

**NEW YORK STATE
TESTING PROGRAM**

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

2009 ADMINISTRATION

TECHNICAL MANUAL

SUBMITTED BY

PEARSON

JANUARY 2010



TABLE OF CONTENTS

<i>OVERVIEW OF THIS MANUAL</i>	1
<i>Part 1 Introduction</i>	1
<i>Part 2 Test Design and Development</i>	1
<i>Part 3 Scoring</i>	1
<i>Part 4 Classical Item-Level Statistics</i>	1
<i>Part 5 Reliability</i>	1
<i>Part 6 Validity</i>	1
<i>Part 7 Calibration, Equating, and Scaling</i>	1
<i>Part 8 Standard Setting</i>	2
<i>Part 9 Summary of Operational Test Results</i>	2
PART 1: INTRODUCTION	3
<i>1.1 Background</i>	3
<i>1.2 Rationale and Purpose</i>	4
<i>1.3 Recommended Test Use</i>	4
<i>1.4 Test Accommodations</i>	4
PART 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 Specification 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
Psychometric Guidelines for Operational Item Selection and Form Construction.....	15
Testing Written Language.....	15
Testing Oral Language.....	156
PART 3: SCORING	18
<i>3.1 NYSESLAT Range Finding</i>	18
<i>3.2 Operational File</i>	19
<i>3.3 Rater Training</i>	19
<i>3.4 Inter-Rater and Intra-Rater Reliability</i>	20
PART 4: CLASSICAL ITEM-LEVEL STATISTICS	21
<i>4.1 Item-Level Descriptive Statistics</i>	21
<i>4.2 Differential Item Functioning</i>	24
<i>Summary of DIF Analyses</i>	27
PART 5: RELIABILITY	29
<i>5.1 Internal Consistency Reliability</i>	29
<i>5.2 Classical SEM (based on Classical Test Theory)</i>	29
<i>5.3 Standard Error of the Mean (SEM_n)</i>	30
<i>5.4 Conditional SEM (Based on Item Response Theory)</i>	31
<i>5.5 Summary of Descriptive and Reliability Statistics</i>	32
<i>5.6 Inter-Rater Reliability</i>	37
<i>Inter-Rater Agreement</i>	38
<i>Intraclass Correlation</i>	38
<i>Kappa Coefficient</i>	39
<i>Summary</i>	42
<i>5.7 Reliability of Classification Decision at Proficient Cut</i>	42

<i>PART 6: VALIDITY</i>	46
6.1 Content Validity.....	46
6.2 Internal Structure	47
Validity of the Scoring Structure via Confirmatory Factor Analyses (CFA).....	51
6.3 External Structure.....	52
6.3.1 Relationship with the New York English Language Arts Assessment	53
Correlations and Other Descriptive Statistics	53
Proficiency Level	56
Summary	59
6.3.2 Relationship with the New York State Regents Examination in Comprehensive English	59
Correlations and Other Descriptive Statistics	59
Proficiency Level Classification	61
Summary	62
<i>PART 7: CALIBRATION, EQUATING, AND SCALING</i>	64
7.1 Item Response Model and Rational for Use	64
7.2 Evidence of Model Fit.....	64
7.3 Description of Calibration Sample, Process, and Results	65
Calibration Sample	65
Calibration Process	65
7.4 Equating of the NYSESLAT	66
Equating Method and Rationale	66
Common Item Linking Design	67
Anchor Item Evaluation	68
Stability of Anchor Items	68
TCC and Standard Error Curves Comparison	69
7.5 Scaling of the NYSESLAT	75
Raw Score-to-Scale Score Tables	75
7.6 Rasch Information	76
<i>PART 8: STANDARD SETTING</i>	78
8.1 Introduction	78
8.2 Standard-Setting Model.....	78
8.3 Committees of Panelists.....	78
8.4 Performance Levels and Cut Scores.....	79
8.5 Standard-Setting Process.....	79
Review of the Assessment	80
Experience the Assessment	80
Scoring the Assessment.....	80
Review of Student Performance Levels.....	80
Three Rounds of Ratings.....	80
8.6 Summary Statistics for the Three Rounds of Ratings.....	81
8.7 Post-Standard-Setting Analyses.....	81
8.8 Final Performance-Level Cut Points	79
<i>PART 9: SUMMARY OF OPERATIONAL TEST RESULTS</i>	84

LIST OF TABLES

Table 2.1 NYSESLAT Operational Test Blueprint	7
Table 2.2 Test Specifications by Modality and Grade Span	9
Table 2.3 Maximum Number of Points by Modality and Grade Span	9
Table 2.4 Test Specifications for New York State ESL Learning Standards by Grade Span	10
Table 2.5 Item Mapping by New York State ESL Learning Standards and Performance Indicators	11
Table 2.6 Four-Year Sampling Matrix for the NYSESLAT Field Test	14
Table 4.1 Summary of Classical Item Difficulty and Discrimination by Grade Span	22
Table 4.2 DIF Classification for Constructed-Response Items	25
Table 4.3 DIF Classification for Multiple-Choice Items	26
Table 4.4 Results of DIF Analyses for 2009 Operational Test Items	28
Table 5.1 Descriptive Statistics and Reliability by Grade Span and Modality	33
Table 5.2 Descriptive Statistics and Reliability by Grade and Modality	33
Table 5.3 Audit Sample	38
Table 5.4 Rater Agreement for Pre-Writing and Writing Prompts	40
Table 5.5 Percentages of the Score Difference Between Raters	41
Table 5.6 Comparison Between Local and Pearson Raters	42
Table 5.7 Classification Accuracy and Consistency by Grade	45
Table 6.1 Intercorrelation Among the Modalities by Grade Span	48
Table 6.2 Intercorrelation Among the Modalities by Grade	48
Table 6.3 Global Fit Indices for the One- and Two-Factor Models (Grades K–1)	51
Table 6.4 Global Fit Indices for the One- and Two-Factor Models (Grades 2–4)	52
Table 6.5 Global Fit Indices for the One- and Two-Factor Models (Grades 5–6)	52
Table 6.6 Global Fit Indices for the One- and Two-Factor Models (Grades 7–8)	52
Table 6.7 Global Fit Indices for the One- and Two-Factor Models (Grades 9–12)	52
Table 6.8 Descriptive Statistics of the New York ELA and the NYSESLAT Total Raw Scores of Reading, Writing, and Listening	54
Table 6.9 Descriptive Statistics of the New York ELA and the NYSESLAT Combined Scale Scores of Listening/Speaking and Reading/Writing	55
Table 6.10 Grade 3: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level	57
Table 6.11 Grade 4: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level	58
Table 6.12 Grade 5: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level	58
Table 6.13 Grade 6: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level	58
Table 6.14 Grade 7: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level	58
Table 6.15 Grade 8: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level	59
Table 6.16 Descriptive Statistics of the NYSESLAT Raw Scores of Listening, Speaking, Reading, and Writing Modalities and Their Correlations With the Regents English Exam Total Scale Scores	60
Table 6.17 Descriptive Statistics of the NYSESLAT Raw Scores of Listening/Speaking, Reading/Writing Modalities and Total Scale Scores and Their Correlations With the Regents English Exam Total Scale Scores	60
Table 6.18 Grade 10: Percentage of English language learners Passing/Failing the Regents English Exam at Each of NYSESLAT Proficiency Level	62
Table 6.19 Grade 11: Percentage of English language learners Passing/Failing the Regents English Exam at Each of NYSESLAT Proficiency Level	62
Table 6.20 Grade 12: Percentage of English language learners Passing/Failing the Regents English Exam at Each of NYSESLAT Proficiency Level	62
Table 7.1 Common Items as Anchor Items by Modality Combination and Grade Span	69
Table 7.2 Average Rasch Difficulty by Grade Span and Modality	77
Table 8.1 Panel Composition for Standard-Setting Committees	79
Table 8.2 Final Performance-Level Cut Points for Listening and Speaking	82
Table 8.3 Final Performance-Level Cut Points for Reading and Writing	83
Table 9.1 Raw Score Summary by Grade	85
Table 9.2 Raw Score Summary by Grade Span	85
Table 9.3 Scale Score Summary by Grade	86
Table 9.4 Scale Score Summary by Grade Span	86
Table 9.5 Percentage of Students in Each Proficiency Level by Grade	87
Table 9.6 Percentage of Students in Each Proficiency Level by Grade Span	88

Table 9.7 Exit Rate by Grade for 2006, 2007, 2008, and 2009 Administrations	88
Table 9.8 Exit Rate by Grade Span for 2006, 2007, 2008, and 2009 Administrations	89
Table F.1 Maximum Number of Points by Modality and Grade Span for the 2009 NYSESLAT	199
Table F.2a Summary Statistics	202
Table F.2b Factor Loading Estimates for the One- and Two-Factor Models	203
Table F.2c Global Fit Indices for the One- and Two-Factor Models.....	203
Table F.3a Summary Statistics	203
Table F.3b Factor Loading Estimates for the One- and Two-Factor Models	204
Table F.3c Global Fit Indices for the One- and Two-Factor Models.....	204
Table F.4a Summary Statistics	204
Table F.4b Factor Loading Estimates for the One- and Two-Factor Models	205
Table F.4c Global Fit Indices for the One- and Two-Factor Models.....	205
Table F.5a Summary Statistics	205
Table F.5b Factor Loading Estimates for the One- and Two-Factor Models	206
Table F.5c Global Fit Indices for the One- and Two-Factor Models.....	206
Table F.6a Summary Statistics	206
Table F.6b Factor Loading Estimates for the One- and Two-Factor Models	207
Table F.6c Global Fit Indices for the One- and Two-Factor Models.....	207

LIST OF FIGURES

Figure 5.1: Classification Accuracy.....	43
Figure 5.2: Classification Consistency	43
Figure F.1: Path Diagram for the One-Factor Model.....	201
Figure F.2: Path Diagram for the Two-Factor Model.....	202

OVERVIEW OF THIS MANUAL

This New York State English as a Second Language Achievement Test (NYSESLAT) Technical Manual for the 2009 administration is organized into nine major parts: Introduction; Test Design and Development; Scoring; Classical Item-Level Statistics; Reliability; Validity; Calibration, Equating, and Scaling (CES); Standard Setting; and Summary of Operational Test Results. An overview of this manual is provided below.

Part 1 Introduction

Part 1 presents the background of the NYSESLAT, its rationale and purpose, recommended test use, and test accommodations. Test accommodations include Large Type and Braille.

Part 2 Test Design and Development

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

Part 3 Scoring

Part 3 provides a description of the scoring process. It includes a description of the range finding meeting in Albany, New York, in 2009. It also provides information about the audit-scoring process that was conducted on a ten-percent sample, and a description of rater training and inter-rater reliability.

Part 4 Classical Item-Level Statistics

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

Part 5 Reliability

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

Part 6 Validity

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

Part 7 Calibration, Equating, and Scaling

Part 7 explains the Rasch and Partial Credit Models and provides sample-item characteristic curves for a one-step item and a two-step item. It also includes the process of the calibration,

equating, and scaling of the 2009 administration of the NYSESLAT. Part 7 also explains the rationale for the use of the IRT model. It includes the IRT model fit statistics and the average Rasch difficulty of the subtests.

Part 8 Standard Setting

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

Part 9 Summary of Operational Test Results

Part 9 presents the raw score summary, scale score summary, percentage of students in each performance category, and exit rate for the 2006, 2007, 2008, and 2009 administrations of the NYSESLAT.

PART 1: INTRODUCTION

1.1 Background

Title III of the Federal *No Child Left Behind* (NCLB) Act of 2001 requires annual assessment of the English language skills of English language learners. 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 English language learner students using a state-approved assessment.

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

In response to NYSED’s request, Pearson¹ offered a solution with two distinct phases. With respect to the first phase, custom items were developed and used to create the 2009 NYSESLAT

¹Pearson purchased Harcourt Assessment, Inc. in 2009. Pearson is now the vendor for the NYSESLAT.

operational forms. Anchor items were pulled from the 2008 test and used to produce custom forms for the 2009 test administration. With respect to the second phase, brand new custom items were developed for the 2009 test administration.

1.2 Rationale and Purpose

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

1.3 Recommended Test Use

The NYSESLAT is 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 English language learners 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 English language learners who score at the proficient level from the ESL/Bilingual programs.

1.4 Test Accommodations

All test items are developed following the guidelines of Universal Design. 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, Pearson utilizes in-house content and fairness experts to ensure that the forms are pulled with concepts of universal design in mind. Pearson 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

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

Braille

Pearson 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. Pearson produces Braille forms for each NYSESLAT subtest and grade span. For the K–1 grade span, a checklist is provided rather than a Braille test.

If a content area is difficult for Braille, Pearson 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.

PART 2: TEST DESIGN AND DEVELOPMENT

2.1 Overview

To meet the requirements of Title III of the Federal *No Child Left Behind* (NCLB) Act and of New York State regulations regarding the assessment of limited English proficient students, NYSED developed *Learning Standards for English as a Second Language*, a comprehensive set of ESL learning standards and performance indicators in the four domains of Speaking, Listening, Reading, and Writing. NYSED then requested test development, research, and scoring based on these standards. The NYSESLAT consists of a test in each of the five grade spans (K–1, 2–4, 5–6, 7–8, 9–12). The five tests are vertically scaled and each consists of four subtests: Listening, Speaking, Reading, and Writing. The tests are designed to assess the English language skills of English language learners in grades K–12 and to capture their progress toward achieving full English language proficiency. 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 four basic domains or modalities—Speaking, Listening, Reading, and Writing—for grades K–12. It includes multiple-choice, constructed-response, short-response, and extended-response items. The total number of items per grade span varies. For grade span K–1, there are a total of 70 items; for grade span 2–4, there are a total of 80 items; and for grade spans 5–6, 7–8, and 9–12, there are a total of 87 items each.

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

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

The operational test blueprint for the 2009 administration of the NYSESLAT is presented in Table 2.1. Test specifications by modality and grade span are shown in Table 2.2. Table 2.3 provides the maximum number of points by modality and grade span. Table 2.4 provides a breakdown of the number of items based on New York State’s ESL learning standards and grade spans. The 2009 test design consists of all custom-developed items for the NYSESLAT.

Table 2.1
2009 NYSESLAT Operational Test Blueprint

NYSESLAT 2009 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					
	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 2.1 (continued)
2009 NYSESLAT Operational Test Blueprint

NYSESLAT 2009 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	Each of the Levels 2-4, 5-6, 7-8, and 9-12 has 3 CR items with 0-2 score points.
	Extended Response	1	1	1	1	1	Each of the five levels has one ER item. The ER item for level K-1 has 0-2 score points. The ER items for 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 2.2
2009 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 2.3
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 2.4
Test Specifications for New York State ESL Learning Standards by Grade Span

Grade Span	Standard*	Number of Items	Percent of Total Items	Number of Score Points	Percent of Total Points
K-1	1	46	66	54	56
	2	6	9	12	12
	3	6	9	10	10
	4	8	11	15	15
	5	4	6	6	6
	Total Test		70	100	97
2-4	1	44	55	50	48
	2	12	15	15	14
	3	9	11	17	16
	4	11	14	17	16
	5	4	5	5	5
	Total Test		80	100	104
5-6	1	48	55	52	47
	2	9	10	9	8
	3	8	9	14	13
	4	19	22	31	28
	5	3	3	5	5
	Total Test		87	100	111
7-8	1	50	57	55	50
	2	14	16	17	15
	3	13	15	20	18
	4	8	9	16	14
	5	2	2	3	3
	Total test		87	100	111
9-12	1	40	46	43	39
	2	9	10	12	11
	3	13	15	19	17
	4	14	16	24	22
	5	11	13	13	12
	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 Specification by New York State Learning Standards for English as a Second Language

Table 2.5 provides an item-mapping summary, and Appendix A.1 provides, in detail, the 2009 item mapping by New York State Learning Standards for each item within each grade span and modality. Item specifications are presented in Appendix A.2.

Table 2.5
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
		Writing	Standard 1	PI 8, 11, 12	22-29
		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
	Writing Conventions	Standard 1	PI 1, 6, 12	25-35	69
		Standard 3	PI 1	36	6
	Writing	Standard 1	PI 6	37	6
Standard 3		PI 1, 6	38-40	19	
Speaking		Standard 1	PI 5, 9	30, 31	13
	Standard 4	PI 3, 5	26-28, 32-37, 39, 41	75	
Listening	Standard 5	PI 3	29, 38	13	
	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 2.5 (continued)
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
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
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
Reading		Standard 1	PI 1, 3, 4, 6, 16	1–7, 13, 14, 16, 17	41
		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 3	PI 1, 3, 5	8, 12, 20	11
		Standard 1	PI 6, 12, 15	28–30, 32–36, 38–42	68
		Standard 5	PI 1	31, 37	11
		Standard 3	PI 1, 6	45, 46	11
Writing Conventions	Standard 5	PI 3	43, 44	11	

2.4 Item Development and Review Processes

All items on the 2009 NYSESLAT were written by educators of English language learners, including a number of New York State teachers. An Item Writers' Training Guide was sent out to all item writers to assist them in developing the items. In addition, Pearson specialists were available via phone and/or e-mail to answer any questions the item writers had. Assessment specialists at Pearson reviewed the passages and items, and in accordance with the item specifications, the assessment specialists ensured the following:

- Absence of bias and sensitive topics in passages
- Item soundness 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

The Reading passages were then reviewed by a committee of New York State ESL teachers. Each test question was also reviewed by grade-level ESL educators from New York State. Only those passages and test questions judged to be of acceptable quality and fair to students from the various ethnic/cultural backgrounds of those who presently live in New York State were approved for use. In addition to the committee reviews, a "Final Eyes" review was conducted by a separate group of New York teachers, to ensure items were error-free and appropriate. After items were field tested, New York teachers were invited to a range-finding session, during which they reviewed potential anchor papers and select practice sets that were used to aid teachers in scoring the 2009 NYSESLAT.

Anchor items serve the purpose of providing statistics that help equate tests from one year to the next, so that the scaled scores between them are comparable. Anchor item selection criteria are described in Section 7.4.

2.5 Field Test

Pearson constructed, and NYSED reviewed and approved, stand-alone field test forms that were administered to a representative sample of limited English proficient students around New York State. There were a total of two field test forms at each grade span, and each form was split into four mini-forms for the four modalities (Listening, Speaking, Reading, and Writing). Each modality mini-form is administered to a representative sample of New York State schools. The field tests are conducted in all New York State schools with English language learner populations. To accommodate the field test needs, school districts have been divided into four stratified random samples (S1, S2, S3, S4), using the Need/Resource Capacity Index as the strata. Each of the four samples is assigned to administer a field test in one modality each year. The field test assignments systematically rotate among the field test samples each year (see Table

2.6). 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 2.6
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

Data from the 2009 field test administration informed construction of the 2010 NYSESLAT operational test. The 2009 field-testing sampling plan and statistics presented to NYSED by Pearson are discussed in the *2009 NYSESLAT Fall Field Test Technical Report*.

2.6 Test Construction

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 on the basis of either gender or ethnicity.

The process for constructing the 2009 NYSESLAT operational forms was as follows: As soon as Pearson had clean statistics from the spring 2008 Field Test, Pearson 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. Pearson produced draft operational forms that were then presented to NYSED. A joint Pearson-NYSED team reviewed the forms, following all item-selection criteria, and made various changes to the draft forms. Final forms were then reviewed by Pearson ESL assessment specialists, editorial staff, psychometricians, and quality control staff, as well as by NYSED.

Psychometric Guidelines for Operational Item Selection and Form Construction

Statistical considerations included item difficulty and other statistical characteristics of the items. P-values were 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 '09 form and spring '08 form. The average p-values should be similar to one another; and
 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 accurately identify students at different levels of language development provides better information to classroom teachers, who must find the most effective way to help their students reach proficiency. It also provides the very important evidence of students' progress toward proficiency, which is required by the NCLB legislation.

To determine the appropriate language for Reading passages, Pearson assessment specialists, 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. Pearson uses the *Educational Developmental Laboratory (EDL) Core Vocabularies in Reading, Mathematics, Science, and Social Studies*, published by Steck-Vaughn, to help determine age- and grade-appropriate language for English language learner items and stimuli for the oral-language subtests. Furthermore, Pearson ESL assessment specialists and editors ensure that the language in all stimuli and items, from kindergarten through grade 12, is both topic- and age-appropriate for test takers.

Testing Oral Language

Recognizing that oral language structure and vocabulary in English differ vastly from the written language, issues of oral language assessment among Kindergarten through Grade 12 English language learners have been the subject of special investigation at Pearson. Pearson's English language proficiency professionals have conducted research on the item types that appear in the NYSESLAT Speaking and Listening subtests by presenting examples of these item types to English language learners in cognitive laboratories. This setting allows for careful observation and recording of student responses and 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 student spoken responses

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

PART 3: SCORING

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

3.1 NYSESLAT Range Finding

Range finding was held in Albany on July 7–10, 2009, before field test items were scored by Pearson. The participants included:

- One full-time Pearson Program Manager, one Pearson Performance Scoring Center (PSC) Project Manager, one full-time PSC Content Specialist, and five temporary Pearson Performance Scoring Center (PSC) facilitators (one for each grade span: K–1, 2–4, 5–6, 7–8, and 9–12)
- Three New York State Education Department representatives who greeted participants and showed support for the range-finding process—two of the three NYSED representatives participated in range finding.
- Four teachers and/or education department representatives for each grade span

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

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

A PowerPoint slide show was presented to all range-finding participants for grades 2 to 12 and a separate presentation was given to grades K–1. These presentations reviewed the rubric and gave examples for each score point. For grades 2–12, three pre-writing items and one writing prompt were reviewed per grade span. For grades K–1, nine items (sentence copying, letter writing, word writing, sentence writing, and picture description) were reviewed. Each group of teachers read two assembled sets of sample papers per item. The two sets ranged from possible low to

high responses. Each set of pre-writing items included at least 15 papers, and each set of writing items included at least 25 papers.

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

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

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

3.2 Operational File

The operational file for the spring 2009 NYSESLAT came from the New York State Data Repository after scoring was completed. The file was verified by Pearson's Assessment and Information Quality (AIQ) Department to ensure data accuracy based on the description values in the file layout. Sections 3.3 through 3.4 indicate the procedures that Pearson conducted to assure reliable and accurate scoring of the items.

3.3 Rater Training

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

All PSC readers had a minimum of a bachelor's degree and had successfully completed generalized workshops in performance assessment scoring before being considered as a potential reader for a specific project such as NYSESLAT. Training of readers was based on anchors and training sets that were part of the sampler incorporated into the training procedures for operational scoring done by New York teachers. The sets were developed by New York teachers during range finding in July 2009.

3.4 Inter-Rater and Intra-Rater Reliability

All readers were trained to score to the same scale to ensure accurate, consistent, reliable scoring. PSC adhered to stringent criteria in its general screening, training, and qualifying procedures as preliminary measures for obtaining high levels of consistency and reliability. Team leaders conducted “read behinds,” reading the same booklets after readers, to check for accuracy of scores. In addition, 10% of all booklets were “second” read by the training team leader to check accuracy. If individuals were not “on track,” retraining ensued, as these readers were being systematically too lenient or too harsh in their ratings, or varied unsystematically and unpredictably from other raters and deviated from the training standards in their scores. The statistics of rater reliability are presented in Section 5.6.

PART 4: CLASSICAL ITEM-LEVEL STATISTICS

4.1 Item-Level Descriptive Statistics

This section presents the raw score summary statistics for all items in the spring 2009 administration of the NYSESLAT within the framework of Classical Test Theory. The compiled file from the New York State Data Repository was used to obtain all raw score statistics. The p-value for each item is defined as the proportion of students that answer an item correctly for the multiple-choice items. A high p-value means that an item is easy; a low p-value means that an item is difficult. For the constructed-response items, the p-value is reported as the average number of points out of the maximum number of possible points.

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

The item-level statistics for the operational 2009 NYSESLAT are presented in Appendices B.1–B.5 by grade span. The tables are grouped by Listening/ Speaking and Reading/Writing modality combinations. The following item information and statistics are presented for each item:

- Item number
- Item format (multiple-choice, constructed-response, short-response, or extended-response)
- Maximum number of possible points
- N-count (number of students)
- Response options for multiple-choice items and percentage of students obtaining each score point for constructed-response items
- Omits (percentage of students omitting an item)
- P-value for multiple-choice items (percentage of examinees that answered the item correctly)
- 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. Such items do not provide adequate information, and their inclusion serves a limited purpose in the measurement process. However, the NYSESLAT, being a standard-referenced examination, allows exception to the rule if the item is deemed absolutely necessary by content experts. Although there is no consensus as to what is an acceptable p-value, the guiding policy during item review and form building is to choose a p-value ≤ 0.90 and ≥ 0.30 .

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

independence of the item (low correlation). A very high correlation, in fact, may mean a redundancy of the item; that is, the item performs the same function as some other items. On the other hand, too low a value for the point biserial may indicate a “shaky” association of the item with the underlying trait.

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

As can be seen in Appendix B, with the exception of some high p-values and point-biserial values, all the items fall well within our pre-set level of acceptance, both in terms of the p-value and the point biserial. The mean p-values and point-biserial indices are presented in Table 4.1. The mean p-values and point-biserial indices at each grade span for the combined modalities of Listening/Speaking and Reading/Writing are in the middle range and are acceptable. The total population had 220,332 records. After removing duplicates (N=3694) and off-level records (N=3733), the total N=212,905. Invalid raw scores were also removed from data for the final sample sizes used in Table 4.1:

	Listening/Speaking			Reading/Writing		
	Initial N	Invalid RS	Final N	Initial N	Invalid RS	Final N
K-1	53210	1614	51596	53210	1603	51607
2-4	64213	1342	62871	64213	3397	60816
5-6	29282	815	28467	29282	1854	27428
7-8	25322	893	24429	25322	2083	23239
9-12	40878	900	39978	40878	2780	38098

Table 4.1
Summary of Classical Item Difficulty and Discrimination by Grade Span

Table 4.1
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	51596	0.85	0.11	0.48	0.05
	Speaking	51596	0.77	0.06	0.76	0.03
	Reading	51607	0.74	0.09	0.55	0.09
	Writing	50607	0.78	0.15	0.59	0.16
	Listening and Speaking	51596	0.82	0.10	0.59	0.14
	Reading and Writing	51607	0.76	0.12	0.57	0.13

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

2–4	Listening	62871	0.80	0.11	0.46	0.04
	Speaking	62871	0.83	0.04	0.76	0.03
	Reading	60816	0.72	0.18	0.48	0.07
	Writing	60816	0.72	0.12	0.56	0.10
	Listening and Speaking	62871	0.81	0.09	0.58	0.15
	Reading and Writing	60816	0.72	0.16	0.51	0.09
5–6	Listening	28467	0.76	0.12	0.47	0.09
	Speaking	28467	0.82	0.04	0.80	0.03
	Reading	27428	0.70	0.13	0.51	0.10
	Writing	27428	0.71	0.15	0.53	0.14
	Listening and Speaking	28467	0.78	0.10	0.60	0.18
	Reading and Writing	27428	0.71	0.14	0.50	0.10
7–8	Listening	24429	0.69	0.14	0.47	0.05
	Speaking	24429	0.79	0.03	0.83	0.02
	Reading	23239	0.64	0.13	0.50	0.06
	Writing	23239	0.69	0.12	0.54	0.12
	Listening and Speaking	24429	0.73	0.12	0.61	0.18
	Reading and Writing	23239	0.66	0.13	0.50	0.07
9–12	Listening	39978	0.70	0.11	0.49	0.07
	Speaking	39978	0.79	0.03	0.80	0.03
	Reading	38098	0.63	0.16	0.44	0.07
	Writing	38098	0.72	0.14	0.53	0.14
	Listening and Speaking	39978	0.74	0.10	0.61	0.16
	Reading and Writing	38098	0.67	0.16	0.45	0.09

Note: The data file used in the table above is the final research file. The total n-counts for grades K–12 are 220,332 and includes private schools. 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).

4.2 Differential Item Functioning

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

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

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

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

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

where y_T represents the T scores that can be obtained on the item, while n_{FTK} denotes the number of focal group members who are on the k th level of the matching variable and received an item score of y_T . The expectation of F_K under the hypothesis of no association is 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 without an indication of the direction of DIF (i.e., whether in favor of the reference or the comparison 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 and

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

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

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

Table 4.2
DIF Classification for Constructed-Response Items

Category	Description	Criterion
AA	No <i>DIF</i>	Non-significant Mantel χ^2 or Significant Mantel χ^2 and $ SMD/SD \leq .17$
BB	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) is used in conjunction with the M-H odds ratio (transformed to what ETS calls the Delta Scale (D)). To calculate the

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

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

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

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

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

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

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

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

The quantity D represents the apparent difference in the difficulty of the items in the Delta metric for the two groups whose performance is being compared. Table 4.3 depicts DIF classifications for MC items based on the M-H χ^2 and the item Delta Scale difference value (D):

Table 4.3
DIF Classification for Multiple-Choice Items

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

For the NYSESLAT, DIF analysis was 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 Caucasian 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 enough n-counts were available for the comparison.

In the separate *2009 Field Test Technical Report*, a DIF analysis results for each grade span by form for the field test items was provided, while Appendix J 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 4.4 provides a summary of DIF analysis of operational test items, including n-counts for each DIF analysis group and number of items flagged by both SMD and/or Mantel for the constructed-response items and ETS Delta Scale metric (D) and M-H criteria for multiple-choice items (see details in Table 4.2 and Table 4.3). All items flagged for DIF were carefully reviewed during operational test construction. Only those items that passed the reviews were included in the operational tests. It is important to note that DIF does not indicate bias. None of the items showing moderate or large DIF in Table 4.4 was identified as “biased” by the content specialists.

Table 4.4
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	51,596	3	3	0
	Reading/Writing	51,607	1	1	0
2-4	Listening/Speaking	62,871	8	8	0
	Reading/Writing	60,816	4	3	1
5-6	Listening/Speaking	28,467	8	7	1
	Reading/Writing	27,428	4	3	1
7-8	Listening/Speaking	24,429	3	2	1
	Reading/Writing	23,239	11	9	2
9-12	Listening/Speaking	39,978	7	5	2
	Reading/Writing	38,098	13	7	6

Note: Detailed results can be found in Appendix J.

PART 5: RELIABILITY

5.1 Internal Consistency Reliability

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

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

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

where

I is the number of items on the test,

s_i^2 is the variance of item i , and

S_x^2 is the total test variance.

Cronbach's coefficient alpha reliability statistic was calculated and presented in Tables 5.1 and 5.2 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)

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

The SEM is another index of reliability that provides an estimate of the amount of error in an individual's observed test score. The individual's observed total score is considered the estimate of the person's true score. Because the SEM is inversely related to the reliability of a test, the greater the reliability, the less the standard error of measurement, and the more confidence one may have in the accuracy, or precision, of the observed test score. The measurement error is commonly expressed in terms of standard deviation units; that is, the SEM is the standard deviation of the measurement error distribution. 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 SEM s are presented in Tables 5.1 and 5.2 in Section 5.5. The smaller the SEM s (close to 0), the higher the quality of the test will be.

5.3 Standard Error of the Mean (SEMn)

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

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

where $SEMn$ 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 5.1 and 5.2 in Section 5.5. The more accurate the estimation of the population mean, the smaller the $SEMn$ values will be.

5.4 Conditional SEM (Based on Item Response Theory)

Unlike the SEM based on the classical test theory, the SEM based on the item-response theory (IRT) is not the same for all students. 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. This implies that 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:

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

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

According to the consensus in measurement, an aspect for the popularity of IRT methods in analyzing data is because classical statistics assume equivalency of students and item measurements for all examinees and items in a test. IRT methods allow for the differentiation of varied student and item performances in estimating the reliability of the measurement (Crocker & Algina, 1986). As evidenced by Equation 13, one reason for the fluctuation in the standard errors of students is that they are a function of the n-counts. As such, the standard errors for each of the ability score estimates are the smallest in the middle of the score distribution (where most examinees perform) and greatest for estimates in the extreme, where subsequently lower numbers of students perform on a test and thus produce less precise estimates. It is for this reason that IRT estimates with individual standard errors at score points, i.e., conditioned on theta (student ability estimates), are preferred to classical SEMs, which do not differentiate between the precision of student estimates at different levels of performances.

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

5.5 Summary of Descriptive and Reliability Statistics

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

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

In analyzing the information provided in Tables 5.1 and 5.2, the measures of reliability should be examined by keeping to the general trend that associates greater reliability with a lower error index for that particular measure. In Table 5.1, the test reliability values of Listening/Speaking and Reading/Writing across the five grade spans are all above 0.80, 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 0.02 to 0.08, which is very small and acceptable. The SEM values of Listening/Speaking and Reading/Writing across the five grade spans ranged from 1.44 to 2.96, 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 5.2, most test reliability values of Listening/Speaking and Reading/ Writing across the twelve grades are above 0.80. The reliability for Grade K Reading/Writing may seem a little low in comparison to the other reliability values, but this is most likely due to the students' ages, as they just started learning reading and writing, even in their mother language. This may also be explained by the relatively erratic learning of reading and writing English in kindergarten, which could be based on students being in a new environment, missing their homes, the level of learning to read and write at home (which may not be the same for all students), etc. This kind of reasoning seems plausible because it is the same reason that kindergarten students are separated from other grade levels and tested on their own by other states (e.g., MI). The SEM_n and SEM values are small and within acceptable ranges at each grade level.

Table 5.1
Descriptive Statistics and Reliability by Grade Span and Modality

Grade	Test	Number of Items	Max Points	N-Count	RS Mean	SD	Mean P-value	SEMn	Reliability	SEM
K-1	Listening	24	24	51,596	20.61	3.43	0.85	0.02	0.81	1.49
	Speaking	16	34	51,596	26.75	7.13	0.77	0.03	0.93	1.90
	Reading	15	15	51,607	11.14	3.50	0.74	0.02	0.83	1.44
	Writing	15	24	51,607	18.45	4.80	0.78	0.02	0.86	1.79
	Listening and Speaking	40	58	51,596	47.36	9.60	0.82	0.04	0.92	2.66
	Reading and Writing	30	39	51,607	29.59	7.83	0.76	0.03	0.91	2.33
2-4	Listening	24	24	62,871	19.36	3.89	0.80	0.02	0.81	1.72
	Speaking	16	34	62,871	28.69	6.36	0.83	0.03	0.93	1.68
	Reading	24	24	60,816	17.62	4.25	0.72	0.02	0.82	1.82
	Writing	16	22	60,816	15.68	4.22	0.73	0.02	0.82	1.80
	Listening and Speaking	40	58	62,871	48.05	9.28	0.81	0.04	0.92	2.58
	Reading and Writing	40	46	60,816	33.30	7.90	0.72	0.03	0.89	2.59
5-6	Listening	25	25	28,467	19.18	4.35	0.76	0.03	0.82	1.86
	Speaking	16	34	28,467	28.33	7.25	0.82	0.04	0.95	1.68
	Reading	27	27	27,428	19.62	5.53	0.70	0.03	0.87	1.98
	Writing	19	25	27,428	17.99	4.53	0.71	0.03	0.82	1.94
	Listening and Speaking	41	59	28,467	47.51	10.59	0.78	0.06	0.93	2.73
	Reading and Writing	46	52	27,428	37.61	9.40	0.71	0.06	0.91	2.80
7-8	Listening	25	25	24,429	17.60	4.90	0.69	0.03	0.84	1.97
	Speaking	16	34	24,429	27.48	8.19	0.79	0.05	0.96	1.68
	Reading	27	27	23,239	18.04	5.90	0.64	0.04	0.87	2.10
	Writing	19	25	23,239	17.67	4.87	0.69	0.03	0.83	2.03
	Listening and Speaking	41	59	24,429	45.08	11.98	0.73	0.08	0.94	2.87
	Reading and Writing	46	52	23,239	35.71	9.99	0.66	0.07	0.91	2.95
9-12	Listening	25	25	39,978	17.67	5.29	0.70	0.03	0.86	1.98
	Speaking	16	34	39,978	27.29	7.77	0.79	0.04	0.95	1.69
	Reading	27	27	38,098	17.41	5.15	0.63	0.03	0.82	2.16
	Writing	19	25	38,098	18.48	4.64	0.72	0.02	0.82	1.97
	Listening and Speaking	41	59	39,978	44.96	11.96	0.74	0.06	0.94	2.83
	Reading and Writing	46	52	38,098	35.89	8.93	0.67	0.05	0.89	2.96

Note: The data file used in the table above is the final research file. The total n-counts for grades K-12 are 220,332 and includes private schools. 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).

Table 5.2
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	24,257	19.38	3.69	0.81	0.02	0.79	1.69
	Speaking	16	34	24,257	24.72	7.47	0.74	0.05	0.93	2.04
	Reading	15	15	24,146	9.19	3.33	0.61	0.02	0.75	1.68
	Writing	15	24	24,146	15.92	4.65	0.84	0.03	0.83	1.89
	Listening and Speaking	40	58	24,257	44.10	10.00	0.78	0.06	0.92	2.90
	Reading and Writing	30	39	24,146	25.11	7.29	0.68	0.05	0.88	2.57
1	Listening	24	24	27,339	21.70	2.76	0.90	0.02	0.78	1.28
	Speaking	16	34	27,339	28.55	6.29	0.85	0.04	0.92	1.74
	Reading	15	15	27,461	12.86	2.66	0.86	0.02	0.80	1.18
	Writing	15	24	27,461	20.68	3.70	0.93	0.02	0.83	1.52
	Listening and Speaking	40	58	27,339	50.25	8.21	0.88	0.05	0.92	2.38
	Reading and Writing	30	39	27,461	33.54	5.92	0.88	0.04	0.89	1.95
2	Listening	24	24	23,653	18.12	3.87	0.75	0.03	0.77	1.87
	Speaking	16	34	23,653	28.11	6.26	0.83	0.04	0.92	1.76
	Reading	24	24	22,603	15.98	4.25	0.67	0.03	0.79	1.94
	Writing	16	22	22,603	14.14	4.22	0.68	0.03	0.80	1.88
	Listening and Speaking	40	58	23,653	46.23	9.03	0.79	0.06	0.91	2.74
	Reading and Writing	40	46	22,603	30.11	7.83	0.67	0.05	0.88	2.73
3	Listening	24	24	21,286	19.85	3.60	0.83	0.02	0.79	1.66
	Speaking	16	34	21,286	29.04	6.15	0.86	0.04	0.93	1.64
	Reading	24	24	20,749	18.18	3.93	0.76	0.03	0.79	1.78
	Writing	16	22	20,749	16.24	3.91	0.79	0.03	0.80	1.74
	Listening and Speaking	40	58	21,286	48.90	8.82	0.84	0.06	0.92	2.51
	Reading and Writing	40	46	20,749	34.42	7.24	0.77	0.05	0.88	2.52
4	Listening	24	24	17,932	20.42	3.82	0.85	0.03	0.84	1.55
	Speaking	16	34	17,932	29.04	6.68	0.86	0.05	0.94	1.60
	Reading	24	24	17,464	19.07	3.91	0.79	0.03	0.81	1.69
	Writing	16	22	17,464	17.01	3.94	0.82	0.03	0.81	1.70
	Listening and Speaking	40	58	17,932	49.46	9.73	0.85	0.07	0.94	2.43
	Reading and Writing	40	46	17,464	36.08	7.32	0.80	0.06	0.89	2.42

Note: The data file used in the table above is the final research file including both public and non-public schools. The total n-counts for grades K–12 are 220,332. 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).

Table 5.2 (continued)
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
5	Listening	25	25	15,425	19.16	4.27	0.77	0.03	0.81	1.87
	Speaking	16	34	15,425	28.55	6.97	0.85	0.06	0.94	1.67
	Reading	27	27	14,832	19.60	5.46	0.73	0.04	0.87	1.99
	Writing	19	25	14,832	18.04	4.39	0.74	0.04	0.81	1.93
	Listening and Speaking	41	59	15,425	47.71	10.24	0.80	0.08	0.93	2.72
	Reading and Writing	46	52	14,832	37.64	9.18	0.73	0.08	0.91	2.79
6	Listening	25	25	13,042	19.21	4.44	0.77	0.04	0.83	1.85
	Speaking	16	34	13,042	28.07	7.55	0.83	0.07	0.95	1.69
	Reading	27	27	12,596	19.63	5.61	0.73	0.05	0.88	1.97
	Writing	19	25	12,596	17.94	4.69	0.73	0.04	0.83	1.96
	Listening and Speaking	41	59	13,042	47.28	11.00	0.79	0.10	0.94	2.75
	Reading and Writing	46	52	12,596	37.57	9.66	0.73	0.09	0.92	2.80
7	Listening	25	25	12,252	17.28	4.87	0.69	0.04	0.83	2.00
	Speaking	16	34	12,252	27.36	8.20	0.81	0.07	0.96	1.71
	Reading	27	27	11,617	17.54	5.91	0.65	0.05	0.87	2.13
	Writing	19	25	11,617	17.46	4.86	0.71	0.05	0.82	2.04
	Listening and Speaking	41	59	12,252	44.64	11.93	0.74	0.11	0.94	2.91
	Reading and Writing	46	52	11,617	34.99	9.96	0.67	0.09	0.91	2.98
8	Listening	25	25	12,177	17.93	4.90	0.72	0.04	0.84	1.94
	Speaking	16	34	12,177	27.59	8.18	0.82	0.07	0.96	1.65
	Reading	27	27	11,622	18.54	5.84	0.69	0.05	0.87	2.07
	Writing	19	25	11,622	17.89	4.87	0.72	0.05	0.83	2.02
	Listening and Speaking	41	59	12,177	45.52	12.02	0.76	0.11	0.94	2.82
	Reading and Writing	46	52	11,622	36.42	9.96	0.70	0.09	0.91	2.92
9	Listening	25	25	13,426	16.51	5.73	0.66	0.05	0.87	2.03
	Speaking	16	34	13,426	25.80	8.95	0.76	0.08	0.96	1.77
	Reading	27	27	12,371	16.14	5.32	0.60	0.05	0.83	2.22
	Writing	19	25	12,371	17.55	5.12	0.71	0.05	0.84	2.03
	Listening and Speaking	41	59	13,426	42.31	13.60	0.70	0.12	0.95	2.98
	Reading and Writing	46	52	12,371	33.68	9.58	0.64	0.09	0.90	3.04

Note: The data file used in the table above is the final research file including both public and non-public schools. The total n-counts for grades K–12 are 220,332. 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).

Table 5.2 (continued)
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
10	Listening	25	25	12,495	17.32	5.35	0.69	0.05	0.86	2.00
	Speaking	16	34	12,495	26.82	7.81	0.79	0.07	0.95	1.73
	Reading	27	27	11,959	17.19	5.08	0.64	0.05	0.82	2.18
	Writing	19	25	11,959	18.19	4.68	0.73	0.04	0.82	1.99
	Listening and Speaking	41	59	12,495	44.13	12.03	0.73	0.11	0.94	2.88
	Reading and Writing	46	52	11,959	35.38	8.88	0.67	0.08	0.89	2.99
11	Listening	25	25	8,295	18.76	4.62	0.75	0.05	0.83	1.92
	Speaking	16	34	8,295	28.52	6.40	0.84	0.07	0.94	1.63
	Reading	27	27	8,088	18.47	4.77	0.68	0.05	0.80	2.12
	Writing	19	25	8,088	19.26	4.06	0.76	0.05	0.79	1.88
	Listening and Speaking	41	59	8,295	47.28	9.83	0.79	0.11	0.92	2.69
	Reading and Writing	46	52	8,088	37.73	7.96	0.71	0.09	0.87	2.86
12	Listening	25	25	5,762	19.57	4.02	0.78	0.05	0.79	1.86
	Speaking	16	34	5,762	29.99	5.16	0.89	0.07	0.92	1.49
	Reading	27	27	5,680	19.14	4.58	0.71	0.06	0.79	2.08
	Writing	19	25	5,680	20.00	3.58	0.79	0.05	0.74	1.82
	Listening and Speaking	41	59	5,762	49.56	7.86	0.82	0.10	0.90	2.52
	Reading and Writing	46	52	5,680	39.14	7.25	0.74	0.10	0.85	2.79

Note: The data file used in the table above is the final research file including both public and non-public schools. The total n-counts for grades K–12 are 220,332. 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).

5.6 Inter-Rater Reliability

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

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

To ensure that local teachers apply the same rigorous scoring standards as intended by the New York State Education Department (NYSED), and to provide evidence of inter-rater reliability, the Department requires that Pearson rescore 10% of all test papers after each test administration. The selection of the 10% audit sample for each grade span followed a stratified random selection procedure based on the Need/Resource Category, which divided the State schools into six categories: New York City, Big 4 Cities (Buffalo, Rochester, Syracuse, and Yonkers), High Need Urban/Suburbs, High Need Rural, Average Need, and Low Need.

Target values were calculated for each level using the percentage of enrollment data for each of the above classifications and was 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 2009 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 Pearson for rescoring. The following sections (or tables) indicate the procedures that Pearson conducted to ensure reliability and accurate scoring of the items.

Table 5.3 provides the audit sample information regarding the number of students obtained for the rater analyses after merging the auditing file (Accudata) and final research file. The Accudata and the research file were merged according to student ID (SID). The n-counts in the table are based on 88% of the Accudata file from auditing, matching the research file database. The percentage of the audit sample ranged from 7% to 18% across the five grade spans.

Table 5.3
Audit Sample

Grade Span	Number of Students in Audit Sample after Merging ¹	Number of Students in Population by Grade Span	Percentage of Total Population ²
K-1	3,964	54,093	7
2-4	4,866	65,327	7
5-6	4,618	29,945	15
7-8	4,649	25,824	18
9-12	4,297	41,571	10
Total	22,394	216,760	10

Note 1: There are 2,923 Accudata students that did not merge with the final research file and 6 Accudata students that did merge with the final research file but had missing data. Note 2: Percentage of Total Population = Number of Students in audit sample/number of students in final research file.

Inter-Rater Agreement

Table 5.4 provides the rater agreement for the Pre-writing and Writing constructed-response items between local raters and Pearson raters. The Pearson raters scored the items independently. When the two raters assigned the same score to a student's paper, the agreement rating was denoted as *exact* (i.e., perfect agreement). Ratings that differed by exactly one score point were denoted as *adjacent*. Ratings that differed by two or more score points were denoted as *non-adjacent*. The results are presented by grade span and 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 Pearson Raters

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

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

In Table 5.4, for the 2-point items, the percent of exact rating ranged from 68.23 to 96.44 across all the grade spans. The percent of adjacent rating ranged from 1.54 to 28.93. The total percent of the exact and adjacent ratings ranged from 97.02 to 99.75. The percent of exact rating is relatively low for item 4 (Grade 2-12) because it has a greater range (0-4). The percent of exact rating ranged from 49.90 to 55.07 across the four grade spans. In addition, the percent of adjacent rating ranged from 37.94 to 42.96. The total percent of the exact and adjacent ratings for the four-point items ranged from 92.85 to 93.88.

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

Intraclass Correlation

The intraclass correlation assesses rating reliability by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects. The intraclass correlations in Table 5.4 were calculated by using the random and fixed effects model introduced by Shrout and Fleiss (1979) to measure the agreement of the local and Pearson raters. The SAS program from Douglas Steinley and Philip Karl Wood (2000) was modified so that large student

samples could be used for the analyses. The intraclass correlations ranged from 0.57 to 0.82 across all the grade spans. The lowest correlations occurred at the 9–12 grade span.

Kappa Coefficient

Kappa coefficient is an index measuring the rater agreement, with the value always less than or equal to 1. A value of 1 implies perfect agreement, and a value less than 1 implies less than perfect agreement. In rare situations, Weighted Kappa can be negative. This is a sign that the two observers agreed that less than would be expected just by chance. It is also rare that we get perfect agreement. Different people have different interpretations as to what is a good level of agreement. Here is an interpretation of the Weighted Kappa coefficient (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 5.4, 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 limited range.

Table 5.4
Rater Agreement for Pre-Writing and Writing Prompts

Grade Span	Item	Max Point	N-count	Exact	Adjacent	Non-Adjacent	Intraclass Correlation	Weighted Kappa	ASE	95% Confidence Limits	
										Lower	Upper
K-1	1	2	3,964	79.94	19.73	0.33	0.58	0.57	0.01	0.55	0.60
	2	2	3,964	77.30	22.45	0.25	0.58	0.56	0.01	0.54	0.59
	3	2	3,964	96.44	1.99	1.56	0.69	0.71	0.03	0.66	0.75
	4	2	3,964	96.42	1.54	2.04	0.71	0.73	0.02	0.68	0.78
	5	2	3,964	87.59	9.54	2.88	0.81	0.76	0.01	0.74	0.78
	6	2	3,964	82.72	14.35	2.93	0.82	0.73	0.01	0.71	0.74
	7	2	3,964	75.78	22.00	2.22	0.75	0.70	0.01	0.68	0.71
	8	2	3,964	73.26	24.67	2.07	0.74	0.67	0.01	0.65	0.69
	9	2	3,964	73.01	24.80	2.19	0.74	0.66	0.01	0.64	0.68
2-4	1	2	4,866	85.02	13.44	1.54	0.58	0.56	0.01	0.53	0.59
	2	2	4,866	79.31	17.82	2.88	0.69	0.64	0.01	0.61	0.66
	3	2	4,866	71.87	25.89	2.24	0.63	0.58	0.01	0.56	0.60
	4	4	4,866	53.27	40.61	6.12	0.60	0.46	0.01	0.44	0.48
5-6	1	2	4,618	87.57	11.15	1.28	0.69	0.67	0.01	0.64	0.69
	2	2	4,618	83.65	14.81	1.54	0.71	0.70	0.01	0.68	0.72
	3	2	4,618	68.23	28.93	2.84	0.61	0.56	0.01	0.53	0.58
	4	4	4,618	53.16	39.69	7.15	0.65	0.49	0.01	0.48	0.51
7-8	1	2	4,649	78.02	19.51	2.47	0.72	0.68	0.01	0.66	0.70
	2	2	4,649	92.17	6.73	1.10	0.81	0.80	0.01	0.78	0.82
	3	2	4,649	77.31	20.89	1.81	0.67	0.65	0.01	0.63	0.67
	4	4	4,649	55.07	37.94	6.99	0.70	0.56	0.01	0.54	0.58
9-12	1	2	4,297	86.85	10.87	2.28	0.71	0.66	0.01	0.63	0.68
	2	2	4,297	85.50	12.12	2.37	0.72	0.68	0.01	0.66	0.70
	3	2	4,297	69.77	27.25	2.98	0.57	0.53	0.01	0.51	0.55
	4	4	4,297	49.90	42.96	7.14	0.66	0.51	0.01	0.49	0.52

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

Table 5.5 provides the frequency distribution of the score point differences between the local raters’ and Pearson raters’ scores by grade span. For the 2-point items, the percentage of 0 differences ranged from 68.23 to 96.44. For the 4-point items, the percentage of 0 differences ranged from 49.90 to 55.07. The above evidence indicates that the rater-agreement percentage is higher for the 2-point item than for the 4-point item.

Table 5.5
Percentages of the Score Difference Between Raters

Grade Span	Item	Percentage of Point Difference								
		Score Difference (Local score minus Pearson score)								
		-4	-3	-2	-1	0	1	2	3	4
K-1	1			0.10	5.50	79.94	14.23	0.23		
	2			0.10	6.61	77.30	15.84	0.15		
	3			0.71	1.06	96.44	0.93	0.86		
	4			1.11	0.76	96.42	0.78	0.93		
	5			1.29	3.83	87.59	5.70	1.59		
	6			1.16	4.54	82.72	9.81	1.77		
	7			1.19	10.54	75.78	11.45	1.03		
	8			1.14	13.19	73.26	11.48	0.93		
	9			0.93	9.56	73.01	15.24	1.26		
2-4	1			0.76	6.31	85.02	7.13	0.78		
	2			1.50	8.78	79.31	9.04	1.38		
	3			0.97	12.43	71.87	13.46	1.27		
	4	0.06	0.16	1.87	15.45	53.27	25.15	3.68	0.35	
5-6	1			0.84	5.26	87.57	5.89	0.43		
	2			0.89	5.13	83.65	9.68	0.65		
	3			1.00	11.17	68.23	17.76	1.84		
	4		0.19	1.17	10.63	53.16	29.06	5.28	0.45	0.04
7-8	1			1.48	11.21	78.02	8.30	0.99		
	2			0.69	2.95	92.17	3.79	0.41		
	3			1.23	10.05	77.31	10.84	0.58		
	4	0.06	0.39	1.40	11.64	55.07	26.31	4.71	0.39	0.04
9-12	1			1.21	5.47	86.85	5.40	1.07		
	2			1.28	5.79	85.50	6.33	1.09		
	3			1.35	9.91	69.77	17.34	1.63		
	4	0.02	0.47	1.58	12.85	49.90	30.11	4.54	0.47	0.07

Table 5.6 provides the mean and standard deviation of each item for both the local raters and the Pearson raters. There is a column for the local raters, a column for Pearson raters, and a third column for the difference of the mean scores and ratio of standard deviations between the local and Pearson raters. The largest mean difference (0.28) occurred at Grade Spans 5–6 (item 4), and SD ratio ranged from 0.90 to 1.28. For all items, the mean difference is close to 0, and SD ratio is close to 1, which is a strong indication of rater agreement.

Table 5.6
Comparison Between Local and Pearson Raters

Grade Span	Item	Local		Pearson		Differences	
		Mean	SD	Mean	SD	Mean	SD Ratio
K-1	1	1.70	0.49	1.61	0.52	0.09	0.94
	2	1.58	0.52	1.49	0.53	0.09	0.98
	3	1.91	0.39	1.91	0.40	0.00	0.99
	4	1.89	0.45	1.89	0.44	0.00	1.01
	5	1.60	0.70	1.58	0.73	0.02	0.97
	6	1.51	0.72	1.44	0.80	0.06	0.90
	7	1.22	0.81	1.21	0.82	0.01	0.98
	8	1.21	0.80	1.23	0.83	-0.02	0.97
	9	1.00	0.78	0.93	0.79	0.06	0.99
2-4	1	1.78	0.50	1.78	0.50	0.01	0.98
	2	1.56	0.69	1.56	0.70	0.00	0.99
	3	1.36	0.69	1.34	0.71	0.02	0.96
	4	1.92	0.99	1.79	0.84	0.14	1.18
5-6	1	1.75	0.54	1.75	0.53	0.00	1.02
	2	1.63	0.65	1.59	0.67	0.04	0.98
	3	1.35	0.73	1.26	0.74	0.08	0.99
	4	2.14	1.09	1.86	0.86	0.28	1.28
7-8	1	1.41	0.77	1.44	0.75	-0.04	1.03
	2	1.73	0.58	1.72	0.58	0.00	1.00
	3	1.47	0.72	1.47	0.70	0.00	1.03
	4	2.24	1.15	2.03	1.02	0.21	1.13
9-12	1	1.73	0.58	1.74	0.60	0.00	0.97
	2	1.67	0.64	1.67	0.64	0.00	0.99
	3	1.47	0.69	1.39	0.70	0.08	0.99
	4	2.47	1.15	2.24	0.97	0.23	1.18

Summary

The evidence of inter-rater reliability statistics (percentage of agreement, intraclass correlation, Weighted Kappa coefficient, percentage of score point difference, and mean difference) indicated that the local raters and Pearson raters generally have good agreement regarding the NYSESLAT scores. However, for a few items (e.g., item 4 in grade span 5-6), the agreement between local raters and Pearson raters was poor. In these cases, the local raters generally gave higher scores than those given by Pearson raters. To improve the agreement between local raters and Pearson raters in later administrations, the scoring guides may need to be improved or the teachers who score the test may need to be trained better, or both.

5.7 Reliability of Classification Decision at Proficient Cut

Based on the NYSESLAT scale scores, student performance is classified into one of four proficiency levels. While it is always important to know the reliability of student scores in any examination, it is of even greater importance to assess the reliability of the decisions based on these scores. Evaluation of the reliability of classification decisions is performed through estimation of the probabilities of correct and consistent classification of student performance.

Procedures from Livingston and Lewis (1995) were applied to derive measures of the accuracy and consistency of the classifications. The brief descriptions of the procedures used and results obtained are presented in this section.

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

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

Note: Adapted from Young and Yoon (1998).

Figure 5. 1: Classification Accuracy

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

Note: Adapted from Young and Yoon (1998).

Figure 5. 2: Classification Consistency

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

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

These analyses make use of the techniques outlined and implemented by Hanson (1991), Haertel (1996), Livingston and Lewis (1995), and Young and Yoon (1998). The software developed by

Hanson (1995) was used for the analyses. Estimates of decision accuracy and consistency were made for the Achieves Proficient Status cut points on the Listening/Speaking and Reading/Writing scores reported in the NYSESLAT.

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

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

- Accuracy
- False positives
- False negatives
- Consistency

Decision accuracy refers to the agreement between the classifications based on the form actually taken and the classifications that would be made if each student could somehow be tested with all possible forms (or true score) of the examination. For example, for K students, 85% 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. Nine 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 we don't like to see high percentages of students who were not proficient classified as proficient and exit the program. Generally speaking, the NYSESLAT shows much higher percentages of accuracy and consistency in comparison to the false positive and negative percentages.

Table 5.7
Classification Accuracy and Consistency by Grade

Grade	Test	Accuracy	False Positives	False Negatives	Consistency
K	Listening and Speaking	0.89	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 table above is the final research file. The total n-counts for grades K–12 are 220,332. The stats 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). Additionally, cases with missing grades are also removed.

PART 6: VALIDITY

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

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

6.1 Content Validity

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

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

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

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

The assigned standards and performance indicators were one of several criteria (e.g., item statistics, content diversity, gender and ethnic balance, etc.) that were then considered by Pearson English language learner assessment specialists and NYSED assessment and content specialists during forms construction to ensure that forms reflected broad coverage of the *Learning Standards*. After forms construction was completed, item maps were developed for each form to show this coverage item by item. English language learner assessment specialists at Pearson and content specialists at NYSED reviewed the item maps to confirm the alignment of a given form to the *New York State Learning Standards for English as a Second Language*. The item mapping provided in Table 6.1 and Appendices A.1–A.2 gives concrete evidence for the alignment to the *New York State Learning Standards for English as a Second Language*.

6.2 Internal Structure

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

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

In order to investigate the answers to these questions, correlations based on raw scores were obtained from the four modalities. Tables 6.1 and 6.2 present the intercorrelation among the four modalities by grade span and grade. The evidence of internal structure of the 2009 NYSESLAT is also depicted by the point-biserial correlation coefficient and fit statistics. Appendices B.1–B.5 and D.1–D.5 provide these statistics for the 2009 NYSESLAT administration.

Table 6.1
Intercorrelation Among the Modalities by Grade Span

Grade Span	Test	Correlation		Coefficient	
		Listening	Speaking	Reading	Writing
K-1	Listening	1.00			
	Speaking	0.59	1.00		
	Reading	0.55	0.45	1.00	
	Writing	0.57	0.51	0.77	1.00
2-4	Listening	1.00			
	Speaking	0.57	1.00		
	Reading	0.63	0.44	1.00	
	Writing	0.64	0.53	0.74	1.00
5-6	Listening	1.00			
	Speaking	0.59	1.00		
	Reading	0.65	0.52	1.00	
	Writing	0.66	0.62	0.74	1.00
7-8	Listening	1.00			
	Speaking	0.61	1.00		
	Reading	0.69	0.52	1.00	
	Writing	0.70	0.67	0.71	1.00
9-12	Listening	1.00			
	Speaking	0.63	1.00		
	Reading	0.66	0.49	1.00	
	Writing	0.69	0.65	0.66	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 220,332. The statistics presented in the table are based on raw scores with the n-counts after cleaning (removal of duplicate cases, off-level cases, and all records with Strand Raw Score = 0 or > 900). Additionally, cases with a missing grade are also removed.

Table 6.2
Intercorrelation Among the Modalities by Grade

Grade	Test	Correlation		Coefficient	
		Listening	Speaking	Reading	Writing
K	Listening	1.00			
	Speaking	0.54	1.00		
	Reading	0.44	0.33	1.00	
	Writing	0.48	0.41	0.66	1.00
1	Listening	1.00			
	Speaking	0.57	1.00		
	Reading	0.50	0.43	1.00	
	Writing	0.51	0.50	0.72	1.00
2	Listening	1.00			
	Speaking	0.51	1.00		
	Reading	0.56	0.39	1.00	
	Writing	0.58	0.48	0.71	1.00

Table 6.2 (continued)
Intercorrelation Among the Modalities by Grade

Grade	Test	Correlation		Coefficient	
		Listening	Speaking	Reading	Writing
3	Listening	1.00			
	Speaking	0.57	1.00		
	Reading	0.59	0.45	1.00	
	Writing	0.60	0.54	0.70	1.00
4	Listening	1.00			
	Speaking	0.65	1.00		
	Reading	0.65	0.52	1.00	
	Writing	0.66	0.62	0.73	1.00
5	Listening	1.00			
	Speaking	0.57	1.00		
	Reading	0.65	0.49	1.00	
	Writing	0.64	0.59	0.73	1.00
6	Listening	1.00			
	Speaking	0.61	1.00		
	Reading	0.66	0.55	1.00	
	Writing	0.68	0.65	0.75	1.00
7	Listening	1.00			
	Speaking	0.60	1.00		
	Reading	0.68	0.51	1.00	
	Writing	0.69	0.66	0.71	1.00
8	Listening	1.00			
	Speaking	0.63	1.00		
	Reading	0.70	0.54	1.00	
	Writing	0.70	0.68	0.72	1.00
9	Listening	1.00			
	Speaking	0.68	1.00		
	Reading	0.68	0.53	1.00	
	Writing	0.72	0.70	0.68	1.00
10	Listening	1.00			
	Speaking	0.63	1.00		
	Reading	0.66	0.50	1.00	
	Writing	0.69	0.66	0.65	1.00
11	Listening	1.00			
	Speaking	0.56	1.00		
	Reading	0.61	0.41	1.00	
	Writing	0.63	0.56	0.62	1.00
12	Listening	1.00			
	Speaking	0.44	1.00		
	Reading	0.58	0.35	1.00	
	Writing	0.55	0.47	0.56	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 220,332. The stats 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). Additionally, cases with missing grade are also removed.

Observations of language proficiency assessment subtests in Table 6.1:

- Listening and Speaking are moderately correlated for Grades K–12 ranging from 0.57–0.63.
- Listening and Reading are moderately correlated for Grades K–12 ranging from 0.55–0.69.
- Listening and Writing are moderately correlated for Grades K–12 ranging from 0.57–0.70.
- Speaking and Reading are moderately correlated for Grades K–12 ranging from 0.44–0.52.
- Speaking and Writing are moderately correlated for Grades K–12 ranging from 0.51–0.67.
- Reading and Writing are moderately correlated for Grades K–12 ranging from 0.66–0.77.
- The language domain pairs of Listening and Speaking and Reading and Writing are moderately to highly correlated, while Speaking and Reading are not highly correlated.

Generally, the internal correlations based on grade spans are within expectation. The Reading and Writing subtests have the highest correlation range (from 0.66 to 0.77) across the five grade spans. Speaking and Reading subtests have the lowest correlation range (from 0.44 to 0.52). Listening and Speaking, Listening and Reading, Listening and Writing, and Speaking and Writing have correlation ranges between 0.51 and 0.70.

Observations of language proficiency assessment subtests in Table 6.2:

Kindergarten

- Students in this age group do not usually read or write yet, but can have Listening and Speaking skills. The expected outcome is that neither Reading nor Writing will correlate highly with Listening or Speaking.
- In Table 6.2, for kindergarten students, Speaking and Listening have a correlation of 0.54, and Reading and Writing have a correlation of 0.66. Speaking and Reading have a low correlation of 0.33, Listening and Reading have a low correlation of 0.44, Speaking and Writing have a correlation of 0.41, and Listening and Writing have a correlation of 0.48.

Grades 1–8

- A trend of increase in the correlation (from 0.51 to 0.70) between Listening and Writing is observed.
- A possible explanation (personal communication with Pearson content specialists) is that, in general, students during this age span experience expanding use and development in their Writing skills. At the same time, demands on the Listening skills of this age group remain static with only moderate development.

Grades 9–12

- A steady decrease in the correlation (from 0.72 to 0.55) between Listening and Writing is observed.
- A possible explanation (personal communication with Pearson content specialists) is that by high school, there is an increased focus on the use of writing skills, especially an increased focus on academic content. Requirements for listening skills in high-school age students also increase, but not at nearly as steep of a curve as writing.
- 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 0.48 to 0.68) for grades 1–8, and there is a steady decrease in the correlation from (0.70 to 0.47) for grades 9–12.

Generally, the internal correlations based on grade levels are within expectation. The Reading and Writing subtests have the highest correlation range (from 0.56 to 0.75) across the thirteen grades. Speaking and Reading have the lowest correlation range (from 0.33 to 0.55). Listening and Speaking, Listening and Reading, Listening and Writing, and Speaking and Writing have correlation ranges between 0.41 and 0.72.

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

In order to assess the validity of the scoring structure (i.e., Speaking and Listening as one score and Reading and Writing as the other), a confirmatory factor analysis was undertaken with the 2009 full population data in order to compare the two-factor modality combinations with the unidimensional model (i.e., the total test providing a single score).

The 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 6.3–6.7 below; Appendix F provides the 2009 CFA analyses in their entirety.)

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

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	7963.838	535.992
GFI	0.930	0.995
AGFI	0.652	0.975
SRMR	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 6.4
Global Fit Indices for the One- and Two-Factor Models (Grades 2–4)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	5855.534	2107.898
GFI	0.957	0.984
AGFI	0.783	0.920
SRMR	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 6.5
Global Fit Indices for the One- and Two-Factor Models (Grades 5–6)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	1572.171	2500.706
GFI	0.974	0.959
AGFI	0.870	0.797
SRMR	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 6.6
Global Fit Indices for the One- and Two-Factor Models (Grades 7–8)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	1661.807	2243.666
GFI	0.968	0.958
AGFI	0.842	0.789
SRMR	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 6.7
Global Fit Indices for the One- and Two-Factor Models (Grades 9–12)

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	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.

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

In order to ascertain the external structure of the NYSESLAT vis-à-vis an external criterion, data from two other tests were used. For English language learners in Grades 3–8, data from those who took both the NYSESLAT in May 2009, and the New York State English Language Assessment (ELA) in January 2009 were analyzed. In New York State, all English language

learners, except for the first year ESL students, are required to take the ELA test. For English language learners in Grades 10–12, data from those who took both the NYSESLAT in May 2009 and the New York State Regents Examination in Comprehensive English (Regents English exam) in June 2009 were analyzed. The Regents English exam is an end-of-course exam. All high school students, including English language learners, are required to pass the Regents English exam for the high school graduation requirement. Students can take the exam in any grade, although most students take the exam in Grade 11.

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

6.3.1 Relationship with the New York English Language Arts Assessment

Correlations and Other Descriptive Statistics

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

The correlations between the New York ELA test and NYSESLAT raw score total of the three modalities (Reading, Writing, and Listening) ranged from 0.67 to 0.73 across the six grades. The correlations are positive between the two tests, which are within expectation because the modalities tested are Reading, Writing, and Listening for the New York ELA test, and the total raw scores include those three modalities. The correlations between New York ELA test and NYSESLAT combined scale scores of the four modalities ranged from 0.64 to 0.68 across the six grades. The correlations in Table 6.9 are positive, but slightly lower than the correlations in Table 6.8, which is within expectation because the NYSESLAT total combined scale scores include the Speaking modality, whereas the New York ELA test does not include the Speaking part.

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

Table 6.8
Descriptive Statistics of the NYSESLAT Raw Scores of Listening, Speaking, Reading, and Writing Modalities and Their Correlation with the ELA Exam Total Scale Scores.

Grade	N	Test	Max Points	Mean	SD	Sample Min.	Sample Max.	Correlation with ELA
03	17848	Listening	24	20.43	2.91	0	24	0.55
03	17848	Speaking	34	30.47	4.17	0	34	0.40
03	17848	Reading	24	18.45	3.73	0	24	0.66
03	17848	Writing	22	16.58	3.82	0	22	0.66
03	17848	LI/RD/WR Total	70	55.46	9.04	3	70	0.73
04	14555	Listening	24	21.25	2.59	2	24	0.52
04	14555	Speaking	34	30.98	3.86	0	34	0.39
04	14555	Reading	24	19.50	3.46	1	24	0.65
04	14555	Writing	22	17.56	3.56	0	22	0.65
04	14555	LI/RD/WR Total	70	58.31	8.28	13	70	0.71
05	12380	Listening	25	19.98	3.28	3	25	0.51
05	12380	Speaking	34	30.60	4.24	0	34	0.40
05	12380	Reading	27	20.16	5.07	0	27	0.62
05	12380	Writing	25	18.56	4.05	0	25	0.61
05	12380	LI/RD/WR Total	77	58.70	10.83	6	77	0.67
06	10172	Listening	25	20.11	3.35	0	25	0.54
06	10172	Speaking	34	30.29	4.95	0	34	0.43
06	10172	Reading	27	20.30	5.16	1	27	0.66
06	10172	Writing	25	18.61	4.27	0	25	0.66
06	10172	LI/RD/WR Total	77	59.02	11.23	12	77	0.71
07	9388	Listening	25	18.21	4.09	0	25	0.56
07	9388	Speaking	34	29.85	5.69	0	34	0.46
07	9388	Reading	27	18.09	5.66	1	27	0.62
07	9388	Writing	25	18.03	4.60	0	25	0.62
07	9388	LI/RD/WR Total	77	54.34	12.56	10	77	0.69
08	9486	Listening	25	18.81	4.18	2	25	0.56
08	9486	Speaking	34	29.74	6.28	0	34	0.50
08	9486	Reading	27	19.02	5.64	1	27	0.62
08	9486	Writing	25	18.35	4.76	0	25	0.62
08	9486	LI/RD/WR Total	77	56.19	12.95	9	77	0.68
234	32411	Listening	24	20.80	2.80	0	24	0.51
234	32411	Speaking	34	30.70	4.04	0	34	0.39
234	32411	Reading	24	18.92	3.65	0	24	0.63
234	32411	Writing	22	17.02	3.74	0	22	0.63
234	32411	LI/RD/WR Total	70	56.74	8.83	3	70	0.69
056	22552	Listening	25	20.04	3.31	0	25	0.51
056	22552	Speaking	34	30.46	4.58	0	34	0.40
056	22552	Reading	27	20.22	5.11	0	27	0.63
056	22552	Writing	25	18.58	4.15	0	25	0.62
056	22552	LI/RD/WR Total	77	58.84	11.02	6	77	0.68

Table 6.9
Descriptive Statistics of the NYSESLAT Raw Scores of Listening, Speaking, Reading, and Writing Modalities and Their Correlation with the ELA Exam Total Scale Scores.

Grade	N	Test	Max Points	Mean	SD	Sample Min.	Sample Max.	Correlation with ELA
078	18874	Listening	25	18.51	4.14	0	25	0.52
078	18874	Speaking	34	29.80	5.99	0	34	0.47
078	18874	Reading	27	18.56	5.67	1	27	0.58
078	18874	Writing	25	18.19	4.68	0	25	0.59
078	18874	LI/RD/WR Total	77	55.27	12.79	9	77	0.64

Table 6.10
Descriptive Statistics of the New York ELA and NYSESLAT Combined Scale Scores of Listening/Speaking and Reading/Writing

Grade	N	Test	Mean	SD	Sample Min.	Sample Max.	Correlation
3	17848	NYSESLAT Scale Score	1339.90	75.02	913	1611	0.67
		NY ELA Scale Score	645.22	31.03	475	780	
4	14555	NYSESLAT Scale Score	1365.50	76.84	1003	1611	0.64
		NY ELA Scale Score	638.99	32.60	430	737	
5	12380	NYSESLAT Scale Score	1373.27	67.65	1100	1657	0.65
		NY ELA Scale Score	646.23	24.41	495	795	
6	10172	NYSESLAT Scale Score	1375.78	73.25	1096	1657	0.68
		NY ELA Scale Score	643.92	17.44	480	785	
7	9388	NYSESLAT Scale Score	1382.33	73.54	1105	1683	0.66
		NY ELA Scale Score	636.67	23.04	470	705	
8	9486	NYSESLAT Scale Score	1393.63	80.91	1124	1683	0.66
		NY ELA Scale Score	623.42	30.13	430	717	

Proficiency Level

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

Table 6.10 shows that for those Grade 3 students who were classified as beginning students on the NYSESLAT, 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, 89% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 46% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 54 % of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 14% of them were classified as not proficient (Level 1 and Level 2) on the ELA test and 86% of them were classified as proficient on the ELA test.

Table 6.11 shows that for those Grade 4 students who were classified as beginning students on the NYSESLAT, 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, 91% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 56% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 44% of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 22% of them were classified as not proficient (Level 1 and Level 2) on the ELA test; and 78% of them were classified as proficient on the ELA test.

Table 6.12 shows that for those Grade 5 students who were classified as beginning students on the NYSESLAT, 98% 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, 90% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 59% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 41% of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 21% of them were classified as not proficient (Level 1 and Level 2) on the ELA test; and 79% of them were classified as proficient on the ELA test.

Table 6.13 shows that for those Grade 6 students who were classified as beginning students on the NYSESLAT, 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, 94% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 60% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 40% of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 24% of them were classified as not proficient (Level 1 and Level 2) on the ELA test; and 76% of them were classified as proficient on the ELA test.

Table 6.14 shows that for those Grade 7 students who were classified as beginning students on the NYSESLAT, 99% 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, 93% of them were also classified as not proficient (Level 1 and Level 2) on the ELA test. For those students who were classified as advanced students on the NYSESLAT test, 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 those students who were classified as proficient students on the NYSESLAT, 30% of them were classified as not proficient (Level 1 and Level 2) on the ELA test; and 70% of them were classified as proficient on the ELA test.

Table 6.15 shows that for those Grade 8 students who were classified as beginning students on the NYSESLAT, 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 test, 88% of them were classified as not proficient (Level 1 and Level 2) on the ELA test, and 12% of them were classified as proficient on the ELA test. For those students who were classified as proficient students on the NYSESLAT, 64% of them were classified as not proficient (Level 1 and Level 2) on the ELA test; and 36% of them were classified as proficient on the ELA test.

About 89%–100% of students who were classified as beginning and intermediate on the NYSESLAT test were also classified as not proficient (Level 1 and Level 2) by the ELA cut scores across the six grade levels. About 46%–88% of students who were classified as advanced on the NYSESLAT test were also classified as not proficient (Level 1 and Level 2) by the ELA cut scores across the six grade levels. About 12%–54% of students who were classified as advanced on the NYSESLAT test were classified as proficient by the ELA cut scores across the six grade levels. The higher the grade level, the lower the percentage. About 14%–64% of students who were classified as proficient on the NYSESLAT test were classified as not proficient by the ELA cut scores across the six grade levels. About 36%–86% of students who were classified as proficient on the NYSESLAT test were also classified as proficient by the ELA cut scores across the six grade levels. The higher the grade level, the lower the percentage. In general, all the evidence above indicates that those English language learners who scored well on the NYSESLAT also scored well on the ELA test, and those who did not score well on the NYSESLAT also did not perform well on the ELA test.

Table 6.11
Grade 3: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=1052)	Intermediate (N=3806)	Advanced (N=9670)	Proficient (N=3321)
Level 1 on ELA	82%	34%	3%	0%
Level 2 on ELA	17%	55%	43%	14%
Level 3 on ELA	1%	11%	53%	81%
Level 4 on ELA	0%	0%	1%	5%
Total	100%	100%	100%	100%

Table 6.12
Grade 4: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=752)	Intermediate (N=3414)	Advanced (N=7164)	Proficient (N=3226)
Level 1 on ELA	87%	34%	4%	0%
Level 2 on ELA	12%	57%	52%	22%
Level 3 on ELA	1%	9%	44%	77%
Level 4 on ELA	0%	0%	0%	1%
Total	100%	100%	100%	100%

Table 6.13
Grade 5: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=758)	Intermediate (N=2317)	Advanced (N=6020)	Proficient (N=3285)
Level 1 on ELA	35%	6%	0%	0%
Level 2 on ELA	63%	84%	59%	21%
Level 3 on ELA	2%	10%	41%	76%
Level 4 on ELA	0%	0%	0%	3%
Total	100%	100%	100%	100%

Table 6.14
Grade 6: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=764)	Intermediate (N=2410)	Advanced (N=4998)	Proficient (N=2001)
Level 1 on ELA	8%	1%	0%	0%
Level 2 on ELA	91%	93%	60%	24%
Level 3 on ELA	1%	6%	40%	75%
Level 4 on ELA	0%	0%	0%	1%
Total	100%	100%	100%	100%

Table 6.15
Grade 7: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=1239)	Intermediate (N=3332)	Advanced (N=3140)	Proficient (N=1677)
Level 1 on ELA	21%	2%	0%	0%
Level 2 on ELA	78%	91%	68%	30%

Level 3 on ELA	1%	7%	32%	70%
Level 4 on ELA	0%	0%	0%	0%
Total	100%	100%	100%	100%

Table 6.16
Grade 8: Percentage of English language learners Scoring at Each NYSESLAT Proficiency Level

	NYSESLAT Proficiency Levels			
	Beginning (N=1288)	Intermediate (N=3423)	Advanced (N=2409)	Proficient (N=2366)
Level 1 on ELA	62%	13%	1%	0%
Level 2 on ELA	38%	85%	87%	64%
Level 3 on ELA	0%	2%	12%	36%
Level 4 on ELA	0%	0%	0%	0%
Total	100%	100%	100%	100%

Summary

The correlation analyses provide evidence of a positive relationship between the NYSESLAT and the ELA assessment (ranging from 0.67 to 0.73). This indicates that higher scores on the NYSESLAT are associated with higher scores on the ELA. The ELA and the NYSESLAT tests were developed to serve different populations of students. Thus, it is not surprising that we see only a moderate, rather than a high, positive correlation between the two tests. The analyses also confirm the validity of the NYSESLAT cut score in that about 95% to 100% (across six grade levels) of those who scored at the proficiency level on the NYSESLAT test were also classified as Level 2 or Level 3 (proficient) by the ELA test. On the other hand, the high percentages of students who were classified as not proficient by both tests is a strong indication that students who cannot do well on one test also did not perform well on the other test.

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

Correlations and Other Descriptive Statistics

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

Table 6.17
Descriptive Statistics of the NYSESLAT Raw Scores of Listening, Speaking, Reading, and Writing Modalities and Their Correlations With the Regents English Exam Total Scale Scores

Grade	N	Test	Max Points	Mean	SD	Sample Min.	Sample Max.	Correlation with Regents English
10	1,726	Listening	25	18.57	4.57	3	25	0.47
10	1,726	Speaking	34	29.08	6.15	0	34	0.39
10	1,726	Reading	27	17.96	4.81	4	27	0.53
10	1,726	Writing	25	19.07	4.44	0	25	0.53
10	1,726	LI/RD/WR Total	77	55.60	11.89	15	76	0.59
11	4,283	Listening	25	19.13	4.16	3	25	0.50
11	4,283	Speaking	34	29.17	5.69	0	34	0.39
11	4,283	Reading	27	18.77	4.51	1	27	0.56
11	4,283	Writing	25	19.47	3.98	0	25	0.57
11	4,283	LI/RD/WR Total	77	57.38	10.81	6	77	0.63
12	1,882	Listening	25	18.59	4.08	4	25	0.37
12	1,882	Speaking	34	28.82	5.71	0	34	0.27
12	1,882	Reading	27	18.04	4.42	2	27	0.43
12	1,882	Writing	25	18.97	3.91	0	25	0.43
12	1,882	LI/RD/WR Total	77	55.60	10.31	17	77	0.50
10–12	7,891	Listening	25	18.88	4.24	3	25	0.46
10–12	7,891	Speaking	34	29.07	5.80	0	34	0.36
10–12	7,891	Reading	27	18.42	4.57	1	27	0.52
10–12	7,891	Writing	25	19.26	4.08	0	25	0.52
10–12	7,891	LI/RD/WR Total	77	56.57	10.98	6	77	0.59

Table 6.18
Descriptive Statistics of the NYSESLAT Scale Scores of Listening/Speaking, Reading/Writing Modalities, and Total Scale Scores and Their Correlations With the Regents English Exam Total Scale Scores

Grade	N	Test	Mean	SD	Sample Min.	Sample Max.	Correlation with Regents English
		Regents English	49.26	20.88	1	98	–
10	1726	Listening and Reading and Writing	718.14	47.23	562	851	0.44
		Total NYSESLAT	702.08	36.30	598	860	0.59
		Total NYSESLAT	1420.22	76.18	1165	1662	0.55
11	4283	Regents English	55.88	18.74	0	100	–
		Listening and Reading and Writing	719.77	44.26	526	851	0.46
		Total NYSESLAT	707.04	34.68	505	860	0.62
		Total NYSESLAT	1426.81	71.56	1083	1711	0.59

Table 6.17 (continued)
Descriptive Statistics of the NYSESLAT Scale Scores of Listening/Speaking, Reading/Writing Modalities, and Total Scale Scores and Their Correlations With the Regents English Exam Total Scale Scores

Grade	N	Test	Mean	SD	Sample Min.	Sample Max.	Correlation with Regents English
12	1882	Regents English	59.20	15.46	2	100	–
		Listening and	714.22	41.66	573	851	0.35
		Reading and Writing	701.03	32.32	598	860	0.48
		Total NYSESLAT	1415.25	66.32	1181	1711	0.45
10–12	7891	Regents English	55.22	18.84	0	100	–
		Listening and	718.09	44.38	526	851	0.42
		Reading and Writing	704.52	34.60	505	860	0.57
		Total NYSESLAT	1422.61	71.56	1083	1711	0.54

The correlations between the scale scores of the Regents English exam and the NYSESLAT raw score total of the three modalities (Reading, Writing, and Listening) ranged from 0.50 to 0.63 across the three grade levels, which is higher than the correlations between the scale scores of the Regents English exam and any modality of the NYSESLAT raw scores. Additionally it is within expectation that the correlations between the scale scores of the Regents English exam and the NYSESLAT combined raw scores of Reading/Writing are higher than the correlations between the scale scores of the Regents English exam and the NYSESLAT combined raw scores of Listening/Speaking.

The correlations between the scale scores of the Regents English exam and the NYSESLAT combined scale scores of the four modalities ranged from 0.45 to 0.59 across the three grade levels. The correlations in Table 6.17 are positive, but slightly lower than the correlations in Table 6.16, which are within expectation because the NYSESLAT total combined scale scores include the Speaking modality, whereas the Regents English exam does not include the Speaking component. Additionally, it is within expectation that the correlations between the scale scores of the Regents English exam and the NYSESLAT combined scale scores of Reading/Writing are higher than the correlations between the scale scores of the Regents English exam and the NYSESLAT combined scale scores of Listening/Speaking.

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

Proficiency Level Classification

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

Table 6.18 shows that 63% of the grade 10 students who scored proficient on the NYSESLAT passed the Regents English exam. The percentage of students who passed the Regents English

exam decreased to 44% for students at the advanced level, and 22% for students at the intermediate and beginning levels.

Table 6.19

Grade 10: Percentage of English language learners Passing/Failing the Regents English Exam at Each NYSESLAT Proficiency Level

Regents English Exam	NYSESLAT Proficiency Levels			
	Beginning (N=124)	Intermediate (N=790)	Advanced (N=526)	Proficient (N=286)
Pass	4%	18%	44%	63%
Fail	96%	82%	56%	37%
Total	100%	100%	100%	100%

Table 6.19 shows that of the Grade 11 students who were classified as proficient on the NYSESLAT, 84% passed the Regents English exam. For those students who were classified as advanced students on the NYSESLAT, 61% passed. For those students who were classified at the beginning and intermediate levels on the NYSESLAT, the percentage of students passing the Regents English were 4% and 27% respectively.

Table 6.20

Grade 11: Percentage of English language learners Passing/Failing the Regents English Exam at Each NYSESLAT Proficiency Level

Regents English Exam	NYSESLAT Proficiency Levels			
	Beginning (N=229)	Intermediate (N=2278)	Advanced (N=1229)	Proficient (N=547)
Pass	4%	27%	61%	84%
Fail	96%	73%	39%	16%
Total	100%	100%	100%	100%

Table 6.20 shows that of the Grade 12 students who were classified as proficient on the NYSESLAT, 86% passed the Regents English exam. For those students who were classified as advanced students on the NYSESLAT, 78% passed. For those students who were classified at the beginning and intermediate levels on the NYSESLAT, the percentage of students passing the Regents English were 17% and 47% respectively.

Table 6.21

Grade 12: Percentage of English language learners Passing/Failing the Regents English Exam at Each NYSESLAT Proficiency Level

Regents English Exam	NYSESLAT Proficiency Levels			
	Beginning (N=120)	Intermediate (N=1258)	Advanced (N=362)	Proficient (N=142)
Pass	17%	47%	78%	86%
Fail	83%	53%	22%	14%
Total	100%	100%	100%	100%

Summary

The correlation analyses provide evidence of a moderately positive relationship between the NYSESLAT and the Regents English exam. The evidence suggests that the higher NYSESLAT

scores are associated with higher scores on the Regents English exam required for high school graduation.

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

PART 7: CALIBRATION, EQUATING, AND SCALING

This part introduces the IRT models and the processes that were used to calibrate, equate, and scale the NYSESLAT using these IRT models. The Rasch model (Rasch, 1960) for dichotomous items and the Partial Credit Model (PCM) (Masters, 1982) for polytomous items were used to calibrate, equate, and scale the NYSESLAT. These measurement models are regularly used to construct test forms, for scaling and equating, and to develop and maintain large item banks. All item and test analyses, including item-fit analysis, scaling, equating, diagnosis, and performance prediction were accomplished within this framework. The statistical software used to calibrate and scale the NYSESLAT was *Winsteps* Version 3.27 (Linacre & Wright, 2000).

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 (Rasch, 1960) for dichotomous items and the Partial Credit Model (Masters, 1982) for polytomous items were used for developing, scoring, and reporting the NYSESLAT assessment. These models were recommended for several reasons.

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

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

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

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

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

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

7.3 Description of Calibration Sample, Process, and Results

Calibration Sample

An equating sample (near 100% of the population) that was used for the equating analyses represented the English language learner 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 field 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), which makes the common items 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 & Wingsky, 1984). True score equating was used for the NYSESLAT by the application

of the Newton-Raphson method for finding the roots of the non-linear function associated in the true score equating process. This procedure allows observed scores to be matched to their respective theta values, which in turn are converted to more acceptable scale score values (see scale score transformation to the SELP scale equations on p. 56). These thetas for the different forms/tests are on the same scale as established by the scale transformation method discussed above. By applying the true score equating process, the 2009 NYSESLAT was successfully equated to the 2008 NYSESLAT while maintaining the established 2008 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) that Pearson received from New York. 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, Pearson used the pre-existing 2008 vertical scale³ that 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.

³ For additional details of how the 2008 vertical scale was established, please see the *2008 NYSESLAT Technical Manual*. Pearson, Inc.

Anchor Item Evaluation

There are some general guidelines (2009 NYSESLAT Form Construction Specifications) in the use of anchor items. These guidelines are as follows:

- Use an adequate number of anchor items.
- Anchor items selected for linking the two forms need to represent a miniature version of the test in terms of content.
- Anchor items need to appear in the same or similar locations as they do in the test that we are equating.
- 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 the 2009 anchor items should be very similar to the mean difficulty of the 2008 operational test.

Stability of Anchor Items

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

For the 2009 NYSESLAT, we used the displacement value of 0.50 (as is done by many large-scale assessments) for an item to be considered as an outlier⁴. We used this criterion consistently across all the grade spans. For all grade spans, the overall numbers of anchor items were around 25% of the total items. See Table 7.1 for the number of items excluded for each modality combination by grade span.

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

⁴ 0.30 was first used as the exclusion criterion; however, when 0.30 was used, after excluding items with displacement values higher than 0.30, the anchor items no longer were 20% of the entire test. After discussions with NYSED research team, 0.50 was used as the exclusion criterion.

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

Grade Span	Modality Combination	Total Number of Items	Common Items ¹ (2008 and 2009)	Percentage of Total Items	Anchor Items (2008 and 2009)	Percentage of Total Items
K-1	Listening/Speaking	40	10	0.25	10	0.25
	Reading/Writing	30	8	0.27	8	0.27
2-4	Listening/Speaking	40	10	0.25	10	0.25
	Reading/Writing	40	10	0.25	9	0.23
5-6	Listening/Speaking	41	10	0.24	10	0.24
	Reading/Writing	46	11	0.24	11	0.24
7-8	Listening/Speaking	41	11	0.27	11	0.27
	Reading/Writing	46	13	0.26	12	0.26
9-12	Listening/Speaking	41	11	0.27	11	0.27
	Reading/Writing	46	11	0.28	11	0.24

Note 1: Common items are the same items appearing in both 2008 and 2009 operational tests.

TCC and Standard Error Curves Comparison

In order to ascertain the maintenance of the scale across years, comparisons between the 2007, 2008, and 2009 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 2009 TCC plots for all grade spans (see Figures 7.1 & 7.2) indicate that the TCC shifts to the right with each progression of grade span across the Listening/Speaking and Reading/Writing vertical scales. The shift to the right with the corresponding increase in grade span clearly indicates that both the Listening/Speaking and Reading/Writing tests are ordinal across the grade spans. The TCC alignment at each grade span seems within our realm of acceptance (See Figures 7.3-7.7 & 7.13-7.17 for all the 2007, 2008, and 2009 plots). Overall, the TCCs of 2008 and 2009 align well.

The differences in terms of the SE curves among 2007, 2008, and 2009 for both the Listening/Speaking and Reading/Writing tests are minimal across all the grade spans (See Figures 7.8-7.12 & 7.18-7.22). The noticeable differences occur only at the tail portion of the SE curves.

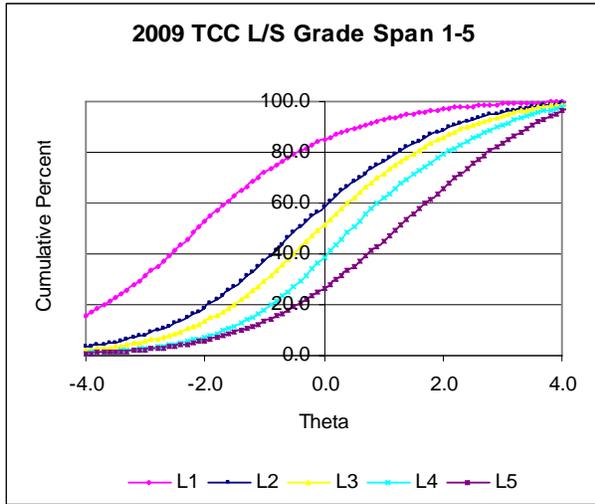


Figure 7.1
2009 TCC L/S Grade Span 1-5

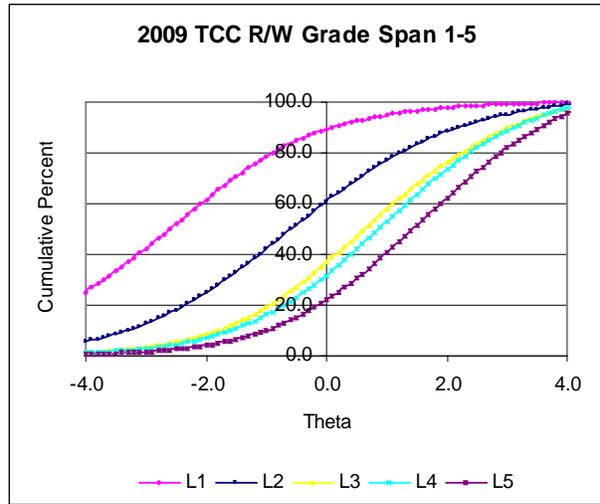


Figure 7.2
2009 TCC R/W Grade Span 1-5

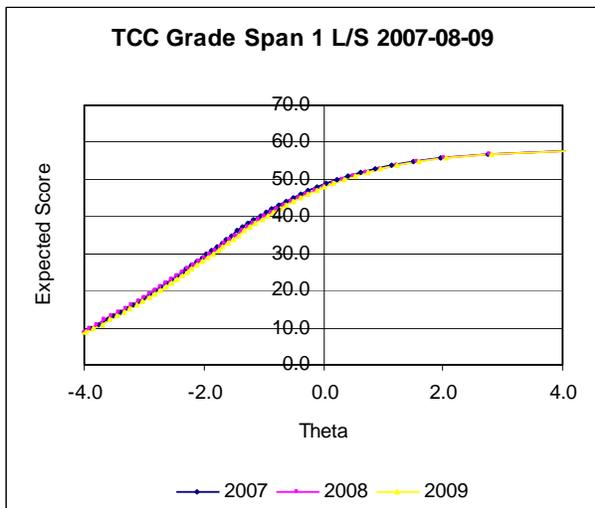


Figure 7.3
2007-08-09 TCC L/S Grade Span 1

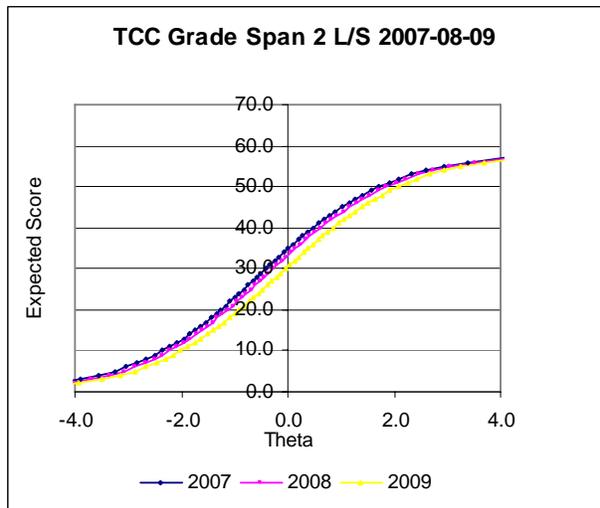


Figure 7.4
2007-08-09 TCC L/S Grade Span 2

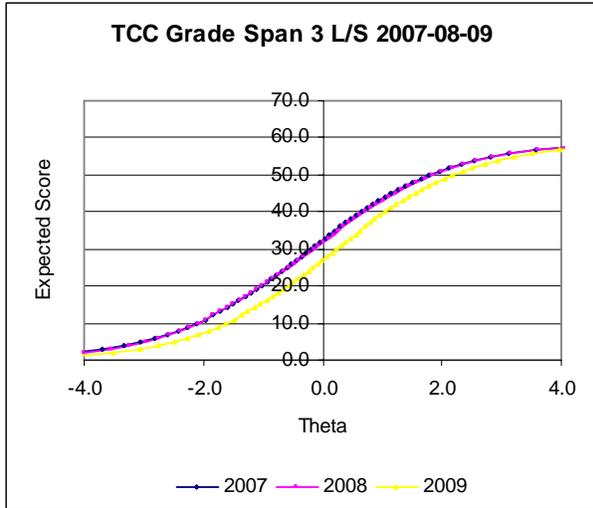


Figure 7.5
2007-08-09 TCC L/S Grade Span 3

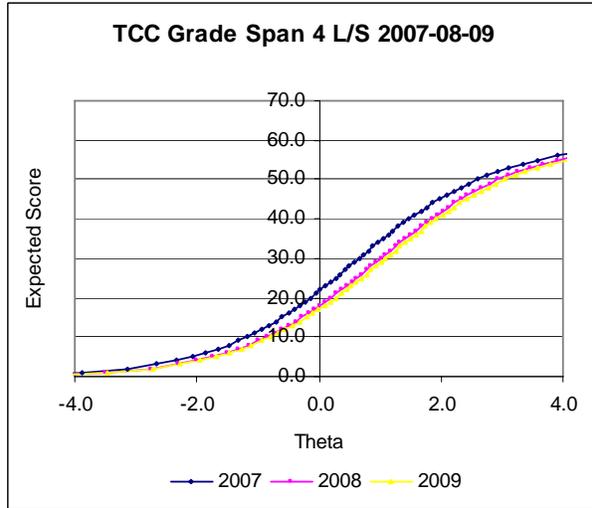


Figure 7.6
2007-08-09 TCC L/S Grade Span 4

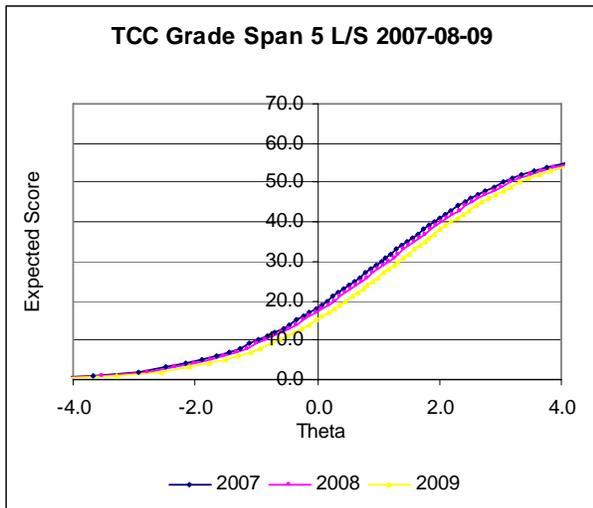


Figure 7.7
2007-08-09 TCC L/S Grade Span 5

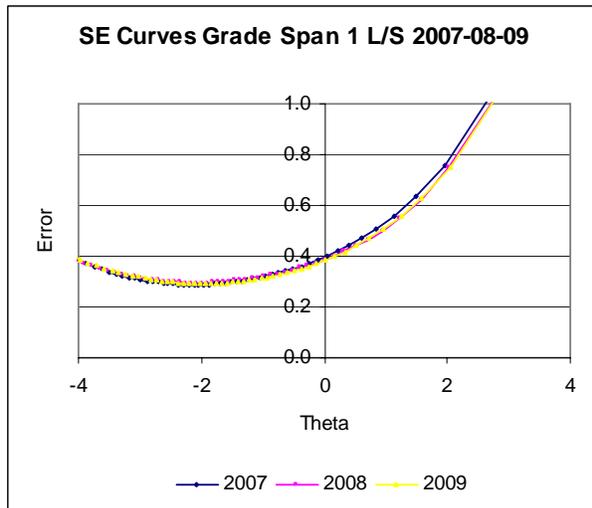


Figure 7.8
2007-08-09 SE Curves L/S Grade Span 1

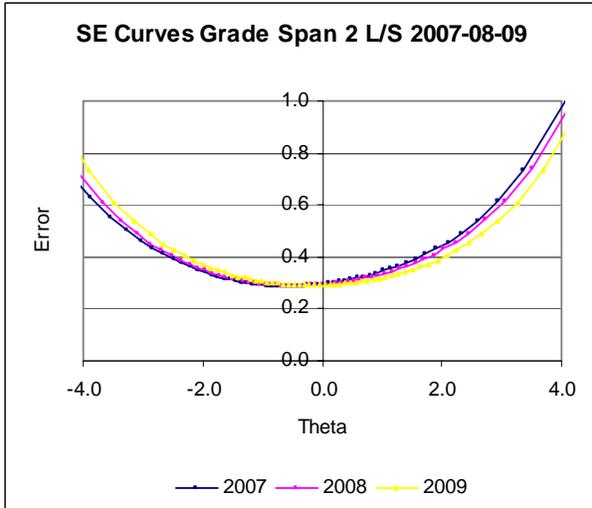


Figure 7.9
2007-08-09 SE Curves L/S Grade Span 2

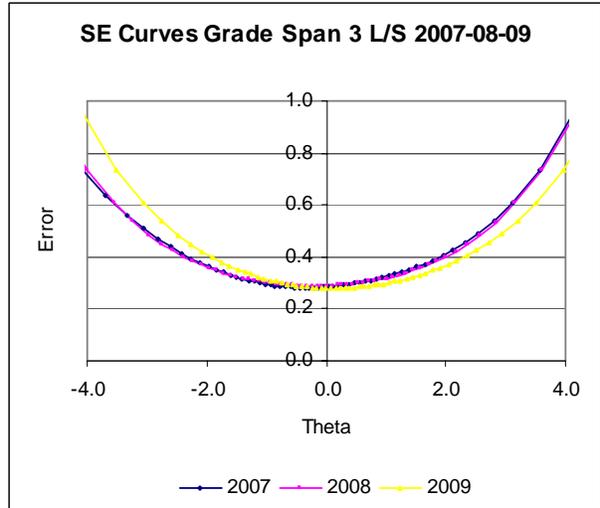


Figure 7.10
2007-08-09 SE Curves L/S Grade Span 3

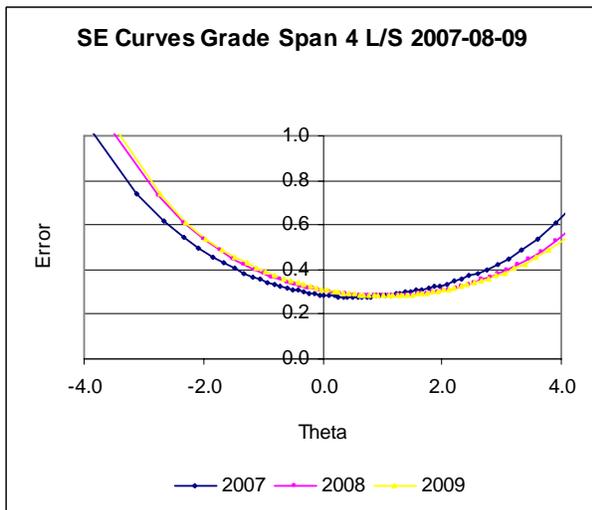


Figure 7.11
2007-08-09 SE Curves L/S Grade Span 4

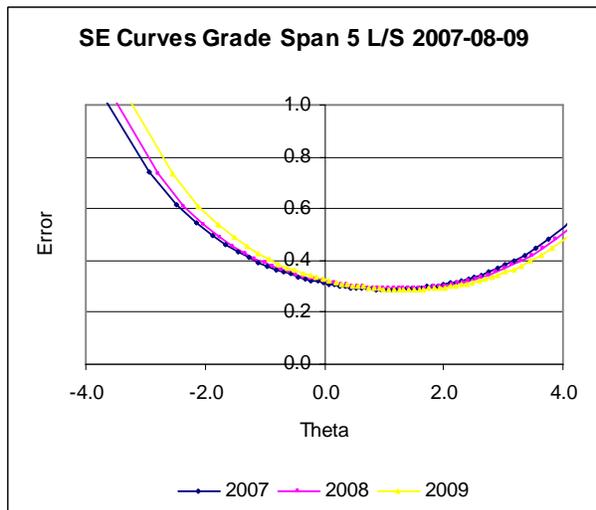


Figure 7.12
2007-08-09 SE Curves L/S Grade Span 5

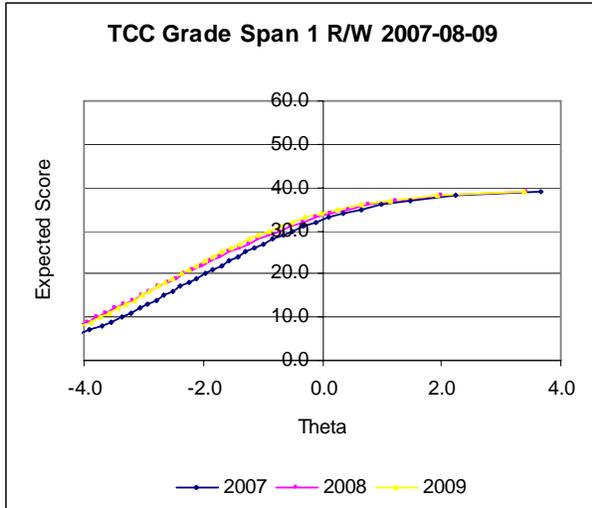


Figure 7.13
2007-08-09 TCC R/W Grade Span 1

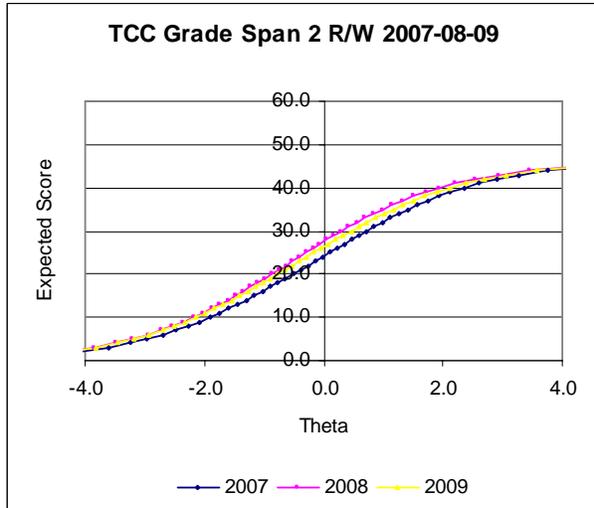


Figure 7.14
2007-08-09 TCC R/W Grade Span 2

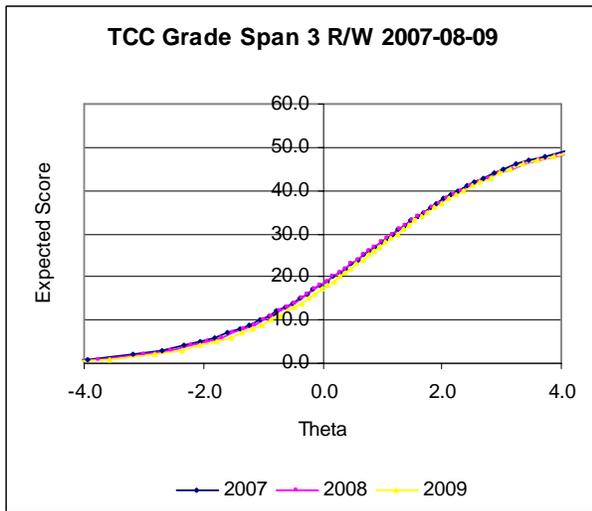


Figure 7.15
2007-08-09 TCC R/W Grade Span 3

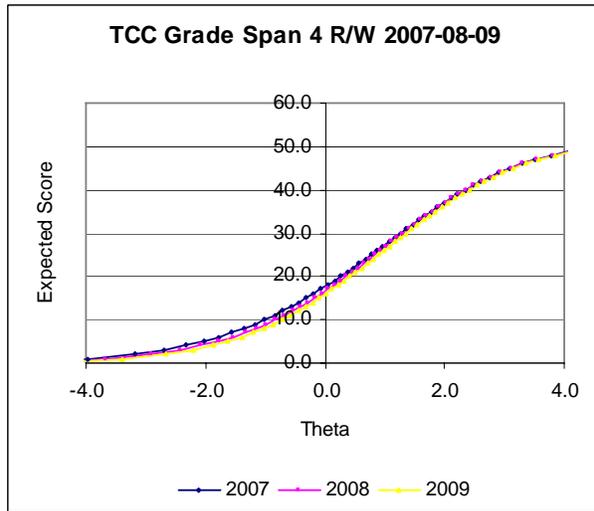


Figure 7.16
2007-08-09 TCC R/W Grade Span 4

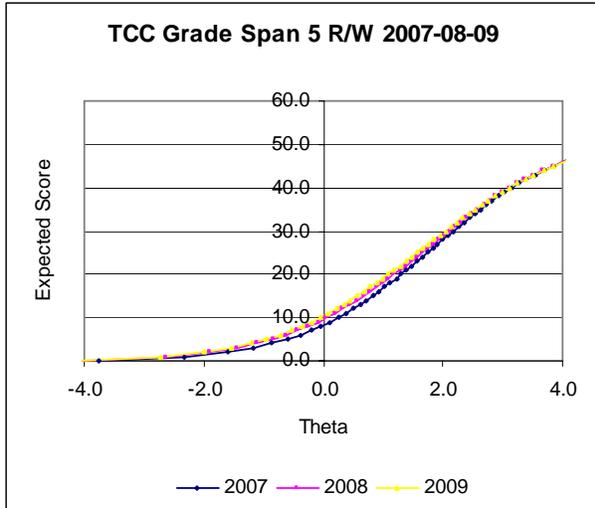


Figure 7.17
2007-08-09 TCC R/W Grade Span 5

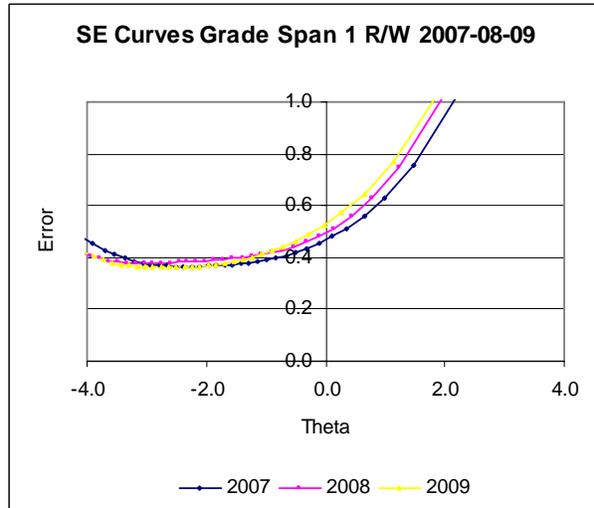


Figure 7.18
2007-08-09 SE Curves R/W Grade Span 1

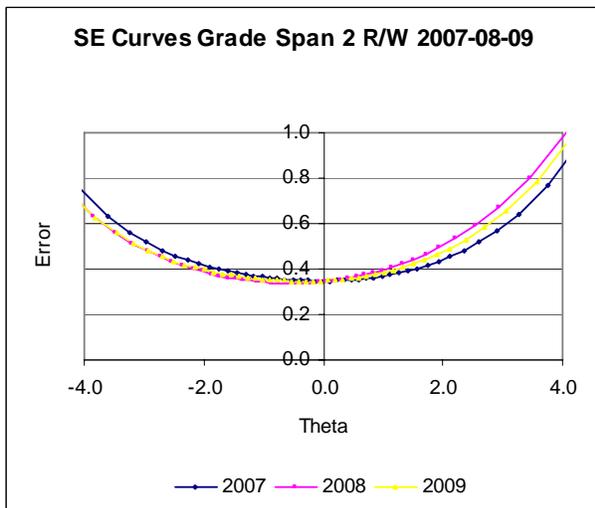


Figure 7.19
2007-08-09 SE Curves R/W Grade Span 2

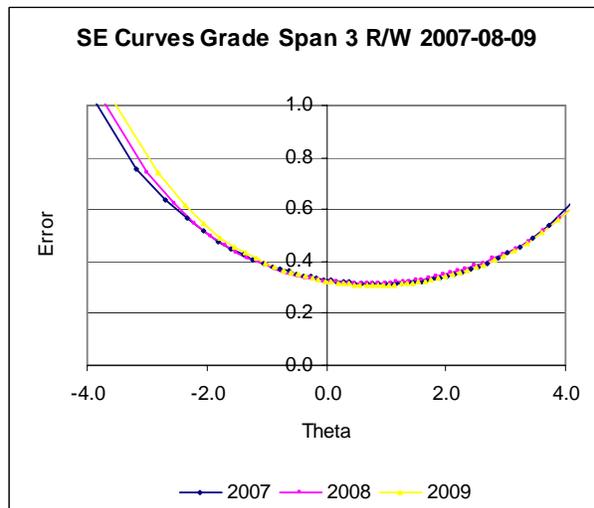


Figure 7.20
2007-08-09 SE Curves R/W Grade Span 3

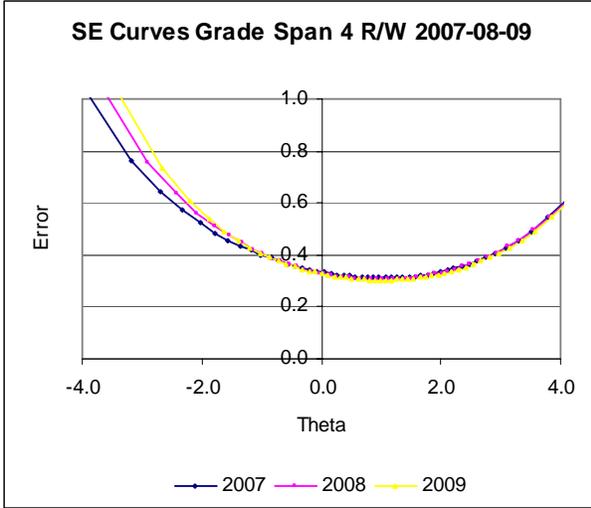


Figure 7.21
2007-08-09 SE Curves R/W Grade Span 4

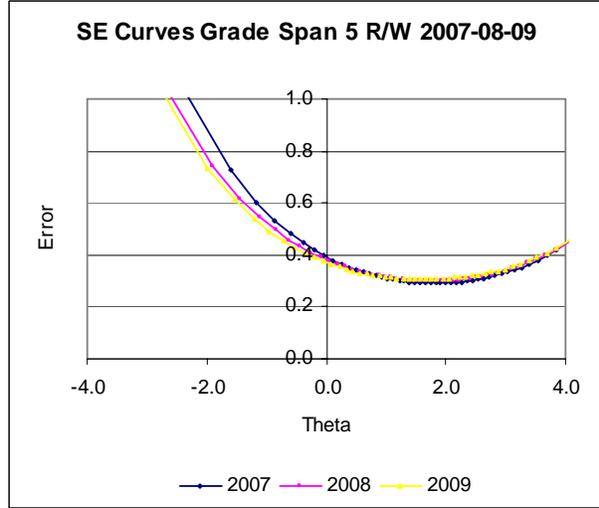


Figure 7.22
2007-08-09 SE Curves R/W Grade Span 5

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:

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

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. Appendices C.1–C.5 provides the raw-to-scale score conversion tables for the Speaking/Listening and Reading/Writing combinations by grade span.

7.6 Rasch Information

Table 7.2 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 D contains the results of the operational items for the NYSESLAT. 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 7.2
Average Rasch Difficulty by Grade Span and Modality

Grade Span	Test	Number of Items	Max Points	Average Rasch Difficulty
K-1	Listening	24	24	-2.37
	Speaking	16	34	-1.65
	Reading	15	15	-2.17
	Writing	15	24	-2.74
	Listening and Speaking	40	58	-2.08
	Reading and Writing	30	39	-2.46
	Total	70	97	-2.24
2-4	Listening	24	24	0.09
	Speaking	16	34	-0.31
	Reading	24	24	-0.51
	Writing	16	22	-0.44
	Listening and Speaking	40	58	-0.07
	Reading and Writing	40	46	-0.52
	Total	80	104	-0.27
5-6	Listening	25	25	0.54
	Speaking	16	34	-0.08
	Reading	27	27	0.89
	Writing	19	25	0.80
	Listening and Speaking	41	59	0.34
	Reading and Writing	46	52	0.82
	Total	87	111	0.59
7-8	Listening	25	25	1.50
	Speaking	16	34	0.71
	Reading	27	27	1.12
	Writing	19	25	0.76
	Listening and Speaking	41	59	1.19
	Reading and Writing	46	51	0.97
	Total	87	111	1.07
9-12	Listening	25	25	1.76
	Speaking	16	34	0.90
	Reading	27	27	1.93
	Writing	19	25	1.33
	Listening and Speaking	41	59	1.42
	Reading and Writing	46	51	1.68
	Total	87	111	1.56

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

PART 8: STANDARD SETTING

8.1 Introduction

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

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

8.2 Standard-Setting Model

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

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

The item-mapping procedure orders items for each test into a booklet according to the difficulty of the items, which is determined by IRT scaling techniques. Easy items are placed in the beginning of the booklet, and subsequent items become increasingly more difficult 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 8.1, the first group recommended standards on Grades K–1; the second group recommended standards on Grades 2–4; the third group recommended standards on Grades 5–6; the fourth group recommended standards on Grades 7–8; and the fifth group recommended standards on Grades 9–12.

The panel members were New York State certified ESL, English language arts, bilingual education, and bilingual special education teachers who were familiar with the content standards. The NYSESLAT panelists were recruited by NYSED to participate in the standard-setting meeting. The panelists possess knowledge of working with students with limited English proficiency and represent all regions of the state of New York. Other educational stakeholders, such as administrators, curriculum specialists, NYSED 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 8.1
Panel Composition for Standard-Setting Committees

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

8.4 Performance Levels and Cut Scores

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

After the orientation, the panelists were separated into specific breakout rooms according to their group assignments. Each group/room was led by a facilitator who was an expert in the standard-setting methodology. Assessment specialists rotated from group to group in order to provide content support. In addition, the panel members were further divided into three smaller table groups within their grade spans, each composed of five to seven members. These small groups worked independently but had the opportunity to collaborate with the other table groups in their grade span during the standard-setting process. The following sequences of tasks were followed.

Review of the Assessment

The first task was to review the assessment blueprint. This was done in order for the panelists to gain an understanding of what the assessment was intended to measure. Discussions about the assessment content, the use of different item types, and 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, Pearson performed several post-standard-setting analyses. The first step was to look up the equivalent scale scores corresponding to the raw score cuts recommended by the committees. Graphs were then plotted using the grades as the independent variable and scale score as the dependent variable. The three cut points were then plotted on the same graph to show that the cuts were monotonically increasing from the lower cuts to the higher cuts. Some technical adjustments were made to the recommended cuts. Any adjustments made were within one standard error of the median. A quadratic equation was then applied to the scaled cut scores across all 13 grades, K–12. Further analyses were performed to smooth the cuts. This was done in order to ensure that there are no reversals.

Next, the proportion of the three cut points in raw score points based on the total test was applied to the Listening/Speaking and Reading/Writing combinations for each of the grades. Impact analysis was conducted on the 2005 operational data. The percentage of students falling into each of the performance levels was calculated for each grade if those cut points were adopted. This information was provided to the NYSED to 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 8.2 and Table 8.3. There are three cut points that correspond to four performance levels. Any score below the Intermediate cut point is the Beginning performance level.

Table 8.2
Final Performance-Level Cut Points for Listening and Speaking

Grade	Listening & Speaking								
	Raw Score 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 8.3
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, P = Proficient.

PART 9: SUMMARY OF OPERATIONAL TEST RESULTS

This part presents both the raw score and scale score summaries for each of the reporting combinations (Listening/Speaking and Reading/Writing). Table 9.1 presents the raw score summary by grade. Table 9.2 presents the raw score summary by grade span. Table 9.3 presents the scale score summary by grade. Table 9.4 presents the scale score summary by grade span. Tables 9.1 through 9.4 include the sample size (n-count), the mean, median, interquartile range (IQR), range, and the standard deviation (SD). Table 9.5 presents the percentage of students in each of the proficiency levels by grade. Table 9.6 presents the percentage of students in each of the proficiency levels by grade span. The overall proficiency level is defined by the lower of the two proficiency level designations (e.g., Listening/Speaking level=3, Reading/Writing level=4, Overall level=3). Table 9.7 presents the exit rate by grade for 2006, 2007, 2008, and 2009 administrations. Table 9.8 presents the exit rate by grade span for 2006, 2007, 2008, and 2009 administrations. These two tables include n-counts and the percentage of English language learners receiving exit status after scoring at the Proficient Level (level 4) on both the LS and RW modality combinations. Additionally, similar types of analyses (See Appendices G, H, and I) were done, according to the following subgroups:

- Gender
- Ethnicity (American Indian or Alaska Native, Black or African-American, Asian, Native Hawaiian/other Pacific Islander, Hispanic or Latino, White, and Multi-Racial [not 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 English language learner
- 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 9.1
Raw Score Summary by Grade

Grade	Test	N- count	Mean	Median	IQR	Range	SD
K	Listening and Speaking	24,257	44.10	46	13	2~58	10.00
	Reading and Writing	24,146	25.11	26	11	3~39	7.29
1	Listening and Speaking	27,339	50.25	53	8	2~58	8.21
	Reading and Writing	27,461	33.54	36	7	4~39	5.92
2	Listening and Speaking	23,653	46.23	49	11	2~58	9.03
	Reading and Writing	22,603	30.11	31	11	3~46	7.83
3	Listening and Speaking	21,286	48.90	52	9	2~58	8.82
	Reading and Writing	20,749	34.42	36	9	3~46	7.24
4	Listening and Speaking	17,932	49.46	53	8	2~58	9.73
	Reading and Writing	17,464	36.08	38	8	6~46	7.32
5	Listening and Speaking	15,425	47.71	51	9	2~59	10.24
	Reading and Writing	14,832	37.64	40	11	3~52	9.18
6	Listening and Speaking	13,042	47.28	51	10	4~59	11.00
	Reading and Writing	12,596	37.57	40	13	5~52	9.66
7	Listening and Speaking	12,252	44.64	49	13	3~59	11.93
	Reading and Writing	11,617	34.99	37	15	6~52	9.96
8	Listening and Speaking	12,177	45.52	50	13	5~59	12.02
	Reading and Writing	11,622	36.42	39	14	6~52	9.96
9	Listening and Speaking	13,426	42.31	47	20	3~59	13.60
	Reading and Writing	12,371	33.68	35	15	3~52	9.58
10	Listening and Speaking	12,495	44.13	48	17	3~59	12.03
	Reading and Writing	11,959	35.38	37	14	7~52	8.88
11	Listening and Speaking	8,295	47.28	50	12	5~59	9.83
	Reading and Writing	8,088	37.73	39	11	8~52	7.96
12	Listening and Speaking	5,762	49.56	52	9	5~59	7.86
	Reading and Writing	5,680	39.14	40	10	7~52	7.25

Table 9.2
Raw Score Summary by Grade Span

Grade	Test	N- count	Mean	Median	IQR	Range	SD
K-1	Listening and Speaking	51,596	47.36	50	11	2~58	9.60
	Reading and Writing	51,607	29.59	31	12	3~39	7.83
2-4	Listening and Speaking	62,871	48.05	51	9	2~58	9.28
	Reading and Writing	60,816	33.30	35	10	3~46	7.90
5-6	Listening and Speaking	28,467	47.51	51	9	2~59	10.59
	Reading and Writing	27,428	37.61	40	12	3~52	9.40
7-8	Listening and Speaking	24,429	45.08	49	14	3~59	11.98
	Reading and Writing	23,239	35.71	38	15	6~52	9.99
9-12	Listening and Speaking	39,978	44.96	49	15	3~59	11.96
	Reading and Writing	38,098	35.89	37	13	3~52	8.93

Note: The data file used in Tables 9.1 and 9.2, above, is the final research file. The total n-counts for grades K-12 are 220,332. The stats 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).

Table 9.3
Scale Score Summary by Grade

Grade	Test	N-count	Mean	Median	IQR	Range	SD
K	Listening and Speaking	24,257	594.18	591	62	398~749	48.80
	Reading and Writing	24,146	548.02	546	58	412~718	46.14
1	Listening and Speaking	27,339	634.70	634	71	398~749	55.92
	Reading and Writing	27,461	619.78	622	92	425~718	61.41
2	Listening and Speaking	23,653	664.20	667	56	463~805	44.02
	Reading and Writing	22,603	620.78	621	52	467~806	40.52
3	Listening and Speaking	21,286	682.17	685	61	463~805	49.55
	Reading and Writing	20,749	644.77	646	53	467~806	42.07
4	Listening and Speaking	17,932	689.46	693	68	463~805	55.25
	Reading and Writing	17,464	656.76	659	53	498~806	45.90
5	Listening and Speaking	15,425	677.72	682	49	477~814	45.54
	Reading and Writing	14,832	679.63	683	50	517~843	42.10
6	Listening and Speaking	13,042	677.32	682	58	504~814	48.86
	Reading and Writing	12,596	680.36	683	59	538~843	45.04
7	Listening and Speaking	12,252	694.76	702	57	520~845	48.98
	Reading and Writing	11,617	670.44	672	58	551~838	41.91
8	Listening and Speaking	12,177	700.74	707	62	541~845	52.17
	Reading and Writing	11,622	677.15	680	57	551~838	43.50
9	Listening and Speaking	13,426	696.90	702	78	526~851	57.29
	Reading and Writing	12,371	688.76	689	56	546~860	39.26
10	Listening and Speaking	12,495	703.62	706	73	526~851	53.09
	Reading and Writing	11,959	695.44	697	55	581~860	37.62
11	Listening and Speaking	8,295	716.16	716	63	547~851	47.62
	Reading and Writing	8,088	705.28	705	47	588~860	36.21
12	Listening and Speaking	5,762	726.61	727	52	547~851	43.84
	Reading and Writing	5,680	711.36	709	46	581~860	35.11

Table 9.4
Scale Score Summary by Grade Span

Grade	Test	N-count	Mean	Median	IQR	Range	SD
K-1	Listening and Speaking	51,596	615.65	612	66	398~749	56.44
	Reading and Writing	51,607	586.20	575	86	412~718	65.46
2-4	Listening and Speaking	62,871	677.49	679	53	463~805	50.45
	Reading and Writing	60,816	639.30	641	54	467~806	45.21
5-6	Listening and Speaking	28,467	677.54	682	49	477~814	47.09
	Reading and Writing	27,428	679.96	683	56	517~843	43.48
7-8	Listening and Speaking	24,429	697.74	702	65	520~845	50.68
	Reading and Writing	23,239	673.80	676	60	551~838	42.84
9-12	Listening and Speaking	39,978	707.28	711	67	526~851	53.30
	Reading and Writing	38,098	697.73	697	51	546~860	38.40

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 8.2 and Table 8.3 of this manual, 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 9.3 and 9.4 are based on the final research file. The total n-counts for grade K-12 are 220,332. The stats 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).

Table 9.5
Percentage of Students in Each Proficiency Level by Grade

Grade	Test	Proficiency Levels			
		Beginning	Intermediate	Advanced	Proficient
K	Listening and Speaking	4.64	27.12	37.84	30.40
	Reading and Writing	49.73	28.90	11.06	10.32
	Overall	49.16	31.37	12.43	7.04
1	Listening and Speaking	2.69	11.91	48.56	36.83
	Reading and Writing	19.40	22.69	27.67	30.24
	Overall	19.01	24.77	37.20	19.02
2	Listening and Speaking	0.78	6.60	37.17	55.45
	Reading and Writing	13.43	37.89	34.85	13.82
	Overall	13.20	38.17	35.83	12.80
3	Listening and Speaking	1.51	6.04	35.74	56.70
	Reading and Writing	8.38	22.29	50.61	18.71
	Overall	8.24	22.72	52.30	16.75
4	Listening and Speaking	3.29	6.75	31.22	58.74
	Reading and Writing	9.20	24.47	44.22	22.11
	Overall	9.12	24.77	46.63	19.49
5	Listening and Speaking	4.00	9.06	39.97	46.96
	Reading and Writing	10.11	19.14	40.14	30.61
	Overall	10.09	20.02	46.25	23.64
6	Listening and Speaking	6.66	11.41	48.81	33.12
	Reading and Writing	12.69	23.34	36.88	27.08
	Overall	12.54	24.34	45.67	17.45
7	Listening and Speaking	4.72	13.55	34.74	47.00
	Reading and Writing	19.56	32.94	29.20	18.30
	Overall	19.11	33.24	31.36	16.29
8	Listening and Speaking	5.49	15.63	26.64	52.25
	Reading and Writing	18.46	34.65	22.04	24.85
	Overall	18.12	35.05	24.12	22.71
9	Listening and Speaking	12.10	23.95	21.93	42.02
	Reading and Writing	15.58	43.51	23.86	17.04
	Overall	16.99	43.19	24.37	15.46
10	Listening and Speaking	8.76	27.37	23.54	40.34
	Reading and Writing	11.78	45.54	26.39	16.29
	Overall	13.47	46.08	26.63	13.83
11	Listening and Speaking	4.76	24.65	32.42	38.17
	Reading and Writing	7.16	49.64	26.48	16.72
	Overall	8.53	51.15	27.52	12.80
12	Listening and Speaking	2.45	27.07	25.18	45.30
	Reading and Writing	4.00	52.02	24.56	19.42
	Overall	4.83	55.24	25.15	14.78

Table 9.6
Percentage of Students in Each Proficiency Level by Grade Span

Grade	Test	Proficiency Levels			
		Beginning	Intermediate	Advanced	Proficient
K-1	Listening and Speaking	3.61	19.06	43.52	33.81
	Reading and Writing	33.59	25.60	19.90	20.92
	Overall	33.06	27.84	25.66	13.44
2-4	Listening and Speaking	1.74	6.45	34.99	56.81
	Reading and Writing	10.49	28.72	42.92	17.87
	Overall	10.33	29.05	44.55	16.07
5-6	Listening and Speaking	5.22	10.14	44.02	40.62
	Reading and Writing	11.30	21.07	38.64	28.99
	Overall	11.21	22.00	45.98	20.81
7-8	Listening and Speaking	5.10	14.59	30.70	49.61
	Reading and Writing	19.01	33.80	25.62	21.58
	Overall	18.62	34.14	27.75	19.50
9-12	Listening and Speaking	8.14	25.62	25.08	41.17
	Reading and Writing	10.87	46.72	25.32	17.09
	Overall	12.27	47.59	25.86	14.28

Note: The statistics in Tables 9.5 and 9.6 are based on the final research file. The total n-counts for grade K-12 are 220,332. The stats 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).

Table 9.7
Exit Rate by Grade for 2006, 2007, 2008, and 2009 Administrations

Grade	2006		2007		2008		2009	
	N-Count	Percent	N-Count	Percent	N-Count	Percent	N-Count	Percent
K	1,265	5.21	1,069	4.41	1,191	4.94	1,663	7.04
1	4,166	16.17	3,265	12.62	4,132	16.05	5,148	19.02
2	4,081	17.59	2,514	11.53	3,256	14.12	2,875	12.80
3	3,692	19.41	3,567	18.51	3,708	19.12	3,455	16.75
4	3,124	20.86	3,486	21.69	3,314	19.88	3,386	19.49
5	2,564	19.17	1,935	14.72	2,442	17.71	3,489	23.64
6	1,765	15.88	1,447	12.63	1,890	15.52	2,172	17.45
7	1,706	13.79	866	7.98	1,192	10.10	1,873	16.29
8	2,150	17.52	1,473	12.77	1,608	14.57	2,607	22.71
9	1,909	13.05	1,589	12.23	1,612	12.29	1,896	15.46
10	1,822	14.15	1,222	10.66	1,142	9.68	1,644	13.83
11	1,021	14.78	725	10.58	897	11.60	1,032	12.80
12	345	12.50	407	12.45	726	14.51	836	14.78
Total	29,610	15.27	23,565	12.48	27,110	13.85	32,076	16.10

Table 9.8
Exit Rate by Grade Span for 2006, 2007, 2008, and 2009 Administrations

Grade	2006		2007		2008		2009	
	N-Count	Percent	N-Count	Percent	N-Count	Percent	N-Count	Percent
K-1	5,431	10.85	4,334	8.65	5,323	10.68	6,811	13.44
2-4	10,897	19.05	9,567	16.74	10,278	17.38	9,716	16.07
5-6	4,329	17.68	3,382	13.75	4,332	16.68	5,661	20.81
7-8	3,856	15.65	2,339	10.45	2,800	12.26	4,480	19.50
9-12	5,097	13.71	3,943	11.40	4,377	11.63	5,408	14.28
Total	29,610	15.27	23,565	12.48	27,110	13.85	32,076	16.10

APPENDIX A: ITEM MAP FOR 2009 NYSESLAT

A. 1: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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: 2009 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. 2: 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

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
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

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
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 #	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	53,547	0	5	92	2	0	1	0.9	0.48
	2	MC	1	53,547	0	5	90	4	0	1	0.9	0.46
	3	MC	1	53,547	0	93	2	3	0	1	0.9	0.48
	4	MC	1	53,547	0	6	92	1	0	1	0.9	0.48
	5	MC	1	53,547	0	2	93	5	0	1	0.9	0.46
	6	MC	1	53,547	0	2	2	96	0	1	0.9	0.51
	7	MC	1	53,547	0	2	94	4	0	0	0.9	0.46
	8	MC	1	53,547	0	15	83	1	0	0	0.8	0.50
	9	MC	1	53,547	0	89	6	4	0	1	0.8	0.51
	10	MC	1	53,547	0	2	3	95	0	0	0.9	0.50
	11	MC	1	53,547	0	21	75	3	0	1	0.7	0.41
	12	MC	1	53,547	0	8	86	6	0	1	0.8	0.43
	13	MC	1	53,547	0	9	3	88	0	0	0.8	0.57
	14	MC	1	53,547	0	82	4	13	0	1	0.8	0.45
	15	MC	1	53,547	0	3	95	2	0	1	0.9	0.52
	16	MC	1	53,547	0	3	4	92	0	1	0.9	0.57
	17	MC	1	53,547	0	6	87	6	0	1	0.8	0.50
	18	MC	1	53,547	0	87	6	6	0	1	0.8	0.46
	19	MC	1	53,547	0	9	2	88	0	1	0.8	0.46
	20	MC	1	53,547	0	82	5	12	0	1	0.8	0.50
	21	MC	1	53,547	0	9	24	66	0	1	0.6	0.54
	22	MC	1	53,547	0	13	31	55	0	1	0.5	0.40
	23	MC	1	53,547	0	13	76	11	0	1	0.7	0.40
	24	MC	1	53,547	0	7	62	30	0	1	0.6	0.42
SP	1	SS	2	53,852	11	30	59	0	0	0	0.7	0.74
	2	SS	2	53,849	7	22	71	0	0	0	0.8	0.75
	3	SS	2	53,840	12	28	59	0	0	0	0.7	0.68
	4	SS	2	53,834	10	28	62	0	0	0	0.7	0.72
	5	SS	2	53,816	9	26	66	0	0	0	0.7	0.77
	6	SE	4	53,576	7	10	26	32	23	1	0.6	0.84
	7	SS	2	53,578	7	32	60	0	0	1	0.7	0.77
	8	SS	2	53,582	7	40	53	0	0	1	0.7	0.75
	9	SS	2	53,591	7	27	65	0	0	1	0.7	0.78
	10	SS	2	53,582	7	37	55	0	0	1	0.7	0.77
	11	SS	2	53,553	8	37	54	0	0	1	0.7	0.77
	12	SS	2	53,565	9	15	76	0	0	1	0.8	0.75
	13	SS	2	53,560	8	19	72	0	0	1	0.8	0.75
	14	SS	2	53,550	7	13	80	0	0	1	0.8	0.75
	15	SS	2	53,550	9	20	71	0	0	1	0.8	0.77
	16	SS	2	53,514	12	19	68	0	0	1	0.7	0.76

Note: The data file used in Tables B.1–B.5 was based on the final research file (N=220,332) after cleaning (with valid corresponding subtest raw score).

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

	Item Number	Item Format	Max Points	N-count	% at 0	% at 1	% at 2	% at 3	% at 4	% Omit	P-value	Point Biserial
RD	1	MC	1	53,356	0	8	87	5	0	1	0.87	0.50
	2	MC	1	53,356	0	83	5	11	0	1	0.83	0.38
	3	MC	1	53,356	0	9	6	84	0	1	0.84	0.58
	4	MC	1	53,356	0	15	14	70	0	1	0.70	0.61
	5	MC	1	53,356	0	22	74	4	0	1	0.74	0.56
	6	MC	1	53,356	0	12	9	78	0	1	0.78	0.67
	7	MC	1	53,356	0	10	83	6	0	1	0.83	0.59
	8	MC	1	53,356	0	66	24	9	0	1	0.66	0.43
	9	MC	1	53,356	0	17	10	72	0	1	0.72	0.68
	10	MC	1	53,356	0	14	69	15	0	1	0.69	0.64
	11	MC	1	53,356	0	10	29	60	0	2	0.60	0.64
	12	MC	1	53,356	0	16	73	9	0	2	0.73	0.52
	13	MC	1	53,356	0	11	21	66	0	2	0.66	0.57
	14	MC	1	53,356	0	24	59	14	0	3	0.59	0.43
	15	MC	1	53,356	0	11	78	8	0	3	0.78	0.48
WR	1	MC	1	53,112	0	4	2	94	0	0	0.94	0.49
	2	MC	1	53,112	0	88	8	3	0	1	0.88	0.47
	3	MC	1	53,112	0	13	81	6	0	1	0.81	0.49
	4	MC	1	53,112	0	6	84	9	0	1	0.84	0.41
	5	MC	1	53,112	0	87	5	7	0	1	0.87	0.44
	6	MC	1	53,112	0	91	3	5	0	1	0.91	0.45
	7	CR	2	53,095	3	32	65	0	0	0	0.81	0.53
	8	CR	2	53,080	3	44	53	0	0	0	0.75	0.53
	9	CR	2	53,080	5	3	92	0	0	0	0.94	0.54
	10	CR	2	53,060	7	1	92	0	0	0	0.92	0.54
	11	CR	2	53,079	15	16	69	0	0	0	0.77	0.77
	12	CR	2	53,066	16	25	59	0	0	0	0.71	0.80
	13	CR	2	53,061	28	30	41	0	0	0	0.56	0.83
	14	CR	2	53,049	28	33	38	0	0	0	0.55	0.83
	15	ER	2	53,041	37	37	26	0	0	0	0.45	0.78

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=220,332) after cleaning (with valid corresponding subtest raw score).

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	65,055	0	95	3	1	0	0	0.95	0.47
	2	MC	1	65,055	0	7	13	80	0	1	0.8	0.39
	3	MC	1	65,055	0	6	3	90	0	1	0.9	0.51
	4	MC	1	65,055	0	89	8	3	0	0	0.89	0.4
	5	MC	1	65,055	0	14	73	13	0	1	0.73	0.47
	6	MC	1	65,055	0	93	3	3	0	1	0.93	0.5
	7	MC	1	65,055	0	1	94	4	0	0	0.94	0.46
	8	MC	1	65,055	0	75	18	7	0	0	0.75	0.44
	9	MC	1	65,055	0	69	25	5	0	1	0.69	0.43
	10	MC	1	65,055	0	3	39	57	0	1	0.57	0.41
	11	MC	1	65,055	0	19	4	70	7	1	0.7	0.4
	12	MC	1	65,055	0	10	80	3	6	1	0.8	0.43
	13	MC	1	65,055	0	2	5	88	4	1	0.88	0.54
	14	MC	1	65,055	0	19	6	60	15	1	0.6	0.5
	15	MC	1	65,055	0	11	7	6	75	1	0.75	0.45
	16	MC	1	65,055	0	4	74	10	10	1	0.74	0.5
	17	MC	1	65,055	0	8	5	80	6	1	0.8	0.52
	18	MC	1	65,055	0	6	12	12	70	1	0.7	0.48
	19	MC	1	65,055	0	7	65	16	11	1	0.65	0.44
	20	MC	1	65,055	0	8	3	87	1	1	0.87	0.43
	21	MC	1	65,055	0	81	6	5	8	1	0.81	0.47
	22	MC	1	65,055	0	2	2	3	92	1	0.92	0.53
	23	MC	1	65,055	0	3	90	4	3	1	0.9	0.49
	24	MC	1	65,055	0	1	2	86	9	1	0.86	0.48
SP	1	SS	2	65,072	5	20	75	0	0	0	0.85	0.73
	2	SS	2	65,067	6	22	72	0	0	0	0.83	0.77
	3	SS	2	65,043	5	18	77	0	0	0	0.86	0.77
	4	SS	2	65,043	7	21	72	0	0	0	0.83	0.76
	5	SS	2	65,035	5	28	67	0	0	0	0.81	0.7
	6	SE	4	65,037	4	5	18	37	36	0	0.74	0.82
	7	SS	2	65,039	4	22	75	0	0	0	0.85	0.77
	8	SS	2	65,037	7	24	69	0	0	0	0.81	0.76
	9	SS	2	65,053	6	25	69	0	0	0	0.82	0.77
	10	SS	2	65,020	6	28	66	0	0	0	0.8	0.77
	11	SS	2	64,969	6	28	65	0	0	0	0.8	0.74
	12	SS	2	65,014	4	18	78	0	0	0	0.87	0.73
	13	SS	2	64,991	6	14	80	0	0	0	0.87	0.76
	14	SS	2	64,974	7	17	75	0	0	0	0.84	0.76
	15	SS	2	64,963	7	14	78	0	0	0	0.85	0.77
	16	SS	2	64,899	5	14	80	0	0	0	0.88	0.75

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=220,332) after cleaning (with valid corresponding subtest raw score).

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 Biseria I
RD	1	MC	1	64973	0	93	2	4	0	0	0.93	0.44
	2	MC	1	64973	0	8	3	89	0	0	0.89	0.47
	3	MC	1	64973	0	6	5	88	0	0	0.88	0.52
	4	MC	1	64973	0	94	3	2	0	0	0.94	0.42
	5	MC	1	64973	0	2	92	6	0	0	0.92	0.46
	6	MC	1	64973	0	92	4	3	0	1	0.92	0.49
	7	MC	1	64973	0	3	90	6	0	1	0.9	0.43
	8	MC	1	64973	0	3	3	90	3	1	0.9	0.56
	9	MC	1	64973	0	6	18	9	67	1	0.67	0.53
	10	MC	1	64973	0	5	79	5	11	1	0.79	0.6
	11	MC	1	64973	0	56	7	15	22	1	0.56	0.48
	12	MC	1	64973	0	85	3	3	8	1	0.85	0.48
	13	MC	1	64973	0	10	8	20	62	1	0.62	0.45
	14	MC	1	64973	0	7	65	8	20	1	0.65	0.57
	15	MC	1	64973	0	29	38	22	11	1	0.38	0.31
	16	MC	1	64973	0	6	30	8	56	1	0.56	0.46
	17	MC	1	64973	0	21	61	9	7	1	0.61	0.44
	18	MC	1	64973	0	53	17	12	16	1	0.53	0.51
	19	MC	1	64973	0	9	10	12	68	1	0.68	0.5
	20	MC	1	64973	0	73	11	7	8	1	0.73	0.6
	21	MC	1	64973	0	21	21	10	47	1	0.47	0.46
	22	MC	1	64973	0	11	60	12	15	1	0.6	0.55
	23	MC	1	64973	0	78	5	10	6	1	0.78	0.39
	24	MC	1	64973	0	18	29	8	44	1	0.44	0.33
WR	1	MC	1	64793	0	5	91	4	0	0	0.91	0.52
	2	MC	1	64793	0	4	2	93	0	0	0.93	0.47
	3	MC	1	64793	0	11	16	73	0	0	0.73	0.48
	4	MC	1	64793	0	84	8	7	0	0	0.84	0.49
	5	MC	1	64793	0	16	59	7	17	0	0.59	0.44
	6	MC	1	64793	0	10	73	8	8	1	0.73	0.57
	7	MC	1	64793	0	7	7	21	64	1	0.64	0.51
	8	MC	1	64793	0	22	68	4	5	1	0.68	0.55
	9	MC	1	64793	0	18	9	7	65	1	0.65	0.53
	10	MC	1	64793	0	4	6	81	9	1	0.81	0.55
	11	MC	1	64793	0	17	9	4	69	1	0.69	0.53
	12	MC	1	64793	0	63	13	6	16	1	0.63	0.5
	13	CR	2	64766	6	16	78	0	0	0	0.86	0.67
	14	CR	2	64761	14	25	60	0	0	0	0.73	0.71
	15	CR	2	64752	15	41	44	0	0	0	0.64	0.72
	16	ER	4	64793	11	25	38	20	5	0	0.46	0.79

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=220,332) after cleaning (with valid corresponding subtest raw score).

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	29,794	0	3	4	92	0	0	0.92	0.58
	2	MC	1	29,794	0	13	74	12	0	0	0.74	0.47
	3	MC	1	29,794	0	73	23	4	0	0	0.73	0.38
	4	MC	1	29,794	0	3	4	4	89	0	0.89	0.61
	5	MC	1	29,794	0	4	5	87	4	0	0.87	0.60
	6	MC	1	29,794	0	3	14	81	2	0	0.81	0.39
	7	MC	1	29,794	0	12	74	10	4	0	0.74	0.46
	8	MC	1	29,794	0	87	4	3	5	0	0.87	0.57
	9	MC	1	29,794	0	6	86	5	3	0	0.86	0.56
	10	MC	1	29,794	0	4	6	8	81	0	0.81	0.55
	11	MC	1	29,794	0	7	5	13	74	0	0.74	0.47
	12	MC	1	29,794	0	76	9	7	9	0	0.76	0.57
	13	MC	1	29,794	0	6	12	69	12	0	0.69	0.47
	14	MC	1	29,794	0	3	3	4	90	0	0.90	0.55
	15	MC	1	29,794	0	64	12	16	8	0	0.64	0.48
	16	MC	1	29,794	0	85	5	5	5	0	0.85	0.53
	17	MC	1	29,794	0	12	56	26	6	0	0.56	0.37
	18	MC	1	29,794	0	17	20	56	7	1	0.56	0.31
	19	MC	1	29,794	0	4	4	8	83	0	0.83	0.45
	20	MC	1	29,794	0	2	2	11	84	0	0.84	0.47
	21	MC	1	29,794	0	10	70	10	10	0	0.70	0.37
	22	MC	1	29,794	0	2	15	80	2	1	0.80	0.34
	23	MC	1	29,794	0	29	47	13	10	1	0.47	0.36
	24	MC	1	29,794	0	78	10	9	3	1	0.78	0.45
	25	MC	1	29,794	0	14	54	15	16	1	0.54	0.40
SP	1	SS	2	29,799	6	20	73	0	0	0	0.83	0.78
	2	SS	2	29,799	6	17	77	0	0	0	0.85	0.81
	3	SS	2	29,801	9	21	70	0	0	0	0.81	0.79
	4	SS	2	29,789	9	18	72	0	0	0	0.82	0.80
	5	SS	2	29,778	9	28	63	0	0	0	0.77	0.77
	6	SE	4	29,781	5	7	19	35	33	0	0.71	0.84
	7	SS	2	29,793	6	21	73	0	0	0	0.83	0.80
	8	SS	2	29,787	8	22	70	0	0	0	0.81	0.82
	9	SS	2	29,791	9	24	67	0	0	0	0.79	0.82
	10	SS	2	29,785	6	23	70	0	0	0	0.82	0.80
	11	SS	2	29,759	6	18	76	0	0	0	0.85	0.81
	12	SS	2	29,783	9	10	81	0	0	0	0.86	0.82
	13	SS	2	29,777	9	22	69	0	0	0	0.80	0.73
	14	SS	2	29,772	6	10	84	0	0	0	0.89	0.80
	15	SS	2	29,761	9	13	78	0	0	0	0.85	0.83
	16	SS	2	29,678	13	17	70	0	0	0	0.79	0.80

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=220,332) after cleaning (with valid corresponding subtest raw score).

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	29,774	0	3	93	4	0	0	0.93	0.42
	2	MC	1	29,774	0	7	6	87	0	0	0.87	0.40
	3	MC	1	29,774	0	7	11	82	0	0	0.82	0.41
	4	MC	1	29,774	0	3	2	92	3	0	0.92	0.44
	5	MC	1	29,774	0	61	24	9	6	0	0.61	0.49
	6	MC	1	29,774	0	3	14	5	77	1	0.77	0.43
	7	MC	1	29,774	0	4	81	8	7	0	0.81	0.57
	8	MC	1	29,774	0	7	6	7	79	0	0.79	0.62
	9	MC	1	29,774	0	27	23	39	10	0	0.39	0.31
	10	MC	1	29,774	0	8	5	79	7	0	0.79	0.64
	11	MC	1	29,774	0	16	66	6	11	1	0.66	0.52
	12	MC	1	29,774	0	73	5	6	16	1	0.73	0.52
	13	MC	1	29,774	0	16	15	23	46	1	0.46	0.40
	14	MC	1	29,774	0	79	8	4	9	1	0.79	0.60
	15	MC	1	29,774	0	10	22	8	59	1	0.59	0.43
	16	MC	1	29,774	0	9	7	68	15	1	0.68	0.49
	17	MC	1	29,774	0	68	13	9	10	1	0.68	0.59
	18	MC	1	29,774	0	8	6	76	9	1	0.76	0.65
	19	MC	1	29,774	0	11	7	77	5	1	0.77	0.58
	20	MC	1	29,774	0	6	13	15	66	1	0.66	0.61
	21	MC	1	29,774	0	53	11	8	28	1	0.53	0.48
	22	MC	1	29,774	0	13	60	15	11	1	0.60	0.55
	23	MC	1	29,774	0	10	10	71	9	1	0.71	0.63
	24	MC	1	29,774	0	13	14	60	13	1	0.60	0.33
	25	MC	1	29,774	0	9	74	7	10	1	0.74	0.67
	26	MC	1	29,774	0	8	66	12	13	1	0.66	0.56
	27	MC	1	29,774	0	8	19	7	65	1	0.65	0.48
WR	1	MC	1	29,681	0	17	78	5	0	0	0.78	0.31
	2	MC	1	29,681	0	7	88	5	0	0	0.88	0.49
	3	MC	1	29,681	0	12	3	82	3	0	0.82	0.46
	4	MC	1	29,681	0	80	5	12	3	0	0.80	0.58
	5	MC	1	29,681	0	6	8	57	29	0	0.57	0.42
	6	MC	1	29,681	0	6	82	6	5	0	0.82	0.60
	7	MC	1	29,681	0	18	9	50	21	1	0.50	0.41
	8	MC	1	29,681	0	88	5	3	4	0	0.88	0.58
	9	MC	1	29,681	0	3	11	6	80	0	0.80	0.54
	10	MC	1	29,681	0	80	9	4	7	0	0.80	0.55
	11	MC	1	29,681	0	25	10	20	45	0	0.45	0.37
	12	MC	1	29,681	0	88	3	3	6	0	0.88	0.54
	13	MC	1	29,681	0	10	66	17	6	0	0.66	0.46
	14	MC	1	29,681	0	6	29	55	10	0	0.55	0.38
	15	MC	1	29,681	0	23	52	6	18	1	0.52	0.39
	16	CR	2	29,669	8	15	77	0	0	0	0.84	0.73
	17	CR	2	29,663	12	21	67	0	0	0	0.77	0.76
	18	CR	2	29,662	18	37	45	0	0	0	0.63	0.71
	19	ER	4	29,681	11	18	36	27	8	0	0.51	0.80

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=220,332) 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	25,684	0	4	94	2	0	0	0.94	0.39
	2	MC	1	25,684	0	90	6	4	0	0	0.90	0.43
	3	MC	1	25,684	0	84	11	4	0	0	0.84	0.48
	4	MC	1	25,684	0	80	9	6	4	0	0.80	0.57
	5	MC	1	25,684	0	12	58	18	11	1	0.58	0.43
	6	MC	1	25,684	0	6	15	55	23	0	0.55	0.42
	7	MC	1	25,684	0	11	19	18	51	1	0.51	0.46
	8	MC	1	25,684	0	8	8	73	11	1	0.73	0.42
	9	MC	1	25,684	0	6	7	83	4	0	0.83	0.54
	10	MC	1	25,684	0	13	63	8	15	1	0.63	0.42
	11	MC	1	25,684	0	13	8	19	59	1	0.59	0.40
	12	MC	1	25,684	0	13	18	58	10	1	0.58	0.48
	13	MC	1	25,684	0	41	6	39	14	0	0.41	0.46
	14	MC	1	25,684	0	15	5	55	24	0	0.55	0.43
	15	MC	1	25,684	0	12	74	7	6	0	0.74	0.56
	16	MC	1	25,684	0	76	13	6	5	0	0.76	0.49
	17	MC	1	25,684	0	82	4	3	10	0	0.82	0.43
	18	MC	1	25,684	0	7	12	15	66	0	0.66	0.51
	19	MC	1	25,684	0	8	8	67	17	0	0.67	0.54
	20	MC	1	25,684	0	73	11	10	6	0	0.73	0.55
	21	MC	1	25,684	0	5	80	10	5	0	0.80	0.51
	22	MC	1	25,684	0	2	3	6	89	0	0.89	0.51
	23	MC	1	25,684	0	10	66	14	9	0	0.66	0.48
	24	MC	1	25,684	0	20	16	43	21	1	0.43	0.42
	25	MC	1	25,684	0	6	7	11	76	0	0.76	0.51
SP	1	SS	2	25,642	8	22	70	0	0	0	0.81	0.81
	2	SS	2	25,639	12	24	65	0	0	0	0.77	0.81
	3	SS	2	25,639	11	22	67	0	0	0	0.78	0.83
	4	SS	2	25,635	7	26	67	0	0	0	0.80	0.81
	5	SS	2	25,632	12	23	64	0	0	0	0.76	0.83
	6	SE	4	25,632	6	6	16	29	43	0	0.74	0.89
	7	SS	2	25,642	7	25	68	0	0	0	0.80	0.82
	8	SS	2	25,637	8	22	70	0	0	0	0.81	0.84
	9	SS	2	25,643	8	23	69	0	0	0	0.80	0.82
	10	SS	2	25,630	8	24	68	0	0	0	0.80	0.83
	11	SS	2	25,614	6	19	75	0	0	0	0.84	0.83
	12	SS	2	25,627	11	22	68	0	0	0	0.79	0.82
	13	SS	2	25,628	13	22	65	0	0	0	0.76	0.83
	14	SS	2	25,623	11	18	71	0	0	0	0.80	0.85
	15	SS	2	25,624	18	18	64	0	0	0	0.73	0.83
	16	SS	2	25,603	12	18	70	0	0	0	0.79	0.83

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=220,332) 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	25,667	0	5	12	83	0	0	0.83	0.41
	2	MC	1	25,667	0	88	9	3	0	0	0.88	0.46
	3	MC	1	25,667	0	5	4	91	0	0	0.91	0.38
	4	MC	1	25,667	0	23	10	60	7	0	0.60	0.54
	5	MC	1	25,667	0	13	73	10	5	0	0.73	0.51
	6	MC	1	25,667	0	4	5	7	84	0	0.84	0.52
	7	MC	1	25,667	0	20	6	60	13	0	0.60	0.51
	8	MC	1	25,667	0	15	68	12	5	0	0.68	0.48
	9	MC	1	25,667	0	12	66	7	14	0	0.66	0.59
	10	MC	1	25,667	0	14	9	66	10	0	0.66	0.54
	11	MC	1	25,667	0	17	64	10	8	0	0.64	0.51
	12	MC	1	25,667	0	4	79	9	8	0	0.79	0.50
	13	MC	1	25,667	0	8	7	9	75	1	0.75	0.58
	14	MC	1	25,667	0	25	15	11	49	1	0.49	0.38
	15	MC	1	25,667	0	10	67	16	7	0	0.67	0.53
	16	MC	1	25,667	0	19	12	8	61	0	0.61	0.48
	17	MC	1	25,667	0	61	9	17	13	0	0.61	0.53
	18	MC	1	25,667	0	56	14	16	13	1	0.56	0.56
	19	MC	1	25,667	0	11	54	27	8	0	0.54	0.42
	20	MC	1	25,667	0	13	53	26	8	1	0.53	0.53
	21	MC	1	25,667	0	20	10	26	44	1	0.44	0.45
	22	MC	1	25,667	0	56	22	10	11	1	0.56	0.58
	23	MC	1	25,667	0	22	42	23	12	1	0.42	0.40
	24	MC	1	25,667	0	68	12	13	7	1	0.68	0.59
	25	MC	1	25,667	0	13	22	11	54	1	0.54	0.50
	26	MC	1	25,667	0	13	10	10	66	1	0.66	0.58
	27	MC	1	25,667	0	12	14	52	21	1	0.52	0.50
WR	1	MC	1	25,504	0	87	9	4	0	0	0.87	0.37
	2	MC	1	25,504	0	6	88	6	0	0	0.88	0.47
	3	MC	1	25,504	0	44	32	18	6	0	0.44	0.35
	4	MC	1	25,504	0	15	4	60	20	0	0.60	0.48
	5	MC	1	25,504	0	4	7	9	80	0	0.80	0.57
	6	MC	1	25,504	0	3	81	3	12	0	0.81	0.48
	7	MC	1	25,504	0	5	7	21	67	1	0.67	0.53
	8	MC	1	25,504	0	8	11	3	78	0	0.78	0.60
	9	MC	1	25,504	0	53	22	10	14	0	0.53	0.37
	10	MC	1	25,504	0	24	68	5	3	0	0.68	0.52
	11	MC	1	25,504	0	7	5	25	62	0	0.62	0.54
	12	MC	1	25,504	0	9	81	5	4	0	0.81	0.60
	13	MC	1	25,504	0	65	5	14	15	1	0.65	0.50
	14	MC	1	25,504	0	22	4	66	7	1	0.66	0.43
	15	MC	1	25,504	0	3	17	20	59	1	0.59	0.48
	16	CR	2	25,489	21	26	53	0	0	0	0.66	0.70
	17	CR	2	25,482	10	15	75	0	0	0	0.83	0.68
	18	CR	2	25,491	18	28	54	0	0	0	0.68	0.72
	19	ER	4	25,504	12	17	33	27	11	0	0.52	0.80

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=220,332) 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	41,248	0	6	88	5	0	1	0.88	0.44
	2	MC	1	41,248	0	81	8	9	0	1	0.81	0.55
	3	MC	1	41,248	0	7	87	5	0	1	0.87	0.51
	4	MC	1	41,248	0	9	8	78	4	0	0.78	0.53
	5	MC	1	41,248	0	67	10	8	15	1	0.67	0.50
	6	MC	1	41,248	0	4	84	5	7	0	0.84	0.38
	7	MC	1	41,248	0	82	8	5	5	0	0.82	0.48
	8	MC	1	41,248	0	74	8	9	9	1	0.74	0.49
	9	MC	1	41,248	0	10	5	4	80	1	0.80	0.57
	10	MC	1	41,248	0	26	56	12	6	1	0.56	0.48
	11	MC	1	41,248	0	7	13	53	26	1	0.53	0.52
	12	MC	1	41,248	0	11	12	72	5	1	0.72	0.56
	13	MC	1	41,248	0	59	19	16	5	0	0.59	0.45
	14	MC	1	41,248	0	60	19	15	5	1	0.60	0.48
	15	MC	1	41,248	0	14	21	11	53	1	0.53	0.35
	16	MC	1	41,248	0	13	13	13	61	1	0.61	0.46
	17	MC	1	41,248	0	13	63	13	11	1	0.63	0.46
	18	MC	1	41,248	0	6	7	23	63	1	0.63	0.55
	19	MC	1	41,248	0	72	15	5	8	0	0.72	0.66
	20	MC	1	41,248	0	66	10	5	19	0	0.66	0.60
	21	MC	1	41,248	0	5	25	12	57	1	0.57	0.48
	22	MC	1	41,248	0	83	8	4	5	0	0.83	0.54
	23	MC	1	41,248	0	4	5	88	3	0	0.88	0.48
	24	MC	1	41,248	0	14	10	13	62	0	0.62	0.48
	25	MC	1	41,248	0	4	67	21	8	1	0.67	0.31
SP	1	SS	2	41,349	6	32	62	0	0	0	0.78	0.76
	2	SS	2	41,346	5	29	65	0	0	0	0.80	0.79
	3	SS	2	41,346	7	18	75	0	0	0	0.84	0.78
	4	SS	2	41,339	9	28	63	0	0	0	0.77	0.77
	5	SS	2	41,336	9	20	71	0	0	0	0.81	0.79
	6	SE	4	41,329	5	6	17	32	40	0	0.74	0.88
	7	SS	2	41,348	5	26	68	0	0	0	0.81	0.81
	8	SS	2	41,345	8	30	62	0	0	0	0.77	0.82
	9	SS	2	41,337	7	28	65	0	0	0	0.79	0.80
	10	SS	2	41,340	5	26	69	0	0	0	0.82	0.80
	11	SS	2	41,319	8	28	63	0	0	0	0.77	0.81
	12	SS	2	41,335	9	24	67	0	0	0	0.79	0.82
	13	SS	2	41,328	11	19	69	0	0	0	0.79	0.80
	14	SS	2	41,322	10	22	68	0	0	0	0.79	0.80
	15	SS	2	41,330	13	25	62	0	0	0	0.75	0.80
	16	SS	2	41,304	7	22	72	0	0	0	0.83	0.78

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=220,332) 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	41,238	0	10	88	2	0	0	0.88	0.39
	2	MC	1	41,238	0	5	4	90	0	0	0.90	0.35
	3	MC	1	41,238	0	86	12	2	0	0	0.86	0.33
	4	MC	1	41,238	0	92	4	2	2	0	0.92	0.36
	5	MC	1	41,238	0	14	65	8	12	0	0.65	0.43
	6	MC	1	41,238	0	20	11	52	16	1	0.52	0.35
	7	MC	1	41,238	0	5	85	3	6	0	0.85	0.46
	8	MC	1	41,238	0	14	53	8	25	0	0.53	0.45
	9	MC	1	41,238	0	69	8	12	11	1	0.69	0.47
	10	MC	1	41,238	0	17	6	31	45	0	0.45	0.40
	11	MC	1	41,238	0	20	9	25	45	1	0.45	0.43
	12	MC	1	41,238	0	64	15	11	9	1	0.64	0.56
	13	MC	1	41,238	0	8	10	3	78	1	0.78	0.38
	14	MC	1	41,238	0	14	8	6	72	1	0.72	0.52
	15	MC	1	41,238	0	9	14	72	4	1	0.72	0.41
	16	MC	1	41,238	0	18	67	13	1	1	0.67	0.52
	17	MC	1	41,238	0	6	20	63	10	1	0.63	0.43
	18	MC	1	41,238	0	5	65	14	15	1	0.65	0.50
	19	MC	1	41,238	0	18	50	17	14	1	0.50	0.52
	20	MC	1	41,238	0	60	7	14	18	1	0.60	0.53
	21	MC	1	41,238	0	18	21	53	7	1	0.53	0.50
	22	MC	1	41,238	0	21	13	53	12	1	0.53	0.43
	23	MC	1	41,238	0	17	36	18	28	1	0.36	0.33
	24	MC	1	41,238	0	45	22	19	13	1	0.45	0.54
	25	MC	1	41,238	0	19	21	10	49	1	0.49	0.44
	26	MC	1	41,238	0	18	10	65	5	1	0.65	0.48
	27	MC	1	41,238	0	26	39	17	17	1	0.39	0.27
WR	1	MC	1	41,173	0	6	89	5	0	0	0.89	0.39
	2	MC	1	41,173	0	3	3	94	0	0	0.94	0.39
	3	MC	1	41,173	0	9	70	13	8	0	0.70	0.54
	4	MC	1	41,173	0	5	85	5	4	0	0.85	0.55
	5	MC	1	41,173	0	18	63	14	4	0	0.63	0.48
	6	MC	1	41,173	0	19	11	18	52	0	0.52	0.48
	7	MC	1	41,173	0	88	6	3	3	0	0.88	0.44
	8	MC	1	41,173	0	11	3	81	5	0	0.81	0.49
	9	MC	1	41,173	0	4	15	61	20	0	0.61	0.42
	10	MC	1	41,173	0	7	81	9	3	0	0.81	0.57
	11	MC	1	41,173	0	84	3	10	3	0	0.84	0.57
	12	MC	1	41,173	0	47	7	5	40	0	0.47	0.30
	13	MC	1	41,173	0	10	15	21	52	1	0.52	0.47
	14	MC	1	41,173	0	17	8	64	10	0	0.64	0.32
	15	MC	1	41,173	0	6	8	69	17	1	0.69	0.59
	16	CR	2	41,167	10	16	74	0	0	0	0.82	0.72
	17	CR	2	41,168	12	19	69	0	0	0	0.79	0.74
	18	CR	2	41,163	14	32	54	0	0	0	0.70	0.72
	19	ER	4	41,173	10	12	28	31	19	0	0.59	0.81

Note: The data file used in Tables B.1–B.5 were based on the final research file (N=220,332) 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	-4.89	2.01	429	70
1	-3.47	1.02	479	36
2	-2.73	0.74	504	26
3	-2.28	0.61	520	21
4	-1.95	0.54	532	19
5	-1.69	0.49	541	17
6	-1.47	0.45	549	16
7	-1.27	0.43	555	15
8	-1.10	0.41	561	14
9	-0.94	0.39	567	14
10	-0.80	0.37	572	13
11	-0.67	0.36	577	13
12	-0.54	0.35	581	12
13	-0.42	0.34	585	12
14	-0.31	0.33	589	12
15	-0.21	0.32	593	11
16	-0.10	0.32	596	11
17	0.00	0.31	600	11
18	0.09	0.31	603	11
19	0.18	0.30	606	11
20	0.27	0.30	610	10
21	0.36	0.29	613	10
22	0.45	0.29	616	10
23	0.53	0.29	619	10
24	0.61	0.29	622	10
25	0.70	0.29	624	10
26	0.78	0.28	627	10
27	0.86	0.28	630	10
28	0.94	0.28	633	10
29	1.02	0.28	636	10
30	1.10	0.28	638	10
31	1.18	0.28	641	10
32	1.26	0.28	644	10
33	1.34	0.28	647	10
34	1.42	0.29	650	10
35	1.50	0.29	652	10
36	1.58	0.29	655	10
37	1.67	0.29	658	10
38	1.75	0.29	661	10
39	1.84	0.30	664	10
40	1.93	0.30	668	11
41	2.02	0.31	671	11
42	2.12	0.31	674	11
43	2.22	0.32	678	11
44	2.32	0.32	681	11
45	2.42	0.33	685	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	--
WR	27	23,627	1.34	0.02	1.07	1.10	--
	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**

**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

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

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
WR	27	1.34	0.22	0.24	0.14
	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
SP	25	1.22	0.19	0.23	0.12
	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
WR	27	1.86	0.21	0.23	0.17
	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
WR	27	3.32	0.10	0.24	0.24
	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
	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
LI	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
	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
SP	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 students from Kindergarten through Grade 12 who are English language learners.

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

Table F.1
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 Department of Education (NYSDE) 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 they are based on the full population responses.

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

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

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

Furthermore, as 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 much as a test of significance for the null of no difference in the fit of the model to the data, but should be used as an indication of how far the data deviate from the model. High values indicate large deviance and small values indicate small deviance. Furthermore, large sample sizes will generally produce larger Chi-square values.

The GFI and AGFI indices have a range of 0 to 1, where values closer to 1 indicate a better fit of the model to the data. On the other hand, the RMR is an index produced by the 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 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 F.1 and F.2:

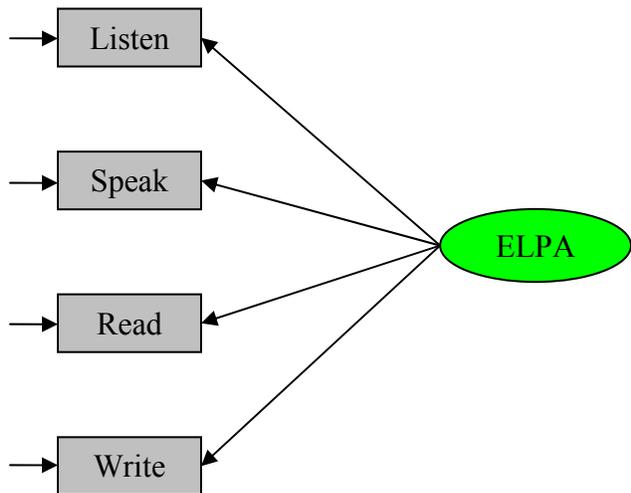


Figure F.1: Path Diagram for the One-Factor Model

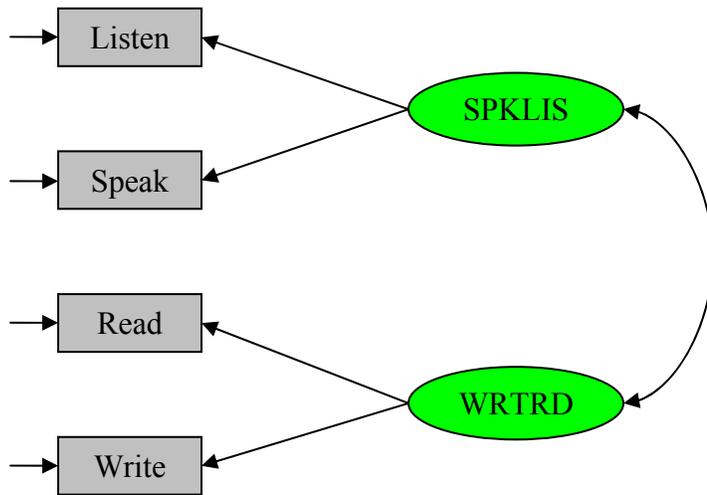


Figure F.2: Path Diagram for the Two-Factor Model

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

F.3: Results and Comments

Grade Span K–1

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

**Table F.2a
Summary Statistics**

Variable	N-count	Mean	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 F.2b
Factor Loading Estimates for the One- and Two-Factor Models

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

Table F.2c
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 0.930 and 0.652, respectively, and the RMR = 0.057 and the RMSEA = 0.274 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 (0.74) is considerably improved with the two-factor model to 0.84. Although the *p*-value of the Chi-square is still in the range of significance at the 0.05 level of rejection ($\chi^2 = 535.992$, $p = 0.00$), the Chi-square itself is reduced tremendously. The RMR and RMSEA values have decreased a great deal to 0.025 and 0.071 respectively, and the GFI as well as the AGFI are approaching the peak of 1.000.

Grade Span 2–4

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

Table F.3a
Summary Statistics

Variable	N-count	Mean	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 F.3b
Factor Loading Estimates for the One- and Two-Factor Models

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

Table F.3c
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 ($\chi^2 = 5855.53$ with 5 *df*, $p = 0.00$). The value of GFI is 0.957 with a poor value for the AGFI, i.e., 0.783. The RMR is acceptable at 0.036, while the RMSEA is large at 0.213. A summary of the fit indices suggest that an alternate model is justifiable.

With the two-factor model, the Chi-square reduces to 2107.90 (with 4 *df*, $p = 0.00$). All the other fit indices also react positively with the GFI value increasing to 0.984. Although the two-factor model shows substantial improvement over the one-factor model, the low values of RMR = 0.034 and RMSEA = 0.128 and the acceptable value of AGFI = 0.920 suggest that the two-factor model is a substantial improvement over the one-factor model and is an acceptable model.

Grade Span 5–6

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

Table F.4a
Summary Statistics

Variable	N-count	Mean	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 F.4b
Factor Loading Estimates for the One- and Two-Factor Models

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

Table F.4c
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 ($\chi^2 = 2500.71$ with 4 *df*, $p = 0.00$ for the two-factor model). The RMR and the RMSEA also show an increase. Similarly, there is not improvement for the GFI and the AGFI with the GFI showing a value of 0.959.

Grade Span 7–8

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

Table F.5a
Summary Statistics

Variable	N-count	Mean	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 F.5b
Factor Loading Estimates for the One- and Two-Factor Models

Modalities	One-Factor Model	Two-Factor Model	
	Λ	Λ_1	Λ_2
Listen	0.84	0.87	–
Speak	0.79	0.80	–
Read	0.82	–	0.78
Write	0.91	–	0.91
			Correlation
			0.98

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

Fit Indices	One-Factor Model	Two-Factor Model
Chi-square	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 not any 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 1661.81 with 5 *df* to 2243.67 with 4 degrees of freedom. Similarly, GFI, AGFI, RMR, and RMSEA all indicate slightly poorer fit over their counterparts in the one-factor model.

Grade Span 9–12

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

Table F.6a
Summary Statistics

Variable	N-count	Mean	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 F.6b
Factor Loading Estimates for the One- and Two-Factor Models

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

Table F.6c
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 did not show any 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 increase ($\chi^2 = 2325.567$ with 5 *df* for the one-factor model and $\chi^2 = 2606.121$ 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 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	11,979	595.28	596	49.75	62
K	LS	Male	12,277	593.11	591	47.84	54
K	LS	American Indian or Alaska Native	32	597.09	589	48.95	55
K	LS	Black or African American	929	592.88	591	46.11	51
K	LS	Asian	4,880	592.51	591	49.36	54
K	LS	Native Hawaiian/Other Pacific Islander	15	603.13	606	37.47	63
K	LS	Hispanic or Latino	15,077	595.39	591	46.66	59
K	LS	White	3,153	591.05	591	57.87	72
K	LS	Multiracial (not Hispanic origin)	168	600.64	606	45.51	53
K	LS	NYC	14,517	596.70	596	48.21	59
K	LS	Big 4 Cities	1,071	584.21	587	48.91	61
K	LS	High Need Urban/Suburban	3,560	581.65	583	52.16	61
K	LS	High Need Rural	195	592.73	591	55.39	62
K	LS	Average Need	2,944	597.57	596	46.79	59
K	LS	Low Need	1,312	600.75	596	44.54	51
K	LS	Charter Schools	222	604.84	603.5	40.49	59
K	LS	0 Years English language learner	13,751	594.19	591	50.20	62
K	LS	1 Years English language learner	10,329	594.29	591	46.94	51
K	LS	Arabic	490	596.93	596	51.06	67
K	LS	Bengali	654	599.04	596	48.72	55
K	LS	Chinese	1,117	590.99	587	47.29	58
K	LS	English	1,531	602.44	601	46.86	63
K	LS	Spanish	13,846	594.88	591	46.67	62
K	LS	Other Language	6,619	590.66	591	53.18	69
K	LS	Students with Disabilities	3,575	573.39	575	47.51	60

G.1b: Kindergarten Reading and Writing

Grade	Test	Group	N- count	Mean	Median	SD	IQR
K	RW	Female	11,909	551.39	546	46.28	53
K	RW	Male	12,236	544.75	541	45.78	57
K	RW	American Indian or Alaska Native	31	548.58	531	47.15	65
K	RW	Black or African American	916	540.05	536	44.74	55
K	RW	Asian	4,963	565.38	563	50.44	59
K	RW	Native Hawaiian/Other Pacific Islander	15	565.13	551	50.96	100
K	RW	Hispanic or Latino	15,188	542.91	541	42.39	57
K	RW	White	2,863	547.43	541	50.30	63
K	RW	Multiracial (not Hispanic origin)	167	550.07	541	46.20	65
K	RW	NYC	14,587	553.29	551	47.86	60
K	RW	Big 4 Cities	1,063	528.78	526	39.73	48
K	RW	High Need Urban/Suburban	3,426	537.28	536	40.19	45
K	RW	High Need Rural	196	532.19	531	42.06	54
K	RW	Average Need	2,930	542.08	536	43.22	51
K	RW	Low Need	1,305	548.31	546	42.99	47
K	RW	Charter Schools	218	555.25	551	41.61	60
K	RW	0 Years English language learner	13,697	553.67	551	48.23	60
K	RW	1 Years English language learner	10,271	540.62	536	42.06	51
K	RW	Arabic	488	544.88	541	45.36	58
K	RW	Bengali	652	568.44	569	49.24	63
K	RW	Chinese	1,129	572.45	569	49.79	63
K	RW	English	1,525	543.69	541	44.46	57
K	RW	Spanish	13,949	543.53	541	42.71	57
K	RW	Other Language	6,403	552.69	546	50.26	65
K	RW	Students with Disabilities	3,222	528.88	526	40.60	48

G.2a: Grade 1 Listening and Speaking

Grade	Test	Group	N- count	Mean	Media n	SD	IQR
1	LS	Female	13,082	637.6	634	55.80	71
1	LS	Male	14,253	631.9	634	55.91	76
1	LS	American Indian or Alaska Native	45	635.7	626	52.53	71
1	LS	Black or African American	1,290	631.9	634	53.61	55
1	LS	Asian	5,022	640.0	634	57.65	71
1	LS	Native Hawaiian/Other Pacific Islander	19	634.5	634	60.96	108
1	LS	Hispanic or Latino	17,261	635.4	634	52.30	71
1	LS	White	3,636	625.1	626	68.36	94
1	LS	Multiracial (not Hispanic origin)	64	627.2	626	58.55	55
1	LS	NYC	17,056	636.3	634	54.75	71
1	LS	Big 4 Cities	1,317	623.5	626	57.91	65
1	LS	High Need Urban/Suburban	4,057	618.8	618	59.06	73
1	LS	High Need Rural	175	646.6	644	57.86	60
1	LS	Average Need	2,926	646.1	644	53.38	60
1	LS	Low Need	1,366	647.4	644	53.89	60
1	LS	Charter Schools	195	637.9	634	39.12	44
1	LS	0 Years English language learner	14,791	634.1	634	56.16	71
1	LS	1 Years English language learner	4,001	620.0	618	60.13	73
1	LS	2 Years English language learner	8,487	642.7	634	51.75	66
1	LS	Arabic	600	640.4	634	59.68	71
1	LS	Bengali	710	636.4	634	55.75	66
1	LS	Chinese	1,103	643.2	634	56.98	66
1	LS	English	1,396	645.9	644	50.95	60
1	LS	Spanish	16,065	634.5	634	52.21	71
1	LS	Other Language	7,465	630.9	634	63.18	81
1	LS	Students with Disabilities	5,418	605.8	606	52.76	59

G.2b: Grade 1 Reading and Writing

Grade	Test	Group	N- count	Mean	Median	SD	IQR
1	RW	Female	13,16	626.77	622	60.11	85
1	RW	Male	14,29	613.34	610	61.90	98
1	RW	American Indian or Alaska Native	45	641.82	640	66.58	119
1	RW	Black or African American	1,287	608.94	610	60.15	71
1	RW	Asian	5,074	645.99	640	60.64	108
1	RW	Native Hawaiian/Other Pacific Islander	19	620.21	610	66.64	98
1	RW	Hispanic or Latino	17,42	613.43	610	59.72	98
1	RW	White	3,550	617.47	610	61.02	92
1	RW	Multiracial (not Hispanic origin)	64	600.98	590	58.67	74
1	RW	NYC	17,22	621.36	622	61.64	92
1	RW	Big 4 Cities	1,318	601.75	599	62.88	83
1	RW	High Need Urban/Suburban	3,985	609.06	610	58.68	71
1	RW	High Need Rural	178	623.57	610	65.46	92
1	RW	Average Need	2,938	627.01	622	60.49	85
1	RW	Low Need	1,375	634.61	640	58.66	77
1	RW	Charter Schools	195	632.88	622	54.71	77
1	RW	0 Years English language learner	14,89	619.86	622	61.99	92
1	RW	1 Years English language learner	4,028	607.75	599	64.16	77
1	RW	2 Years English language learner	8,483	625.53	622	58.04	85
1	RW	Arabic	609	621.22	622	64.03	92
1	RW	Bengali	715	644.97	640	62.23	119
1	RW	Chinese	1,116	651.54	667	56.56	108
1	RW	English	1,397	622.14	622	59.79	85
1	RW	Spanish	16,21	612.95	610	59.66	98
1	RW	Other Language	7,406	626.95	622	62.89	85
1	RW	Students with Disabilities	5,296	590.01	590	54.97	71

G.3a: Grade 2 Listening and Speaking

Grade	Test	Group	N- count	Mean	Median	SD	IQR
2	LS	Female	11,229	665.68	667	43.54	52
2	LS	Male	12,423	662.86	662	44.41	56
2	LS	American Indian or Alaska Native	27	656.07	657	27.78	32
2	LS	Black or African American	1,099	659.69	662	44.01	52
2	LS	Asian	3,955	668.41	667	49.87	52
2	LS	Native Hawaiian/Other Pacific Islander	11	705.45	693	51.10	35
2	LS	Hispanic or Latino	15,496	664.71	667	40.64	44
2	LS	White	3,033	657.69	657	50.99	70
2	LS	Multiracial (not Hispanic origin)	32	663.97	667	59.07	77
2	LS	NYC	15,056	663.31	662	42.41	48
2	LS	Big 4 Cities	1,258	655.34	657	47.45	55
2	LS	High Need Urban/Suburban	3,475	658.58	657	44.90	55
2	LS	High Need Rural	153	664.80	667	38.98	52
2	LS	Average Need	2,329	676.22	673	45.92	53
2	LS	Low Need	1,045	681.16	679	47.58	49
2	LS	Charter Schools	143	670.20	667	36.93	40
2	LS	0 Years English language learner	13,034	661.80	662	43.08	48
2	LS	1 Years English language learner	2,028	649.32	653	51.11	59
2	LS	2 Years English language learner	3,001	662.93	662	44.81	48
2	LS	3 Years English language learner	5,542	676.04	673	40.01	53
2	LS	Arabic	530	657.46	662	45.67	52
2	LS	Bengali	604	665.17	667	47.19	52
2	LS	Chinese	673	672.78	673	49.57	57
2	LS	English	1,058	676.74	673	44.46	53
2	LS	Spanish	14,561	664.01	662	40.28	44
2	LS	Other Language	6,227	662.08	662	50.29	63
2	LS	Students with Disabilities	5,389	645.96	645	40.14	53

G.3 b: Grade 2 Reading and Writing

Grade	Test	Group	N-	Mean	Median	SD	IQR
2	RW	Female	10,848	623.48	625	39.81	47
2	RW	Male	11,753	618.29	621	40.99	56
2	RW	American Indian or Alaska Native	27	618.96	630	35.29	64
2	RW	Black or African American	1,008	610.31	612	40.08	53
2	RW	Asian	3,823	635.30	635	43.19	52
2	RW	Native Hawaiian/Other Pacific Islander	11	657.73	646	39.39	69
2	RW	Hispanic or Latino	14,842	617.24	616	38.38	51
2	RW	White	2,863	623.25	625	42.89	58
2	RW	Multiracial (not Hispanic origin)	29	620.41	625	40.88	62
2	RW	NYC	14,419	619.38	621	39.16	52
2	RW	Big 4 Cities	1,139	612.79	616	41.41	59
2	RW	High Need Urban/Suburban	3,260	620.28	621	39.16	52
2	RW	High Need Rural	146	621.19	618.5	42.33	52
2	RW	Average Need	2,278	626.49	630	44.31	58
2	RW	Low Need	1,032	639.59	641	47.14	54
2	RW	Charter Schools	144	628.13	627.5	34.45	49
2	RW	0 Years English language learner	12,441	618.06	621	38.74	56
2	RW	1 Years English language learner	1,858	615.57	616	45.80	64
2	RW	2 Years English language learner	2,852	624.16	625	42.79	58
2	RW	3 Years English language learner	5,408	627.03	630	40.48	49
2	RW	Arabic	489	614.42	616	40.65	55
2	RW	Bengali	580	631.30	635	37.60	45
2	RW	Chinese	664	636.89	638	43.55	54
2	RW	English	1,026	627.36	630	44.44	53
2	RW	Spanish	13,932	616.71	616	38.06	51
2	RW	Other Language	5,912	626.89	630	43.66	53
2	RW	Students with Disabilities	4,884	602.58	603	37.76	52

G.4a: Grade 3 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
3	LS	Female	9,783	683.90	685	49.09	57
3	LS	Male	11,499	680.72	679	49.89	61
3	LS	American Indian or Alaska Native	25	705.12	714	61.42	63
3	LS	Black or African American	1,064	678.30	679	49.32	49
3	LS	Asian	3,491	686.95	693	54.10	57
3	LS	Native Hawaiian/Other Pacific Islander	9	664.67	693	65.77	70
3	LS	Hispanic or Latino	13,964	683.78	685	45.80	57
3	LS	White	2,704	669.05	667	58.90	76
3	LS	Multiracial (not Hispanic origin)	29	684.90	673	45.01	57
3	LS	NYC	13,619	682.73	685	47.43	57
3	LS	Big 4 Cities	1,147	671.59	673	52.15	57
3	LS	High Need Urban/Suburban	3,184	671.57	673	51.75	61
3	LS	High Need Rural	144	681.99	679	48.24	57
3	LS	Average Need	2,060	695.83	693	51.60	63
3	LS	Low Need	892	695.68	693	53.73	68
3	LS	Charter Schools	97	687.01	679	35.78	52
3	LS	0 Years English language learner	12,299	681.98	685	48.53	61
3	LS	1 Years English language learner	1,706	656.23	657	57.18	73
3	LS	2 Years English language learner	1,486	680.90	679	48.73	65
3	LS	3 Years English language learner	1,654	679.51	679	53.58	65
3	LS	4 Years English language learner	4,112	695.02	693	42.75	47
3	LS	Arabic	482	680.31	679	54.04	65
3	LS	Bengali	476	692.34	693	50.81	63
3	LS	Chinese	505	686.26	685	51.15	57
3	LS	English	980	694.92	693	49.58	63
3	LS	Spanish	13,149	683.16	685	45.62	57
3	LS	Other Language	5,694	676.64	679	56.43	73
3	LS	Students with Disabilities	5,386	665.00	662	45.58	56

G.4b: Grade 3 Reading and Writing

Grade	Test	Group	N- count	Mean	Median	SD	IQR
3	RW	Female	9,613	648.35	652	40.97	49
3	RW	Male	11,133	641.69	646	42.76	50
3	RW	American Indian or Alaska Native	25	643.32	641	44.13	41
3	RW	Black or African American	1,010	635.07	635	45.38	59
3	RW	Asian	3,412	657.26	659	44.74	53
3	RW	Native Hawaiian/Other Pacific Islander	7	661.57	666	16.96	33
3	RW	Hispanic or Latino	13,656	642.72	646	39.61	45
3	RW	White	2,610	642.85	646	46.26	58
3	RW	Multiracial (not Hispanic origin)	29	651.17	646	45.78	62
3	RW	NYC	13,303	644.06	646	40.69	45
3	RW	Big 4 Cities	1,093	632.40	641	42.63	52
3	RW	High Need Urban/Suburban	3,073	642.69	646	41.67	45
3	RW	High Need Rural	139	641.30	646	41.36	50
3	RW	Average Need	2,040	652.22	652	46.32	58
3	RW	Low Need	874	661.64	659	45.88	48
3	RW	Charter Schools	96	656.07	659	29.96	46
3	RW	0 Years English language learner	11,987	643.52	646	40.62	45
3	RW	1 Years English language learner	1,591	631.25	630	50.08	67
3	RW	2 Years English language learner	1,458	645.55	646	47.22	58
3	RW	3 Years English language learner	1,617	647.39	646	44.27	53
3	RW	4 Years English language learner	4,069	652.36	652	38.07	44
3	RW	Arabic	456	641.68	646	44.10	62
3	RW	Bengali	470	658.32	659	41.28	48
3	RW	Chinese	496	658.99	659	42.92	48
3	RW	English	953	651.72	652	43.18	49
3	RW	Spanish	12,865	642.26	646	39.39	45
3	RW	Other Language	5,509	647.26	652	46.73	53
3	RW	Students with Disabilities	5,160	624.31	625	40.36	53

G.5a: Grade 4 Listening and Speaking

Grade	Test	Group	N- count	Mean	Median	SD	IQR
4	LS	Female	7,961	692.04	693	52.39	68
4	LS	Male	9,970	687.40	693	57.35	73
4	LS	American Indian or Alaska Native	21	681.29	685	46.37	65
4	LS	Black or African American	967	686.98	685	53.11	57
4	LS	Asian	2,611	690.51	693	60.33	73
4	LS	Native Hawaiian/Other Pacific Islander	6	623.17	628	34.28	42
4	LS	Hispanic or Latino	11,897	694.10	693	50.03	63
4	LS	White	2,405	666.65	667	67.89	88
4	LS	Multiracial (not Hispanic origin)	24	688.04	685	49.69	51.5
4	LS	NYC	11,554	692.37	693	51.49	68
4	LS	Big 4 Cities	1,061	682.93	685	54.67	61
4	LS	High Need Urban/Suburban	2,818	669.98	673	65.34	81
4	LS	High Need Rural	127	694.47	702	53.96	68
4	LS	Average Need	1,523	701.30	702	53.87	57
4	LS	Low Need	643	705.18	702	53.16	57
4	LS	Charter Schools	95	710.16	702	44.23	51
4	LS	0 Years English language learner	10,454	692.09	693	52.69	63
4	LS	1 Years English language learner	1,643	656.86	657	62.26	76
4	LS	2 Years English language learner	1,139	686.18	685	49.84	61
4	LS	3 Years English language learner	837	671.29	685	73.75	93
4	LS	4 Years English language learner	1,192	683.44	685	55.02	69
4	LS	5 Years English language learner	2,619	709.42	702	43.76	51
4	LS	Arabic	397	685.94	685	53.50	52
4	LS	Bengali	336	693.08	702	58.86	68
4	LS	Chinese	366	684.80	685	66.97	81
4	LS	English	752	708.42	702	50.41	51
4	LS	Spanish	11,195	693.24	693	49.95	63
4	LS	Other Language	4,886	678.26	679	63.82	73
4	LS	Students with Disabilities	5,375	674.98	679	53.84	57

G.5b: Grade 4 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
4	RW	Female	7,840	661.59	666	44.19	48
4	RW	Male	9,623	652.83	659	46.88	58
4	RW	American Indian or Alaska Native	20	661.75	652	48.88	54
4	RW	Black or African American	932	649.81	652	46.60	62
4	RW	Asian	2,552	665.23	666	48.46	53
4	RW	Native Hawaiian/Other Pacific Islander	6	621.00	629	35.24	52
4	RW	Hispanic or Latino	11,682	656.62	659	42.89	48
4	RW	White	2,248	650.86	659	55.37	58
4	RW	Multiracial (not Hispanic origin)	23	653.30	652	34.26	31
4	RW	NYC	11,343	656.97	659	43.35	53
4	RW	Big 4 Cities	1,026	646.03	652	48.26	58
4	RW	High Need Urban/Suburban	2,643	650.45	659	51.86	58
4	RW	High Need Rural	120	661.07	662.5	38.15	53
4	RW	Average Need	1,502	665.40	666	47.31	53
4	RW	Low Need	634	676.46	674	47.96	61
4	RW	Charter Schools	95	670.31	666	31.32	48
4	RW	0 Years English language learner	10,260	656.69	659	43.32	48
4	RW	1 Years English language learner	1,527	638.99	641	53.85	75
4	RW	2 Years English language learner	1,114	655.74	659	47.99	58
4	RW	3 Years English language learner	790	645.78	655.5	63.36	71
4	RW	4 Years English language learner	1,131	664.27	666	46.00	53
4	RW	5 Years English language learner	2,597	667.83	666	38.95	48
4	RW	Arabic	385	649.88	652	49.67	62
4	RW	Bengali	323	668.92	674	45.10	53
4	RW	Chinese	351	664.60	666	47.56	59
4	RW	English	743	664.48	666	42.52	53
4	RW	Spanish	10,987	656.10	659	42.89	53
4	RW	Other Language	4,675	656.23	659	52.13	53
4	RW	Students with Disabilities	5,117	638.86	641	44.33	50

G.6a: Grade 5 Listening and Speaking

Grade	Test	Group	N- count	Mean	Median	SD	IQR
5	LS	Female	6,961	679.35	682	43.24	45
5	LS	Male	8,463	676.38	682	47.31	49
5	LS	American Indian or Alaska Native	21	696.95	688	45.06	56
5	LS	Black or African American	912	675.80	677	43.22	49
5	LS	Asian	2,172	675.92	682	49.54	61
5	LS	Native Hawaiian/Other Pacific Islander	7	674.86	682	38.89	48
5	LS	Hispanic or Latino	10,174	682.48	688	40.86	41
5	LS	White	2,117	657.85	662	56.36	73
5	LS	Multiracial (not Hispanic origin)	22	630.18	628	52.84	91
5	LS	NYC	10,202	681.49	682	42.19	41
5	LS	Big 4 Cities	915	666.11	671	49.27	58
5	LS	High Need Urban/Suburban	2,279	660.07	667	53.93	64
5	LS	High Need Rural	99	684.51	688	44.48	54
5	LS	Average Need	1,214	683.94	688	44.91	50
5	LS	Low Need	488	685.45	688	42.73	47.5
5	LS	Charter Schools	117	687.46	688	38.34	36
5	LS	0 Years English language learner	9,152	681.94	688	42.71	50
5	LS	1 Years English language learner	1,602	650.71	654	51.68	72
5	LS	2 Years English language learner	1,130	676.21	677	41.21	52
5	LS	3 Years English language learner	522	682.55	682	41.06	41
5	LS	4 Years English language learner	764	649.74	658	57.73	75.5
5	LS	5 Years English language learner	665	674.85	682	47.11	52
5	LS	6 Years English language learner or More	1,551	695.25	695	33.00	35
5	LS	Arabic	384	666.53	677	46.47	58
5	LS	Bengali	265	681.80	688	44.79	58
5	LS	Chinese	269	673.83	682	53.17	66
5	LS	English	514	687.91	688	38.16	45
5	LS	Spanish	9,673	682.43	682	41.08	41
5	LS	Other Language	4,320	666.95	671	52.64	66
5	LS	Students with Disabilities	5,038	670.03	677	43.49	44

G.6b: Grade 5 Reading and Writing

Grade	Test	Group	N- count	Mean	Median	SD	IQR
5	RW	Female	6,756	683.17	687	40.60	52
5	RW	Male	8,076	676.66	678	43.10	53
5	RW	American Indian or Alaska Native	20	678.20	687	41.49	53
5	RW	Black or African American	869	671.47	674	42.73	54
5	RW	Asian	2,047	686.69	692	46.14	60
5	RW	Native Hawaiian/Other Pacific Islander	7	702.29	698	53.29	97
5	RW	Hispanic or Latino	9,915	680.37	683	39.29	50
5	RW	White	1,957	672.39	678	48.88	64
5	RW	Multiracial (not Hispanic origin)	17	641.41	640	42.22	56
5	RW	NYC	9,886	680.39	683	40.42	50
5	RW	Big 4 Cities	835	669.48	674	40.59	58
5	RW	High Need Urban/Suburban	2,139	674.76	683	47.58	60
5	RW	High Need Rural	102	677.38	683	45.15	57
5	RW	Average Need	1,182	684.32	687	43.34	52
5	RW	Low Need	470	692.55	692	43.43	52
5	RW	Charter Schools	114	690.84	692	32.69	48
5	RW	0 Years English language learner	8,873	680.82	683	39.87	46
5	RW	1 Years English language learner	1,446	663.51	662	47.84	71
5	RW	2 Years English language learner	1,095	680.20	683	45.38	63
5	RW	3 Years English language learner	506	685.15	687	42.32	52
5	RW	4 Years English language learner	714	666.72	674	54.10	74
5	RW	5 Years English language learner	621	681.45	683	36.97	46
5	RW	6 Years English language learner or More	1,542	690.80	692	36.04	40
5	RW	Arabic	351	667.79	670	45.13	67
5	RW	Bengali	255	684.59	687	44.11	64
5	RW	Chinese	250	686.20	692	48.05	67
5	RW	English	504	687.06	687	44.02	48
5	RW	Spanish	9,422	680.24	683	39.30	50
5	RW	Other Language	4,050	677.58	683	46.80	63
5	RW	Students with Disabilities	4,807	666.11	670	39.47	52

G.7a: Grade 6 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
6	LS	Female	5,806	679.23	682	46.45	58
6	LS	Male	7,235	675.79	682	50.67	52
6	LS	American Indian or Alaska Native	18	680.00	674	51.01	58
6	LS	Black or African American	832	672.22	677	45.41	56
6	LS	Asian	2,034	673.42	677	54.32	72
6	LS	Native Hawaiian/Other Pacific Islander	5	703.80	703	19.12	35
6	LS	Hispanic or Latino	8,396	681.65	688	45.08	54
6	LS	White	1,748	663.67	671	57.18	72
6	LS	Multiracial (not Hispanic origin)	8	626.00	639	71.67	121
6	LS	NYC	8,732	679.40	682	46.92	58
6	LS	Big 4 Cities	789	674.60	682	49.02	56
6	LS	High Need Urban/Suburban	1,822	663.72	671	55.38	69
6	LS	High Need Rural	86	685.57	695	60.11	54
6	LS	Average Need	960	682.47	688	48.60	56
6	LS	Low Need	407	683.83	688	48.35	50
6	LS	Charter Schools	77	694.97	688	28.50	35
6	LS	0 Years English language learner	7,606	680.55	688	47.20	54
6	LS	1 Years English language learner	1,593	649.50	651	52.12	74
6	LS	2 Years English language learner	1,082	675.36	677	43.44	56
6	LS	3 Years English language learner	442	686.64	688	40.21	45
6	LS	4 Years English language learner	604	654.54	667	59.51	82.5
6	LS	5 Years English language learner	331	686.71	688	38.57	45
6	LS	6 Years English language learner or More	1,347	698.47	695	39.45	46
6	LS	Arabic	345	669.25	677	49.11	63
6	LS	Bengali	211	678.70	682	53.66	83
6	LS	Chinese	280	666.35	667	60.90	84
6	LS	English	379	691.87	695	39.80	41
6	LS	Spanish	7,997	681.17	688	45.31	54
6	LS	Other Language	3,830	669.27	677	53.96	63
6	LS	Students with Disabilities	4,181	675.71	682	44.95	49

G.7b: Grade 6 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
6	RW	Female	5,689	684.77	687	43.88	52
6	RW	Male	6,907	676.73	683	45.66	63
6	RW	American Indian or Alaska Native	18	680.06	680.5	44.52	52
6	RW	Black or African American	782	675.61	678	42.46	57
6	RW	Asian	1,925	688.04	692	49.68	64
6	RW	Native Hawaiian/Other Pacific Islander	5	763.00	765	31.37	44
6	RW	Hispanic or Latino	8,190	679.97	683	42.05	56
6	RW	White	1,669	675.58	683	52.82	70
6	RW	Multiracial (not Hispanic origin)	6	650.17	650.5	33.34	59
6	RW	NYC	8,406	680.88	683	43.99	56
6	RW	Big 4 Cities	747	674.84	678	43.27	60
6	RW	High Need Urban/Suburban	1,783	673.05	683	49.77	73
6	RW	High Need Rural	80	692.26	692	46.47	64
6	RW	Average Need	952	685.27	687	44.46	60
6	RW	Low Need	389	699.09	704	42.05	52
6	RW	Charter Schools	76	702.11	698	35.64	43
6	RW	0 Years English language learner	7,329	681.32	683	43.48	56
6	RW	1 Years English language learner	1,482	663.06	662	48.59	74
6	RW	2 Years English language learner	1,058	680.52	683	46.79	63
6	RW	3 Years English language learner	435	691.10	692	42.56	52
6	RW	4 Years English language learner	599	668.09	678	56.94	80
6	RW	5 Years English language learner	329	693.26	698	40.16	48
6	RW	6 Years English language learner or More	1,328	692.83	692	35.50	48
6	RW	Arabic	313	674.73	674	49.83	70
6	RW	Bengali	198	688.94	692	49.47	64
6	RW	Chinese	265	683.93	687	53.88	71
6	RW	English	371	692.72	692	42.81	52
6	RW	Spanish	7,800	679.50	683	42.08	50
6	RW	Other Language	3,649	680.69	687	49.53	63
6	RW	Students with Disabilities	4,042	670.40	674	41.80	54

G.8a: Grade 7 Listening and Speaking

Grade	Test	Group	N- count	Mean	Median	SD	IQR
7	LS	Female	5,559	696.05	702	47.23	57
7	LS	Male	6,693	693.68	702	50.36	61
7	LS	American Indian or Alaska Native	10	696.60	693.5	65.10	54
7	LS	Black or African American	844	694.36	697	44.09	57
7	LS	Asian	2,126	685.65	693	54.72	81
7	LS	Native Hawaiian/Other Pacific Islander	4	678.00	686.5	56.73	89
7	LS	Hispanic or Latino	7,723	698.83	707	45.78	51
7	LS	White	1,528	687.53	697	55.36	73
7	LS	Multiracial (not Hispanic origin)	17	655.06	650	62.93	80
7	LS	NYC	8,242	696.57	702	47.05	54
7	LS	Big 4 Cities	790	683.48	689	49.93	66
7	LS	High Need Urban/Suburban	1,710	686.21	697	55.34	78
7	LS	High Need Rural	85	691.41	702	57.04	81
7	LS	Average Need	887	699.25	707	49.92	65
7	LS	Low Need	417	703.74	707	45.42	52
7	LS	Charter Schools	50	721.04	721.5	37.67	53
7	LS	0 Years English language learner	6,971	698.50	707	47.04	51
7	LS	1 Years English language learner	1,676	665.04	664	51.64	75
7	LS	2 Years English language learner	1,140	693.17	693	45.56	64
7	LS	3 Years English language learner	466	708.77	707	39.71	48
7	LS	4 Years English language learner	562	677.85	685	58.93	91
7	LS	5 Years English language learner	294	711.70	712	36.81	44
7	LS	6 Years English language learner or More	1,115	715.97	718	37.66	45
7	LS	Arabic	306	687.58	693	49.05	63
7	LS	Bengali	214	690.77	695	56.32	78
7	LS	Chinese	328	675.91	672.5	57.53	85
7	LS	English	351	707.57	712	45.24	48
7	LS	Spanish	7,384	698.36	702	45.69	51
7	LS	Other Language	3,669	688.79	697	53.12	73
7	LS	Students with Disabilities	3,528	695.63	702	43.52	51

G.8b: Grade 7 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
7	RW	Female	5,347	675.44	676	41.90	57
7	RW	Male	6,270	666.18	668	41.45	55
7	RW	American Indian or Alaska Native	10	668.60	678	50.81	68
7	RW	Black or African American	802	670.10	670	41.70	58
7	RW	Asian	1,936	672.32	672	45.66	66
7	RW	Native Hawaiian/Other Pacific Islander	4	654.75	652.5	30.31	44.5
7	RW	Hispanic or Latino	7,434	669.88	672	39.23	55
7	RW	White	1,420	671.18	676	49.49	66
7	RW	Multiracial (not Hispanic origin)	11	656.91	665	49.31	91
7	RW	NYC	7,857	670.57	672	40.29	55
7	RW	Big 4 Cities	712	660.15	661	43.02	64
7	RW	High Need Urban/Suburban	1,598	667.37	672	46.74	65
7	RW	High Need Rural	73	669.16	676	41.40	52
7	RW	Average Need	859	675.46	680	43.34	60
7	RW	Low Need	404	686.14	689	42.34	59
7	RW	Charter Schools	49	691.90	693	37.55	36
7	RW	0 Years English language learner	6,661	671.94	672	40.16	55
7	RW	1 Years English language learner	1,503	651.88	647	42.72	62
7	RW	2 Years English language learner	1,087	670.37	668	42.21	58
7	RW	3 Years English language learner	457	682.11	684	41.56	52
7	RW	4 Years English language learner	515	660.91	665	52.33	81
7	RW	5 Years English language learner	290	684.08	684	39.14	52
7	RW	6 Years English language learner or More	1,077	683.08	684	36.45	49
7	RW	Arabic	275	663.85	665	43.19	62
7	RW	Bengali	197	669.02	676	41.06	61
7	RW	Chinese	286	669.62	666.5	47.97	68
7	RW	English	339	678.21	680	41.89	63
7	RW	Spanish	7,101	669.44	672	39.10	49
7	RW	Other Language	3,419	672.42	676	46.56	63
7	RW	Students with Disabilities	3,338	662.50	665	38.74	51

G.9a: Grade 8 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
8	LS	Female	5,560	702.23	707	51.91	62
8	LS	Male	6,617	699.48	707	52.36	65
8	LS	American Indian or Alaska Native	14	707.57	718	44.19	74
8	LS	Black or African American	819	694.84	697	48.97	61
8	LS	Asian	2,129	691.46	697	56.61	83
8	LS	Native Hawaiian/Other Pacific Islander	4	688.75	704.5	52.61	67.5
8	LS	Hispanic or Latino	7,826	704.05	712	49.79	55
8	LS	White	1,367	700.05	707	57.93	69
8	LS	Multiracial (not Hispanic origin)	18	675.61	683	42.96	77
8	LS	NYC	8,466	702.31	707	49.82	59
8	LS	Big 4 Cities	716	688.68	693	55.19	75
8	LS	High Need Urban/Suburban	1,558	690.18	697	58.27	86
8	LS	High Need Rural	86	709.29	709.5	62.07	90
8	LS	Average Need	833	707.83	712	54.45	68
8	LS	Low Need	398	708.03	712	55.07	68
8	LS	Charter Schools	47	733.83	733	44.30	67
8	LS	0 Years English language learner	7,052	705.05	712	49.35	52
8	LS	1 Years English language learner	1,761	667.46	664	54.26	72
8	LS	2 Years English language learner	1,244	694.40	693	50.61	64
8	LS	3 Years English language learner	443	709.04	712	46.28	55
8	LS	4 Years English language learner	526	699.45	707	56.59	65
8	LS	5 Years English language learner	215	729.22	733	39.59	52
8	LS	6 Years English language learner or More	911	730.51	733	39.25	47
8	LS	Arabic	310	694.59	695	51.78	75
8	LS	Bengali	217	707.54	712	48.47	64
8	LS	Chinese	327	681.70	678	56.28	74
8	LS	English	334	715.85	718	49.17	57
8	LS	Spanish	7,510	703.15	712	49.46	55
8	LS	Other Language	3,479	695.99	702	56.90	75
8	LS	Students with Disabilities	2,928	704.84	707	44.80	48

G.9b: Grade 8 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
8	RW	Female	5,383	682.12	684	43.45	59
8	RW	Male	6,239	672.87	676	43.08	63
8	RW	American Indian or Alaska Native	13	682.54	689	39.22	46
8	RW	Black or African American	768	675.29	676	42.23	63
8	RW	Asian	1,976	679.25	680	45.67	66
8	RW	Native Hawaiian/Other Pacific Islander	4	684.00	707	58.67	67
8	RW	Hispanic or Latino	7,552	675.70	676	41.33	57
8	RW	White	1,294	683.81	689	51.67	66
8	RW	Multiracial (not Hispanic origin)	15	647.20	644	37.83	36
8	RW	NYC	8,097	677.61	680	42.06	57
8	RW	Big 4 Cities	648	664.35	665	42.60	59
8	RW	High Need Urban/Suburban	1,500	671.98	676	47.01	66
8	RW	High Need Rural	82	675.29	678	50.07	76
8	RW	Average Need	802	683.15	684	45.46	66
8	RW	Low Need	383	694.88	693	45.43	60
8	RW	Charter Schools	46	705.17	699	41.53	58
8	RW	0 Years English language learner	6,741	679.47	680	41.42	59
8	RW	1 Years English language learner	1,601	658.55	654	45.72	64
8	RW	2 Years English language learner	1,181	672.68	668	45.08	66
8	RW	3 Years English language learner	431	682.08	680	44.40	66
8	RW	4 Years English language learner	539	677.63	684	49.71	59
8	RW	5 Years English language learner	209	692.23	689	37.62	45
8	RW	6 Years English language learner or More	895	692.66	693	37.84	49
8	RW	Arabic	282	671.50	670	44.03	58
8	RW	Bengali	209	686.43	684	41.13	63
8	RW	Chinese	300	676.08	672	46.39	63
8	RW	English	325	688.53	689	44.48	59
8	RW	Spanish	7,235	675.14	676	41.16	57
8	RW	Other Language	3,271	680.46	684	47.63	63
8	RW	Students with Disabilities	2,812	671.24	672	40.52	55

G.10a: Grade 9 Listening and Speaking

Grade	Test	Group	N- count	Mean	Median	SD	IQR
9	LS	Female	6,178	697.71	702	56.64	75
9	LS	Male	7,248	696.20	702	57.83	81
9	LS	American Indian or Alaska Native	23	701.83	694	55.96	63
9	LS	Black or African American	1,039	696.42	698	51.00	75
9	LS	Asian	2,697	683.35	684	54.91	77
9	LS	Native Hawaiian/Other Pacific Islander	2	709.00	709	45.25	64
9	LS	Hispanic or Latino	8,314	699.49	706	58.13	82
9	LS	White	1,304	709.79	711	56.25	64
9	LS	Multiracial (not Hispanic origin)	45	667.16	656	55.14	52
9	LS	NYC	8,750	698.96	702	56.45	82
9	LS	Big 4 Cities	856	695.77	698	59.45	91
9	LS	High Need Urban/Suburban	1,630	691.08	698	57.90	84
9	LS	High Need Rural	85	707.68	716	62.41	73
9	LS	Average Need	1,113	699.24	702	57.68	82
9	LS	Low Need	487	699.68	702	55.01	70
9	LS	Charter Schools	44	715.48	713.5	57.66	48.5
9	LS	0 Years English language learner	7,354	700.04	706	58.76	82
9	LS	1 Years English language learner	2,574	666.58	664	51.02	69
9	LS	2 Years English language learner	1,490	692.56	691	49.86	68
9	LS	3 Years English language learner	470	713.26	711	45.82	54
9	LS	4 Years English language learner	491	721.09	721	39.12	43
9	LS	5 Years English language learner	323	729.84	727	43.76	52
9	LS	6 Years English language learner or More	671	744.25	741	38.02	40
9	LS	Arabic	374	690.57	691	50.22	68
9	LS	Bengali	260	692.61	691	62.46	87
9	LS	Chinese	365	679.85	684	54.13	77
9	LS	English	425	702.68	711	55.30	77
9	LS	Spanish	7,913	699.28	706	58.15	82
9	LS	Other Language	4,089	694.06	694	55.89	78
9	LS	Students with Disabilities	2,060	716.52	721	47.34	50

G.10b: Grade 9 Reading and Writing

Grade	Test	Group	N- count	Mean	Median	SD	IQR
9	RW	Female	5,853	690.18	689	38.96	52
9	RW	Male	6,518	687.50	689	39.48	56
9	RW	American Indian or Alaska Native	22	686.41	689	35.02	33
9	RW	Black or African American	963	687.91	689	39.05	52
9	RW	Asian	2,457	685.89	685	39.22	59
9	RW	Native Hawaiian/Other Pacific Islander	2	707.00	707	2.83	4
9	RW	Hispanic or Latino	7,638	688.17	689	38.59	52
9	RW	White	1,248	699.44	701	41.72	57
9	RW	Multiracial (not Hispanic origin)	39	666.95	655	41.49	44
9	RW	NYC	8,080	689.96	689	38.30	52
9	RW	Big 4 Cities	769	680.17	678	38.89	62
9	RW	High Need Urban/Suburban	1,504	685.08	685	39.92	59
9	RW	High Need Rural	73	695.11	697	40.40	58
9	RW	Average Need	1,042	691.85	693	41.00	56
9	RW	Low Need	467	699.07	701	43.82	57
9	RW	Charter Schools	41	710.10	705	40.20	50
9	RW	0 Years English language learner	6,767	690.66	693	38.83	56
9	RW	1 Years English language learner	2,243	673.49	668	38.92	56
9	RW	2 Years English language learner	1,395	684.18	682	38.60	51
9	RW	3 Years English language learner	461	694.33	693	37.24	50
9	RW	4 Years English language learner	484	701.63	701	36.29	48
9	RW	5 Years English language learner	318	704.82	705	35.37	47
9	RW	6 Years English language learner or More	659	710.45	714	29.71	36
9	RW	Arabic	333	676.92	675	35.85	49
9	RW	Bengali	239	686.10	685	37.71	59
9	RW	Chinese	331	688.65	685	39.98	63
9	RW	English	401	689.31	689	37.04	49
9	RW	Spanish	7,259	688.06	689	38.59	52
9	RW	Other Language	3,808	691.26	693	40.79	56
9	RW	Students with Disabilities	1,938	692.34	691	35.80	46

G.11a: Grade 10 Listening and Speaking

Grade	Test	Group	N- count	Mean	Median	SD	IQR
10	LS	Female	5,947	703.35	706	51.76	73
10	LS	Male	6,547	703.87	706	54.27	73
10	LS	American Indian or Alaska Native	16	705.81	716	48.09	60.5
10	LS	Black or African American	1,074	706.39	706	44.63	60
10	LS	Asian	3,051	687.24	687	50.98	68
10	LS	Native Hawaiian/Other Pacific Islander	13	721.00	711	53.52	56
10	LS	Hispanic or Latino	7,106	707.90	711	54.05	70
10	LS	White	1,191	718.38	721	49.86	63
10	LS	Multiracial (not Hispanic origin)	43	673.72	668	50.86	73
10	LS	NYC	9,217	702.38	702	53.56	77
10	LS	Big 4 Cities	458	708.77	711	53.99	67
10	LS	High Need Urban/Suburban	1,238	704.65	706	50.82	60
10	LS	High Need Rural	63	719.79	721	39.51	59
10	LS	Average Need	830	712.63	716	50.74	70
10	LS	Low Need	441	709.84	711	49.99	61
10	LS	Charter Schools	28	730.36	730.5	48.47	61
10	LS	0 Years English language learner	6,958	706.86	711	55.81	73
10	LS	1 Years English language learner	2,294	683.87	684	47.11	63
10	LS	2 Years English language learner	1,619	694.95	694	46.78	57
10	LS	3 Years English language learner	412	710.36	711	44.72	54
10	LS	4 Years English language learner	410	724.10	721	39.38	39
10	LS	5 Years English language learner	276	729.46	730.5	36.76	44
10	LS	6 Years English language learner or More	490	741.86	741	41.54	45
10	LS	Arabic	293	707.99	711	45.17	57
10	LS	Bengali	294	696.88	694	48.96	66
10	LS	Chinese	461	677.13	677	47.67	67
10	LS	English	297	719.50	721	49.27	59
10	LS	Spanish	6,816	707.54	711	54.27	70
10	LS	Other Language	4,334	699.35	698	51.46	70
10	LS	Students with Disabilities	1,495	721.28	721	46.62	56

G.11b: Grade 10 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
10	RW	Female	5,769	696.43	697	37.12	51
10	RW	Male	6,189	694.50	693	38.06	50
10	RW	American Indian or Alaska Native	17	687.12	689	36.45	47
10	RW	Black or African American	1,055	693.45	693	36.38	50
10	RW	Asian	2,870	691.88	689	37.71	53
10	RW	Native Hawaiian/Other Pacific Islander	13	708.92	705	36.00	60
10	RW	Hispanic or Latino	6,805	695.08	697	37.43	50
10	RW	White	1,160	708.33	709	36.83	50
10	RW	Multiracial (not Hispanic origin)	38	688.03	685	42.26	56
10	RW	NYC	8,805	695.19	693	37.12	50
10	RW	Big 4 Cities	437	687.89	689	36.53	52
10	RW	High Need Urban/Suburban	1,197	695.07	697	38.76	55
10	RW	High Need Rural	59	702.69	697	40.16	41
10	RW	Average Need	804	701.47	701	38.44	45
10	RW	Low Need	429	704.00	705	39.46	51
10	RW	Charter Schools	26	718.85	726	30.31	34
10	RW	0 Years English language learner	6,616	697.59	697	37.21	51
10	RW	1 Years English language learner	2,177	685.60	682	37.59	51
10	RW	2 Years English language learner	1,556	690.17	689	38.12	49
10	RW	3 Years English language learner	402	693.51	693	38.18	50
10	RW	4 Years English language learner	408	708.10	709	35.12	44
10	RW	5 Years English language learner	280	707.42	709	33.50	44
10	RW	6 Years English language learner or More	485	711.83	714	30.72	42
10	RW	Arabic	284	689.88	689	33.73	50.5
10	RW	Bengali	283	688.00	685	35.57	47
10	RW	Chinese	430	694.10	689	39.83	53
10	RW	English	288	703.97	705	37.22	48
10	RW	Spanish	6,521	694.89	697	37.39	50
10	RW	Other Language	4,153	696.72	697	38.00	55
10	RW	Students with Disabilities	1,434	700.66	701	35.66	45

G.12a: Grade 11 Listening and Speaking

Grade	Test	Group	N-count	Mean	Median	SD	IQR
11	LS	Female	4,289	714.46	716	47.88	66
11	LS	Male	4,006	717.98	716	47.29	59
11	LS	American Indian or Alaska Native	15	715.40	702	49.68	90
11	LS	Black or African American	739	719.04	716	39.90	47
11	LS	Asian	1,977	705.26	702	42.94	57
11	LS	Native Hawaiian/Other Pacific Islander	2	781.00	781	98.99	140
11	LS	Hispanic or Latino	4,607	718.49	721	48.32	59
11	LS	White	915	726.66	727	53.64	63
11	LS	Multiracial (not Hispanic origin)	39	689.31	680	65.04	89
11	LS	NYC	5,780	715.74	716	47.48	57
11	LS	Big 4 Cities	300	713.56	711	49.64	57
11	LS	High Need Urban/Suburban	935	711.15	716	48.75	54
11	LS	High Need Rural	52	722.37	716	40.41	59.5
11	LS	Average Need	705	722.13	727	48.03	59
11	LS	Low Need	465	720.86	721	43.45	56
11	LS	Charter Schools	18	747.50	750	45.75	66
11	LS	0 Years English language learner	4,068	721.08	721	49.20	59
11	LS	1 Years English language learner	1,370	697.39	694	47.47	59
11	LS	2 Years English language learner	1,306	708.02	706	43.99	57
11	LS	3 Years English language learner	525	717.72	716	42.57	59
11	LS	4 Years English language learner	410	723.21	721	35.72	39
11	LS	5 Years English language learner	269	730.39	734	37.40	44
11	LS	6 Years English language learner or More	324	744.47	741	36.05	34
11	LS	Arabic	184	722.25	721	42.75	52
11	LS	Bengali	193	708.49	706	43.58	50
11	LS	Chinese	301	699.93	698	36.46	50
11	LS	English	252	722.99	721	44.68	59
11	LS	Spanish	4,344	718.38	721	48.36	59
11	LS	Other Language	3,021	714.14	711	47.83	57
11	LS	Students with Disabilities	946	721.58	727	50.35	56

G.12b: Grade 11 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
11	RW	Female	4,205	705.22	705	35.51	47
11	RW	Male	3,883	705.34	705	36.96	51
11	RW	American Indian or Alaska Native	15	702.40	697	35.92	41
11	RW	Black or African American	719	704.17	705	35.84	45
11	RW	Asian	1,951	704.98	705	36.90	51
11	RW	Native Hawaiian/Other Pacific Islander	2	714.00	714	50.91	72
11	RW	Hispanic or Latino	4,463	703.65	705	35.53	45
11	RW	White	903	714.96	718	36.66	53
11	RW	Multiracial (not Hispanic origin)	34	702.56	699	42.76	46
11	RW	NYC	5,630	705.16	705	35.48	47
11	RW	Big 4 Cities	282	692.27	693	36.82	56
11	RW	High Need Urban/Suburban	917	701.71	705	34.23	45
11	RW	High Need Rural	52	710.04	707	35.00	37
11	RW	Average Need	693	709.92	709	39.44	50
11	RW	Low Need	460	713.30	714	40.07	46
11	RW	Charter Schools	17	726.29	723	47.78	61
11	RW	0 Years English language learner	3,960	707.27	705	35.14	44
11	RW	1 Years English language learner	1,320	700.22	701	38.63	51
11	RW	2 Years English language learner	1,276	701.21	701	37.77	48
11	RW	3 Years English language learner	519	700.85	701	37.61	48
11	RW	4 Years English language learner	406	707.24	709	34.15	44
11	RW	5 Years English language learner	269	712.59	714	32.25	42
11	RW	6 Years English language learner or More	319	716.65	714	29.73	34
11	RW	Arabic	180	700.21	701	36.28	55.5
11	RW	Bengali	190	696.32	693	31.85	40
11	RW	Chinese	295	708.46	709	38.71	47
11	RW	English	250	708.68	709	35.09	44
11	RW	Spanish	4,202	703.52	705	35.66	45
11	RW	Other Language	2,971	708.05	705	36.81	53
11	RW	Students with Disabilities	919	704.46	705	36.01	47

G.13a: Grade 12 Listening and Speaking

Grade	Test	Group	N- count	Mean	Median	SD	IQR
12	LS	Female	2,888	723.65	721	43.36	56
12	LS	Male	2,874	729.59	727	44.12	59
12	LS	American Indian or Alaska Native	6	689.00	687.5	38.05	71
12	LS	Black or African American	559	726.84	721	36.75	48
12	LS	Asian	1,555	717.56	716	41.67	50
12	LS	Native Hawaiian/Other Pacific Islander	3	748.33	750	13.58	27
12	LS	Hispanic or Latino	3,086	730.40	727	44.40	59
12	LS	White	542	731.35	734	49.51	59
12	LS	Multiracial (not Hispanic origin)	11	713.36	711	26.55	39
12	LS	NYC	4,150	725.44	721	43.94	52
12	LS	Big 4 Cities	224	727.41	727	46.90	56
12	LS	High Need Urban/Suburban	517	723.66	727	45.55	52
12	LS	High Need Rural	28	740.25	737.5	34.64	34.5
12	LS	Average Need	515	733.81	734	41.97	55
12	LS	Low Need	292	735.66	734	37.92	45
12	LS	Charter Schools	9	695.00	698	37.54	70
12	LS	0 Years English language learner	3,163	728.92	727	44.38	48
12	LS	1 Years English language learner	612	712.62	711	45.23	57
12	LS	2 Years English language learner	807	716.53	716	41.67	50
12	LS	3 Years English language learner	317	731.29	727	39.82	44
12	LS	4 Years English language learner	355	731.23	727	36.82	44
12	LS	5 Years English language learner	177	737.84	734	42.22	50
12	LS	6 Years English language learner or More	315	739.93	741	41.56	45
12	LS	Arabic	126	735.29	734	37.81	44
12	LS	Bengali	136	718.26	721	37.15	47
12	LS	Chinese	294	707.05	706	39.45	43
12	LS	English	161	733.07	734	39.59	44
12	LS	Spanish	2,913	730.12	727	44.73	59
12	LS	Other Language	2,132	724.05	721	43.24	52
12	LS	Students with Disabilities	621	725.41	727	51.89	56

G.13b: Grade 12 Reading and Writing

Grade	Test	Group	N-count	Mean	Median	SD	IQR
12	RW	Female	2,856	710.63	709	34.13	40
12	RW	Male	2,824	712.10	714	36.07	46
12	RW	American Indian or Alaska Native	6	693.33	689	35.46	29
12	RW	Black or African American	556	708.16	705	33.54	44
12	RW	Asian	1,542	711.96	714	35.30	46
12	RW	Native Hawaiian/Other Pacific Islander	3	742.33	750	21.55	41
12	RW	Hispanic or Latino	3,025	711.19	709	33.99	40
12	RW	White	537	713.76	714	41.64	46
12	RW	Multiracial (not Hispanic origin)	11	718.55	723	32.18	46
12	RW	NYC	4,084	711.11	709	35.09	42
12	RW	Big 4 Cities	220	703.51	701	36.15	45
12	RW	High Need Urban/Suburban	508	707.59	709	33.59	44
12	RW	High Need Rural	28	720.07	723	30.18	47
12	RW	Average Need	512	715.64	718	35.45	42
12	RW	Low Need	292	720.97	723	33.54	34
12	RW	Charter Schools	9	678.33	665	39.03	34
12	RW	0 Years English language learner	3,116	711.04	709	35.20	46
12	RW	1 Years English language learner	592	711.75	709	37.92	50
12	RW	2 Years English language learner	800	709.97	709	34.76	44
12	RW	3 Years English language learner	312	712.54	714	33.12	46
12	RW	4 Years English language learner	354	714.03	714	34.47	42
12	RW	5 Years English language learner	174	712.01	709	29.88	40
12	RW	6 Years English language learner or More	316	712.94	714	34.90	36
12	RW	Arabic	124	709.95	711.5	34.53	40
12	RW	Bengali	135	704.50	705	36.50	45
12	RW	Chinese	296	708.00	709	34.13	47
12	RW	English	159	715.87	718	32.37	53
12	RW	Spanish	2,852	710.97	709	34.25	40
12	RW	Other Language	2,114	712.54	714	36.44	46
12	RW	Students with Disabilities	612	702.10	701	42.73	45

APPENDIX H: PROFICIENCY PERCENTAGES**H.1a: Kindergarten Listening and Speaking**

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
K	LS	Female	11,979	4.87	25.90	37.49	31.75
K	LS	Male	12,277	4.42	28.30	38.19	29.09
K	LS	American Indian or Alaska Native	32	0.00	37.50	37.50	25.00
K	LS	Black or African American	929	3.77	27.88	38.75	29.60
K	LS	Asian	4,880	4.75	29.04	37.46	28.75
K	LS	Native Hawaiian/Other Pacific Islander	15	0.00	26.67	40.00	33.33
K	LS	Hispanic or Latino	15,077	3.50	26.90	39.12	30.48
K	LS	White	3,153	10.21	25.21	31.97	32.60
K	LS	Multiracial (not Hispanic origin)	168	3.57	21.43	39.88	35.12
K	LS	NYC	14,517	3.50	26.85	37.74	31.91
K	LS	Big 4 Cities	1,071	7.66	29.23	37.91	25.21
K	LS	High Need Urban/Suburban	3,560	10.56	31.12	35.03	23.29
K	LS	High Need Rural	195	8.21	23.59	37.95	30.26
K	LS	Average Need	2,944	3.13	25.10	39.64	32.13
K	LS	Low Need	1,312	2.21	22.48	42.00	33.31
K	LS	Charter Schools	222	0.00	24.32	38.29	37.39
K	LS	0 Years English language learner	13,751	5.18	27.37	36.16	31.28
K	LS	1 Years English language learner	10,329	3.93	26.70	40.06	29.31
K	LS	Arabic	490	5.92	24.29	35.10	34.69
K	LS	Bengali	654	4.43	22.32	38.53	34.71
K	LS	Chinese	1,117	3.76	32.32	36.62	27.31
K	LS	English	1,531	2.16	22.99	40.37	34.49
K	LS	Spanish	13,846	3.61	27.13	39.04	30.22
K	LS	Other Language	6,619	7.45	27.84	35.10	29.61
K	LS	Students with Disabilities	3,575	10.69	37.59	34.80	16.92

H.1b: Kindergarten Reading and Writing

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
K	RW	Female	11,909	46.54	30.23	11.95	11.28
K	RW	Male	12,236	52.83	27.60	10.19	9.38
K	RW	American Indian or Alaska Native	31	58.06	16.13	9.68	16.13
K	RW	Black or African American	916	55.46	27.62	9.61	7.31
K	RW	Asian	4,963	34.58	29.98	16.40	19.04
K	RW	Native Hawaiian/Other Pacific Islander	15	40.00	33.33	0.00	26.67
K	RW	Hispanic or Latino	15,188	54.09	29.06	9.39	7.46
K	RW	White	2,863	50.93	26.96	11.00	11.11
K	RW	Multiracial (not Hispanic origin)	167	50.30	23.95	14.37	11.38
K	RW	NYC	14,587	45.17	29.32	12.73	12.78
K	RW	Big 4 Cities	1,063	68.11	22.30	5.46	4.14
K	RW	High Need Urban/Suburban	3,426	59.19	27.96	7.38	5.46
K	RW	High Need Rural	196	60.71	26.02	9.18	4.08
K	RW	Average Need	2,930	55.53	28.46	8.81	7.20
K	RW	Low Need	1,305	47.20	33.10	10.73	8.97
K	RW	Charter Schools	218	44.04	29.82	16.97	9.17
K	RW	0 Years English language learner	13,697	44.89	29.15	12.97	13.00
K	RW	1 Years English language learner	10,271	56.02	28.68	8.53	6.77
K	RW	Arabic	488	51.84	27.66	12.30	8.20
K	RW	Bengali	652	31.29	30.06	18.10	20.55
K	RW	Chinese	1,129	29.05	29.85	18.78	22.32
K	RW	English	1,525	54.43	27.34	10.03	8.20
K	RW	Spanish	13,949	53.55	29.13	9.54	7.77
K	RW	Other Language	6,403	45.65	28.55	12.43	13.37
K	RW	Students with Disabilities	3,222	68.87	21.20	5.59	4.35

H.1c: Kindergarten Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
K	Overall	Female	11,628	45.91	32.94	13.37	7.78
K	Overall	Male	11,978	52.31	29.85	11.51	6.33
K	Overall	American Indian or Alaska Native	31	58.06	22.58	6.45	12.90
K	Overall	Black or African American	902	55.21	28.82	11.09	4.88
K	Overall	Asian	4,834	33.95	35.04	19.03	11.98
K	Overall	Native Hawaiian/Other Pacific Islander	15	40.00	33.33	6.67	20.00
K	Overall	Hispanic or Latino	14,845	53.43	30.87	10.41	5.30
K	Overall	White	2,811	50.62	29.03	12.10	8.25
K	Overall	Multiracial (not Hispanic origin)	166	51.20	24.70	15.66	8.43
K	Overall	NYC	14,227	44.49	32.24	14.54	8.74
K	Overall	Big 4 Cities	1,039	67.66	23.68	5.77	2.89
K	Overall	High Need Urban/Suburban	3,345	58.59	29.69	8.01	3.71
K	Overall	High Need Rural	188	60.11	27.13	9.04	3.72
K	Overall	Average Need	2,896	55.39	30.39	9.29	4.94
K	Overall	Low Need	1,284	46.73	35.36	11.84	6.07
K	Overall	Charter Schools	218	44.04	30.73	18.81	6.42
K	Overall	0 Years English language learner	13,337	44.14	32.34	14.73	8.79
K	Overall	1 Years English language learner	10,098	55.62	30.21	9.40	4.76
K	Overall	Arabic	478	51.05	29.71	12.55	6.69
K	Overall	Bengali	644	31.06	33.70	22.20	13.04
K	Overall	Chinese	1,107	28.73	36.86	21.41	13.01
K	Overall	English	1,506	54.12	28.49	11.09	6.31
K	Overall	Spanish	13,631	52.87	31.03	10.62	5.48
K	Overall	Other Language	6,241	45.20	31.71	14.10	8.99
K	Overall	Students with Disabilities	3,169	69.20	23.10	5.71	1.99

H.2a: Grade 1 Listening and Speaking

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
1	LS	Female	13,082	2.10	11.44	47.46	38.99
1	LS	Male	14,253	3.23	12.34	49.57	34.86
1	LS	American Indian or Alaska Native	45	0.00	13.33	51.11	35.56
1	LS	Black or African American	1,290	2.33	11.78	53.33	32.56
1	LS	Asian	5,022	2.65	10.47	45.52	41.36
1	LS	Native Hawaiian/Other Pacific Islander	19	5.26	5.26	47.37	42.11
1	LS	Hispanic or Latino	17,261	1.65	10.98	51.34	36.03
1	LS	White	3,636	7.78	18.43	37.76	36.03
1	LS	Multiracial (not Hispanic origin)	64	4.69	9.38	56.25	29.69
1	LS	NYC	17,056	2.10	11.43	48.84	37.62
1	LS	Big 4 Cities	1,317	5.01	14.43	50.80	29.76
1	LS	High Need Urban/Suburban	4,057	5.79	18.24	47.97	28.00
1	LS	High Need Rural	175	2.29	10.29	40.57	46.86
1	LS	Average Need	2,926	1.50	7.35	46.82	44.33
1	LS	Low Need	1,366	1.32	6.44	47.73	44.51
1	LS	Charter Schools	195	0.00	7.69	56.92	35.38
1	LS	0 Years English language learner	14,791	2.90	12.17	47.92	37.01
1	LS	1 Years English language learner	4,001	6.27	16.97	48.06	28.69
1	LS	2 Years English language learner	8,487	0.59	9.04	49.90	40.47
1	LS	Arabic	600	2.50	11.17	45.50	40.83
1	LS	Bengali	710	3.10	10.14	47.89	38.87
1	LS	Chinese	1,103	1.45	11.60	44.06	42.88
1	LS	English	1,396	1.15	6.52	48.57	43.77
1	LS	Spanish	16,065	1.67	11.25	51.67	35.41
1	LS	Other Language	7,465	5.33	14.63	42.85	37.19
1	LS	Students with Disabilities	5,418	5.96	24.46	51.44	18.14

H.2b: Grade 1 Reading and Writing

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
1	RW	Female	13,161	15.73	21.77	28.57	33.93
1	RW	Male	14,296	22.78	23.54	26.84	26.84
1	RW	American Indian or Alaska Native	45	11.11	20.00	22.22	46.67
1	RW	Black or African American	1,287	23.00	25.95	27.74	23.31
1	RW	Asian	5,074	10.29	13.70	27.59	48.42
1	RW	Native Hawaiian/Other Pacific Islander	19	21.05	21.05	26.32	31.58
1	RW	Hispanic or Latino	17,420	21.62	24.68	27.98	25.72
1	RW	White	3,550	20.14	24.45	26.51	28.90
1	RW	Multiracial (not Hispanic origin)	64	25.00	35.94	18.75	20.31
1	RW	NYC	17,220	18.82	22.16	27.66	31.36
1	RW	Big 4 Cities	1,318	30.42	22.84	24.13	22.61
1	RW	High Need Urban/Suburban	3,985	23.49	26.40	27.30	22.81
1	RW	High Need Rural	178	16.85	25.84	23.60	33.71
1	RW	Average Need	2,938	15.66	21.95	28.32	34.07
1	RW	Low Need	1,375	11.85	19.71	30.04	38.40
1	RW	Charter Schools	195	10.26	22.56	32.31	34.87
1	RW	0 Years English language learner	14,890	19.64	22.16	27.44	30.75
1	RW	1 Years English language learner	4,028	26.69	23.56	24.85	24.90
1	RW	2 Years English language learner	8,483	15.36	23.26	29.51	31.88
1	RW	Arabic	609	19.54	21.67	25.94	32.84
1	RW	Bengali	715	12.03	15.94	23.22	48.81
1	RW	Chinese	1,116	7.17	12.28	29.93	50.63
1	RW	English	1,397	16.54	24.98	27.70	30.78
1	RW	Spanish	16,218	21.88	24.56	28.18	25.38
1	RW	Other Language	7,406	17.04	20.48	26.79	35.69
1	RW	Students with Disabilities	5,296	34.01	29.81	23.06	13.12

H.2c: Grade 1 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
1	Overall	Female	12,978	15.52	24.02	39.14	21.32
1	Overall	Male	14,083	22.24	25.46	35.40	16.91
1	Overall	American Indian or Alaska Native	45	11.11	26.67	35.56	26.67
1	Overall	Black or African American	1,275	22.43	27.29	36.78	13.49
1	Overall	Asian	4,992	9.44	17.27	43.91	29.39
1	Overall	Native Hawaiian/Other Pacific Islander	19	21.05	21.05	36.84	21.05
1	Overall	Hispanic or Latino	17,181	21.05	26.34	36.59	16.02
1	Overall	White	3,487	21.39	26.64	30.97	20.99
1	Overall	Multiracial (not Hispanic origin)	64	25.00	35.94	25.00	14.06
1	Overall	NYC	16,950	18.22	24.27	38.13	19.38
1	Overall	Big 4 Cities	1,297	29.53	24.75	32.38	13.34
1	Overall	High Need Urban/Suburban	3,928	23.88	29.23	33.32	13.57
1	Overall	High Need Rural	174	16.09	27.59	32.18	24.14
1	Overall	Average Need	2,914	15.75	23.13	37.13	23.99
1	Overall	Low Need	1,364	11.73	21.11	41.13	26.03
1	Overall	Charter Schools	195	10.26	25.64	45.64	18.46
1	Overall	0 Years English language learner	14,624	19.00	24.38	37.55	19.08
1	Overall	1 Years English language learner	3,929	26.62	25.81	33.14	14.43
1	Overall	2 Years English language learner	8,453	15.37	25.00	38.54	21.09
1	Overall	Arabic	595	18.15	23.70	36.97	21.18
1	Overall	Bengali	707	11.46	17.82	41.44	29.28
1	Overall	Chinese	1,099	6.46	17.38	45.31	30.85
1	Overall	English	1,391	16.46	25.59	36.59	21.35
1	Overall	Spanish	15,989	21.29	26.30	36.84	15.57
1	Overall	Other Language	7,284	17.20	23.13	36.48	23.19
1	Overall	Students with Disabilities	5,232	34.92	32.99	25.88	6.21

H.3a: Grade 2 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
2	LS	Female	11,229	0.81	5.69	36.90	56.60
2	LS	Male	12,423	0.76	7.43	37.41	54.40
2	LS	American Indian or Alaska Native	27	0.00	7.41	48.15	44.44
2	LS	Black or African American	1,099	1.36	8.46	37.76	52.41
2	LS	Asian	3,955	1.49	7.43	31.76	59.32
2	LS	Native Hawaiian/Other Pacific Islander	11	0.00	0.00	0.00	100.00
2	LS	Hispanic or Latino	15,496	0.55	4.92	38.62	55.92
2	LS	White	3,033	0.86	13.35	36.80	48.99
2	LS	Multiracial (not Hispanic origin)	32	0.00	18.75	21.88	59.38
2	LS	NYC	15,056	0.77	6.20	38.22	54.80
2	LS	Big 4 Cities	1,258	2.78	9.38	38.63	49.21
2	LS	High Need Urban/Suburban	3,475	0.49	9.47	40.58	49.47
2	LS	High Need Rural	153	0.00	5.88	36.60	57.52
2	LS	Average Need	2,329	0.47	4.89	28.21	66.42
2	LS	Low Need	1,045	0.48	3.54	26.32	69.67
2	LS	Charter Schools	143	0.00	2.80	37.06	60.14
2	LS	0 Years English language learner	13,034	0.85	7.10	38.21	53.84
2	LS	1 Years English language learner	2,028	2.86	14.15	40.34	42.65
2	LS	2 Years English language learner	3,001	0.43	7.40	39.15	53.02
2	LS	3 Years English language learner	5,542	0.04	2.15	32.61	65.21
2	LS	Arabic	530	3.02	7.36	37.55	52.08
2	LS	Bengali	604	2.15	7.12	32.28	58.44
2	LS	Chinese	673	1.04	6.69	30.16	62.11
2	LS	English	1,058	0.09	3.69	31.19	65.03
2	LS	Spanish	14,561	0.55	5.01	39.22	55.22
2	LS	Other Language	6,227	1.09	10.70	34.58	53.64
2	LS	Students with Disabilities	5,389	0.39	12.45	52.20	34.96

H.3b: Grade 2 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
2	RW	Female	10,848	11.43	37.36	36.59	14.62
2	RW	Male	11,753	15.27	38.38	33.26	13.09
2	RW	American Indian or Alaska Native	27	14.81	33.33	48.15	3.70
2	RW	Black or African American	1,008	19.64	42.06	29.27	9.03
2	RW	Asian	3,823	8.74	27.81	38.95	24.51
2	RW	Native Hawaiian/Other Pacific Islander	11	0.00	18.18	36.36	45.45
2	RW	Hispanic or Latino	14,842	14.21	40.82	34.04	10.93
2	RW	White	2,863	13.52	34.86	35.42	16.21
2	RW	Multiracial (not Hispanic origin)	29	13.79	34.48	37.93	13.79
2	RW	NYC	14,419	13.26	39.66	34.73	12.35
2	RW	Big 4 Cities	1,139	20.54	37.14	32.13	10.18
2	RW	High Need Urban/Suburban	3,260	13.59	38.10	35.12	13.19
2	RW	High Need Rural	146	11.64	39.73	33.56	15.07
2	RW	Average Need	2,278	12.60	31.65	36.30	19.45
2	RW	Low Need	1,032	8.24	27.13	35.27	29.36
2	RW	Charter Schools	144	5.56	38.89	39.58	15.97
2	RW	0 Years English language learner	12,441	13.89	40.23	34.45	11.43
2	RW	1 Years English language learner	1,858	20.94	35.36	28.74	14.96
2	RW	2 Years English language learner	2,852	12.80	35.52	35.13	16.55
2	RW	3 Years English language learner	5,408	10.13	34.67	37.67	17.53
2	RW	Arabic	489	17.59	40.70	30.27	11.45
2	RW	Bengali	580	8.10	30.00	44.83	17.07
2	RW	Chinese	664	8.43	25.60	40.81	25.15
2	RW	English	1,026	12.67	32.65	34.50	20.18
2	RW	Spanish	13,932	14.31	41.37	33.83	10.49
2	RW	Other Language	5,912	12.25	32.53	36.06	19.16
2	RW	Students with Disabilities	4,884	24.86	44.74	25.12	5.28

H.3c: Grade 2 Overall

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
2	Overall	Female	10,774	11.28	37.53	37.82	13.37
2	Overall	Male	11,687	14.96	38.77	34.00	12.27
2	Overall	American Indian or Alaska Native	27	14.81	33.33	48.15	3.70
2	Overall	Black or African American	1,005	19.60	42.09	30.25	8.06
2	Overall	Asian	3,807	8.59	27.90	41.32	22.20
2	Overall	Native Hawaiian/Other Pacific Islander	11	0.00	18.18	36.36	45.45
2	Overall	Hispanic or Latino	14,777	13.96	40.94	34.78	10.32
2	Overall	White	2,806	13.15	36.28	35.82	14.75
2	Overall	Multiracial (not Hispanic origin)	29	13.79	34.48	37.93	13.79
2	Overall	NYC	14,333	13.05	39.94	35.61	11.40
2	Overall	Big 4 Cities	1,134	20.46	37.21	32.80	9.52
2	Overall	High Need Urban/Suburban	3,234	13.39	38.65	36.05	11.90
2	Overall	High Need Rural	146	11.64	39.73	34.25	14.38
2	Overall	Average Need	2,263	12.15	31.73	37.83	18.29
2	Overall	Low Need	1,028	7.98	27.24	37.06	27.72
2	Overall	Charter Schools	143	5.59	38.46	40.56	15.38
2	Overall	0 Years English language learner	12,354	13.62	40.58	35.20	10.60
2	Overall	1 Years English language learner	1,832	20.14	35.97	30.73	13.16
2	Overall	2 Years English language learner	2,830	12.69	35.72	36.68	14.91
2	Overall	3 Years English language learner	5,403	10.14	34.72	38.50	16.64
2	Overall	Arabic	487	17.66	40.66	30.39	11.29
2	Overall	Bengali	580	8.28	29.83	46.21	15.69
2	Overall	Chinese	657	7.91	25.72	43.07	23.29
2	Overall	English	1,024	12.60	32.81	35.94	18.65
2	Overall	Spanish	13,870	14.06	41.49	34.58	9.88
2	Overall	Other Language	5,844	11.96	33.26	37.41	17.37
2	Overall	Students with Disabilities	4,821	24.81	45.65	25.43	4.11

H.4a: Grade 3 Listening and Speaking

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
3	LS	Female	9,783	1.28	5.50	35.70	57.52
3	LS	Male	11,499	1.71	6.50	35.76	56.03
3	LS	American Indian or Alaska Native	25	4.00	0.00	24.00	72.00
3	LS	Black or African American	1,064	1.32	8.27	36.65	53.76
3	LS	Asian	3,491	2.09	6.90	29.16	61.84
3	LS	Native Hawaiian/Other Pacific Islander	9	11.11	0.00	22.22	66.67
3	LS	Hispanic or Latino	13,964	1.14	4.17	36.82	57.87
3	LS	White	2,704	2.74	13.83	38.28	45.16
3	LS	Multiracial (not Hispanic origin)	29	0.00	3.45	51.72	44.83
3	LS	NYC	13,619	1.20	5.39	36.18	57.22
3	LS	Big 4 Cities	1,147	3.31	9.24	38.45	49.00
3	LS	High Need Urban/Suburban	3,184	2.51	9.61	39.48	48.40
3	LS	High Need Rural	144	2.08	6.25	32.64	59.03
3	LS	Average Need	2,060	0.92	4.13	27.28	67.67
3	LS	Low Need	892	1.57	3.25	30.16	65.02
3	LS	Charter Schools	97	0.00	0.00	42.27	57.73
3	LS	0 Years English language learner	12,299	1.41	6.14	35.43	57.02
3	LS	1 Years English language learner	1,706	5.80	15.06	42.85	36.28
3	LS	2 Years English language learner	1,486	0.61	6.26	38.96	54.17
3	LS	3 Years English language learner	1,654	2.36	6.95	36.28	54.41
3	LS	4 Years English language learner	4,112	0.02	1.53	32.32	66.12
3	LS	Arabic	482	2.49	5.60	38.38	53.53
3	LS	Bengali	476	1.26	6.51	25.21	67.02
3	LS	Chinese	505	2.18	3.96	34.06	59.80
3	LS	English	980	1.12	3.06	30.92	64.90
3	LS	Spanish	13,149	1.15	4.24	37.21	57.40
3	LS	Other Language	5,694	2.30	10.89	33.98	52.83
3	LS	Students with Disabilities	5,386	1.41	9.51	49.74	39.34

H.4b: Grade 3 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
3	RW	Female	9,613	6.69	20.68	52.01	20.62
3	RW	Male	11,133	9.84	23.68	49.40	17.08
3	RW	American Indian or Alaska Native	25	8.00	24.00	52.00	16.00
3	RW	Black or African American	1,010	13.56	29.21	42.67	14.55
3	RW	Asian	3,412	6.27	15.65	48.07	30.01
3	RW	Native Hawaiian/Other Pacific Islander	7	0.00	0.00	85.71	14.29
3	RW	Hispanic or Latino	13,656	8.13	23.24	52.61	16.02
3	RW	White	2,610	10.54	23.30	46.51	19.66
3	RW	Multiracial (not Hispanic origin)	29	3.45	31.03	44.83	20.69
3	RW	NYC	13,303	7.95	23.34	50.97	17.73
3	RW	Big 4 Cities	1,093	15.65	23.88	48.49	11.99
3	RW	High Need Urban/Suburban	3,073	9.34	21.97	51.51	17.18
3	RW	High Need Rural	139	6.47	26.62	51.80	15.11
3	RW	Average Need	2,040	7.35	18.48	48.43	25.74
3	RW	Low Need	874	5.03	14.65	48.28	32.04
3	RW	Charter Schools	96	1.04	17.71	56.25	25.00
3	RW	0 Years English language learner	11,987	8.20	23.10	51.41	17.29
3	RW	1 Years English language learner	1,591	18.35	28.47	36.90	16.28
3	RW	2 Years English language learner	1,458	11.11	21.47	45.75	21.67
3	RW	3 Years English language learner	1,617	7.11	22.39	49.35	21.15
3	RW	4 Years English language learner	4,069	4.57	17.74	55.84	21.85
3	RW	Arabic	456	10.31	25.00	46.05	18.64
3	RW	Bengali	470	4.89	15.53	47.23	32.34
3	RW	Chinese	496	4.23	15.12	51.21	29.44
3	RW	English	953	4.72	22.46	49.32	23.50
3	RW	Spanish	12,865	8.37	23.12	52.97	15.54
3	RW	Other Language	5,509	9.55	21.33	45.94	23.18
3	RW	Students with Disabilities	5,160	15.95	35.16	41.45	7.44

H.4c: Grade 3 Overall

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
3	Overall	Female	9,554	6.56	21.29	53.86	18.29
3	Overall	Male	11,070	9.68	23.95	50.94	15.43
3	Overall	American Indian or Alaska Native	25	8.00	24.00	52.00	16.00
3	Overall	Black or African American	1,009	13.48	29.34	44.00	13.18
3	Overall	Asian	3,399	6.30	15.71	50.54	27.45
3	Overall	Native Hawaiian/Other Pacific Islander	7	0.00	0.00	85.71	14.29
3	Overall	Hispanic or Latino	13,602	7.91	23.35	54.43	14.31
3	Overall	White	2,556	10.56	26.02	46.52	16.90
3	Overall	Multiracial (not Hispanic origin)	29	3.45	31.03	48.28	17.24
3	Overall	NYC	13,238	7.80	23.60	52.69	15.91
3	Overall	Big 4 Cities	1,090	15.50	23.94	50.09	10.46
3	Overall	High Need Urban/Suburban	3,040	9.57	23.59	52.60	14.24
3	Overall	High Need Rural	139	6.47	28.06	50.36	15.11
3	Overall	Average Need	2,021	6.58	18.60	50.22	24.59
3	Overall	Low Need	873	5.04	14.66	51.20	29.10
3	Overall	Charter Schools	96	1.04	17.71	62.50	18.75
3	Overall	0 Years English language learner	11,917	8.05	23.47	53.03	15.45
3	Overall	1 Years English language learner	1,574	17.98	28.72	39.01	14.29
3	Overall	2 Years English language learner	1,445	10.38	21.80	48.37	19.45
3	Overall	3 Years English language learner	1,599	7.57	23.95	50.34	18.14
3	Overall	4 Years English language learner	4,065	4.55	18.03	57.42	20.00
3	Overall	Arabic	456	10.31	25.22	47.15	17.32
3	Overall	Bengali	469	5.12	15.57	48.83	30.49
3	Overall	Chinese	493	3.85	15.21	54.77	26.17
3	Overall	English	952	4.62	22.58	50.53	22.27
3	Overall	Spanish	12,810	8.13	23.24	54.82	13.81
3	Overall	Other Language	5,447	9.62	22.60	47.16	20.62
3	Overall	Students with Disabilities	5,102	16.01	36.71	41.63	5.64

H.5a: Grade 4 Listening and Speaking

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
4	LS	Female	7,961	1.92	6.31	31.97	59.80
4	LS	Male	9,970	4.38	7.10	30.63	57.88
4	LS	American Indian or Alaska Native	21	0.00	9.52	33.33	57.14
4	LS	Black or African American	967	1.96	9.00	32.37	56.67
4	LS	Asian	2,611	3.87	8.58	28.84	58.71
4	LS	Native Hawaiian/Other Pacific Islander	6	16.67	16.67	66.67	0.00
4	LS	Hispanic or Latino	11,897	1.86	4.92	30.93	62.29
4	LS	White	2,405	10.27	12.89	34.59	42.25
4	LS	Multiracial (not Hispanic origin)	24	4.17	4.17	37.50	54.17
4	LS	NYC	11,554	2.06	5.85	31.43	60.66
4	LS	Big 4 Cities	1,061	3.20	9.80	32.33	54.67
4	LS	High Need Urban/Suburban	2,818	9.62	10.18	35.02	45.17
4	LS	High Need Rural	127	1.57	8.66	27.56	62.20
4	LS	Average Need	1,523	1.58	5.98	24.49	67.96
4	LS	Low Need	643	1.24	4.51	25.35	68.90
4	LS	Charter Schools	95	0.00	1.05	26.32	72.63
4	LS	0 Years English language learner	10,454	2.49	6.20	30.09	61.22
4	LS	1 Years English language learner	1,643	9.01	18.14	39.26	33.60
4	LS	2 Years English language learner	1,139	1.05	6.76	41.53	50.66
4	LS	3 Years English language learner	837	17.08	4.66	25.81	52.45
4	LS	4 Years English language learner	1,192	1.93	9.23	37.58	51.26
4	LS	5 Years English language learner	2,619	0.04	1.34	25.12	73.50
4	LS	Arabic	397	2.77	7.81	31.23	58.19
4	LS	Bengali	336	2.08	10.12	27.08	60.71
4	LS	Chinese	366	7.65	9.29	27.32	55.74
4	LS	English	752	0.27	4.26	24.07	71.41
4	LS	Spanish	11,195	1.95	5.00	31.26	61.79
4	LS	Other Language	4,886	6.63	10.62	32.81	49.94
4	LS	Students with Disabilities	5,375	4.78	8.09	41.51	45.62

H.5b: Grade 4 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
4	RW	Female	7,840	7.13	22.77	45.31	24.80
4	RW	Male	9,623	10.88	25.87	43.32	19.93
4	RW	American Indian or Alaska Native	20	5.00	35.00	35.00	25.00
4	RW	Black or African American	932	12.66	28.43	40.67	18.24
4	RW	Asian	2,552	8.19	19.51	40.56	31.74
4	RW	Native Hawaiian/Other Pacific Islander	6	16.67	33.33	50.00	0.00
4	RW	Hispanic or Latino	11,682	8.06	25.51	46.05	20.38
4	RW	White	2,248	14.86	22.86	40.39	21.89
4	RW	Multiracial (not Hispanic origin)	23	8.70	34.78	39.13	17.39
4	RW	NYC	11,343	8.00	25.73	45.17	21.10
4	RW	Big 4 Cities	1,026	14.52	29.24	39.38	16.86
4	RW	High Need Urban/Suburban	2,643	14.38	21.83	43.13	20.66
4	RW	High Need Rural	120	7.50	20.83	45.00	26.67
4	RW	Average Need	1,502	7.72	19.77	43.41	29.09
4	RW	Low Need	634	3.94	17.98	40.22	37.85
4	RW	Charter Schools	95	0.00	18.95	54.74	26.32
4	RW	0 Years English language learner	10,260	8.19	25.09	46.01	20.71
4	RW	1 Years English language learner	1,527	21.94	29.60	31.63	16.83
4	RW	2 Years English language learner	1,114	10.32	27.20	38.78	23.70
4	RW	3 Years English language learner	790	19.62	21.14	35.06	24.18
4	RW	4 Years English language learner	1,131	6.10	22.19	45.98	25.73
4	RW	5 Years English language learner	2,597	3.47	19.95	48.90	27.69
4	RW	Arabic	385	14.55	27.53	37.40	20.52
4	RW	Bengali	323	5.88	20.74	39.32	34.06
4	RW	Chinese	351	7.12	22.22	39.89	30.77
4	RW	English	743	5.65	20.73	46.84	26.78
4	RW	Spanish	10,987	8.28	25.75	45.85	20.11
4	RW	Other Language	4,675	11.85	22.25	41.18	24.73
4	RW	Students with Disabilities	5,117	14.29	36.11	39.26	10.34

H.5c: Grade 4 Overall

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
4	Overall	Female	7,793	7.11	23.10	48.15	21.65
4	Overall	Male	9,581	10.75	26.12	45.39	17.73
4	Overall	American Indian or Alaska Native	20	5.00	35.00	35.00	25.00
4	Overall	Black or African American	928	12.61	28.23	41.92	17.24
4	Overall	Asian	2,539	8.31	19.65	44.23	27.81
4	Overall	Native Hawaiian/Other Pacific Islander	6	16.67	33.33	50.00	0.00
4	Overall	Hispanic or Latino	11,640	7.91	25.58	48.24	18.26
4	Overall	White	2,218	14.92	24.66	43.06	17.36
4	Overall	Multiracial (not Hispanic origin)	23	8.70	34.78	39.13	17.39
4	Overall	NYC	11,288	7.94	25.80	47.64	18.62
4	Overall	Big 4 Cities	1,024	14.75	28.91	40.33	16.02
4	Overall	High Need Urban/Suburban	2,621	14.27	23.12	46.24	16.37
4	Overall	High Need Rural	120	7.50	20.83	45.83	25.83
4	Overall	Average Need	1,493	7.43	20.29	44.94	27.33
4	Overall	Low Need	633	3.79	18.64	43.60	33.97
4	Overall	Charter Schools	95	0.00	18.95	56.84	24.21
4	Overall	0 Years English language learner	10,202	8.11	25.20	48.42	18.27
4	Overall	1 Years English language learner	1,516	21.70	30.21	33.44	14.64
4	Overall	2 Years English language learner	1,113	10.42	27.22	42.50	19.86
4	Overall	3 Years English language learner	788	19.67	21.32	37.18	21.83
4	Overall	4 Years English language learner	1,120	5.89	24.64	49.02	20.45
4	Overall	5 Years English language learner	2,592	3.47	19.98	50.93	25.62
4	Overall	Arabic	385	14.55	27.79	39.74	17.92
4	Overall	Bengali	323	6.19	20.74	42.41	30.65
4	Overall	Chinese	346	7.23	22.25	43.06	27.46
4	Overall	English	743	5.65	20.86	48.05	25.44
4	Overall	Spanish	10,945	8.12	25.83	48.10	17.94
4	Overall	Other Language	4,633	11.91	23.10	44.05	20.94
4	Overall	Students with Disabilities	5,087	14.31	37.02	40.85	7.82

H.6a: Grade 5 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
5	LS	Female	6,961	2.63	9.18	40.77	47.42
5	LS	Male	8,463	5.13	8.97	39.31	46.59
5	LS	American Indian or Alaska Native	21	0.00	4.76	42.86	52.38
5	LS	Black or African American	912	1.97	11.73	42.87	43.42
5	LS	Asian	2,172	4.93	12.02	35.77	47.28
5	LS	Native Hawaiian/Other Pacific Islander	7	0.00	14.29	57.14	28.57
5	LS	Hispanic or Latino	10,174	2.37	6.44	41.02	50.18
5	LS	White	2,117	11.62	17.29	38.03	33.07
5	LS	Multiracial (not Hispanic origin)	22	22.73	31.82	31.82	13.64
5	LS	NYC	10,202	2.44	7.76	40.34	49.46
5	LS	Big 4 Cities	915	6.34	14.97	42.30	36.39
5	LS	High Need Urban/Suburban	2,279	11.36	13.30	40.46	34.88
5	LS	High Need Rural	99	1.01	7.07	41.41	50.51
5	LS	Average Need	1,214	2.31	8.98	35.09	53.62
5	LS	Low Need	488	2.87	7.17	37.50	52.46
5	LS	Charter Schools	117	1.71	1.71	43.59	52.99
5	LS	0 Years English language learner	9,152	2.74	7.56	39.06	50.63
5	LS	1 Years English language learner	1,602	10.86	21.72	42.38	25.03
5	LS	2 Years English language learner	1,130	1.15	10.71	47.35	40.80
5	LS	3 Years English language learner	522	3.07	4.79	43.10	49.04
5	LS	4 Years English language learner	764	17.02	15.05	38.22	29.71
5	LS	5 Years English language learner	665	4.66	10.53	40.15	44.66
5	LS	6 Years English language learner or More	1,551	0.06	1.35	37.33	61.25
5	LS	Arabic	384	5.21	16.67	39.06	39.06
5	LS	Bengali	265	1.51	11.32	35.47	51.70
5	LS	Chinese	269	5.58	15.61	30.48	48.33
5	LS	English	514	1.17	4.47	39.11	55.25
5	LS	Spanish	9,673	2.43	6.43	41.16	49.98
5	LS	Other Language	4,320	7.80	14.28	38.38	39.54
5	LS	Students with Disabilities	5,038	4.96	9.45	47.88	37.71

H.6b: Grade 5 Reading and Writing

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
5	RW	Female	6,756	7.74	18.46	41.39	32.42
5	RW	Male	8,076	12.10	19.71	39.09	29.10
5	RW	American Indian or Alaska Native	20	10.00	15.00	40.00	35.00
5	RW	Black or African American	869	13.92	23.48	38.67	23.94
5	RW	Asian	2,047	10.36	15.88	34.64	39.13
5	RW	Native Hawaiian/Other Pacific Islander	7	0.00	28.57	28.57	42.86
5	RW	Hispanic or Latino	9,915	8.25	19.45	42.44	29.86
5	RW	White	1,957	17.42	18.96	35.16	28.46
5	RW	Multiracial (not Hispanic origin)	17	35.29	35.29	11.76	17.65
5	RW	NYC	9,886	8.79	19.54	41.26	30.41
5	RW	Big 4 Cities	835	14.73	22.04	42.51	20.72
5	RW	High Need Urban/Suburban	2,139	16.08	17.91	34.88	31.14
5	RW	High Need Rural	102	14.71	16.67	38.24	30.39
5	RW	Average Need	1,182	8.54	17.43	40.61	33.42
5	RW	Low Need	470	5.32	16.81	34.89	42.98
5	RW	Charter Schools	114	2.63	14.04	46.49	36.84
5	RW	0 Years English language learner	8,873	8.54	18.80	42.20	30.46
5	RW	1 Years English language learner	1,446	22.48	25.38	30.57	21.58
5	RW	2 Years English language learner	1,095	10.96	21.37	34.89	32.79
5	RW	3 Years English language learner	506	9.09	16.60	38.14	36.17
5	RW	4 Years English language learner	714	22.97	17.93	30.67	28.43
5	RW	5 Years English language learner	621	6.12	20.13	44.28	29.47
5	RW	6 Years English language learner or More	1,542	3.05	14.59	44.42	37.94
5	RW	Arabic	351	18.23	21.37	35.04	25.36
5	RW	Bengali	255	10.59	17.25	36.86	35.29
5	RW	Chinese	250	11.60	16.00	30.80	41.60
5	RW	English	504	7.14	16.27	42.26	34.33
5	RW	Spanish	9,422	8.32	19.51	42.27	29.90
5	RW	Other Language	4,050	13.83	18.77	36.12	31.28
5	RW	Students with Disabilities	4,807	14.17	26.98	41.40	17.45

H.6c: Grade 5 Overall

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
5	Overall	Female	6,720	7.60	19.64	48.21	24.54
5	Overall	Male	8,042	12.17	20.34	44.60	22.88
5	Overall	American Indian or Alaska Native	20	10.00	15.00	40.00	35.00
5	Overall	Black or African American	866	13.74	24.02	42.49	19.75
5	Overall	Asian	2,039	10.30	16.43	42.28	31.00
5	Overall	Native Hawaiian/Other Pacific Islander	7	0.00	28.57	57.14	14.29
5	Overall	Hispanic or Latino	9,884	8.12	19.90	48.60	23.37
5	Overall	White	1,929	18.09	22.60	40.33	18.97
5	Overall	Multiracial (not Hispanic origin)	17	41.18	29.41	17.65	11.76
5	Overall	NYC	9,844	8.67	20.23	46.98	24.13
5	Overall	Big 4 Cities	833	14.77	22.69	48.62	13.93
5	Overall	High Need Urban/Suburban	2,129	17.05	20.15	42.88	19.92
5	Overall	High Need Rural	95	8.42	17.89	47.37	26.32
5	Overall	Average Need	1,175	8.09	17.87	45.45	28.60
5	Overall	Low Need	469	5.33	17.48	43.28	33.90
5	Overall	Charter Schools	114	2.63	14.04	55.26	28.07
5	Overall	0 Years English language learner	8,829	8.40	19.66	47.72	24.22
5	Overall	1 Years English language learner	1,436	22.35	26.53	36.42	14.69
5	Overall	2 Years English language learner	1,094	11.06	21.57	41.86	25.50
5	Overall	3 Years English language learner	505	9.50	16.24	44.75	29.50
5	Overall	4 Years English language learner	702	21.65	22.65	39.89	15.81
5	Overall	5 Years English language learner	620	9.19	20.32	49.35	21.13
5	Overall	6 Years English language learner or More	1,542	3.05	14.79	52.33	29.83
5	Overall	Arabic	350	18.29	22.57	37.71	21.43
5	Overall	Bengali	254	10.24	18.50	39.37	31.89
5	Overall	Chinese	246	10.98	16.26	36.18	36.59
5	Overall	English	502	6.97	16.53	47.01	29.48
5	Overall	Spanish	9,394	8.21	19.95	48.56	23.28
5	Overall	Other Language	4,016	14.12	20.74	42.53	22.61
5	Overall	Students with Disabilities	4,778	14.46	28.59	45.17	11.78

H.7a: Grade 6 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
6	LS	Female	5,806	5.18	11.02	50.38	33.41
6	LS	Male	7,235	7.84	11.71	47.56	32.90
6	LS	American Indian or Alaska Native	18	5.56	11.11	44.44	38.89
6	LS	Black or African American	832	5.29	15.50	51.32	27.88
6	LS	Asian	2,034	8.36	15.59	43.12	32.94
6	LS	Native Hawaiian/Other Pacific Islander	5	0.00	0.00	40.00	60.00
6	LS	Hispanic or Latino	8,396	4.68	9.43	51.01	34.87
6	LS	White	1,748	14.65	14.13	43.82	27.40
6	LS	Multiracial (not Hispanic origin)	8	37.50	12.50	37.50	12.50
6	LS	NYC	8,732	5.35	10.82	50.11	33.72
6	LS	Big 4 Cities	789	5.83	15.46	47.28	31.43
6	LS	High Need Urban/Suburban	1,822	14.60	12.62	45.39	27.39
6	LS	High Need Rural	86	6.98	11.63	34.88	46.51
6	LS	Average Need	960	4.90	12.60	45.63	36.88
6	LS	Low Need	407	5.41	9.34	47.42	37.84
6	LS	Charter Schools	77	0.00	1.30	55.84	42.86
6	LS	0 Years English language learner	7,606	5.81	9.40	49.65	35.14
6	LS	1 Years English language learner	1,593	13.75	27.37	42.25	16.64
6	LS	2 Years English language learner	1,082	2.40	16.08	55.64	25.88
6	LS	3 Years English language learner	442	0.90	10.41	53.17	35.52
6	LS	4 Years English language learner	604	23.34	7.62	44.70	24.34
6	LS	5 Years English language learner	331	1.21	8.76	54.38	35.65
6	LS	6 Years English language learner or More	1,347	2.15	2.75	45.29	49.81
6	LS	Arabic	345	6.96	17.39	47.83	27.83
6	LS	Bengali	211	7.11	17.06	35.55	40.28
6	LS	Chinese	280	12.14	19.64	35.71	32.50
6	LS	English	379	1.58	7.92	50.66	39.84
6	LS	Spanish	7,997	4.85	9.48	51.01	34.66
6	LS	Other Language	3,830	10.47	14.33	45.82	29.37
6	LS	Students with Disabilities	4,181	6.03	9.26	55.99	28.73

H.7b: Grade 6 Reading and Writing

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
6	RW	Female	5,689	10.58	21.08	39.08	29.27
6	RW	Male	6,907	14.43	25.21	35.08	25.28
6	RW	American Indian or Alaska Native	18	16.67	22.22	33.33	27.78
6	RW	Black or African American	782	14.07	27.37	36.19	22.38
6	RW	Asian	1,925	12.62	19.27	31.48	36.62
6	RW	Native Hawaiian/Other Pacific Islander	5	0.00	0.00	0.00	100.00
6	RW	Hispanic or Latino	8,190	11.14	24.48	39.32	25.07
6	RW	White	1,669	19.65	20.67	31.64	28.04
6	RW	Multiracial (not Hispanic origin)	6	33.33	16.67	50.00	0.00
6	RW	NYC	8,406	11.72	24.08	37.70	26.50
6	RW	Big 4 Cities	747	14.19	26.64	37.08	22.09
6	RW	High Need Urban/Suburban	1,783	19.80	20.25	34.32	25.63
6	RW	High Need Rural	80	7.50	22.50	35.00	35.00
6	RW	Average Need	952	10.61	22.37	35.61	31.41
6	RW	Low Need	389	5.91	15.68	34.96	43.44
6	RW	Charter Schools	76	1.32	14.47	42.11	42.11
6	RW	0 Years English language learner	7,329	11.27	23.39	38.93	26.42
6	RW	1 Years English language learner	1,482	25.24	30.63	24.29	19.84
6	RW	2 Years English language learner	1,058	13.33	26.84	31.57	28.26
6	RW	3 Years English language learner	435	6.67	18.62	40.23	34.48
6	RW	4 Years English language learner	599	25.88	17.03	32.72	24.37
6	RW	5 Years English language learner	329	6.08	16.41	41.03	36.47
6	RW	6 Years English language learner or More	1,328	3.61	18.52	44.05	33.81
6	RW	Arabic	313	18.85	26.20	30.03	24.92
6	RW	Bengali	198	11.62	20.71	31.82	35.86
6	RW	Chinese	265	16.23	21.13	29.81	32.83
6	RW	English	371	5.39	22.64	36.66	35.31
6	RW	Spanish	7,800	11.42	24.51	39.24	24.82
6	RW	Other Language	3,649	15.43	20.96	33.24	30.36
6	RW	Students with Disabilities	4,042	14.72	29.94	38.35	17.00

H.7c: Grade 6 Overall

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
6	Overall	Female	5,597	10.34	22.48	49.08	18.10
6	Overall	Male	6,849	14.34	25.86	42.88	16.92
6	Overall	American Indian or Alaska Native	18	16.67	22.22	38.89	22.22
6	Overall	Black or African American	778	13.88	28.66	42.80	14.65
6	Overall	Asian	1,905	12.07	20.89	43.41	23.62
6	Overall	Native Hawaiian/Other Pacific Islander	5	0.00	0.00	40.00	60.00
6	Overall	Hispanic or Latino	8,107	10.76	25.04	48.03	16.17
6	Overall	White	1,626	21.22	22.94	38.01	17.84
6	Overall	Multiracial (not Hispanic origin)	6	33.33	16.67	50.00	0.00
6	Overall	NYC	8,326	11.40	24.92	46.12	17.56
6	Overall	Big 4 Cities	740	13.92	26.89	44.46	14.73
6	Overall	High Need Urban/Suburban	1,733	19.79	22.62	43.28	14.31
6	Overall	High Need Rural	80	7.50	22.50	42.50	27.50
6	Overall	Average Need	943	11.45	23.33	44.86	20.36
6	Overall	Low Need	389	6.17	17.22	49.36	27.25
6	Overall	Charter Schools	76	1.32	14.47	59.21	25.00
6	Overall	0 Years English language learner	7,251	10.92	24.16	47.28	17.64
6	Overall	1 Years English language learner	1,450	24.97	32.21	31.66	11.17
6	Overall	2 Years English language learner	1,049	12.87	28.31	42.14	16.68
6	Overall	3 Years English language learner	434	6.68	20.51	52.07	20.74
6	Overall	4 Years English language learner	578	26.64	17.99	42.21	13.15
6	Overall	5 Years English language learner	326	5.52	20.25	55.83	18.40
6	Overall	6 Years English language learner or More	1,323	4.91	18.82	51.78	24.49
6	Overall	Arabic	312	18.59	27.24	38.14	16.03
6	Overall	Bengali	197	11.17	22.84	38.07	27.92
6	Overall	Chinese	261	14.94	24.90	37.93	22.22
6	Overall	English	370	5.68	22.97	48.11	23.24
6	Overall	Spanish	7,716	11.02	25.08	47.77	16.14
6	Overall	Other Language	3,590	15.91	22.67	42.53	18.89
6	Overall	Students with Disabilities	3,993	15.28	31.43	43.35	9.94

H.8a: Grade 7 Listening and Speaking

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
7	LS	Female	5,559	3.69	13.11	35.89	47.31
7	LS	Male	6,693	5.57	13.91	33.78	46.74
7	LS	American Indian or Alaska Native	10	0.00	20.00	40.00	40.00
7	LS	Black or African American	844	2.84	13.03	41.71	42.42
7	LS	Asian	2,126	6.68	21.45	31.89	39.98
7	LS	Native Hawaiian/Other Pacific Islander	4	0.00	25.00	25.00	50.00
7	LS	Hispanic or Latino	7,723	3.63	10.79	35.32	50.27
7	LS	White	1,528	8.38	16.56	32.00	43.06
7	LS	Multiracial (not Hispanic origin)	17	23.53	29.41	23.53	23.53
7	LS	NYC	8,242	3.70	12.65	36.00	47.65
7	LS	Big 4 Cities	790	8.23	16.46	37.09	38.23
7	LS	High Need Urban/Suburban	1,710	9.24	16.73	30.76	43.27
7	LS	High Need Rural	85	9.41	15.29	28.24	47.06
7	LS	Average Need	887	3.04	15.11	30.44	51.41
7	LS	Low Need	417	2.40	10.55	33.33	53.72
7	LS	Charter Schools	50	0.00	4.00	26.00	70.00
7	LS	0 Years English language learner	6,971	3.86	11.55	33.97	50.62
7	LS	1 Years English language learner	1,676	11.04	29.77	36.99	22.20
7	LS	2 Years English language learner	1,140	2.46	14.65	43.33	39.56
7	LS	3 Years English language learner	466	0.43	6.87	38.84	53.86
7	LS	4 Years English language learner	562	15.66	14.06	32.92	37.37
7	LS	5 Years English language learner	294	0.34	4.76	35.03	59.86
7	LS	6 Years English language learner or More	1,115	0.27	5.38	26.28	68.07
7	LS	Arabic	306	4.58	17.32	39.54	38.56
7	LS	Bengali	214	5.61	20.56	30.37	43.46
7	LS	Chinese	328	9.76	26.52	32.62	31.10
7	LS	English	351	3.13	7.69	29.63	59.54
7	LS	Spanish	7,384	3.63	10.79	36.08	49.50
7	LS	Other Language	3,669	6.57	17.77	32.57	43.09
7	LS	Students with Disabilities	3,528	3.83	9.38	40.90	45.89

H.8b: Grade 7 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
7	RW	Female	5,347	16.10	31.91	30.39	21.60
7	RW	Male	6,270	22.50	33.83	28.18	15.49
7	RW	American Indian or Alaska Native	10	40.00	10.00	30.00	20.00
7	RW	Black or African American	802	19.95	33.54	27.31	19.20
7	RW	Asian	1,936	22.26	28.15	27.84	21.75
7	RW	Native Hawaiian/Other Pacific Islander	4	25.00	50.00	25.00	0.00
7	RW	Hispanic or Latino	7,434	18.08	35.35	30.25	16.32
7	RW	White	1,420	23.17	26.62	26.69	23.52
7	RW	Multiracial (not Hispanic origin)	11	27.27	36.36	18.18	18.18
7	RW	NYC	7,857	18.29	35.37	29.25	17.09
7	RW	Big 4 Cities	712	29.07	32.30	23.88	14.75
7	RW	High Need Urban/Suburban	1,598	24.72	26.53	28.91	19.84
7	RW	High Need Rural	73	21.92	27.40	32.88	17.81
7	RW	Average Need	859	17.69	27.71	30.97	23.63
7	RW	Low Need	404	11.39	25.99	31.19	31.44
7	RW	Charter Schools	49	6.12	20.41	48.98	24.49
7	RW	0 Years English language learner	6,661	17.23	34.42	30.60	17.75
7	RW	1 Years English language learner	1,503	36.99	33.27	18.76	10.98
7	RW	2 Years English language learner	1,087	19.50	36.34	25.48	18.68
7	RW	3 Years English language learner	457	13.35	25.82	33.92	26.91
7	RW	4 Years English language learner	515	30.87	26.21	23.69	19.22
7	RW	5 Years English language learner	290	9.66	29.66	33.10	27.59
7	RW	6 Years English language learner or More	1,077	9.38	27.21	38.35	25.07
7	RW	Arabic	275	26.18	33.82	23.64	16.36
7	RW	Bengali	197	22.34	27.41	32.99	17.26
7	RW	Chinese	286	25.17	29.72	25.52	19.58
7	RW	English	339	15.34	28.02	30.38	26.25
7	RW	Spanish	7,101	18.29	35.67	30.18	15.86
7	RW	Other Language	3,419	21.44	28.28	27.58	22.70
7	RW	Students with Disabilities	3,338	21.75	40.17	26.18	11.89

H.8c: Grade 7 Overall

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
7	Overall	Female	5,285	15.63	32.19	33.26	18.92
7	Overall	Male	6,213	22.07	34.14	29.74	14.05
7	Overall	American Indian or Alaska Native	10	40.00	10.00	30.00	20.00
7	Overall	Black or African American	799	19.65	34.17	30.16	16.02
7	Overall	Asian	1,922	22.06	28.20	29.92	19.82
7	Overall	Native Hawaiian/Other Pacific Islander	4	25.00	50.00	25.00	0.00
7	Overall	Hispanic or Latino	7,362	17.52	35.66	32.07	14.75
7	Overall	White	1,390	22.88	26.98	30.43	19.71
7	Overall	Multiracial (not Hispanic origin)	11	27.27	36.36	18.18	18.18
7	Overall	NYC	7,776	17.73	35.69	31.24	15.34
7	Overall	Big 4 Cities	707	28.71	32.53	26.59	12.16
7	Overall	High Need Urban/Suburban	1,576	24.87	26.78	31.79	16.56
7	Overall	High Need Rural	73	21.92	27.40	32.88	17.81
7	Overall	Average Need	849	16.73	28.15	32.98	22.14
7	Overall	Low Need	404	11.63	25.74	34.41	28.22
7	Overall	Charter Schools	49	6.12	20.41	48.98	24.49
7	Overall	0 Years English language learner	6,577	16.59	34.80	32.58	16.03
7	Overall	1 Years English language learner	1,488	36.83	33.53	20.36	9.27
7	Overall	2 Years English language learner	1,086	19.52	36.37	28.18	15.93
7	Overall	3 Years English language learner	456	13.38	25.88	36.40	24.34
7	Overall	4 Years English language learner	511	30.92	26.03	26.81	16.24
7	Overall	5 Years English language learner	287	9.76	29.62	36.93	23.69
7	Overall	6 Years English language learner or More	1,066	8.63	27.77	40.90	22.70
7	Overall	Arabic	274	25.91	33.94	25.55	14.60
7	Overall	Bengali	197	22.34	28.43	32.49	16.75
7	Overall	Chinese	281	24.56	29.89	28.11	17.44
7	Overall	English	339	15.34	28.32	32.15	24.19
7	Overall	Spanish	7,029	17.71	35.97	32.07	14.26
7	Overall	Other Language	3,378	21.20	28.57	30.49	19.75
7	Overall	Students with Disabilities	3,306	21.66	40.47	28.10	9.77

H.9a: Grade 8 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
8	LS	Female	5,560	4.87	15.13	27.39	52.61
8	LS	Male	6,617	6.00	16.05	26.01	51.94
8	LS	American Indian or Alaska Native	14	0.00	14.29	28.57	57.14
8	LS	Black or African American	819	4.15	19.17	32.48	44.20
8	LS	Asian	2,129	7.66	22.36	26.68	43.31
8	LS	Native Hawaiian/Other Pacific Islander	4	0.00	25.00	25.00	50.00
8	LS	Hispanic or Latino	7,826	4.59	13.62	26.02	55.78
8	LS	White	1,367	8.19	14.26	26.48	51.06
8	LS	Multiracial (not Hispanic origin)	18	0.00	33.33	38.89	27.78
8	LS	NYC	8,466	4.41	15.02	27.11	53.46
8	LS	Big 4 Cities	716	8.10	22.63	26.82	42.46
8	LS	High Need Urban/Suburban	1,558	11.62	17.01	25.29	46.08
8	LS	High Need Rural	86	3.49	17.44	25.58	53.49
8	LS	Average Need	833	4.68	14.77	23.29	57.26
8	LS	Low Need	398	3.27	13.57	29.40	53.77
8	LS	Charter Schools	47	0.00	2.13	25.53	72.34
8	LS	0 Years English language learner	7,052	4.55	12.71	25.78	56.96
8	LS	1 Years English language learner	1,761	13.12	33.90	29.36	23.62
8	LS	2 Years English language learner	1,244	3.94	20.90	34.65	40.51
8	LS	3 Years English language learner	443	1.35	13.32	26.86	58.47
8	LS	4 Years English language learner	526	10.46	9.51	28.33	51.71
8	LS	5 Years English language learner	215	0.47	2.79	22.33	74.42
8	LS	6 Years English language learner or More	911	0.22	3.51	17.01	79.25
8	LS	Arabic	310	4.19	20.97	32.26	42.58
8	LS	Bengali	217	2.30	16.59	25.81	55.30
8	LS	Chinese	327	7.65	28.44	34.86	29.05
8	LS	English	334	2.10	11.38	20.66	65.87
8	LS	Spanish	7,510	4.62	13.85	26.64	54.89
8	LS	Other Language	3,479	7.79	18.14	25.98	48.09
8	LS	Students with Disabilities	2,928	3.96	9.29	31.80	54.95

H.9b: Grade 8 Reading and Writing

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
8	RW	Female	5,383	15.83	32.36	22.76	29.05
8	RW	Male	6,239	20.72	36.62	21.43	21.22
8	RW	American Indian or Alaska Native	13	15.38	30.77	30.77	23.08
8	RW	Black or African American	768	19.79	34.38	21.35	24.48
8	RW	Asian	1,976	19.33	33.05	19.03	28.59
8	RW	Native Hawaiian/Other Pacific Islander	4	25.00	0.00	25.00	50.00
8	RW	Hispanic or Latino	7,552	18.11	36.43	22.97	22.48
8	RW	White	1,294	18.08	26.97	21.64	33.31
8	RW	Multiracial (not Hispanic origin)	15	40.00	40.00	13.33	6.67
8	RW	NYC	8,097	17.23	35.85	22.55	24.37
8	RW	Big 4 Cities	648	27.78	37.96	18.36	15.90
8	RW	High Need Urban/Suburban	1,500	24.53	31.13	20.47	23.87
8	RW	High Need Rural	82	25.61	25.61	20.73	28.05
8	RW	Average Need	802	16.21	30.55	21.57	31.67
8	RW	Low Need	383	8.88	29.24	24.80	37.08
8	RW	Charter Schools	46	4.35	26.09	21.74	47.83
8	RW	0 Years English language learner	6,741	15.46	35.28	23.87	25.40
8	RW	1 Years English language learner	1,601	35.92	35.98	13.30	14.80
8	RW	2 Years English language learner	1,181	22.95	35.82	18.37	22.86
8	RW	3 Years English language learner	431	15.31	35.27	19.49	29.93
8	RW	4 Years English language learner	539	19.11	29.50	23.93	27.46
8	RW	5 Years English language learner	209	8.13	31.10	28.23	32.54
8	RW	6 Years English language learner or More	895	7.37	29.39	27.82	35.42
8	RW	Arabic	282	19.86	40.78	17.73	21.63
8	RW	Bengali	209	11.96	33.49	20.10	34.45
8	RW	Chinese	300	21.33	34.67	19.33	24.67
8	RW	English	325	11.69	33.54	19.08	35.69
8	RW	Spanish	7,235	18.38	36.66	22.85	22.11
8	RW	Other Language	3,271	19.32	29.87	21.31	29.50
8	RW	Students with Disabilities	2,812	19.17	39.90	23.01	17.92

H.9c: Grade 8 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
8	Overall	Female	5,306	15.49	32.74	25.82	25.95
8	Overall	Male	6,176	20.39	37.03	22.67	19.92
8	Overall	American Indian or Alaska Native	13	15.38	30.77	30.77	23.08
8	Overall	Black or African American	763	19.79	34.73	24.25	21.23
8	Overall	Asian	1,963	19.46	33.62	22.26	24.66
8	Overall	Native Hawaiian/Other Pacific Islander	4	25.00	0.00	25.00	50.00
8	Overall	Hispanic or Latino	7,467	17.60	36.68	24.56	21.16
8	Overall	White	1,257	17.90	27.84	24.42	29.83
8	Overall	Multiracial (not Hispanic origin)	15	40.00	40.00	13.33	6.67
8	Overall	NYC	8,018	16.89	36.16	24.62	22.34
8	Overall	Big 4 Cities	639	27.54	38.50	18.78	15.18
8	Overall	High Need Urban/Suburban	1,455	24.47	31.82	22.54	21.17
8	Overall	High Need Rural	82	25.61	25.61	20.73	28.05
8	Overall	Average Need	796	15.58	31.03	23.74	29.65
8	Overall	Low Need	382	8.64	30.10	30.10	31.15
8	Overall	Charter Schools	46	4.35	26.09	26.09	43.48
8	Overall	0 Years English language learner	6,675	15.06	35.63	26.02	23.30
8	Overall	1 Years English language learner	1,566	35.31	37.04	15.39	12.26
8	Overall	2 Years English language learner	1,177	23.02	35.85	20.73	20.39
8	Overall	3 Years English language learner	431	15.55	35.03	21.58	27.84
8	Overall	4 Years English language learner	512	19.73	30.08	25.39	24.80
8	Overall	5 Years English language learner	209	8.13	31.58	29.19	31.10
8	Overall	6 Years English language learner or More	889	6.97	29.47	29.47	34.08
8	Overall	Arabic	281	19.93	40.57	19.57	19.93
8	Overall	Bengali	208	12.02	34.62	23.08	30.29
8	Overall	Chinese	297	21.55	34.68	23.57	20.20
8	Overall	English	324	11.73	33.64	20.68	33.95
8	Overall	Spanish	7,154	17.88	36.89	24.53	20.70
8	Overall	Other Language	3,218	19.24	30.67	24.08	26.01
8	Overall	Students with Disabilities	2,769	19.07	40.30	25.24	15.38

H.10a: Grade 9 Listening and Speaking

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
9	LS	Female	6,178	11.17	24.41	22.40	42.02
9	LS	Male	7,248	12.90	23.57	21.52	42.01
9	LS	American Indian or Alaska Native	23	8.70	26.09	26.09	39.13
9	LS	Black or African American	1,039	7.51	28.78	25.31	38.40
9	LS	Asian	2,697	14.35	32.52	22.80	30.33
9	LS	Native Hawaiian/Other Pacific Islander	2	0.00	50.00	0.00	50.00
9	LS	Hispanic or Latino	8,314	12.44	21.51	20.78	45.27
9	LS	White	1,304	8.74	17.10	24.77	49.39
9	LS	Multiracial (not Hispanic origin)	45	20.00	46.67	20.00	13.33
9	LS	NYC	8,750	10.80	23.49	22.27	43.44
9	LS	Big 4 Cities	856	11.10	29.79	18.22	40.89
9	LS	High Need Urban/Suburban	1,630	15.52	23.01	22.27	39.20
9	LS	High Need Rural	85	11.76	18.82	16.47	52.94
9	LS	Average Need	1,113	12.04	23.63	22.64	41.69
9	LS	Low Need	487	11.91	19.92	25.46	42.71
9	LS	Charter Schools	44	6.82	9.09	34.09	50.00
9	LS	0 Years English language learner	7,354	12.51	20.98	20.45	46.06
9	LS	1 Years English language learner	2,574	21.45	39.70	20.86	17.99
9	LS	2 Years English language learner	1,490	8.19	30.34	27.58	33.89
9	LS	3 Years English language learner	470	2.55	17.87	32.13	47.45
9	LS	4 Years English language learner	491	1.02	9.57	31.36	58.04
9	LS	5 Years English language learner	323	0.62	10.22	25.39	63.78
9	LS	6 Years English language learner or More	671	0.30	2.98	14.01	82.71
9	LS	Arabic	374	9.09	29.14	29.14	32.62
9	LS	Bengali	260	13.46	25.38	24.23	36.92
9	LS	Chinese	365	15.34	32.05	24.38	28.22
9	LS	English	425	11.29	19.53	20.00	49.18
9	LS	Spanish	7,913	12.44	21.69	20.99	44.89
9	LS	Other Language	4,089	11.45	27.51	22.92	38.13
9	LS	Students with Disabilities	2,060	4.95	11.21	28.45	55.39

H.10b: Grade 9 Reading and Writing

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
9	RW	Female	5,853	13.82	44.11	24.82	17.24
9	RW	Male	6,518	17.17	42.97	23.00	16.86
9	RW	American Indian or Alaska Native	22	18.18	50.00	18.18	13.64
9	RW	Black or African American	963	15.78	46.00	21.81	16.41
9	RW	Asian	2,457	17.42	45.14	21.16	16.28
9	RW	Native Hawaiian/Other Pacific Islander	2	0.00	0.00	100.00	0.00
9	RW	Hispanic or Latino	7,638	15.70	43.51	25.02	15.78
9	RW	White	1,248	10.58	38.14	24.12	27.16
9	RW	Multiracial (not Hispanic origin)	39	33.33	48.72	10.26	7.69
9	RW	NYC	8,080	14.01	44.41	24.81	16.77
9	RW	Big 4 Cities	769	22.50	43.82	20.55	13.13
9	RW	High Need Urban/Suburban	1,504	19.28	41.49	22.74	16.49
9	RW	High Need Rural	73	12.33	39.73	23.29	24.66
9	RW	Average Need	1,042	13.24	43.67	23.03	20.06
9	RW	Low Need	467	15.63	32.55	23.34	28.48
9	RW	Charter Schools	41	7.32	31.71	34.15	26.83
9	RW	0 Years English language learner	6,767	13.96	43.05	25.43	17.56
9	RW	1 Years English language learner	2,243	28.40	45.83	16.01	9.76
9	RW	2 Years English language learner	1,395	17.06	49.39	19.43	14.12
9	RW	3 Years English language learner	461	9.98	44.69	24.08	21.26
9	RW	4 Years English language learner	484	4.13	45.45	25.62	24.79
9	RW	5 Years English language learner	318	4.72	37.11	31.76	26.42
9	RW	6 Years English language learner or More	659	2.58	28.68	39.15	29.59
9	RW	Arabic	333	21.02	53.75	14.11	11.11
9	RW	Bengali	239	15.48	47.28	22.59	14.64
9	RW	Chinese	331	16.62	45.92	20.24	17.22
9	RW	English	401	14.46	42.64	26.68	16.21
9	RW	Spanish	7,259	15.65	43.77	24.85	15.73
9	RW	Other Language	3,808	15.02	41.78	22.93	20.27
9	RW	Students with Disabilities	1,938	9.80	48.19	25.23	16.77

H.10c: Grade 9 Overall

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
9	Overall	Female	5,789	15.62	43.51	25.57	15.30
9	Overall	Male	6,474	18.21	42.89	23.29	15.60
9	Overall	American Indian or Alaska Native	22	22.73	45.45	18.18	13.64
9	Overall	Black or African American	957	16.61	47.02	21.84	14.52
9	Overall	Asian	2,436	19.38	45.24	21.35	14.04
9	Overall	Native Hawaiian/Other Pacific Islander	2	0.00	50.00	50.00	0.00
9	Overall	Hispanic or Latino	7,569	16.86	42.90	25.45	14.80
9	Overall	White	1,236	12.54	37.86	26.21	23.38
9	Overall	Multiracial (not Hispanic origin)	39	38.46	43.59	10.26	7.69
9	Overall	NYC	8,010	15.11	44.16	25.24	15.49
9	Overall	Big 4 Cities	765	24.18	42.75	20.39	12.68
9	Overall	High Need Urban/Suburban	1,487	21.65	40.75	23.27	14.32
9	Overall	High Need Rural	73	13.70	39.73	21.92	24.66
9	Overall	Average Need	1,035	14.49	44.06	23.48	17.97
9	Overall	Low Need	464	16.81	33.41	26.94	22.84
9	Overall	Charter Schools	41	7.32	31.71	36.59	24.39
9	Overall	0 Years English language learner	6,697	15.47	42.15	25.97	16.41
9	Overall	1 Years English language learner	2,223	30.99	45.39	15.74	7.87
9	Overall	2 Years English language learner	1,393	17.73	50.39	19.45	12.42
9	Overall	3 Years English language learner	457	10.28	45.95	24.95	18.82
9	Overall	4 Years English language learner	480	4.38	45.21	29.38	21.04
9	Overall	5 Years English language learner	316	5.06	38.29	32.91	23.73
9	Overall	6 Years English language learner or More	653	2.45	29.86	39.97	27.72
9	Overall	Arabic	332	21.69	53.01	15.06	10.24
9	Overall	Bengali	235	17.02	46.38	22.13	14.47
9	Overall	Chinese	329	18.24	47.42	20.67	13.68
9	Overall	English	400	16.00	43.75	24.50	15.75
9	Overall	Spanish	7,192	16.73	43.20	25.33	14.74
9	Overall	Other Language	3,775	17.06	41.67	23.79	17.48
9	Overall	Students with Disabilities	1,926	11.16	48.08	26.53	14.23

H.11a: Grade 10 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
10	LS	Female	5,947	8.22	27.63	23.91	40.24
10	LS	Male	6,547	9.24	27.13	23.20	40.43
10	LS	American Indian or Alaska Native	16	6.25	31.25	25.00	37.50
10	LS	Black or African American	1,074	3.82	28.68	27.56	39.94
10	LS	Asian	3,051	12.32	37.07	24.48	26.12
10	LS	Native Hawaiian/Other Pacific Islander	13	0.00	23.08	30.77	46.15
10	LS	Hispanic or Latino	7,106	8.64	24.44	22.22	44.69
10	LS	White	1,191	4.53	18.14	25.52	51.81
10	LS	Multiracial (not Hispanic origin)	43	18.60	46.51	16.28	18.60
10	LS	NYC	9,217	9.08	28.85	22.65	39.42
10	LS	Big 4 Cities	458	7.64	25.33	21.40	45.63
10	LS	High Need Urban/Suburban	1,238	8.64	22.46	29.16	39.74
10	LS	High Need Rural	63	1.59	12.70	30.16	55.56
10	LS	Average Need	830	6.27	22.29	22.65	48.80
10	LS	Low Need	441	5.90	22.22	28.34	43.54
10	LS	Charter Schools	28	3.57	7.14	28.57	60.71
10	LS	0 Years English language learner	6,958	9.60	24.48	21.26	44.67
10	LS	1 Years English language learner	2,294	12.12	40.45	24.72	22.71
10	LS	2 Years English language learner	1,619	7.91	34.34	28.91	28.84
10	LS	3 Years English language learner	412	2.91	26.46	27.43	43.20
10	LS	4 Years English language learner	410	0.73	10.73	35.12	53.41
10	LS	5 Years English language learner	276	0.36	8.70	29.35	61.59
10	LS	6 Years English language learner or More	490	0.41	8.98	16.33	74.29
10	LS	Arabic	293	3.41	27.99	26.28	42.32
10	LS	Bengali	294	6.80	38.10	22.79	32.31
10	LS	Chinese	461	15.84	40.78	23.86	19.52
10	LS	English	297	5.05	15.82	27.27	51.85
10	LS	Spanish	6,816	8.83	24.75	21.89	44.53
10	LS	Other Language	4,334	8.63	30.09	25.70	35.58
10	LS	Students with Disabilities	1,495	3.55	15.18	28.09	53.18

H.11b: Grade 10 Reading and Writing

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
10	RW	Female	5,769	10.92	45.14	27.70	16.24
10	RW	Male	6,189	12.59	45.92	25.16	16.34
10	RW	American Indian or Alaska Native	17	11.76	41.18	35.29	11.76
10	RW	Black or African American	1,055	10.62	48.91	25.31	15.17
10	RW	Asian	2,870	12.93	50.73	21.25	15.09
10	RW	Native Hawaiian/Other Pacific Islander	13	0.00	46.15	23.08	30.77
10	RW	Hispanic or Latino	6,805	12.40	44.69	27.48	15.43
10	RW	White	1,160	6.21	34.66	33.71	25.43
10	RW	Multiracial (not Hispanic origin)	38	21.05	44.74	23.68	10.53
10	RW	NYC	8,805	11.36	46.78	25.84	16.02
10	RW	Big 4 Cities	437	15.56	47.37	26.09	10.98
10	RW	High Need Urban/Suburban	1,197	14.12	40.94	28.24	16.71
10	RW	High Need Rural	59	11.86	44.07	25.42	18.64
10	RW	Average Need	804	8.71	41.92	29.35	20.02
10	RW	Low Need	429	8.62	40.33	29.84	21.21
10	RW	Charter Schools	26	3.85	26.92	34.62	34.62
10	RW	0 Years English language learner	6,616	10.64	43.95	27.99	17.41
10	RW	1 Years English language learner	2,177	17.50	51.54	19.75	11.21
10	RW	2 Years English language learner	1,556	13.95	51.86	20.44	13.75
10	RW	3 Years English language learner	402	12.94	45.02	28.61	13.43
10	RW	4 Years English language learner	408	5.39	34.80	38.24	21.57
10	RW	5 Years English language learner	280	3.57	41.43	32.50	22.50
10	RW	6 Years English language learner or More	485	3.51	31.75	38.35	26.39
10	RW	Arabic	284	10.92	52.11	26.06	10.92
10	RW	Bengali	283	11.66	56.89	20.85	10.60
10	RW	Chinese	430	13.72	47.91	19.07	19.30
10	RW	English	288	6.94	41.32	30.90	20.83
10	RW	Spanish	6,521	12.59	44.58	27.60	15.23
10	RW	Other Language	4,153	10.72	45.87	25.33	18.08
10	RW	Students with Disabilities	1,434	7.74	43.72	31.66	16.88

H.11c: Grade 10 Overall

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
10	Overall	Female	5,728	12.76	45.67	27.83	13.74
10	Overall	Male	6,162	14.14	46.45	25.51	13.91
10	Overall	American Indian or Alaska Native	16	6.25	56.25	31.25	6.25
10	Overall	Black or African American	1,052	11.60	49.81	25.95	12.64
10	Overall	Asian	2,860	15.84	52.10	21.26	10.80
10	Overall	Native Hawaiian/Other Pacific Islander	13	0.00	46.15	30.77	23.08
10	Overall	Hispanic or Latino	6,761	13.80	44.59	27.56	14.05
10	Overall	White	1,151	7.38	36.32	35.10	21.20
10	Overall	Multiracial (not Hispanic origin)	37	21.62	43.24	24.32	10.81
10	Overall	NYC	8,765	13.33	47.35	25.68	13.65
10	Overall	Big 4 Cities	433	16.63	47.34	26.79	9.24
10	Overall	High Need Urban/Suburban	1,183	15.55	40.91	30.35	13.19
10	Overall	High Need Rural	58	10.34	44.83	25.86	18.97
10	Overall	Average Need	802	9.10	43.27	29.55	18.08
10	Overall	Low Need	428	9.11	41.59	31.78	17.52
10	Overall	Charter Schools	26	3.85	26.92	42.31	26.92
10	Overall	0 Years English language learner	6,578	12.59	44.39	27.76	15.26
10	Overall	1 Years English language learner	2,163	19.69	52.10	19.88	8.32
10	Overall	2 Years English language learner	1,553	15.33	52.22	21.12	11.33
10	Overall	3 Years English language learner	402	13.43	46.27	28.86	11.44
10	Overall	4 Years English language learner	405	5.43	34.81	43.70	16.05
10	Overall	5 Years English language learner	275	3.64	42.55	35.64	18.18
10	Overall	6 Years English language learner or More	481	3.53	33.89	37.63	24.95
10	Overall	Arabic	283	11.31	54.06	24.03	10.60
10	Overall	Bengali	282	13.12	57.45	20.21	9.22
10	Overall	Chinese	430	19.07	50.23	19.53	11.16
10	Overall	English	287	7.67	41.46	31.01	19.86
10	Overall	Spanish	6,480	14.07	44.46	27.67	13.80
10	Overall	Other Language	4,129	12.52	47.18	26.04	14.26
10	Overall	Students with Disabilities	1,421	8.30	45.60	32.79	13.30

H.12a: Grade 11 Listening and Speaking

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
11	LS	Female	4,289	5.11	26.18	31.52	37.19
11	LS	Male	4,006	4.39	23.02	33.37	39.22
11	LS	American Indian or Alaska Native	15	6.67	26.67	26.67	40.00
11	LS	Black or African American	739	1.62	23.14	38.29	36.94
11	LS	Asian	1,977	4.20	34.80	34.45	26.56
11	LS	Native Hawaiian/Other Pacific Islander	2	0.00	0.00	50.00	50.00
11	LS	Hispanic or Latino	4,607	5.19	22.34	30.76	41.72
11	LS	White	915	5.57	15.30	32.13	46.99
11	LS	Multiracial (not Hispanic origin)	39	23.08	33.33	20.51	23.08
11	LS	NYC	5,780	4.45	25.80	32.42	37.34
11	LS	Big 4 Cities	300	5.67	27.33	30.33	36.67
11	LS	High Need Urban/Suburban	935	7.91	20.96	34.65	36.47
11	LS	High Need Rural	52	1.92	23.08	32.69	42.31
11	LS	Average Need	705	4.40	22.41	28.65	44.54
11	LS	Low Need	465	2.80	21.08	35.27	40.86
11	LS	Charter Schools	18	0.00	16.67	16.67	66.67
11	LS	0 Years English language learner	4,068	4.70	20.50	31.71	43.09
11	LS	1 Years English language learner	1,370	9.20	38.54	29.49	22.77
11	LS	2 Years English language learner	1,306	4.29	32.01	34.23	29.48
11	LS	3 Years English language learner	525	2.48	24.19	36.38	36.95
11	LS	4 Years English language learner	410	0.49	16.10	45.85	37.56
11	LS	5 Years English language learner	269	0.74	14.87	31.97	52.42
11	LS	6 Years English language learner or More	324	0.31	7.72	23.77	68.21
11	LS	Arabic	184	1.63	17.39	44.02	36.96
11	LS	Bengali	193	4.15	30.57	38.86	26.42
11	LS	Chinese	301	4.65	33.55	38.87	22.92
11	LS	English	252	2.38	23.81	31.75	42.06
11	LS	Spanish	4,344	5.23	22.28	30.71	41.78
11	LS	Other Language	3,021	4.53	27.31	33.17	34.99
11	LS	Students with Disabilities	946	5.60	17.12	33.72	43.55

H.12b: Grade 11 Reading and Writing

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
11	RW	Female	4,205	6.78	50.51	26.80	15.91
11	RW	Male	3,883	7.57	48.70	26.14	17.59
11	RW	American Indian or Alaska Native	15	6.67	53.33	26.67	13.33
11	RW	Black or African American	719	6.95	52.16	26.15	14.74
11	RW	Asian	1,951	6.92	51.10	24.71	17.27
11	RW	Native Hawaiian/Other Pacific Islander	2	0.00	50.00	0.00	50.00
11	RW	Hispanic or Latino	4,463	7.84	50.59	26.46	15.10
11	RW	White	903	4.65	39.09	31.23	25.03
11	RW	Multiracial (not Hispanic origin)	34	2.94	67.65	11.76	17.65
11	RW	NYC	5,630	6.48	50.71	26.29	16.52
11	RW	Big 4 Cities	282	15.60	53.90	20.21	10.28
11	RW	High Need Urban/Suburban	917	9.60	49.62	28.03	12.76
11	RW	High Need Rural	52	3.85	55.77	19.23	21.15
11	RW	Average Need	693	6.06	46.46	25.97	21.50
11	RW	Low Need	460	7.61	38.70	31.52	22.17
11	RW	Charter Schools	17	5.88	23.53	29.41	41.18
11	RW	0 Years English language learner	3,960	6.01	48.46	27.75	17.78
11	RW	1 Years English language learner	1,320	10.61	52.20	21.97	15.23
11	RW	2 Years English language learner	1,276	8.78	53.37	22.57	15.28
11	RW	3 Years English language learner	519	10.40	51.64	22.93	15.03
11	RW	4 Years English language learner	406	4.43	48.77	32.51	14.29
11	RW	5 Years English language learner	269	2.97	45.35	31.97	19.70
11	RW	6 Years English language learner or More	319	2.51	39.81	38.87	18.81
11	RW	Arabic	180	10.00	52.78	22.78	14.44
11	RW	Bengali	190	7.89	63.16	20.00	8.95
11	RW	Chinese	295	6.44	46.78	26.44	20.34
11	RW	English	250	5.60	48.80	28.00	17.60
11	RW	Spanish	4,202	7.85	51.02	26.15	14.97
11	RW	Other Language	2,971	6.16	46.99	27.47	19.39
11	RW	Students with Disabilities	919	7.62	49.18	26.77	16.43

H.12c: Grade 11 Overall

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
11	Overall	Female	4,184	8.44	52.13	27.92	11.52
11	Overall	Male	3,877	8.64	50.09	27.08	14.19
11	Overall	American Indian or Alaska Native	15	6.67	60.00	26.67	6.67
11	Overall	Black or African American	718	7.52	54.18	27.16	11.14
11	Overall	Asian	1,947	8.58	54.75	26.09	10.58
11	Overall	Native Hawaiian/Other Pacific Islander	2	0.00	50.00	0.00	50.00
11	Overall	Hispanic or Latino	4,454	8.91	51.06	27.23	12.80
11	Overall	White	890	7.19	41.01	32.92	18.88
11	Overall	Multiracial (not Hispanic origin)	34	14.71	55.88	11.76	17.65
11	Overall	NYC	5,621	7.97	52.07	27.49	12.47
11	Overall	Big 4 Cities	282	16.67	53.55	20.57	9.22
11	Overall	High Need Urban/Suburban	903	11.96	50.94	27.91	9.19
11	Overall	High Need Rural	52	3.85	57.69	21.15	17.31
11	Overall	Average Need	690	6.67	48.99	26.52	17.83
11	Overall	Low Need	459	7.41	42.05	33.77	16.78
11	Overall	Charter Schools	17	5.88	23.53	35.29	35.29
11	Overall	0 Years English language learner	3,949	7.57	49.46	28.97	14.00
11	Overall	1 Years English language learner	1,313	13.18	54.68	22.39	9.75
11	Overall	2 Years English language learner	1,276	9.72	54.94	23.51	11.83
11	Overall	3 Years English language learner	518	10.81	52.90	23.36	12.93
11	Overall	4 Years English language learner	406	4.43	50.74	35.71	9.11
11	Overall	5 Years English language learner	265	3.40	47.55	33.21	15.85
11	Overall	6 Years English language learner or More	315	2.54	42.22	38.73	16.51
11	Overall	Arabic	180	11.11	52.78	23.89	12.22
11	Overall	Bengali	190	8.95	64.74	20.00	6.32
11	Overall	Chinese	295	8.81	50.51	29.83	10.85
11	Overall	English	250	5.60	49.60	30.40	14.40
11	Overall	Spanish	4,193	8.99	51.35	27.05	12.62
11	Overall	Other Language	2,953	7.92	50.08	28.41	13.58
11	Overall	Students with Disabilities	904	9.85	51.44	26.88	11.84

H.13a: Grade 12 Listening and Speaking

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
12	LS	Female	2,888	2.35	29.64	26.25	41.76
12	LS	Male	2,874	2.54	24.50	24.11	48.85
12	LS	American Indian or Alaska Native	6	0.00	66.67	16.67	16.67
12	LS	Black or African American	559	0.54	24.69	31.66	43.11
12	LS	Asian	1,555	2.70	35.05	27.27	34.98
12	LS	Native Hawaiian/Other Pacific Islander	3	0.00	0.00	0.00	100.00
12	LS	Hispanic or Latino	3,086	2.30	24.30	23.75	49.64
12	LS	White	542	4.61	21.77	20.85	52.77
12	LS	Multiracial (not Hispanic origin)	11	0.00	45.45	27.27	27.27
12	LS	NYC	4,150	2.22	28.75	25.40	43.64
12	LS	Big 4 Cities	224	2.68	30.80	25.00	41.52
12	LS	High Need Urban/Suburban	517	4.64	25.73	25.34	44.29
12	LS	High Need Rural	28	0.00	7.14	28.57	64.29
12	LS	Average Need	515	2.52	20.19	21.17	56.12
12	LS	Low Need	292	1.37	15.75	29.45	53.42
12	LS	Charter Schools	9	0.00	66.67	0.00	33.33
12	LS	0 Years English language learner	3,163	2.15	25.17	25.48	47.20
12	LS	1 Years English language learner	612	4.90	38.56	22.55	33.99
12	LS	2 Years English language learner	807	3.35	36.31	25.77	34.57
12	LS	3 Years English language learner	317	1.58	22.08	27.44	48.90
12	LS	4 Years English language learner	355	0.56	23.10	28.45	47.89
12	LS	5 Years English language learner	177	0.00	19.21	26.55	54.24
12	LS	6 Years English language learner or More	315	2.22	15.24	18.41	64.13
12	LS	Arabic	126	0.79	21.43	26.98	50.79
12	LS	Bengali	136	2.21	37.50	17.65	42.65
12	LS	Chinese	294	3.40	46.26	26.53	23.81
12	LS	English	161	1.24	20.50	26.09	52.17
12	LS	Spanish	2,913	2.40	24.58	23.45	49.57
12	LS	Other Language	2,132	2.58	28.00	27.67	41.74
12	LS	Students with Disabilities	621	6.28	23.99	23.03	46.70

H.13b: Grade 12 Reading and Writing

Grade	Test	Group	N- count	Proficiency Levels			
				1	2	3	4
12	RW	Female	2,856	3.40	53.82	24.79	18.00
12	RW	Male	2,824	4.60	50.21	24.33	20.86
12	RW	American Indian or Alaska Native	6	0.00	83.33	0.00	16.67
12	RW	Black or African American	556	3.24	56.83	23.74	16.19
12	RW	Asian	1,542	4.22	50.65	25.03	20.10
12	RW	Native Hawaiian/Other Pacific Islander	3	0.00	0.00	33.33	66.67
12	RW	Hispanic or Latino	3,025	3.37	53.36	24.56	18.71
12	RW	White	537	7.82	43.76	23.84	24.58
12	RW	Multiracial (not Hispanic origin)	11	0.00	36.36	45.45	18.18
12	RW	NYC	4,084	3.65	53.09	24.00	19.27
12	RW	Big 4 Cities	220	5.45	62.27	18.64	13.64
12	RW	High Need Urban/Suburban	508	6.30	54.13	23.82	15.75
12	RW	High Need Rural	28	0.00	42.86	25.00	32.14
12	RW	Average Need	512	3.32	46.29	26.95	23.44
12	RW	Low Need	292	3.08	38.36	34.25	24.32
12	RW	Charter Schools	9	33.33	44.44	11.11	11.11
12	RW	0 Years English language learner	3,116	4.01	52.28	24.39	19.32
12	RW	1 Years English language learner	592	4.39	52.20	20.44	22.97
12	RW	2 Years English language learner	800	4.38	53.88	24.50	17.25
12	RW	3 Years English language learner	312	3.21	49.68	28.21	18.91
12	RW	4 Years English language learner	354	3.11	49.15	27.12	20.62
12	RW	5 Years English language learner	174	2.87	55.17	25.29	16.67
12	RW	6 Years English language learner or More	316	4.43	48.10	27.85	19.62
12	RW	Arabic	124	3.23	55.65	21.77	19.35
12	RW	Bengali	135	4.44	59.26	20.74	15.56
12	RW	Chinese	296	5.07	53.72	26.35	14.86
12	RW	English	159	2.52	44.03	28.30	25.16
12	RW	Spanish	2,852	3.44	53.93	24.12	18.51
12	RW	Other Language	2,114	4.73	49.15	25.02	21.10
12	RW	Students with Disabilities	612	11.44	55.23	19.28	14.05

H.13c: Grade 12 Overall

Grade	Test	Group	N-count	Proficiency Levels			
				1	2	3	4
12	Overall	Female	2,842	4.57	56.83	25.30	13.30
12	Overall	Male	2,813	5.08	53.64	24.99	16.28
12	Overall	American Indian or Alaska Native	6	0.00	83.33	0.00	16.67
12	Overall	Black or African American	553	3.62	59.67	23.87	12.84
12	Overall	Asian	1,540	5.65	55.58	26.62	12.14
12	Overall	Native Hawaiian/Other Pacific Islander	3	0.00	0.00	33.33	66.67
12	Overall	Hispanic or Latino	3,019	4.24	55.45	24.74	15.57
12	Overall	White	523	7.27	48.37	24.67	19.69
12	Overall	Multiracial (not Hispanic origin)	11	0.00	54.55	27.27	18.18
12	Overall	NYC	4,072	4.72	56.41	24.31	14.56
12	Overall	Big 4 Cities	216	6.94	64.35	18.06	10.65
12	Overall	High Need Urban/Suburban	506	6.52	58.89	23.72	10.87
12	Overall	High Need Rural	27	0.00	40.74	33.33	25.93
12	Overall	Average Need	506	2.57	49.01	28.46	19.96
12	Overall	Low Need	292	3.42	40.75	37.67	18.15
12	Overall	Charter Schools	9	33.33	44.44	11.11	11.11
12	Overall	0 Years English language learner	3,104	4.77	55.19	24.84	15.21
12	Overall	1 Years English language learner	591	6.77	55.50	22.34	15.40
12	Overall	2 Years English language learner	799	5.76	59.07	23.90	11.26
12	Overall	3 Years English language learner	311	3.22	52.73	29.90	14.15
12	Overall	4 Years English language learner	353	3.40	51.84	28.05	16.71
12	Overall	5 Years English language learner	174	2.87	55.75	28.16	13.22
12	Overall	6 Years English language learner or More	307	3.26	51.79	27.36	17.59
12	Overall	Arabic	124	3.23	56.45	24.19	16.13
12	Overall	Bengali	135	5.19	62.22	19.26	13.33
12	Overall	Chinese	294	7.82	60.88	24.83	6.46
12	Overall	English	157	2.55	47.13	29.30	21.02
12	Overall	Spanish	2,847	4.36	55.78	24.48	15.38
12	Overall	Other Language	2,098	5.29	53.81	26.22	14.68
12	Overall	Students with Disabilities	595	11.76	58.82	18.15	11.26

APPENDIX I: Exit Rate by Subgroup**I.1: Kindergarten**

Grade	Group	N-count	Percent
K	Female	905	7.78
K	Male	758	6.33
K	American Indian or Alaska Native	4	12.90
K	Black or African American	44	4.88
K	Asian	579	11.98
K	Native Hawaiian/Other Pacific Islander	3	20.00
K	Hispanic or Latino	787	5.30
K	White	232	8.25
K	Multiracial (not Hispanic origin)	14	8.43
K	NYC	1,243	8.74
K	Big 4 Cities	30	2.89
K	High Need Urban/Suburban	124	3.71
K	High Need Rural	7	3.72
K	Average Need	143	4.94
K	Low Need	78	6.07
K	Charter Schools	14	6.42
K	0 Years English language learner	1,172	8.79
K	1 Years English language learner	481	4.76
K	Arabic	32	6.69
K	Bengali	84	13.04
K	Chinese	144	13.01
K	English	95	6.31
K	Spanish	747	5.48
K	Other Language	561	8.99
K	Students with Disabilities	63	1.99

I.2: Grade 1

Grade	Group	N-count	Percent
1	Female	2,767	21.32
1	Male	2,381	16.91
1	American Indian or Alaska Native	12	26.67
1	Black or African American	172	13.49
1	Asian	1,467	29.39
1	Native Hawaiian/Other Pacific Islander	4	21.05
1	Hispanic or Latino	2,752	16.02
1	White	732	20.99
1	Multiracial (not Hispanic origin)	9	14.06
1	NYC	3,285	19.38
1	Big 4 Cities	173	13.34
1	High Need Urban/Suburban	533	13.57
1	High Need Rural	42	24.14
1	Average Need	699	23.99
1	Low Need	355	26.03
1	Charter Schools	36	18.46
1	0 Years English language learner	2,790	19.08
1	1 Years English language learner	567	14.43
1	2 Years English language learner	1,783	21.09
1	Arabic	126	21.18
1	Bengali	207	29.28
1	Chinese	339	30.85
1	English	297	21.35
1	Spanish	2,490	15.57
1	Other Language	1,689	23.19
1	Students with Disabilities	325	6.21

I.3: Grade 2

Grade	Group	N-count	Percent
2	Female	1,441	13.37
2	Male	1,434	12.27
2	American Indian or Alaska Native	1	3.70
2	Black or African American	81	8.06
2	Asian	845	22.20
2	Native Hawaiian/Other Pacific Islander	5	45.45
2	Hispanic or Latino	1,525	10.32
2	White	414	14.75
2	Multiracial (not Hispanic origin)	4	13.79
2	NYC	1,634	11.40
2	Big 4 Cities	108	9.52
2	High Need Urban/Suburban	385	11.90
2	High Need Rural	21	14.38
2	Average Need	414	18.29
2	Low Need	285	27.72
2	Charter Schools	22	15.38
2	0 Years English language learner	1,310	10.60
2	1 Years English language learner	241	13.16
2	2 Years English language learner	422	14.91
2	3 Years English language learner	899	16.64
2	Arabic	55	11.29
2	Bengali	91	15.69
2	Chinese	153	23.29
2	English	191	18.65
2	Spanish	1,370	9.88
2	Other Language	1,015	17.37
2	Students with Disabilities	198	4.11

I.4: Grade 3

Grade	Group	N-count	Percent
3	Female	1,747	18.29
3	Male	1,708	15.43
3	American Indian or Alaska Native	4	16.00
3	Black or African American	133	13.18
3	Asian	933	27.45
3	Native Hawaiian/Other Pacific Islander	1	14.29
3	Hispanic or Latino	1,947	14.31
3	White	432	16.90
3	Multiracial (not Hispanic origin)	5	17.24
3	NYC	2,106	15.91
3	Big 4 Cities	114	10.46
3	High Need Urban/Suburban	433	14.24
3	High Need Rural	21	15.11
3	Average Need	497	24.59
3	Low Need	254	29.10
3	Charter Schools	18	18.75
3	0 Years English language learner	1,841	15.45
3	1 Years English language learner	225	14.29
3	2 Years English language learner	281	19.45
3	3 Years English language learner	290	18.14
3	4 Years English language learner	813	20.00
3	Arabic	79	17.32
3	Bengali	143	30.49
3	Chinese	129	26.17
3	English	212	22.27
3	Spanish	1,769	13.81
3	Other Language	1,123	20.62
3	Students with Disabilities	288	5.64

I.5: Grade 4

Grade	Group	N-count	Percent
4	Female	1,687	21.65
4	Male	1,699	17.73
4	American Indian or Alaska Native	5	25.00
4	Black or African American	160	17.24
4	Asian	706	27.81
4	Native Hawaiian/Other Pacific Islander		
4	Hispanic or Latino	2,126	18.26
4	White	385	17.36
4	Multiracial (not Hispanic origin)	4	17.39
4	NYC	2,102	18.62
4	Big 4 Cities	164	16.02
4	High Need Urban/Suburban	429	16.37
4	High Need Rural	31	25.83
4	Average Need	408	27.33
4	Low Need	215	33.97
4	Charter Schools	23	24.21
4	0 Years English language learner	1,864	18.27
4	1 Years English language learner	222	14.64
4	2 Years English language learner	221	19.86
4	3 Years English language learner	172	21.83
4	4 Years English language learner	229	20.45
4	5 Years English language learner	664	25.62
4	Arabic	69	17.92
4	Bengali	99	30.65
4	Chinese	95	27.46
4	English	189	25.44
4	Spanish	1,964	17.94
4	Other Language	970	20.94
4	Students with Disabilities	398	7.82

I.6: Grade 5

Grade	Group	N-count	Percent
5	Female	1,649	24.54
5	Male	1,840	22.88
5	American Indian or Alaska Native	7	35.00
5	Black or African American	171	19.75
5	Asian	632	31.00
5	Native Hawaiian/Other Pacific Islander	1	14.29
5	Hispanic or Latino	2,310	23.37
5	White	366	18.97
5	Multiracial (not Hispanic origin)	2	11.76
5	NYC	2,375	24.13
5	Big 4 Cities	116	13.93
5	High Need Urban/Suburban	424	19.92
5	High Need Rural	25	26.32
5	Average Need	336	28.60
5	Low Need	159	33.90
5	Charter Schools	32	28.07
5	0 Years English language learner	2,138	24.22
5	1 Years English language learner	211	14.69
5	2 Years English language learner	279	25.50
5	3 Years English language learner	149	29.50
5	4 Years English language learner	111	15.81
5	5 Years English language learner	131	21.13
5	6 Years English language learner or More	460	29.83
5	Arabic	75	21.43
5	Bengali	81	31.89
5	Chinese	90	36.59
5	English	148	29.48
5	Spanish	2,187	23.28
5	Other Language	908	22.61
5	Students with Disabilities	563	11.78

I.7: Grade 6

Grade	Group	N-count	Percent
6	Female	1,013	18.10
6	Male	1,159	16.92
6	American Indian or Alaska Native	4	22.22
6	Black or African American	114	14.65
6	Asian	450	23.62
6	Native Hawaiian/Other Pacific Islander	3	60.00
6	Hispanic or Latino	1,311	16.17
6	White	290	17.84
6	Multiracial (not Hispanic origin)		
6	NYC	1,462	17.56
6	Big 4 Cities	109	14.73
6	High Need Urban/Suburban	248	14.31
6	High Need Rural	22	27.50
6	Average Need	192	20.36
6	Low Need	106	27.25
6	Charter Schools	19	25.00
6	0 Years English language learner	1,279	17.64
6	1 Years English language learner	162	11.17
6	2 Years English language learner	175	16.68
6	3 Years English language learner	90	20.74
6	4 Years English language learner	76	13.15
6	5 Years English language learner	60	18.40
6	6 Years English language learner or More	324	24.49
6	Arabic	50	16.03
6	Bengali	55	27.92
6	Chinese	58	22.22
6	English	86	23.24
6	Spanish	1,245	16.14
6	Other Language	678	18.89
6	Students with Disabilities	397	9.94

I.8: Grade 7

Grade	Group	N-count	Percent
7	Female	1,000	18.92
7	Male	873	14.05
7	American Indian or Alaska Native	2	20.00
7	Black or African American	128	16.02
7	Asian	381	19.82
7	Native Hawaiian/Other Pacific Islander		
7	Hispanic or Latino	1,086	14.75
7	White	274	19.71
7	Multiracial (not Hispanic origin)	2	18.18
7	NYC	1,193	15.34
7	Big 4 Cities	86	12.16
7	High Need Urban/Suburban	261	16.56
7	High Need Rural	13	17.81
7	Average Need	188	22.14
7	Low Need	114	28.22
7	Charter Schools	12	24.49
7	0 Years English language learner	1,054	16.03
7	1 Years English language learner	138	9.27
7	2 Years English language learner	173	15.93
7	3 Years English language learner	111	24.34
7	4 Years English language learner	83	16.24
7	5 Years English language learner	68	23.69
7	6 Years English language learner or More	242	22.70
7	Arabic	40	14.60
7	Bengali	33	16.75
7	Chinese	49	17.44
7	English	82	24.19
7	Spanish	1,002	14.26
7	Other Language	667	19.75
7	Students with Disabilities	323	9.77

I.9: Grade 8

Grade	Group	N-count	Percent
8	Female	1,377	25.95
8	Male	1,230	19.92
8	American Indian or Alaska Native	3	23.08
8	Black or African American	162	21.23
8	Asian	484	24.66
8	Native Hawaiian/Other Pacific Islander	2	50.00
8	Hispanic or Latino	1,580	21.16
8	White	375	29.83
8	Multiracial (not Hispanic origin)	1	6.67
8	NYC	1,791	22.34
8	Big 4 Cities	97	15.18
8	High Need Urban/Suburban	308	21.17
8	High Need Rural	23	28.05
8	Average Need	236	29.65
8	Low Need	119	31.15
8	Charter Schools	20	43.48
8	0 Years English language learner	1,555	23.30
8	1 Years English language learner	192	12.26
8	2 Years English language learner	240	20.39
8	3 Years English language learner	120	27.84
8	4 Years English language learner	127	24.80
8	5 Years English language learner	65	31.10
8	6 Years English language learner or More	303	34.08
8	Arabic	56	19.93
8	Bengali	63	30.29
8	Chinese	60	20.20
8	English	110	33.95
8	Spanish	1,481	20.70
8	Other Language	837	26.01
8	Students with Disabilities	426	15.38

I.10: Grade 9

Grade	Group	N-count	Percent
9	Female	886	15.30
9	Male	1,010	15.60
9	American Indian or Alaska Native	3	13.64
9	Black or African American	139	14.52
9	Asian	342	14.04
9	Native Hawaiian/Other Pacific Islander		
9	Hispanic or Latino	1,120	14.80
9	White	289	23.38
9	Multiracial (not Hispanic origin)	3	7.69
9	NYC	1,241	15.49
9	Big 4 Cities	97	12.68
9	High Need Urban/Suburban	213	14.32
9	High Need Rural	18	24.66
9	Average Need	186	17.97
9	Low Need	106	22.84
9	Charter Schools	10	24.39
9	0 Years English language learner	1,099	16.41
9	1 Years English language learner	175	7.87
9	2 Years English language learner	173	12.42
9	3 Years English language learner	86	18.82
9	4 Years English language learner	101	21.04
9	5 Years English language learner	75	23.73
9	6 Years English language learner or More	181	27.72
9	Arabic	34	10.24
9	Bengali	34	14.47
9	Chinese	45	13.68
9	English	63	15.75
9	Spanish	1,060	14.74
9	Other Language	660	17.48
9	Students with Disabilities	274	14.23

I.11: Grade 10

Grade	Group	N-count	Percent
10	Female	787	13.74
10	Male	857	13.91
10	American Indian or Alaska Native	1	6.25
10	Black or African American	133	12.64
10	Asian	309	10.80
10	Native Hawaiian/Other Pacific Islander	3	23.08
10	Hispanic or Latino	950	14.05
10	White	244	21.20
10	Multiracial (not Hispanic origin)	4	10.81
10	NYC	1,196	13.65
10	Big 4 Cities	40	9.24
10	High Need Urban/Suburban	156	13.19
10	High Need Rural	11	18.97
10	Average Need	145	18.08
10	Low Need	75	17.52
10	Charter Schools	7	26.92
10	0 Years English language learner	1,004	15.26
10	1 Years English language learner	180	8.32
10	2 Years English language learner	176	11.33
10	3 Years English language learner	46	11.44
10	4 Years English language learner	65	16.05
10	5 Years English language learner	50	18.18
10	6 Years English language learner or More	120	24.95
10	Arabic	30	10.60
10	Bengali	26	9.22
10	Chinese	48	11.16
10	English	57	19.86
10	Spanish	894	13.80
10	Other Language	589	14.26
10	Students with Disabilities	189	13.30

I.12: Grade 11

Grade	Group	N-count	Percent
11	Female	482	11.52
11	Male	550	14.19
11	American Indian or Alaska Native	1	6.67
11	Black or African American	80	11.14
11	Asian	206	10.58
11	Native Hawaiian/Other Pacific Islander	1	50.00
11	Hispanic or Latino	570	12.80
11	White	168	18.88
11	Multiracial (not Hispanic origin)	6	17.65
11	NYC	701	12.47
11	Big 4 Cities	26	9.22
11	High Need Urban/Suburban	83	9.19
11	High Need Rural	9	17.31
11	Average Need	123	17.83
11	Low Need	77	16.78
11	Charter Schools	6	35.29
11	0 Years English language learner	553	14.00
11	1 Years English language learner	128	9.75
11	2 Years English language learner	151	11.83
11	3 Years English language learner	67	12.93
11	4 Years English language learner	37	9.11
11	5 Years English language learner	42	15.85
11	6 Years English language learner or More	52	16.51
11	Arabic	22	12.22
11	Bengali	12	6.32
11	Chinese	32	10.85
11	English	36	14.40
11	Spanish	529	12.62
11	Other Language	401	13.58
11	Students with Disabilities	107	11.84

I.13: Grade 12

Grade	Group	N-count	Percent
12	Female	378	13.30
12	Male	458	16.28
12	American Indian or Alaska Native	1	16.67
12	Black or African American	71	12.84
12	Asian	187	12.14
12	Native Hawaiian/Other Pacific Islander	2	66.67
12	Hispanic or Latino	470	15.57
12	White	103	19.69
12	Multiracial (not Hispanic origin)	2	18.18
12	NYC	593	14.56
12	Big 4 Cities	23	10.65
12	High Need Urban/Suburban	55	10.87
12	High Need Rural	7	25.93
12	Average Need	101	19.96
12	Low Need	53	18.15
12	Charter Schools	1	11.11
12	0 Years English language learner	472	15.21
12	1 Years English language learner	91	15.40
12	2 Years English language learner	90	11.26
12	3 Years English language learner	44	14.15
12	4 Years English language learner	59	16.71
12	5 Years English language learner	23	13.22
12	6 Years English language learner or More	54	17.59
12	Arabic	20	16.13
12	Bengali	18	13.33
12	Chinese	19	6.46
12	English	33	21.02
12	Spanish	438	15.38
12	Other Language	308	14.68
12	Students with Disabilities	67	11.26

APPENDIX J: DIF CATEGORIES FOR OPERATIONAL ITEMS**J. 1 Results of DIF Analyses**

Grade Span	Modality	Item Seq.	DIF–	DIF–	DIF–	DIF–
		Number	Male/Female	Asian/Other	Hispanic/Other	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
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

J.1 Results of DIF Analyses (Continued)

Grade Span	Modality	Item Seq.	DIF–	DIF–	DIF–	DIF–
		Number	Male/Female	Asian/Other	Hispanic/Other	White/Other
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
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*. Orland, 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. (2004). *Liseral 8.7 for windows/computer software*. Lincolnwood, IL: Scientific Software International, Inc.
- Kolen, M. J. & Brennan, R. L. (1995). *Test equating methods and practices*. New York: Springer-Verlag.
- Kolen, M. J. & Brennan R. L. (2004). *Test equating, scaling, and linking: methods and practices* (2nd ed.). New York: Springer-Verlag.
- Linacre, J. M. & 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.
- 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. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- 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 testlets scored in more than two categories. In D. Thissen and H. Wainer (Eds.), *Test scoring*. Mahwah, NJ: Lawrence Erlbaum Associates. 217–250.
- Taherbhai, H. & Seo, D. (2007). Comparing concurrent vs. fixed parameter equating with common items, Using the Rasch dichotomous and the partial credit models in a mixed-item format test. *Journal of Applied Measurement* 8(1).
- Thissen, D. & Steinberg, L. (1983). *A response model for multiple choice items. Psychometric technical report 1*. Chicago: National Opinion Research Center.
- Young, M. J. & Yoon, B. (1998, April). Estimating the consistency and accuracy of classifications in a standards-referenced assessment. *CSE technical report 475*. Los Angeles: University of California, Center for the Study of Evaluation, Standards, and Student Test

