	AA
		37	AA	<BB	AA	AA
		40	AA	<BB	AA	AA
5-6	Reading/Writing	21	A	>B	A	A
		28	AA	<BB	>BB	<BB
		33	AA	<BB	AA	AA
		39	AA	<CC	>BB	AA
7-8	Listen/Speaking	2	A	<C	>C	A
		3	A	<B	A	>B
		39	AA	<BB	AA	AA
7-8	Reading/Writing	2	A	>B	<C	A
		3	<B	B	A
		10	A	>B	A	A
		16	A	C	<B
		25	A	>B	A	A
		29	AA	<BB	AA	AA
		31	AA	AA	AA	>BB
		35	AA	<BB	>BB	AA
		37	AA	AA	AA	>B
		38	AA	AA	AA	>B
39	AA	AA	AA	>B		

J. 2 Results of DIF Analyses (continued)

Grade Span	Modality	Item Seq. Number	DIF Male/Female	DIF Asian/Other	DIF Hispanic/Other	DIF White/Other
9-12	Listen/Speaking	9	A	>B	A	A
		10	A	>C	<B	A
		16	A	<C	>B	A
		19	A	A	A	>B
		20	A	>B	A	A
		26	AA	<BB	AA	AA
		32	AA	<BB	AA	AA
9-12	Reading/Writing	1	A	>C	<B	A
		2	A	>C	<C	A
		3	A	<C	>C	A
		5	A	<B	A	A
		15	A	<C	>C	<B
		28	AA	<CC	>CC	AA
		32	AA	<BB	AA	AA
		37	AA	>BB	<BB	AA
		38	AA	>BB	<CC	AA
		39	AA	>BB	AA	AA
		41	AA	>BB	AA	AA
		42	AA	AA	<BB	AA
		46	>BB	AA	AA	AA

Note: > indicates in favor of Males or Other groups while < indicates in favor of the focal groups (i.e., Female, Asian, Hispanic, and White).

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.
- 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.
- Camilli, G. & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart and Winston.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334.
- 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. (2006). *Liseral 8.8 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. & Wright, B. D. (2000). *A user's guide to WINSTEPS: Rasch-Model computer program*. Chicago: MESA Press
- Linacre, J. M. (2003). *A user's guide to WINSTEPS: MINISTEP: Rasch-Model computer program*. Chicago: MESA Press.
- Linacre, J. M. (2009). *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., Browne, M. & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods* 2(1):130–149.
- Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22:719–748.
- Mantel, N. (1963). Chi-square tests with one degree of freedom: Extension of the Mantel-Haenszel procedure. *Journal of the American Statistical Association* 58:690–700.
- Masters, G. N. (1982). “A Rasch model for partial credit scoring.” *Psychometrika* 47:149–174.
- 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. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 86(2):420–428.
- Steinley, D. & Wood, P. (2000). *Icc.sas—a program to calculate intraclass correlations and confidence intervals*. Columbia, MO: University of Missouri.
- Swygert, K. A., McLeod, L. D. & Thissen, D. (2001). Factor analysis for items or tests 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).
- 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 Test.