

TECHNICAL REPORT
PART I – SUMMATIVE ASSESSMENT

**Ohio English Language Proficiency Assessment –
Listening, Reading, Speaking, and Writing**

Grades K–12

Spring 2019 Administration

Submitted to:

Ohio Department of Education

Submitted by:

American Institutes for Research
1000 Thomas Jefferson Street, NW
Washington, DC 20007

December 2019

TABLE OF CONTENTS

Chapter 1. Introduction.....	1
Chapter 2. Test Administration.....	3
2.1 Testing Window.....	3
2.2 Summative Tests.....	3
2.3 Test Administration Manual.....	4
2.3.1 Directions for Administration	4
2.3.2 Training/Practice Tests.....	4
2.3.3 Summative Tests	5
2.4 Test Security	5
Chapter 3. Scoring.....	6
3.1 Estimating Student Ability for Summative Assessments.....	6
3.2 Additional Scoring Rules for Summative Assessments	6
3.3 Theta to Scale Score Transformation.....	6
3.4 Lowest/Highest Obtainable Scores	7
3.5 Handscoring.....	7
Chapter 4. Standard Setting.....	8
Chapter 5. 2019 Summary.....	10
5.1 2019 Student Participation	10
5.2 2019 Student Scale Score and Performance Summary	15
5.3 2019 Testing Time	21
Chapter 6. Reliability.....	23
6.1 Internal Consistency	25
6.2 Marginal Reliability and Conditional Standard Error of Measurement	27
6.3 Marginal Standard Error of Measurement.....	29
6.4 Classification Accuracy and Consistency.....	31
Chapter 7. Validity.....	39
7.1 Dimensionality Analysis.....	41
7.2 Student Abilities vs. Test Difficulties.....	41
7.3 Inter-Rater Analysis.....	41
7.4 Relationship of Test Scores to External Variables.....	44
7.5 Student Progress.....	46
Chapter 8. Reporting	50
8.1 Online Reporting System.....	50
8.1.1 Types of Online Score Reports	50

8.1.2 Subgroup Reports 53

8.2 Paper Reports 54

Chapter 9. Quality Control57

9.1 Quality Control in Test Configuration..... 57

9.1.1 Platform Review 57

9.1.2 User Acceptance Testing and Final Review 58

9.2 Quality Assurance in Scoring..... 58

9.2.1 Quality Assurance in Online Data..... 59

9.2.2 Quality Assurance in Handscoring..... 59

9.2.3 Handscoring Quality Assurance Monitoring Reports 61

9.2.4 Quality Control on Final Scores 63

9.3 Quality Assurance in Reporting..... 63

9.3.1 Online Report Quality Assurance..... 63

9.3.2 Paper Report Quality Assurance 63

References65

List of Tables

Table 2.1 Number of Items and Score Points—Online Summative	3
Table 2.2 Number of Items and Score Points—Paper Summative	3
Table 2.3 Number of Items and Score Points—Braille Summative	4
Table 3.1 Scaling Constants on the Reporting Metric	7
Table 4.1 ELPA21 Domain Cut Scores by Grade	8
Table 5.1 Student Participation by Test Mode.....	12
Table 5.2 Student Participation by Subgroups in Each Test.....	13
Table 5.3 Scale Score Summary	16
Table 5.4 Percentage of Students by Performance Level—Listening and Reading	17
Table 5.5 Percentage of Students by Performance Level—Speaking and Writing	17
Table 5.6 Percentage of Students by Overall Proficiency Level	19
Table 5.7 2019 Testing Time	22
Table 6.1 Cronbach’s Alpha	26
Table 6.2 Marginal Reliability	28
Table 6.3 Overall Classification Accuracy and Consistency	34
Table 6.4 Classification Accuracy for Each Cut.....	35
Table 6.5 Classification Consistency for Each Cut	37
Table 7.1 Overall Inter-Rater Analysis	43
Table 7.2 Correlations by Domain, Overall, and Comprehension Scores	45
Table 7.3 Correlations by Domain Performance Levels and Overall Proficiency Levels	45
Table 8.1 Types of Online Score Reports by Level of Aggregation	50
Table 8.2 Types of Subgroups	53

List of Figures

Figure 5.1 Percentage of Students in Performance Levels	20
Figure 5.2 Percentage of Students in Proficiency Levels	21
Figure 6.1 Cronbach’s Alpha by Domain	27
Figure 6.2 Marginal Reliability by Test.....	29
Figure 6.3 Ratio of Standard Error of Measurement Over Standard Deviation by Test	30
Figure 6.4 Domain Classification Accuracy	36
Figure 6.5 Domain Classification Consistency.....	38
Figure 7.1 Average Domain Scale Score Progress	46
Figure 7.2 Average Overall and Comprehension Scale Score Progress	47
Figure 7.3 Percentage of Domain Performance Level from Screener to Summative.....	48
Figure 7.4 Percentages of Overall Proficiency from Screener to Summative	49
Figure 8.1 ORS Page for Online Reporting System User Guide	52
Figure 8.2. Mock-Up for Family Report.....	55

Chapter 1. Introduction

The English Language Proficiency Assessment for the 21st Century (ELPA21) is a testing program in which multiple states participate to support educators, member states, and members of the public as they adopt and implement the English Language Proficiency (ELP) Standards and college- and career-readiness standards. The ELPA21 Program, called the Program in this report, provides an assessment system that measures growth based on the 2014 ELP Standards and provides valuable information that informs instruction and facilitates academic English proficiency so that all English language learners (ELLs) leave high school prepared for college and career success. The assessment includes tests on listening, reading, speaking, and writing for students in kindergarten, grade 1, grades 2–3, grades 4–5, grades 6–8, and grades 9–12. Ohio is a member of the ELPA21 consortium, and the assessment is called the Ohio English Language Proficiency Assessment (OELPA).

The Program conducted test development and item development for the summative EL assessment, as part of a U.S. Department of Education grant, commencing in 2013 and running through the first operational administration of the assessment in 2016. As part of the development process, Questar Assessment Inc., built multiple fixed-length forms for each test. Items were field tested in spring 2015 with the Center for Research on Evaluation, Standards, and Student Testing (CRESST)—the Program’s psychometrics partner—conducting analysis of item performance and data reviews. The first operational administration was implemented in spring 2016, and again, CRESST further analyzed items, conducted item data review meetings, and finalized item parameters. Pacific Metrics, the organization contracted for standard setting, held a standard-setting workshop in July 2016. Based on recommendations from the workshop, the Program made decisions with respect to domain cut scores that further translated into performance levels for each grade. After the 2017 administration, CRESST calibrated the items field tested in 2017. Based on the finalized item parameters for scoring provided by CRESST and the cut scores decisions from the Program, the American Institutes for Research (AIR) scored and reported the testing results.

Details about test development, item development, field-test form building, item data review, item calibration, and standard setting can be found in their respective reports provided by the Program or the respective supporting vendors.

In 2017, along with the summative assessment, the Program developed the ELPA21 screener. The purpose of the screener is to identify students who require English language development instruction. It is an assessment of a student’s language proficiency in the required domains of listening, reading, writing, and speaking. The test questions are based on the same ELP Standards adopted by the ELPA21 member states. Ohio first administered the screener assessment in the 2018-2019 school year.

This technical report focuses on the 2018-2019 test administration, test form reliability and validity, scoring, reporting, and quality control. This technical report has two parts.

Part I includes chapters that delineate different aspects of the 2019 administration of the summative assessment, including:

- Chapter 1. Introduction

- Chapter 2. Test Administration
- Chapter 3. Scoring
- Chapter 4. Standard Setting
- Chapter 5. 2019 Summary
- Chapter 6. Reliability
- Chapter 7. Validity
- Chapter 8. Reporting
- Chapter 9. Quality Control
- Appendix Part I: Analysis for Summative Assessment—2019 Summary

Part II includes chapters that delineate different aspects of the 2018-2019 administration of the screener assessment, including:

- Chapter 1. Introduction
- Chapter 2. Test Administration
- Chapter 3. 2019 Summary
- Chapter 4. Reliability
- Chapter 5. Validity
- Chapter 6. Reporting
- Appendix Part II: Analysis for Screener Assessment—2019 Summary

Chapter 2. Test Administration

The 2018-2019 OELPA assessments included summative tests and screener tests. The screener test is used to identify students who require English language development instruction. Each of the summative and screener tests were administered to students in kindergarten, grade 1, grades 2–3, grades 4–5, grades 6–8, and grades 9–12. The tests for kindergarten and grade 1 were grade-specific. The tests for grades 2 and above were grade-band tests. None of the tests are timed.

Each form of both assessments involves four domain tests. Students could be exempted from as many as three domain tests.

2.1 TESTING WINDOW

The 2018–2019 testing windows for OELPA was from August 1, 2018 to July 12, 2019 for the screener assessment and from February 4, 2019 to March 29, 2019 for the summative assessment.

2.2 SUMMATIVE TESTS

The 2019 summative assessment includes one online form, one paper form, and one braille form. Each form had separate tests for the four language domains. There were no field-test items in the 2019 summative tests.

Table 2.1-Table 2.3 list the number of operational items and score points in each online, paper-pencil, and braille form. The tables show that listening and reading have comparable numbers of items in each test. Writing and speaking have fewer but comparable numbers of items in each test.

Table 2.1 Number of Items and Score Points—Online Summative

	Grade/Grade Band											
	K		1		2–3		4–5		6–8		9–12	
Domain	Items	Score Points	Items	Score Points	Items	Score Points	Items	Score Points	Items	Score Points	Items	Score Points
Listening	29	29	25	25	25	27	29	32	33	35	23	24
Reading	23	23	29	29	28	33	27	30	29	32	36	38
Speaking	11	27	9	25	9	25	8	30	7	27	7	27
Writing	14	21	14	21	16	32	18	48	13	45	12	42
Writing Supplement	5	12	4	11								

Table 2.2 Number of Items and Score Points—Paper Summative

	Grade/Grade Band											
	K		1		2–3		4–5		6–8		9–12	
Domain	Items	Score Points	Items	Score Points	Items	Score Points	Items	Score Points	Items	Score Points	Items	Score Points
Listening	28	28	22	22	23	24	24	27	30	31	20	21
Reading	23	23	29	29	26	28	26	28	28	32	35	38
Speaking	11	27	9	25	9	25	8	30	7	27	7	27

Writing 6 6 5 5 6 6 5 9 2 6 2 6

Table 2.3 Number of Items and Score Points—Braille Summative

Domain	Grade/Grade Band											
	K		1		2–3		4–5		6–8		9–12	
	Items	Score Points	Items	Score Points	Items	Score Points	Items	Score Points	Items	Score Points	Items	Score Points
Listening	17	19	21	21	20	20	22	24	22	23	18	19
Reading	13	13	22	22	22	24	23	23	25	29	34	37
Speaking	4	12	7	17	8	20	7	25	6	22	5	19
Writing	10	21	7	19	9	19	10	27	8	28	8	28

2.3 TEST ADMINISTRATION MANUAL

2.3.1 Directions for Administration

For the 2018-2019 administrations, the Test Administration Manual (TAM) was developed to guide test administrators (TAs) in test administrations for the summative test. The TAM usually includes the following key points:

- Overview of the OELPA summative assessment
- Test administrator qualifications
- Preliminary planning
- Materials required
- Administrative considerations
- Student preparation/guidance for practice tests
- Detailed instructions for preparing and administering the training tests and summative tests
- Test security instructions
- Contact information for user support

2.3.2 Training/Practice Tests

To help TAs and students familiarize themselves with the online registration and test delivery systems, training or practice tests were provided before and during the testing windows. Training/practice tests can be accessed through a non-secure or secure browser.

The summative training tests have two components, one for TAs to create and manage the training/practice test sessions and the other for students to take an actual training/practice test.

The *Practice Test Administration* site introduces TAs to

- logging in;
- starting a test session;
- providing the session ID to the students signing in to the TA session;
- monitoring students’ progress throughout their tests; and

- stopping the test.

The *Practice Tests* site introduces students to

- signing in;
- verifying student information;
- selecting a test;
- waiting for the TA to check the test settings and approve participation;
- starting the test (adjusting the audio sound, checking the microphone for recording speaking responses, and reviewing test instructions);
- taking the test; and
- submitting the test.

2.3.3 Summative Tests

The instructions for summative tests include a brief direction for each domain test. They also provide the detailed instructions for

- logging in to the secure browser;
- starting a test session;
- providing the session ID to the students;
- approving student test sessions, including reviewing and editing students' test settings and accommodations;
- monitoring students' progress throughout their tests by checking their testing statuses; and
- stopping the session and logging out.

2.4 TEST SECURITY

The test security procedures are consistent across all assessments. Details about student-level testing confidentiality, system security, testing environment security, and documentation of test security violations can be found in Chapter 9. Quality Control.

Chapter 3. Scoring

For summative tests, four domain scores and two composite scores are computed. The composite scores include a comprehension score for listening and reading and an overall score that comprises all four domains.

3.1 ESTIMATING STUDENT ABILITY FOR SUMMATIVE ASSESSMENTS

ELPA21 reported scale scores for each domain test, the overall scores for the whole test that includes four domains, and the comprehension scores for the partial test that includes reading and listening domains. Multidimensional item response theory (MIRT) is used to estimate domain scores. The bi-factor model is used to estimate the overall and comprehension scores. The details of the estimations can be found in the *ELPA21 Scoring Specifications*.

In addition, business rules were established to specify rules about domain exemptions and attemptedness at the item, domain, and test levels. The additional scoring rules for the 2019 OELPA summative tests are outlined in the following section.

3.2 ADDITIONAL SCORING RULES FOR SUMMATIVE ASSESSMENTS

Additional rules applied to the 2019 OELPA scoring include:

1. If a domain test was attempted (even one item), the items without responses in that domain should be scored as 0.
2. If a domain test was not attempted (no item responses in the domain) and the student was not marked as exempt in that domain, set the domain score and performance level to N.
3. If any domain tests are exempted before a student takes his or her first test, they should be left out of the computation of the overall scores. If the exempted domain test is reading or listening, the test should be left out of the computation of the comprehension score. In this case, set the score and performance level to E. However, if a test in an exempt domain is started in AIR's test delivery system (TDS), we will assume the test was not exempt.
4. If one of the reading and listening domains is exempted and the other one is not attempted, set the comprehension score to E.
5. If one of the domains is exempted and the other three domains are not attempted or exempted, set the overall score to N.
6. Rules for Scoring K–grade 1 writing:
 - For Ohio, if either the online or paper supplement is omitted, set the items in the omitted section to 0 and score the writing domain accordingly.
 - a. If the form ID for the omitted writing domain is known, set the scores of items in that form to 0.
 - b. If the form ID for the omitted writing domain is unknown, set the scores of items in Form A to 0.

3.3 THETA TO SCALE SCORE TRANSFORMATION

The student's performance is summarized in an individual domain score for each domain, a comprehension score that includes listening and reading, and an overall score that includes all four

domains. Each theta score is linearly transformed to a scale score using the formula $SS = a * \theta + b$, where a is the slope and b is the intercept. As decided in the 2016 standard-setting meeting, there is one set of scaling constants for the domain scores and another set of constants for the composite scores, as shown in Table 3.1. Scale scores are rounded to an integer.

Table 3.1 Scaling Constants on the Reporting Metric

Subject	Grade	Slope (a)	Intercept (b)
Domain Scores (listening, reading, speaking, and writing)	K–12	80	550
Comprehension Scores	K–12	600	5500
Overall Scores	K–12	600	5500

3.4 LOWEST/HIGHEST OBTAINABLE SCORES

There are no minimum or maximum scale scores for any tests in the 2018-2019 administration.

3.5 HANDSCORING

For OELPA summative tests, all items in speaking and some items in writing are handscored. Measurement Incorporated (MI) provides all handscoring except for screeners administered in Ohio, which are scored locally. The procedure for handscoring items is provided by the ELPA21 Program. The scoring rubrics and item content are reviewed by content experts as a part of the item review meetings. A key facet of reliability is whether scoring rules are applied with fidelity during scoring sessions.

Chapter 4. Standard Setting

For the summative assessment, the domain cut scores and the overall proficiency levels were set through a standard setting meeting convened by the ELPA21 Program on July 19–22, 2016. Details about the standard setting can be found in the ELPA21 standard-setting report.

Five performance levels were established for each domain. The cut scores were set by grade, as listed in Table 4.1. The four cut scores set for each domain identify students with five performance levels, Levels 1–5. If a student scores below the first cut, the student is classified in performance Level 1. If a student scores at or above the first cut but below the second cut, the student is classified in performance Level 2; this pattern continues for performance Levels 3 and 4. If a student scores at or above the fourth cut, the student is classified in performance Level 5.

Table 4.1 ELPA21 Domain Cut Scores by Grade

Grade	Domain	Cut 1	Cut 2	Cut 3	Cut 4	Grade	Domain	Cut 1	Cut 2	Cut 3	Cut 4
K	Listening	467	507	613	645	5	Listening	413	455	498	581
	Reading	473	514	592	627		Reading	468	511	588	627
	Speaking	487	535	598	625		Speaking	483	526	573	607
	Writing	497	562	651	673		Writing	438	486	598	628
1	Listening	435	467	549	594	6	Listening	410	440	498	565
	Reading	479	515	584	629		Reading	461	496	565	604
	Speaking	528	577	593	619		Speaking	465	511	562	595
	Writing	498	548	613	641		Writing	425	472	564	594
2	Listening	408	438	512	564	7	Listening	430	473	553	597
	Reading	457	489	555	595		Reading	486	534	609	642
	Speaking	490	529	555	588		Speaking	475	527	582	611
	Writing	452	493	555	591		Writing	474	520	597	625
3	Listening	409	448	536	598	8	Listening	432	478	565	613
	Reading	495	541	610	644		Reading	494	547	640	669
	Speaking	500	538	572	612		Speaking	476	528	590	619
	Writing	498	542	603	636		Writing	484	533	619	647
4	Listening	398	431	492	563	9–12	Listening	451	491	571	613
	Reading	453	488	550	594		Reading	488	539	631	662
	Speaking	462	506	544	584		Speaking	481	536	593	619
	Writing	437	481	568	600		Writing	485	533	615	641

Overall proficiency, defined as “proficiency determination,” for a given student is established on the basis of a profile of domain performance levels across all four tested domains. There are three proficiency determination levels: Emerging, Progressing, and Proficient. The following rules determine a student’s overall proficiency:

- Students whose domain performance levels are 1 or 2 across all nonexempt domains are

identified as Emerging.

- Students whose domain performance levels are 4 or 5 across all nonexempt domains are identified as Proficient. Students cannot be proficient if any domains are untested in the absence of a valid exemption or invalidated
- Students with domain performance levels that do not fit with Emerging or Proficient are identified as Progressing.

Chapter 5. 2019 Summary

The 2019 student participation, accommodation, and performance statistics for OELPA summative assessment are presented in this chapter and in Sections 1–4 of the appendix. The figures and tables included in Sections 1–4 are listed below:

- Section 1. Summative Assessment—Accommodation Summary
 - Table S1.1 shows the types of the accommodations, the sub-type (Value) within each type of accommodation, the number of students accommodated in each subtype by grade, and the total number of students in each subtype of accommodation.
- Section 2. Summative Assessment—Scale Score Summary
 - Tables S2.1–S2.13 present the number of students taking each domain test, their average scores, and standard deviation of scores and by subgroups in each grade (K–12). At the end of each table, the total number of students who have valid overall and comprehension scores, their average scores, and the standard deviations of their scores are also presented.
- Section 3. Summative Assessment—Percentage of Students by Domain Performance Level
 - Tables S3.1–S3.13 show the total number of students taking each domain test and the percentage of students in each performance level by domain test and by subgroups.
- Section 4. Summative Assessment—Percentage of Students by Overall Proficiency Level
 - Tables S4.1–S4.13 show the total number of students who are categorized in each of the overall proficiency levels: Emerging, Progressing, and Proficient, or not assigned an overall proficiency level (N) by subgroups.

5.1 2019 STUDENT PARTICIPATION

Table 5.1 summarizes student participation in each mode of testing. Three students tested with braille forms, 132 students tested with paper-pencil forms, and all other students tested with online forms.

Table 5.2 shows student participation overall and by subgroups. It shows the number of students in each subgroup who took a particular test. Generally, the number of students who participate in the 2019 OELPA administration decreases as the grade level increases, with some fluctuation at grades 9 and 10. More male students than female students take the test. By test, most students are in the Hispanic or Latino group (26.5%–35.2%), followed by African American students (18.6%–24%) and Asian students (15.6%–24.1%). About 0.1%–0.7% of the students have migrated to the United States from other countries. The percentage of students who have individualized education

programs (IEPs) increased from 5.1% to 18.4% between kindergarten and middle school, and it dropped to 14% in high school. Between kindergarten and middle school, the percentages of students with disabilities increased from 5.1% to 18.2% and dropped to 14.1% in high school. Most students identified as having a speech or language impairment are in kindergarten (2.2%) and grade 1 (2%); while more students identified as having a specific learning disability are in upper grades: 4.6% in grade band 2–3, 10.3% in grade band 4–5, 11.1% in middle school, and 8.8% in high school.

In addition, Section 1 of the Appendix Part I shows the number of students who used various available accommodations. The descriptions of accommodations can be found in the [ELPA21 Accessibility and Accommodations Manual](#).

Table 5.1 Student Participation by Test Mode

Grade	Total	Braille		Online		Paper	
		N	%	N	%	N	%
K	9170			9170	100.00		
1	8036			8036	100.00		
2	6904			6870	99.51	34	0.49
3	5716			5693	99.60	23	0.4
4	4692			4668	99.49	24	0.51
5	3687			3673	99.62	14	0.38
6	3432			3424	99.77	8	0.23
7	3412	1	0.03	3404	99.77	7	0.21
8	3372			3359	99.61	13	0.39
9	4335	1	0.02	4331	99.91	3	0.07
10	3505	1	0.03	3501	99.89	3	0.09
11	2785			2783	99.93	2	0.07
12	2077			2076	99.95	1	0.05
Total	61123	3	0	60988	99.78	132	0.22

Table 5.2 Student Participation by Subgroups in Each Test

Subgroup	Status	K		1		2-3		4-5		6-8		9-12	
		N	%	N	%	N	%	N	%	N	%	N	%
Total		9170	100.0	8036	100.0	12620	100.0	8379	100.0	10216	100.0	12702	100.0
Grade	K	9170	100.0										
	1			8036	100.0								
	2					6904	54.7						
	3					5716	45.3						
	4							4692	56.0				
	5							3687	44.0				
	6									3432	33.6		
	7									3412	33.4		
	8									3372	33.0		
	9											4335	34.1
	10											3505	27.6
	11											2785	21.9
Gender	12											2077	16.4
	Female	4202	45.8	3719	46.3	5802	46.0	3720	44.4	4411	43.2	5667	44.6
	Male	4525	49.3	4105	51.1	6680	52.9	4516	53.9	5655	55.4	6962	54.8
	Missing	443	4.8	212	2.6	138	1.1	143	1.7	150	1.5	73	0.6
Ethnicity	African American	1705	18.6	1529	19.0	2491	19.7	1619	19.3	2265	22.2	3047	24.0
	American Indian or Alaskan Native	27	0.3	20	0.2	24	0.2	18	0.2	14	0.1	33	0.3
	Asian	2206	24.1	1816	22.6	2354	18.7	1323	15.8	1593	15.6	2345	18.5
	Hispanic or Latino	2427	26.5	2267	28.2	4073	32.3	2948	35.2	3440	33.7	3957	31.2
	Multiple Races	1005	11.0	894	11.1	1687	13.4	1219	14.5	1522	14.9	1751	13.8
	Native Hawaiian or Other Pacific Islander	40	0.4	29	0.4	69	0.5	46	0.5	57	0.6	71	0.6

Subgroup	Status	K		1		2-3		4-5		6-8		9-12	
		N	%	N	%	N	%	N	%	N	%	N	%
Other	Other/Unknown	41	0.4	25	0.3	24	0.2	14	0.2	23	0.2	37	0.3
	White	1420	15.5	1300	16.2	1898	15.0	1192	14.2	1302	12.7	1461	11.5
	Missing	299	3.3	156	1.9								
	Gifted	14	0.2	38	0.5	74	0.6	24	0.3	30	0.3	18	0.1
	IEP	468	5.1	497	6.2	1357	10.8	1490	17.8	1881	18.4	1783	14.0
	Migrant	10	0.1	15	0.2	63	0.5	29	0.3	34	0.3	95	0.7
Primary Disability	Autism	97	1.1	83	1.0	166	1.3	106	1.3	90	0.9	73	0.6
	Deaf-blindness					2	0.0			2	0.0		
	Developmental delay	6	0.1	4	0.0	11	0.1	10	0.1	18	0.2	25	0.2
	Emotional disturbance	7	0.1	4	0.0	30	0.2	32	0.4	33	0.3	41	0.3
	Hearing Impaired	9	0.1	4	0.0	20	0.2	10	0.1	33	0.3	22	0.2
	Intellectual Disability	22	0.2	38	0.5	81	0.6	116	1.4	157	1.5	169	1.3
	Multiple disabilities	32	0.3	31	0.4	58	0.5	57	0.7	76	0.7	90	0.7
	Orthopedic impairment	8	0.1	15	0.2	12	0.1	11	0.1	11	0.1	8	0.1
	Other health impairment	37	0.4	52	0.6	151	1.2	180	2.1	248	2.4	192	1.5
	Specific learning disability	33	0.4	73	0.9	580	4.6	864	10.3	1136	11.1	1122	8.8
	Speech or language impairment	206	2.2	185	2.3	238	1.9	83	1.0	48	0.5	20	0.2
	Traumatic brain injury	1	0.0	3	0.0	6	0.0	11	0.1	15	0.1	11	0.1
	Visual impairment	6	0.1	3	0.0	3	0.0	2	0.0	6	0.1	8	0.1

5.2 2019 STUDENT SCALE SCORE AND PERFORMANCE SUMMARY

Student performance in the 2019 OELPA administration is summarized for the entire population and by subgroup. Table 5.3 shows the number of students, mean and standard deviation of scale scores in each grade and domain. The tests are not vertically linked across all grades. Scale scores can be compared only within grade-band tests (2–3, 4–5, 6–8, and 9–12). In general, the mean scale scores increase as grades increase within each grand band. A disaggregated summary based on gender, ethnicity, and other subgroups is also available in Section 2 of the Appendix Part I.

Table 5.4 and Table 5.5 display the percentage of students in each performance level in each grade and domain (see Section 3 of the Appendix Part I for disaggregated summary). In addition, Table 5.6 shows the percentage of student in each overall proficiency level in each grade. The percentage of students by performance level in each domain test and by overall proficiency level are also plotted in Figure 5.1 and 5.2, respectively. Section 4 of the Appendix Part I displays the percentage of students by proficiency level by gender, ethnicity, and other subgroups.

For both reading and writing, Figure 5.1 shows that most students are in performance level 3 except for writing kindergarten and reading grade 9 tests. In listening and writing, the percentages of students in performance level 4 and performance level 5 increase between kindergarten and grade 2 or grade 3, remain stable up to grade 6, then drop up to grade 9, and slightly increase again up to grade 11. In reading and speaking, the percentage of students in the same performance levels (4 and 5) follows the same pattern but remains highest in grades 3 or 4 only.

Figure 5.2 shows that most students are in the Progressing category in all grades (63.4% to 77.7%). The percentage of students who are proficient increases from kindergarten to grade 4 (with slight drop in grade 3), then consistently decreases until grade 9, and slightly increases after grade 9. The percentage of students in the Emerging category is relatively stable until grade 6, increases until grade 9, and then consistently drops afterwards.

Table 5.3 Scale Score Summary

Grade	Listening			Reading			Speaking			Writing		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
K	9144	556	78.2	9077	558	74.3	9025	578	76.3	9068	544	76.8
1	8013	554	69.4	7948	550	75.0	7935	577	66.3	7947	541	72.7
2	6884	540	65.6	6815	526	67.2	6790	545	66.0	6808	523	70.9
3	5704	570	69.9	5638	566	71.6	5628	574	69.9	5636	562	72.3
4	4669	528	73.6	4591	530	69.8	4603	548	72.9	4598	524	74.3
5	3670	540	79.0	3607	542	75.6	3621	554	76.5	3615	536	79.4
6	3415	527	67.4	3372	524	63.0	3368	551	68.9	3367	519	71.7
7	3399	538	74.9	3360	540	70.1	3330	558	74.0	3357	533	79.0
8	3342	553	81.2	3314	557	76.4	3276	569	77.4	3306	549	85.2
9	4290	520	76.2	4242	522	71.2	4162	548	70.9	4234	518	80.4
10	3475	539	76.5	3450	540	73.9	3402	563	65.8	3443	537	76.3
11	2751	551	74.0	2733	550	73.6	2723	573	60.4	2733	547	71.8
12	2058	553	67.8	2040	551	68.1	2022	575	55.7	2031	549	63.5

* Scale scores cannot be compared across grade bands

Table 5.4 Percentage of Students by Performance Level—Listening and Reading

Grade	Count	Listening								Reading							
		1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
K	9170	13.0	14.3	48.3	10.3	13.9	0.3	0.0	9170	13.5	14.4	38.3	14.3	18.6	0.6	0.4	
1	8036	6.2	5.4	30.6	25.8	31.6	0.3	0.0	8036	17.8	17.9	30.6	15.1	17.6	0.5	0.6	
2	6904	4.6	3.1	20.1	33.8	38.1	0.2	0.1	6904	15.0	15.9	33.6	17.4	16.7	0.6	0.7	
3	5716	3.7	3.5	17.4	35.4	39.8	0.2	0.0	5716	18.0	14.9	35.5	17.1	13.2	0.6	0.8	
4	4692	6.9	4.6	14.5	37.9	35.6	0.4	0.1	4692	14.9	11.0	31.6	22.4	17.9	1.0	1.2	
5	3687	8.7	7.0	9.8	39.3	34.8	0.4	0.1	3687	18.4	12.7	36.8	17.7	12.2	1.1	1.1	
6	3432	6.2	6.3	17.9	37.9	31.1	0.3	0.2	3432	16.7	15.5	40.1	15.5	10.5	0.7	1.0	
7	3412	11.0	9.4	29.8	25.9	23.6	0.3	0.1	3412	23.8	20.6	36.9	10.4	6.9	0.6	0.9	
8	3372	11.0	9.0	27.8	25.3	26.1	0.6	0.3	3372	22.4	19.0	42.5	8.6	5.8	0.9	0.9	
9	4335	21.3	14.3	36.2	15.8	11.3	0.9	0.1	4335	35.1	24.0	31.0	4.6	3.1	1.2	0.9	
10	3505	14.2	13.8	35.1	17.7	18.3	0.7	0.2	3505	28.3	23.1	33.6	7.2	6.2	0.7	0.9	
11	2785	10.0	13.4	33.6	19.0	22.8	1.2	0.0	2785	23.4	23.1	35.7	8.4	7.6	1.1	0.8	
12	2077	7.1	11.0	40.3	20.1	20.6	0.7	0.2	2077	19.4	26.5	37.9	8.4	6.1	0.8	1.0	

* N = Not Attempted; E = Exempted.

Table 5.5 Percentage of Students by Performance Level—Speaking and Writing

Grade	Count	Speaking								Writing							
		1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
K	9170	11.5	10.3	32.4	17.0	27.3	1.1	0.5	9170	30.5	29.3	28.9	4.5	5.6	0.7	0.4	
1	8036	17.0	26.9	11.4	17.6	25.8	0.8	0.5	8036	25.6	23.7	33.5	8.5	7.5	0.5	0.6	
2	6904	16.6	17.5	16.6	22.0	25.6	1.0	0.7	6904	16.4	14.2	33.1	18.0	16.9	0.7	0.7	
3	5716	11.8	10.9	17.4	28.4	29.9	0.9	0.7	5716	17.5	14.6	35.5	18.0	13.0	0.7	0.7	
4	4692	11.0	9.1	16.6	27.7	33.7	1.1	0.8	4692	13.6	8.8	47.1	16.7	11.7	0.9	1.1	
5	3687	14.6	11.1	24.4	24.3	23.7	1.0	0.8	3687	13.0	9.0	55.2	12.5	8.4	1.0	0.9	
6	3432	11.2	8.9	26.8	25.8	25.4	1.1	0.7	3432	11.9	8.8	50.5	15.9	11.0	0.8	1.1	
7	3412	12.6	11.9	31.4	19.0	22.7	1.7	0.7	3412	19.3	14.2	45.8	10.7	8.4	0.7	0.9	

Grade	Count	Speaking								Writing							
		1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
8	3372	11.8	9.2	32.2	19.2	24.9	2.1	0.7	3372	19.6	14.1	45.3	9.8	9.2	0.9	1.0	
9	4335	17.4	16.4	34.5	15.1	12.5	3.5	0.5	4335	29.6	20.2	38.9	5.6	3.4	1.5	0.8	
10	3505	11.7	16.0	33.6	17.1	18.6	2.3	0.7	3505	22.2	19.9	41.4	8.0	6.6	0.9	0.9	
11	2785	7.6	15.7	33.0	19.5	22.0	1.7	0.6	2785	17.8	20.5	42.4	9.8	7.6	1.1	0.7	
12	2077	5.6	13.9	38.1	19.0	20.7	1.8	0.8	2077	13.3	23.4	46.1	9.5	5.5	1.2	1.0	

* N = Not Attempted; E = Exempted.

Table 5.6 Percentage of Students by Overall Proficiency Level

Grade	Number Tested	Emerging	Progressing	Proficient	N*
K	9170	14.5	77.7	7.8	0.1
1	8036	9.4	76.6	13.9	0.1
2	6904	7.3	66.7	25.9	0.1
3	5716	7.0	69.0	23.9	0.0
4	4692	10.9	63.4	25.5	0.2
5	3687	14.3	68.4	17.2	0.1
6	3432	11.5	70.4	18.1	0.0
7	3412	18.0	70.9	11.0	0.0
8	3372	17.3	72.0	10.6	0.1
9	4335	30.7	64.3	5.0	0.0
10	3505	23.4	67.5	9.1	0.0
11	2785	18.5	70.7	10.7	0.1
12	2077	14.2	76.8	8.9	0.1

* N = Not Assigned.

Figure 5.1 Percentage of Students in Performance Levels

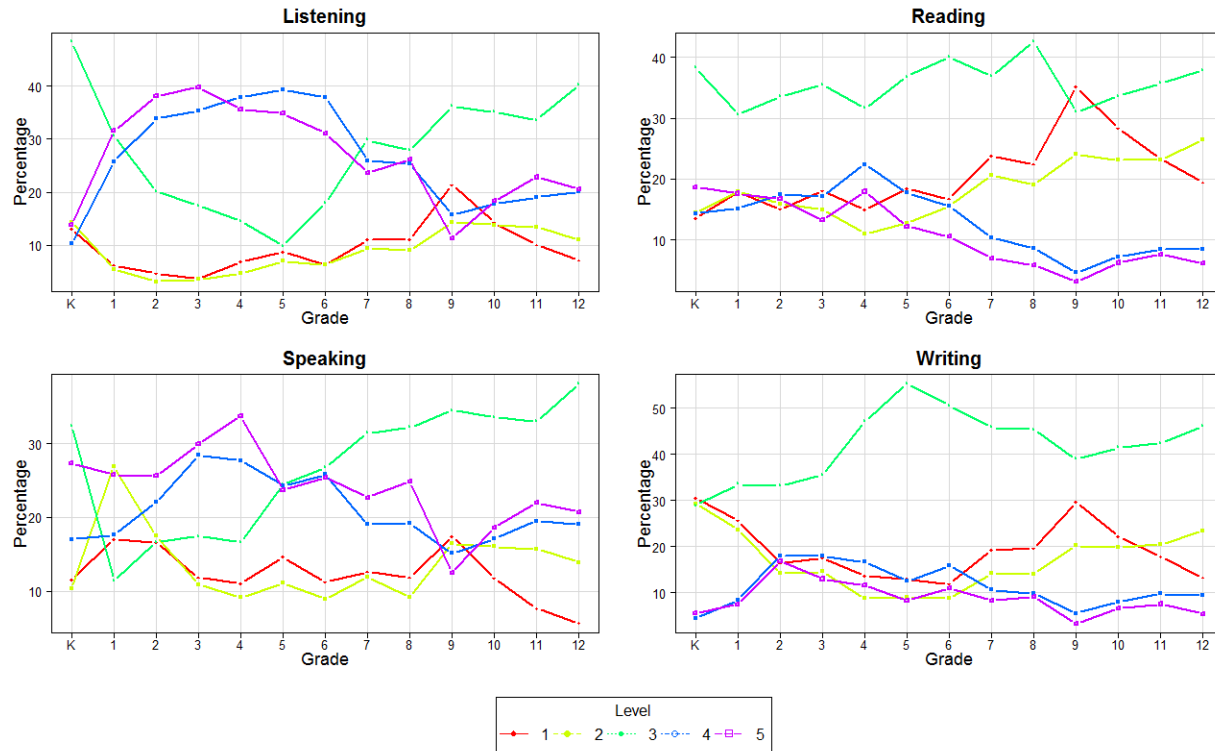
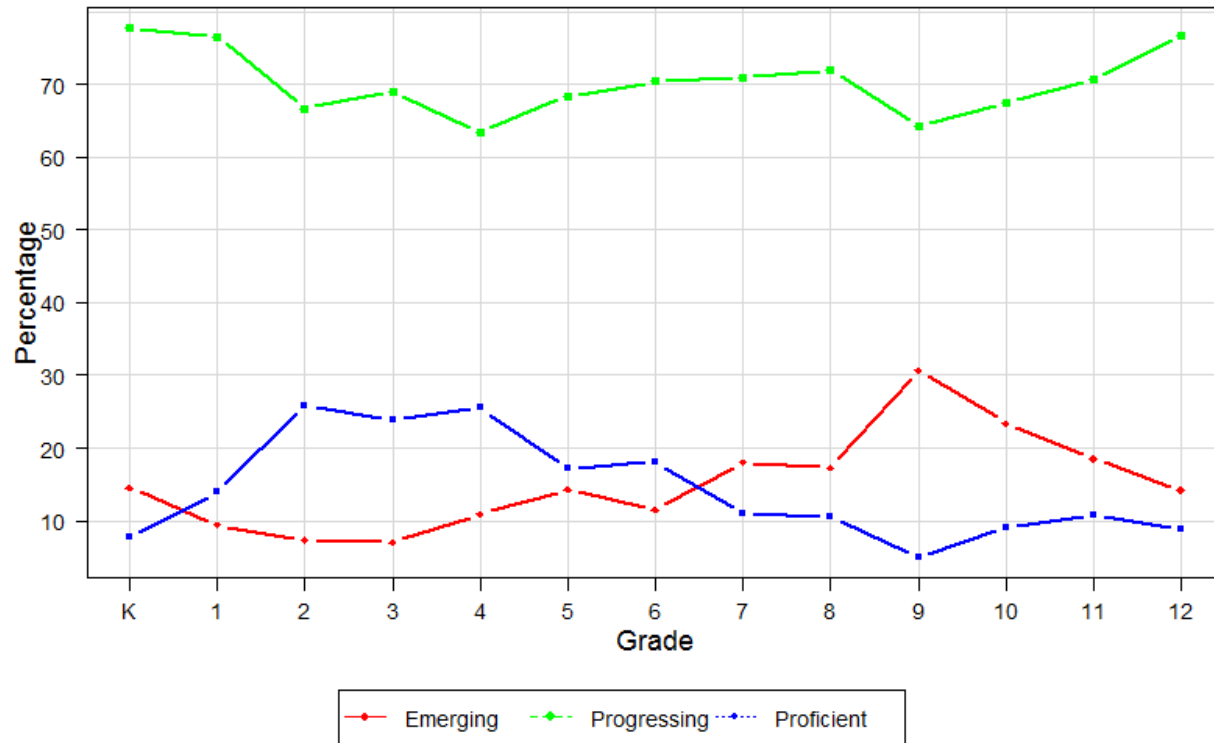


Figure 5.2 Percentage of Students in Proficiency Levels



5.3 2019 TESTING TIME

Table 5.7 shows testing time per each grade or grade band. In general, higher grade tests showed longer testing time than the lower grade.

Table 5.7 2019 Testing Time

Grade/Grade Band	N Students	N of Items			Testing Time (Minutes)							
		Min	Mean	Max	Mean	Min	P25	P50	P75	P80	P90	Max
K	8666	78	91	91	54.37	9.69	42.31	50.98	62.72	66.24	76.08	186.80
1	7626	76	80	80	56.93	19.97	46.08	53.86	63.98	67.16	76.71	185.56
2-3	12286	72	78	78	86.46	11.81	65.87	80.49	99.31	105.57	124.09	385.91
4-5	8090	63	80	80	97.34	9.73	73.74	89.70	112.78	120.34	142.88	381.62
6-8	9855	72	77	77	138.45	21.00	94.76	122.88	164.33	177.23	219.84	671.81
9-12	12142	54	75	75	149.31	7.84	99.52	136.40	184.31	199.43	242.53	759.76

Chapter 6. Reliability

Reliability refers to consistency in test scores. Reliability can be defined as the degree to which individuals' deviation scores remain relatively consistent over repeated administrations of the same test or alternate test forms (Crocker & Algina, 1986). For example, if a person takes the same or parallel tests repeatedly, he or she should receive consistent results. The reliability coefficient refers to the ratio of true score variance to observed score variance:

$$\rho_{XX'} = \frac{\sigma_T^2}{\sigma_X^2}$$

There are various approaches for estimating the reliability of scores. The conventional approaches used are characterized as follows:

- The *test-retest* method measures stability over time. With this method, the same test is administered twice to the same group at two different points in time. If test scores from the two administrations are highly correlated, then the test scores are deemed to have a high level of stability. For example, if the result is highly stable, those who scored high on the first administration tend to obtain a high score on the second administration. The critical factor, however, is the time interval. The time interval should not be too long, which could allow for changes in the test takers' true scores. Likewise, it should not be too short, in which case memory and practice may confound the results. The test-retest method is most effective for measuring constructs that are stable over time, such as intelligence or personality traits.
- The *parallel-forms* method is used for measuring equivalence. With this design, two parallel forms of the test are administered to the same group. This method requires two similar forms of a test. However, it is very difficult to create two strictly parallel forms. When this method is applied, the effects of memory or practice can be eliminated or reduced, since the tests are not purely identical as with the test-retest method. The reliability coefficient from this method indicates the degree to which the two tests are measuring the same construct. While there are a wide variety of possible items to administer to measure any particular construct, it is only feasible to administer a sample of items on any given test. If there is a high correlation between the scores of the two tests, then inferences regarding high reliability of scores can be substantiated. This method is commonly used to estimate the reliability of achievement or aptitude tests.
- The *split-half* method utilizes one test divided into two halves within a single test administration. It is crucial to make the two half-tests as parallel as possible, as the correlation between the two half-tests is used to estimate reliability of the whole test. In general, this method produces a coefficient that underestimates the reliability for the full test. To correct the estimate, the Spearman-Brown prophecy formula (Brown, 1910; Spearman, 1910) can be applied. While this method is convenient, varying splits of the items may yield different reliability estimates.
- The *internal consistency* method can be employed when it is not possible to conduct repeated test administrations. Whereas other methods often compute the correlation between two separate tests, this method considers each item within a test to be a one-item

test. There are several other statistical methods based on this idea: Coefficient alpha (Cronbach & Chavelson, 2004), Kuder-Richardson Formula 20 (Kuder & Richardson, 1937), Kuder-Richardson Formula 21 (Kuder & Richardson, 1937), stratified coefficient alpha (Qualls, 1995), and Feldt-Raju coefficient (Feldt & Qualls, 1996; Feldt & Brennan, 1989).

- *Inter-rater reliability* is the extent to which two or more individuals (coders or raters) agree. Inter-rater reliability addresses the consistency of the implementation of a rating system.

Another way to view reliability is to consider its relationship with the standard errors of measurement (SEM)—the smaller the standard error, the higher the precision of the test scores. For example, classical test theory assumes that an observed score (X) of each individual can be expressed as a true score (T) plus some error (E), $X = T + E$. The variance of X can be shown to be the sum of two orthogonal variance components:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2.$$

Returning to the definition of reliability as the ratio of true score variance to observed score variance, we can arrive at:

$$\rho_{XX'} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_X^2 - \sigma_E^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}.$$

As the fraction of error variance to observed score variance approaches 0, the reliability then approaches 1. The Classical Test Theory (CTT) standard error of measurement (SEM), which assumes a homogeneous error, is derived from the classical notion expressed earlier as $\sigma_X \sqrt{1 - \rho_{XX'}}$, where σ_X is the standard deviation of the scaled score and $\rho_{XX'}$ is a reliability coefficient. Based on the definition of reliability, this formula can be derived:

$$\rho_{XX'} = 1 - \frac{\sigma_E^2}{\sigma_X^2},$$

$$\frac{\sigma_E^2}{\sigma_X^2} = 1 - \rho_{XX'},$$

$$\sigma_E^2 = \sigma_X^2(1 - \rho_{XX'}),$$

$$\sigma_E = \sigma_X \sqrt{(1 - \rho_{XX'})}.$$

In general, the SEM is relatively constant across samples as the group dependent term, σ_X , can be shown to cancel out:

$$\sigma_E = \sigma_X \sqrt{(1 - \rho_{XX'})} = \sigma_X \sqrt{(1 - (1 - \frac{\sigma_E^2}{\sigma_X^2}))} = \sigma_X \sqrt{\frac{\sigma_E^2}{\sigma_X^2}} = \sigma_X \cdot \frac{\sigma_E}{\sigma_X} = \sigma_E.$$

This shows that the SEM in the classical test theory is assumed to be homogeneous error irrespective of the standard deviation of a group.

In contrast, the SEMs in IRT vary over the ability continuum. These heterogeneous errors are a function of a test information function (TIF) that provides different information about test takers depending on their estimated abilities. Often, the TIF is maximized over an important performance cut, such as the proficient cut score.

Because the TIF indicates the amount of information provided by the test at different points along the ability scale, its inverse indicates the “lack” of information at different points along the ability scale. This lack of information is the uncertainty, or the SEM, of the score at various score points. Conventionally, fixed-form tests are maximized near the middle of the score distribution, or near an important classification cut, and have less information at the tails of the score distribution.

In this chapter, test reliability for the OLEPA summative tests is provided using

- Cronbach’s alpha;
- marginal reliability based on all students in the state and by subgroup of students;
- marginal SEM;
- conditional SEM; and
- classification accuracy and consistency.

The following sections describes the methods used in the computation. The results for the 2019 OELPA summative assessment are provided in this chapter and Sections 5-6 of the Appendix Part I.

- Section 5. Summative Assessment—Marginal Reliability
 - Figure S5.1 presents the marginal reliability by gender for each domain test across grades.
 - Figure S5.2 presents the marginal reliability by ethnicity for each domain test across grades.
 - Figures S5.3 and S5.4 present the marginal reliability by IEP and primary disability for each domain test across grades, respectively
- Section 6. Summative Assessment—Conditional Standard Error of Measurement (CSEM)
 - Figures S6.1–S6.13 show the CSEM plots for each domain, overall, and comprehension tests.

6.1 INTERNAL CONSISTENCY

Cronbach’s alpha (Cronbach & Shavelson, 2004) is used to assess the internal consistency of items in each test for each domain. It is conceptually defined as “the degree to which measures are free

from error and therefore yield consistent results” (Peter, 1979, p.6). As such, it places a limit on the construct validity of a test (Peterson, 1994).

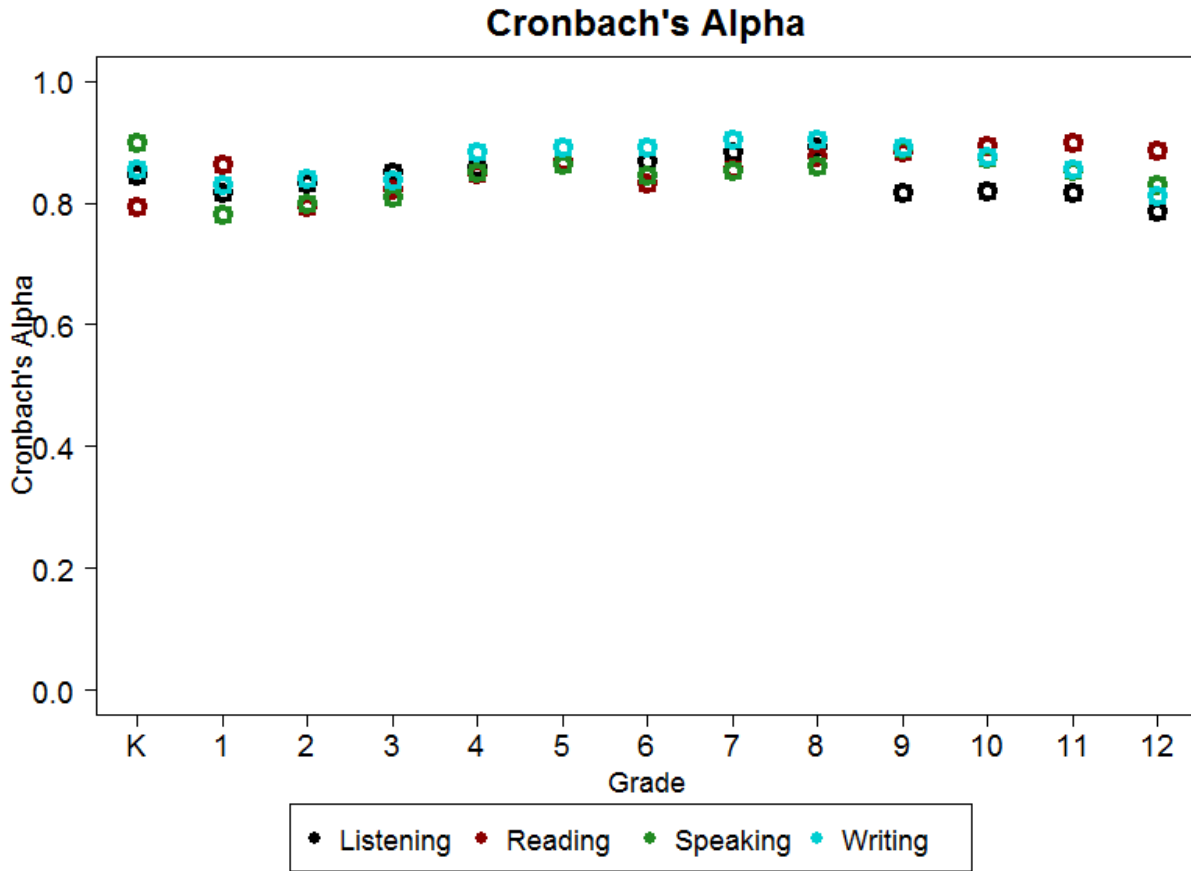
Tests with an *n* count at or above 50 are included in the analysis. By this criterion, the paper-pencil and braille forms are excluded from the analysis. Additionally, items with less than a 10% response rate are also excluded from the analysis.

Table 6.1 and Figure 6.1 show the result of analysis on Cronbach’s alpha for each domain test. The listening domain in grade 12 has the lowest alpha coefficient, 0.78. According to Nunnally (1978), 0.7 is the minimum acceptable alpha coefficient. All domain tests meet the minimum acceptable requirement.

Table 6.1 Cronbach’s Alpha

Grade	Cronbach’s Alpha			
	Listening	Reading	Speaking	Writing
K	0.84	0.79	0.90	0.86
1	0.82	0.86	0.78	0.83
2	0.83	0.79	0.80	0.84
3	0.85	0.82	0.81	0.84
4	0.86	0.85	0.85	0.88
5	0.87	0.87	0.86	0.89
6	0.87	0.83	0.84	0.89
7	0.88	0.85	0.85	0.90
8	0.89	0.88	0.86	0.91
9	0.82	0.88	0.89	0.89
10	0.82	0.89	0.87	0.88
11	0.82	0.90	0.85	0.85
12	0.78	0.89	0.83	0.81

Figure 6.1 Cronbach's Alpha by Domain



6.2 MARGINAL RELIABILITY AND CONDITIONAL STANDARD ERROR OF MEASUREMENT

Marginal reliability (Sireci, Thissen, & Wainer, 1991) assesses the precision of scoring. It is based on the average of the conditional standard errors of measurement (CSEM) for the estimated theta scores. By definition, it is the proportion of true score variance among the observed score variance.

Expected a posteriori (EAP) is used to estimate the domain scores. EAP is the estimate of true score, but its variance underestimates the true score variance, so the marginal reliability within domain can be estimated by

$$\bar{\rho} = \left(\frac{\sigma_{true}^2}{\sigma_{total}^2} \right) = 1 - \frac{\bar{\sigma}_{error}^2}{\sigma_{total}^2}$$

where $\bar{\sigma}_{error}^2$ is the average error variance (variance of the measurement error), $\sigma_{total}^2 = \sigma_{EAP}^2 + \bar{\sigma}_{error}^2$, σ_{EAP}^2 is the variance of the EAP estimate.

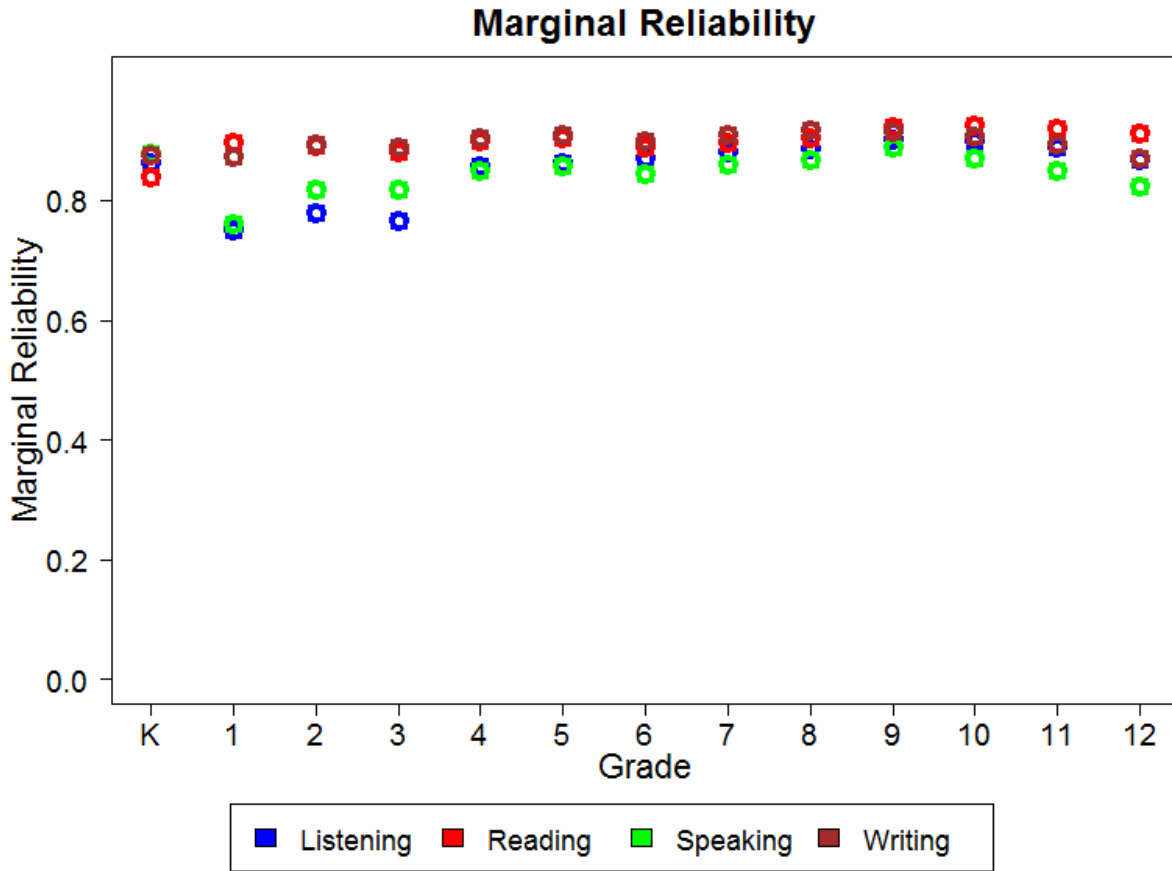
The maximum value for the marginal reliability is 1. A higher reliability coefficient indicates a greater precision of scoring. The marginal reliability for the 2019 OELPA summative tests is

presented in Table 6.2 and Figure 6.2. The results show that the listening tests at grades 1–5 have the lowest reliabilities, followed by the speaking tests. The reliability for the speaking domain in the middle and high school tests are lower than the other domains. All the reliability indices are above 0.8, except for listening tests in grades 1–3 and the speaking test in grade 1. In addition, Section 5 of the Appendix Part I presents marginal reliability by gender, ethnicity, IEP, and primary disability.

Table 6.2 Marginal Reliability

Grade	N	Marginal Reliability			
		Listening	Reading	Speaking	Writing
K	8980	0.86	0.84	0.88	0.88
1	7898	0.75	0.90	0.76	0.87
2	6754	0.78	0.89	0.82	0.89
3	5587	0.77	0.88	0.82	0.89
4	4558	0.86	0.90	0.85	0.90
5	3570	0.86	0.91	0.86	0.91
6	3332	0.87	0.89	0.85	0.90
7	3304	0.88	0.90	0.86	0.91
8	3242	0.89	0.90	0.87	0.92
9	4098	0.90	0.92	0.89	0.92
10	3362	0.90	0.92	0.87	0.91
11	2687	0.89	0.92	0.85	0.89
12	2002	0.87	0.91	0.82	0.87

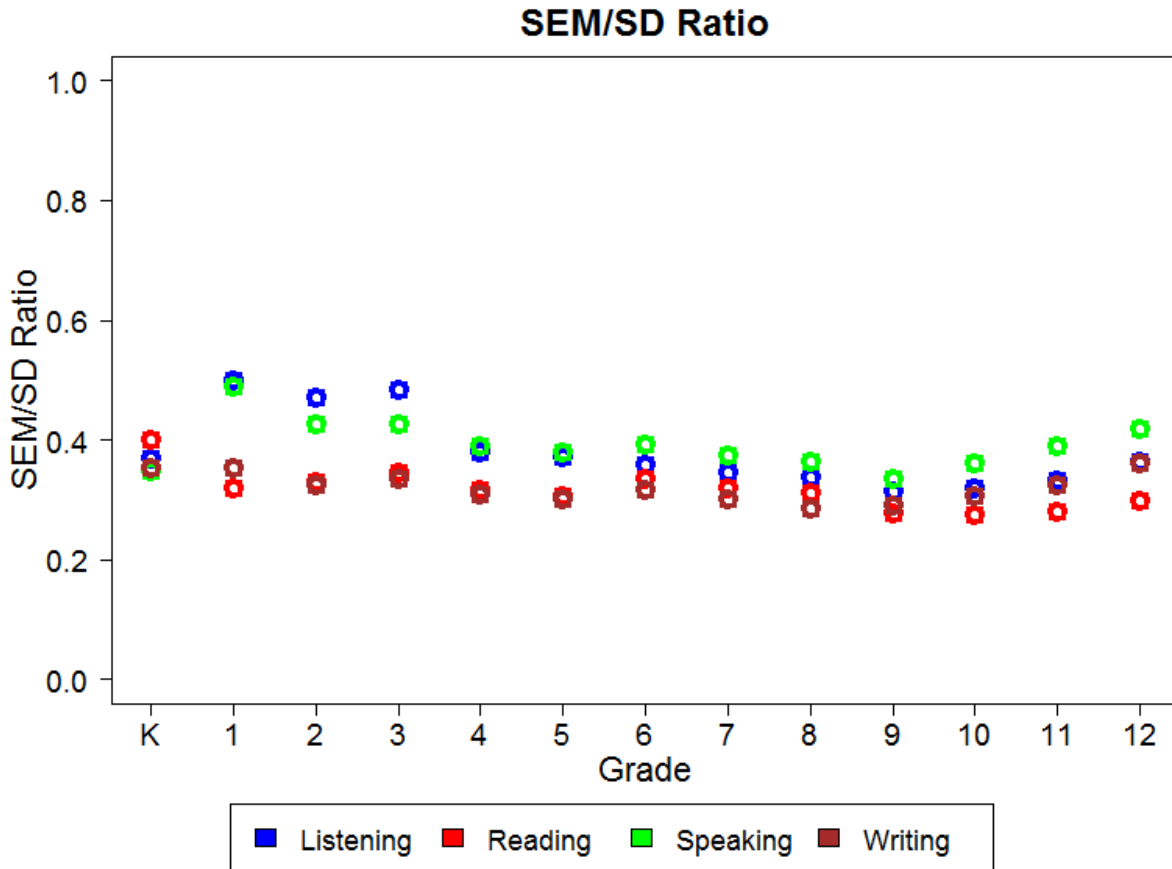
Figure 6.2 Marginal Reliability by Test



6.3 MARGINAL STANDARD ERROR OF MEASUREMENT

Another way to examine score reliability is with the marginal standard error of measurement (MSEM), computed as the square root of $\bar{\sigma}_{error}^2$. A smaller value of MSEM indicates a greater accuracy of test scores. The marginal reliability $\bar{\rho}$ and the test MSEM behave oppositely: The higher the $\bar{\rho}$, the lower the MSEM and vice versa. The ratio of MSEM and the standard deviation of scale scores can also indicate the measurement errors. The analysis for the ratio is displayed in Figure 6.3.

Figure 6.3 Ratio of Standard Error of Measurement Over Standard Deviation by Test



Within the item response theory (IRT) framework, measurement error varies across the range of ability as a result of the test information function (TIF). The TIF describes the amount of information provided by the test at each score point along the ability continuum. The inverse of the TIF is characterized as the conditional measurement error at each score point. For instance, if the measurement error is large, then less information is being provided by the assessment at the specific ability level.

The standard error for estimated student ability (theta score) is the square root of the reciprocal of the TIF:

$$se(\theta_i) = \frac{1}{\sqrt{TIF(\theta_i)}}$$

It is typically more useful to consider the inverse of the TIF rather than the TIF itself, as the standard errors are more useful for score interpretation. For this reason, standard error plots for each grade are presented in Section 6 of the Appendix Part I. In general, the standard error curves

follow the typical expected trends with more test information regarding scores observed near the middle of the score scale.

6.4 CLASSIFICATION ACCURACY AND CONSISTENCY

When student performance is reported in terms of achievement levels, a reliability of achievement classification is computed in terms of the probabilities of consistent classification of students as specified in Standard 2.16 in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). This index considers the consistency of classifications for the percentage of test takers who would, hypothetically, be classified in the same category on an alternate, equivalent form.

Classification accuracy (CA) analysis investigates how precisely students are classified into each performance level. The CA index evaluates the degree to which the fixed true score and the random observed score from the same measure would both result in the same performance level classification. Classification consistency (CC) analysis investigates how consistently students are classified into each performance level across two independent administrations of equivalent forms. Since obtaining test scores from two independent administrations is not feasible due to issues such as logistics and cost constraints, the CC index is computed with the assumption that the same test is independently administered twice to the same group of students.

For OELPA, since the overall proficiency is based on domain performance level, the CA and CC are examined at each cut score in each domain test. Five performance levels divided by four cut scores, cuts 1–4, are established for each domain test. Forms with an n count fewer than 50 are excluded from the analysis; for this reason, the paper-pencil and braille forms are excluded. Four domain cut scores divided the scale range into five performance levels.

In general, we can estimate the CA and CC using the following approach.

Let p_{il} be the probability of the true score at performance level l for the i th student, and define the following matrix based on L performance levels

$$\begin{pmatrix} n_{a11} & \cdots & n_{a1L} \\ \vdots & \vdots & \vdots \\ n_{aL1} & \cdots & n_{aLL} \end{pmatrix}$$

where $n_{alm} = \sum_{p_{li}=l} p_{im}$ which is the sum of the probabilities for each expected performance level at each observed performance level. In the matrix, the row represents the observed level and the column represents the expected level.

Based on the above matrix, the CA for the cut c_l ($l = 1, \dots, L - 1$) is:

$$CA_{c_l} = \frac{\sum_{k,m=1}^l n_{akm} + \sum_{k,m=l+1}^L n_{akm}}{N}$$

where N is the total number of students.

The CA for a single cut, for example, the CA at the cut 2 is the sum of the n_{alm} values in blue ($\sum_{k,m=1}^l n_{akm}$) assigned in the levels equal to or below cut 2 at both expected and observed levels and in green ($\sum_{k,m=l+1}^L n_{akm}$) assigned in the levels above cut 2 at both expected and observed levels divided by the total number of students.

$$\begin{pmatrix} n_{a11} & n_{a12} & n_{a13} & \cdots & n_{a1L} \\ n_{a21} & n_{a22} & n_{a23} & \cdots & n_{a2L} \\ n_{a31} & n_{a32} & n_{a33} & \cdots & n_{a3L} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ n_{a51} & n_{a52} & n_{a53} & \cdots & n_{a5L} \end{pmatrix}$$

For CC using p_{il} , similar to CA, a similar $L \times L$ table is constructed by assuming the test is administered twice independently to the same student group,

$$\begin{pmatrix} n_{c11} & \cdots & n_{c1L} \\ \vdots & \vdots & \vdots \\ n_{cL1} & \cdots & n_{cLL} \end{pmatrix}$$

where $n_{clm} = \sum_{i=1}^N p_{il} p_{im}$ which is the sum of the probabilities multiplied by each paired combination of performance.

The CC for the cut c_l ($l = 1, \dots, L - 1$) is:

$$CC_{c_l} = \frac{\sum_{k,m=1}^l n_{ckm} + \sum_{k,m=l+1}^L n_{ckm}}{N}.$$

Similarly, we can estimate CC for a single cut.

The CA and CC indices are affected by the interaction of the magnitude of $se(\theta)$, the distance between adjacent cuts, the location of the cuts on the ability scale, and the proportion of students around a cut point. The larger the $se(\theta)$, the closer the two adjacent cuts, and the greater the proportion of students around a cut point, the lower the indices.

At domain Level l , we can estimate p_{il} for student i as: Assuming the estimated ability of student i is $\hat{\theta}_i$ with SEM of $se(\hat{\theta}_i)$, and the estimated ability is distributed as $\hat{\theta}_i \sim N(\theta_i, se(\hat{\theta}_i))$, assuming a normal distribution, where θ_i is the unknown true ability of the i th student. The probability of the true score at performance Level l based on the cut scores c_{l-1} and c_l is estimated as

$$\begin{aligned} p_{il} &= p(c_{l-1} \leq \theta_i < c_l) = p\left(\frac{c_{l-1} - \hat{\theta}_i}{e(\hat{\theta}_i)} \leq \frac{\theta_i - \hat{\theta}_i}{e(\hat{\theta}_i)} < \frac{c_l - \hat{\theta}_i}{e(\hat{\theta}_i)}\right) \\ &= p\left(\frac{\hat{\theta}_i - c_l}{e(\hat{\theta}_i)} \leq \frac{\hat{\theta}_i - \theta_i}{e(\hat{\theta}_i)} < \frac{\hat{\theta}_i - c_{l-1}}{e(\hat{\theta}_i)}\right) = \Phi\left(\frac{\hat{\theta}_i - c_{l-1}}{e(\hat{\theta}_i)}\right) - \Phi\left(\frac{\hat{\theta}_i - c_l}{e(\hat{\theta}_i)}\right). \end{aligned}$$

For level 1, $c_0 = -\infty$, and for level L , $c_L = \infty$. If scaled score is to be used, the above formula can be used based on scored score distribution.

Table 6.3 shows overall classification accuracy and consistency. Classification consistency rates can be lower than classification accuracy because consistency is based on two tests with measurement errors, while accuracy is based on one test with a measurement error and the true score. The accuracy and consistency rates for each performance level are higher for the levels with smaller standard error.

The analysis results for each cut are presented in Table 6.4 and Figure 6.4 as well as Table 6.5 and Figure 6.5. For each cut, all CAs are above 0.8 and all CCs are above 0.75. In listening and speaking, both indices for cut 3 and/or cut 4 are relatively lower in elementary and middle school grades, which indicates lack of difficult items.

Table 6.3 Overall Classification Accuracy and Consistency

Grade	Listening		Reading		Speaking		Writing	
	Accuracy	Consistency	Accuracy	Consistency	Accuracy	Consistency	Accuracy	Consistency
K	0.71	0.63	0.67	0.57	0.68	0.58	0.74	0.65
1	0.63	0.53	0.71	0.61	0.53	0.46	0.69	0.59
2	0.68	0.58	0.69	0.59	0.57	0.48	0.69	0.59
3	0.68	0.57	0.68	0.58	0.58	0.49	0.67	0.57
4	0.74	0.64	0.72	0.62	0.62	0.52	0.73	0.65
5	0.73	0.64	0.72	0.63	0.60	0.51	0.77	0.70
6	0.74	0.65	0.71	0.62	0.61	0.52	0.74	0.66
7	0.70	0.61	0.73	0.64	0.63	0.53	0.73	0.64
8	0.71	0.62	0.75	0.67	0.64	0.55	0.74	0.66
9	0.72	0.63	0.80	0.73	0.67	0.57	0.76	0.68
10	0.72	0.62	0.78	0.70	0.66	0.56	0.73	0.64
11	0.71	0.61	0.77	0.69	0.64	0.54	0.72	0.63
12	0.70	0.60	0.76	0.67	0.64	0.53	0.70	0.61

Table 6.4 Classification Accuracy for Each Cut

Grade	Listening				Reading				Speaking				Writing			
	Cut1	Cut2	Cut3	Cut4	Cut1	Cut2	Cut3	Cut4	Cut1	Cut2	Cut3	Cut4	Cut1	Cut2	Cut3	Cut4
K	0.95	0.92	0.90	0.92	0.95	0.91	0.88	0.90	0.97	0.93	0.87	0.87	0.91	0.92	0.94	0.95
1	0.97	0.95	0.84	0.83	0.93	0.92	0.92	0.93	0.90	0.82	0.81	0.83	0.93	0.89	0.91	0.94
2	0.98	0.97	0.88	0.83	0.94	0.92	0.91	0.92	0.92	0.87	0.84	0.86	0.95	0.92	0.89	0.92
3	0.99	0.98	0.89	0.81	0.95	0.92	0.88	0.91	0.95	0.90	0.84	0.84	0.95	0.91	0.88	0.91
4	0.98	0.97	0.92	0.86	0.96	0.94	0.90	0.91	0.97	0.92	0.86	0.84	0.97	0.95	0.88	0.92
5	0.98	0.96	0.94	0.85	0.96	0.94	0.90	0.92	0.95	0.91	0.84	0.85	0.98	0.96	0.89	0.93
6	0.98	0.97	0.92	0.87	0.94	0.92	0.91	0.94	0.97	0.92	0.84	0.85	0.97	0.95	0.88	0.92
7	0.97	0.96	0.89	0.88	0.94	0.91	0.92	0.95	0.96	0.91	0.85	0.87	0.96	0.91	0.90	0.94
8	0.98	0.97	0.89	0.87	0.95	0.92	0.92	0.95	0.97	0.93	0.85	0.86	0.96	0.92	0.91	0.94
9	0.94	0.93	0.91	0.94	0.93	0.93	0.96	0.98	0.95	0.91	0.87	0.91	0.94	0.91	0.94	0.97
10	0.95	0.93	0.91	0.92	0.93	0.93	0.95	0.96	0.96	0.92	0.86	0.89	0.94	0.91	0.92	0.95
11	0.96	0.93	0.90	0.91	0.94	0.93	0.94	0.96	0.96	0.92	0.85	0.87	0.94	0.91	0.91	0.94
12	0.96	0.94	0.89	0.91	0.93	0.92	0.94	0.96	0.97	0.91	0.85	0.87	0.94	0.89	0.91	0.94

* Cuts 1 to 4 fall between performance levels 1 and 2, 2 and 3, 3 and 4, 4 and 5, respectively.

Figure 6.4 Domain Classification Accuracy

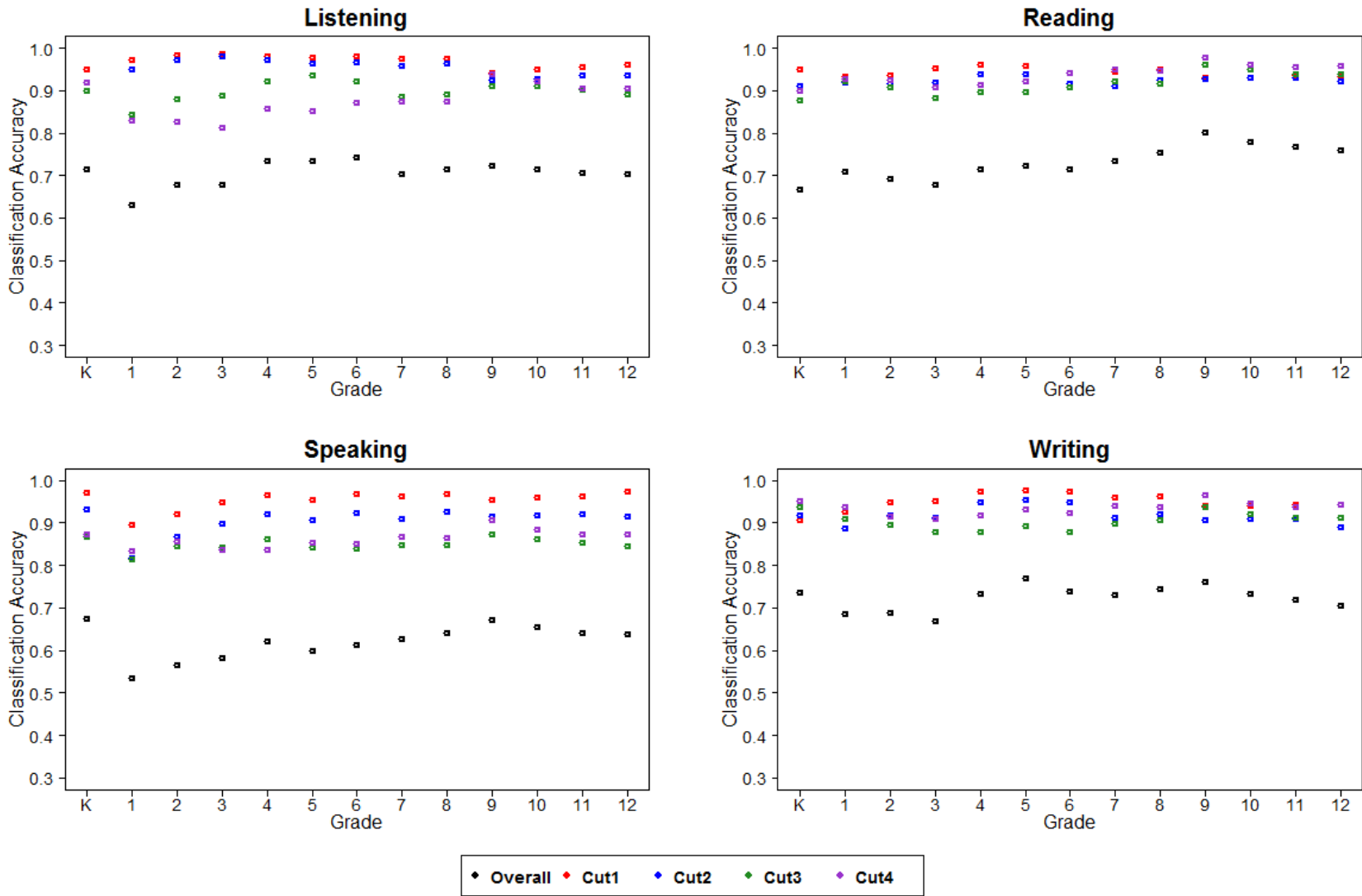
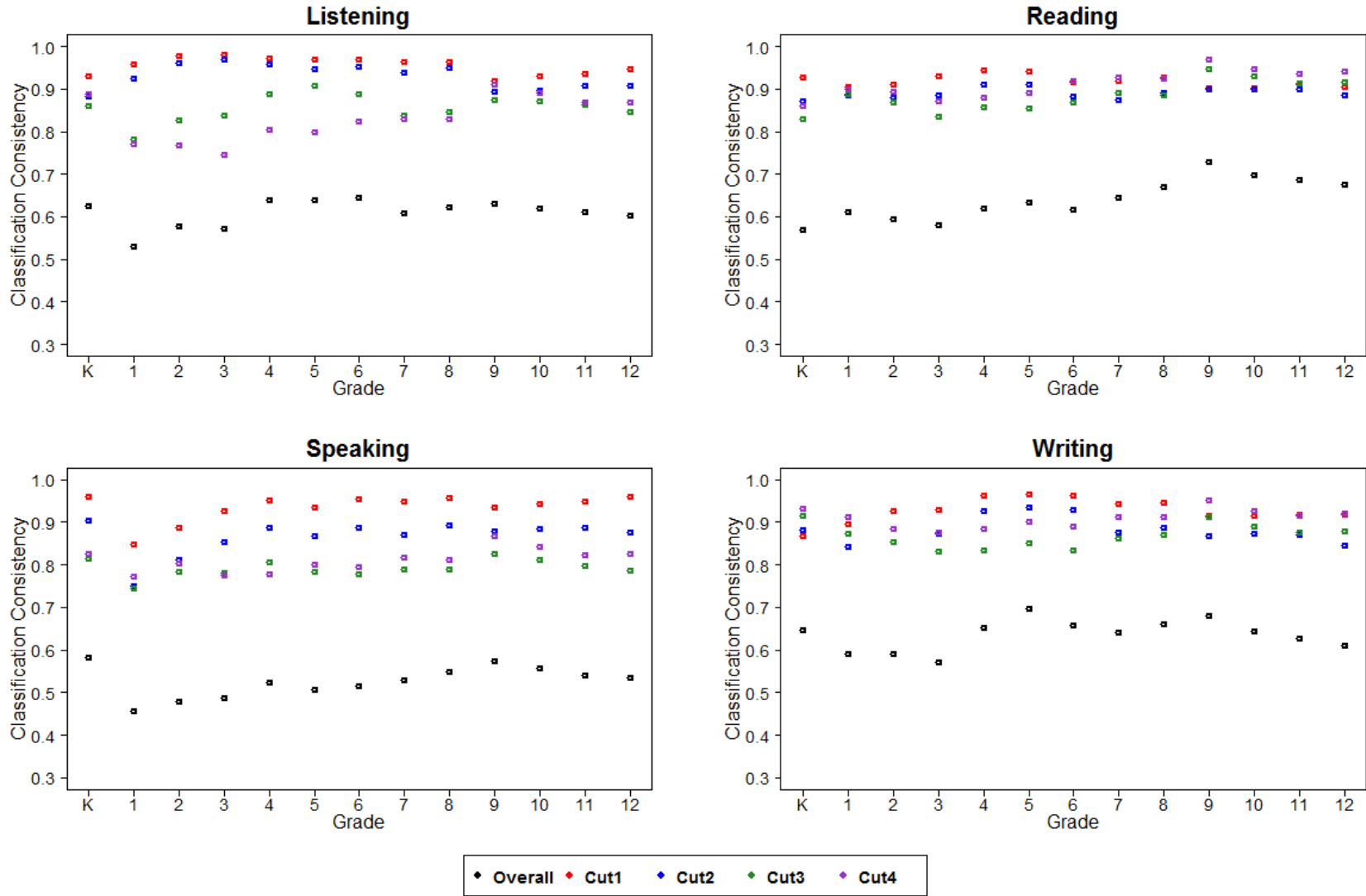


Table 6.5 Classification Consistency for Each Cut

Grade	Listening				Reading				Speaking				Writing			
	Cut1	Cut2	Cut3	Cut4	Cut1	Cut2	Cut3	Cut4	Cut1	Cut2	Cut3	Cut4	Cut1	Cut2	Cut3	Cut4
K	0.93	0.88	0.86	0.89	0.93	0.87	0.83	0.86	0.96	0.90	0.81	0.82	0.87	0.88	0.91	0.93
1	0.96	0.92	0.78	0.77	0.91	0.89	0.89	0.90	0.85	0.75	0.75	0.77	0.90	0.84	0.87	0.91
2	0.98	0.96	0.83	0.77	0.91	0.88	0.87	0.89	0.89	0.81	0.78	0.80	0.93	0.88	0.85	0.88
3	0.98	0.97	0.84	0.75	0.93	0.88	0.84	0.87	0.93	0.85	0.78	0.78	0.93	0.87	0.83	0.88
4	0.97	0.96	0.89	0.80	0.94	0.91	0.86	0.88	0.95	0.89	0.81	0.78	0.96	0.93	0.83	0.89
5	0.97	0.95	0.91	0.80	0.94	0.91	0.85	0.89	0.93	0.87	0.78	0.80	0.97	0.94	0.85	0.90
6	0.97	0.95	0.89	0.82	0.92	0.88	0.87	0.92	0.95	0.89	0.78	0.80	0.96	0.93	0.83	0.89
7	0.96	0.94	0.84	0.83	0.92	0.88	0.89	0.93	0.95	0.87	0.79	0.82	0.94	0.88	0.86	0.91
8	0.96	0.95	0.85	0.83	0.93	0.89	0.89	0.93	0.96	0.89	0.79	0.81	0.94	0.89	0.87	0.91
9	0.92	0.89	0.88	0.91	0.90	0.90	0.95	0.97	0.93	0.88	0.83	0.87	0.92	0.87	0.91	0.95
10	0.93	0.90	0.87	0.89	0.90	0.90	0.93	0.95	0.94	0.88	0.81	0.84	0.91	0.87	0.89	0.93
11	0.94	0.91	0.86	0.87	0.91	0.90	0.91	0.94	0.95	0.89	0.80	0.82	0.92	0.87	0.88	0.91
12	0.95	0.91	0.85	0.87	0.91	0.89	0.92	0.94	0.96	0.88	0.79	0.82	0.92	0.84	0.88	0.92

*Cuts 1 to 4 fall between performance levels 1 and 2, 2 and 3, 3 and 4, 4 and 5, respectively.

Figure 6.5 Domain Classification Consistency



Chapter 7. Validity

Validity refers to the degree to which “evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). Messick (1989) defines validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment.” Both definitions emphasize evidence and theory to support inferences and interpretations of test scores. *The Standards* (AERA, APA, & NCME, 2014) suggests five sources of validity evidence that can be used in evaluating a proposed interpretation of test scores. When validating test scores, these sources of evidence should be carefully considered.

The first source of evidence for validity is the relationship between the test content and the intended test construct. In order for test score inferences to support a validity claim, the items should be representative of the content domain, and the content domain should be relevant to the proposed interpretation of test scores. To determine content representativeness, diverse panels of content experts conduct alignment studies, in which experts review individual items and rate them based on how well they match the test specifications or cognitive skills required for a particular construct (discussions about test development, form construction, scaling, equating, and standard setting can be found in related ELPA21 documents). Test scores can be used to support an intended validity claim when they contain minimal construct irrelevant variance. For example, a mathematics item targeting a specific mathematics skill that requires advanced reading proficiency and vocabulary has a high level of construct irrelevant variance. Thus, the intended construct of measurement is confounded, which impedes the validity of the test scores. Statistical analyses, such as factor analysis or multi-dimensional scaling of relevance, are also used to evaluate content relevance. Evidence based on test content is a crucial component of validity, because construct underrepresentation or irrelevancy could result in unfair advantages or disadvantages to one or more groups of test takers (see Section 7.1).

The second source of evidence for validity is based on “the fit between the construct and the detailed nature of performance or response actually engaged in by examinees” (AERA, APA, & NCME, 2014). This evidence is collected by surveying test takers about their performance strategies or responses to particular items. Because items are developed to measure particular constructs and intellectual processes, evidence that test takers have engaged in relevant performance strategies to correctly answer the items supports the validity of the test scores.

The third source of evidence for validity is based on internal structure: the degree to which the relationships among test items and test components relate to the construct on which the proposed test scores are interpreted. Differential item functioning, which determines whether particular items may function differently for subgroups of test takers, is one method for analyzing the internal structure of tests. Other possible analyses to examine internal structure are dimensionality assessment, goodness-of-model-fit to data, and reliability analysis (see Chapter 6 for details).

A fourth source of evidence for validity is the relationship of test scores to external variables (see Section 7.4 for details). *The Standards* (AERA, APA, & NCME, 2014) divides this source of evidence into three parts: convergent and discriminant evidence, test-criterion relationships, and validity generalization. Convergent evidence supports the relationship between the test and other

measures intended to assess similar constructs. Conversely, discriminant evidence delineates the test from other measures intended to assess different constructs. To analyze both convergent and discriminant evidence, a multi-trait, multi-method matrix can be used. Additionally, test-criterion relationships indicate how accurately test scores predict criterion performance. The degree of accuracy mainly depends upon the purpose of the test, such as classification, diagnosis, or selection. Test-criterion evidence is also used to investigate predictions of favoring different groups. Due to construct underrepresentation or construct-irrelevant components, the relation of test scores to a relevant criterion may differ from one group to another. Furthermore, validity generalization is related to whether the evidence is situation-specific or can be generalized across different settings and times. For example, sampling errors or range restriction may need to be considered to determine whether the conclusions of a test can be assumed for the larger population.

The fifth source of evidence for validity is based on whether the intended and unintended consequences of the test use should be included in the test-validation process. Determining the validity of the test should depend upon evidence directly related to the test; this process should not be influenced by external factors. For example, if an employer administers a test to determine hiring rates for different groups of people, an unequal distribution of skills related to the measurement construct does not necessarily imply a lack of validity for the test. However, if the unequal distribution of scores is in fact due to an unintended, confounding aspect of the test, this would interfere with the test's validity. As described in this document, test use should align with the intended purpose of the test.

Supporting a validity argument requires multiple sources of validity evidence. This then allows for one to evaluate if sufficient evidence has been presented to support the intended uses and interpretations of the test scores. Thus, determining the validity of a test first requires an explicit statement regarding the intended uses of the test scores, and subsequently, evidence that the scores can be used to support these inferences. In this technical report, the standardized test administration can be found in Chapter 2, scoring in Chapter 3, reporting in Chapter 8, and quality control in Chapter 9.

In this chapter, validity for the summative assessment is measured on the internal structure of the items, the comparison of student abilities versus the difficulty of the items, and handscoring precision analysis for the summative assessment. The domain test internal structure is measured using domain dimensionality. The appropriateness of the assessment for the student population is assessed by comparing student abilities vs. test difficulties. Handscoring precision is measured by the consistencies of first raters and second raters. The correlations between summative and screener assessments are presented to provide relationship of test scores to external variables.

Additional results for assessing validity are summarized in the Sections 7-10 in Part I Appendix:

- Section 7. Summative Assessment—Dimensionality
 - Figures S7.1–S7.6 present the scree plots for each domain test. If a test involves multiple forms, the plots are broken down by grade.
- Section 8. Summative Assessment—Ability vs. Difficulty

- Figures S8.1–S8.6 present the comparison of student ability vs. test difficulty on the logit scale for each domain test for each grade of students, respectively.
- Section 9. Summative Assessment—Inter-Rater Analysis
 - Tables S9.1–9.6 display the inter-rater analysis result for each handscored item in each grade.
- Section 10. Student Progress from Screener to Summative – Scatter Plots of Scale Score Change
 - Figures S10.1–S10.6 present the scatter plots of scale score change from screener to summative for each grade band
- Section 11. Student Progress from Screener to Summative – Comparison of Scale Scores
 - Tables S10–S10.6 summarize comparisons of scale score summary statistics between screener and summative for listening, reading, speaking, writing, overall and comprehension scores

7.1 DIMENSIONALITY ANALYSIS

The graded response model (Samejima, 1969) is used for each domain test. One of the underlying assumptions of the graded response model is unidimensionality. For OELPA, the principal components analysis (PCA) with an orthogonal rotation method (Jolliffe, 2002; Cook, Kallen, & Amtmann, 2009) is used to investigate the dimensionality for each domain test.

The results are presented in the scree plots in Section 7 in the Appendix Part I. The graphs show that the magnitude of the first eigenvalue is always noticeably larger than the magnitude of the second factor in all tests, which indicates that the test has a dominant factor.

7.2 STUDENT ABILITIES VS. TEST DIFFICULTIES

When student abilities are well matched to test difficulties, the measurement errors are reduced. Therefore, it is desired that the test difficulty matches student ability. To examine this aspect of the test, domain difficulties were plotted versus student abilities.

The results are included in Section 8 in the Appendix Part I. It shows that the student abilities are generally higher than the test difficulties in all domain tests, except the reading tests grades 9–12, where the test difficulties match student abilities well.

7.3 INTER-RATER ANALYSIS

The fidelity of handscoring is monitored using a second rater. Scores from first and second raters were compared. Second raters are usually teachers, test administrators, administrators, or other qualified staff.

In 2019 OELPA summative assessment, 16 handscored items in kindergarten and 13 handscored items in each of the other grade or grade-band online tests were scored by second raters. Twenty percent of handscores were monitored by second raters. In the screener assessment, OELPA adopted local TA's for scorings and hence there were no second rater scores.

Handscorer score reliability is examined using Cohen's weighted Kappa coefficient. The coefficient allows differential weighting of disagreement and a measure of agreement corrected for chance. In addition, the frequencies and percentages of the exact match between first rater and second rater, the exact match plus +1/-1 score differences, +2/-2 and above differences are computed. Handscored items on paper and braille forms are not included in the results due to the small sample size.

Table 7.1 contains the number of handscores that were monitored by second raters, the weighted Kappa coefficient, and the frequencies of exact matches, exact matches plus +1/-1 off, and +2/-2 or more off on item scores. The table shows that, across the states, 69–83% of handscores are consistent between the first rater and the second rater; 15–29% of handscores are off by one point between the two raters, and less than 2% of handscores are off by two or more points across the six tests. The weighted Kappa coefficients range from 0.82 to 0.87.

The inter-rater consistencies are also assessed by item and are summarized in Section 9 of the Appendix Part I.

Table 7.1 Overall Inter-Rater Analysis

Grade/Grade Band	Score point	Number Items	Number Students	Weighted Kappa	Count			Percent		
					Exact	Exact +1/-1	More than +2/-2	Exact	Exact +1/-1	More than +2/-2
K	2	9	15828	0.826	14115	15691	137	89.2	99.1	0.9
	3	7	12328	0.8526	8811	12174	154	71.5	98.8	1.2
1	2	7	10893	0.7459	9994	10838	55	91.7	99.5	0.5
	3	3	4666	0.828	3600	4600	66	77.2	98.6	1.4
	4	2	3106	0.906	2300	3052	54	74.1	98.3	1.7
	5	1	1543	0.843	871	1484	59	56.4	96.2	3.8
	2	5	12264	0.7319	10915	12213	51	89	99.6	0.4
2-3	3	4	9827	0.8645	7632	9769	58	77.7	99.4	0.6
	4	3	7325	0.8737	5240	7245	80	71.5	98.9	1.1
	5	1	2437	0.8425	1527	2377	60	62.7	97.5	2.5
4-5	3	8	12949	0.8112	9847	12824	125	76	99	1
	4	1	1615	0.806	1006	1594	21	62.3	98.7	1.3
	5	4	6470	0.83	4040	6303	167	62.4	97.4	2.6
6-8	3	8	15784	0.8373	12342	15641	143	78.2	99.1	0.9
	5	5	9861	0.8852	6714	9663	198	68.1	98	2
9-12	3	8	19407	0.8214	14659	19250	157	75.5	99.2	0.8
	5	5	11985	0.8585	7064	11642	343	58.9	97.1	2.9

7.4 RELATIONSHIP OF TEST SCORES TO EXTERNAL VARIABLES

The screener assessment was administered in addition to the summative assessment. The test questions of the screener tests are based on the same English Language Proficiency (ELP) Standards adopted by the ELPA21 member states, including OELPA. Test takers who took both the summative assessment in spring 2019 and optional screener assessments were identified for conducting the cross-test set of correlations.

The correlation between the scale scores from summative and screener tests is assessed using Pearson correlations. The correlation between the performance levels from both tests is assessed using gamma correlation. The gamma correlation, or gamma statistics, is for ordinal level data with a small number of response categories. It is designed to determine how effectively a researcher can use the information about an individual measured on one variable to predict the measure of the individual on another variable. The significance of the correlation can be tested via z-test. The computations of gamma statistics and its z-test can be found in <https://www.harding.edu/sbreezeel/460%20files/statbook/chapter14.pdf>. The correlation results are presented in Table 7.2 and Table 7.3.

Table 7.2 shows the correlation between the screener and the summative tests in domain and composite scores. Correlations of all types of scores are the lowest in the kindergarten test, followed by the grade 1 test; the correlations are above 0.8 in listening, reading, writing, and overall scale scores in grade 2 and above, except for grades 2-3 listening and grades 4-5 reading tests. The speaking tests have relatively lower correlations than the other three domains except those taken at the kindergarten level.

Table 7.3 shows the correlations between domain performance levels and test proficiency levels. Similar to the correlations between scale scores presented in Table 7.2, kindergarten has the lowest correlations in all domain performance levels and overall proficiency level. For grade 2 and above, the correlations are about 0.8 except for the speaking domain. In addition, the correlations between proficiency levels are overall higher than those between domain performance levels. This is because there are three levels in overall proficiency while there are five levels in domain performance. These correlations show predictive validity between the two OELPA tests because they were given to the same students at different times.

Table 7.2 Correlations by Domain, Overall, and Comprehension Scores

Grade/Grade Band	Listening		Reading		Speaking		Writing		Comprehension		Overall	
	N	Corr	N	Corr	N	Corr	N	Corr	N	Corr	N	Corr
K	5027	0.57	5001	0.53	4976	0.65	4991	0.42	5030	0.49	5034	0.64
1	941	0.77	939	0.72	931	0.71	936	0.72	942	0.74	943	0.79
2-3	1190	0.79	1177	0.80	1172	0.76	1176	0.81	1190	0.77	1190	0.84
4-5	785	0.80	786	0.78	785	0.74	780	0.80	788	0.78	792	0.82
6-8	985	0.82	982	0.80	970	0.77	978	0.83	989	0.80	992	0.85
9-12	1171	0.83	1163	0.83	1128	0.76	1164	0.83	1175	0.81	1178	0.86

Table 7.3 Correlations by Domain Performance Levels and Overall Proficiency Levels

Grade/Grade Band	Listening		Reading		Speaking		Writing		Overall	
	N	Corr	N	Corr	N	Corr	N	Corr	N	Corr
K	5027	0.67	5001	0.57	4976	0.61	4991	0.48	5034	0.88
1	941	0.80	939	0.75	931	0.61	936	0.76	943	0.94
2-3	1190	0.79	1177	0.82	1172	0.68	1176	0.84	1190	0.92
4-5	785	0.82	786	0.82	785	0.69	780	0.84	792	0.92
6-8	985	0.84	982	0.85	970	0.70	978	0.86	992	0.96
9-12	1171	0.85	1163	0.90	1128	0.76	1164	0.89	1178	0.94

7.5 STUDENT PROGRESS

Student progress from the time they took screener tests to the time they took summative tests is evaluated by the changes in scale scores and performance levels. The major confounding factor in this result is the measurement error in both assessments. Given the acceptable marginal reliability indices described in Chapter 6 of the technical report Part I and Chapter 4 **Error! Reference source not found.** of the technical report Part II, we can still see the trend of student progress.

Figure 7.1 and Figure 7.2 show the growth of the average domain scores and composite scores, respectively from the 2018-19 OELPA school year. The average scale scores in the summative assessment are in general higher than those in the screener assessment, except for the writing domain in the grades 9–12 test, where the screener test has slightly higher average scale score than the summative test. In addition, Section 10 of the Appendix Part I presents a scatter plot of scale score change from screener to summative assessments for each grade, and Section 11 summarizes a comparison of scale scores between screener and summative assessments.

Figure 7.1 Average Domain Scale Score Progress

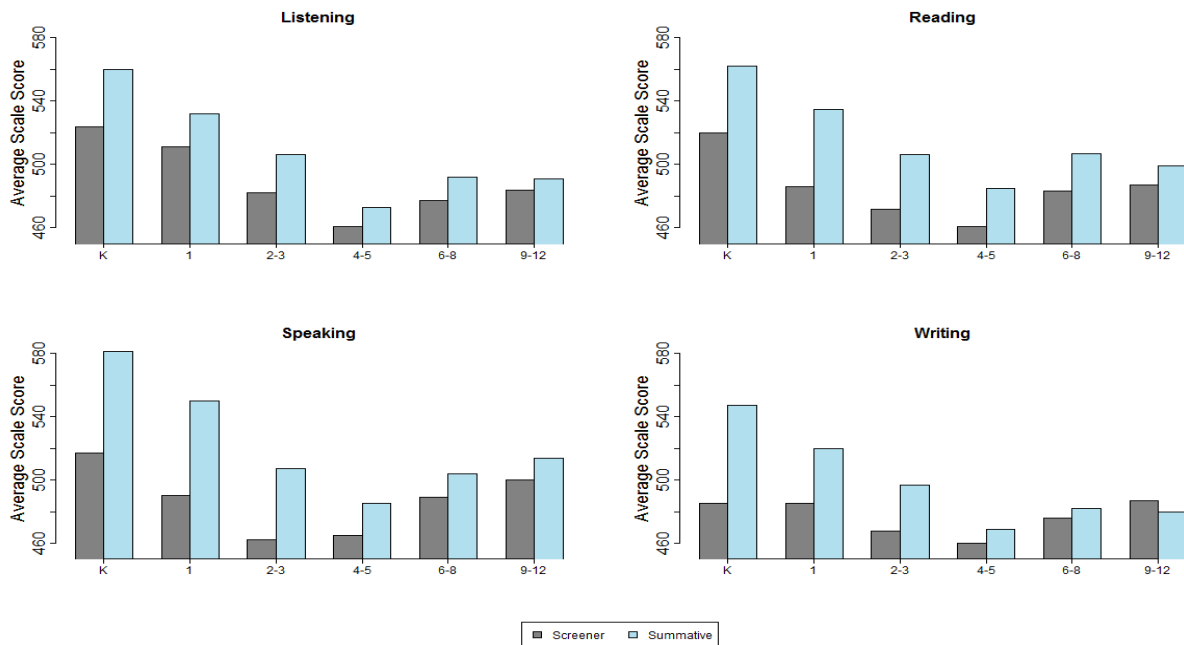


Figure 7.2 Average Overall and Comprehension Scale Score Progress

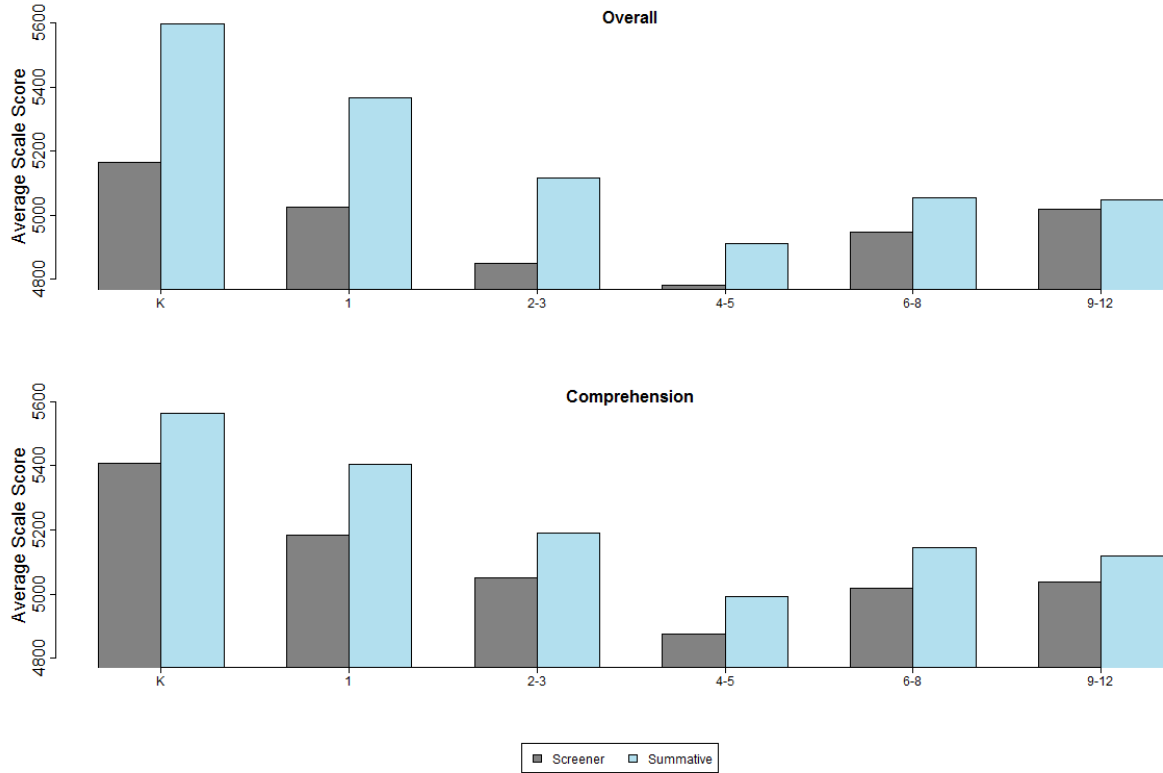


Figure 7.3 and Figure 7.4 display the percentage of students in each domain performance level and overall proficiency level, respectively. In each pair of bars, the left bar is from the screener test and the right bar is from the corresponding summative test. The plots indicate that more students are in higher domain performance levels and overall proficiency levels in the summative tests.

Figure 7.3 Percentage of Domain Performance Level from Screener to Summative

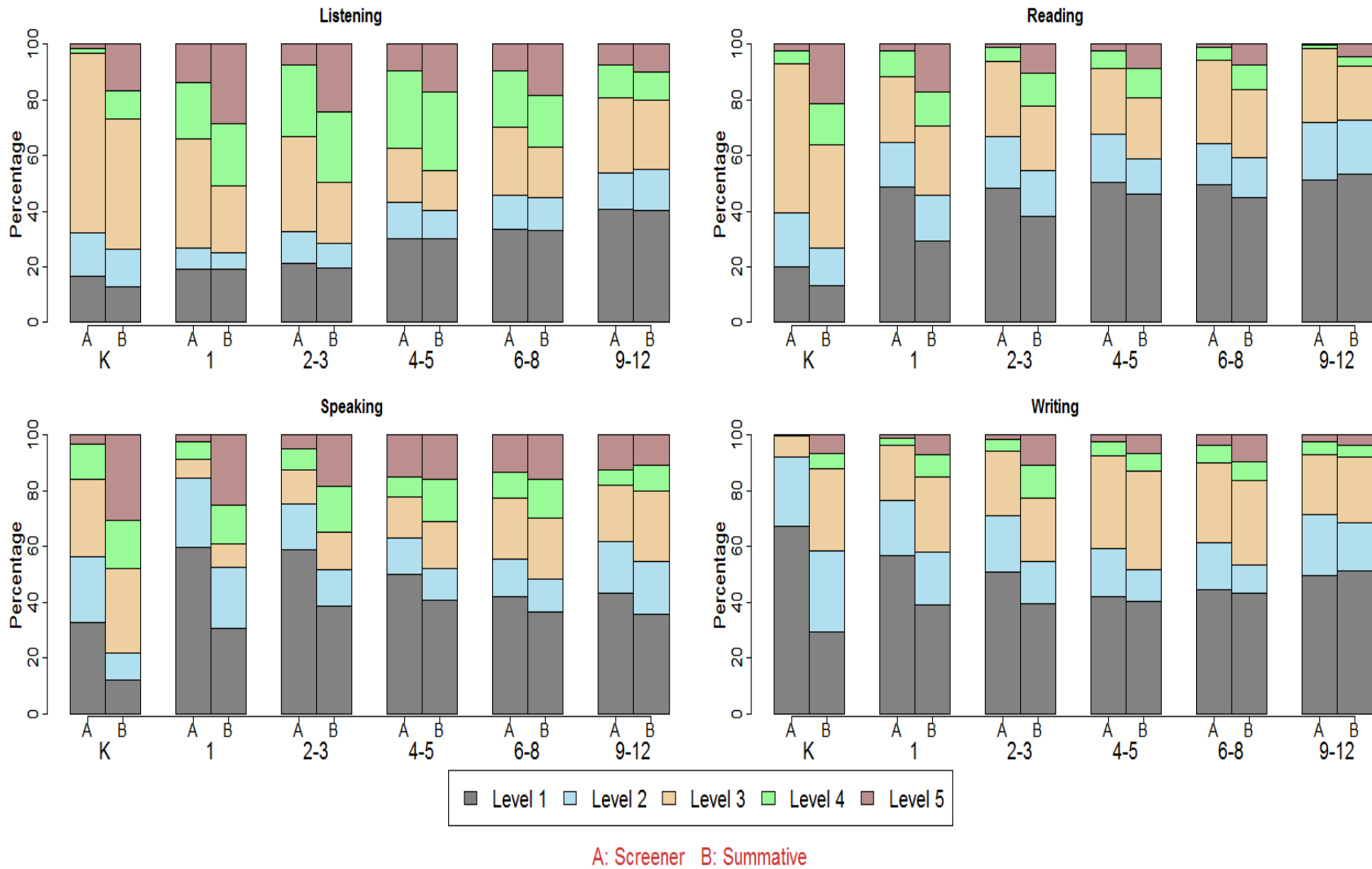
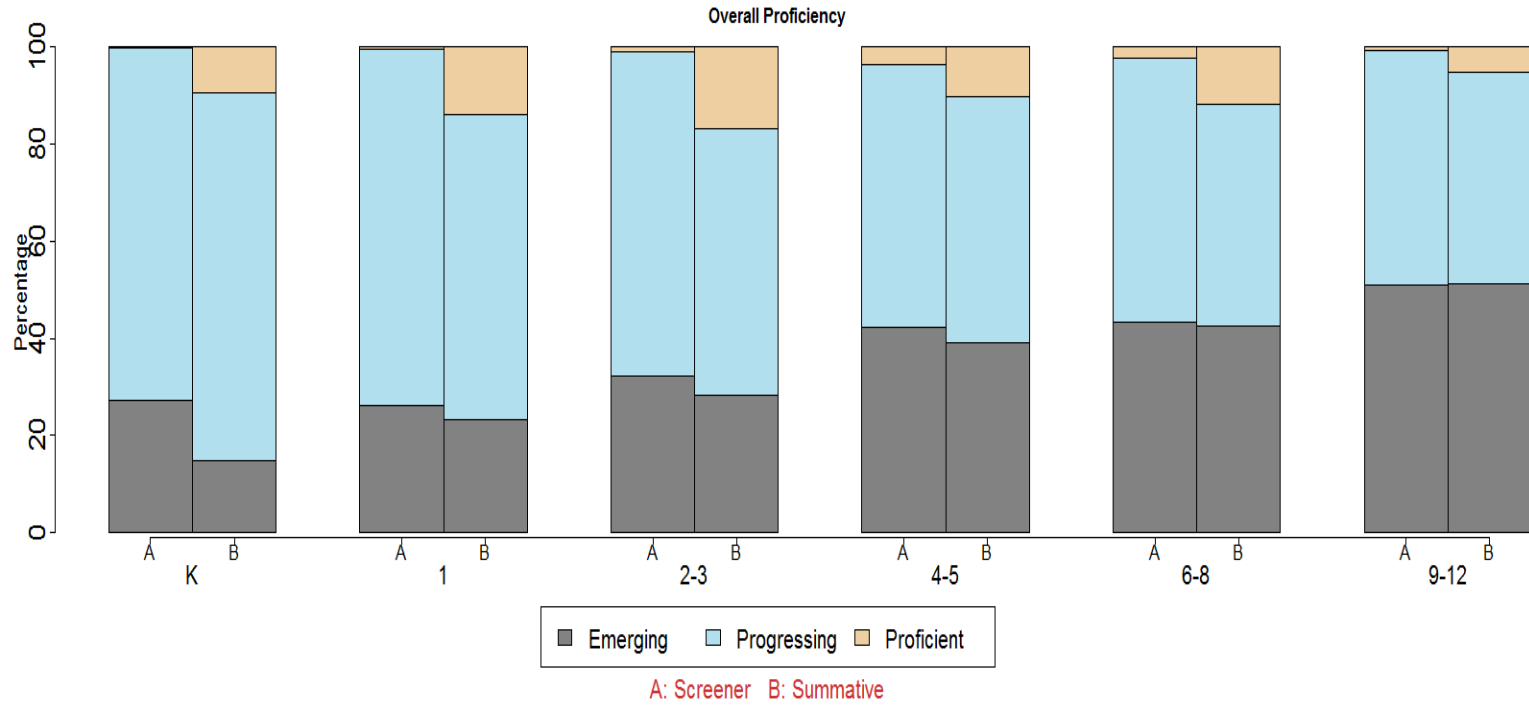


Figure 7.4 Percentages of Overall Proficiency from Screener to Summative



Chapter 8. Reporting

For both summative and screener tests, the OELPA results are available in the Online Reporting System (ORS) and ORS-generated paper family reports to be sent home with the students.

8.1 ONLINE REPORTING SYSTEM

The ORS generates a set of online score reports that includes reliable and valid information describing student performance for students, parents, educators, and other stakeholders for both summative and screener assessments. Because the score reports on student performance are updated each time students complete tests, and those tests are handscored, authorized users (e.g., school principals, teachers) may view student performance on the tests and can use the results to improve student learning. In addition to the individual student’s score report, the ORS produces aggregate score reports for teachers, schools, districts, and states. Additionally, the ORS provides participation data that help monitor the student participation rate.

Furthermore, to facilitate comparisons, each aggregate report contains summary results for the selected aggregate unit, as well as all aggregate units above the selected aggregate. For example, if a school is selected, the summary results of the district to which the school belongs and the summary results of the state are also provided so that the school performance can be compared with district and state performance. If a teacher is selected, the summary results for the school, the district, and the state are also provided for comparison purposes. Table 8.1 typical types of online reports and the levels at which they can be viewed (student, roster, teacher, school, and district).

Table 8.1 Types of Online Score Reports by Level of Aggregation

Level of Aggregation	Types of Online Score Reports
State District School Teacher Roster	Number of students tested and percentage of students determined proficient (overall and by subgroup) Average overall and comprehension scale scores and standard errors of the average scale scores (overall and by subgroup) Percentage of students at each domain performance level (overall and by subgroup) Average domain scale scores and the standard errors of average scale scores (overall and by subgroup) On-demand student roster report
Student	Overall and comprehension scale scores and standard errors of the scale scores Proficiency status based on the domain performance levels Domain scale scores with domain performance levels and level descriptors

8.1.1 Types of Online Score Reports

The Online Reporting System (ORS) is designed to help educators, students, and parents answer questions regarding how well students have performed in the assessment for each domain. The ORS is designed with great consideration for stakeholders who are not technical measurement

experts (e.g., teachers, parents, students). It ensures that test results are easily readable. Simple language is used so that users can quickly understand assessment results and make valid inferences about student achievement. In addition, the ORS is designed to present student performance in a uniform format. For example, similar colors are used for groups of similar elements, such as achievement levels, throughout the design. This design strategy allows state-, district-, and school-level users to compare similar elements and to avoid comparing dissimilar elements.

Once authorized users log in to the ORS and select Score Reports, the online score reports are presented hierarchically. The ORS starts by presenting summaries on student performance by grade at a selected aggregate level. In order to view student performance for a specific aggregate unit, users can select the specific aggregate unit from a drop-down menu with a list of aggregate units (e.g., schools within a district or teachers within a school) to choose from. For more detailed student assessment results for a school, a teacher, and a roster, users can select the grade on the online score reports.

Generally, the ORS provides two categories of online score reports: aggregate score reports and student score reports. Table 8.1 summarizes the typical types of online score reports available at the aggregate level and the individual student level. Detailed information about the online score reports and instructions on how to navigate the online score reporting system can be found in the *Online Reporting System User Guide* for each state, accessible by the help button in the ORS, as an example shown in Figure 8.1.

Figure 8.1 ORS Page for Online Reporting System User Guide

Ohio | Department of Education

Score Reports
Reports & Files ▾
State at a Glance

Inbox
Search Students
View/Edit Rosters
This Page: Help | Print | Export

📍 Now viewing: Scores for students who were mine when they tested during the selected administration

Home Page Dashboard

Select Test and Year

Test: OELPA ▾
 Administration: Spring2019 ▾
 Scores for students who were mine at the end of the selected administration
 Scores for my current students
 Scores for students who were mine when they tested during the selected administration

Select

Ohio Department of Education ▾

Select a district and then click on a grade and subject to view more information.

Overall Performance on the OELPA test, by Grade: Ohio Department of Education, Spring2019

OELPA

Grade	Number of Students Tested	Percent Determined Proficient
Kindergarten	9136	8%
Grade 1	8009	14%
Grade 2	6897	26%
Grade 3	5712	24%
Grade 4	4683	25%
Grade 5	3680	17%
Grade 6	3432	18%
Grade 7	3411	11%
Grade 8	3366	11%
Grade 9	4329	5%
Grade 10	3503	9%
Grade 11	2783	11%
Grade 12	2075	9%

8.1.2 Subgroup Reports

The aggregate score reports at a selected aggregate level are provided for students overall and by subgroups. Users can see student assessment results by any subgroup. Table 8.2 presents the typical subgroups and subgroup categories provided in the ORS.

Table 8.2 Types of Subgroups

Breakdown by Category	Displayed Category
Ethnicity	Hispanic or Latino
	American Indian or Alaska Native
	Asian
	Black or African American
	White
	Native Hawaiian or Other Pacific Islander
	Two or More Races
Gender	Male
	Female
Limited English Proficiency Status	Yes
	No
IDEA Indicator	Yes
	No
Enrolled Grade	Kindergarten
	Grade 01
	Grade 02
	Grade 03
	Grade 04
	Grade 05
	Grade 06
	Grade 07
	Grade 08
	Grade 09
	Grade 10
	Grade 11
Grade 12	

8.2 PAPER REPORTS

The Online Reporting System (ORS) provides the functionality for users to print reports described earlier. The ORS also allows users to print the family report for each student. Figure 8.2 shows the mock-up of the paper family report.

Figure 8.2. Mock-Up for Family Report

Student Name: Jane E. Doe
 Grade: 5 School: Demo School
 Test Date: Spring 2019 District: Demo District

Ohio | Department of Education | Ohio English Language Proficiency Assessment

Dear Doe Family,

This report contains Jane’s results from the spring 2019 Ohio English Language Proficiency Assessment (OELPA). The OELPA measures reading, listening, writing, and speaking for students who are learning English. A student can score from level 1 to level 5 on each of these tests. Schools give the OELPA to determine whether to exit a student from the English language learning program.

A student’s overall performance level is based on scores on the reading, listening, writing, and speaking tests. It shows whether a student is Emerging, Progressing, or Proficient in the English language.

Your child Jane is **Proficient** in the English language. See below to learn what this means.

Overall Performance Levels



Proficient – The student can work without help in a classroom where English is used for teaching. The student may now exit the English learning program. The student will no longer be identified as an English learner. The student scored any combination of level 4s and 5s on the four tests.

Progressing – The student needs English language services to be able to work in the classroom. The student will continue to receive English language services. The student scored a combination of levels that do not meet the Proficient or Emerging levels.

Emerging – The student is in the beginning stages of learning English. The student will continue to receive English language services. The student scored any combination of level 1s and 2s on the four tests.

Test Performance Levels

A student can score from Level 1 to Level 5 on the reading, listening, writing, and speaking tests. Level 5 is the highest. The table below shows you the four tests and how Jane did on each of them. See the back page to learn more about each level on each test.

	Level 1	Level 2	Level 3	Level 4	Level 5
READING					
LISTENING					
WRITING					
SPEAKING					

What can I do to help my child?


Talk with your child’s teacher for more details on what your child’s performance levels mean. Discuss services provided in the classroom to help with learning English. Ask your child’s teacher for home activities to improve English learning.

66 6666 6666

Student Name: **Jane E. Doe**
 Grade: **5**
 Test Date: **Spring 2019**

School: **Demo School**
 District: **Demo District**

Ohio | Department of Education | **Ohio English Language Proficiency Assessment**



Reading


Level 5: Student reads many types of written words and sentences at the level expected for the grade the student is in. Student understands difficult written English. Student understands social and classroom language used at this grade level.

Level 4: Student understands most grade-level books and nonfiction written words and sentences but may need help with new written words and sentences. Student corrects mistakes while reading.

Level 3: Student understands written directions and can draw simple conclusions. Student understands most social language and some classroom language.

Level 2: Student understands short written words and sentences and simple wording. Student uses pictures or what the student knows to help understand the main ideas of simple reading passages.

Level 1: Student identifies single words and some phrases. Student may understand simple written directions. Students may understand the meanings of common words.



Listening


Level 5: Student identifies the main ideas and important details of discussions and lectures. Student identifies a speaker’s purpose and understands social and classroom language.

Level 4: Student understands directions that have more than one step and understands point of view. Student understands a range of social and classroom language.

Level 3: Student understands main ideas in short conversations and understands some classroom language.

Level 2: Student understands simple spoken statements, questions and one-step directions.

Level 1: Student understands some common words and one-phrase directions.



Writing


Level 5: Student writes for a variety of purposes and audiences without help.

Level 4: Student writes using different types of words and sentences making some errors that do not affect meaning.

Level 3: Student writes using different types of words and sentences making many errors in spelling, punctuation and capitalization.

Level 2: Student writes simple stories or descriptions with limited language and needs help to revise work.

Level 1: Student may be able to copy letters and write some words, but the wording does not give a clear message.



Speaking

Level 5: Student speaks well with language and grammar that is similar to native English speakers in the same grade.

Level 4: Student’s speech is mostly clear and organized with some mistakes that seldom interfere with speech.

Level 3: Student speaks well in simple situations but mistakes in grammar and word sounds make it difficult to express difficult ideas.

Level 2: Student speaks in simple words and phrases but makes many mistakes in language, word order and grammar that cause problems.

Level 1: Student repeats short, common phrases and may be able to provide simple answers to requests.

6666666666

Chapter 9. Quality Control

Thorough quality control has been integrated into every aspect of ELPA21 (including OELPA) summative and screener tests. ELPA21, the states (including Ohio Department of Education), Questar, AIR, and Measurement Incorporated (MI) have built in multiple layers of reviews and verifications to ensure that outputs are of the highest quality in areas such as materials prepared for item-writing workshops, test form constructions, test booklet development and printing, post-test score quality control processes, and reporting. The quality control for item-writing workshops, test form construction, and test booklet development and printing can be found in the related documents prepared by ELPA21 and associated vendors. This chapter describes the quality control procedures related to test administration, scoring, and reporting. The service was provided by AIR and MI.

9.1 QUALITY CONTROL IN TEST CONFIGURATION

For online summative and screener testing, the test configuration files contain the complete information required for test administration and scoring, such as the test blueprint specifications, slopes and intercepts for theta-to-scale score transformation, cut scores, and item information (i.e., answer keys, item attributes, item parameters, passage information). The accuracy of the configuration file is checked and confirmed independently numerous times by multiple teams prior to the testing window. Scoring is also verified before the testing windows open.

9.1.1 Platform Review

AIR's online test delivery system supports a variety of item layouts for online test administration to many populations of students, including students who need designated supports and accommodations to test online. Each item on the assessment goes through an extensive platform device review on different operating systems, including Windows, Linux, and iOS, to ensure that the item displays consistently across all platforms. Some of the layouts have the stimulus and item response options/response area displayed side by side. In each of these layouts, both the stimulus and the response options have independent scroll bars.

Platform review is a process in which each item is checked to ensure that it is displayed appropriately (i.e., rendered) on each tested platform. A platform is a combination of a hardware device and an operating system. In recent years, the number of platforms has proliferated, and platform review now takes place on various platforms that are significantly different from one another.

Platform review is conducted by AIR's quality assurance team. The team leader projects every item from AIR's item tracking system (ITS¹), and team members, each behind a different platform, look at the same item to ensure that it renders as expected.

¹ ITS is AIR's item bank for ELPA21. It contains all information that relates to each item, such as item content categories at all levels, item type, maximum score points, item statistics from each administration, etc.

9.1.2 User Acceptance Testing and Final Review

Both internal and external user acceptance testing (UAT), usually the state's, were conducted before the testing window opened. Detailed protocols were developed for the review process of the test delivery system (TDS), and reviewers were given thorough instructions to note or report issues related to system functionality, item display, and scoring.

During the internal UAT, AIR staff took all OELPA online tests that covered the entire range of possibilities of item responses and the complete set of scoring rules in TDS. When issues were found, AIR took immediate actions to address them. When TDS was updated, the tests were taken again to ensure the issues were fixed. The process was repeated until all issues were resolved during the UAT period prior to operational testing.

State staff were also able to conduct a hands-on review of the system prior to the testing window opening. The states approved the TDS before the system was opened for testing.

Before the ORS opened, AIR and the state staff conducted internal and external UAT of the system similar with that of TDS to ensure that the ORS would function as intended when opened to the public for score reporting.

9.2 QUALITY ASSURANCE IN SCORING

The quality assurance of scoring includes the assurance of the online data, the precision of handscoring, the correctness of machine scoring, and the strictness when applying the business rules in scoring. This section describes the details of scoring quality assurance.

The kindergarten and grade 1 writing supplements and the writing constructed-response items were handscored by Measurement Incorporated (MI). For online tests, the responses for the handscored items were transferred between AIR and MI on a rolling basis via Ledger.² Therefore, as soon as a student submitted a test to TDS, the responses to handscored items were transformed into XLM format, and were then sent to Ledger, from which MI retrieved responses for handscoring. When scoring was complete, the record was sent to Ledger, from which AIR downloaded the record for final scoring. The data transmission process was automatic.

For paper-pencil tests, after test administration, student responses were entered into the AIR Data Entry Interface (DEI) on the state testing portal for all ELPA21 domain tests, with the exception of the kindergarten and grade 1 writing supplements and the writing constructed-response items. The item responses of the writing supplements and writing constructed-response items were mailed to MI for scoring via secure shipping. After scoring, MI transmitted the scores to the Ledger system, from which AIR retrieved the item scores for final scoring. To answer speaking items, students who took paper-pencil tests spoke into the DEI directly, and the item responses followed the online procedure for scoring.

² Ledger is an electronic system that AIR and MI use to transmit data from one vendor to the other for purposes of transmitting and reporting handscored item scores. Individual responses can be tracked at all times through Ledger before a record is reported.

For braille tests, item responses were entered into the braille DEI by test administrators. The data were processed following the online data processing procedure, and the secure testing materials were returned to the scoring vendor, MI.

9.2.1 Quality Assurance in Online Data

AIR’s test delivery system (TDS) has a real-time, built-in quality monitoring component. After a test is administered to a student, TDS passes the resulting data to our quality assurance (QA) system. QA conducts a series of data integrity checks, ensuring, for example, that the record for each test contains information for each item, keys for multiple-choice items, score points in each item, and total number of items, and that the test record contains no data from items that have been invalidated.

Data pass directly from the Quality Monitoring System (QMS³) to the Database of Record (DoR), which serves as the repository for all test information and from which all test information for reporting is pulled. The Data Extract Generator (DEG) is the tool that is used to pull data from the DoR for delivery to each state. AIR staff ensure that data in the extracted files match the DoR prior to delivery to the state.

9.2.2 Quality Assurance in Handscoring

MI’s scoring process is designed to employ a high level of quality control. The quality control procedures are implemented at each stage of the scoring process, which includes scorer recruitment, leader recruitment, training, and various reports that help to ensure scoring quality.

Scorer Recruitment/Qualifications

MI retains scorers who have years of experience in handscoring, and those scorers made up approximately 65% of the scorer pool. To complete the scorer staffing for this project, MI placed advertisements on various job boards, in local papers, in publications, and at regional colleges and universities. Recruiting events were held, and applications for scorer positions were screened by MI recruiting staff. Candidates were personally interviewed, and references and proof of a four-year college degree were collected. Candidates completed placement tests for English language arts (ELA; reading and writing) and mathematics. In this screening process, preference was given to candidates with previous experience scoring large-scale assessments. The scorer pool consisted of educators, writers, editors, and other professionals who were valued for their experience, but who were also required to set aside their own biases about student performance and accept the scoring standards.

Leadership Recruitment/Qualifications

Scoring directors and team leaders had experience as successful scorers and leaders on previous MI projects and had strong backgrounds in scoring content-specific projects. These individuals demonstrated strong organization, leadership, and management skills. All scoring directors, team

³ QMS is AIR’s quality monitoring system. It ensures that the information in a student record, such as item key or score point, is correct.

leaders, and scorers were required to sign confidentiality agreements prior to training with ELPA21 materials or handling secure materials.

Each room of scorers was assigned a scoring director or assistant scoring director. This individual led the handscoring for the duration of the project and was monitored by the scoring project manager. The scoring director conducted the team leader training and was responsible for training the scorers.

In general, team leaders assisted the scoring directors/assistant scoring directors with scorer training and monitoring by working with their teams in small group discussions and answering individual questions that scorers may not have felt comfortable asking in a large group. Once scorers were qualified, the team leaders were responsible for maintaining the accuracy and workload of team members. The ongoing monitoring identified those scorers who were having difficulty scoring and resulted in individual scorers receiving one-on-one retraining. If this process did not correct inaccuracies in scoring, individual scorers were released from the project.

Training

To train ELPA21 scorers, MI scoring staff used approved rubrics and training materials. The training materials were composed of anchor, qualifying, and training responses provided by the ELPA21 Program. Training materials included a comprehensive annotated scoring guide for each item. The guide contained the anchor set scorers referenced while evaluating live student responses. The scoring guides also contained several typical student responses presented in score point order.

Guides included detailed annotations explaining how the scoring criteria applied to each response's specific features and why the response merited a particular score. Guides included responses that were the most useful in making scoring decisions, including some that fell within the upper and lower ranges of the score point to help scorers define the lines between score points.

Anchor and qualifying sets were designed to help the scorers learn to apply the criteria illustrated in the scoring guide, ensure that they become familiar with the process of scoring student responses, and assess the scorers' understanding of the ELPA21 scoring criteria before they were allowed to begin live scoring.

The item-specific rubrics served as the scorers' constant reference. Scorers were instructed on how to apply the rubrics and were required to demonstrate a clear comprehension of each anchor set by performing well on the training materials that were presented for each grade and item.

Team leaders assisted the scoring directors with the training and monitoring of scorers. The scoring director conducted the team leader training before the scorer training. This training followed much of the same process as the scorer training, but additional time was allotted for review, discussion, and addressing anticipated scorer questions and concerns. To facilitate scoring consistency, it was imperative that each team leader imparted the same rationale for each response that other team leaders used. Once team leaders qualified, leadership responsibilities were reviewed and team assignments were given. A ratio of one team leader for 8–10 scorers ensured adequate monitoring of the scorers.

Scorer training involved an intensive review of the rubric and anchor responses, provided by the scoring director, to help the scorers internalize the scoring criteria. The scoring director and team leaders led a thorough discussion of the training materials with the entire group. All responses

were discussed using the annotations from rangefinding, which was used to identify the full range of responses that represent each score point and produce scoring training materials including qualification, anchor, practice, and validity sets. A similar process was followed in training for writing and speaking items.

Once the scoring guidelines were discussed, scorers were required to apply the scoring criteria by qualifying (i.e., scoring with acceptable agreement to the “true” scores decided upon at rangefinding) on at least one of the qualifying sets. Scorers who failed to achieve the qualifying criteria were given additional training. Scorers who did not perform at the required level of agreement for a given item or related group of items by the end of the qualifying process were not permitted to score live student work.

Training was an ongoing process that did not end after the qualifying rounds. Feedback was an integral part of several reliability checks that were performed throughout the project. Primarily, team leaders monitored scorers’ reliability by conducting read-behinds/listen-behinds on an as-needed basis. This was a process whereby team leaders reread or re-listened to the responses and checked the scores of each scorer on their team to catch potential scorer drift so that the scorer could have immediate feedback and be retrained in a timely fashion. The percentage of read-behinds/listen-behinds conducted for an individual scorer was not fixed but varied based on current levels of performance. Scorers were removed from scoring an item or a related group of items if they were unable to score consistently with the rubric and the anchor responses after retraining. The scores were reset so the items would be scored correctly by other scorers.

Development and rangefinding of the materials used with the 2017 administration were completed by a previous vendor. For 2018–2019, a field-test score validation of the new short response speaking items was conducted by MI. This information is available from the Program.

9.2.3 Handscoring Quality Assurance Monitoring Reports

Measurement Incorporated (MI) scorer accuracy was monitored throughout the scoring sessions by producing real-time, on-demand reports to ensure that an acceptable level of scoring accuracy was maintained. Interscorer reliability was tracked and monitored with multiple quality control reports that were reviewed by MI scoring staff. These reports were reviewed by the program manager, scoring project director, scoring directors, and team leaders. The following reports, available in daily, cumulative, and summary formats, were used during handscoring:

- *Interscorer Reliability Reports* displayed how often scorers were in exact agreement and supported maintaining an acceptable agreement rate. These reports provided exact, adjacent, and nonadjacent interscorer agreement, as well as mismatches between scores and nonscorable codes and within nonscorable codes. They also indicated the number of responses read by each scorer.
- *Score Point Distribution Reports* displayed the percentage of responses that had been assigned each of the score points and nonscorable codes.
- *Validity Reports* tracked how the scorers performed by comparing predetermined scored responses to scores assigned by the selected scorer on the same set of responses. If the assigned score of the selected scorer fell outside of a determined

percentage of agreement, remediation occurred and additional responses were reviewed by the team leader of the individual(s) who needed to be monitored more closely.

- *Item Status Reports* tracked each item and indicated the status (e.g., “first read complete,” “tabled”). This report was used to monitor the overall status and progress of handscoring.

Maintaining Consistency

MI utilized numerous processes to ensure scorer accuracy and detect drift. The objective of the scoring process is to make sure that scorers rate student responses in a manner consistent with ELPA21 standards, within a single administration of ELPA21, as well as across multiple administrations.

The validity selection process involved MI scoring staff selecting 30–75 responses per item from live responses from the current administration to serve as validity responses. Validity responses were selected to illustrate trends identified by leadership in live responses but not strongly reflected in the anchor sets, represent particular types of responses identified as challenging to score during training, and assess transfer of scorers’ knowledge of the anchor responses. Vetting of new validity responses involved identification and recommendation by team leaders while conducting read-behinds/listen-behinds, review and approval by scoring directors, and review and approval by the scoring project director.

The validity responses were used during handscoring to verify scorer accuracy. Validity responses were dispersed intermittently to the scorers throughout scoring at a rate of at least 10% of the total responses. These validity responses were blind reads, meaning that scorers saw these responses the same as they saw the actual live student responses; there was no distinguishable difference. This helped ensure the internal validity of the process. All scorers who received validity responses had already successfully completed the training and qualifying process.

Next, the scores that the scorers assigned to the validity responses were compared to the predetermined scores in order to determine the validity of the scorers’ scores. For each item, the percentage of exact agreement and the percentage of high and low scores were computed. The same data were also computed for each specific scorer. Using these pieces of data, various validity reports could be produced in real time and used to monitor for potential drift.

If results indicated that there was drift, or shifts in scoring over time, for a particular response, item, or scorer, action was taken to correct it. This was in the form of individual scorer retraining, room-wide retraining/recalibration, and/or rescoring responses where it was determined a scorer had been errantly assigning scores. Sometimes, when a particular validity response generated low agreement, an example of a similar response could be found in the existing training materials. If this was the case, a review of that particular training response was pursued in order to realign the scorer.

Recalibration sets consisting of a validation set representing a variety of score points in random score point order were also used to maintain consistency. Sets varied in size from three to five responses based on particular issues observed during scoring. The recalibration sets were distributed at the beginning of the morning on a weekly basis. MI also recalibrated approximately once a week with

scorers who had missed a required day’s scoring session and were required to recalibrate. Those scorers achieving less than an acceptable percentage of correct scores on these responses were monitored closely throughout that day. Scorers who did not demonstrate improvement received personal and extensive retraining. These scorers continued to be monitored on an individual basis until the next recalibration round took place.

By implementing these scoring procedures—using the same training materials whenever possible, utilizing a suite of real-time reports, and making training decisions based on report data—MI maximized scoring reliability and validity.

9.2.4 Quality Control on Final Scores

AIR’s scoring engine is used to produce final scores upon receiving handscores. Before operational scoring, AIR created mock-ups of student records to verify the accuracy of the scoring engine. Both AIR’s analysis team (responsible for the scoring engine) and psychometricians independently computed scores on the mock-ups of student records. They compared their results iteratively until a 100% match was reached.

During operational scoring, AIR’s psychometricians independently scored students and compared the scores with the results from the scoring engine. Discrepancies were iteratively resolved until a 100% match was reached.

Before final scores were delivered to the state, they were also compared with the unofficial scores from CRESST, if needed. Discrepancies were again investigated and resolved until a 100% match was reached.

9.3 QUALITY ASSURANCE IN REPORTING

In 2018–2019, two types of score reports were produced for both summative and screener tests: online reports and printed reports (family reports only).

9.3.1 Online Report Quality Assurance

Every test undergoes a series of validation checks. Once the Quality Monitoring System (QMS) signs off, data are passed to the Database of Record (DoR), which serves as the centralized location for all student scores and responses, ensuring that there is only one place where the official record is stored. Only after scores have passed the QA checks and are uploaded to the DoR are they passed to the ORS, which is responsible for presenting individual-level results and calculating and presenting aggregate results. Absolutely no score is reported in the Online Reporting System (ORS) until it passes all of the QA system’s validation checks.

9.3.2 Paper Report Quality Assurance

Statistical Programming

The family reports contain custom programming and require rigorous quality assurance processes to ensure their accuracy. All custom programming is guided by detailed and precise specifications in our reporting specifications document. Upon approval of the specifications, analytic rules are

programmed and each program is extensively tested on test decks and real data from other programs. The final programs are reviewed by two senior statisticians and one senior programmer to ensure that they implement agreed-upon procedures. Custom programming is implemented independently by two statistical programming teams working from the specifications. Only when the output from both teams matches exactly are the scripts released for production. Quality control, however, does not stop there.

Much of the statistical processing is repeated, and AIR has implemented a structured software development process to ensure that the repeated tasks are implemented correctly and identically each time. AIR's software developers write small programs called macros that take specified data as input and produce data sets containing derived variables as output. Approximately 30 such macros reside in AIR's library. Each macro is extensively tested and stored in a central development server. Once a macro is tested and stored, changes to the macro must be approved by the director of score reporting and the director of psychometrics, as well as by the project directors for affected projects.

Each change is followed by a complete retesting with the entire collection of scenarios on which the macro was originally tested. The main statistical program is made up mostly of calls to various macros, including macros that read in and verify the data and conversion tables and macros that do the many complex calculations. This program is developed and tested using artificial data generated to test both typical and extreme cases. In addition, the program goes through a rigorous code review by a senior statistician.

Display Programming

The paper report development process uses graphical programming, which takes place in a Xerox-developed programming language called Variable Data Intelligent PostScript Printware (VIPP) and allows virtually infinite control of the visual appearance of the reports. After designers at AIR create backgrounds, our VIPP programmers write code that indicates where to place all variable information (data, graphics, and text) on the reports. The VIPP code is tested using both artificial and real data. AIR's data generation utilities can read the output layout specifications and generate artificial data for direct input into the VIPP programs. This allows the testing of these programs to begin before the statistical programming is complete. In later stages, artificial data are generated according to the input layout and run through the score reporting statistical programs, and the output is formatted as VIPP input. This enables us to test the entire system. Programmed output goes through multiple stages of review and revision by graphics editors and the score reporting team to ensure that design elements are accurately reproduced and data are correctly displayed.

Once we receive final data and VIPP programs, the AIR score reporting team reviews proofs that contain actual data based on our standard quality assurance documentation. In addition, we compare data independently calculated by AIR psychometricians with data on the reports. A large sample of reports is reviewed by several AIR staff members to make sure that all data are correctly placed on reports. This rigorous review is typically conducted over several days and takes place in a secure location at AIR. All reports containing actual data are stored in a locked storage area. Prior to printing the reports, AIR provides a live data file and individual student reports with sample districts for the state staff review. AIR works closely with each state to resolve questions and correct any problems. The reports are not delivered until the state approves the sample reports and data file.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Institutes for Research. (2017). *Additional rules for 2019 spring ELPA21 summative test scoring*. Washington, DC: American Institutes for Research.
- Brown, W. (1910). Some Experimental Results in the Correlation of Mental Abilities. *British Journal of Psychology*, 3, 296–322.
- Cook, K. F., Kallen, M., & Amtmann, D. (2009). Having a fit: Impact of number of items and non-normality on tests of IRT's unidimensionality assumption. *Quality of Life Research*, 18(4), 447–460.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston
- Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on Coefficient Alpha and successor procedures. *Educational and Psychological Measurement*, 64(3), 391–418.
- Feldt, L. S., & Brennan, R. (1989). *Reliability*. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.), (pp. 105–146). New York: Macmillan.
- Feldt, L. S., & Qualls, A. L. (1996). Bias in coefficient Alpha arising from heterogeneity of test content. *Applied Measurement in Education*, 9, 277–286.
- Jolliffe, I. (2002). *Principal component analysis (2nd ed.)*. New York: Springer.
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160.
- Messick, S. (1989). *Validity*. In R.L. Linn (Ed.), *Educational Measurement*. New York: Macmillan.
- Nunnally, J. C. (2nd ed.). (1978). *Psychometric Theory*. New York: McGraw-Hill.
- Peter, J. P. (1979). Reliability: A review of Psychometric basics and recent marketing practices. *Journal of Marketing Research*, 16(1), 6–17.
- Peterson, R. A. (1994). A meta-analysis of Cronbach's Coefficient Alpha. *Journal of Consumer Research*, 21, 381–391.
- Qualls, A. L. (1995). Estimating the reliability of a test containing multiple item formats. *Applied Measurement in Education*, 8, 111–120.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores (Series 17) *Psychometric Monographs*. Richmond, VA: Psychometric Society.
- Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28(3), 234–247.

Spearman, C. (1910). Correlation Calculated From Faulty Data. *British Journal of Psychology*, 3, 271–295.

Table of Contents

Section 1: Summative Assessment - Accommodation Summary	1
Table S1.1: 2019 Accommodation Table	2
Section 2: Summative Assessment - Scale Score Summary	6
Table S2.1: Scale Score Summary by Subgroup - Kindergarten	7
Table S2.2: Scale Score Summary by Subgroup - Grade 1	9
Table S2.3: Scale Score Summary by Subgroup - Grade 2	11
Table S2.4: Scale Score Summary by Subgroup - Grade 3	13
Table S2.5: Scale Score Summary by Subgroup - Grade 4	15
Table S2.6: Scale Score Summary by Subgroup - Grade 5	17
Table S2.7: Scale Score Summary by Subgroup - Grade 6	19
Table S2.8: Scale Score Summary by Subgroup - Grade 7	21
Table S2.9: Scale Score Summary by Subgroup - Grade 8	23
Table S2.10: Scale Score Summary by Subgroup - Grade 9	25
Table S2.11: Scale Score Summary by Subgroup - Grade 10	27
Table S2.12: Scale Score Summary by Subgroup - Grade 11	29
Table S2.13: Scale Score Summary by Subgroup - Grade 12	31
Section 3: Summative Assessment - Percentage of Students by Domain Performance Level	33
Table S3.1: Performance Level by Domain - Kindergarten	34
Table S3.1: Performance Level by Domain - Kindergarten(cont.)	36
Table S3.2: Performance Level by Domain - Grade 1	38
Table S3.2: Performance Level by Domain - Grade 1(cont.)	40
Table S3.3: Performance Level by Domain - Grade 2	42
Table S3.3: Performance Level by Domain - Grade 2(cont.)	44
Table S3.4: Performance Level by Domain - Grade 3	46
Table S3.4: Performance Level by Domain - Grade 3(cont.)	48
Table S3.5: Performance Level by Domain - Grade 4	50
Table S3.5: Performance Level by Domain - Grade 4(cont.)	52
Table S3.6: Performance Level by Domain - Grade 5	54
Table S3.6: Performance Level by Domain - Grade 5(cont.)	56
Table S3.7: Performance Level by Domain - Grade 6	58
Table S3.7: Performance Level by Domain - Grade 6(cont.)	60
Table S3.8: Performance Level by Domain - Grade 7	62
Table S3.8: Performance Level by Domain - Grade 7(cont.)	64

Table S3.9: Performance Level by Domain - Grade 8.....	66
Table S3.9: Performance Level by Domain - Grade 8(cont.).....	68
Table S3.10: Performance Level by Domain - Grade 9.....	70
Table S3.10: Performance Level by Domain - Grade 9(cont.).....	72
Table S3.11: Performance Level by Domain - Grade 10.....	74
Table S3.11: Performance Level by Domain - Grade 10(cont.).....	76
Table S3.12: Performance Level by Domain - Grade 11.....	78
Table S3.12: Performance Level by Domain - Grade 11(cont.).....	80
Table S3.13: Performance Level by Domain - Grade 12.....	82
Table S3.13: Performance Level by Domain - Grade 12(cont.).....	84
Section 4: Summative Assessment - Percentage of Students by Overall Proficiency Level.....	86
Table S4.1: Overall Proficiency Levels - Kindergarten.....	87
Table S4.2: Overall Proficiency Levels - Grade 1.....	89
Table S4.3: Overall Proficiency Levels - Grade 2.....	91
Table S4.4: Overall Proficiency Levels - Grade 3.....	93
Table S4.5: Overall Proficiency Levels - Grade 4.....	95
Table S4.6: Overall Proficiency Levels - Grade 5.....	97
Table S4.7: Overall Proficiency Levels - Grade 6.....	99
Table S4.8: Overall Proficiency Levels - Grade 7.....	101
Table S4.9: Overall Proficiency Levels - Grade 8.....	103
Table S4.10: Overall Proficiency Levels - Grade 9.....	105
Table S4.11: Overall Proficiency Levels - Grade 10.....	107
Table S4.12: Overall Proficiency Levels - Grade 11.....	109
Table S4.13: Overall Proficiency Levels - Grade 12.....	111
Section 5: Summative Assessment - Marginal Reliability.....	112
Figure S5.1: Marginal Reliability by Gender.....	113
Figure S5.2: Marginal Reliability by Ethnicity.....	114
Figure S5.3: Marginal Reliability by IEP.....	115
Figure S5.4: Marginal Reliability by Primary Disability.....	116
Section 6: Summative Assessment - Conditional Standard Error of Measurement (CSEM).....	117
Figure S6.1: Conditional Standard Error of Measurement for Kindergarten.....	118
Figure S6.2: Conditional Standard Error of Measurement for Grade 1.....	119
Figure S6.3: Conditional Standard Error of Measurement for Grade 2.....	120
Figure S6.4: Conditional Standard Error of Measurement for Grade 3.....	121

Figure S6.5: Conditional Standard Error of Measurement for Grade 4.....	122
Figure S6.6: Conditional Standard Error of Measurement for Grade 5.....	123
Figure S6.7: Conditional Standard Error of Measurement for Grade 6.....	124
Figure S6.8: Conditional Standard Error of Measurement for Grade 7.....	125
Figure S6.9: Conditional Standard Error of Measurement for Grade 8.....	126
Figure S6.10: Conditional Standard Error of Measurement for Grade 9.....	127
Figure S6.11: Conditional Standard Error of Measurement for Grade 10.....	128
Figure S6.12: Conditional Standard Error of Measurement for Grade 11.....	129
Figure S6.13: Conditional Standard Error of Measurement for Grade 12.....	130
Section 7: Summative Assessment - Dimensionality	131
Figure S7.1: Validity - Dimensionality for Kindergarten.....	132
Figure S7.2: Validity - Dimensionality for Grade 1	133
Figure S7.3: Validity - Dimensionality for Grade Band 2-3	134
Figure S7.4: Validity - Dimensionality for Grade Band 4-5	135
Figure S7.5: Validity - Dimensionality for Grade Band 6-8	136
Figure S7.6: Validity - Dimensionality for Grade Band 9-12	137
Section 8: Summative Assessment - Ability vs. Difficulty	138
Figure S8.1: Validity - Ability vs. Difficulty for Kindergarten.....	139
Figure S8.2: Validity - Ability vs. Difficulty for Grade 1	140
Figure S8.3: Validity - Ability vs. Difficulty for Grade Band 2-3	141
Figure S8.4: Validity - Ability vs. Difficulty for Grade Band 4-5	142
Figure S8.5: Validity - Ability vs. Difficulty for Grade Band 6-8	143
Figure S8.6: Validity - Ability vs. Difficulty for Grade Band 9-12	144
Section 9: Summative Assessment - Inter-Rater Analysis	145
Table S9.1: Inter-Rater Result - Kindergarten.....	146
Table S9.2: Inter-Rater Result - Grade 1	147
Table S9.3: Inter-Rater Result - Grade Band 2-3	148
Table S9.4: Inter-Rater Result - Grade Band 4-5	149
Table S9.5: Inter-Rater Result - Grade Band 6-8	150
Table S9.6: Inter-Rater Result - Grade Band 9-12	151
Section 10: Student Progress from Screener to Summative - Scatter Plots of Scale Score Change	152
Figure S10.1: Scatter Plot of Scale Score Change from Screener to Summative for Grade K ..	153
Figure S10.2: Scatter Plot of Scale Score Change from Screener to Summative for Grade 1 ...	154

Figure S10.3: Scatter Plot of Scale Score Change from Screener to Summative for Grade Band 2-3	155
Figure S10.4: Scatter Plot of Scale Score Change from Screener to Summative for Grade Band 4-5	156
Figure S10.5: Scatter Plot of Scale Score Change from Screener to Summative for Grade Band 6-8	157
Figure S10.6: Scatter Plot of Scale Score Change from Screener to Summative for Grade Band 9-12	158
Section 11: Student Progress from Screener to Summative - Comparison of Scale Scores.....	159
Table S11.1: Comparison of Scale Scores Between Screener and Summative - Listening	160
Table S11.2: Comparison of Scale Scores Between Screener and Summative - Reading	161
Table S11.3: Comparison of Scale Scores Between Screener and Summative - Speaking.....	162
Table S11.4: Comparison of Scale Scores Between Screener and Summative - Writing	163
Table S11.5: Comparison of Scale Scores Between Screener and Summative - Comprehension	164
Table S11.6: Comparison of Scale Scores Between Screener and Summative - Overall.....	165

Section 1: Summative Assessment - Accommodation Summary

Table SI.1: 2019 Accommodation Table

Type	Value	K	1	2	3	4	5	6	7	8	9	10	11	12	Total
Audio Playback Controls	Audio Scrubber	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
	Inverted (White on Black)				1	1	1								3
	Light Yellow											1			1
Color Choices	None	9092	7960	6870	5693	4668	3673	3424	3403	3359	4331	3500	2783	2076	60832
	White on Navy				1	1			1						3
Cover Pages Zooming	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Enhanced Contrast	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Enhanced Security Options	ESO DetectBrowserBreach	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Exempted Domains	Listening	1	1	5	1	3	3	7	4	11	6	6	1	4	53
	No Exemptions	9114	7975	6845	5665	4634	3634	3387	3374	3330	4292	3467	2762	2053	60532
	Reading	41	51	52	46	55	44	36	32	29	40	33	22	20	501
	Speaking	47	37	45	39	39	28	25	24	23	22	23	16	17	385
	Writing	41	51	51	42	50	39	39	31	35	36	32	20	21	488
Expandable Items	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Expandable Passages	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Font Type	Verdana	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Global Notes	Off													1	1
	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
HTML Editor Button Groups	SpellChecker Disabled & Print Disabled	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Hardware Checks	Check Playback Capabilities & Check Playback Vol Capabilities	9158	8027	6904	5715	4686	3683	3429	3408	3368	4318	3497	2769	2070	61032

Type	Value	K	1	2	3	4	5	6	7	8	9	10	11	12	Total
	Check Recording Capabilities&Check Playback Capabilities	9061	7954	6808	5645	4612	3630	3379	3359	3295	4236	3438	2740	2036	60193
Highlight	Off	1	1		2	1			2				1		8
	On	9160	8028	6904	5714	4691	3687	3432	3409	3371	4334	3504	2785	2077	61096
IntraTest TTS Controls	Allow Volume, Pitch, and Rate Adjustments	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Item Badges	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Item Font Size	16pt	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Item Response Recovery	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Item Tools Menu	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
ItemSecurity	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Language	English	9170	8036	6904	5716	4692	3687	3432	3411	3372	4334	3504	2785	2077	61120
Line Reader	Off	1			2	1			2				1		7
	On	9091	7960	6870	5691	4667	3673	3424	3402	3359	4331	3501	2783	2076	60828
Mark for Review	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Masking	Off	1	1		2	2			2				1		9
	On	9160	8028	6904	5714	4690	3687	3432	3409	3371	4334	3504	2785	2077	61095
MathML Rendering	Disabled	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
	Extra Large Black											1			1
	Large Black			1					2						3
Mouse Pointer	Large Red								1						1
	System Default	9161	8029	6903	5716	4692	3687	3432	3409	3371	4334	3503	2785	2077	61099
Navigation Dropdown	Show Question Numbers	9092	7960	6870	5693	4668	3673	3424	3404	3359	4331	3501	2783	2076	60834
Paginate Item Groups	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103

Type	Value	K	1	2	3	4	5	6	7	8	9	10	11	12	Total
Passage Font Size	16pt	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Pause Test Button	Legacy Pause Button	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Permissive Mode	Off	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
	Level 1			1		2					1				4
	Level 2		1					1							2
Print Size	Level 3							1	2			1			4
	Level 4	2		1		2	2				1		1		9
	No Default Zoom Applied	9161	8029	6903	5716	4692	3687	3431	3410	3371	4333	3503	2784	2077	61097
Require Page Visit	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Review Screen Layout	List	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Streamlined Mode	Off	9092	7960	6870	5693	4668	3673	3424	3404	3359	4331	3501	2783	2076	60834
Strikethrough	Off	1	1		2	1			2				1		8
	On	9160	8028	6904	5714	4691	3687	3432	3409	3371	4334	3504	2785	2077	61096
System Volume Control	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
TTS Audio Adjustments	Allow TTS Volume Adjustments&Allow TTS Pitch Adjustments&Allow TTS Rate Adjustments&Allow Voice Pack Selection	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
TTX Business Rules	A203	9092	7960	6870	5693	4668	3673	3424	3404	3359	4331	3501	2783	2076	60834
Test Notifications	Do Not Show Notifications	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Test Progress Indicator	Show indicator as a fraction	9092	7960	6870	5693	4668	3673	3424	3404	3359	4331	3501	2783	2076	60834
Tutorial	On	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103
Video Playback Controls	Play and Pause&Mute and	9161	8029	6904	5716	4692	3687	3432	3411	3371	4334	3504	2785	2077	61103

Type	Value	K	1	2	3	4	5	6	7	8	9	10	11	12	Total
	Unmute&Video Scrubber&Fullscree n View Mode														

Section 2: Summative Assessment - Scale Score Summary

* The mean and SD are suppressed when the N count is fewer than 10.

Table S2.1: Scale Score Summary by Subgroup - Kindergarten

	Listening			Reading			Speaking			Writing				
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD		
Domain Total	9144	556	78.2	9077	558	74.3	9025	578	76.3	9068	544	76.8		
Gender	Female	4195	562	77.6	4174	564	74.2	4149	586	74.2	4166	549	75.4	
	Male	4507	551	77.0	4462	554	72.7	4443	573	74.9	4459	540	76.8	
	Missing	442	538	89.5	441	543	85.2	433	555	97.6	443	530	84.5	
	African American	1694	553	75.9	1679	554	72.8	1675	584	70.3	1680	542	76.7	
Ethnicity	American Indian or Alaskan Native	27	550	67.2	26	553	62.8	25	591	50.1	26	552	81.6	
	Asian	2201	572	81.2	2188	575	77.4	2176	588	76.3	2186	571	80.4	
	Hispanic or Latino	2422	544	72.5	2405	547	68.3	2393	569	74.2	2395	524	67.8	
	Multiple Races	1004	542	76.5	997	543	72.1	985	565	80.5	998	523	70.3	
	Native Hawaiian or Other Pacific Islander	40	555	67.4	40	561	66.7	40	573	71.2	40	542	59.8	
	Other/Unknown	41	584	71.3	40	580	70.3	40	606	58.7	40	577	81.9	
	White	1417	568	76.4	1405	569	72.4	1400	587	72.1	1404	555	73.0	
	Missing	298	526	94.0	297	533	88.2	291	538	103.0	299	521	85.9	
	Other	Gifted	14	624	85.1	14	633	74.1	14	625	70.2	14	612	69.6
		IEP	458	502	83.2	401	522	69.1	401	520	82.0	405	508	78.6
Migrant		10	536	65.9	9			10	546	79.2	10	550	75.3	
Autism		93	477	93.8	67	510	78.7	66	480	95.6	69	514	100.1	
Primary Disability	Developmental delay	6			3			3			3			
	Emotional disturbance	6			5			6			7			

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Hearing Impaired	9			8			8			8		
	Intellectual Disability	22	443	75.3	20	454	76.2	20	460	90.6	20	436	86.7
	Multiple disabilities	28	395	86.9	9			11	396	84.9	9		
	Orthopedic impairment	8			6			6			6		
	Other health impairment	37	518	81.8	34	531	69.5	34	527	79.9	34	502	60.3
	Specific learning disability	33	511	58.0	33	520	52.2	32	534	57.7	33	490	62.2
	Speech or language impairment	205	531	65.5	205	538	60.1	205	541	66.0	205	521	69.2
	Traumatic brain injury	1			1			1			1		
	Visual impairment	6			6			6			6		
Not Attempted		25			52			101			61		
Exempted		1			41			44			41		
Overall Score		9162	5565	562.4									
Comprehension Score		9150	5532	545.2									

Table S2.2: Scale Score Summary by Subgroup - Grade 1

	Listening			Reading			Speaking			Writing			
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	
Domain Total	8013	554	69.4	7948	550	75.0	7935	577	66.3	7947	541	72.7	
Gender	Female	3711	560	67.3	3692	557	74.6	3689	584	62.9	3690	548	70.6
	Male	4094	551	68.9	4048	546	74.0	4043	573	65.4	4046	537	72.4
	Missing	208	515	95.3	208	518	87.5	203	532	107.3	211	502	92.4
Ethnicity	African American	1520	551	67.1	1509	550	71.5	1505	581	59.1	1512	539	69.6
	American Indian or Alaskan Native	20	556	67.8	20	553	75.8	20	579	76.5	20	542	70.0
	Asian	1811	568	70.7	1799	577	77.6	1797	591	68.3	1800	565	71.7
	Hispanic or Latino	2265	547	65.1	2251	533	68.2	2252	569	63.9	2245	526	69.0
	Multiple Races	892	545	69.8	874	531	70.7	871	568	66.3	871	525	70.9
	Native Hawaiian or Other Pacific Islander	29	566	56.8	29	557	72.8	29	587	46.2	29	549	68.7
	Other/Unknown	25	577	63.0	25	571	66.3	25	584	53.8	25	567	62.1
	White	1298	561	67.8	1288	560	74.2	1288	583	61.4	1290	552	70.5
	Missing	153	503	99.7	153	511	88.6	148	516	114.0	155	495	94.6
	Gifted	38	619	54.9	38	633	79.4	38	617	56.6	38	612	73.8
Other	IEP	488	503	82.5	429	508	75.7	438	530	83.1	430	493	79.7
	Migrant	15	534	96.0	15	525	74.4	15	540	85.8	15	512	85.1
	Autism	81	454	82.3	59	503	66.7	62	482	91.4	59	486	66.3
Primary Disability	Developmental delay	2			2			2			2		
	Emotional disturbance	4			4			4			4		

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Hearing Impaired	4			4			4			4		
	Intellectual Disability	37	446	71.5	33	453	76.5	34	466	88.5	34	424	87.2
	Multiple disabilities	29	390	58.7	12	437	59.0	12	396	89.9	10	369	45.2
	Orthopedic impairment	15	537	97.2	10	544	122.9	11	563	102.5	9		
	Other health impairment	52	506	65.5	49	490	66.2	52	525	73.3	50	471	71.5
	Specific learning disability	73	517	56.9	68	496	64.7	69	534	64.4	68	480	66.7
	Speech or language impairment	184	544	66.1	181	532	72.7	181	565	61.2	183	523	71.2
	Traumatic brain injury	2			2			2			2		
	Visual impairment	3			3			3			3		
Not Attempted		22			39			64			40		
Exempted		1			49			37			49		
Overall Score		8027	5544	542.6									
Comprehension Score		8019	5530	516.8									

Table S2.3: Scale Score Summary by Subgroup - Grade 2

	Listening			Reading			Speaking			Writing				
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD		
Domain Total	6884	540	65.6	6815	526	67.2	6790	545	66.0	6808	523	70.9		
Gender	Female	3118	544	65.0	3094	533	67.7	3085	550	65.8	3090	531	69.9	
	Male	3649	536	66.0	3604	521	66.3	3588	541	66.2	3601	516	71.2	
	Missing	117	537	58.5	117	514	62.5	117	530	54.0	117	512	65.3	
Ethnicity	African American	1364	535	65.6	1355	522	65.3	1352	540	62.8	1347	517	69.9	
	American Indian or Alaskan Native	12	531	61.0	12	523	63.8	12	545	59.1	12	516	61.0	
	Asian	1389	550	67.3	1376	545	69.6	1369	556	67.0	1374	541	70.4	
	Hispanic or Latino	2155	535	63.2	2130	516	64.6	2124	539	66.4	2133	514	69.6	
	Multiple Races	846	533	69.5	833	517	67.7	831	536	69.3	834	512	73.5	
	Native Hawaiian or Other Pacific Islander	41	526	56.2	41	510	55.4	41	538	59.6	41	519	66.4	
	Other/Unknown	18	546	87.5	18	534	78.3	18	538	66.4	18	536	87.7	
	White	1059	549	62.3	1050	536	65.0	1043	557	61.9	1049	535	67.2	
	Other	Gifted	38	603	52.9	38	602	65.8	37	586	58.0	38	597	61.3
		IEP	570	491	73.5	509	477	59.9	505	503	67.1	508	467	69.0
Migrant		29	527	58.2	29	506	48.3	28	541	54.6	28	508	51.4	
Autism		78	453	90.9	55	479	74.6	51	475	81.4	51	473	79.8	
Primary Disability	Deaf-blindness	1			1			1			1			
	Developmental delay	3			2			2			2			
	Emotional disturbance	17	500	36.2	16	467	44.4	16	506	37.0	16	442	59.1	

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Hearing Impaired	8			9			8			9		
	Intellectual Disability	34	462	52.5	30	451	55.6	31	460	56.0	30	435	63.0
	Multiple disabilities	32	379	51.7	12	398	27.9	13	399	63.8	14	384	34.4
	Orthopedic impairment	6			6			6			6		
	Other health impairment	61	475	57.7	55	455	40.3	57	489	58.4	57	439	60.6
	Specific learning disability	191	503	51.2	188	468	45.4	187	506	51.2	186	459	54.6
	Speech or language impairment	129	537	59.2	128	509	61.2	127	533	65.0	129	506	67.2
	Traumatic brain injury	3			3			2			3		
	Visual impairment	1			0			0			0		
Not Attempted		15			38			69			46		
Exempted		5			51			45			50		
Overall Score		6899	5360	521.0									
Comprehension Score		6895	5382	488.7									

Table S2.4: Scale Score Summary by Subgroup - Grade 3

	Listening			Reading			Speaking			Writing				
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD		
Domain Total	5704	570	69.9	5638	566	71.6	5628	574	69.9	5636	562	72.3		
Gender	Female	2671	572	70.3	2644	571	71.4	2646	577	69.4	2644	569	71.4	
	Male	3012	569	69.0	2973	561	71.2	2963	570	70.0	2971	556	72.3	
	Missing	21	554	113.8	21	557	102.8	19	582	82.4	21	550	108.2	
Ethnicity	African American	1118	571	67.0	1112	566	69.8	1110	573	65.7	1109	562	69.8	
	American Indian or Alaskan Native	12	578	106.6	11	590	82.6	11	603	52.3	11	588	67.2	
	Asian	962	577	71.3	948	581	70.2	949	582	69.6	952	575	70.7	
	Hispanic or Latino	1907	566	70.3	1886	559	71.4	1882	570	70.7	1886	555	72.5	
	Multiple Races	837	566	70.8	823	557	73.8	819	564	74.0	822	553	76.1	
	Native Hawaiian or Other Pacific Islander	28	563	55.3	28	569	60.5	28	588	60.4	28	573	54.7	
	Other/Unknown	6			6			6			6			
	White	834	576	69.1	824	572	71.0	823	582	68.2	822	569	71.0	
	Other	Gifted	36	628	46.7	36	633	54.1	36	632	49.6	36	623	56.5
		IEP	770	531	73.1	705	515	68.1	717	538	69.8	713	512	71.7
Migrant		34	551	80.4	34	547	79.3	33	558	62.2	33	544	81.9	
Autism		82	479	94.9	64	502	81.8	61	496	85.2	63	498	85.4	
Primary Disability	Deaf-blindness	1			1			1			1			
	Developmental delay	8			8			8			8			
	Emotional disturbance	13	564	64.9	12	548	61.7	13	552	73.9	12	537	75.8	

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Hearing Impaired	10	532	88.7	11	526	66.8	11	531	80.6	11	526	58.7
	Intellectual Disability	47	484	63.4	40	456	44.5	41	485	66.5	41	447	54.9
	Multiple disabilities	23	386	65.7	9			9			10	376	88.8
	Orthopedic impairment	5			5			5			5		
	Other health impairment	87	527	62.4	82	511	64.6	84	526	77.1	81	506	73.2
	Specific learning disability	387	547	52.3	370	515	58.0	382	548	57.0	378	513	60.3
	Speech or language impairment	109	578	52.4	107	571	58.2	106	580	46.0	107	568	52.6
	Traumatic brain injury	3			2			2			2		
	Visual impairment	2			1			1			1		
Not Attempted		11			33			50			39		
Exempted		1			45			38			41		
Overall Score		5715	5639	553.9									
Comprehension Score		5712	5643	530.3									

Table S2.5: Scale Score Summary by Subgroup - Grade 4

	Listening			Reading			Speaking			Writing				
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD		
Domain Total	4669	528	73.6	4591	530	69.8	4603	548	72.9	4598	524	74.3		
Gender	Female	2107	531	75.2	2085	532	70.0	2085	552	73.3	2085	529	74.9	
	Male	2494	525	72.1	2438	527	69.3	2450	544	72.3	2445	519	73.4	
	Missing	68	543	77.4	68	545	77.2	68	563	75.3	68	541	74.2	
Ethnicity	African American	877	525	74.0	866	527	71.6	866	545	73.4	865	521	74.9	
	American Indian or Alaskan Native	9			9			9			9			
	Asian	747	534	75.9	736	537	74.6	735	550	74.9	738	530	77.6	
	Hispanic or Latino	1660	524	73.9	1624	526	66.7	1632	544	73.9	1627	520	73.3	
	Multiple Races	687	528	72.8	678	527	69.2	680	547	72.2	678	520	74.9	
	Native Hawaiian or Other Pacific Islander	27	517	73.1	26	521	72.0	27	538	78.6	26	517	78.6	
	Other/Unknown	8			8			8			8			
	White	654	537	69.6	644	537	68.4	646	560	65.0	647	534	69.7	
	Other	Gifted	12	605	44.7	12	603	69.6	12	614	35.1	12	590	62.1
		IEP	749	490	71.3	676	491	59.8	694	523	66.9	683	483	68.4
Migrant		17	542	43.7	16	534	39.0	16	561	38.8	16	521	43.5	
Primary Disability	Autism	53	445	69.5	33	480	54.6	34	482	69.0	33	470	63.7	
	Developmental delay	5			6			5			5			
	Emotional disturbance	14	489	70.1	15	476	79.3	15	502	67.8	15	464	79.4	
	Hearing Impaired	5			6			5			6			

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Intellectual Disability	53	422	55.1	39	435	57.9	43	452	64.4	39	414	61.2
	Multiple disabilities	25	335	48.9	4			8			7		
	Orthopedic impairment	9			8			8			8		
	Other health impairment	86	494	70.7	79	487	58.7	82	522	64.1	79	474	71.4
	Specific learning disability	434	509	54.7	421	497	52.3	432	538	53.4	426	492	59.7
	Speech or language impairment	50	528	51.7	51	538	53.1	50	555	48.3	51	533	44.7
	Traumatic brain injury	7			6			6			6		
	Visual impairment	1			1			1			1		
	Not Attempted	20			47			50			44		
	Exempted	3			54			39			50		
	Overall Score	4684	5353	571.6									
	Comprehension Score	4677	5350	539.1									

Table S2.6: Scale Score Summary by Subgroup - Grade 5

	Listening			Reading			Speaking			Writing				
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD		
Domain Total	3670	540	79.0	3607	542	75.6	3621	554	76.5	3615	536	79.4		
Gender	Female	1596	543	78.9	1573	544	76.2	1580	558	77.3	1577	541	80.2	
	Male	1999	538	79.2	1960	540	75.4	1967	551	75.9	1964	532	78.8	
	Missing	75	545	74.1	74	553	64.9	74	567	68.8	74	549	71.0	
Ethnicity	African American	732	539	76.2	725	539	74.7	726	553	72.7	723	534	75.9	
	American Indian or Alaskan Native	9			9			9			9			
	Asian	567	540	86.9	557	544	83.0	550	554	78.7	558	538	84.1	
	Hispanic or Latino	1275	543	77.8	1253	546	72.5	1265	555	76.1	1260	539	77.7	
	Multiple Races	530	531	78.4	516	534	75.6	522	545	81.3	516	527	81.5	
	Native Hawaiian or Other Pacific Islander	19	533	69.8	19	523	76.5	19	539	68.6	19	529	85.4	
	Other/Unknown	6			6			6			6			
	White	532	544	77.7	522	544	75.1	524	562	75.4	524	541	80.3	
	Other	Gifted	12	586	68.3	12	575	77.8	12	599	49.3	12	579	61.2
		IEP	715	507	77.8	658	509	68.3	678	535	75.1	667	504	74.6
Migrant		12	556	80.3	12	550	83.3	11	566	91.8	12	546	84.5	
Primary Disability	Autism	49	422	88.5	33	447	63.7	34	439	96.1	34	441	79.0	
	Developmental delay	4			3			4			4			
	Emotional disturbance	17	496	68.3	16	502	68.3	15	490	72.9	16	493	69.5	
	Hearing Impaired	2			4			2			4			

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Intellectual Disability	60	437	75.9	52	444	67.6	55	465	81.7	51	434	75.4
	Multiple disabilities	22	357	60.2	14	389	36.2	10	395	66.8	12	352	53.7
	Orthopedic impairment	2			1			2			1		
	Other health impairment	92	517	63.0	87	518	64.0	90	534	65.2	89	505	72.9
	Specific learning disability	429	530	58.6	410	523	59.2	429	555	58.0	419	520	61.7
	Speech or language impairment	31	543	70.4	31	541	65.3	30	563	64.4	31	540	67.8
	Traumatic brain injury	4			4			4			4		
	Visual impairment	1			1			1			1		
	Not Attempted	14			39			38			37		
	Exempted	3			41			28			35		
	Overall Score	3682	5441	613.2									
	Comprehension Score	3676	5442	582.9									

Table S2.7: Scale Score Summary by Subgroup - Grade 6

	Listening			Reading			Speaking			Writing				
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD		
Domain Total	3415	527	67.4	3372	524	63.0	3368	551	68.9	3367	519	71.7		
Gender	Female	1477	531	67.6	1473	529	61.3	1465	555	69.4	1469	526	71.8	
	Male	1890	524	67.2	1851	521	64.1	1855	547	68.3	1850	514	71.1	
	Missing	48	520	67.2	48	516	63.6	48	547	71.7	48	511	74.7	
Ethnicity	African American	759	529	65.1	752	523	62.1	750	552	67.2	750	520	69.2	
	American Indian or Alaskan Native	4			4			4			4			
	Asian	514	535	68.7	506	535	66.5	504	557	68.3	504	526	70.7	
	Hispanic or Latino	1157	522	65.8	1142	519	59.9	1142	547	68.7	1140	517	70.6	
	Multiple Races	495	519	71.3	488	518	63.6	491	539	74.0	489	509	77.1	
	Native Hawaiian or Other Pacific Islander	22	497	56.2	22	492	50.9	22	521	56.1	22	492	60.4	
	Other/Unknown	8			8			8			8			
	White	456	538	68.2	450	535	65.2	447	564	65.7	450	530	72.6	
	Other	Gifted	13	580	43.1	13	581	45.6	13	590	48.3	13	580	43.2
		IEP	656	505	60.5	617	498	55.8	626	537	61.9	614	494	65.4
Migrant		11	527	74.8	11	523	68.4	11	559	60.1	11	514	73.8	
Autism		30	463	68.5	23	483	69.4	25	488	96.8	22	473	72.9	
Primary Disability		Developmental delay	6			5			5			5		
	Emotional disturbance	7			7			7			7			
	Hearing Impaired	3			8			4			6			

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Intellectual Disability	52	461	46.6	40	463	41.9	44	492	58.1	41	445	60.1
	Multiple disabilities	24	400	52.5	13	410	36.5	11	415	42.8	12	360	28.7
	Orthopedic impairment	2			2			3			2		
	Other health impairment	102	508	52.0	97	497	48.5	99	542	58.0	95	495	59.5
	Specific learning disability	385	517	50.6	377	504	49.8	384	549	51.0	380	504	57.3
	Speech or language impairment	29	551	60.2	29	549	62.9	28	565	41.1	29	542	53.7
	Traumatic brain injury	5			5			5			5		
	Visual impairment	4			4			4			4		
	Not Attempted	10			24			39			26		
	Exempted	7			36			25			39		
	Overall Score	3432	5341	526.9									
	Comprehension Score	3425	5324	482.0									

Table S2.8: Scale Score Summary by Subgroup - Grade 7

	Listening			Reading			Speaking			Writing			
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	
Domain Total	3398	538	74.9	3359	540	70.1	3329	558	74.0	3356	533	79.0	
Gender	Female	1467	543	76.3	1455	546	69.5	1442	563	75.8	1450	542	79.3
	Male	1872	535	73.7	1844	536	70.2	1829	555	72.5	1847	527	78.1
	Missing	59	526	69.7	60	522	70.6	58	543	69.5	59	513	78.3
Ethnicity	African American	714	544	73.7	710	546	70.5	703	562	69.2	709	538	78.4
	American Indian or Alaskan Native	5			6			6			5		
	Asian	515	543	73.6	509	544	73.4	502	562	72.4	510	533	76.5
	Hispanic or Latino	1201	532	75.1	1184	535	68.7	1177	552	78.4	1184	528	81.1
	Multiple Races	510	536	78.2	506	536	69.7	500	552	74.9	504	529	80.9
	Native Hawaiian or Other Pacific Islander	17	541	65.9	17	535	65.2	17	554	79.1	17	540	77.1
	Other/Unknown	9			9			9			9		
	White	427	546	73.2	418	549	68.8	415	572	67.2	418	543	74.2
	Gifted	9			9			9			9		
	Other	IEP	607	515	67.2	568	515	63.1	573	549	67.1	569	512
Migrant		12	532	55.6	11	535	64.6	12	544	50.4	11	514	58.1
Autism		28	491	97.2	23	529	81.9	21	538	69.7	24	501	96.9
Primary Disability	Developmental delay	7			6			7			6		
	Emotional disturbance	14	503	47.9	13	495	55.8	14	524	56.0	13	493	56.2
	Hearing Impaired	3			7			4			5		

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Intellectual Disability	51	472	56.9	37	473	46.3	41	496	75.7	37	460	66.2
	Multiple disabilities	27	407	77.6	17	425	88.5	14	417	111.8	15	399	100.5
	Orthopedic impairment	4			4			4			4		
	Other health impairment	71	531	51.9	68	530	56.1	69	561	57.6	69	533	51.1
	Specific learning disability	385	530	55.7	377	522	56.7	384	560	56.1	381	521	55.9
	Speech or language impairment	12	531	59.1	12	540	57.5	12	588	44.9	12	551	34.9
	Traumatic brain injury	3			2			1			1		
Not Attempted	10			20			59			25			
Exempted	3			32			23			30			
Overall Score		3410	5435	582.2									
Comprehension Score		3407	5425	538.2									

Table S2.9: Scale Score Summary by Subgroup - Grade 8

	Listening			Reading			Speaking			Writing			
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	
Domain Total	3342	553	81.2	3314	557	76.4	3276	569	77.4	3306	549	85.2	
Gender	Female	1440	557	82.0	1430	563	74.1	1412	574	75.7	1430	556	84.5
	Male	1860	550	80.1	1842	554	77.6	1822	565	78.1	1834	542	84.9
	Missing	42	549	95.1	42	555	87.0	42	562	90.7	42	542	102.2
Ethnicity	African American	772	550	79.6	775	553	76.8	760	566	74.2	771	543	85.4
	American Indian or Alaskan Native	4			4			4			4		
	Asian	557	564	78.8	546	570	76.8	547	580	73.9	546	559	79.1
	Hispanic or Latino	1071	550	82.1	1062	555	75.8	1058	563	80.8	1064	547	87.1
	Multiple Races	503	541	85.4	496	549	77.4	475	556	79.9	493	535	90.3
	Native Hawaiian or Other Pacific Islander	18	555	58.6	18	556	56.4	18	576	44.5	18	570	56.3
	Other/Unknown	6			6			6			6		
	White	411	563	78.5	407	566	73.9	408	585	73.3	404	563	79.4
	Gifted	8			8			8			8		
	Other	IEP	578	531	75.2	545	530	70.2	548	557	72.6	541	527
Migrant		11	530	110.6	11	540	89.2	11	535	107.3	11	522	108.2
Autism		28	477	96.2	19	508	89.1	19	524	94.9	19	477	112.0
Primary Disability	Deaf-blindness	0			1			0			0		
	Developmental delay	5			5			5			5		
	Emotional disturbance	12	522	68.3	11	522	69.3	12	514	85.6	12	502	93.2

	Listening			Reading			Speaking			Writing		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
Hearing Impaired	9			14	517	60.0	9			10	535	93.5
Intellectual Disability	53	477	50.7	44	477	50.2	48	501	70.1	46	471	65.1
Multiple disabilities	21	399	83.4	10	444	79.7	10	445	113.5	10	405	90.5
Orthopedic impairment	4			4			4			3		
Other health impairment	73	534	61.7	69	525	70.6	72	555	65.1	69	523	72.1
Specific learning disability	359	549	64.7	355	542	66.0	355	571	63.8	354	542	69.6
Speech or language impairment	7			7			7			7		
Traumatic brain injury	7			6			7			6		
Visual impairment	1			1			1			1		
Not Attempted	19			29			72			32		
Exempted	11			29			24			34		
Overall Score	3368	5549	628.5									
Comprehension Score	3362	5539	583.4									

Table S2.10: Scale Score Summary by Subgroup - Grade 9

	Listening			Reading			Speaking			Writing				
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD		
Domain Total	4289	520	76.2	4241	522	71.2	4161	548	70.9	4233	518	80.4		
Gender	Female	1813	523	76.3	1802	525	71.3	1767	550	69.7	1798	524	80.8	
	Male	2441	518	75.7	2408	521	71.0	2363	546	71.6	2405	513	79.5	
	Missing	35	480	90.6	31	501	73.4	31	524	80.3	30	478	91.8	
Ethnicity	African American	1021	517	72.4	1013	519	67.1	999	550	63.0	1008	519	73.4	
	American Indian or Alaskan Native	12	461	80.2	10	475	56.9	10	512	79.2	10	487	77.0	
	Asian	652	547	73.3	651	543	71.6	637	569	60.7	649	540	71.0	
	Hispanic or Latino	1365	512	79.1	1343	516	72.4	1311	537	79.2	1344	511	87.5	
	Multiple Races	749	509	70.7	744	513	67.1	725	538	68.5	741	504	80.0	
	Native Hawaiian or Other Pacific Islander	27	505	69.1	26	518	59.8	26	543	68.1	26	515	74.6	
	Other/Unknown	16	517	65.1	16	513	63.9	16	540	70.5	16	511	77.6	
	White	447	529	79.0	438	532	76.5	437	559	70.0	439	528	79.1	
	Other	Gifted	7			7			7			7		
		IEP	588	514	65.3	540	519	57.8	554	555	53.5	548	521	65.4
Primary Disability	Migrant	21	482	74.3	21	485	65.6	19	516	68.3	21	473	67.6	
	Autism	29	473	98.7	21	519	86.3	21	527	80.3	20	513	97.9	
	Developmental delay	7			5			5			5			
	Emotional disturbance	9			10	538	61.4	10	570	62.5	10	529	75.5	
	Hearing Impaired	4			6			4			5			

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Intellectual Disability	41	473	38.5	35	472	35.9	40	509	53.6	38	469	57.5
	Multiple disabilities	24	397	54.5	11	431	40.9	11	437	49.1	12	384	50.3
	Orthopedic impairment	2			0			0			0		
	Other health impairment	65	511	53.1	62	513	55.3	66	550	48.6	64	511	54.9
	Specific learning disability	391	529	54.5	375	525	54.4	382	565	44.9	379	532	58.0
	Speech or language impairment	12	566	45.4	12	567	53.1	12	588	20.2	12	571	44.1
	Traumatic brain injury	4			4			4			4		
	Visual impairment	3			2			2			2		
	Not Attempted	39			53			151			66		
	Exempted	6			40			22			35		
	Overall Score	4332	5306	586.3									
	Comprehension Score	4313	5282	544.0									

Table S2.11: Scale Score Summary by Subgroup - Grade 10

	Listening			Reading			Speaking			Writing				
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD		
Domain Total	3474	539	76.5	3449	540	73.9	3401	563	65.8	3442	537	76.3		
Gender	Female	1555	541	77.3	1545	541	73.0	1526	565	65.6	1540	541	76.3	
	Male	1908	538	75.7	1893	538	74.5	1865	562	65.7	1891	534	75.7	
	Missing	11	485	104.6	11	498	94.3	10	487	95.1	11	469	120.8	
Ethnicity	African American	853	532	73.1	855	532	68.7	842	561	63.6	851	533	69.6	
	American Indian or Alaskan Native	9			9			9			9			
	Asian	632	557	76.9	628	557	76.9	626	576	58.2	626	552	71.0	
	Hispanic or Latino	1093	530	79.8	1078	532	76.5	1064	554	74.3	1078	527	86.0	
	Multiple Races	454	542	70.6	447	543	68.7	432	564	60.5	448	540	69.8	
	Native Hawaiian or Other Pacific Islander	21	518	66.6	21	523	66.9	21	560	63.7	21	537	64.1	
	Other/Unknown	11	535	72.7	11	526	76.9	11	560	49.0	11	532	67.0	
	White	401	546	75.7	400	545	74.4	396	573	59.3	398	548	72.9	
	Other	Gifted	5			5			5			5		
		IEP	465	516	64.2	436	519	58.1	438	556	55.3	437	524	65.1
Primary Disability	Migrant	32	476	76.0	30	489	67.8	32	498	77.0	30	468	94.4	
	Autism	16	465	56.9	10	498	43.1	14	519	55.2	10	498	64.1	
	Developmental delay	12	465	73.1	13	458	67.7	12	499	70.4	12	439	80.0	
	Emotional disturbance	15	495	55.9	14	501	57.5	14	563	55.1	13	498	64.1	
	Hearing Impaired	4			9			4			8			

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Intellectual Disability	48	477	54.9	39	485	42.7	43	515	54.5	39	484	65.4
	Multiple disabilities	19	409	52.4	11	452	49.2	12	451	66.3	11	407	56.0
	Orthopedic impairment	3			2			2			2		
	Other health impairment	57	520	58.9	52	523	56.2	52	555	54.0	53	520	66.7
	Specific learning disability	287	535	54.1	283	531	55.2	282	572	42.6	286	541	52.4
	Speech or language impairment	2			2			2			2		
	Traumatic brain injury	3			3			3			3		
	Visual impairment	2			1			1			1		
	Not Attempted	24			23			80			32		
	Exempted	6			32			23			30		
	Overall Score	3504	5453	569.6									
	Comprehension Score	3494	5418	571.5									

Table S2.12: Scale Score Summary by Subgroup - Grade 11

	Listening			Reading			Speaking			Writing			
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	
Domain Total	2751	551	74.0	2733	550	73.6	2723	573	60.4	2733	547	71.8	
Gender	Female	1277	552	75.6	1266	550	73.6	1259	573	60.7	1263	550	73.4
	Male	1461	550	72.6	1454	550	73.6	1451	573	60.1	1457	544	70.5
	Missing	13	552	72.4	13	548	79.9	13	557	62.0	13	551	67.1
Ethnicity	African American	629	545	74.2	625	545	73.3	625	571	56.4	627	542	69.5
	American Indian or Alaskan Native	7			7			7			7		
	Asian	586	567	71.3	576	563	73.7	585	583	53.9	579	559	65.9
	Hispanic or Latino	852	538	75.2	847	538	72.8	838	559	66.3	845	535	77.4
	Multiple Races	323	556	69.3	322	556	71.1	317	579	59.3	321	552	65.4
	Native Hawaiian or Other Pacific Islander	14	563	79.7	14	568	83.7	14	584	50.8	14	563	80.9
	Other/Unknown	5			5			5			5		
	White	335	562	73.6	337	559	74.1	332	589	56.3	335	558	72.5
	Gifted	4			4			4			4		
	IEP	372	526	69.4	355	529	65.1	352	561	60.5	355	530	72.4
Other	Migrant	22	496	85.5	22	499	84.0	21	528	63.9	22	487	91.3
	Autism	11	525	106.1	9			10	557	92.4	9		
	Developmental delay	3			2			3			2		
	Emotional disturbance	13	524	61.6	13	521	65.0	13	550	37.6	13	512	58.9
	Hearing Impaired	1			3			1			3		
Primary Disability													

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Intellectual Disability	48	470	56.0	43	476	45.1	45	512	58.1	43	461	71.2
	Multiple disabilities	17	422	49.5	10	450	34.6	8			10	409	59.8
	Orthopedic impairment	3			2			2			2		
	Other health impairment	33	512	65.6	30	519	62.1	30	558	56.1	31	524	63.7
	Specific learning disability	233	547	59.4	233	543	61.0	230	575	51.5	233	550	60.3
	Speech or language impairment	2			2			2			2		
	Traumatic brain injury	3			3			3			2		
	Not Attempted	33			30			46			32		
Exempted	1			22			16			20			
Overall Score		2783	5541	541.6									
Comprehension Score		2765	5505	570.7									

Table S2.13: Scale Score Summary by Subgroup - Grade 12

	Listening			Reading			Speaking			Writing			
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	
Domain Total	2058	553	67.8	2040	551	68.1	2022	575	55.7	2031	549	63.5	
Gender	Female	967	554	65.9	960	551	64.9	945	575	55.0	955	552	62.3
	Male	1078	552	69.5	1067	551	71.0	1064	576	56.3	1063	548	64.4
	Missing	13	533	72.6	13	528	65.1	13	565	62.2	13	529	75.8
Ethnicity	African American	504	550	67.3	504	547	67.1	499	575	53.8	499	547	60.2
	American Indian or Alaskan Native	5			5			5			5		
	Asian	444	562	68.2	438	559	69.8	435	579	52.7	438	555	59.3
	Hispanic or Latino	625	544	68.0	619	544	67.2	613	566	60.0	617	541	68.3
	Multiple Races	199	556	63.8	197	555	65.7	194	576	53.2	196	554	61.7
	Native Hawaiian or Other Pacific Islander	9			9			9			9		
	Other/Unknown	5			5			5			5		
	White	267	562	68.2	263	557	69.7	262	590	52.8	262	559	64.4
	Gifted	2			2			2			2		
	Other	IEP	314	525	64.5	295	528	57.4	298	561	56.9	293	533
Migrant		20	545	83.3	20	542	87.1	20	573	48.5	20	543	82.0
Autism		13	515	94.0	11	535	87.9	13	541	75.9	11	546	85.1
Primary Disability	Developmental delay	2			2			2			2		
	Emotional disturbance	3			3			3			3		
	Hearing Impaired	1			4			1			2		

		Listening			Reading			Speaking			Writing		
		N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
	Intellectual Disability	31	474	38.6	27	483	34.3	28	517	42.9	28	482	46.0
	Multiple disabilities	19	420	55.7	10	446	37.4	11	450	64.8	10	433	60.1
	Other health impairment	35	525	64.3	35	523	56.7	34	555	70.5	33	531	53.4
	Specific learning disability	202	543	52.9	195	539	53.6	198	575	45.2	196	546	53.9
	Speech or language impairment	4			4			4			4		
	Traumatic brain injury	1			1			1			1		
Not Attempted		15			17			38			25		
Exempted		4			20			17			21		
Overall Score		2075	5556	491.9									
Comprehension Score		2067	5513	533.5									

Section 3: Summative Assessment - Percentage of Students by Domain Performance Level

Table S3.1: Performance Level by Domain - Kindergarten

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
Gender	Total	9170	13.0	14.3	48.3	10.3	13.9	0.3	0.0	9170	13.5	14.4	38.3	14.3	18.6	0.6	0.4
	Female	4202	11.5	13.5	47.7	11.0	16.1	0.2	0.0	4202	11.7	14.2	37.4	15.2	20.9	0.5	0.2
	Male	4525	13.5	15.2	48.9	9.9	12.2	0.4	0.0	4525	14.3	14.9	39.3	13.5	16.7	0.7	0.7
	Missing	443	21.9	11.7	47.4	8.1	10.6	0.2	0.0	443	22.1	11.3	36.1	14.0	16.0	0.5	0.0
Ethnicity	African American	1705	11.6	16.1	50.0	8.9	12.8	0.6	0.0	1705	13.2	16.4	38.9	13.4	16.6	1.1	0.5
	American Indian or Alaskan Native	27	14.8	7.4	63.0	7.4	7.4	0.0	0.0	27	11.1	7.4	51.9	18.5	7.4	0.0	3.7
	Asian	2206	10.4	12.5	43.9	12.3	20.7	0.2	0.0	2206	10.6	12.1	34.7	14.2	27.6	0.4	0.4
	Hispanic or Latino	2427	14.8	16.4	50.6	9.6	8.4	0.2	0.0	2427	15.1	15.7	41.9	14.1	12.2	0.5	0.4
	Multiple Races	1005	17.0	16.1	47.7	9.3	9.9	0.1	0.0	1005	17.9	17.6	37.5	12.9	13.2	0.4	0.4
	Native Hawaiian or Other Pacific Islander	40	12.5	5.0	52.5	20.0	10.0	0.0	0.0	40	15.0	7.5	45.0	12.5	20.0	0.0	0.0
	Other/Unknown	41	4.9	9.8	51.2	12.2	22.0	0.0	0.0	41	9.8	4.9	41.5	14.6	26.8	2.4	0.0
	White	1420	9.9	10.8	49.9	11.2	18.0	0.2	0.0	1420	9.7	12.5	37.0	17.3	22.5	0.4	0.6
	Missing	299	26.8	12.0	43.8	7.7	9.4	0.3	0.0	299	26.1	10.7	36.8	11.7	14.0	0.7	0.0
	Other	Gifted	14	7.1	0.0	28.6	21.4	42.9	0.0	0.0	14	7.1	0.0	7.1	28.6	57.1	0.0
	IEP	468	31.0	19.2	40.6	3.0	4.1	1.9	0.2	468	17.7	19.2	38.7	3.8	6.2	6.0	8.3
	Migrant	10	10.0	40.0	30.0	10.0	10.0	0.0	0.0	10	20.0	20.0	30.0	20.0	0.0	10.0	0.0
Primary Disability	Autism	97	44.3	13.4	33.0	1.0	4.1	3.1	1.0	97	19.6	14.4	25.8	4.1	5.2	16.5	14.4
	Developmental delay	6	83.3	0.0	16.7	0.0	0.0	0.0	0.0	6	33.3	0.0	16.7	0.0	0.0	0.0	50.0
	Emotional disturbance	7	14.3	28.6	42.9	0.0	0.0	14.3	0.0	7	28.6	14.3	28.6	0.0	0.0	28.6	0.0
	Hearing Impaired	9	33.3	44.4	22.2	0.0	0.0	0.0	0.0	9	44.4	22.2	22.2	0.0	0.0	0.0	11.1
	Intellectual Disability	22	63.6	22.7	13.6	0.0	0.0	0.0	0.0	22	50.0	22.7	18.2	0.0	0.0	0.0	9.1

	Count	Listening								Reading							
		1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Multiple disabilities	32	71.9	3.1	12.5	0.0	0.0	12.5	0.0	32	12.5	6.3	9.4	0.0	0.0	25.0	46.9	
Orthopedic impairment	8	12.5	12.5	75.0	0.0	0.0	0.0	0.0	8	12.5	12.5	37.5	0.0	12.5	12.5	12.5	
Other health impairment	37	32.4	16.2	37.8	10.8	2.7	0.0	0.0	37	13.5	27.0	32.4	8.1	10.8	0.0	8.1	
Specific learning disability	33	18.2	33.3	42.4	6.1	0.0	0.0	0.0	33	21.2	24.2	48.5	0.0	6.1	0.0	0.0	
Speech or language impairment	206	16.5	20.4	52.9	3.4	6.3	0.5	0.0	206	11.7	22.3	52.4	5.3	7.8	0.5	0.0	
Traumatic brain injury	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
Visual impairment	6	16.7	50.0	33.3	0.0	0.0	0.0	0.0	6	16.7	16.7	66.7	0.0	0.0	0.0	0.0	

* N = Not Attempted; E = Exempted.

Table S3.1: Performance Level by Domain - Kindergarten(cont.)

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
Gender	Total	9170	11.5	10.3	32.4	17.0	27.3	1.1	0.5	9170	30.5	29.3	28.9	4.5	5.6	0.7	0.4
	Female	4202	9.5	9.3	31.5	16.9	31.6	1.0	0.2	4202	27.3	30.2	30.5	5.2	5.9	0.6	0.2
	Male	4525	12.5	11.2	33.7	17.2	23.6	1.1	0.8	4525	32.7	28.8	27.7	4.0	5.3	0.8	0.7
	Missing	443	20.5	10.4	28.0	14.9	23.9	2.3	0.0	443	38.1	25.5	26.9	3.6	5.9	0.0	0.0
Ethnicity	African American	1705	9.0	8.4	35.7	19.1	26.0	1.3	0.5	1705	32.6	27.2	28.9	4.0	5.8	1.0	0.5
	American Indian or Alaskan Native	27	0.0	14.8	33.3	11.1	33.3	3.7	3.7	27	29.6	22.2	29.6	3.7	11.1	0.0	3.7
	Asian	2206	10.2	8.7	29.0	17.0	33.6	0.9	0.5	2206	20.9	23.8	35.7	8.1	10.6	0.5	0.4
	Hispanic or Latino	2427	12.6	13.3	33.8	17.2	21.7	0.9	0.5	2427	36.9	34.0	22.8	2.9	2.1	0.9	0.4
	Multiple Races	1005	15.3	10.6	32.8	15.0	24.2	1.6	0.4	1005	41.0	30.6	22.7	2.1	2.9	0.3	0.4
	Native Hawaiian or Other Pacific Islander	40	12.5	12.5	35.0	12.5	27.5	0.0	0.0	40	25.0	45.0	25.0	5.0	0.0	0.0	0.0
	Other/Unknown	41	4.9	4.9	26.8	22.0	39.0	2.4	0.0	41	19.5	19.5	39.0	2.4	17.1	2.4	0.0
	White	1420	9.2	9.9	31.7	16.3	31.6	0.7	0.7	1420	22.5	32.5	33.9	4.6	5.4	0.5	0.6
	Missing	299	26.8	9.0	29.4	12.4	19.7	2.7	0.0	299	43.5	23.1	24.7	3.0	5.7	0.0	0.0
	Other	Gifted	14	7.1	0.0	7.1	14.3	71.4	0.0	0.0	14	7.1	14.3	57.1	7.1	14.3	0.0
	IEP	468	25.6	16.7	30.1	6.6	6.6	5.6	8.8	468	38.9	26.1	17.3	1.7	2.6	5.1	8.3
	Migrant	10	20.0	30.0	10.0	20.0	20.0	0.0	0.0	10	20.0	40.0	30.0	0.0	10.0	0.0	0.0
Primary Disability	Autism	97	34.0	11.3	15.5	4.1	3.1	16.5	15.5	97	30.9	14.4	20.6	1.0	4.1	15.5	13.4
	Developmental delay	6	33.3	0.0	16.7	0.0	0.0	0.0	50.0	6	16.7	33.3	0.0	0.0	0.0	0.0	50.0
	Emotional disturbance	7	14.3	14.3	42.9	14.3	0.0	14.3	0.0	7	57.1	28.6	14.3	0.0	0.0	0.0	0.0
	Hearing Impaired	9	11.1	44.4	22.2	0.0	11.1	0.0	11.1	9	44.4	44.4	0.0	0.0	0.0	0.0	11.1
	Intellectual Disability	22	59.1	13.6	13.6	0.0	4.5	0.0	9.1	22	68.2	13.6	9.1	0.0	0.0	0.0	9.1

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Multiple disabilities	32	28.1	3.1	3.1	0.0	0.0	18.8	46.9	32	18.8	3.1	6.3	0.0	0.0	21.9	50.0
	Orthopedic impairment	8	12.5	0.0	25.0	25.0	12.5	12.5	12.5	8	37.5	12.5	12.5	0.0	12.5	12.5	12.5
	Other health impairment	37	27.0	16.2	29.7	10.8	8.1	0.0	8.1	37	45.9	24.3	21.6	0.0	0.0	0.0	8.1
	Specific learning disability	33	18.2	18.2	51.5	0.0	9.1	3.0	0.0	33	63.6	18.2	18.2	0.0	0.0	0.0	0.0
	Speech or language impairment	206	19.9	20.9	41.3	8.7	8.7	0.5	0.0	206	37.9	36.9	18.0	3.4	3.4	0.5	0.0
	Traumatic brain injury	1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	Visual impairment	6	16.7	16.7	33.3	33.3	0.0	0.0	0.0	6	33.3	33.3	33.3	0.0	0.0	0.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.2: Performance Level by Domain - Grade 1

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
Gender	Total	8036	6.2	5.4	30.6	25.8	31.6	0.3	0.0	8036	17.8	17.9	30.6	15.1	17.6	0.5	0.6
	Female	3719	4.7	5.1	29.2	26.5	34.3	0.2	0.0	3719	15.0	17.9	30.6	15.7	20.1	0.3	0.5
	Male	4105	6.6	5.7	32.2	25.5	29.7	0.3	0.0	4105	19.4	18.0	30.8	14.8	15.6	0.6	0.8
	Missing	212	23.6	6.6	25.5	19.8	22.6	1.9	0.0	212	35.4	14.6	25.5	11.8	10.8	1.9	0.0
Ethnicity	African American	1529	5.4	6.3	32.7	25.4	29.6	0.6	0.0	1529	16.4	17.0	32.6	17.4	15.2	0.8	0.5
	American Indian or Alaskan Native	20	5.0	0.0	45.0	10.0	40.0	0.0	0.0	20	15.0	15.0	35.0	20.0	15.0	0.0	0.0
	Asian	1816	4.7	4.6	27.0	22.3	41.1	0.3	0.0	1816	10.9	12.8	27.6	17.6	30.1	0.4	0.5
	Hispanic or Latino	2267	6.2	6.0	33.6	28.5	25.7	0.0	0.0	2267	22.0	22.6	31.7	12.9	10.1	0.3	0.4
	Multiple Races	894	7.8	5.6	32.6	28.0	25.8	0.2	0.0	894	24.3	21.9	28.1	11.2	12.3	0.4	1.8
	Native Hawaiian or Other Pacific Islander	29	3.4	0.0	31.0	37.9	27.6	0.0	0.0	29	17.2	10.3	41.4	13.8	17.2	0.0	0.0
	Other/Unknown	25	4.0	0.0	28.0	16.0	52.0	0.0	0.0	25	16.0	8.0	24.0	24.0	28.0	0.0	0.0
	White	1300	5.2	4.8	27.5	26.5	35.8	0.2	0.0	1300	14.5	15.8	32.8	15.9	20.2	0.5	0.5
	Missing	156	30.8	6.4	22.4	16.0	22.4	1.9	0.0	156	40.4	13.5	23.1	10.9	10.3	1.9	0.0
	Other	Gifted	38	0.0	0.0	10.5	15.8	73.7	0.0	0.0	38	7.9	2.6	15.8	5.3	68.4	0.0
IEP		497	20.9	10.5	35.0	17.7	14.1	1.6	0.2	497	35.8	17.1	19.1	6.6	7.6	4.2	9.5
Migrant		15	20.0	0.0	26.7	13.3	40.0	0.0	0.0	15	26.7	20.0	46.7	0.0	6.7	0.0	0.0
Primary Disability	Autism	83	39.8	16.9	28.9	6.0	6.0	1.2	1.2	83	28.9	15.7	20.5	1.2	4.8	12.0	16.9
	Developmental delay	4	0.0	50.0	0.0	0.0	0.0	50.0	0.0	4	50.0	0.0	0.0	0.0	0.0	50.0	0.0
	Emotional disturbance	4	0.0	25.0	25.0	25.0	25.0	0.0	0.0	4	75.0	0.0	25.0	0.0	0.0	0.0	0.0
	Hearing Impaired	4	50.0	0.0	50.0	0.0	0.0	0.0	0.0	4	75.0	0.0	0.0	25.0	0.0	0.0	0.0
	Intellectual Disability	38	47.4	7.9	34.2	5.3	2.6	2.6	0.0	38	63.2	5.3	13.2	2.6	2.6	2.6	10.5

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Multiple disabilities	31	77.4	6.5	6.5	3.2	0.0	6.5	0.0	31	32.3	0.0	6.5	0.0	0.0	12.9	48.4
	Orthopedic impairment	15	13.3	13.3	26.7	13.3	33.3	0.0	0.0	15	26.7	0.0	13.3	0.0	26.7	13.3	20.0
	Other health impairment	52	13.5	19.2	30.8	28.8	7.7	0.0	0.0	52	46.2	19.2	21.2	5.8	1.9	0.0	5.8
	Specific learning disability	73	6.8	11.0	50.7	26.0	5.5	0.0	0.0	73	46.6	24.7	11.0	4.1	6.8	0.0	6.8
	Speech or language impairment	185	7.0	5.4	38.9	21.6	26.5	0.5	0.0	185	27.6	19.5	26.5	13.0	11.4	1.1	1.1
	Traumatic brain injury	3	0.0	0.0	33.3	33.3	0.0	33.3	0.0	3	0.0	33.3	33.3	0.0	0.0	0.0	33.3
	Visual impairment	3	0.0	0.0	33.3	66.7	0.0	0.0	0.0	3	33.3	33.3	0.0	0.0	33.3	0.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.2: Performance Level by Domain - Grade 1(cont.)

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
Gender	Total	8036	17.0	26.9	11.4	17.6	25.8	0.8	0.5	8036	25.6	23.7	33.5	8.5	7.5	0.5	0.6
	Female	3719	13.8	26.6	11.6	18.2	29.0	0.5	0.3	3719	22.6	23.2	35.6	8.8	9.0	0.3	0.5
	Male	4105	19.1	27.4	11.5	17.3	23.2	0.9	0.6	4105	27.4	24.3	32.4	8.3	6.2	0.7	0.8
	Missing	212	34.9	21.2	7.5	12.7	19.3	4.2	0.0	212	45.8	21.2	18.9	6.6	7.1	0.5	0.0
Ethnicity	African American	1529	14.5	27.1	12.8	18.8	25.4	1.1	0.5	1529	25.8	23.2	36.9	8.0	5.0	0.6	0.5
	American Indian or Alaskan Native	20	10.0	25.0	20.0	25.0	20.0	0.0	0.0	20	20.0	30.0	40.0	5.0	5.0	0.0	0.0
	Asian	1816	14.9	21.2	9.3	17.5	36.1	0.6	0.4	1816	18.1	18.1	35.9	13.2	13.8	0.3	0.6
	Hispanic or Latino	2267	18.4	32.3	13.1	16.3	19.3	0.3	0.4	2267	31.2	27.7	30.4	5.3	4.5	0.5	0.5
	Multiple Races	894	18.5	30.2	11.4	18.1	19.2	1.5	1.1	894	30.4	30.0	26.5	6.0	4.5	0.9	1.7
	Native Hawaiian or Other Pacific Islander	29	13.8	17.2	10.3	37.9	20.7	0.0	0.0	29	24.1	13.8	48.3	3.4	10.3	0.0	0.0
	Other/Unknown	25	20.0	24.0	12.0	12.0	32.0	0.0	0.0	25	20.0	16.0	28.0	32.0	4.0	0.0	0.0
	White	1300	17.0	24.0	10.5	18.8	28.8	0.7	0.2	1300	20.4	21.9	37.8	9.6	9.5	0.4	0.4
	Missing	156	41.0	19.9	7.1	9.0	17.9	5.1	0.0	156	48.1	19.2	19.2	7.1	5.8	0.6	0.0
	Other	Gifted	38	10.5	10.5	2.6	21.1	55.3	0.0	0.0	38	7.9	10.5	23.7	13.2	44.7	0.0
IEP		497	38.8	22.7	5.8	9.5	11.3	4.8	7.0	497	45.9	18.9	14.7	3.8	3.2	3.8	9.7
Migrant		15	33.3	26.7	20.0	6.7	13.3	0.0	0.0	15	46.7	20.0	26.7	0.0	6.7	0.0	0.0
Primary Disability	Autism	83	48.2	14.5	1.2	9.6	1.2	12.0	13.3	83	42.2	14.5	13.3	0.0	1.2	10.8	18.1
	Developmental delay	4	50.0	0.0	0.0	0.0	0.0	50.0	0.0	4	50.0	0.0	0.0	0.0	0.0	50.0	0.0
	Emotional disturbance	4	50.0	25.0	25.0	0.0	0.0	0.0	0.0	4	75.0	0.0	25.0	0.0	0.0	0.0	0.0
	Hearing Impaired	4	50.0	25.0	25.0	0.0	0.0	0.0	0.0	4	75.0	0.0	25.0	0.0	0.0	0.0	0.0
	Intellectual Disability	38	73.7	10.5	0.0	0.0	5.3	0.0	10.5	38	71.1	7.9	10.5	0.0	0.0	0.0	10.5

	Count	Speaking								Writing							
		1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Multiple disabilities	31	32.3	6.5	0.0	0.0	0.0	19.4	41.9	31	32.3	0.0	0.0	0.0	0.0	16.1	51.6	
Orthopedic impairment	15	33.3	0.0	6.7	6.7	26.7	13.3	13.3	15	26.7	0.0	6.7	6.7	20.0	20.0	20.0	
Other health impairment	52	48.1	28.8	5.8	7.7	9.6	0.0	0.0	52	57.7	25.0	9.6	1.9	1.9	0.0	3.8	
Specific learning disability	73	43.8	24.7	8.2	9.6	8.2	2.7	2.7	73	61.6	17.8	8.2	1.4	4.1	0.0	6.8	
Speech or language impairment	185	24.3	31.4	8.6	14.1	19.5	1.1	1.1	185	35.1	27.6	24.3	8.1	3.8	0.0	1.1	
Traumatic brain injury	3	0.0	33.3	0.0	33.3	0.0	0.0	33.3	3	66.7	0.0	0.0	0.0	0.0	0.0	33.3	
Visual impairment	3	33.3	33.3	0.0	0.0	33.3	0.0	0.0	3	33.3	33.3	0.0	33.3	0.0	0.0	0.0	

* N = Not Attempted; E = Exempted.

Table S3.3: Performance Level by Domain - Grade 2

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	6904	4.6	3.1	20.1	33.8	38.1	0.2	0.1	6904	15.0	15.9	33.6	17.4	16.7	0.6	0.7	
	Female	3127	4.0	2.7	20.1	31.6	41.4	0.2	0.1	3127	12.8	14.8	33.1	18.5	19.8	0.5	0.5	
	Male	3660	5.1	3.4	20.2	35.4	35.5	0.2	0.1	3660	16.9	16.8	34.0	16.6	14.3	0.6	1.0	
	Missing	117	3.4	5.1	17.9	41.9	31.6	0.0	0.0	117	19.7	19.7	34.2	12.8	13.7	0.0	0.0	
Ethnicity	African American	1372	4.7	3.1	22.3	34.8	34.5	0.5	0.1	1372	15.8	17.4	32.2	18.4	14.9	0.6	0.7	
	American Indian or Alaskan Native	12	0.0	8.3	33.3	25.0	33.3	0.0	0.0	12	16.7	8.3	41.7	25.0	8.3	0.0	0.0	
	Asian	1391	4.4	2.2	17.6	29.5	46.2	0.0	0.1	1391	10.4	12.2	30.6	19.7	26.0	0.2	0.9	
	Hispanic or Latino	2159	4.7	3.3	21.2	36.2	34.4	0.1	0.0	2159	17.8	18.3	34.8	16.0	11.7	0.7	0.6	
	Multiple Races	849	6.6	3.9	20.6	33.7	34.9	0.2	0.1	849	18.1	17.1	34.4	14.8	13.7	0.8	1.1	
	Native Hawaiian or Other Pacific Islander	41	2.4	7.3	24.4	43.9	22.0	0.0	0.0	41	19.5	17.1	43.9	9.8	9.8	0.0	0.0	
	Other/Unknown	18	5.6	11.1	11.1	33.3	38.9	0.0	0.0	18	16.7	22.2	22.2	11.1	27.8	0.0	0.0	
	White	1062	3.0	2.7	17.8	33.2	42.9	0.3	0.0	1062	12.0	13.1	35.7	18.3	19.9	0.4	0.8	
	Other	Gifted	38	0.0	0.0	5.3	18.4	76.3	0.0	0.0	38	2.6	2.6	13.2	18.4	63.2	0.0	0.0
		IEP	581	13.6	7.1	34.4	28.9	14.1	1.0	0.9	581	35.1	19.3	23.9	6.5	2.8	4.3	8.1
Migrant		29	6.9	0.0	24.1	34.5	34.5	0.0	0.0	29	17.2	31.0	34.5	13.8	3.4	0.0	0.0	
Primary Disability	Autism	83	28.9	12.0	24.1	20.5	8.4	3.6	2.4	83	24.1	8.4	25.3	7.2	1.2	9.6	24.1	
	Deaf-blindness	1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
	Developmental delay	3	100.0	0.0	0.0	0.0	0.0	0.0	0.0	3	66.7	0.0	0.0	0.0	0.0	33.3	0.0	
	Emotional disturbance	17	0.0	0.0	58.8	35.3	5.9	0.0	0.0	17	41.2	29.4	17.6	5.9	0.0	5.9	0.0	
	Hearing Impaired	9	11.1	0.0	22.2	44.4	11.1	0.0	11.1	9	11.1	22.2	44.4	22.2	0.0	0.0	0.0	
	Intellectual Disability	34	20.6	17.6	44.1	17.6	0.0	0.0	0.0	34	47.1	17.6	20.6	2.9	0.0	5.9	5.9	

	Count	Listening								Reading							
		1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Multiple disabilities	35	65.7	11.4	14.3	0.0	0.0	2.9	5.7	35	34.3	0.0	0.0	0.0	0.0	25.7	40.0	
Orthopedic impairment	7	0.0	0.0	14.3	28.6	42.9	14.3	0.0	7	14.3	0.0	42.9	0.0	28.6	14.3	0.0	
Other health impairment	61	11.5	8.2	49.2	29.5	1.6	0.0	0.0	61	50.8	23.0	13.1	3.3	0.0	1.6	8.2	
Specific learning disability	192	3.6	6.3	44.8	32.8	12.0	0.5	0.0	192	44.8	25.5	22.9	3.6	1.0	0.0	2.1	
Speech or language impairment	129	2.3	2.3	23.3	39.5	32.6	0.0	0.0	129	19.4	23.3	34.9	13.2	8.5	0.8	0.0	
Traumatic brain injury	3	66.7	0.0	0.0	33.3	0.0	0.0	0.0	3	66.7	0.0	0.0	33.3	0.0	0.0	0.0	
Visual impairment	1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	

* N = Not Attempted; E = Exempted.

Table S3.3: Performance Level by Domain - Grade 2(cont.)

		Speaking								Writing								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	6904	16.6	17.5	16.6	22.0	25.6	1.0	0.7	6904	16.4	14.2	33.1	18.0	16.9	0.7	0.7	
	Female	3127	14.2	17.1	15.8	23.1	28.5	0.9	0.4	3127	13.1	13.6	33.5	18.5	20.0	0.7	0.5	
	Male	3660	18.7	17.5	17.2	21.0	23.6	1.1	0.8	3660	19.0	14.6	32.7	17.8	14.3	0.7	0.9	
	Missing	117	17.1	29.1	18.8	23.1	12.0	0.0	0.0	117	20.5	17.1	35.9	11.1	15.4	0.0	0.0	
Ethnicity	African American	1372	17.8	22.2	15.7	19.9	22.9	0.9	0.6	1372	17.6	16.9	31.4	17.5	14.8	1.1	0.7	
	American Indian or Alaskan Native	12	16.7	16.7	25.0	16.7	25.0	0.0	0.0	12	16.7	16.7	33.3	25.0	8.3	0.0	0.0	
	Asian	1391	14.7	13.0	14.9	22.1	33.6	0.7	0.9	1391	11.4	10.9	31.1	20.3	25.2	0.4	0.9	
	Hispanic or Latino	2159	17.7	18.3	18.7	22.7	20.9	1.1	0.6	2159	19.3	15.1	35.2	16.5	12.7	0.7	0.5	
	Multiple Races	849	20.0	18.0	17.1	20.8	21.9	1.5	0.6	849	19.3	16.5	32.2	17.3	13.0	0.7	1.1	
	Native Hawaiian or Other Pacific Islander	41	14.6	26.8	22.0	12.2	24.4	0.0	0.0	41	14.6	17.1	41.5	12.2	14.6	0.0	0.0	
	Other/Unknown	18	22.2	16.7	16.7	27.8	16.7	0.0	0.0	18	22.2	11.1	16.7	11.1	38.9	0.0	0.0	
	White	1062	12.8	15.1	15.0	24.3	31.1	1.0	0.8	1062	12.8	11.3	34.7	19.5	20.4	0.5	0.8	
	Other	Gifted	38	5.3	10.5	5.3	26.3	50.0	2.6	0.0	38	2.6	2.6	13.2	10.5	71.1	0.0	0.0
		IEP	581	36.3	19.6	12.6	11.0	7.4	5.3	7.7	581	39.1	16.2	21.5	7.9	2.8	4.6	7.9
Migrant		29	20.7	20.7	10.3	20.7	24.1	3.4	0.0	29	13.8	17.2	44.8	20.7	0.0	3.4	0.0	
Primary Disability	Autism	83	37.3	7.2	4.8	9.6	2.4	10.8	27.7	83	24.1	13.3	14.5	7.2	2.4	10.8	27.7	
	Deaf-blindness	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
	Developmental delay	3	66.7	0.0	0.0	0.0	0.0	33.3	0.0	3	66.7	0.0	0.0	0.0	0.0	33.3	0.0	
	Emotional disturbance	17	41.2	29.4	11.8	5.9	5.9	5.9	0.0	17	58.8	5.9	23.5	5.9	0.0	5.9	0.0	
	Hearing Impaired	9	22.2	22.2	0.0	33.3	11.1	0.0	11.1	9	11.1	22.2	44.4	22.2	0.0	0.0	0.0	
	Intellectual Disability	34	64.7	17.6	8.8	0.0	0.0	5.9	2.9	34	61.8	8.8	11.8	2.9	2.9	5.9	5.9	

	Speaking									Writing							
	Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Multiple disabilities	35	37.1	0.0	0.0	0.0	0.0	25.7	37.1	35	37.1	2.9	0.0	0.0	0.0	25.7	34.3	
Orthopedic impairment	7	0.0	14.3	28.6	14.3	28.6	14.3	0.0	7	0.0	28.6	28.6	0.0	28.6	14.3	0.0	
Other health impairment	61	42.6	26.2	16.4	4.9	3.3	1.6	4.9	61	55.7	16.4	16.4	3.3	1.6	1.6	4.9	
Specific learning disability	192	39.6	24.5	14.1	13.5	5.7	2.1	0.5	192	49.5	20.3	21.9	4.7	0.5	1.0	2.1	
Speech or language impairment	129	23.3	24.8	17.8	14.7	17.8	0.8	0.8	129	21.7	19.4	33.3	17.8	7.8	0.0	0.0	
Traumatic brain injury	3	33.3	0.0	0.0	33.3	0.0	33.3	0.0	3	66.7	0.0	0.0	33.3	0.0	0.0	0.0	
Visual impairment	1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	

* N = Not Attempted; E = Exempted.

Table S3.4: Performance Level by Domain - Grade 3

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	5716	3.7	3.5	17.4	35.4	39.8	0.2	0.0	5716	18.0	14.9	35.5	17.1	13.2	0.6	0.8	
	Female	2675	3.4	3.4	16.9	35.1	40.9	0.1	0.0	2675	15.9	14.4	36.0	17.3	15.2	0.5	0.6	
	Male	3020	3.8	3.6	17.9	35.7	38.7	0.2	0.0	3020	19.9	15.3	35.1	16.9	11.3	0.6	0.9	
	Missing	21	14.3	0.0	14.3	19.0	52.4	0.0	0.0	21	14.3	19.0	23.8	28.6	14.3	0.0	0.0	
Ethnicity	African American	1119	2.9	3.7	17.6	37.5	38.2	0.1	0.0	1119	16.7	14.5	39.8	16.8	11.6	0.4	0.2	
	American Indian or Alaskan Native	12	8.3	0.0	16.7	25.0	50.0	0.0	0.0	12	16.7	8.3	25.0	0.0	41.7	8.3	0.0	
	Asian	963	2.4	3.9	18.0	29.0	46.6	0.1	0.0	963	14.0	12.5	31.7	20.8	19.5	0.6	0.9	
	Hispanic or Latino	1914	4.6	3.3	18.3	36.2	37.3	0.3	0.1	1914	20.0	16.6	35.8	15.7	10.4	0.7	0.7	
	Multiple Races	838	4.7	3.9	15.6	39.5	36.2	0.1	0.0	838	22.4	15.2	33.1	16.9	10.6	0.5	1.3	
	Native Hawaiian or Other Pacific Islander	28	3.6	0.0	21.4	57.1	17.9	0.0	0.0	28	7.1	10.7	64.3	7.1	10.7	0.0	0.0	
	Other/Unknown	6	0.0	0.0	33.3	16.7	50.0	0.0	0.0	6	16.7	16.7	33.3	33.3	0.0	0.0	0.0	
	White	836	3.2	3.0	16.0	33.4	44.1	0.2	0.0	836	15.7	14.4	35.0	17.1	16.4	0.4	1.1	
	Other	Gifted	36	0.0	0.0	2.8	27.8	69.4	0.0	0.0	36	0.0	5.6	27.8	22.2	44.4	0.0	0.0
		IEP	776	7.7	6.7	29.4	39.4	16.0	0.6	0.1	776	38.1	19.7	24.4	5.8	2.8	3.6	5.5
Migrant		34	5.9	11.8	17.6	35.3	29.4	0.0	0.0	34	29.4	8.8	38.2	11.8	11.8	0.0	0.0	
Primary Disability	Autism	83	27.7	15.7	21.7	19.3	14.5	1.2	0.0	83	36.1	15.7	18.1	4.8	2.4	8.4	14.5	
	Deaf-blindness	1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
	Developmental delay	8	12.5	50.0	25.0	12.5	0.0	0.0	0.0	8	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
	Emotional disturbance	13	0.0	0.0	30.8	38.5	30.8	0.0	0.0	13	23.1	23.1	23.1	23.1	0.0	7.7	0.0	
	Hearing Impaired	11	9.1	9.1	27.3	18.2	27.3	9.1	0.0	11	27.3	36.4	27.3	9.1	0.0	0.0	0.0	
	Intellectual Disability	47	8.5	23.4	42.6	23.4	2.1	0.0	0.0	47	68.1	10.6	6.4	0.0	0.0	6.4	8.5	

	Count	Listening								Reading							
		1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Multiple disabilities	23	69.6	17.4	13.0	0.0	0.0	0.0	0.0	23	34.8	0.0	4.3	0.0	0.0	17.4	43.5	
Orthopedic impairment	5	0.0	20.0	20.0	40.0	20.0	0.0	0.0	5	40.0	40.0	0.0	0.0	20.0	0.0	0.0	
Other health impairment	90	3.3	8.9	37.8	35.6	11.1	2.2	1.1	90	36.7	16.7	33.3	3.3	1.1	4.4	4.4	
Specific learning disability	388	2.8	1.0	32.2	50.5	13.1	0.3	0.0	388	41.8	24.2	23.2	4.4	1.8	2.1	2.6	
Speech or language impairment	109	0.0	3.7	15.6	39.4	41.3	0.0	0.0	109	11.9	13.8	42.2	19.3	11.0	0.9	0.9	
Traumatic brain injury	3	0.0	33.3	33.3	33.3	0.0	0.0	0.0	3	33.3	0.0	33.3	0.0	0.0	0.0	33.3	
Visual impairment	2	0.0	0.0	50.0	50.0	0.0	0.0	0.0	2	0.0	0.0	50.0	0.0	0.0	0.0	50.0	

* N = Not Attempted; E = Exempted.

Table S3.4: Performance Level by Domain - Grade 3(cont.)

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
Gender	Total	5716	11.8	10.9	17.4	28.4	29.9	0.9	0.7	5716	17.5	14.6	35.5	18.0	13.0	0.7	0.7
	Female	2675	11.5	9.2	16.6	28.9	32.7	0.7	0.4	2675	15.4	13.6	34.9	19.6	15.5	0.6	0.6
	Male	3020	12.1	12.4	18.2	27.9	27.4	1.0	0.9	3020	19.4	15.6	36.0	16.6	10.9	0.8	0.9
	Missing	21	4.8	9.5	4.8	42.9	28.6	9.5	0.0	21	19.0	9.5	28.6	28.6	14.3	0.0	0.0
Ethnicity	African American	1119	11.3	12.0	19.2	28.8	28.0	0.5	0.3	1119	15.6	14.9	38.0	19.0	11.5	0.7	0.2
	American Indian or Alaskan Native	12	0.0	8.3	8.3	41.7	33.3	8.3	0.0	12	8.3	16.7	25.0	16.7	25.0	8.3	0.0
	Asian	963	12.3	8.7	14.6	25.0	37.9	0.5	0.9	963	14.8	12.6	32.0	21.3	18.2	0.4	0.7
	Hispanic or Latino	1914	11.3	11.6	18.3	30.6	26.5	1.1	0.6	1914	19.4	15.7	36.5	16.0	10.9	0.8	0.7
	Multiple Races	838	15.0	11.8	19.6	26.4	24.9	1.4	0.8	838	21.6	14.9	32.9	17.7	11.0	0.7	1.2
	Native Hawaiian or Other Pacific Islander	28	3.6	7.1	7.1	50.0	32.1	0.0	0.0	28	3.6	17.9	57.1	14.3	7.1	0.0	0.0
	Other/Unknown	6	0.0	16.7	0.0	33.3	50.0	0.0	0.0	6	16.7	16.7	33.3	33.3	0.0	0.0	0.0
	White	836	10.3	9.7	14.7	28.1	35.6	0.6	1.0	836	15.2	13.6	35.6	17.7	16.1	0.6	1.1
	Other	Gifted	36	2.8	0.0	5.6	22.2	69.4	0.0	0.0	36	0.0	11.1	22.2	22.2	44.4	0.0
IEP		776	20.7	19.7	21.5	19.3	11.1	2.8	4.8	776	37.1	22.4	22.9	6.6	2.8	3.1	5.0
Migrant		34	14.7	14.7	23.5	26.5	17.6	2.9	0.0	34	26.5	11.8	35.3	14.7	8.8	2.9	0.0
Primary Disability	Autism	83	37.3	10.8	10.8	7.2	7.2	10.8	15.7	83	33.7	18.1	12.0	12.0	0.0	9.6	14.5
	Deaf-blindness	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	Developmental delay	8	62.5	25.0	12.5	0.0	0.0	0.0	0.0	8	100.0	0.0	0.0	0.0	0.0	0.0	0.0
	Emotional disturbance	13	15.4	23.1	15.4	30.8	15.4	0.0	0.0	13	30.8	15.4	23.1	15.4	7.7	7.7	0.0
	Hearing Impaired	11	27.3	18.2	27.3	9.1	18.2	0.0	0.0	11	18.2	36.4	36.4	9.1	0.0	0.0	0.0
	Intellectual Disability	47	42.6	23.4	12.8	6.4	2.1	6.4	6.4	47	74.5	8.5	4.3	0.0	0.0	4.3	8.5

	Speaking									Writing							
	Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Multiple disabilities	23	30.4	4.3	0.0	4.3	0.0	13.0	47.8	23	39.1	0.0	4.3	0.0	0.0	13.0	43.5	
Orthopedic impairment	5	20.0	20.0	20.0	20.0	20.0	0.0	0.0	5	40.0	40.0	0.0	20.0	0.0	0.0	0.0	
Other health impairment	90	21.1	25.6	20.0	17.8	8.9	3.3	3.3	90	34.4	23.3	27.8	3.3	1.1	5.6	4.4	
Specific learning disability	388	16.2	23.2	25.0	22.7	11.3	0.8	0.8	388	39.4	27.8	22.7	5.2	2.3	1.0	1.5	
Speech or language impairment	109	6.4	9.2	26.6	29.4	25.7	0.9	1.8	109	11.9	16.5	43.1	16.5	10.1	0.9	0.9	
Traumatic brain injury	3	33.3	0.0	33.3	0.0	0.0	0.0	33.3	3	33.3	33.3	0.0	0.0	0.0	0.0	33.3	
Visual impairment	2	0.0	0.0	0.0	50.0	0.0	0.0	50.0	2	0.0	0.0	50.0	0.0	0.0	0.0	50.0	

* N = Not Attempted; E = Exempted.

Table S3.5: Performance Level by Domain - Grade 4

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	4692	6.9	4.6	14.5	37.9	35.6	0.4	0.1	4692	14.9	11.0	31.6	22.4	17.9	1.0	1.2	
	Female	2118	7.2	4.2	13.7	37.4	37.0	0.4	0.1	2118	14.4	9.8	33.7	21.6	19.0	0.8	0.8	
	Male	2506	6.7	5.0	15.2	38.5	34.0	0.4	0.0	2506	15.4	12.1	30.0	23.0	16.8	1.2	1.5	
	Missing	68	4.4	4.4	13.2	27.9	50.0	0.0	0.0	68	14.7	8.8	26.5	25.0	25.0	0.0	0.0	
Ethnicity	African American	882	6.9	5.2	15.8	35.8	35.7	0.6	0.0	882	16.1	11.7	29.1	23.8	17.5	1.0	0.8	
	American Indian or Alaskan Native	9	0.0	0.0	22.2	44.4	33.3	0.0	0.0	9	11.1	11.1	22.2	33.3	22.2	0.0	0.0	
	Asian	751	6.0	4.7	14.5	34.4	39.9	0.3	0.3	751	15.0	10.7	26.9	21.2	24.2	0.7	1.3	
	Hispanic or Latino	1666	8.0	4.7	14.1	39.6	33.1	0.3	0.1	1666	14.3	11.5	34.8	22.4	14.5	1.3	1.2	
	Multiple Races	689	7.8	4.5	13.2	39.3	34.8	0.3	0.0	689	16.4	11.5	32.1	21.8	16.7	0.4	1.2	
	Native Hawaiian or Other Pacific Islander	27	11.1	0.0	14.8	48.1	25.9	0.0	0.0	27	7.4	18.5	40.7	14.8	14.8	0.0	3.7	
	Other/Unknown	8	25.0	25.0	12.5	12.5	25.0	0.0	0.0	8	50.0	12.5	12.5	25.0	0.0	0.0	0.0	
	White	660	3.9	3.6	15.0	38.5	38.0	0.9	0.0	660	13.2	8.3	31.8	22.7	21.5	1.2	1.2	
	Other	Gifted	12	0.0	0.0	0.0	25.0	75.0	0.0	0.0	12	0.0	8.3	8.3	33.3	50.0	0.0	0.0
		IEP	766	9.9	9.8	25.5	39.6	13.1	1.8	0.4	766	22.2	19.2	32.8	10.4	3.7	4.7	7.0
Migrant		17	0.0	0.0	11.8	58.8	29.4	0.0	0.0	17	0.0	11.8	64.7	5.9	11.8	5.9	0.0	
Primary Disability	Autism	56	14.3	25.0	33.9	17.9	3.6	3.6	1.8	56	16.1	12.5	25.0	3.6	1.8	14.3	26.8	
	Developmental delay	6	16.7	33.3	33.3	0.0	0.0	16.7	0.0	6	83.3	16.7	0.0	0.0	0.0	0.0	0.0	
	Emotional disturbance	15	6.7	13.3	26.7	26.7	20.0	6.7	0.0	15	40.0	20.0	13.3	26.7	0.0	0.0	0.0	
	Hearing Impaired	6	33.3	0.0	0.0	16.7	33.3	0.0	16.7	6	33.3	16.7	16.7	33.3	0.0	0.0	0.0	
	Intellectual Disability	55	29.1	25.5	34.5	5.5	1.8	1.8	1.8	55	43.6	12.7	12.7	1.8	0.0	16.4	12.7	
	Multiple disabilities	31	71.0	9.7	0.0	0.0	0.0	19.4	0.0	31	12.9	0.0	0.0	0.0	0.0	25.8	61.3	

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	9	11.1	11.1	11.1	55.6	11.1	0.0	0.0	9	33.3	22.2	11.1	11.1	11.1	11.1	0.0
	Other health impairment	87	10.3	11.5	26.4	33.3	17.2	1.1	0.0	87	27.6	16.1	33.3	12.6	1.1	3.4	5.7
	Specific learning disability	435	3.2	5.1	26.2	51.5	13.8	0.2	0.0	435	18.6	23.9	40.0	10.8	3.4	1.4	1.8
	Speech or language impairment	52	0.0	3.8	21.2	38.5	32.7	3.8	0.0	52	5.8	9.6	42.3	21.2	19.2	1.9	0.0
	Traumatic brain injury	7	14.3	57.1	0.0	28.6	0.0	0.0	0.0	7	57.1	14.3	14.3	0.0	0.0	14.3	0.0
	Visual impairment	1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.5: Performance Level by Domain - Grade 4(cont.)

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
Gender	Total	4692	11.0	9.1	16.6	27.7	33.7	1.1	0.8	4692	13.6	8.8	47.1	16.7	11.7	0.9	1.1
	Female	2118	10.4	8.6	14.5	27.8	37.2	0.9	0.7	2118	13.0	7.6	47.4	16.1	14.4	0.8	0.7
	Male	2506	11.6	9.7	18.5	27.5	30.6	1.2	1.0	2506	14.2	9.9	47.2	16.8	9.4	1.0	1.4
	Missing	68	10.3	4.4	11.8	35.3	38.2	0.0	0.0	68	13.2	7.4	36.8	27.9	14.7	0.0	0.0
Ethnicity	African American	882	12.2	9.5	17.6	27.6	31.3	1.0	0.8	882	14.1	9.1	45.5	18.4	11.1	1.1	0.8
	American Indian or Alaskan Native	9	0.0	11.1	22.2	11.1	55.6	0.0	0.0	9	11.1	0.0	66.7	11.1	11.1	0.0	0.0
	Asian	751	11.1	9.5	15.0	26.4	36.0	0.8	1.3	751	13.2	8.8	43.1	16.5	16.6	0.5	1.2
	Hispanic or Latino	1666	11.3	9.0	17.8	28.1	31.7	1.4	0.7	1666	14.1	8.8	49.3	16.1	9.2	1.2	1.1
	Multiple Races	689	11.6	9.3	15.7	27.9	34.3	0.7	0.6	689	15.7	7.5	48.9	16.7	9.6	0.6	1.0
	Native Hawaiian or Other Pacific Islander	27	11.1	22.2	18.5	22.2	25.9	0.0	0.0	27	11.1	11.1	55.6	3.7	14.8	0.0	3.7
	Other/Unknown	8	50.0	0.0	37.5	12.5	0.0	0.0	0.0	8	50.0	12.5	37.5	0.0	0.0	0.0	0.0
	White	660	7.6	7.9	14.5	28.9	38.9	1.1	1.1	660	9.8	9.8	46.1	16.7	15.6	0.9	1.1
	Other	Gifted	12	0.0	0.0	8.3	8.3	83.3	0.0	0.0	12	0.0	8.3	25.0	25.0	41.7	0.0
IEP		766	15.1	15.7	20.0	23.4	16.4	4.3	5.1	766	22.5	15.9	44.0	5.2	1.6	4.3	6.5
Migrant		17	0.0	0.0	35.3	29.4	29.4	5.9	0.0	17	0.0	17.6	64.7	11.8	0.0	5.9	0.0
Primary Disability	Autism	56	21.4	14.3	12.5	7.1	5.4	16.1	23.2	56	16.1	10.7	32.1	0.0	0.0	14.3	26.8
	Developmental delay	6	83.3	0.0	0.0	0.0	0.0	16.7	0.0	6	66.7	16.7	0.0	0.0	0.0	16.7	0.0
	Emotional disturbance	15	26.7	13.3	33.3	13.3	13.3	0.0	0.0	15	40.0	6.7	46.7	6.7	0.0	0.0	0.0
	Hearing Impaired	6	33.3	0.0	0.0	16.7	33.3	0.0	16.7	6	50.0	0.0	50.0	0.0	0.0	0.0	0.0
	Intellectual Disability	55	45.5	18.2	9.1	3.6	1.8	16.4	5.5	55	43.6	16.4	9.1	1.8	0.0	16.4	12.7
	Multiple disabilities	31	25.8	0.0	0.0	0.0	0.0	19.4	54.8	31	22.6	0.0	0.0	0.0	0.0	19.4	58.1

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	9	22.2	11.1	11.1	22.2	22.2	11.1	0.0	9	22.2	11.1	44.4	11.1	0.0	11.1	0.0
	Other health impairment	87	14.9	18.4	26.4	17.2	17.2	2.3	3.4	87	28.7	18.4	36.8	6.9	0.0	3.4	5.7
	Specific learning disability	435	8.3	17.2	23.7	30.1	20.0	0.5	0.2	435	18.6	19.1	52.6	5.1	2.5	0.9	1.1
	Speech or language impairment	52	3.8	13.5	15.4	36.5	26.9	1.9	1.9	52	3.8	7.7	67.3	17.3	1.9	1.9	0.0
	Traumatic brain injury	7	71.4	0.0	0.0	14.3	0.0	14.3	0.0	7	57.1	0.0	28.6	0.0	0.0	14.3	0.0
	Visual impairment	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.6: Performance Level by Domain - Grade 5

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	3687	8.7	7.0	9.8	39.3	34.8	0.4	0.1	3687	18.4	12.7	36.8	17.7	12.2	1.1	1.1	
	Female	1602	8.7	5.9	9.9	39.6	35.6	0.3	0.1	1602	17.6	12.9	36.4	17.8	13.5	1.1	0.7	
	Male	2010	8.8	8.1	9.5	39.2	34.0	0.4	0.1	2010	19.3	12.6	36.8	17.6	11.2	1.0	1.4	
	Missing	75	6.7	2.7	16.0	37.3	37.3	0.0	0.0	75	13.3	9.3	42.7	20.0	13.3	1.3	0.0	
Ethnicity	African American	737	7.2	7.2	11.3	40.4	33.2	0.7	0.0	737	19.5	11.9	40.2	15.3	11.4	0.9	0.7	
	American Indian or Alaskan Native	9	0.0	0.0	11.1	44.4	44.4	0.0	0.0	9	11.1	11.1	44.4	0.0	33.3	0.0	0.0	
	Asian	572	10.1	8.9	9.8	32.9	37.4	0.5	0.3	572	21.3	11.7	29.5	18.7	16.1	1.0	1.6	
	Hispanic or Latino	1282	8.8	6.0	8.1	40.6	35.9	0.5	0.1	1282	15.8	11.8	39.3	19.8	11.0	1.1	1.2	
	Multiple Races	530	9.6	7.5	11.9	40.6	30.4	0.0	0.0	530	22.3	13.6	35.3	15.1	11.1	0.8	1.9	
	Native Hawaiian or Other Pacific Islander	19	0.0	10.5	26.3	31.6	31.6	0.0	0.0	19	31.6	21.1	21.1	15.8	10.5	0.0	0.0	
	Other/Unknown	6	0.0	0.0	33.3	50.0	16.7	0.0	0.0	6	0.0	16.7	50.0	33.3	0.0	0.0	0.0	
	White	532	8.5	6.6	8.8	40.2	35.9	0.0	0.0	532	16.0	15.8	35.3	17.9	13.2	1.5	0.4	
	Other	Gifted	12	0.0	0.0	8.3	41.7	50.0	0.0	0.0	12	8.3	16.7	41.7	8.3	25.0	0.0	0.0
		IEP	724	11.5	12.2	16.4	42.4	16.3	0.8	0.4	724	26.0	18.8	34.1	8.8	3.2	3.5	5.7
Migrant		12	8.3	0.0	0.0	58.3	33.3	0.0	0.0	12	16.7	16.7	41.7	0.0	25.0	0.0	0.0	
Primary Disability	Autism	50	46.0	22.0	10.0	16.0	4.0	2.0	0.0	50	42.0	16.0	6.0	2.0	0.0	10.0	24.0	
	Developmental delay	4	25.0	25.0	25.0	25.0	0.0	0.0	0.0	4	75.0	0.0	0.0	0.0	0.0	25.0	0.0	
	Emotional disturbance	17	11.8	5.9	41.2	23.5	17.6	0.0	0.0	17	35.3	29.4	17.6	11.8	0.0	0.0	5.9	
	Hearing Impaired	4	0.0	0.0	0.0	25.0	25.0	0.0	50.0	4	0.0	0.0	50.0	50.0	0.0	0.0	0.0	
	Intellectual Disability	61	34.4	29.5	19.7	8.2	6.6	0.0	1.6	61	65.6	8.2	8.2	1.6	1.6	6.6	8.2	
Multiple disabilities	26	65.4	15.4	3.8	0.0	0.0	15.4	0.0	26	53.8	0.0	0.0	0.0	0.0	19.2	26.9		

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	2	50.0	50.0	0.0	0.0	0.0	0.0	0.0	2	50.0	0.0	0.0	0.0	0.0	0.0	50.0
	Other health impairment	93	3.2	15.1	18.3	45.2	17.2	1.1	0.0	93	20.4	26.9	28.0	12.9	5.4	4.3	2.2
	Specific learning disability	429	3.3	7.9	17.0	52.0	19.8	0.0	0.0	429	18.4	20.5	42.9	9.8	4.0	1.4	3.0
	Speech or language impairment	31	0.0	12.9	12.9	48.4	25.8	0.0	0.0	31	12.9	16.1	54.8	9.7	6.5	0.0	0.0
	Traumatic brain injury	4	25.0	0.0	0.0	75.0	0.0	0.0	0.0	4	25.0	25.0	50.0	0.0	0.0	0.0	0.0
	Visual impairment	1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	1	0.0	0.0	100.0	0.0	0.0	0.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.6: Performance Level by Domain - Grade 5(cont.)

		Speaking								Writing								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	3687	14.6	11.1	24.4	24.3	23.7	1.0	0.8	3687	13.0	9.0	55.2	12.5	8.4	1.0	0.9	
	Female	1602	13.7	10.4	22.6	26.4	25.5	0.9	0.4	1602	11.9	8.3	55.6	12.4	10.3	0.9	0.6	
	Male	2010	15.5	11.7	25.9	22.8	21.9	1.1	1.0	2010	14.1	9.5	55.0	12.4	6.8	1.0	1.2	
	Missing	75	10.7	10.7	25.3	18.7	33.3	1.3	0.0	75	8.0	12.0	53.3	17.3	8.0	1.3	0.0	
Ethnicity	African American	737	14.4	13.0	23.3	25.2	22.5	0.8	0.7	737	12.6	10.6	57.4	10.7	6.8	1.1	0.8	
	American Indian or Alaskan Native	9	0.0	11.1	22.2	44.4	22.2	0.0	0.0	9	0.0	11.1	55.6	11.1	22.2	0.0	0.0	
	Asian	572	16.4	11.2	21.9	22.7	24.0	1.6	2.3	572	15.2	8.7	49.7	11.7	12.2	0.9	1.6	
	Hispanic or Latino	1282	13.3	10.8	26.1	25.2	23.2	0.9	0.5	1282	11.9	7.5	57.7	13.7	7.4	1.0	0.7	
	Multiple Races	530	18.3	9.8	26.0	24.7	19.6	0.9	0.6	530	15.7	10.4	53.2	10.4	7.7	0.9	1.7	
	Native Hawaiian or Other Pacific Islander	19	26.3	5.3	26.3	31.6	10.5	0.0	0.0	19	21.1	0.0	52.6	15.8	10.5	0.0	0.0	
	Other/Unknown	6	0.0	16.7	50.0	16.7	16.7	0.0	0.0	6	0.0	0.0	66.7	33.3	0.0	0.0	0.0	
	White	532	12.6	10.9	22.7	21.6	30.6	1.3	0.2	532	11.1	9.8	54.1	14.5	9.0	1.1	0.4	
	Other	Gifted	12	0.0	0.0	25.0	41.7	33.3	0.0	0.0	12	0.0	0.0	66.7	8.3	25.0	0.0	0.0
		IEP	724	18.5	15.7	26.9	20.0	12.4	2.5	3.9	724	18.5	14.6	52.3	4.8	1.8	3.0	4.8
Migrant		12	8.3	0.0	41.7	16.7	25.0	8.3	0.0	12	8.3	8.3	58.3	0.0	25.0	0.0	0.0	
Primary Disability	Autism	50	48.0	8.0	4.0	6.0	2.0	12.0	20.0	50	40.0	2.0	26.0	0.0	0.0	14.0	18.0	
	Developmental delay	4	75.0	25.0	0.0	0.0	0.0	0.0	0.0	4	100.0	0.0	0.0	0.0	0.0	0.0	0.0	
	Emotional disturbance	17	23.5	35.3	29.4	0.0	0.0	5.9	5.9	17	11.8	35.3	47.1	0.0	0.0	0.0	5.9	
	Hearing Impaired	4	0.0	0.0	0.0	25.0	25.0	0.0	50.0	4	0.0	0.0	50.0	25.0	25.0	0.0	0.0	
	Intellectual Disability	61	47.5	19.7	16.4	4.9	1.6	4.9	4.9	61	44.3	16.4	21.3	1.6	0.0	6.6	9.8	
	Multiple disabilities	26	30.8	7.7	0.0	0.0	0.0	23.1	38.5	26	42.3	3.8	0.0	0.0	0.0	19.2	34.6	

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	2	100.0	0.0	0.0	0.0	0.0	0.0	0.0	2	50.0	0.0	0.0	0.0	0.0	0.0	50.0
	Other health impairment	93	16.1	18.3	37.6	12.9	11.8	1.1	2.2	93	18.3	19.4	49.5	7.5	1.1	2.2	2.2
	Specific learning disability	429	11.2	15.9	30.1	27.0	15.9	0.0	0.0	429	11.4	15.9	62.7	5.1	2.6	0.7	1.6
	Speech or language impairment	31	6.5	12.9	29.0	25.8	22.6	3.2	0.0	31	9.7	9.7	64.5	12.9	3.2	0.0	0.0
	Traumatic brain injury	4	25.0	25.0	25.0	25.0	0.0	0.0	0.0	4	25.0	0.0	75.0	0.0	0.0	0.0	0.0
	Visual impairment	1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	1	0.0	0.0	100.0	0.0	0.0	0.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.7: Performance Level by Domain - Grade 6

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	3432	6.2	6.3	17.9	37.9	31.1	0.3	0.2	3432	16.7	15.5	40.1	15.5	10.5	0.7	1.0	
	Female	1488	6.0	5.8	16.1	37.6	33.7	0.5	0.3	1488	14.1	15.3	41.9	16.3	11.4	0.5	0.5	
	Male	1896	6.3	6.7	19.5	37.9	29.3	0.2	0.2	1896	18.5	15.8	38.7	14.8	9.9	0.9	1.5	
	Missing	48	8.3	6.3	14.6	47.9	22.9	0.0	0.0	48	22.9	12.5	39.6	18.8	6.3	0.0	0.0	
Ethnicity	African American	760	4.7	7.2	16.4	39.1	32.4	0.1	0.0	760	17.5	15.1	40.9	15.1	10.3	0.5	0.5	
	American Indian or Alaskan Native	4	0.0	0.0	0.0	25.0	75.0	0.0	0.0	4	0.0	0.0	50.0	0.0	50.0	0.0	0.0	
	Asian	517	4.4	4.3	19.0	36.6	35.2	0.6	0.0	517	13.5	14.7	35.4	18.6	15.7	0.6	1.5	
	Hispanic or Latino	1160	7.4	6.0	19.0	39.8	27.5	0.0	0.3	1160	17.4	16.2	42.9	14.0	7.9	0.6	0.9	
	Multiple Races	502	9.2	8.2	16.5	36.5	28.3	1.0	0.4	502	19.3	14.5	40.6	14.7	8.0	1.2	1.6	
	Native Hawaiian or Other Pacific Islander	22	4.5	9.1	40.9	31.8	13.6	0.0	0.0	22	31.8	31.8	31.8	0.0	4.5	0.0	0.0	
	Other/Unknown	8	0.0	0.0	12.5	50.0	37.5	0.0	0.0	8	25.0	0.0	50.0	25.0	0.0	0.0	0.0	
	White	459	4.8	5.7	17.2	34.6	37.0	0.2	0.4	459	13.3	15.9	36.2	18.3	14.4	0.9	1.1	
	Other	Gifted	13	0.0	0.0	7.7	30.8	61.5	0.0	0.0	13	0.0	0.0	38.5	23.1	38.5	0.0	0.0
		IEP	668	5.8	9.1	28.0	40.1	15.1	0.7	1.0	668	24.6	20.7	36.8	6.4	3.9	2.4	5.2
Migrant		11	0.0	0.0	45.5	18.2	36.4	0.0	0.0	11	18.2	18.2	36.4	18.2	9.1	0.0	0.0	
Primary Disability	Autism	32	15.6	25.0	25.0	18.8	9.4	3.1	3.1	32	31.3	18.8	12.5	3.1	6.3	6.3	21.9	
	Developmental delay	6	33.3	33.3	16.7	16.7	0.0	0.0	0.0	6	66.7	0.0	16.7	0.0	0.0	16.7	0.0	
	Emotional disturbance	7	0.0	28.6	28.6	28.6	14.3	0.0	0.0	7	28.6	14.3	42.9	0.0	14.3	0.0	0.0	
	Hearing Impaired	9	0.0	11.1	11.1	0.0	11.1	11.1	55.6	9	44.4	11.1	11.1	11.1	11.1	0.0	11.1	
	Intellectual Disability	52	15.4	23.1	34.6	26.9	0.0	0.0	0.0	52	40.4	17.3	19.2	0.0	0.0	5.8	17.3	
	Multiple disabilities	24	50.0	33.3	12.5	4.2	0.0	0.0	0.0	24	50.0	4.2	0.0	0.0	0.0	20.8	25.0	

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	3	0.0	0.0	33.3	0.0	33.3	33.3	0.0	3	33.3	0.0	0.0	0.0	33.3	33.3	0.0
	Other health impairment	102	2.9	6.9	33.3	43.1	13.7	0.0	0.0	102	23.5	21.6	44.1	3.9	2.0	1.0	3.9
	Specific learning disability	388	2.1	4.6	28.9	45.9	17.8	0.5	0.3	388	19.6	23.7	42.5	8.2	3.1	0.8	2.1
	Speech or language impairment	29	0.0	6.9	6.9	48.3	37.9	0.0	0.0	29	10.3	10.3	44.8	10.3	24.1	0.0	0.0
	Traumatic brain injury	5	0.0	20.0	20.0	60.0	0.0	0.0	0.0	5	20.0	20.0	60.0	0.0	0.0	0.0	0.0
	Visual impairment	4	25.0	0.0	25.0	25.0	25.0	0.0	0.0	4	50.0	0.0	25.0	25.0	0.0	0.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.7: Performance Level by Domain - Grade 6(cont.)

		Speaking								Writing								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	3432	11.2	8.9	26.8	25.8	25.4	1.1	0.7	3432	11.9	8.8	50.5	15.9	11.0	0.8	1.1	
	Female	1488	10.2	8.1	26.1	24.9	29.1	1.2	0.3	1488	10.6	7.1	51.0	16.1	13.8	0.7	0.6	
	Male	1896	11.9	9.7	27.5	26.3	22.5	1.1	1.1	1896	12.9	10.2	50.0	15.7	8.9	0.8	1.6	
	Missing	48	14.6	8.3	18.8	31.3	27.1	0.0	0.0	48	14.6	6.3	54.2	20.8	4.2	0.0	0.0	
Ethnicity	African American	760	11.3	9.2	25.3	25.0	27.9	0.8	0.5	760	12.1	7.5	52.9	15.3	10.9	0.7	0.7	
	American Indian or Alaskan Native	4	0.0	25.0	25.0	50.0	0.0	0.0	0.0	4	0.0	0.0	50.0	25.0	25.0	0.0	0.0	
	Asian	517	10.3	8.9	23.2	25.5	29.6	1.4	1.2	517	10.1	8.5	47.8	16.6	14.5	1.0	1.5	
	Hispanic or Latino	1160	11.0	8.5	30.1	27.1	21.7	1.0	0.5	1160	11.3	10.3	52.2	15.9	8.7	0.8	0.9	
	Multiple Races	502	16.5	9.4	25.1	25.9	20.9	1.6	0.6	502	17.1	7.6	48.4	15.9	8.4	0.6	2.0	
	Native Hawaiian or Other Pacific Islander	22	13.6	22.7	45.5	9.1	9.1	0.0	0.0	22	13.6	27.3	50.0	4.5	4.5	0.0	0.0	
	Other/Unknown	8	0.0	12.5	50.0	25.0	12.5	0.0	0.0	8	0.0	25.0	50.0	12.5	12.5	0.0	0.0	
	White	459	7.0	8.3	25.7	24.4	32.0	1.3	1.3	459	9.8	7.8	47.7	17.0	15.7	0.9	1.1	
	Other	Gifted	13	0.0	7.7	15.4	30.8	46.2	0.0	0.0	13	0.0	0.0	23.1	38.5	38.5	0.0	0.0
		IEP	668	12.1	12.6	31.9	24.6	12.6	2.7	3.6	668	14.5	15.0	51.5	7.8	3.1	2.4	5.7
Migrant		11	0.0	36.4	18.2	18.2	27.3	0.0	0.0	11	9.1	18.2	36.4	27.3	9.1	0.0	0.0	
Primary Disability	Autism	32	25.0	12.5	15.6	21.9	3.1	6.3	15.6	32	21.9	9.4	31.3	3.1	3.1	6.3	25.0	
	Developmental delay	6	66.7	0.0	0.0	16.7	0.0	16.7	0.0	6	66.7	0.0	16.7	0.0	0.0	16.7	0.0	
	Emotional disturbance	7	28.6	14.3	42.9	14.3	0.0	0.0	0.0	7	28.6	14.3	42.9	14.3	0.0	0.0	0.0	
	Hearing Impaired	9	11.1	11.1	0.0	22.2	0.0	11.1	44.4	9	22.2	11.1	11.1	11.1	11.1	0.0	33.3	
	Intellectual Disability	52	28.8	25.0	15.4	13.5	1.9	5.8	9.6	52	30.8	19.2	28.8	0.0	0.0	5.8	15.4	
	Multiple disabilities	24	41.7	4.2	0.0	0.0	0.0	25.0	29.2	24	45.8	4.2	0.0	0.0	0.0	20.8	29.2	

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	3	0.0	0.0	66.7	0.0	0.0	33.3	0.0
	Other health impairment	102	10.8	12.7	31.4	25.5	16.7	1.0	2.0	102	12.7	14.7	53.9	10.8	1.0	2.9	3.9
	Specific learning disability	388	7.0	11.1	38.4	27.8	14.7	0.8	0.3	388	9.5	16.2	60.8	8.2	3.1	0.0	2.1
	Speech or language impairment	29	0.0	13.8	27.6	31.0	24.1	3.4	0.0	29	3.4	6.9	55.2	13.8	20.7	0.0	0.0
	Traumatic brain injury	5	20.0	0.0	80.0	0.0	0.0	0.0	0.0	5	20.0	20.0	60.0	0.0	0.0	0.0	0.0
	Visual impairment	4	50.0	25.0	0.0	25.0	0.0	0.0	0.0	4	50.0	0.0	25.0	25.0	0.0	0.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.8: Performance Level by Domain - Grade 7

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	3411	11.0	9.4	29.8	25.9	23.6	0.3	0.1	3411	23.7	20.6	36.9	10.4	6.9	0.6	0.9	
	Female	1474	11.0	8.4	28.1	25.2	26.8	0.4	0.1	1474	20.8	20.7	37.9	11.3	8.0	0.5	0.7	
	Male	1877	10.9	10.2	30.9	26.5	21.3	0.2	0.1	1877	25.7	20.7	36.0	9.8	6.1	0.6	1.1	
	Missing	60	13.3	8.3	36.7	23.3	16.7	1.7	0.0	60	35.0	15.0	40.0	6.7	3.3	0.0	0.0	
Ethnicity	African American	718	8.9	9.2	29.4	27.9	24.1	0.4	0.1	718	21.9	19.4	37.5	12.0	8.2	0.8	0.3	
	American Indian or Alaskan Native	6	0.0	0.0	16.7	33.3	33.3	16.7	0.0	6	16.7	16.7	33.3	33.3	0.0	0.0	0.0	
	Asian	517	7.9	11.8	29.8	23.0	27.1	0.4	0.0	517	23.6	21.3	31.1	13.2	9.3	0.8	0.8	
	Hispanic or Latino	1203	13.1	9.0	32.1	25.3	20.4	0.1	0.1	1203	25.2	21.5	38.0	8.7	5.0	0.4	1.2	
	Multiple Races	512	14.1	8.4	25.4	27.3	24.4	0.2	0.2	512	25.8	19.7	39.1	9.0	5.3	0.4	0.8	
	Native Hawaiian or Other Pacific Islander	17	0.0	29.4	11.8	41.2	17.6	0.0	0.0	17	29.4	23.5	35.3	5.9	5.9	0.0	0.0	
	Other/Unknown	9	11.1	22.2	44.4	22.2	0.0	0.0	0.0	9	33.3	33.3	33.3	0.0	0.0	0.0	0.0	
	White	429	8.9	8.4	29.8	25.4	27.0	0.5	0.0	429	20.3	19.8	37.3	11.0	9.1	0.7	1.9	
	Other	Gifted	9	11.1	0.0	33.3	11.1	44.4	0.0	0.0	9	11.1	33.3	11.1	11.1	33.3	0.0	0.0
		IEP	614	10.7	15.8	39.9	23.1	9.3	0.7	0.5	614	32.9	23.9	29.2	4.4	2.1	2.3	5.2
Migrant		12	8.3	0.0	50.0	33.3	8.3	0.0	0.0	12	25.0	16.7	33.3	16.7	0.0	8.3	0.0	
Primary Disability	Autism	28	28.6	17.9	21.4	21.4	10.7	0.0	0.0	28	28.6	17.9	17.9	10.7	7.1	7.1	10.7	
	Developmental delay	7	71.4	14.3	14.3	0.0	0.0	0.0	0.0	7	85.7	0.0	0.0	0.0	0.0	14.3	0.0	
	Emotional disturbance	14	7.1	14.3	64.3	14.3	0.0	0.0	0.0	14	50.0	14.3	21.4	7.1	0.0	0.0	7.1	
	Hearing Impaired	7	14.3	0.0	0.0	28.6	0.0	28.6	28.6	7	57.1	0.0	42.9	0.0	0.0	0.0	0.0	
	Intellectual Disability	52	19.2	38.5	28.8	9.6	1.9	1.9	0.0	52	51.9	7.7	11.5	0.0	0.0	9.6	19.2	
	Multiple disabilities	28	75.0	10.7	7.1	0.0	3.6	3.6	0.0	28	57.1	0.0	0.0	0.0	3.6	17.9	21.4	

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	4	0.0	25.0	25.0	25.0	25.0	0.0	0.0	4	50.0	0.0	50.0	0.0	0.0	0.0	0.0
	Other health impairment	71	2.8	11.3	49.3	28.2	8.5	0.0	0.0	71	21.1	31.0	36.6	4.2	2.8	0.0	4.2
	Specific learning disability	386	3.9	13.7	44.0	26.9	11.1	0.0	0.3	386	29.0	29.0	32.6	5.2	1.8	0.3	2.1
	Speech or language impairment	12	8.3	8.3	50.0	25.0	8.3	0.0	0.0	12	16.7	33.3	41.7	0.0	8.3	0.0	0.0
	Traumatic brain injury	3	33.3	33.3	33.3	0.0	0.0	0.0	0.0	3	66.7	0.0	0.0	0.0	0.0	0.0	33.3

* N = Not Attempted; E = Exempted.

Table S3.8: Performance Level by Domain - Grade 7(cont.)

		Speaking								Writing								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	3411	12.6	11.8	31.4	19.0	22.7	1.7	0.7	3411	19.3	14.2	45.9	10.7	8.4	0.7	0.9	
	Female	1474	13.0	9.7	29.6	18.5	27.1	1.7	0.5	1474	17.6	11.9	45.1	12.2	11.6	0.8	0.8	
	Male	1877	12.4	13.4	32.6	19.3	19.8	1.7	0.9	1877	20.4	16.0	46.4	9.5	6.1	0.6	1.0	
	Missing	60	13.3	16.7	35.0	21.7	10.0	3.3	0.0	60	26.7	13.3	48.3	8.3	1.7	1.7	0.0	
Ethnicity	African American	718	11.3	12.7	32.6	17.7	23.7	1.5	0.6	718	17.8	15.6	45.0	10.3	10.0	0.8	0.4	
	American Indian or Alaskan Native	6	16.7	16.7	50.0	16.7	0.0	0.0	0.0	6	0.0	0.0	50.0	33.3	0.0	16.7	0.0	
	Asian	517	12.0	12.4	29.2	18.2	25.3	2.1	0.8	517	20.1	18.0	40.6	11.0	8.9	0.6	0.8	
	Hispanic or Latino	1203	15.1	11.1	31.7	19.0	20.9	1.7	0.5	1203	20.1	13.3	47.7	10.4	6.9	0.7	0.8	
	Multiple Races	512	14.1	11.9	33.0	19.5	19.1	1.8	0.6	512	20.3	13.5	46.5	11.1	7.0	0.6	1.0	
	Native Hawaiian or Other Pacific Islander	17	11.8	17.6	11.8	41.2	17.6	0.0	0.0	17	23.5	11.8	35.3	23.5	5.9	0.0	0.0	
	Other/Unknown	9	0.0	22.2	55.6	22.2	0.0	0.0	0.0	9	33.3	22.2	44.4	0.0	0.0	0.0	0.0	
	White	429	7.2	11.2	29.1	20.7	28.4	1.9	1.4	429	17.0	10.7	48.0	10.5	11.2	0.7	1.9	
	Other	Gifted	9	0.0	11.1	44.4	11.1	33.3	0.0	0.0	9	11.1	11.1	44.4	11.1	22.2	0.0	0.0
		IEP	614	11.7	16.4	34.0	15.0	16.1	2.9	3.7	614	21.2	22.6	42.8	4.2	1.8	2.4	4.9
Migrant		12	8.3	16.7	58.3	8.3	8.3	0.0	0.0	12	8.3	25.0	58.3	0.0	0.0	8.3	0.0	
Primary Disability	Autism	28	14.3	10.7	25.0	17.9	7.1	14.3	10.7	28	21.4	21.4	28.6	10.7	3.6	3.6	10.7	
	Developmental delay	7	57.1	14.3	28.6	0.0	0.0	0.0	0.0	7	85.7	0.0	0.0	0.0	0.0	14.3	0.0	
	Emotional disturbance	14	7.1	35.7	50.0	7.1	0.0	0.0	0.0	14	35.7	14.3	42.9	0.0	0.0	0.0	7.1	
	Hearing Impaired	7	14.3	0.0	28.6	14.3	0.0	14.3	28.6	7	42.9	0.0	28.6	0.0	0.0	14.3	14.3	
	Intellectual Disability	52	32.7	23.1	11.5	1.9	9.6	7.7	13.5	52	42.3	15.4	11.5	1.9	0.0	9.6	19.2	
	Multiple disabilities	28	42.9	0.0	3.6	0.0	3.6	25.0	25.0	28	42.9	7.1	0.0	0.0	3.6	21.4	25.0	

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	4	0.0	50.0	25.0	0.0	25.0	0.0	0.0	4	25.0	25.0	25.0	25.0	0.0	0.0	0.0
	Other health impairment	71	8.5	14.1	38.0	16.9	19.7	1.4	1.4	71	11.3	22.5	56.3	5.6	1.4	0.0	2.8
	Specific learning disability	386	6.5	17.9	38.3	17.6	19.2	0.3	0.3	386	16.8	26.2	49.5	4.1	2.1	0.3	1.0
	Speech or language impairment	12	0.0	0.0	58.3	25.0	16.7	0.0	0.0	12	0.0	16.7	75.0	8.3	0.0	0.0	0.0
	Traumatic brain injury	3	0.0	33.3	0.0	0.0	0.0	0.0	66.7	3	33.3	0.0	0.0	0.0	0.0	0.0	66.7

* N = Not Attempted; E = Exempted.

Table S3.9: Performance Level by Domain - Grade 8

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	3372	11.0	9.0	27.8	25.3	26.1	0.6	0.3	3372	22.4	19.0	42.5	8.6	5.8	0.9	0.9	
	Female	1449	10.6	9.0	25.7	24.8	29.1	0.4	0.2	1449	20.1	19.3	44.0	8.3	7.0	0.9	0.4	
	Male	1881	11.2	8.9	29.2	25.8	23.7	0.7	0.4	1881	24.1	18.8	41.3	9.0	4.7	0.9	1.2	
	Missing	42	14.3	7.1	35.7	14.3	28.6	0.0	0.0	42	21.4	21.4	45.2	4.8	7.1	0.0	0.0	
Ethnicity	African American	787	10.9	10.2	27.3	25.8	23.9	1.1	0.8	787	24.7	18.8	43.1	7.0	5.0	0.9	0.6	
	American Indian or Alaskan Native	4	0.0	0.0	25.0	50.0	25.0	0.0	0.0	4	0.0	0.0	75.0	25.0	0.0	0.0	0.0	
	Asian	559	6.4	9.7	28.3	23.6	31.7	0.0	0.4	559	19.0	18.8	39.5	10.7	9.7	1.1	1.3	
	Hispanic or Latino	1076	12.1	9.3	27.3	25.5	25.4	0.5	0.0	1076	22.8	19.4	43.7	7.7	5.1	0.7	0.6	
	Multiple Races	508	16.7	7.7	27.8	24.4	22.4	0.8	0.2	508	26.2	19.5	39.0	8.9	4.1	1.0	1.4	
	Native Hawaiian or Other Pacific Islander	18	5.6	5.6	44.4	27.8	16.7	0.0	0.0	18	5.6	38.9	50.0	5.6	0.0	0.0	0.0	
	Other/Unknown	6	0.0	0.0	33.3	33.3	33.3	0.0	0.0	6	16.7	16.7	50.0	16.7	0.0	0.0	0.0	
	White	414	8.0	6.8	28.7	26.6	29.2	0.2	0.5	414	17.9	17.6	45.9	10.9	6.0	0.7	1.0	
	Other	Gifted	8	0.0	12.5	12.5	37.5	37.5	0.0	0.0	8	12.5	12.5	50.0	0.0	25.0	0.0	0.0
		IEP	598	11.9	11.0	38.1	22.1	13.5	1.5	1.8	598	30.6	22.2	32.9	4.0	1.3	4.0	4.8
Migrant		11	27.3	9.1	18.2	27.3	18.2	0.0	0.0	11	36.4	18.2	36.4	0.0	9.1	0.0	0.0	
Primary Disability	Autism	30	40.0	13.3	20.0	6.7	13.3	3.3	3.3	30	36.7	6.7	13.3	6.7	0.0	10.0	26.7	
	Deaf-blindness	2	0.0	0.0	0.0	0.0	0.0	100.0	0.0	2	50.0	0.0	0.0	0.0	0.0	50.0	0.0	
	Developmental delay	5	0.0	60.0	20.0	20.0	0.0	0.0	0.0	5	60.0	20.0	20.0	0.0	0.0	0.0	0.0	
	Emotional disturbance	12	16.7	0.0	50.0	33.3	0.0	0.0	0.0	12	41.7	8.3	41.7	0.0	0.0	8.3	0.0	
	Hearing Impaired	17	5.9	5.9	11.8	17.6	11.8	0.0	47.1	17	29.4	23.5	29.4	0.0	0.0	0.0	17.6	
	Intellectual Disability	53	17.0	39.6	39.6	3.8	0.0	0.0	0.0	53	58.5	18.9	3.8	1.9	0.0	7.5	9.4	

	Listening									Reading							
	Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Multiple disabilities	24	66.7	4.2	8.3	8.3	0.0	12.5	0.0	24	33.3	0.0	8.3	0.0	0.0	29.2	29.2	
Orthopedic impairment	4	0.0	0.0	50.0	25.0	25.0	0.0	0.0	4	25.0	25.0	25.0	0.0	25.0	0.0	0.0	
Other health impairment	75	8.0	9.3	49.3	17.3	13.3	1.3	1.3	75	36.0	22.7	29.3	4.0	0.0	4.0	4.0	
Specific learning disability	362	6.4	7.5	40.6	27.6	17.1	0.6	0.3	362	24.0	25.7	41.7	4.7	1.9	1.1	0.8	
Speech or language impairment	7	0.0	0.0	28.6	57.1	14.3	0.0	0.0	7	14.3	14.3	71.4	0.0	0.0	0.0	0.0	
Traumatic brain injury	7	28.6	28.6	14.3	0.0	28.6	0.0	0.0	7	57.1	14.3	0.0	14.3	0.0	14.3	0.0	
Visual impairment	1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0	

* N = Not Attempted; E = Exempted.

Table S3.9: Performance Level by Domain - Grade 8(cont.)

		Speaking								Writing								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	3372	11.8	9.2	32.2	19.2	24.9	2.1	0.7	3372	19.6	14.1	45.3	9.8	9.2	0.9	1.0	
	Female	1449	10.9	8.1	31.8	19.8	26.8	2.2	0.3	1449	17.9	12.9	45.2	11.0	11.7	0.9	0.4	
	Male	1881	12.3	10.0	32.4	18.8	23.3	2.1	1.0	1881	20.8	15.0	45.3	9.1	7.2	1.0	1.5	
	Missing	42	16.7	7.1	38.1	14.3	23.8	0.0	0.0	42	23.8	11.9	50.0	2.4	11.9	0.0	0.0	
Ethnicity	African American	787	11.8	11.4	32.7	17.5	23.1	2.0	1.4	787	22.1	14.0	45.5	8.6	7.8	0.8	1.3	
	American Indian or Alaskan Native	4	0.0	0.0	75.0	0.0	25.0	0.0	0.0	4	0.0	0.0	75.0	0.0	25.0	0.0	0.0	
	Asian	559	9.1	9.7	29.3	18.6	31.1	1.3	0.9	559	17.0	16.6	41.0	10.7	12.3	1.1	1.3	
	Hispanic or Latino	1076	12.9	9.2	33.4	20.1	22.8	1.3	0.4	1076	19.9	13.7	46.2	10.3	8.8	0.7	0.5	
	Multiple Races	508	15.2	8.1	33.3	19.3	17.7	6.1	0.4	508	23.2	13.8	45.7	7.7	6.7	1.4	1.6	
	Native Hawaiian or Other Pacific Islander	18	5.6	0.0	50.0	22.2	22.2	0.0	0.0	18	5.6	11.1	61.1	16.7	5.6	0.0	0.0	
	Other/Unknown	6	0.0	0.0	66.7	16.7	16.7	0.0	0.0	6	0.0	16.7	66.7	0.0	16.7	0.0	0.0	
	White	414	8.7	6.0	29.2	20.5	34.1	1.0	0.5	414	14.3	12.6	46.9	12.1	11.8	1.4	1.0	
	Other	Gifted	8	0.0	0.0	62.5	12.5	25.0	0.0	0.0	8	12.5	0.0	50.0	12.5	25.0	0.0	0.0
		IEP	598	12.2	12.5	35.3	13.9	17.7	4.3	4.0	598	22.1	18.1	41.5	5.9	3.0	3.8	5.7
Migrant		11	27.3	9.1	36.4	0.0	27.3	0.0	0.0	11	36.4	9.1	36.4	9.1	9.1	0.0	0.0	
Primary Disability	Autism	30	23.3	6.7	10.0	13.3	10.0	10.0	26.7	30	30.0	10.0	16.7	3.3	3.3	6.7	30.0	
	Deaf-blindness	2	0.0	0.0	0.0	0.0	0.0	100.0	0.0	2	0.0	0.0	0.0	0.0	0.0	100.0	0.0	
	Developmental delay	5	0.0	60.0	40.0	0.0	0.0	0.0	0.0	5	60.0	0.0	40.0	0.0	0.0	0.0	0.0	
	Emotional disturbance	12	25.0	16.7	41.7	16.7	0.0	0.0	0.0	12	25.0	16.7	58.3	0.0	0.0	0.0	0.0	
	Hearing Impaired	17	5.9	5.9	11.8	5.9	23.5	0.0	47.1	17	11.8	11.8	29.4	0.0	5.9	0.0	41.2	
	Intellectual Disability	53	32.1	17.0	37.7	0.0	3.8	9.4	0.0	53	45.3	22.6	18.9	0.0	0.0	5.7	7.5	

	Speaking									Writing								
	Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*		
Multiple disabilities	24	25.0	4.2	4.2	4.2	4.2	33.3	25.0	24	33.3	0.0	8.3	0.0	0.0	29.2	29.2		
Orthopedic impairment	4	0.0	25.0	50.0	0.0	25.0	0.0	0.0	4	0.0	25.0	25.0	0.0	25.0	0.0	25.0		
Other health impairment	75	12.0	13.3	45.3	9.3	16.0	1.3	2.7	75	28.0	18.7	38.7	5.3	1.3	2.7	5.3		
Specific learning disability	362	7.5	11.9	38.1	18.2	22.4	1.9	0.0	362	16.0	20.2	50.0	8.0	3.6	1.7	0.6		
Speech or language impairment	7	0.0	14.3	57.1	28.6	0.0	0.0	0.0	7	0.0	0.0	85.7	14.3	0.0	0.0	0.0		
Traumatic brain injury	7	42.9	14.3	14.3	0.0	28.6	0.0	0.0	7	57.1	0.0	14.3	0.0	14.3	14.3	0.0		
Visual impairment	1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0	0.0	0.0		

* N = Not Attempted; E = Exempted.

Table S3.10: Performance Level by Domain - Grade 9

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	4334	21.3	14.3	36.2	15.9	11.4	0.9	0.1	4334	35.0	24.0	31.1	4.6	3.1	1.2	0.9	
	Female	1830	20.4	13.7	36.4	16.0	12.6	0.8	0.1	1830	33.4	24.2	32.6	4.8	3.5	0.8	0.8	
	Male	2468	21.7	14.7	36.2	15.8	10.5	0.9	0.2	2468	36.1	23.9	30.1	4.5	2.9	1.4	1.1	
	Missing	36	38.9	13.9	27.8	8.3	8.3	2.8	0.0	36	44.4	19.4	16.7	2.8	2.8	13.9	0.0	
Ethnicity	African American	1029	20.3	16.1	38.9	13.9	10.0	0.8	0.0	1029	35.7	27.5	28.8	3.9	2.6	1.3	0.3	
	American Indian or Alaskan Native	12	41.7	25.0	16.7	16.7	0.0	0.0	0.0	12	66.7	0.0	16.7	0.0	0.0	16.7	0.0	
	Asian	663	11.9	10.4	38.3	19.2	18.6	1.4	0.3	663	25.0	22.8	38.2	6.5	5.7	0.6	1.2	
	Hispanic or Latino	1377	25.5	13.7	34.2	16.2	9.6	0.8	0.1	1377	37.3	23.7	29.9	4.0	2.6	1.3	1.2	
	Multiple Races	761	24.8	15.5	36.3	14.2	7.6	1.3	0.3	761	39.4	23.1	30.2	3.3	1.7	1.6	0.7	
	Native Hawaiian or Other Pacific Islander	27	25.9	7.4	44.4	18.5	3.7	0.0	0.0	27	33.3	25.9	33.3	3.7	0.0	3.7	0.0	
	Other/Unknown	16	12.5	31.3	43.8	6.3	6.3	0.0	0.0	16	43.8	37.5	12.5	0.0	6.3	0.0	0.0	
	White	449	18.0	14.9	32.7	17.4	16.5	0.2	0.2	449	33.2	20.3	31.6	7.8	4.7	0.7	1.8	
	Other	Gifted	7	0.0	0.0	14.3	28.6	57.1	0.0	0.0	7	0.0	0.0	57.1	0.0	42.9	0.0	0.0
		IEP	599	14.5	19.7	44.2	14.2	5.5	1.2	0.7	599	27.7	29.9	29.7	2.5	0.3	3.3	6.5
Migrant		21	33.3	9.5	42.9	14.3	0.0	0.0	0.0	21	52.4	28.6	19.0	0.0	0.0	0.0	0.0	
Primary Disability	Autism	31	38.7	9.7	22.6	22.6	0.0	3.2	3.2	31	19.4	12.9	35.5	0.0	0.0	6.5	25.8	
	Developmental delay	7	57.1	42.9	0.0	0.0	0.0	0.0	0.0	7	42.9	28.6	0.0	0.0	0.0	14.3	14.3	
	Emotional disturbance	10	10.0	20.0	40.0	10.0	10.0	10.0	0.0	10	30.0	10.0	60.0	0.0	0.0	0.0	0.0	
	Hearing Impaired	6	16.7	0.0	50.0	0.0	0.0	0.0	33.3	6	33.3	50.0	16.7	0.0	0.0	0.0	0.0	
	Intellectual Disability	42	28.6	38.1	31.0	0.0	0.0	0.0	2.4	42	54.8	23.8	4.8	0.0	0.0	0.0	16.7	
	Multiple disabilities	26	76.9	7.7	7.7	0.0	0.0	7.7	0.0	26	38.5	3.8	0.0	0.0	0.0	23.1	34.6	

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	2	100.0	0.0	0.0	0.0	0.0	0.0	0.0	2	0.0	0.0	0.0	0.0	0.0	50.0	50.0
	Other health impairment	67	13.4	22.4	43.3	11.9	6.0	3.0	0.0	67	29.9	35.8	23.9	1.5	1.5	6.0	1.5
	Specific learning disability	391	6.6	19.2	50.6	16.9	6.6	0.0	0.0	391	24.8	33.5	33.8	3.6	0.3	1.0	3.1
	Speech or language impairment	12	0.0	0.0	58.3	25.0	16.7	0.0	0.0	12	8.3	25.0	58.3	8.3	0.0	0.0	0.0
	Traumatic brain injury	4	0.0	25.0	75.0	0.0	0.0	0.0	0.0	4	50.0	25.0	25.0	0.0	0.0	0.0	0.0
	Visual impairment	4	0.0	25.0	25.0	0.0	25.0	25.0	0.0	4	0.0	0.0	50.0	0.0	0.0	50.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.10: Performance Level by Domain - Grade 9(cont.)

		Speaking								Writing								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	4334	17.4	16.4	34.5	15.1	12.6	3.5	0.5	4334	29.6	20.2	38.9	5.6	3.4	1.5	0.8	
	Female	1830	17.0	16.1	33.9	16.1	13.3	3.1	0.3	1830	27.0	19.4	40.5	7.0	4.3	1.2	0.5	
	Male	2468	17.6	16.6	35.0	14.5	12.0	3.6	0.6	2468	31.3	20.8	38.0	4.5	2.8	1.5	1.0	
	Missing	36	27.8	16.7	25.0	5.6	11.1	13.9	0.0	36	41.7	13.9	22.2	2.8	2.8	16.7	0.0	
Ethnicity	African American	1029	16.0	18.2	37.0	13.5	12.3	2.6	0.3	1029	27.1	25.0	37.8	4.5	3.6	1.7	0.3	
	American Indian or Alaskan Native	12	33.3	16.7	16.7	8.3	8.3	16.7	0.0	12	41.7	25.0	16.7	0.0	0.0	16.7	0.0	
	Asian	663	7.7	16.1	35.0	19.0	18.3	2.9	1.1	663	19.9	22.0	42.1	7.2	6.6	1.1	1.1	
	Hispanic or Latino	1377	22.1	14.4	33.8	14.5	10.5	4.4	0.4	1377	33.0	16.8	39.1	5.8	2.8	1.5	0.9	
	Multiple Races	761	21.0	18.5	33.0	13.9	8.8	4.3	0.4	761	35.9	18.9	37.1	4.3	1.2	1.8	0.8	
	Native Hawaiian or Other Pacific Islander	27	14.8	22.2	37.0	14.8	7.4	3.7	0.0	27	40.7	14.8	33.3	7.4	0.0	3.7	0.0	
	Other/Unknown	16	18.8	18.8	43.8	18.8	0.0	0.0	0.0	16	31.3	25.0	37.5	6.3	0.0	0.0	0.0	
	White	449	14.5	14.9	32.5	17.1	18.3	2.0	0.7	449	26.9	18.7	40.5	6.9	4.7	0.7	1.6	
	Other	Gifted	7	0.0	0.0	42.9	14.3	42.9	0.0	0.0	7	0.0	0.0	57.1	14.3	28.6	0.0	0.0
		IEP	599	8.7	17.0	43.4	16.7	6.7	3.8	3.7	599	22.5	23.2	42.1	2.3	1.3	2.8	5.7
Migrant		21	19.0	28.6	38.1	4.8	0.0	9.5	0.0	21	61.9	19.0	19.0	0.0	0.0	0.0	0.0	
Primary Disability	Autism	31	16.1	0.0	48.4	0.0	3.2	9.7	22.6	31	16.1	6.5	41.9	0.0	0.0	9.7	25.8	
	Developmental delay	7	14.3	42.9	14.3	0.0	0.0	14.3	14.3	7	42.9	28.6	0.0	0.0	0.0	14.3	14.3	
	Emotional disturbance	10	20.0	0.0	30.0	20.0	30.0	0.0	0.0	10	20.0	20.0	50.0	10.0	0.0	0.0	0.0	
	Hearing Impaired	6	16.7	16.7	33.3	0.0	0.0	0.0	33.3	6	33.3	33.3	16.7	0.0	0.0	0.0	16.7	
	Intellectual Disability	42	19.0	38.1	35.7	2.4	0.0	2.4	2.4	42	47.6	35.7	7.1	0.0	0.0	0.0	9.5	
	Multiple disabilities	26	38.5	0.0	3.8	0.0	0.0	26.9	30.8	26	42.3	3.8	0.0	0.0	0.0	23.1	30.8	

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	2	0.0	0.0	0.0	0.0	0.0	50.0	50.0	2	0.0	0.0	0.0	0.0	0.0	50.0	50.0
	Other health impairment	67	9.0	25.4	46.3	11.9	6.0	1.5	0.0	67	29.9	34.3	29.9	0.0	1.5	3.0	1.5
	Specific learning disability	391	4.9	16.4	47.3	21.5	7.7	1.8	0.5	391	17.9	23.3	51.2	3.1	1.5	0.5	2.6
	Speech or language impairment	12	0.0	0.0	58.3	33.3	8.3	0.0	0.0	12	8.3	0.0	75.0	16.7	0.0	0.0	0.0
	Traumatic brain injury	4	0.0	50.0	50.0	0.0	0.0	0.0	0.0	4	25.0	50.0	25.0	0.0	0.0	0.0	0.0
	Visual impairment	4	0.0	0.0	0.0	25.0	25.0	50.0	0.0	4	0.0	0.0	25.0	0.0	25.0	50.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.11: Performance Level by Domain - Grade 10

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	3504	14.2	13.8	35.1	17.7	18.4	0.7	0.2	3504	28.3	23.1	33.6	7.2	6.3	0.7	0.9	
	Female	1573	14.0	13.5	33.2	18.4	19.6	0.9	0.3	1573	26.9	22.9	34.8	7.9	5.7	0.7	1.1	
	Male	1920	14.1	14.2	36.7	17.1	17.3	0.5	0.1	1920	29.3	23.3	32.7	6.7	6.7	0.6	0.8	
	Missing	11	45.5	0.0	36.4	0.0	18.2	0.0	0.0	11	54.5	18.2	18.2	0.0	9.1	0.0	0.0	
Ethnicity	African American	867	14.5	14.4	37.9	17.4	14.1	1.2	0.5	867	30.3	27.0	30.4	6.7	4.2	0.7	0.7	
	American Indian or Alaskan Native	9	11.1	0.0	33.3	44.4	11.1	0.0	0.0	9	11.1	44.4	33.3	11.1	0.0	0.0	0.0	
	Asian	638	8.9	12.5	33.4	19.4	24.8	0.9	0.0	638	21.9	22.1	34.2	9.7	10.5	0.8	0.8	
	Hispanic or Latino	1095	19.4	13.5	33.3	16.3	17.3	0.2	0.0	1095	32.3	21.7	31.6	6.7	6.1	0.2	1.4	
	Multiple Races	459	10.9	14.8	36.8	18.3	18.1	1.1	0.0	459	23.7	22.7	41.0	5.2	4.8	1.7	0.9	
	Native Hawaiian or Other Pacific Islander	21	19.0	14.3	42.9	14.3	9.5	0.0	0.0	21	38.1	23.8	28.6	9.5	0.0	0.0	0.0	
	Other/Unknown	11	9.1	27.3	27.3	9.1	27.3	0.0	0.0	11	36.4	18.2	36.4	9.1	0.0	0.0	0.0	
	White	404	11.4	14.1	34.7	18.1	21.0	0.2	0.5	404	27.7	20.0	36.9	7.7	6.7	0.5	0.5	
	Other	Gifted	5	0.0	20.0	40.0	20.0	20.0	0.0	0.0	5	20.0	0.0	60.0	20.0	0.0	0.0	0.0
		IEP	477	14.7	18.4	45.1	11.9	7.3	1.3	1.3	477	29.4	29.8	28.1	3.4	0.8	1.9	6.7
Migrant		32	46.9	15.6	18.8	15.6	3.1	0.0	0.0	32	53.1	12.5	28.1	0.0	0.0	0.0	6.3	
Primary Disability	Autism	16	43.8	25.0	31.3	0.0	0.0	0.0	0.0	16	25.0	25.0	12.5	0.0	0.0	0.0	37.5	
	Developmental delay	13	38.5	23.1	23.1	0.0	7.7	7.7	0.0	13	76.9	15.4	7.7	0.0	0.0	0.0	0.0	
	Emotional disturbance	15	20.0	33.3	33.3	6.7	6.7	0.0	0.0	15	40.0	33.3	13.3	6.7	0.0	6.7	0.0	
	Hearing Impaired	9	22.2	0.0	22.2	0.0	0.0	0.0	55.6	9	55.6	33.3	11.1	0.0	0.0	0.0	0.0	
	Intellectual Disability	48	31.3	35.4	27.1	4.2	2.1	0.0	0.0	48	37.5	37.5	6.3	0.0	0.0	6.3	12.5	
	Multiple disabilities	22	63.6	22.7	0.0	0.0	0.0	9.1	4.5	22	40.9	4.5	4.5	0.0	0.0	9.1	40.9	

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	3	33.3	0.0	66.7	0.0	0.0	0.0	0.0	3	33.3	33.3	0.0	0.0	0.0	0.0	33.3
	Other health impairment	57	14.0	14.0	57.9	5.3	8.8	0.0	0.0	57	24.6	36.8	22.8	5.3	1.8	0.0	8.8
	Specific learning disability	290	5.2	16.2	51.4	17.6	8.6	1.0	0.0	290	25.9	29.0	37.6	4.1	1.0	0.7	1.7
	Speech or language impairment	2	0.0	0.0	50.0	0.0	50.0	0.0	0.0	2	0.0	50.0	0.0	0.0	50.0	0.0	0.0
	Traumatic brain injury	3	0.0	33.3	66.7	0.0	0.0	0.0	0.0	3	33.3	66.7	0.0	0.0	0.0	0.0	0.0
	Visual impairment	2	0.0	50.0	50.0	0.0	0.0	0.0	0.0	2	0.0	50.0	0.0	0.0	0.0	50.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.11: Performance Level by Domain - Grade 10(cont.)

		Speaking								Writing								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	3504	11.7	16.0	33.6	17.1	18.6	2.3	0.7	3504	22.2	19.9	41.4	8.0	6.6	0.9	0.9	
	Female	1573	12.1	15.3	32.2	18.7	18.8	2.1	0.9	1573	20.7	19.0	42.1	9.3	6.7	1.0	1.1	
	Male	1920	11.3	16.7	34.8	15.8	18.6	2.4	0.5	1920	23.3	20.7	40.9	7.0	6.5	0.8	0.7	
	Missing	11	45.5	0.0	27.3	18.2	0.0	9.1	0.0	11	45.5	27.3	9.1	0.0	18.2	0.0	0.0	
Ethnicity	African American	867	11.3	17.4	34.3	17.8	16.4	1.8	1.0	867	21.9	24.0	41.4	6.8	4.0	1.0	0.8	
	American Indian or Alaskan Native	9	11.1	22.2	44.4	11.1	11.1	0.0	0.0	9	33.3	11.1	44.4	11.1	0.0	0.0	0.0	
	Asian	638	7.4	14.9	35.1	19.1	21.6	1.1	0.8	638	15.4	21.8	41.2	8.6	11.1	1.1	0.8	
	Hispanic or Latino	1095	16.5	15.5	31.2	15.1	18.8	2.3	0.5	1095	28.7	16.1	39.0	8.1	6.6	0.5	1.1	
	Multiple Races	459	10.2	15.5	35.5	16.6	16.3	5.7	0.2	459	19.8	20.0	44.0	8.9	4.8	1.5	0.9	
	Native Hawaiian or Other Pacific Islander	21	9.5	23.8	33.3	19.0	14.3	0.0	0.0	21	23.8	19.0	52.4	0.0	4.8	0.0	0.0	
	Other/Unknown	11	9.1	18.2	54.5	9.1	9.1	0.0	0.0	11	27.3	27.3	27.3	18.2	0.0	0.0	0.0	
	White	404	8.4	16.1	33.2	18.8	21.5	1.5	0.5	404	18.3	18.8	44.8	8.7	7.9	1.0	0.5	
	Other	Gifted	5	20.0	0.0	20.0	40.0	20.0	0.0	0.0	5	20.0	0.0	60.0	20.0	0.0	0.0	0.0
		IEP	477	9.9	15.1	44.4	13.4	9.0	3.4	4.8	477	20.3	25.8	40.0	3.8	1.7	2.1	6.3
Migrant		32	43.8	25.0	12.5	15.6	3.1	0.0	0.0	32	59.4	9.4	15.6	9.4	0.0	3.1	3.1	
Primary Disability	Autism	16	12.5	43.8	25.0	0.0	6.3	0.0	12.5	16	18.8	18.8	25.0	0.0	0.0	0.0	37.5	
	Developmental delay	13	30.8	30.8	30.8	0.0	0.0	7.7	0.0	13	61.5	23.1	7.7	0.0	0.0	7.7	0.0	
	Emotional disturbance	15	6.7	0.0	60.0	20.0	6.7	6.7	0.0	15	26.7	40.0	20.0	0.0	0.0	13.3	0.0	
	Hearing Impaired	9	22.2	0.0	11.1	11.1	0.0	0.0	55.6	9	44.4	22.2	22.2	0.0	0.0	0.0	11.1	
	Intellectual Disability	48	29.2	22.9	29.2	8.3	0.0	4.2	6.3	48	39.6	18.8	22.9	0.0	0.0	6.3	12.5	
	Multiple disabilities	22	40.9	4.5	9.1	0.0	0.0	9.1	36.4	22	45.5	4.5	0.0	0.0	0.0	9.1	40.9	

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	3	0.0	0.0	66.7	0.0	0.0	0.0	33.3	3	33.3	33.3	0.0	0.0	0.0	0.0	33.3
	Other health impairment	57	8.8	15.8	42.1	17.5	7.0	1.8	7.0	57	21.1	28.1	38.6	1.8	3.5	0.0	7.0
	Specific learning disability	290	3.4	14.1	51.0	15.5	13.1	2.8	0.0	290	13.1	27.9	50.3	5.2	2.1	0.3	1.0
	Speech or language impairment	2	0.0	0.0	50.0	50.0	0.0	0.0	0.0	2	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	Traumatic brain injury	3	0.0	0.0	100.0	0.0	0.0	0.0	0.0	3	0.0	66.7	33.3	0.0	0.0	0.0	0.0
	Visual impairment	2	0.0	0.0	50.0	0.0	0.0	50.0	0.0	2	0.0	0.0	50.0	0.0	0.0	50.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.12: Performance Level by Domain - Grade 11

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	2785	10.0	13.4	33.6	19.0	22.8	1.2	0.0	2785	23.4	23.1	35.7	8.4	7.6	1.1	0.8	
	Female	1291	10.9	13.0	32.0	19.0	24.0	1.1	0.0	1291	23.2	23.5	36.0	8.0	7.4	1.2	0.7	
	Male	1481	9.2	13.6	35.0	19.1	21.7	1.3	0.1	1481	23.5	23.0	35.2	8.8	7.7	0.9	0.9	
	Missing	13	7.7	15.4	46.2	7.7	23.1	0.0	0.0	13	30.8	7.7	46.2	0.0	15.4	0.0	0.0	
Ethnicity	African American	638	11.4	15.7	34.0	16.9	20.5	1.4	0.0	638	26.6	22.7	33.4	8.0	7.2	1.7	0.3	
	American Indian or Alaskan Native	7	0.0	0.0	57.1	28.6	14.3	0.0	0.0	7	0.0	28.6	71.4	0.0	0.0	0.0	0.0	
	Asian	595	6.1	10.1	33.9	18.8	29.6	1.5	0.0	595	18.2	21.8	36.1	10.1	10.6	1.7	1.5	
	Hispanic or Latino	855	14.4	14.2	34.9	19.1	17.2	0.4	0.0	855	29.0	23.2	35.8	5.6	5.5	0.1	0.8	
	Multiple Races	331	6.0	12.7	35.6	20.2	23.0	2.1	0.3	331	16.9	27.5	36.0	7.9	9.1	2.1	0.6	
	Native Hawaiian or Other Pacific Islander	14	14.3	14.3	14.3	7.1	50.0	0.0	0.0	14	21.4	21.4	28.6	21.4	7.1	0.0	0.0	
	Other/Unknown	5	0.0	0.0	60.0	20.0	20.0	0.0	0.0	5	0.0	40.0	60.0	0.0	0.0	0.0	0.0	
	White	340	7.1	13.8	27.4	22.1	28.2	1.5	0.0	340	19.4	21.5	37.6	13.2	7.4	0.3	0.6	
	Other	Gifted	4	0.0	0.0	25.0	0.0	75.0	0.0	0.0	4	0.0	25.0	0.0	25.0	50.0	0.0	0.0
		IEP	382	14.4	16.5	38.7	16.8	11.0	2.4	0.3	382	26.4	28.0	32.5	3.7	2.4	1.3	5.8
Migrant		22	36.4	27.3	18.2	4.5	13.6	0.0	0.0	22	63.6	13.6	13.6	0.0	9.1	0.0	0.0	
Primary Disability	Autism	11	27.3	9.1	36.4	0.0	27.3	0.0	0.0	11	18.2	36.4	9.1	0.0	18.2	0.0	18.2	
	Developmental delay	3	33.3	33.3	33.3	0.0	0.0	0.0	0.0	3	33.3	0.0	33.3	0.0	0.0	0.0	33.3	
	Emotional disturbance	13	7.7	38.5	23.1	23.1	7.7	0.0	0.0	13	46.2	15.4	30.8	7.7	0.0	0.0	0.0	
	Hearing Impaired	3	0.0	0.0	33.3	0.0	0.0	33.3	33.3	3	33.3	0.0	66.7	0.0	0.0	0.0	0.0	
	Intellectual Disability	48	33.3	35.4	22.9	6.3	2.1	0.0	0.0	48	60.4	22.9	6.3	0.0	0.0	2.1	8.3	
	Multiple disabilities	20	70.0	10.0	5.0	0.0	0.0	15.0	0.0	20	45.0	5.0	0.0	0.0	0.0	10.0	40.0	

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	3	66.7	0.0	33.3	0.0	0.0	0.0	0.0	3	33.3	33.3	0.0	0.0	0.0	0.0	33.3
	Other health impairment	33	18.2	21.2	39.4	15.2	6.1	0.0	0.0	33	27.3	33.3	21.2	9.1	0.0	0.0	9.1
	Specific learning disability	238	5.5	12.2	45.0	20.6	14.7	2.1	0.0	238	18.5	29.4	43.3	3.8	2.9	0.8	1.3
	Speech or language impairment	2	0.0	0.0	50.0	50.0	0.0	0.0	0.0	2	0.0	50.0	50.0	0.0	0.0	0.0	0.0
	Traumatic brain injury	3	0.0	0.0	33.3	33.3	33.3	0.0	0.0	3	0.0	33.3	66.7	0.0	0.0	0.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.12: Performance Level by Domain - Grade 11(cont.)

		Speaking								Writing								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	2785	7.6	15.7	33.0	19.5	22.0	1.7	0.6	2785	17.8	20.5	42.4	9.8	7.6	1.1	0.7	
	Female	1291	7.9	16.4	31.6	19.1	22.5	1.8	0.7	1291	17.7	20.1	39.6	11.6	8.8	1.5	0.7	
	Male	1481	7.2	15.2	34.1	19.9	21.7	1.6	0.5	1481	18.0	20.9	44.8	8.1	6.5	0.9	0.7	
	Missing	13	23.1	0.0	53.8	15.4	7.7	0.0	0.0	13	15.4	15.4	46.2	15.4	7.7	0.0	0.0	
Ethnicity	African American	638	7.2	18.3	33.1	19.7	19.6	1.9	0.2	638	18.8	23.5	40.6	8.8	6.6	1.4	0.3	
	American Indian or Alaskan Native	7	0.0	0.0	57.1	0.0	42.9	0.0	0.0	7	0.0	0.0	85.7	14.3	0.0	0.0	0.0	
	Asian	595	4.5	12.6	34.5	20.7	26.1	0.5	1.2	595	12.8	19.3	46.2	10.1	8.9	1.3	1.3	
	Hispanic or Latino	855	12.2	19.2	33.0	16.4	17.3	1.4	0.6	855	23.9	20.1	39.9	8.9	6.1	0.5	0.7	
	Multiple Races	331	6.0	13.9	32.3	19.9	23.6	3.6	0.6	331	13.6	22.7	43.2	9.1	8.5	2.4	0.6	
	Native Hawaiian or Other Pacific Islander	14	0.0	14.3	35.7	21.4	28.6	0.0	0.0	14	14.3	21.4	21.4	28.6	14.3	0.0	0.0	
	Other/Unknown	5	0.0	0.0	60.0	20.0	20.0	0.0	0.0	5	0.0	0.0	80.0	20.0	0.0	0.0	0.0	
	White	340	4.1	9.7	30.3	24.7	28.8	2.1	0.3	340	14.7	16.8	44.1	12.9	10.0	0.9	0.6	
	Other	Gifted	4	0.0	0.0	25.0	0.0	75.0	0.0	0.0	4	0.0	25.0	0.0	50.0	25.0	0.0	0.0
		IEP	382	8.4	17.3	37.2	15.7	13.6	3.7	4.2	382	19.6	22.3	41.9	5.0	4.2	1.8	5.2
Migrant		22	22.7	31.8	27.3	0.0	13.6	4.5	0.0	22	54.5	18.2	13.6	0.0	13.6	0.0	0.0	
Primary Disability	Autism	11	18.2	18.2	18.2	9.1	27.3	0.0	9.1	11	27.3	9.1	27.3	0.0	18.2	0.0	18.2	
	Developmental delay	3	33.3	0.0	33.3	33.3	0.0	0.0	0.0	3	33.3	33.3	0.0	0.0	0.0	0.0	33.3	
	Emotional disturbance	13	0.0	38.5	38.5	23.1	0.0	0.0	0.0	13	30.8	30.8	38.5	0.0	0.0	0.0	0.0	
	Hearing Impaired	3	0.0	0.0	0.0	0.0	33.3	33.3	33.3	3	33.3	33.3	33.3	0.0	0.0	0.0	0.0	
	Intellectual Disability	48	20.8	37.5	29.2	6.3	0.0	2.1	4.2	48	47.9	27.1	14.6	0.0	0.0	2.1	8.3	
	Multiple disabilities	20	25.0	5.0	5.0	5.0	0.0	20.0	40.0	20	40.0	10.0	0.0	0.0	0.0	10.0	40.0	

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Orthopedic impairment	3	0.0	33.3	33.3	0.0	0.0	0.0	33.3	3	66.7	0.0	0.0	0.0	0.0	0.0	33.3
	Other health impairment	33	9.1	18.2	39.4	9.1	15.2	6.1	3.0	33	18.2	30.3	42.4	3.0	0.0	0.0	6.1
	Specific learning disability	238	4.6	13.4	41.6	19.3	17.6	2.5	0.8	238	11.8	20.6	52.1	7.6	5.9	1.3	0.8
	Speech or language impairment	2	0.0	0.0	100.0	0.0	0.0	0.0	0.0	2	0.0	50.0	50.0	0.0	0.0	0.0	0.0
	Traumatic brain injury	3	0.0	0.0	33.3	33.3	33.3	0.0	0.0	3	0.0	33.3	33.3	0.0	0.0	33.3	0.0

* N = Not Attempted; E = Exempted.

Table S3.13: Performance Level by Domain - Grade 12

		Listening								Reading								
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*	
Gender	Total	2077	7.1	11.0	40.3	20.1	20.6	0.7	0.2	2077	19.4	26.5	37.9	8.4	6.1	0.8	1.0	
	Female	972	6.8	10.4	40.0	22.2	20.1	0.3	0.2	972	18.8	26.6	40.0	7.8	5.5	0.3	0.9	
	Male	1092	7.4	11.5	40.2	18.3	21.2	1.1	0.2	1092	20.0	26.1	36.1	9.0	6.6	1.3	1.0	
	Missing	13	7.7	7.7	69.2	7.7	7.7	0.0	0.0	13	15.4	46.2	30.8	0.0	7.7	0.0	0.0	
Ethnicity	African American	511	7.0	12.7	42.1	16.4	20.4	1.2	0.2	511	21.3	27.6	35.0	9.2	5.5	0.8	0.6	
	American Indian or Alaskan Native	5	0.0	20.0	20.0	0.0	60.0	0.0	0.0	5	40.0	0.0	40.0	0.0	20.0	0.0	0.0	
	Asian	449	5.6	9.4	36.5	22.9	24.5	0.9	0.2	449	15.6	28.1	36.3	10.2	7.3	1.1	1.3	
	Hispanic or Latino	630	10.0	11.1	40.6	21.7	15.7	0.5	0.3	630	21.9	26.7	38.7	6.5	4.4	1.0	0.8	
	Multiple Races	200	3.5	12.5	44.5	20.5	18.5	0.5	0.0	200	17.0	23.5	45.0	6.0	7.0	0.0	1.5	
	Native Hawaiian or Other Pacific Islander	9	0.0	0.0	55.6	22.2	22.2	0.0	0.0	9	11.1	33.3	44.4	0.0	11.1	0.0	0.0	
	Other/Unknown	5	0.0	20.0	40.0	20.0	20.0	0.0	0.0	5	0.0	60.0	20.0	0.0	20.0	0.0	0.0	
	White	268	6.3	9.0	39.2	18.3	26.9	0.4	0.0	268	18.3	23.1	38.8	10.4	7.5	0.7	1.1	
	Other	Gifted	2	0.0	0.0	0.0	50.0	50.0	0.0	0.0	2	0.0	0.0	50.0	50.0	0.0	0.0	0.0
		IEP	323	11.8	16.7	44.3	17.3	7.1	1.5	1.2	323	23.2	29.7	35.0	2.2	1.2	2.8	5.9
Migrant		20	15.0	15.0	35.0	20.0	15.0	0.0	0.0	20	25.0	30.0	35.0	0.0	10.0	0.0	0.0	
Primary Disability	Autism	15	20.0	20.0	26.7	6.7	13.3	6.7	6.7	15	26.7	20.0	13.3	6.7	6.7	6.7	20.0	
	Developmental delay	2	50.0	50.0	0.0	0.0	0.0	0.0	0.0	2	50.0	50.0	0.0	0.0	0.0	0.0	0.0	
	Emotional disturbance	3	0.0	33.3	66.7	0.0	0.0	0.0	0.0	3	0.0	100.0	0.0	0.0	0.0	0.0	0.0	
	Hearing Impaired	4	0.0	0.0	25.0	0.0	0.0	0.0	75.0	4	50.0	0.0	50.0	0.0	0.0	0.0	0.0	
	Intellectual Disability	31	32.3	32.3	35.5	0.0	0.0	0.0	0.0	31	48.4	32.3	6.5	0.0	0.0	3.2	9.7	
	Multiple disabilities	22	59.1	22.7	4.5	0.0	0.0	13.6	0.0	22	40.9	4.5	0.0	0.0	0.0	22.7	31.8	

		Listening								Reading							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Other health impairment	35	8.6	20.0	51.4	11.4	8.6	0.0	0.0	35	25.7	31.4	42.9	0.0	0.0	0.0	0.0
	Specific learning disability	203	4.4	11.8	50.7	24.6	7.9	0.5	0.0	203	16.3	31.5	44.3	2.5	1.5	1.0	3.0
	Speech or language impairment	4	0.0	50.0	25.0	25.0	0.0	0.0	0.0	4	50.0	25.0	25.0	0.0	0.0	0.0	0.0
	Traumatic brain injury	1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	1	0.0	0.0	100.0	0.0	0.0	0.0	0.0

* N = Not Attempted; E = Exempted.

Table S3.13: Performance Level by Domain - Grade 12(cont.)

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
Gender	Total	2077	5.6	13.9	38.1	19.0	20.7	1.8	0.8	2077	13.3	23.4	46.1	9.5	5.5	1.2	1.0
	Female	972	5.8	14.0	38.2	18.4	20.9	2.1	0.7	972	12.9	22.5	47.5	9.4	6.0	0.8	0.9
	Male	1092	5.4	13.9	37.9	19.7	20.5	1.6	0.9	1092	13.6	24.0	44.9	9.7	5.1	1.6	1.1
	Missing	13	7.7	7.7	53.8	7.7	23.1	0.0	0.0	13	15.4	30.8	46.2	0.0	7.7	0.0	0.0
Ethnicity	African American	511	4.7	16.0	38.0	18.2	20.7	1.8	0.6	511	12.3	25.8	45.8	10.4	3.3	1.6	0.8
	American Indian or Alaskan Native	5	0.0	20.0	40.0	0.0	40.0	0.0	0.0	5	20.0	20.0	40.0	20.0	0.0	0.0	0.0
	Asian	449	3.8	13.8	38.5	20.9	19.8	1.8	1.3	449	10.2	25.2	45.4	9.4	7.3	1.3	1.1
	Hispanic or Latino	630	8.6	15.4	38.3	18.1	17.0	2.1	0.6	630	17.1	22.2	45.9	8.4	4.3	1.1	1.0
	Multiple Races	200	5.5	10.5	42.0	20.0	19.0	2.5	0.5	200	12.5	25.0	44.0	10.0	6.5	0.5	1.5
	Native Hawaiian or Other Pacific Islander	9	0.0	0.0	55.6	22.2	22.2	0.0	0.0	9	0.0	22.2	55.6	0.0	22.2	0.0	0.0
	Other/Unknown	5	0.0	0.0	60.0	20.0	20.0	0.0	0.0	5	0.0	20.0	60.0	0.0	20.0	0.0	0.0
	White	268	3.7	9.7	33.6	19.0	31.7	1.1	1.1	268	12.3	17.2	49.6	10.4	8.2	1.1	1.1
	Other	Gifted	2	0.0	0.0	50.0	0.0	50.0	0.0	0.0	2	0.0	0.0	0.0	50.0	50.0	0.0
IEP		323	8.7	14.6	43.0	15.8	10.2	2.8	5.0	323	18.3	22.0	44.0	5.0	1.5	3.1	6.2
Migrant		20	5.0	15.0	60.0	0.0	20.0	0.0	0.0	20	20.0	25.0	40.0	5.0	10.0	0.0	0.0
Primary Disability	Autism	15	13.3	13.3	40.0	0.0	20.0	6.7	6.7	15	13.3	13.3	33.3	0.0	13.3	6.7	20.0
	Developmental delay	2	50.0	50.0	0.0	0.0	0.0	0.0	0.0	2	50.0	50.0	0.0	0.0	0.0	0.0	0.0
	Emotional disturbance	3	0.0	33.3	66.7	0.0	0.0	0.0	0.0	3	33.3	33.3	33.3	0.0	0.0	0.0	0.0
	Hearing Impaired	4	0.0	0.0	25.0	0.0	0.0	0.0	75.0	4	25.0	25.0	0.0	0.0	0.0	0.0	50.0
	Intellectual Disability	31	22.6	32.3	35.5	0.0	0.0	3.2	6.5	31	51.6	25.8	12.9	0.0	0.0	3.2	6.5
	Multiple disabilities	22	31.8	13.6	4.5	0.0	0.0	13.6	36.4	22	36.4	9.1	0.0	0.0	0.0	22.7	31.8

		Speaking								Writing							
		Count	1	2	3	4	5	N*	E*	Count	1	2	3	4	5	N*	E*
	Other health impairment	35	11.4	11.4	42.9	20.0	11.4	2.9	0.0	35	14.3	37.1	37.1	5.7	0.0	5.7	0.0
	Specific learning disability	203	3.4	12.3	48.8	20.2	12.8	1.5	1.0	203	11.3	20.2	57.1	6.9	1.0	0.5	3.0
	Speech or language impairment	4	25.0	0.0	50.0	25.0	0.0	0.0	0.0	4	50.0	25.0	25.0	0.0	0.0	0.0	0.0
	Traumatic brain injury	1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	1	0.0	0.0	0.0	0.0	100.0	0.0	0.0

* N = Not Attempted; E = Exempted.

Section 4: Summative Assessment - Percentage of Students by Overall Proficiency Level

Table S4.1: Overall Proficiency Levels - Kindergarten

		Count	Emerging	Progressing	Proficient	N*
Total		9170	14.5	77.7	7.8	0.1
Gender	Female	4202	12.2	79.1	8.7	0.0
	Male	4525	15.8	77.2	6.9	0.1
	Missing	443	22.8	69.5	7.7	0.0
Ethnicity	African American	1705	12.6	79.6	7.4	0.4
	American Indian or Alaskan Native	27	7.4	81.5	11.1	0.0
	Asian	2206	11.4	74.5	14.1	0.0
	Hispanic or Latino	2427	17.6	78.7	3.6	0.0
	Multiple Races	1005	18.9	76.9	4.2	0.0
	Native Hawaiian or Other Pacific Islander	40	15.0	82.5	2.5	0.0
	Other/Unknown	41	4.9	75.6	19.5	0.0
	White	1420	10.4	81.8	7.8	0.0
Other	Missing	299	28.1	64.9	7.0	0.0
	Gifted	14	7.1	71.4	21.4	0.0
	IEP	468	37.0	59.4	2.1	1.5
	Migrant	10	30.0	60.0	10.0	0.0
	Autism	97	51.5	42.3	3.1	3.1
	Developmental delay	6	83.3	16.7	0.0	0.0
	Emotional disturbance	7	28.6	71.4	0.0	0.0
	Hearing Impaired	9	44.4	55.6	0.0	0.0
	Intellectual Disability	22	72.7	27.3	0.0	0.0
	Multiple disabilities	32	75.0	15.6	0.0	9.4
Primary Disability	Orthopedic impairment	8	25.0	75.0	0.0	0.0
	Other health impairment	37	37.8	62.2	0.0	0.0
	Specific learning disability	33	24.2	75.8	0.0	0.0

	Count	Emerging	Progressing	Proficient	N*
Speech or language impairment	206	21.4	74.8	3.4	0.5
Traumatic brain injury	1	100.0	0.0	0.0	0.0
Visual impairment	6	16.7	83.3	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.2: Overall Proficiency Levels - Grade 1

		Count	Emerging	Progressing	Proficient	N*
Total		8036	9.4	76.6	13.9	0.1
Gender	Female	3719	7.7	77.0	15.3	0.1
	Male	4105	10.0	77.2	12.7	0.1
	Missing	212	29.7	57.1	13.2	0.0
Ethnicity	African American	1529	8.7	79.9	11.2	0.2
	American Indian or Alaskan Native	20	5.0	85.0	10.0	0.0
	Asian	1816	7.3	68.4	24.0	0.2
	Hispanic or Latino	2267	10.2	81.3	8.5	0.0
	Multiple Races	894	11.1	79.4	9.5	0.0
	Native Hawaiian or Other Pacific Islander	29	3.4	82.8	13.8	0.0
	Other/Unknown	25	4.0	72.0	24.0	0.0
	White	1300	7.7	76.8	15.4	0.1
Other	Missing	156	36.5	51.3	12.2	0.0
	Gifted	38	0.0	50.0	50.0	0.0
	IEP	497	29.0	62.8	7.0	1.2
	Migrant	15	20.0	73.3	6.7	0.0
	Autism	83	53.0	44.6	1.2	1.2
	Developmental delay	4	50.0	0.0	0.0	50.0
	Emotional disturbance	4	25.0	75.0	0.0	0.0
	Hearing Impaired	4	50.0	50.0	0.0	0.0
	Intellectual Disability	38	57.9	42.1	0.0	0.0
	Multiple disabilities	31	80.6	12.9	0.0	6.5
Primary Disability	Orthopedic impairment	15	26.7	40.0	33.3	0.0
	Other health impairment	52	25.0	69.2	5.8	0.0
	Specific learning disability	73	16.4	76.7	6.8	0.0

	Count	Emerging	Progressing	Proficient	N*
Speech or language impairment	185	10.3	79.5	10.3	0.0
Traumatic brain injury	3	0.0	66.7	0.0	33.3
Visual impairment	3	0.0	66.7	33.3	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.3: Overall Proficiency Levels - Grade 2

		Count	Emerging	Progressing	Proficient	N*
Total		6904	7.3	66.7	25.9	0.1
Gender	Female	3127	6.3	64.1	29.5	0.1
	Male	3660	8.1	68.8	23.0	0.1
	Missing	117	8.5	73.5	17.9	0.0
Ethnicity	African American	1372	7.9	68.6	23.4	0.1
	American Indian or Alaskan Native	12	8.3	75.0	16.7	0.0
	Asian	1391	6.4	56.8	36.8	0.0
	Hispanic or Latino	2159	7.4	71.8	20.7	0.0
	Multiple Races	849	9.9	69.1	20.8	0.1
	Native Hawaiian or Other Pacific Islander	41	9.8	73.2	17.1	0.0
	Other/Unknown	18	16.7	55.6	27.8	0.0
Other	White	1062	5.4	64.9	29.7	0.1
	Gifted	38	0.0	31.6	68.4	0.0
	IEP	581	20.7	71.9	6.5	0.9
	Migrant	29	6.9	82.8	10.3	0.0
	Autism	83	42.2	48.2	6.0	3.6
	Deaf-blindness	1	100.0	0.0	0.0	0.0
	Developmental delay	3	100.0	0.0	0.0	0.0
Primary Disability	Emotional disturbance	17	0.0	100.0	0.0	0.0
	Hearing Impaired	9	11.1	66.7	22.2	0.0
	Intellectual Disability	34	38.2	61.8	0.0	0.0
	Multiple disabilities	35	82.9	14.3	0.0	2.9
	Orthopedic impairment	7	0.0	71.4	14.3	14.3
	Other health impairment	61	16.4	83.6	0.0	0.0
	Specific learning disability	192	9.4	85.9	4.7	0.0

	Count	Emerging	Progressing	Proficient	N*
Speech or language impairment	129	4.7	79.8	15.5	0.0
Traumatic brain injury	3	66.7	0.0	33.3	0.0
Visual impairment	1	0.0	100.0	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.4: Overall Proficiency Levels - Grade 3

		Count	Emerging	Progressing	Proficient	N*
Total		5716	7.0	69.0	23.9	0.0
Gender	Female	2675	6.8	66.4	26.8	0.0
	Male	3020	7.2	71.5	21.3	0.0
	Missing	21	14.3	47.6	38.1	0.0
Ethnicity	African American	1119	6.3	71.5	22.2	0.0
	American Indian or Alaskan Native	12	8.3	50.0	41.7	0.0
	Asian	963	6.1	60.9	32.9	0.1
	Hispanic or Latino	1914	7.7	71.7	20.6	0.0
	Multiple Races	838	8.5	70.6	20.9	0.0
	Native Hawaiian or Other Pacific Islander	28	3.6	78.6	17.9	0.0
	Other/Unknown	6	0.0	66.7	33.3	0.0
Other	White	836	6.2	67.2	26.6	0.0
	Gifted	36	0.0	44.4	55.6	0.0
	IEP	776	14.3	79.3	6.4	0.0
	Migrant	34	14.7	70.6	14.7	0.0
	Autism	83	44.6	49.4	6.0	0.0
	Deaf-blindness	1	100.0	0.0	0.0	0.0
	Developmental delay	8	62.5	37.5	0.0	0.0
Primary Disability	Emotional disturbance	13	0.0	76.9	23.1	0.0
	Hearing Impaired	11	18.2	72.7	9.1	0.0
	Intellectual Disability	47	31.9	68.1	0.0	0.0
	Multiple disabilities	23	87.0	13.0	0.0	0.0
	Orthopedic impairment	5	20.0	60.0	20.0	0.0
	Other health impairment	90	14.4	83.3	2.2	0.0
	Specific learning disability	388	3.6	91.2	5.2	0.0

	Count	Emerging	Progressing	Proficient	N*
Speech or language impairment	109	0.9	79.8	19.3	0.0
Traumatic brain injury	3	33.3	66.7	0.0	0.0
Visual impairment	2	0.0	100.0	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.5: Overall Proficiency Levels - Grade 4

		Count	Emerging	Progressing	Proficient	N*
Total		4692	10.9	63.4	25.5	0.2
Gender	Female	2118	10.9	62.4	26.6	0.1
	Male	2506	11.1	64.4	24.3	0.2
	Missing	68	7.4	58.8	33.8	0.0
Ethnicity	African American	882	11.5	62.5	26.0	0.1
	American Indian or Alaskan Native	9	0.0	77.8	22.2	0.0
	Asian	751	10.4	59.0	30.4	0.3
	Hispanic or Latino	1666	11.9	65.6	22.4	0.1
	Multiple Races	689	11.6	64.0	24.1	0.3
	Native Hawaiian or Other Pacific Islander	27	11.1	70.4	18.5	0.0
	Other/Unknown	8	50.0	50.0	0.0	0.0
Other	White	660	7.3	63.3	29.1	0.3
	Gifted	12	0.0	33.3	66.7	0.0
	IEP	766	19.3	73.1	6.5	1.0
	Migrant	17	0.0	88.2	11.8	0.0
	Autism	56	41.1	53.6	1.8	3.6
	Developmental delay	6	66.7	33.3	0.0	0.0
	Emotional disturbance	15	26.7	66.7	6.7	0.0
Primary Disability	Hearing Impaired	6	50.0	50.0	0.0	0.0
	Intellectual Disability	55	58.2	40.0	1.8	0.0
	Multiple disabilities	31	80.6	0.0	0.0	19.4
	Orthopedic impairment	9	22.2	66.7	11.1	0.0
	Other health impairment	87	20.7	71.3	8.0	0.0
	Specific learning disability	435	6.9	86.0	7.1	0.0
	Speech or language impairment	52	1.9	82.7	15.4	0.0

	Count	Emerging	Progressing	Proficient	N*
Traumatic brain injury	7	71.4	28.6	0.0	0.0
Visual impairment	1	100.0	0.0	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.6: Overall Proficiency Levels - Grade 5

		Count	Emerging	Progressing	Proficient	N*
Total		3687	14.3	68.4	17.2	0.1
Gender	Female	1602	13.4	67.4	19.0	0.1
	Male	2010	15.2	69.1	15.6	0.1
	Missing	75	9.3	70.7	20.0	0.0
Ethnicity	African American	737	13.4	72.9	13.7	0.0
	American Indian or Alaskan Native	9	0.0	66.7	33.3	0.0
	Asian	572	18.2	60.8	20.5	0.5
	Hispanic or Latino	1282	13.7	69.3	16.8	0.2
	Multiple Races	530	14.9	69.6	15.5	0.0
	Native Hawaiian or Other Pacific Islander	19	10.5	68.4	21.1	0.0
	Other/Unknown	6	0.0	66.7	33.3	0.0
	White	532	13.0	66.7	20.3	0.0
Other	Gifted	12	0.0	66.7	33.3	0.0
	IEP	724	21.1	72.1	6.2	0.6
	Migrant	12	8.3	75.0	16.7	0.0
	Autism	50	68.0	32.0	0.0	0.0
	Developmental delay	4	50.0	50.0	0.0	0.0
	Emotional disturbance	17	11.8	88.2	0.0	0.0
	Hearing Impaired	4	0.0	50.0	50.0	0.0
	Intellectual Disability	61	62.3	36.1	1.6	0.0
Primary Disability	Multiple disabilities	26	80.8	3.8	0.0	15.4
	Orthopedic impairment	2	100.0	0.0	0.0	0.0
	Other health impairment	93	14.0	78.5	7.5	0.0
	Specific learning disability	429	8.6	83.7	7.7	0.0
	Speech or language impairment	31	9.7	83.9	6.5	0.0

	Count	Emerging	Progressing	Proficient	N*
Traumatic brain injury	4	25.0	75.0	0.0	0.0
Visual impairment	1	0.0	100.0	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.7: Overall Proficiency Levels - Grade 6

		Count	Emerging	Progressing	Proficient	N*
Total		3432	11.5	70.4	18.1	0.0
Gender	Female	1488	10.7	69.8	19.5	0.0
	Male	1896	12.1	71.0	16.9	0.0
	Missing	48	14.6	66.7	18.8	0.0
Ethnicity	African American	760	10.9	72.5	16.6	0.0
	American Indian or Alaskan Native	4	0.0	75.0	25.0	0.0
	Asian	517	8.1	66.9	25.0	0.0
	Hispanic or Latino	1160	12.1	73.1	14.8	0.0
	Multiple Races	502	16.5	68.7	14.7	0.0
	Native Hawaiian or Other Pacific Islander	22	13.6	81.8	4.5	0.0
	Other/Unknown	8	0.0	75.0	25.0	0.0
Other	White	459	9.6	65.4	25.1	0.0
	Gifted	13	0.0	46.2	53.8	0.0
	IEP	668	15.1	77.5	7.3	0.0
	Migrant	11	0.0	72.7	27.3	0.0
	Autism	32	43.8	43.8	12.5	0.0
	Developmental delay	6	66.7	33.3	0.0	0.0
	Emotional disturbance	7	28.6	57.1	14.3	0.0
Primary Disability	Hearing Impaired	9	44.4	44.4	11.1	0.0
	Intellectual Disability	52	38.5	55.8	5.8	0.0
	Multiple disabilities	24	83.3	12.5	4.2	0.0
	Orthopedic impairment	3	0.0	100.0	0.0	0.0
	Other health impairment	102	9.8	87.3	2.9	0.0
	Specific learning disability	388	5.9	87.1	7.0	0.0
	Speech or language impairment	29	6.9	69.0	24.1	0.0

	Count	Emerging	Progressing	Proficient	N*
Traumatic brain injury	5	20.0	80.0	0.0	0.0
Visual impairment	4	25.0	50.0	25.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.8: Overall Proficiency Levels - Grade 7

		Count	Emerging	Progressing	Proficient	N*
Total		3411	18.0	70.9	11.1	0.0
Gender	Female	1474	16.8	69.5	13.6	0.1
	Male	1877	18.8	72.1	9.2	0.0
	Missing	60	23.3	68.3	8.3	0.0
Ethnicity	African American	718	15.9	70.5	13.5	0.1
	American Indian or Alaskan Native	6	16.7	83.3	0.0	0.0
	Asian	517	17.4	70.4	12.2	0.0
	Hispanic or Latino	1203	19.9	70.8	9.3	0.0
	Multiple Races	512	19.9	71.5	8.6	0.0
	Native Hawaiian or Other Pacific Islander	17	23.5	64.7	11.8	0.0
	Other/Unknown	9	22.2	77.8	0.0	0.0
Other	White	429	14.5	71.8	13.8	0.0
	Gifted	9	0.0	66.7	33.3	0.0
	IEP	614	23.0	73.9	2.9	0.2
	Migrant	12	8.3	91.7	0.0	0.0
	Autism	28	42.9	50.0	7.1	0.0
Primary Disability	Developmental delay	7	71.4	28.6	0.0	0.0
	Emotional disturbance	14	21.4	78.6	0.0	0.0
	Hearing Impaired	7	57.1	42.9	0.0	0.0
	Intellectual Disability	52	57.7	42.3	0.0	0.0
	Multiple disabilities	28	85.7	7.1	3.6	3.6
	Orthopedic impairment	4	25.0	75.0	0.0	0.0
	Other health impairment	71	9.9	87.3	2.8	0.0
	Specific learning disability	386	13.2	83.7	3.1	0.0
	Speech or language impairment	12	0.0	91.7	8.3	0.0

	Count	Emerging	Progressing	Proficient	N*
Traumatic brain injury	3	66.7	33.3	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.9: Overall Proficiency Levels - Grade 8

		Count	Emerging	Progressing	Proficient	N*
Total		3372	17.3	72.0	10.6	0.1
Gender	Female	1449	16.3	71.5	12.1	0.1
	Male	1881	17.9	72.5	9.5	0.2
	Missing	42	21.4	69.0	9.5	0.0
Ethnicity	African American	787	19.3	71.8	8.5	0.4
	American Indian or Alaskan Native	4	0.0	75.0	25.0	0.0
	Asian	559	13.8	71.4	14.8	0.0
	Hispanic or Latino	1076	18.2	71.6	10.1	0.1
	Multiple Races	508	21.7	70.1	8.3	0.0
	Native Hawaiian or Other Pacific Islander	18	5.6	88.9	5.6	0.0
	Other/Unknown	6	0.0	100.0	0.0	0.0
Other	White	414	11.1	75.6	13.3	0.0
	Gifted	8	0.0	75.0	25.0	0.0
	IEP	598	21.7	73.6	4.0	0.7
	Migrant	11	27.3	63.6	9.1	0.0
	Autism	30	53.3	36.7	6.7	3.3
	Deaf-blindness	2	50.0	0.0	0.0	50.0
	Developmental delay	5	60.0	40.0	0.0	0.0
Primary Disability	Emotional disturbance	12	16.7	83.3	0.0	0.0
	Hearing Impaired	17	47.1	47.1	5.9	0.0
	Intellectual Disability	53	47.2	52.8	0.0	0.0
	Multiple disabilities	24	75.0	16.7	0.0	8.3
	Orthopedic impairment	4	0.0	75.0	25.0	0.0
	Other health impairment	75	16.0	84.0	0.0	0.0
	Specific learning disability	362	11.3	83.4	5.2	0.0

	Count	Emerging	Progressing	Proficient	N*
Speech or language impairment	7	0.0	100.0	0.0	0.0
Traumatic brain injury	7	57.1	28.6	14.3	0.0
Visual impairment	1	100.0	0.0	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.10: Overall Proficiency Levels - Grade 9

		Count	Emerging	Progressing	Proficient	N*
Total		4334	30.7	64.3	5.0	0.0
Gender	Female	1830	29.5	64.5	6.0	0.1
	Male	2468	31.2	64.4	4.3	0.0
	Missing	36	52.8	44.4	2.8	0.0
Ethnicity	African American	1029	29.8	65.6	4.6	0.0
	American Indian or Alaskan Native	12	66.7	33.3	0.0	0.0
	Asian	663	19.3	70.7	9.8	0.2
	Hispanic or Latino	1377	34.5	61.6	3.8	0.1
	Multiple Races	761	36.4	61.2	2.4	0.0
	Native Hawaiian or Other Pacific Islander	27	33.3	66.7	0.0	0.0
	Other/Unknown	16	31.3	68.8	0.0	0.0
Other	White	449	26.9	65.5	7.6	0.0
	Gifted	7	0.0	71.4	28.6	0.0
	IEP	599	25.5	72.1	2.0	0.3
	Migrant	21	42.9	57.1	0.0	0.0
	Autism	31	48.4	51.6	0.0	0.0
	Developmental delay	7	85.7	14.3	0.0	0.0
	Emotional disturbance	10	20.0	80.0	0.0	0.0
Primary Disability	Hearing Impaired	6	50.0	50.0	0.0	0.0
	Intellectual Disability	42	54.8	45.2	0.0	0.0
	Multiple disabilities	26	84.6	7.7	0.0	7.7
	Orthopedic impairment	2	100.0	0.0	0.0	0.0
	Other health impairment	67	26.9	71.6	1.5	0.0
	Specific learning disability	391	15.3	81.8	2.8	0.0
	Speech or language impairment	12	0.0	91.7	8.3	0.0

	Count	Emerging	Progressing	Proficient	N*
Traumatic brain injury	4	25.0	75.0	0.0	0.0
Visual impairment	4	25.0	75.0	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.11: Overall Proficiency Levels - Grade 10

		Count	Emerging	Progressing	Proficient	N*
Total		3504	23.4	67.5	9.1	0.0
Gender	Female	1573	23.2	67.3	9.5	0.0
	Male	1920	23.4	67.8	8.9	0.0
	Missing	11	45.5	45.5	9.1	0.0
Ethnicity	African American	867	23.2	70.0	6.8	0.0
	American Indian or Alaskan Native	9	11.1	88.9	0.0	0.0
	Asian	638	17.7	67.6	14.7	0.0
	Hispanic or Latino	1095	28.6	62.6	8.8	0.0
	Multiple Races	459	21.8	72.1	6.1	0.0
	Native Hawaiian or Other Pacific Islander	21	28.6	71.4	0.0	0.0
	Other/Unknown	11	18.2	72.7	9.1	0.0
Other	White	404	20.5	69.1	10.4	0.0
	Gifted	5	20.0	60.0	20.0	0.0
	IEP	477	26.8	71.3	1.9	0.0
	Migrant	32	62.5	37.5	0.0	0.0
	Autism	16	68.8	31.3	0.0	0.0
Primary Disability	Developmental delay	13	61.5	38.5	0.0	0.0
	Emotional disturbance	15	13.3	86.7	0.0	0.0
	Hearing Impaired	9	77.8	22.2	0.0	0.0
	Intellectual Disability	48	58.3	41.7	0.0	0.0
	Multiple disabilities	22	90.9	9.1	0.0	0.0
	Orthopedic impairment	3	33.3	66.7	0.0	0.0
	Other health impairment	57	21.1	73.7	5.3	0.0
	Specific learning disability	290	13.4	84.5	2.1	0.0
	Speech or language impairment	2	0.0	100.0	0.0	0.0

	Count	Emerging	Progressing	Proficient	N*
Traumatic brain injury	3	0.0	100.0	0.0	0.0
Visual impairment	2	50.0	50.0	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.12: Overall Proficiency Levels - Grade 11

		Count	Emerging	Progressing	Proficient	N*
Total		2785	18.5	70.7	10.7	0.1
Gender	Female	1291	19.4	68.9	11.6	0.1
	Male	1481	17.8	72.2	9.9	0.1
	Missing	13	15.4	69.2	15.4	0.0
Ethnicity	African American	638	21.3	69.1	9.6	0.0
	American Indian or Alaskan Native	7	0.0	100.0	0.0	0.0
	Asian	595	12.8	74.3	12.8	0.2
	Hispanic or Latino	855	24.2	67.5	8.3	0.0
	Multiple Races	331	15.7	73.1	11.2	0.0
	Native Hawaiian or Other Pacific Islander	14	14.3	57.1	28.6	0.0
	Other/Unknown	5	0.0	100.0	0.0	0.0
Other	White	340	12.6	72.4	14.7	0.3
	Gifted	4	0.0	25.0	75.0	0.0
	IEP	382	23.8	71.5	4.2	0.5
	Migrant	22	54.5	36.4	9.1	0.0
	Autism	11	36.4	45.5	18.2	0.0
	Developmental delay	3	33.3	66.7	0.0	0.0
	Emotional disturbance	13	38.5	61.5	0.0	0.0
Primary Disability	Hearing Impaired	3	33.3	66.7	0.0	0.0
	Intellectual Disability	48	54.2	43.8	2.1	0.0
	Multiple disabilities	20	80.0	10.0	0.0	10.0
	Orthopedic impairment	3	66.7	33.3	0.0	0.0
	Other health impairment	33	24.2	75.8	0.0	0.0
	Specific learning disability	238	11.8	83.2	5.0	0.0
	Speech or language impairment	2	0.0	100.0	0.0	0.0

	Count	Emerging	Progressing	Proficient	N*
Traumatic brain injury	3	0.0	100.0	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Table S4.13: Overall Proficiency Levels - Grade 12

		Count	Emerging	Progressing	Proficient	N*
Gender	Total	2077	14.2	76.8	8.9	0.1
	Female	972	13.9	77.4	8.7	0.0
	Male	1092	14.6	76.2	9.1	0.2
	Missing	13	7.7	84.6	7.7	0.0
Ethnicity	African American	511	14.1	77.9	8.0	0.0
	American Indian or Alaskan Native	5	20.0	60.0	20.0	0.0
	Asian	449	12.5	77.1	10.5	0.0
	Hispanic or Latino	630	18.7	75.1	6.2	0.0
	Multiple Races	200	10.0	78.5	11.0	0.5
	Native Hawaiian or Other Pacific Islander	9	0.0	88.9	11.1	0.0
	Other/Unknown	5	0.0	80.0	20.0	0.0
Other	White	268	10.4	76.9	12.3	0.4
	Gifted	2	0.0	50.0	50.0	0.0
	IEP	323	23.5	73.4	2.5	0.6
Primary Disability	Migrant	20	20.0	70.0	10.0	0.0
	Autism	15	40.0	46.7	13.3	0.0
	Developmental delay	2	100.0	0.0	0.0	0.0
	Emotional disturbance	3	0.0	100.0	0.0	0.0
	Hearing Impaired	4	50.0	50.0	0.0	0.0
	Intellectual Disability	31	51.6	48.4	0.0	0.0
	Multiple disabilities	22	86.4	4.5	0.0	9.1
	Other health impairment	35	20.0	80.0	0.0	0.0
	Specific learning disability	203	11.3	85.7	3.0	0.0
	Speech or language impairment	4	25.0	75.0	0.0	0.0
	Traumatic brain injury	1	0.0	100.0	0.0	0.0

* N = All domains are either Non-Attempted or Exempted

Section 5: Summative Assessment - Marginal Reliability

Figure S5.1: Marginal Reliability by Gender

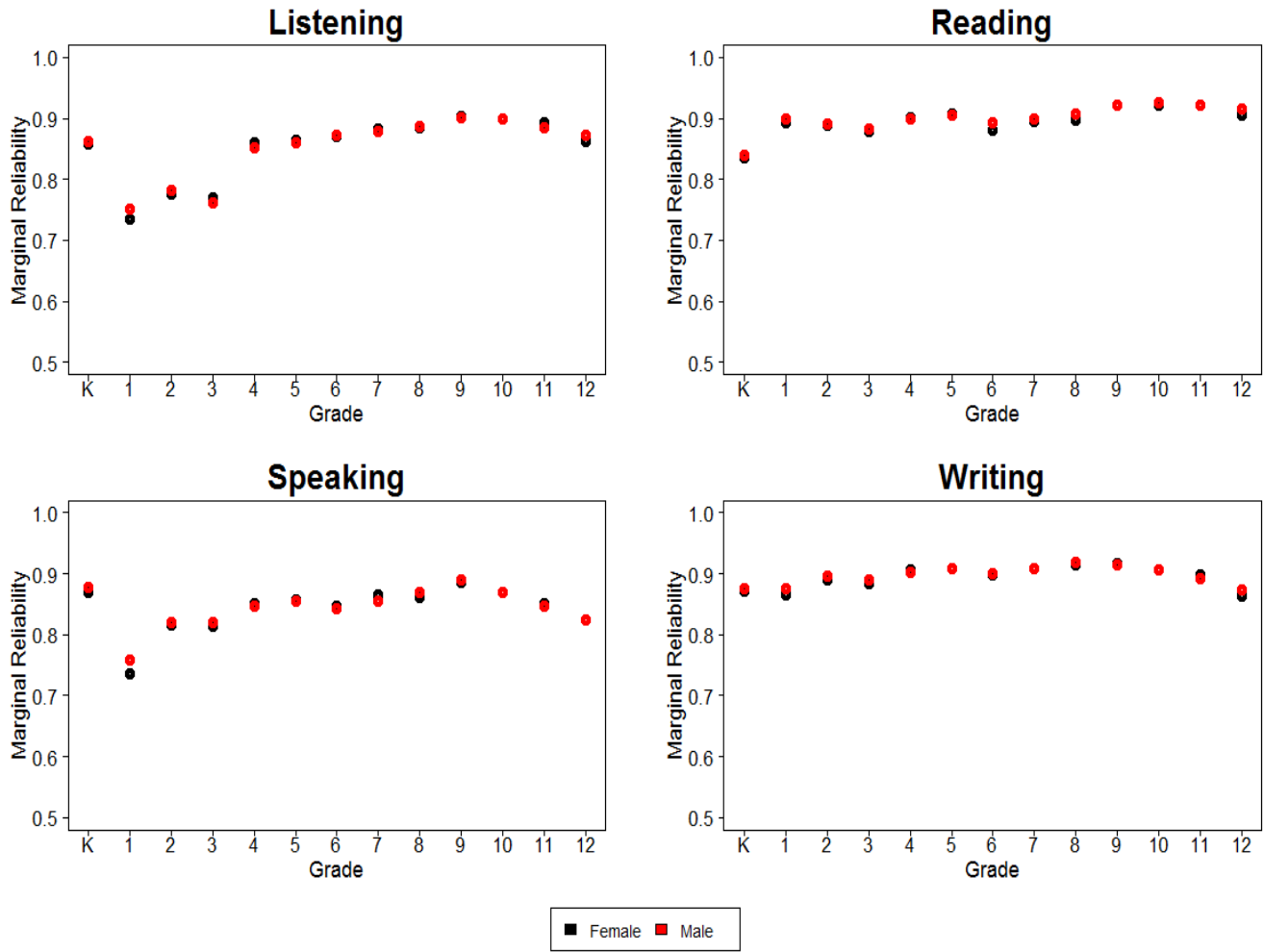


Figure S5.2: Marginal Reliability by Ethnicity

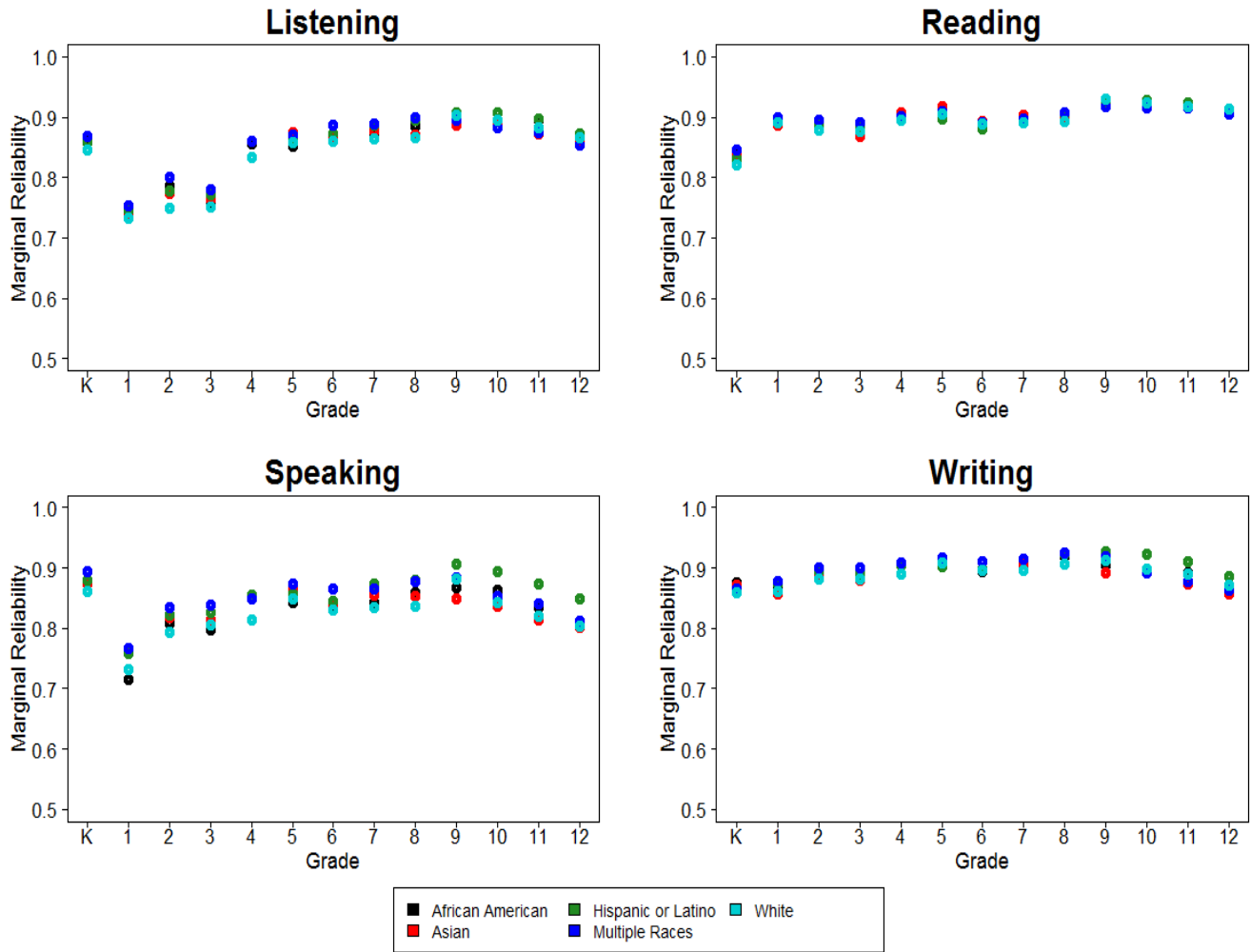


Figure S5.3: Marginal Reliability by IEP

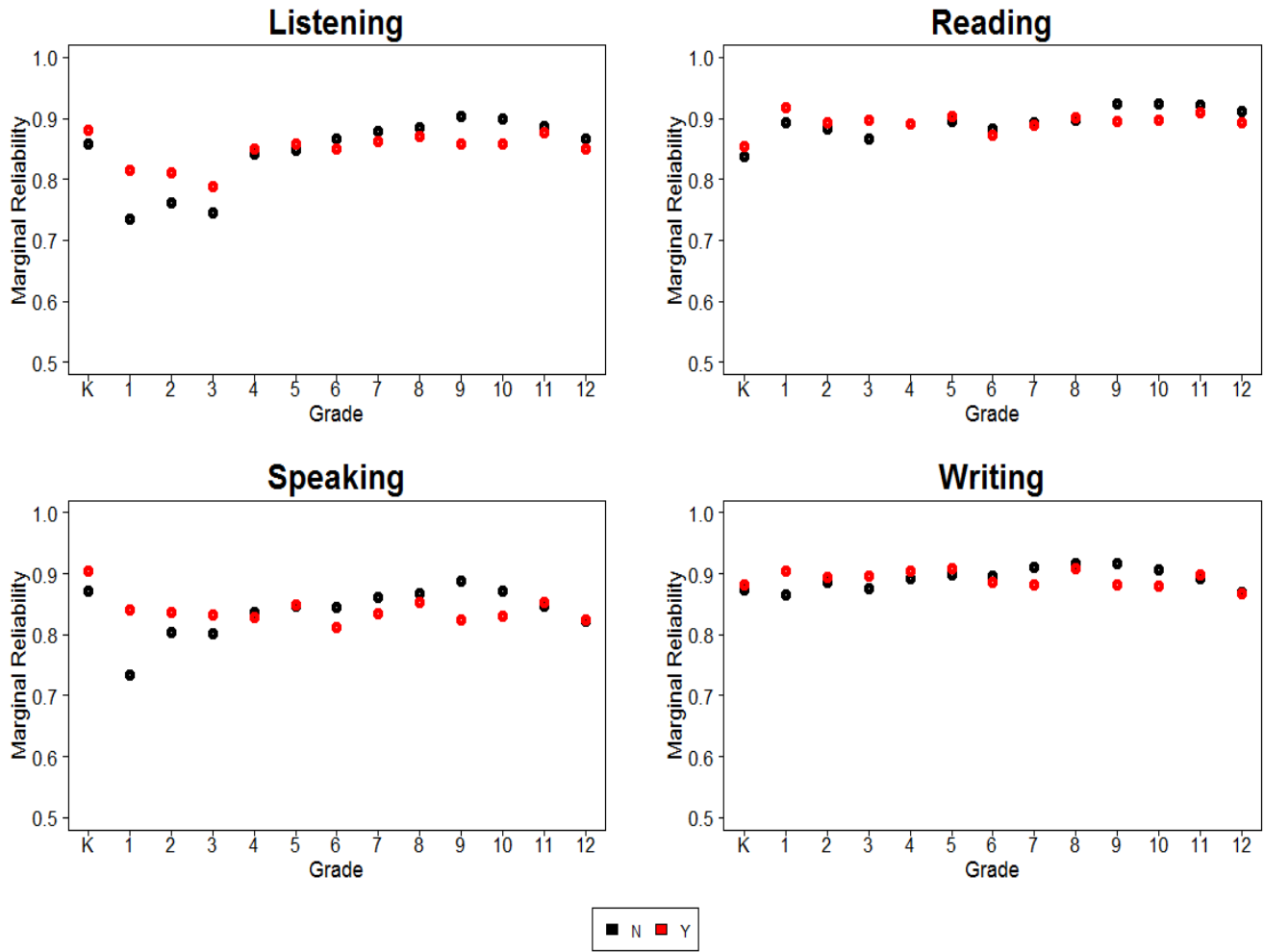
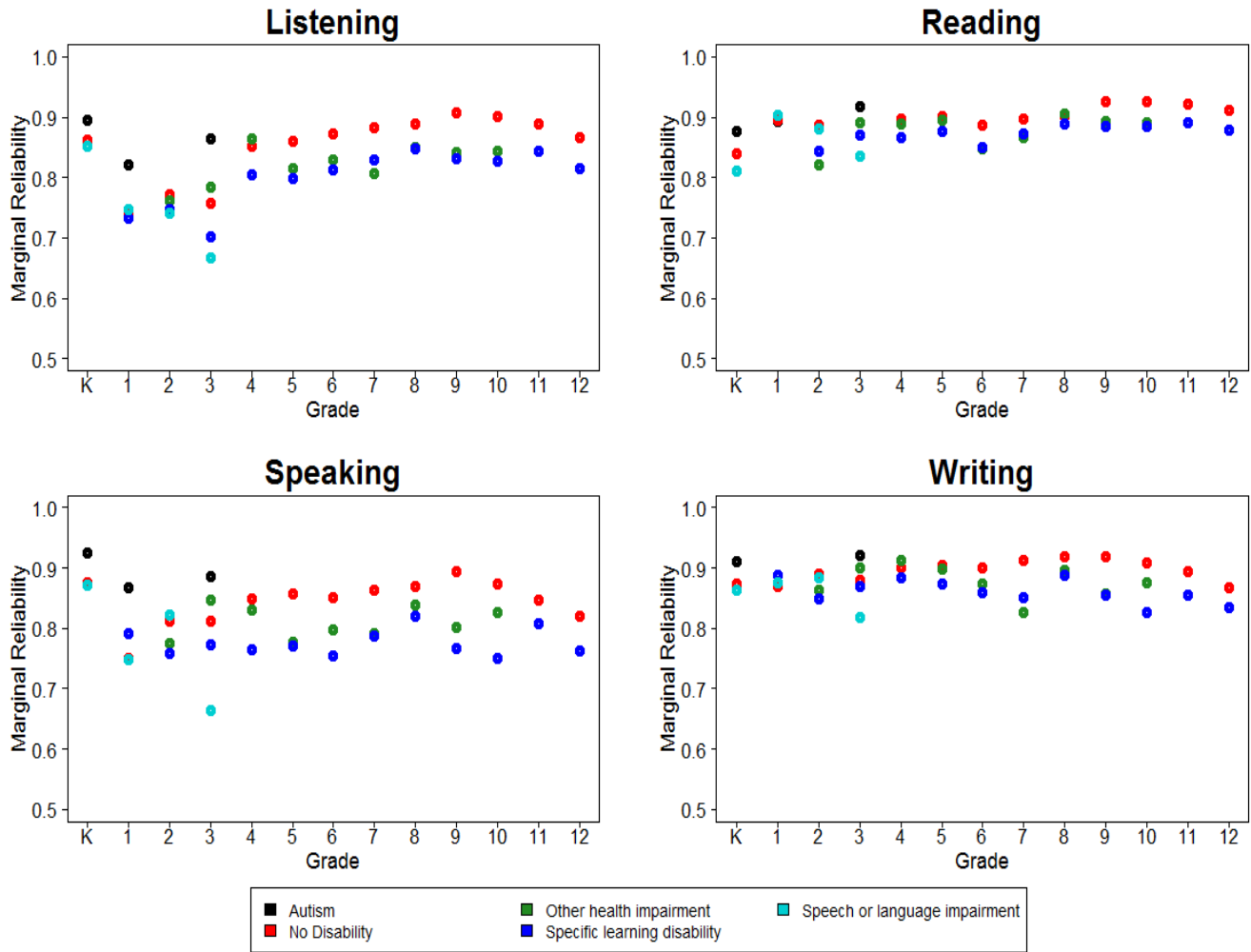


Figure S5.4: Marginal Reliability by Primary Disability



Section 6: Summative Assessment - Conditional Standard Error of Measurement (CSEM)

Figure S6.1: Conditional Standard Error of Measurement for Kindergarten

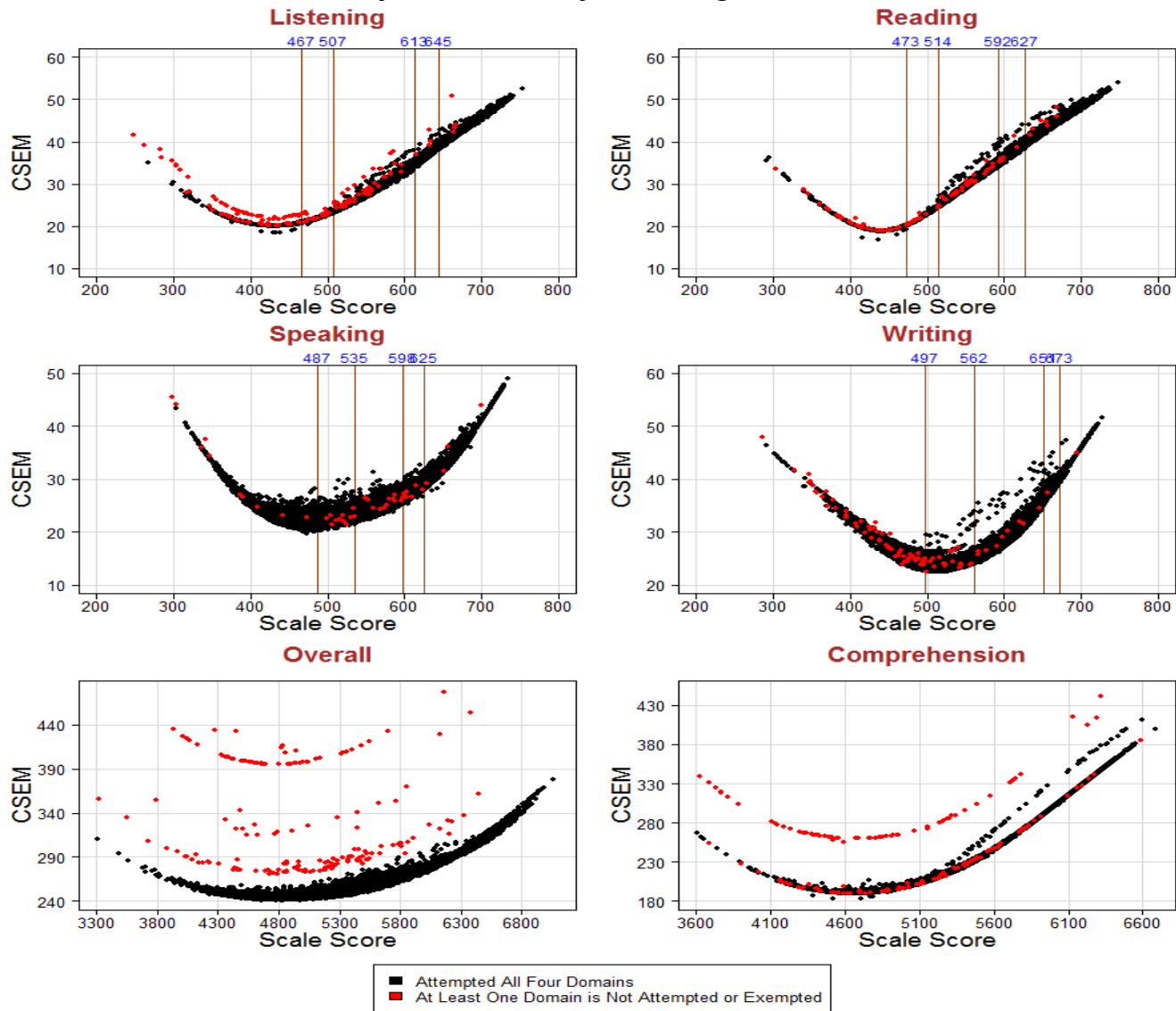


Figure S6.2: Conditional Standard Error of Measurement for Grade 1

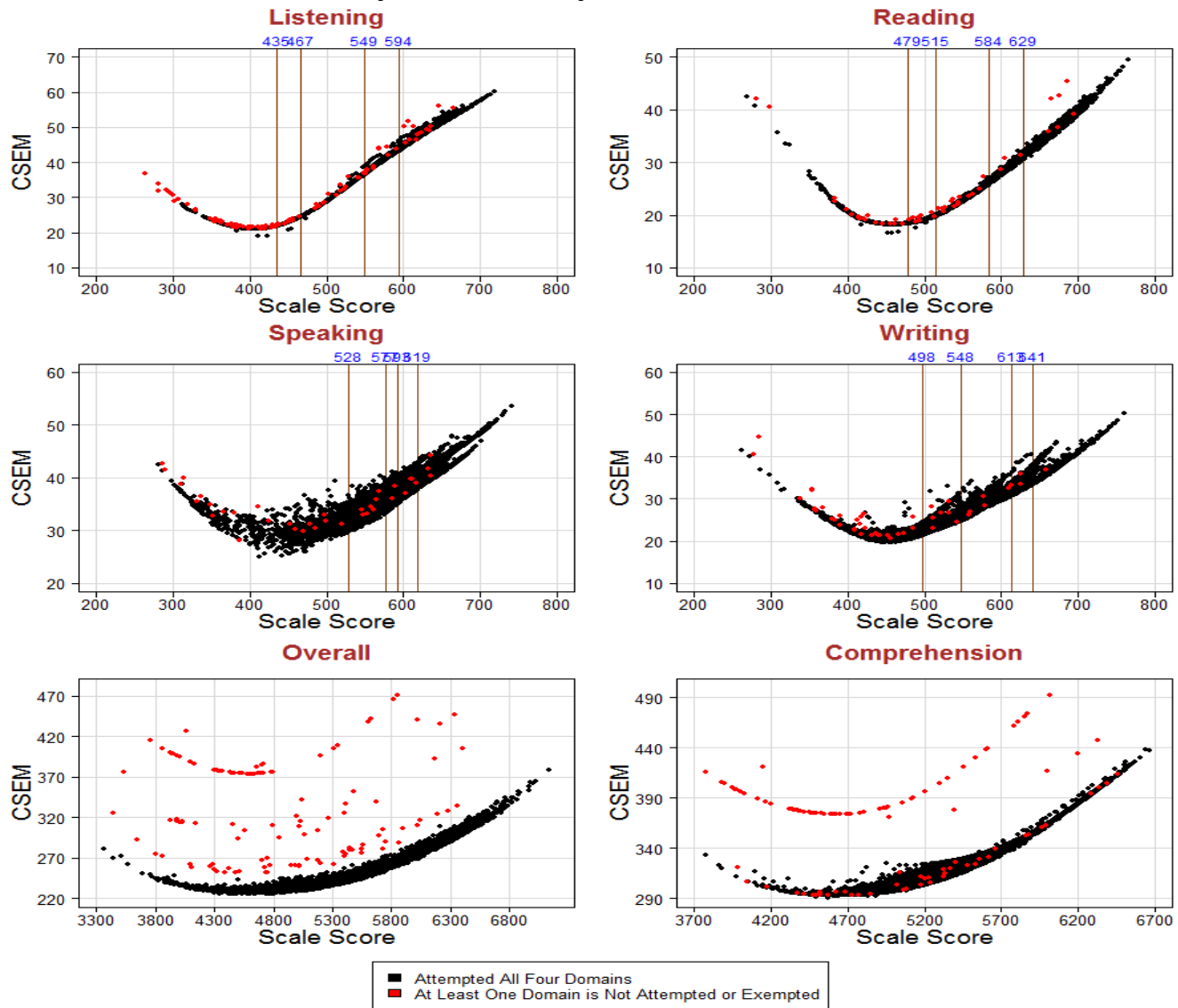


Figure S6.3: Conditional Standard Error of Measurement for Grade 2

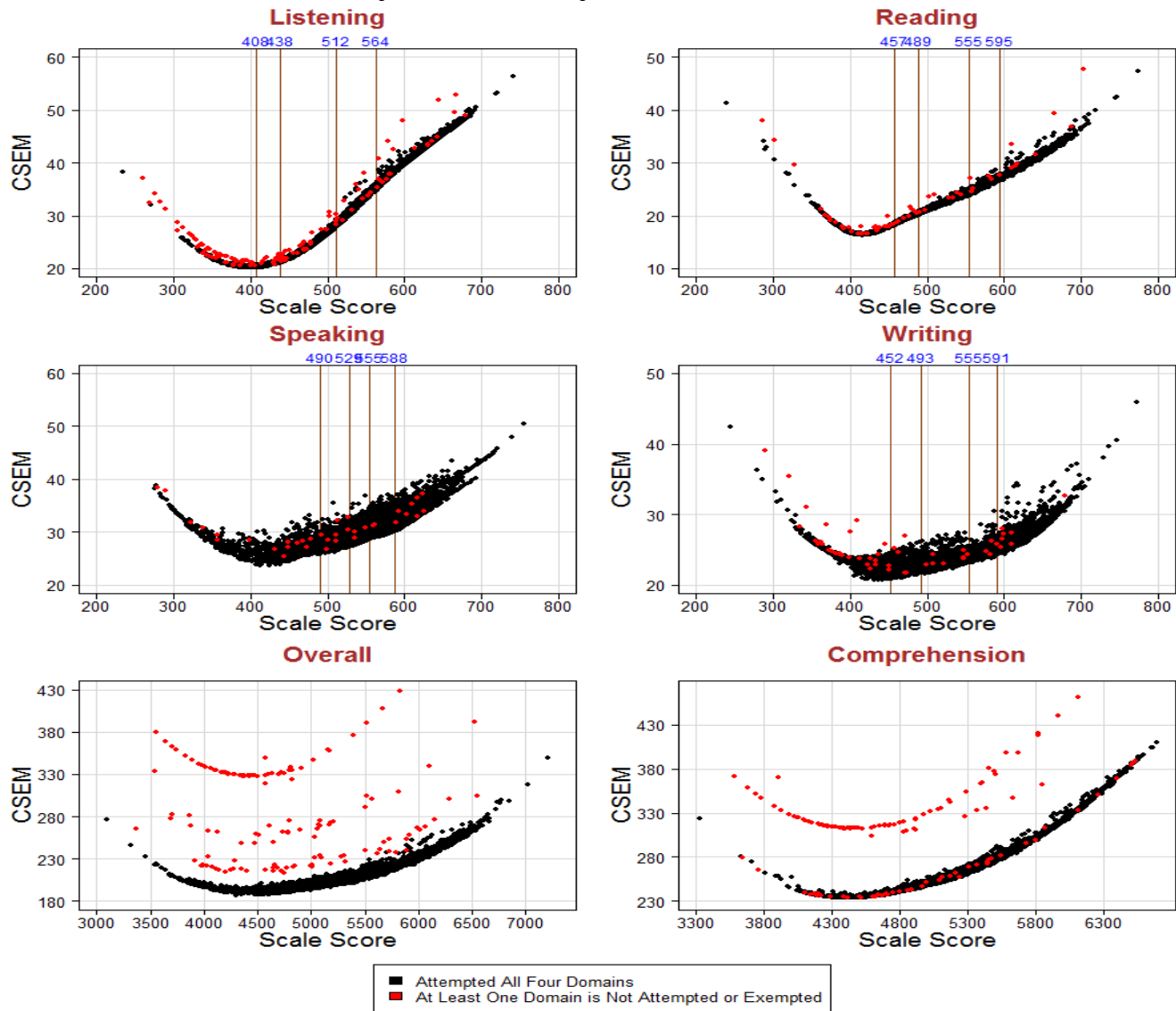


Figure S6.4: Conditional Standard Error of Measurement for Grade 3

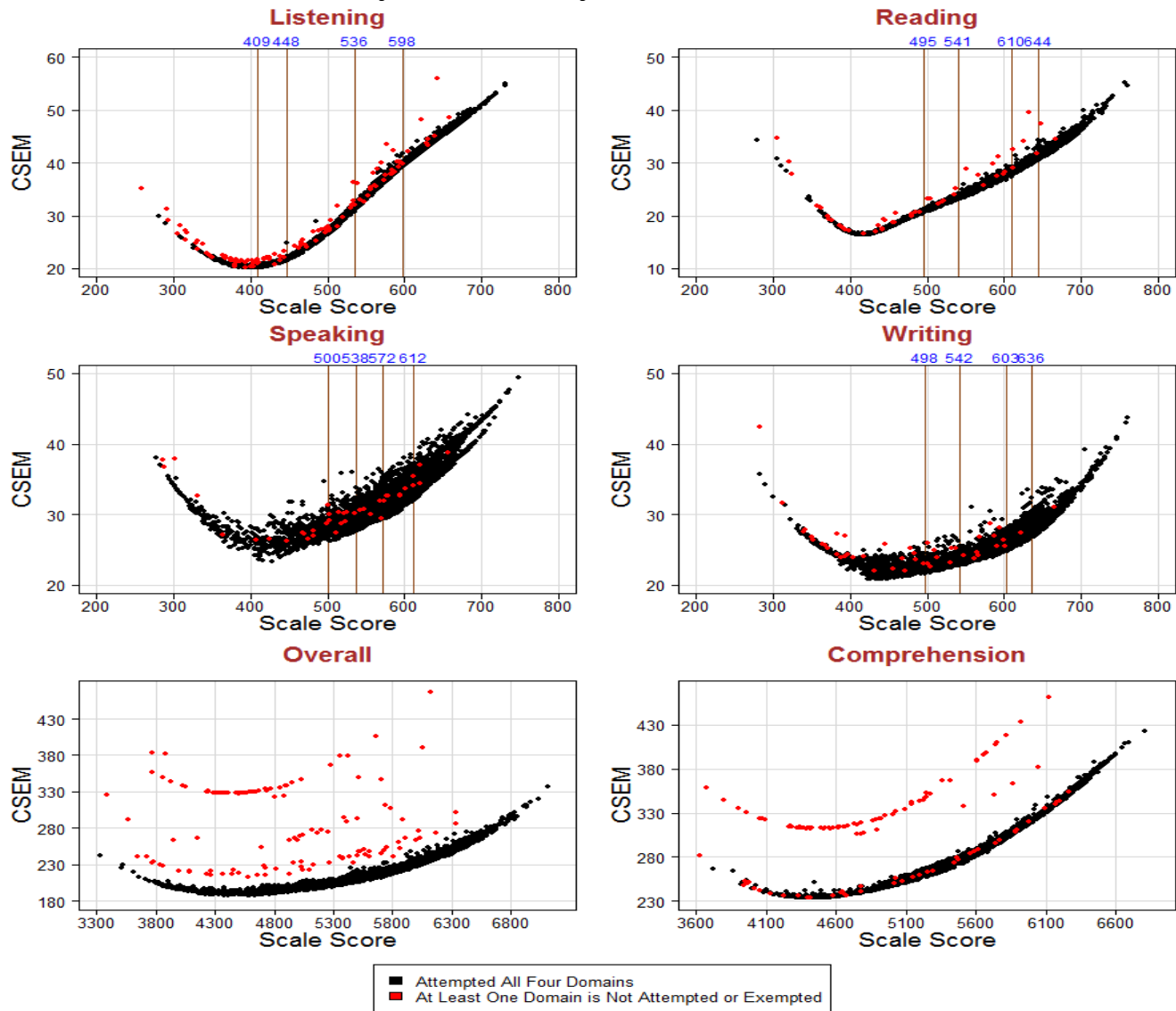


Figure S6.5: Conditional Standard Error of Measurement for Grade 4

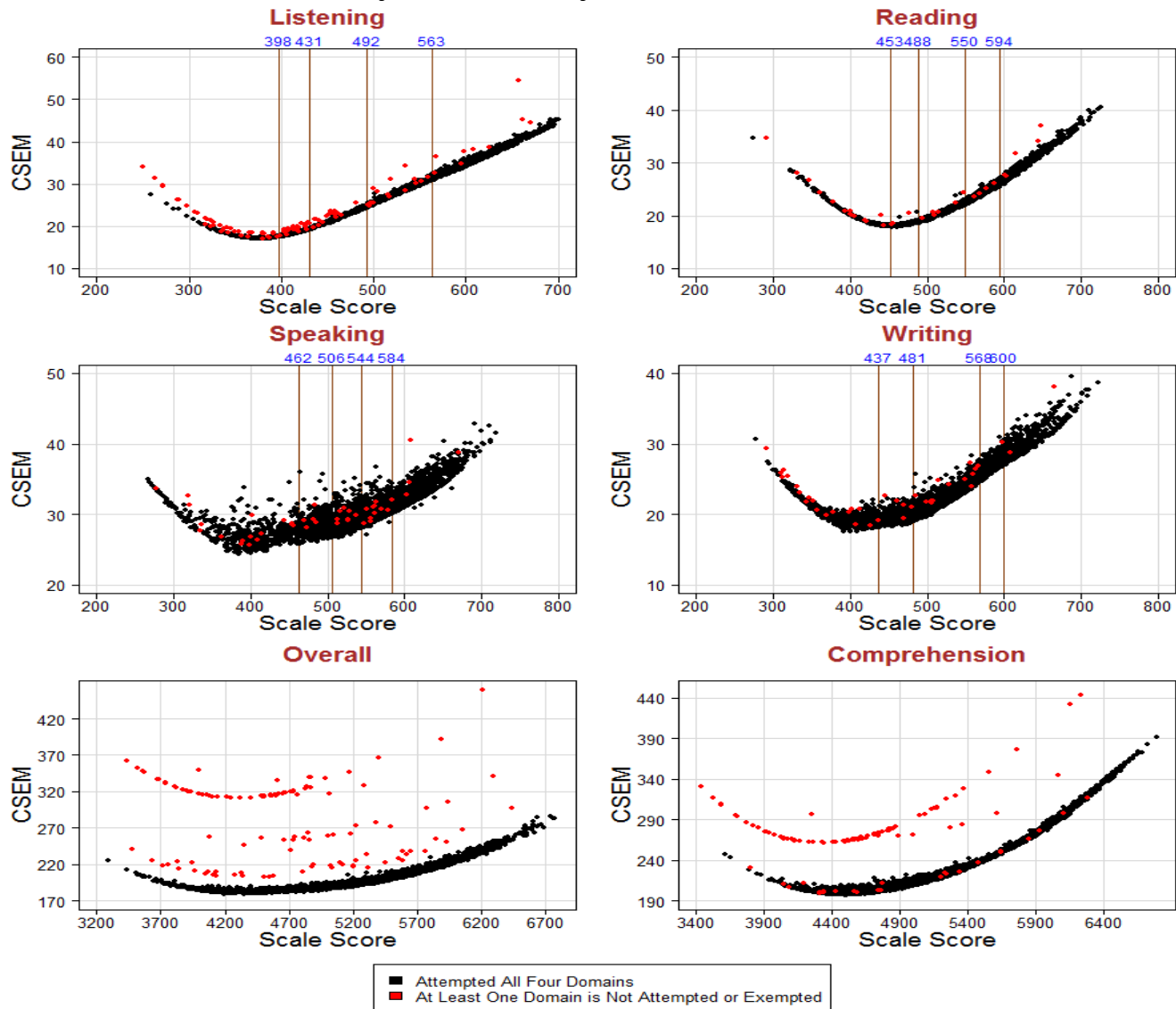


Figure S6.6: Conditional Standard Error of Measurement for Grade 5

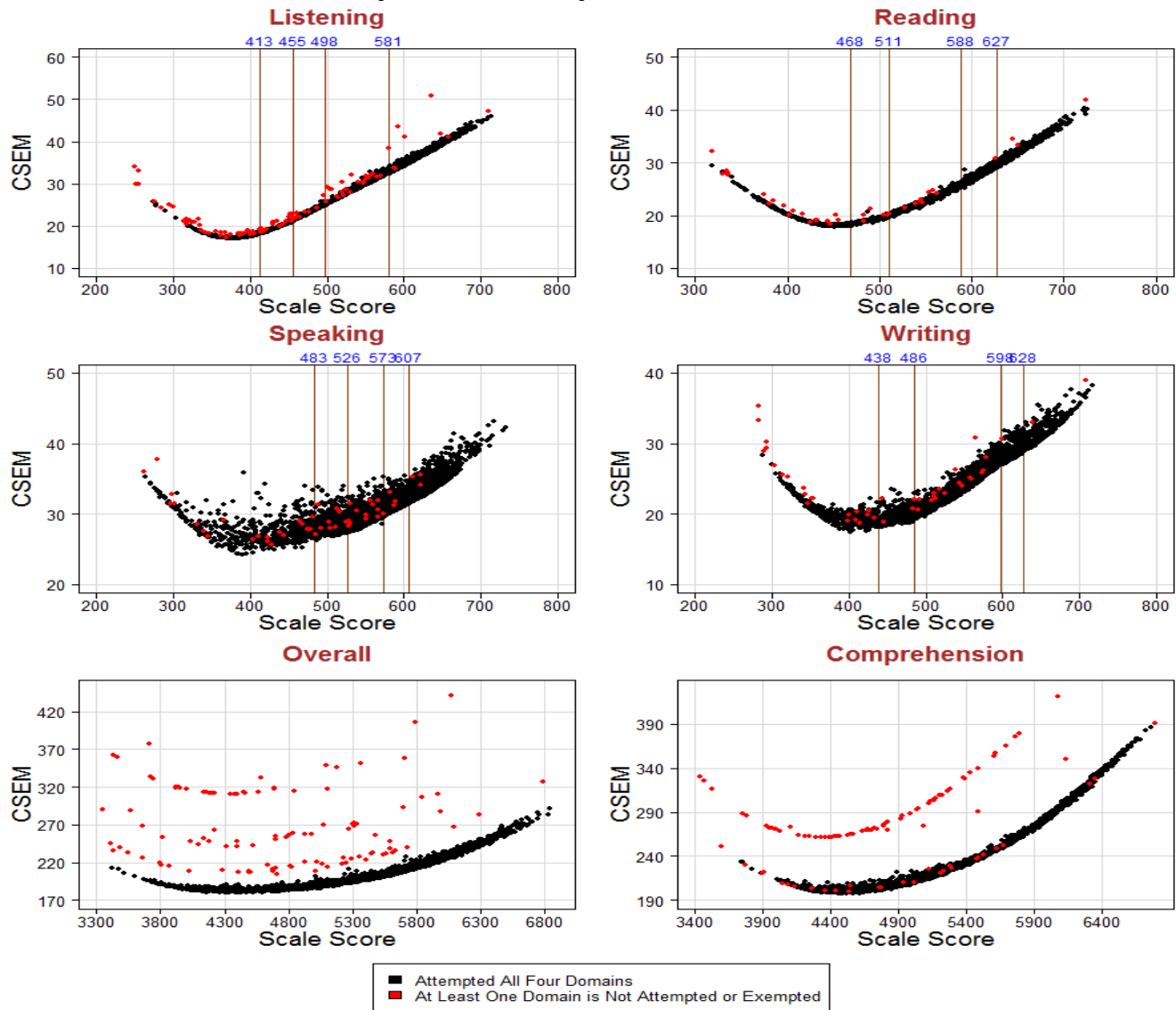


Figure S6.7: Conditional Standard Error of Measurement for Grade 6

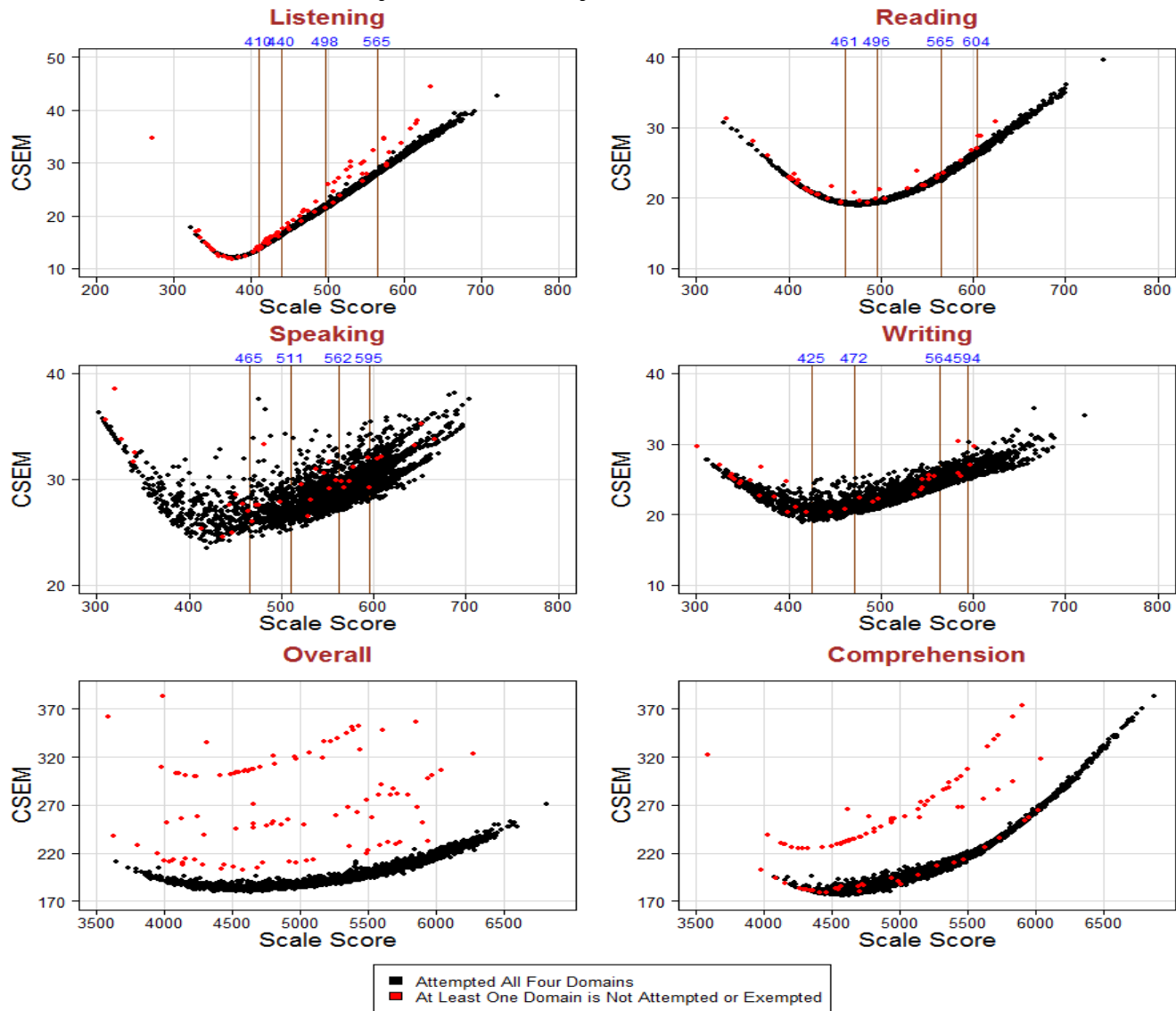


Figure S6.8: Conditional Standard Error of Measurement for Grade 7

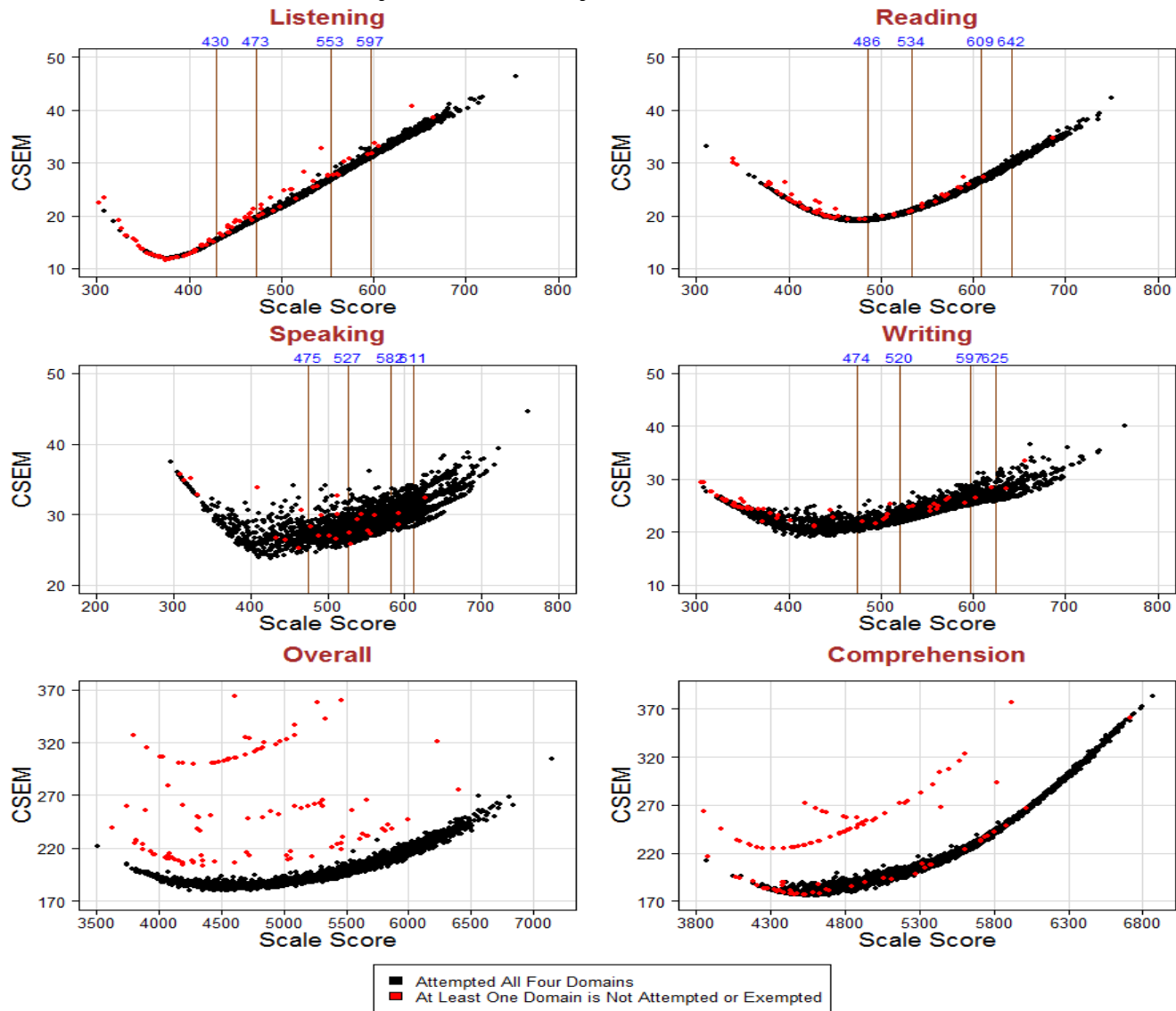


Figure S6.9: Conditional Standard Error of Measurement for Grade 8

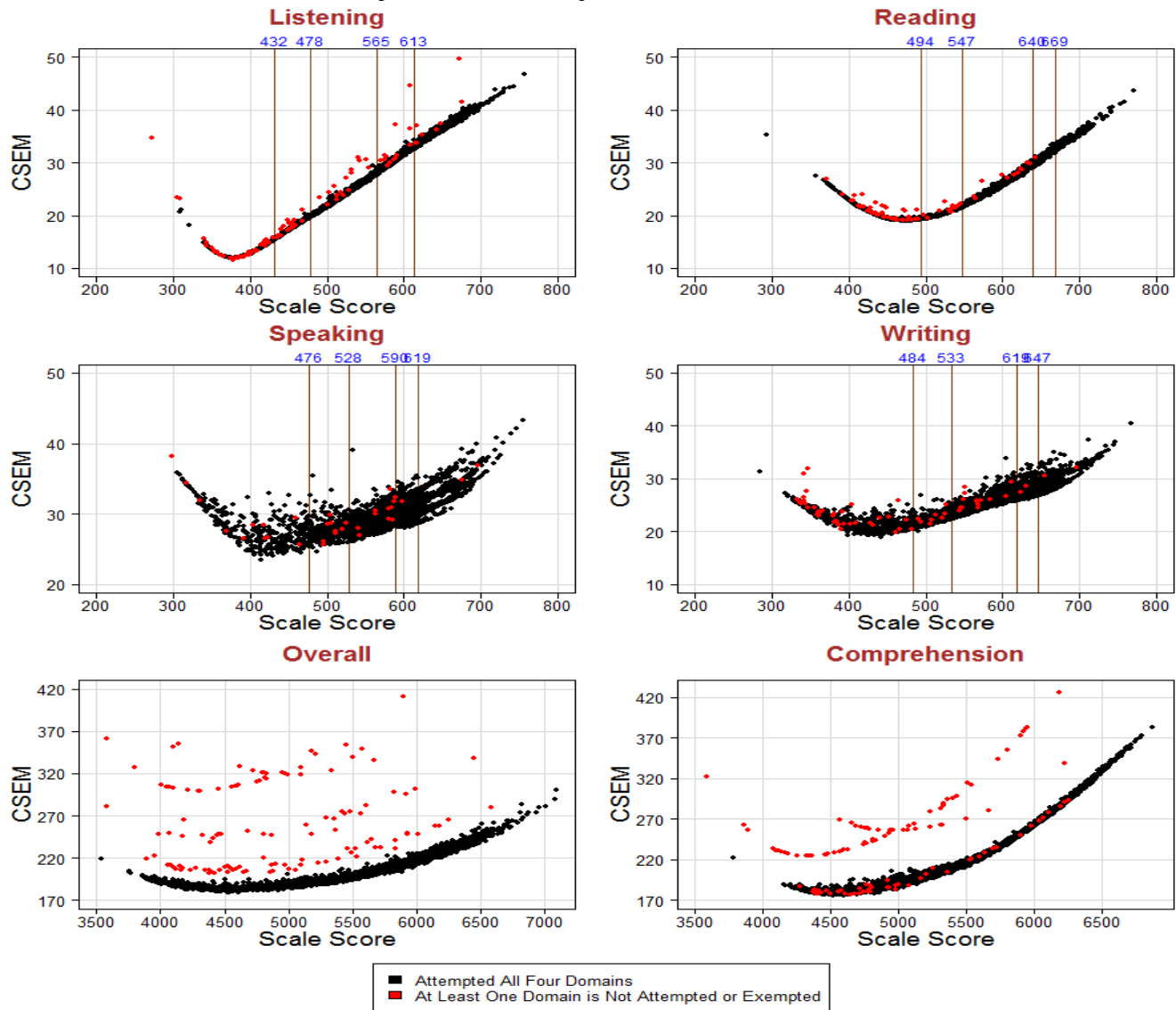


Figure S6.10: Conditional Standard Error of Measurement for Grade 9

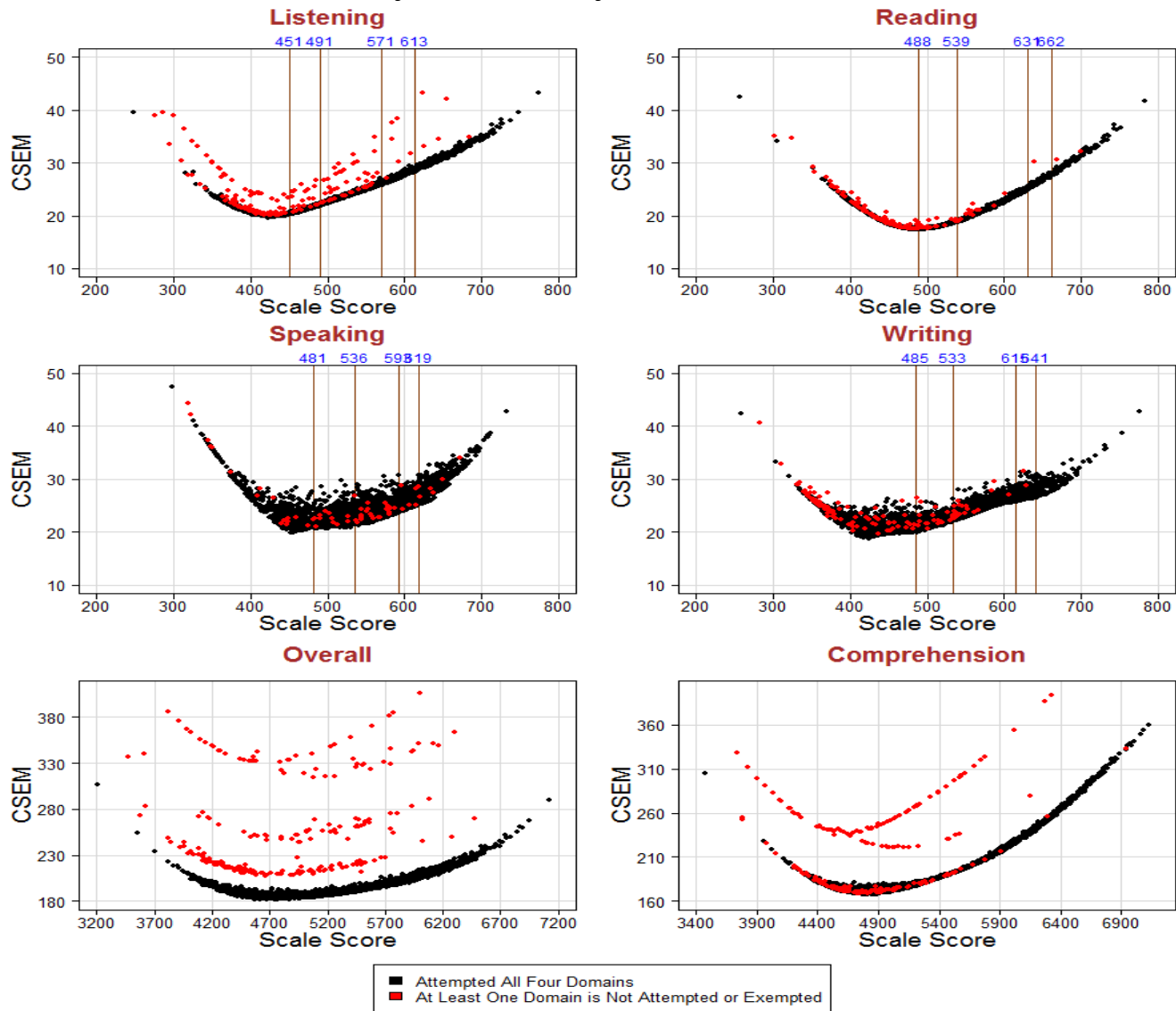


Figure S6.11: Conditional Standard Error of Measurement for Grade 10

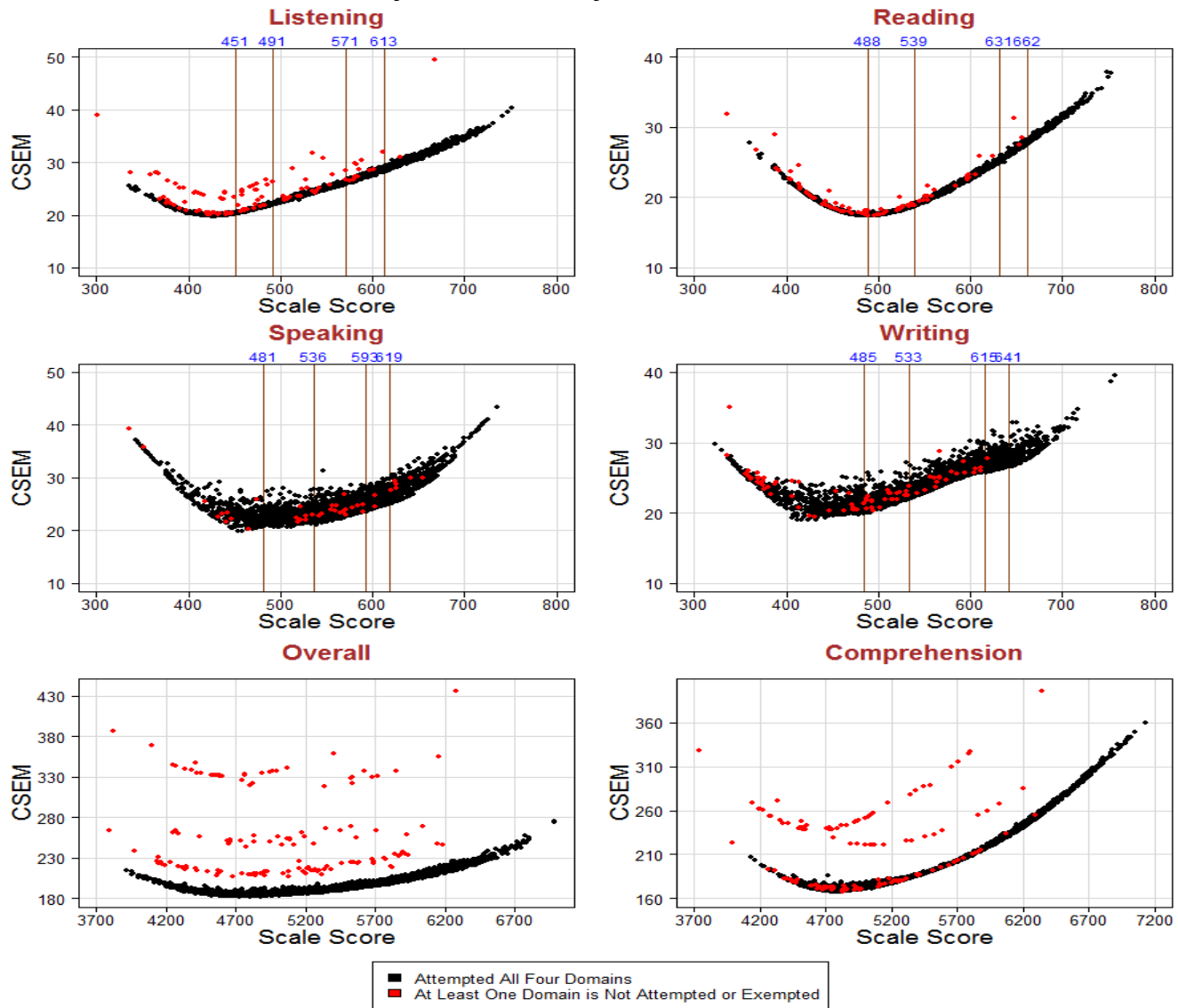


Figure S6.12: Conditional Standard Error of Measurement for Grade 11

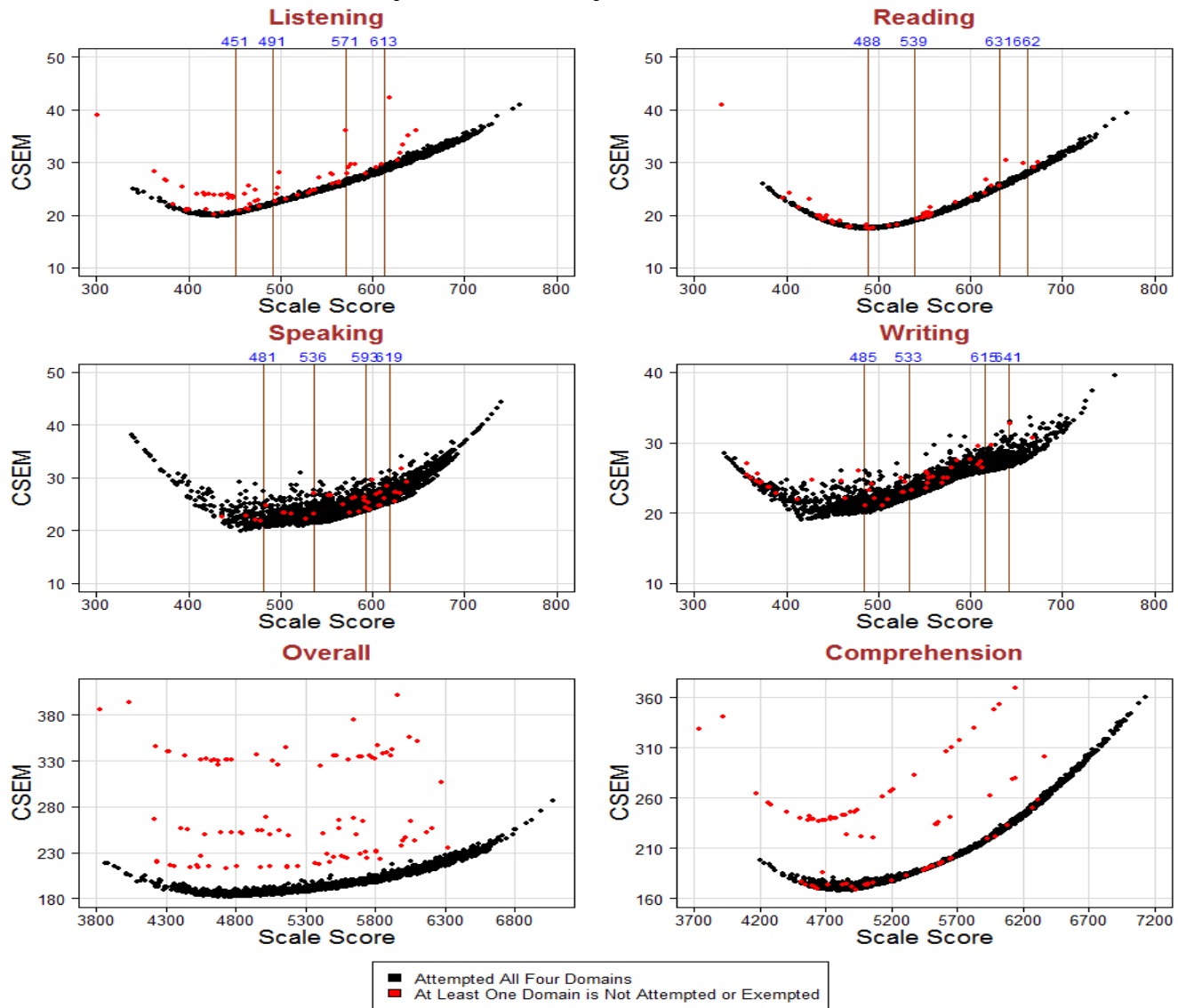
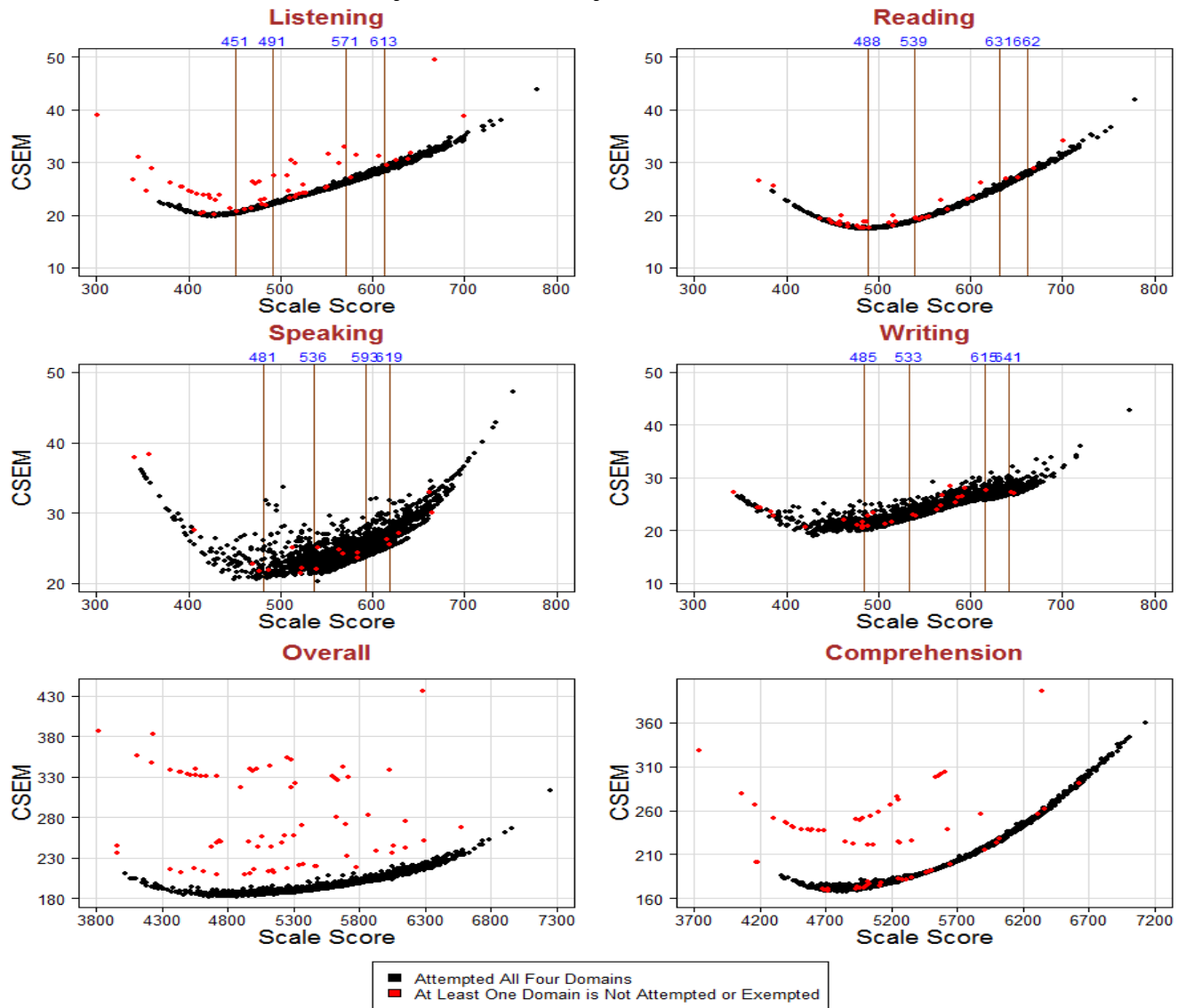


Figure S6.13: Conditional Standard Error of Measurement for Grade 12



Section 7: Summative Assessment - Dimensionality

Figure S7.1: Validity - Dimensionality for Kindergarten

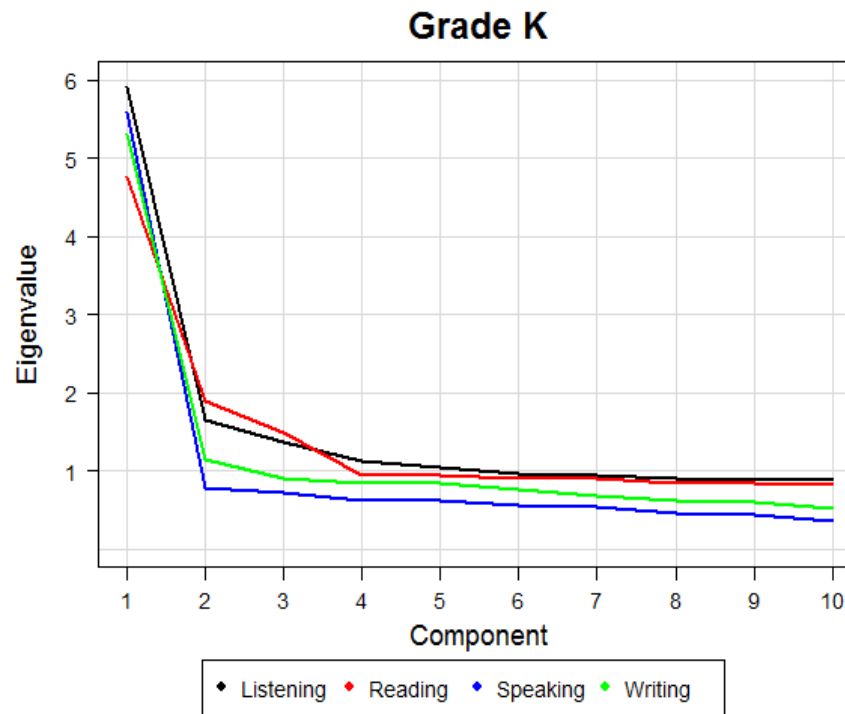


Figure S7.2: Validity - Dimensionality for Grade 1

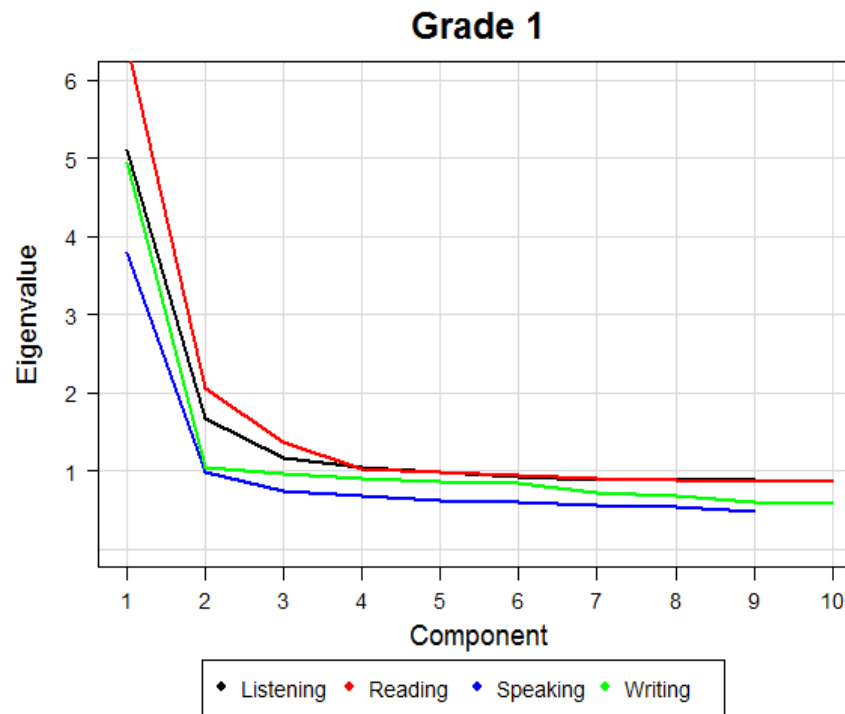


Figure S7.3: Validity - Dimensionality for Grade Band 2-3

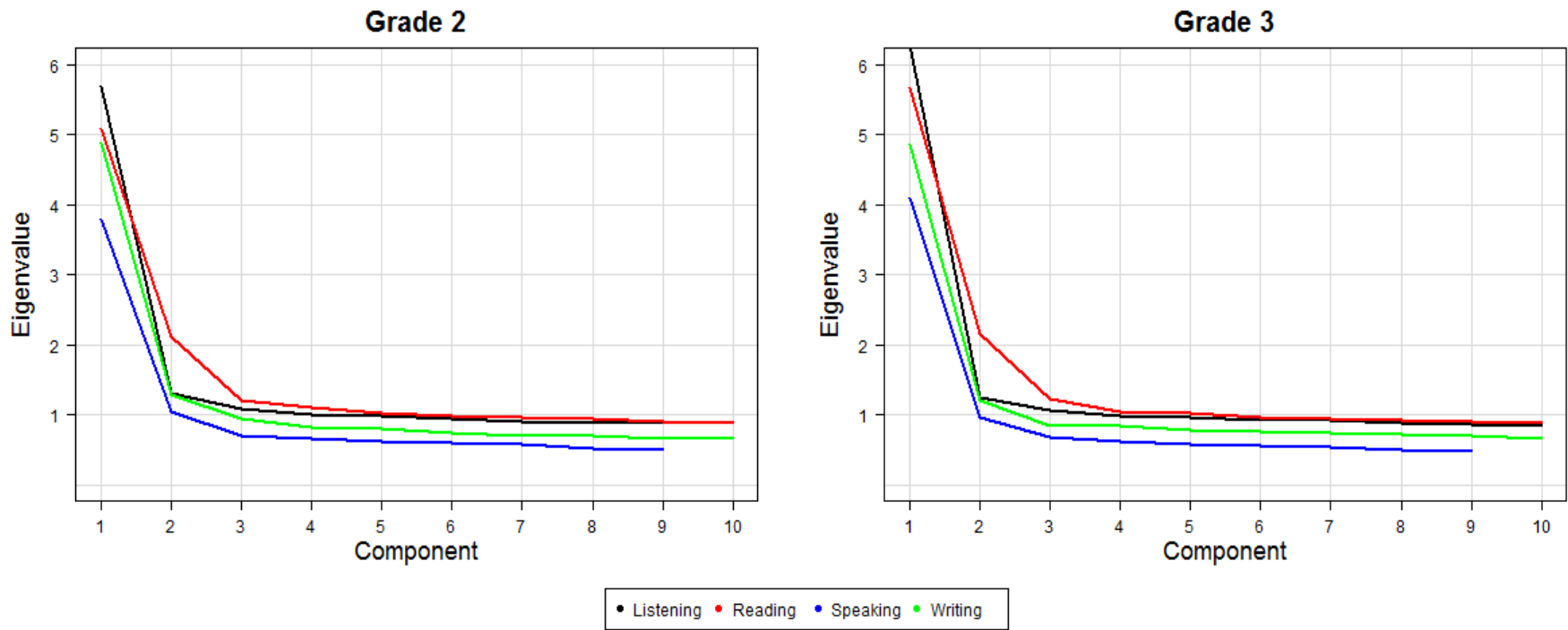


Figure S7.4: Validity - Dimensionality for Grade Band 4-5

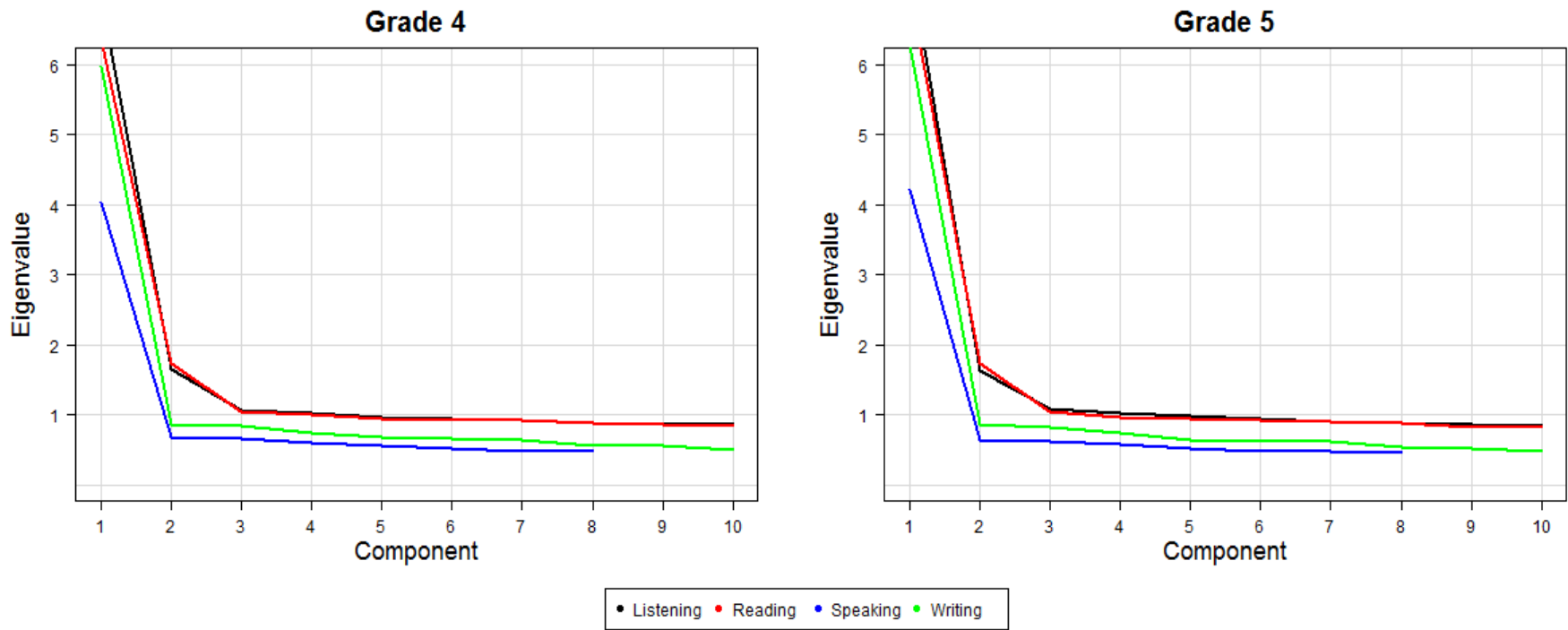


Figure S7.5: Validity - Dimensionality for Grade Band 6-8

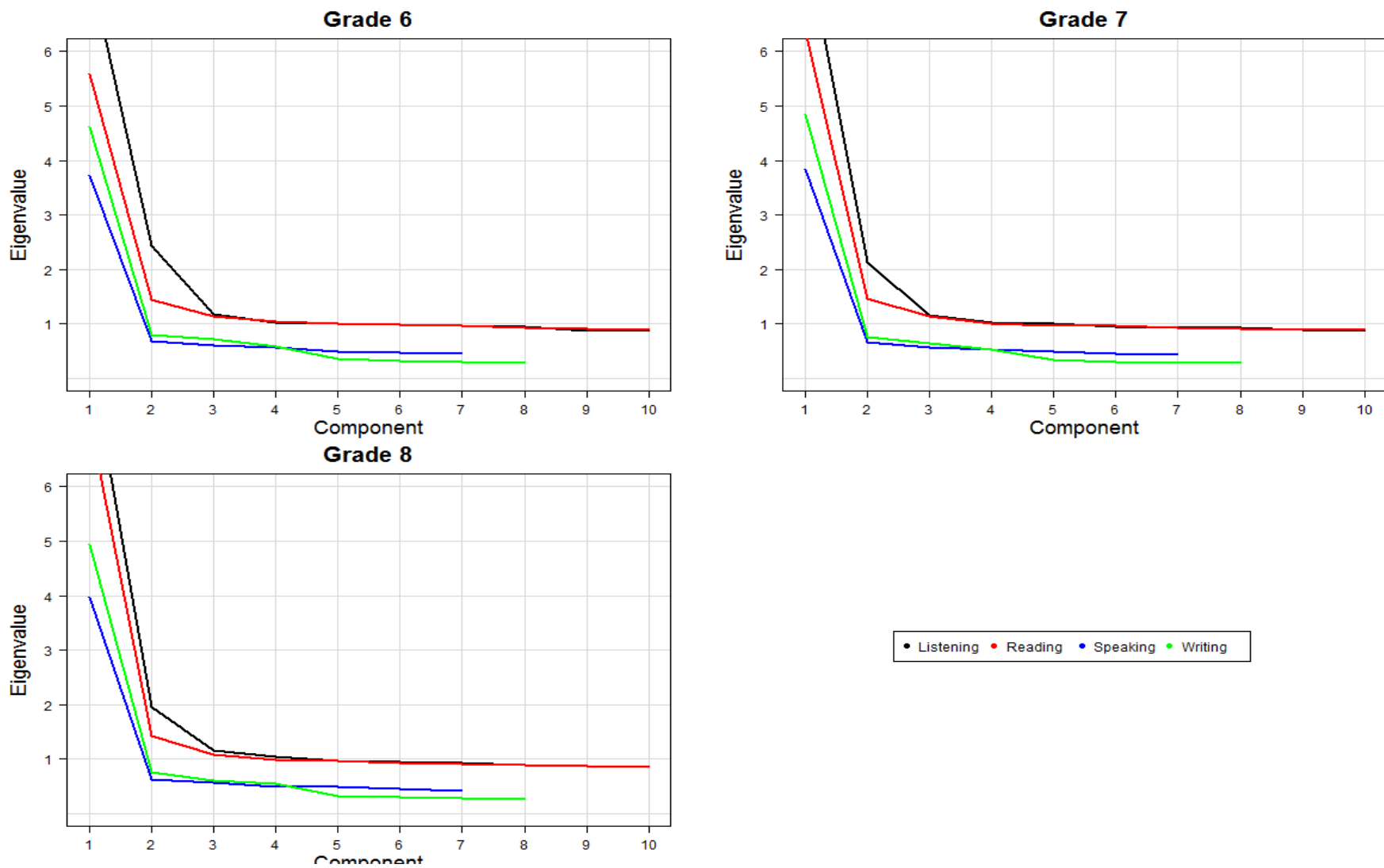
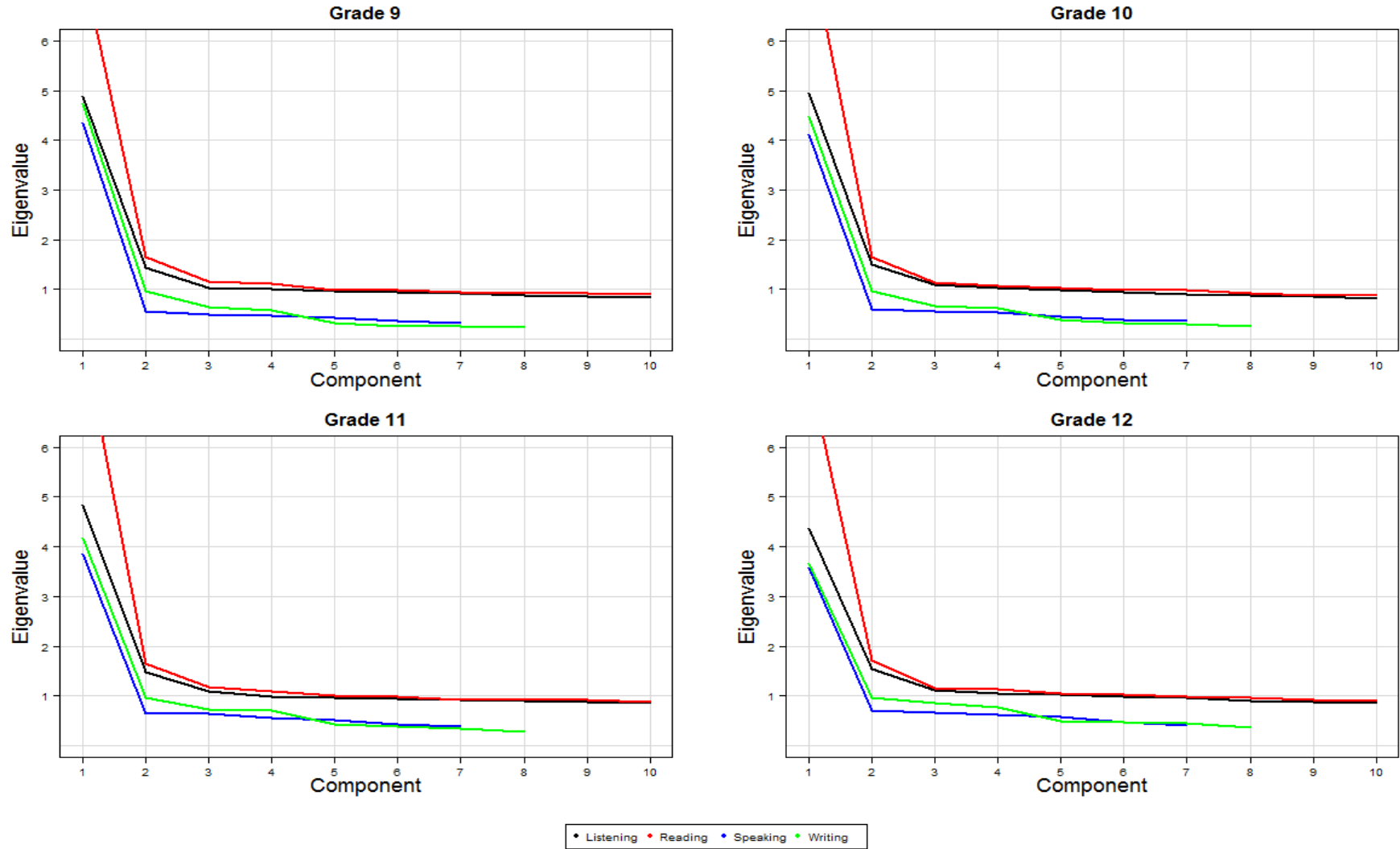
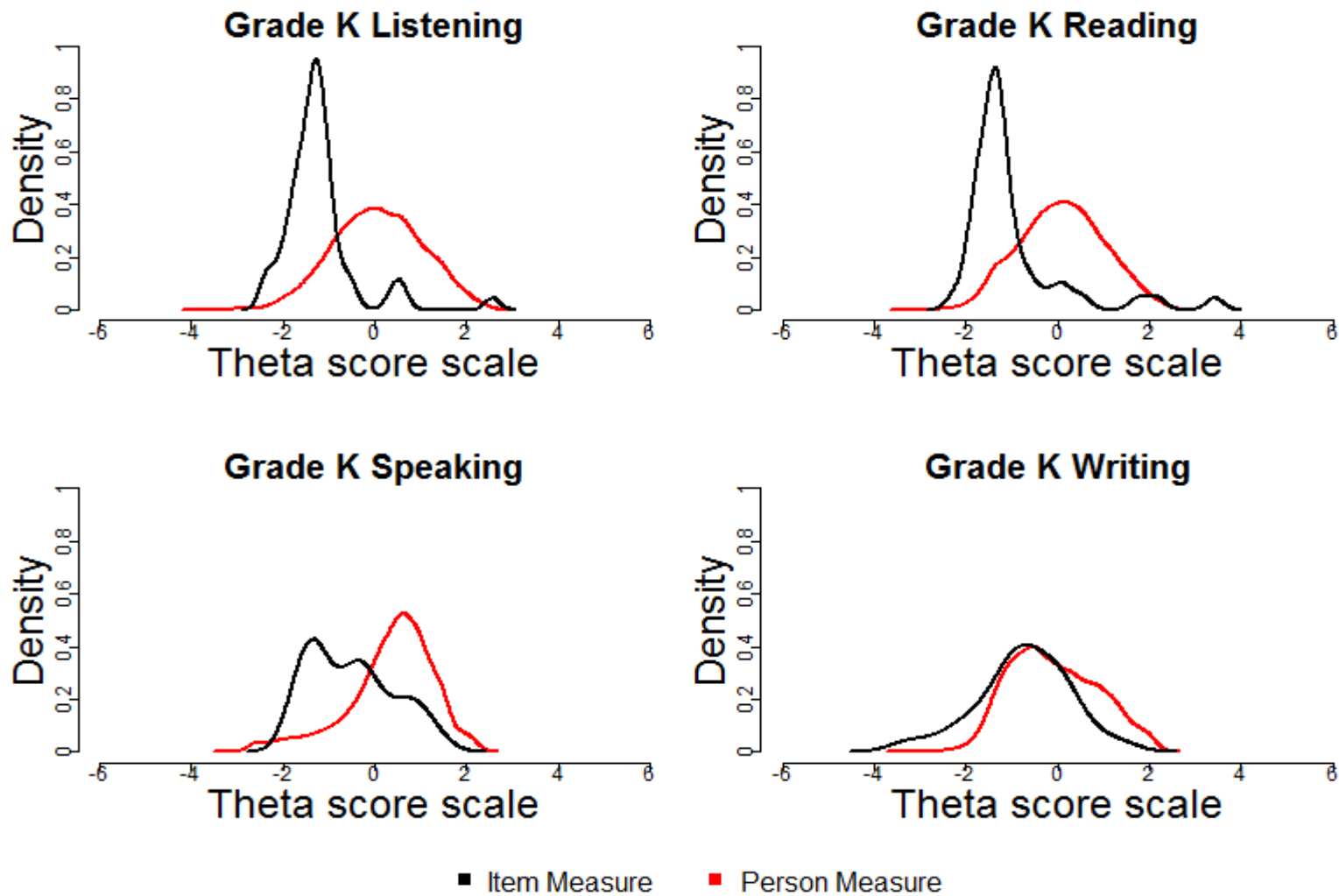


Figure S7.6: Validity - Dimensionality for Grade Band 9-12



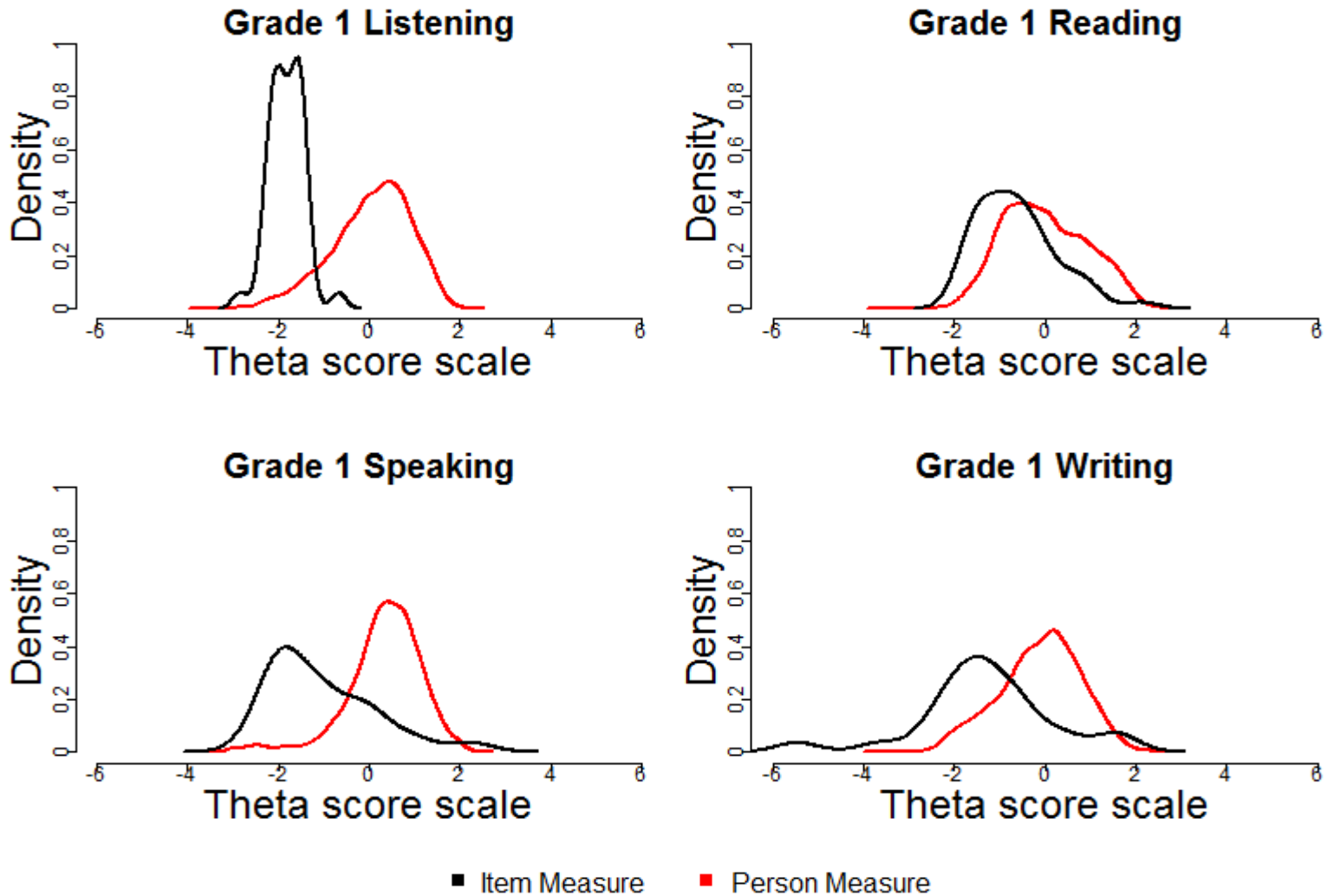
Section 8: Summative Assessment - Ability vs. Difficulty

Figure S8.1: Validity - Ability vs. Difficulty for Kindergarten



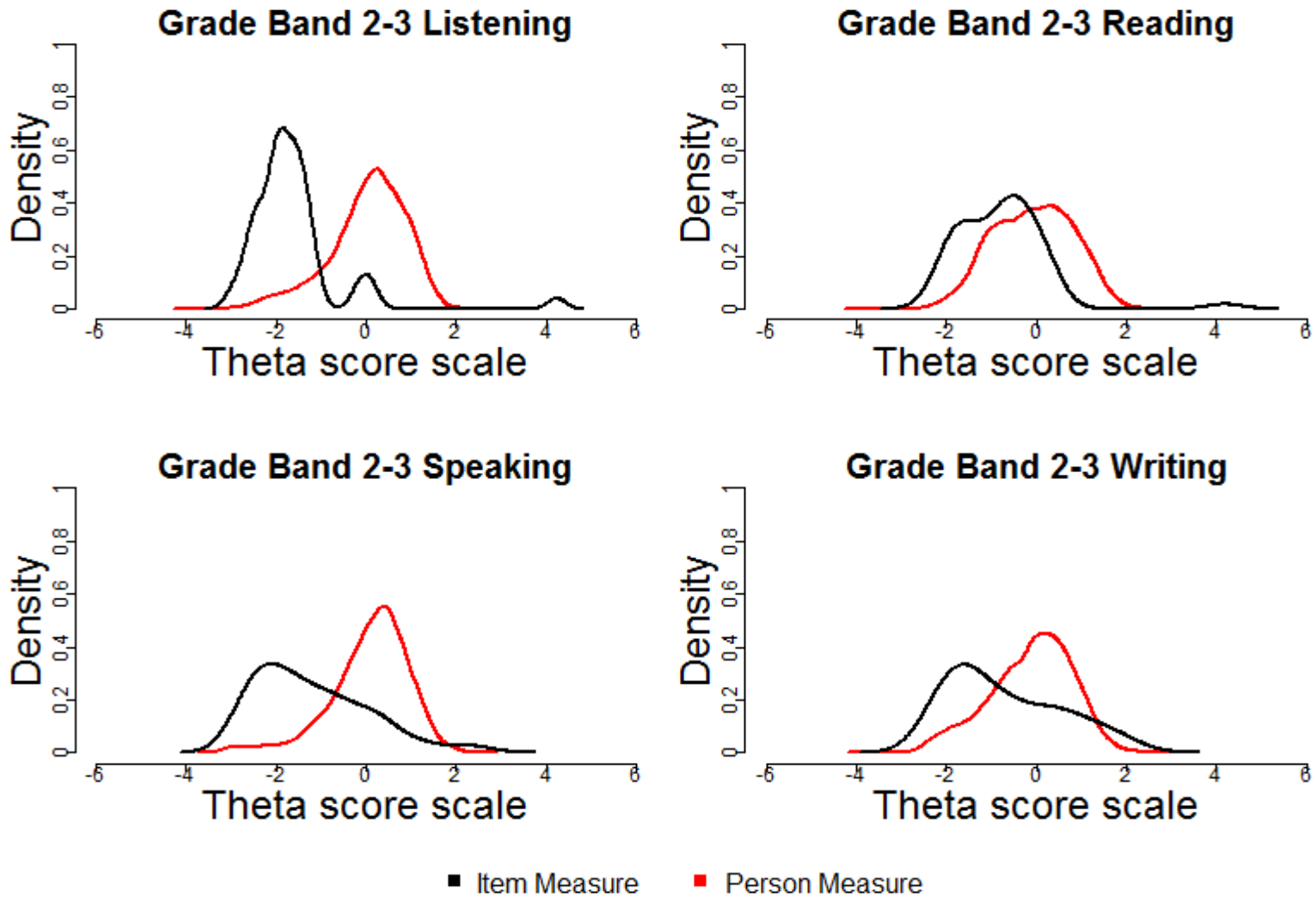
* The probability density is the probability per unit on the x-axis (in this case theta). The peaks of the plot, for example, 0.50 or 0.75, are the locations where there is the highest concentrations of items or persons. The feature of the density plot is that the total area under the curve integrates to one.

Figure S8.2: Validity - Ability vs. Difficulty for Grade 1



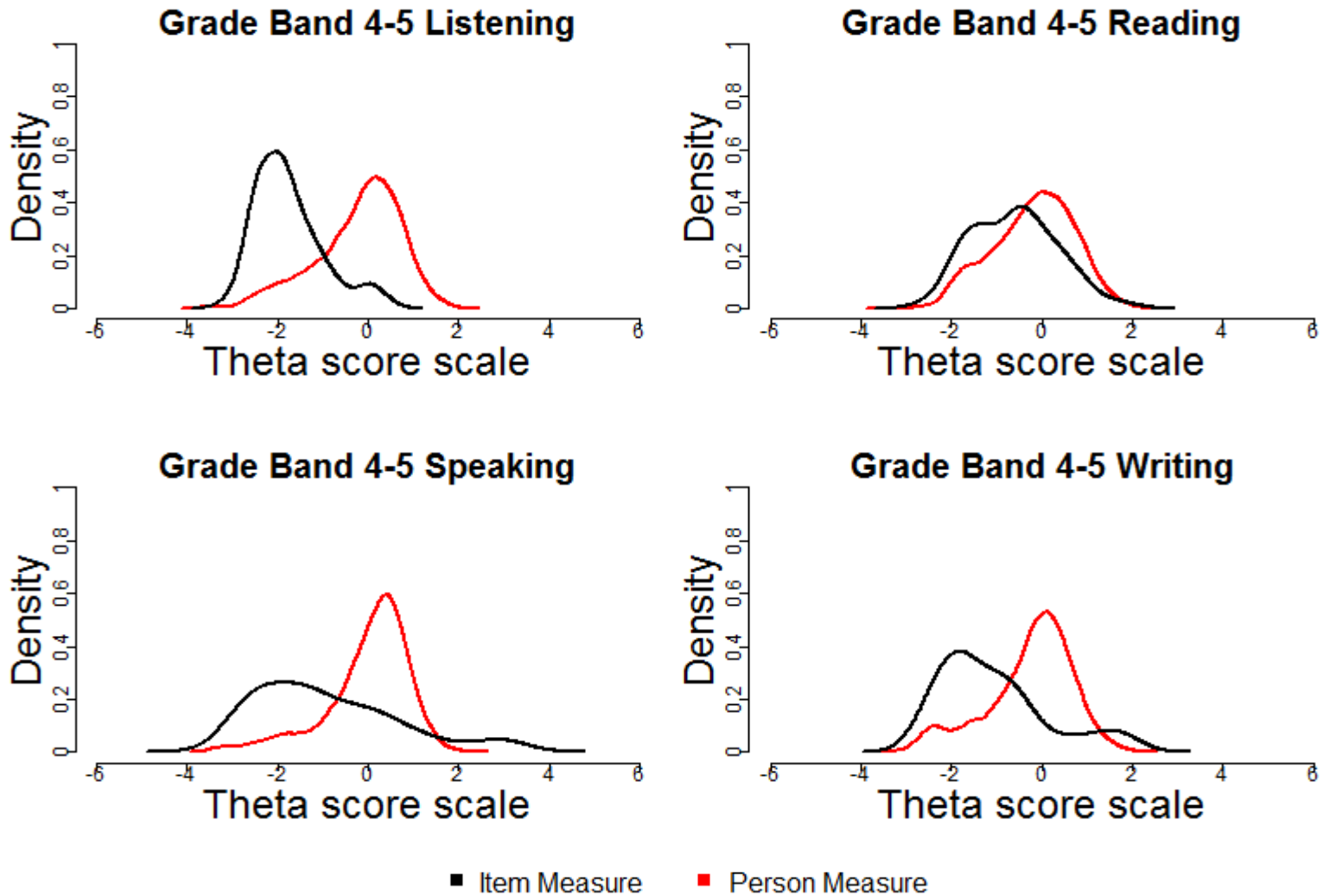
* The probability density is the probability per unit on the x-axis (in this case theta). The peaks of the plot, for example, 0.50 or 0.75, are the locations where there is the highest concentrations of items or persons. The feature of the density plot is that the total area under the curve integrates to one.

Figure S8.3: Validity - Ability vs. Difficulty for Grade Band 2-3



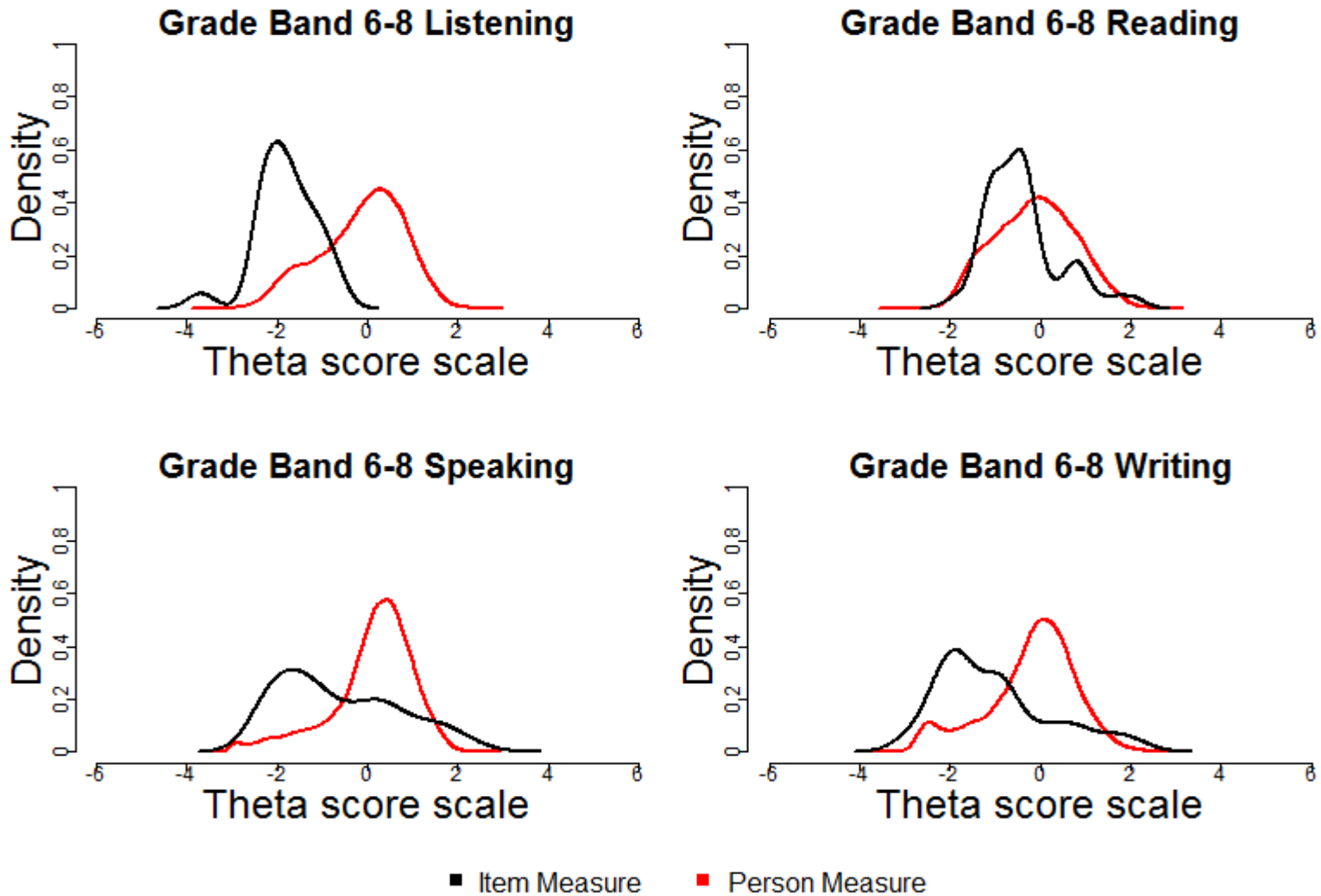
* The probability density is the probability per unit on the x-axis (in this case theta). The peaks of the plot, for example, 0.50 or 0.75, are the locations where there is the highest concentrations of items or persons. The feature of the density plot is that the total area under the curve integrates to one.

Figure S8.4: Validity - Ability vs. Difficulty for Grade Band 4-5



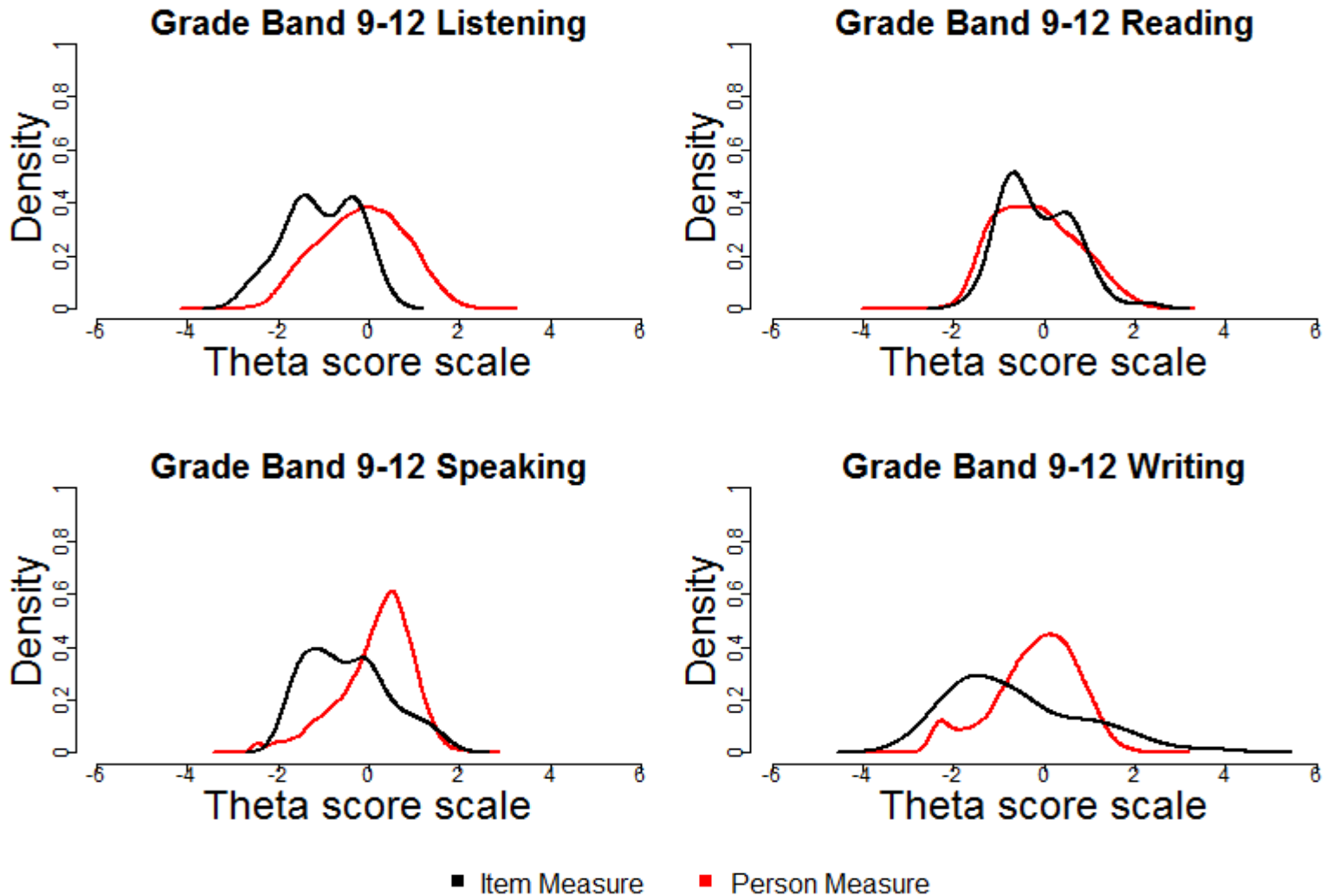
* The probability density is the probability per unit on the x-axis (in this case theta). The peaks of the plot, for example, 0.50 or 0.75, are the locations where there is the highest concentrations of items or persons. The feature of the density plot is that the total area under the curve integrates to one.

Figure S8.5: Validity - Ability vs. Difficulty for Grade Band 6-8



* The probability density is the probability per unit on the x-axis (in this case theta). The peaks of the plot, for example, 0.50 or 0.75, are the locations where there is the highest concentrations of items or persons. The feature of the density plot is that the total area under the curve integrates to one.

Figure S8.6: Validity - Ability vs. Difficulty for Grade Band 9-12



* The probability density is the probability per unit on the x-axis (in this case theta). The peaks of the plot, for example, 0.50 or 0.75, are the locations where there is the highest concentrations of items or persons. The feature of the density plot is that the total area under the curve integrates to one.

Section 9: Summative Assessment - Inter-Rater Analysis

Table S9.1: Inter-Rater Result - Kindergarten

Domain	ItemID	ScorePoints	Total	Weighted Kappa	Count			Percent		
					Exact	Exact +1/-1	More than +2/-2	Exact	Exact +1/-1	More than +2/-2
Speaking	4156	2	1755	0.8446	1548	1729	26	88.2	98.5	1.5
	4158	2	1759	0.7900	1539	1745	14	87.5	99.2	0.8
	4160	2	1761	0.7618	1626	1732	29	92.3	98.4	1.6
	4162	2	1754	0.7329	1464	1743	11	83.5	99.4	0.6
	4164	2	1758	0.7800	1521	1748	10	86.5	99.4	0.6
	4166	2	1739	0.7281	1378	1713	26	79.2	98.5	1.5
	4252	3	1760	0.8016	1244	1744	16	70.7	99.1	0.9
	4260	3	1769	0.7657	1173	1750	19	66.3	98.9	1.1
	4338	3	1777	0.8053	1233	1755	22	69.4	98.8	1.2
	4370	3	1775	0.7899	1224	1754	21	69.0	98.8	1.2
4382	3	1757	0.8349	1366	1738	19	77.7	98.9	1.1	
Writing	4877	2	1776	0.8589	1705	1774	2	96.0	99.9	0.1
	4878	2	1763	0.9223	1713	1758	5	97.2	99.7	0.3
	4879	2	1763	0.9173	1621	1749	14	91.9	99.2	0.8
	4880	3	1735	0.8853	1278	1712	23	73.7	98.7	1.3
	4881	3	1755	0.8362	1293	1721	34	73.7	98.1	1.9

Table S9.2: Inter-Rater Result - Grade 1

Domain	ItemID	ScorePoints	Total	Weighted Kappa	Count			Percent		
					Exact	Exact +1/-1	More than +2/-2	Exact	Exact +1/-1	More than +2/-2
Speaking	578	2	1562	0.7995	1490	1551	11	95.4	99.3	0.7
	580	2	1559	0.6135	1243	1545	14	79.7	99.1	0.9
	582	2	1560	0.7260	1482	1549	11	95.0	99.3	0.7
	584	2	1555	0.6755	1302	1549	6	83.7	99.6	0.4
	586	2	1555	0.7889	1480	1545	10	95.2	99.4	0.6
	628	3	1562	0.6277	1048	1532	30	67.1	98.1	1.9
	664	5	1543	0.8430	871	1484	59	56.4	96.2	3.8
	678	4	1561	0.7324	1199	1520	41	76.8	97.4	2.6
	706	3	1556	0.8750	1312	1545	11	84.3	99.3	0.7
Writing	4894	2	1554	0.7208	1509	1553	1	97.1	99.9	0.1
	4895	2	1548	0.8999	1488	1546	2	96.1	99.9	0.1
	4896	3	1548	0.8834	1240	1523	25	80.1	98.4	1.6
	4897	4	1545	0.8405	1101	1532	13	71.3	99.2	0.8

Table S9.3: Inter-Rater Result - Grade Band 2-3

Domain	ItemID	ScorePoints	Total	Weighted Kappa	Count			Percent		
					Exact	Exact +1/-1	More than +2/-2	Exact	Exact +1/-1	More than +2/-2
Speaking	1338	2	2451	0.7346	2083	2440	11	85.0	99.6	0.4
	1340	2	2459	0.6350	2164	2451	8	88.0	99.7	0.3
	1342	2	2449	0.8026	2231	2435	14	91.1	99.4	0.6
	1344	2	2455	0.6826	2194	2447	8	89.4	99.7	0.3
	1346	2	2450	0.7253	2243	2440	10	91.6	99.6	0.4
	1396	3	2443	0.7896	1748	2422	21	71.6	99.1	0.9
	1400	3	2461	0.8026	1863	2447	14	75.7	99.4	0.6
	1436	5	2437	0.8425	1527	2377	60	62.7	97.5	2.5
	1454	4	2447	0.8929	1820	2393	54	74.4	97.8	2.2
Writing	1462	4	2435	0.8519	1722	2417	18	70.7	99.3	0.7
	1468	3	2461	0.9103	2052	2449	12	83.4	99.5	0.5
	1492	3	2462	0.9006	1969	2451	11	80.0	99.6	0.4
	1582	4	2443	0.8474	1698	2435	8	69.5	99.7	0.3

Table S9.4: Inter-Rater Result - Grade Band 4-5

Domain	ItemID	ScorePoints	Total	Weighted Kappa	Count			Percent		
					Exact	Exact +1/-1	More than +2/-2	Exact	Exact +1/-1	More than +2/-2
Speaking	2090	5	1620	0.8213	1105	1544	76	68.2	95.3	4.7
	2092	5	1620	0.7796	968	1569	51	59.8	96.9	3.1
	2112	3	1618	0.7803	1195	1606	12	73.9	99.3	0.7
	2118	3	1636	0.5984	1055	1623	13	64.5	99.2	0.8
	2172	3	1622	0.8693	1440	1613	9	88.8	99.4	0.6
	2174	3	1619	0.7901	1215	1582	37	75.0	97.7	2.3
	2176	3	1615	0.6931	1103	1587	28	68.3	98.3	1.7
	2204	5	1614	0.7955	915	1581	33	56.7	98.0	2.0
Writing	2312	5	1616	0.8702	1052	1609	7	65.1	99.6	0.4
	2340	4	1615	0.8060	1006	1594	21	62.3	98.7	1.3
	2374	3	1614	0.8819	1302	1607	7	80.7	99.6	0.4
	2376	3	1612	0.8564	1249	1602	10	77.5	99.4	0.6
	2378	3	1613	0.8820	1288	1604	9	79.9	99.4	0.6

Table S9.5: Inter-Rater Result - Grade Band 6-8

Domain	ItemID	ScorePoints	Total	Weighted Kappa	Count			Percent		
					Exact	Exact +1/-1	More than +2/-2	Exact	Exact +1/-1	More than +2/-2
Speaking	2650	5	1969	0.9063	1524	1922	47	77.4	97.6	2.4
	2652	5	1961	0.8059	1327	1912	49	67.7	97.5	2.5
	2670	3	1963	0.7497	1418	1958	5	72.2	99.7	0.3
	2700	3	1967	0.9508	1863	1954	13	94.7	99.3	0.7
	2702	3	1955	0.7513	1448	1907	48	74.1	97.5	2.5
	2704	3	1957	0.7960	1514	1928	29	77.4	98.5	1.5
	2708	5	1961	0.8639	1271	1908	53	64.8	97.3	2.7
Writing	2754	5	1984	0.8757	1299	1958	26	65.5	98.7	1.3
	2784	3	1981	0.8342	1371	1974	7	69.2	99.6	0.4
	2798	5	1986	0.8680	1293	1963	23	65.1	98.8	1.2
	2818	3	1989	0.8475	1570	1978	11	78.9	99.4	0.6
	2820	3	1987	0.8449	1578	1973	14	79.4	99.3	0.7
	2822	3	1985	0.8610	1580	1969	16	79.6	99.2	0.8

Table S9.6: Inter-Rater Result - Grade Band 9-12

Domain	ItemID	ScorePoints	Total	Weighted Kappa	Count			Percent		
					Exact	Exact +1/-1	More than +2/-2	Exact	Exact +1/-1	More than +2/-2
Speaking	3360	5	2374	0.8115	1302	2236	138	54.8	94.2	5.8
	3362	5	2363	0.8468	1367	2286	77	57.9	96.7	3.3
	3370	3	2414	0.7604	1774	2393	21	73.5	99.1	0.9
	3388	3	2416	0.7988	1922	2393	23	79.6	99.0	1.0
	3390	3	2407	0.8552	1937	2377	30	80.5	98.8	1.2
	3392	3	2386	0.7235	1679	2341	45	70.4	98.1	1.9
	3410	5	2380	0.8284	1290	2279	101	54.2	95.8	4.2
Writing	3458	5	2433	0.8740	1530	2416	17	62.9	99.3	0.7
	3488	3	2433	0.7937	1642	2427	6	67.5	99.8	0.2
	3496	5	2435	0.8831	1575	2425	10	64.7	99.6	0.4
	3522	3	2453	0.8499	1915	2448	5	78.1	99.8	0.2
	3524	3	2447	0.8342	1887	2432	15	77.1	99.4	0.6
	3526	3	2451	0.8389	1903	2439	12	77.6	99.5	0.5

Section 10: Student Progress from Screener to Summative - Scatter Plots of Scale Score Change

Figure S10.1: Scatter Plot of Scale Score Change from Screener to Summative for Grade K

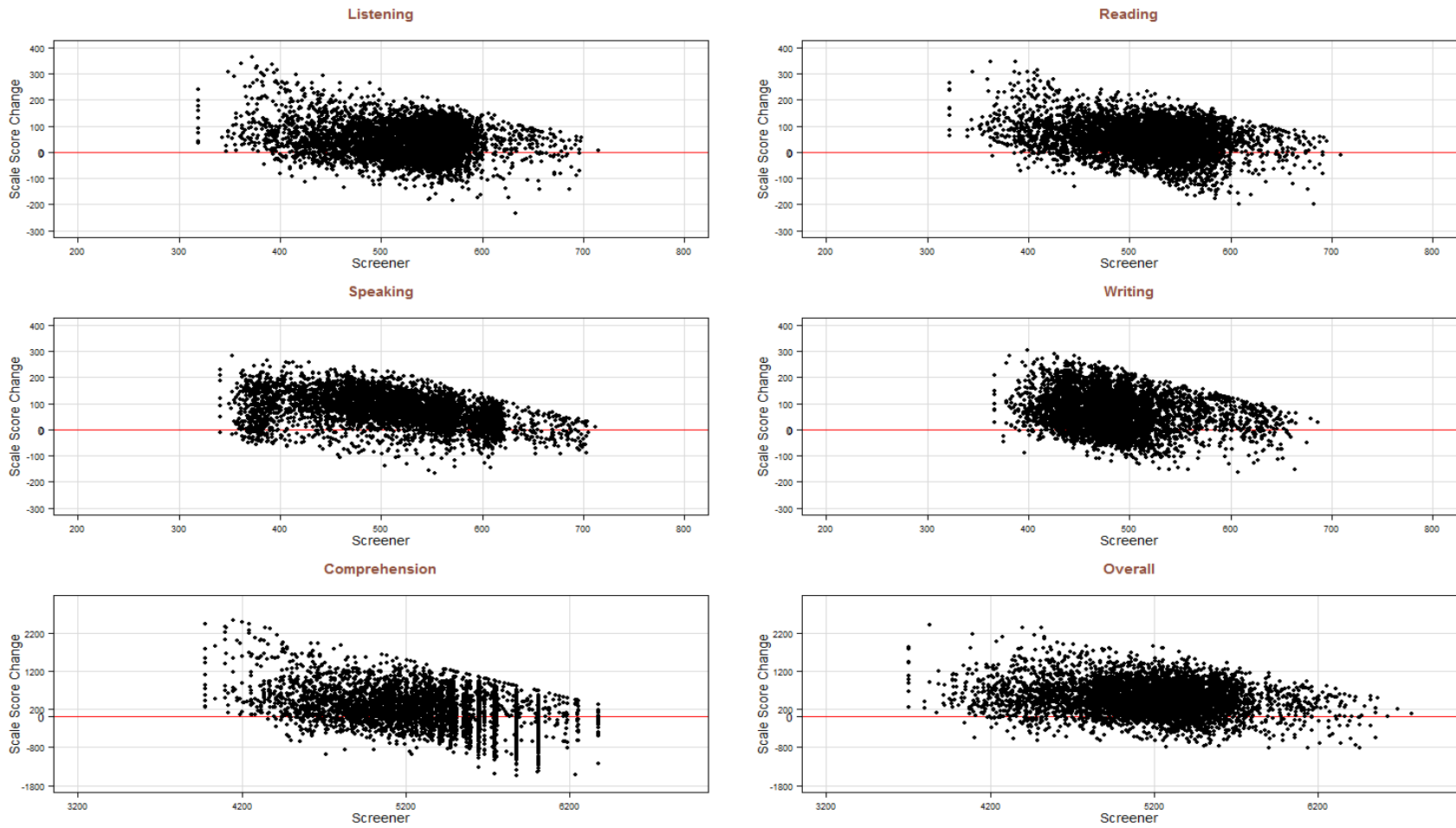


Figure S10.2: Scatter Plot of Scale Score Change from Screener to Summative for Grade 1

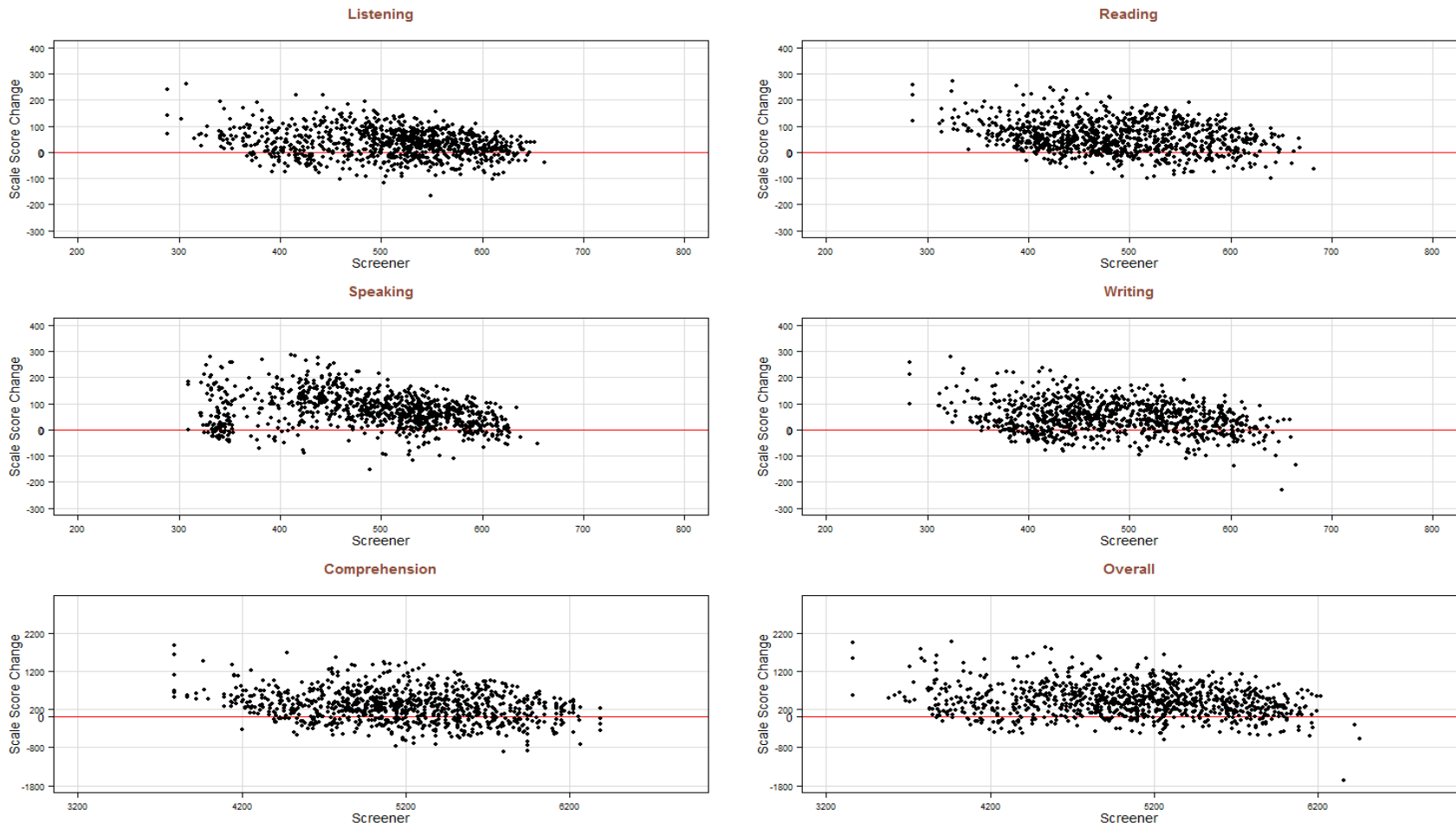


Figure S10.3: Scatter Plot of Scale Score Change from Screener to Summative for Grade Band 2-3

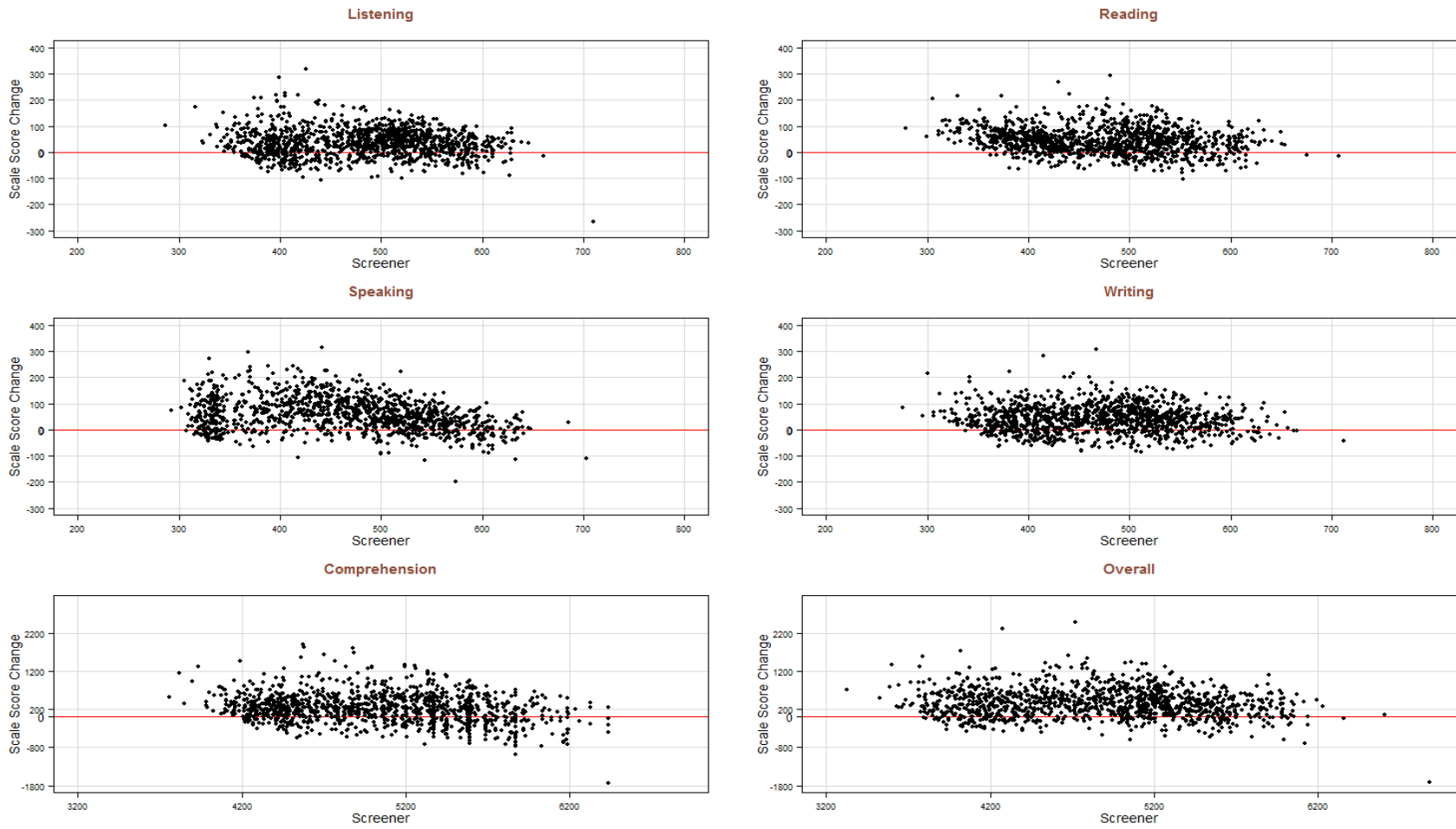


Figure S10.4: Scatter Plot of Scale Score Change from Screener to Summative for Grade Band 4-5

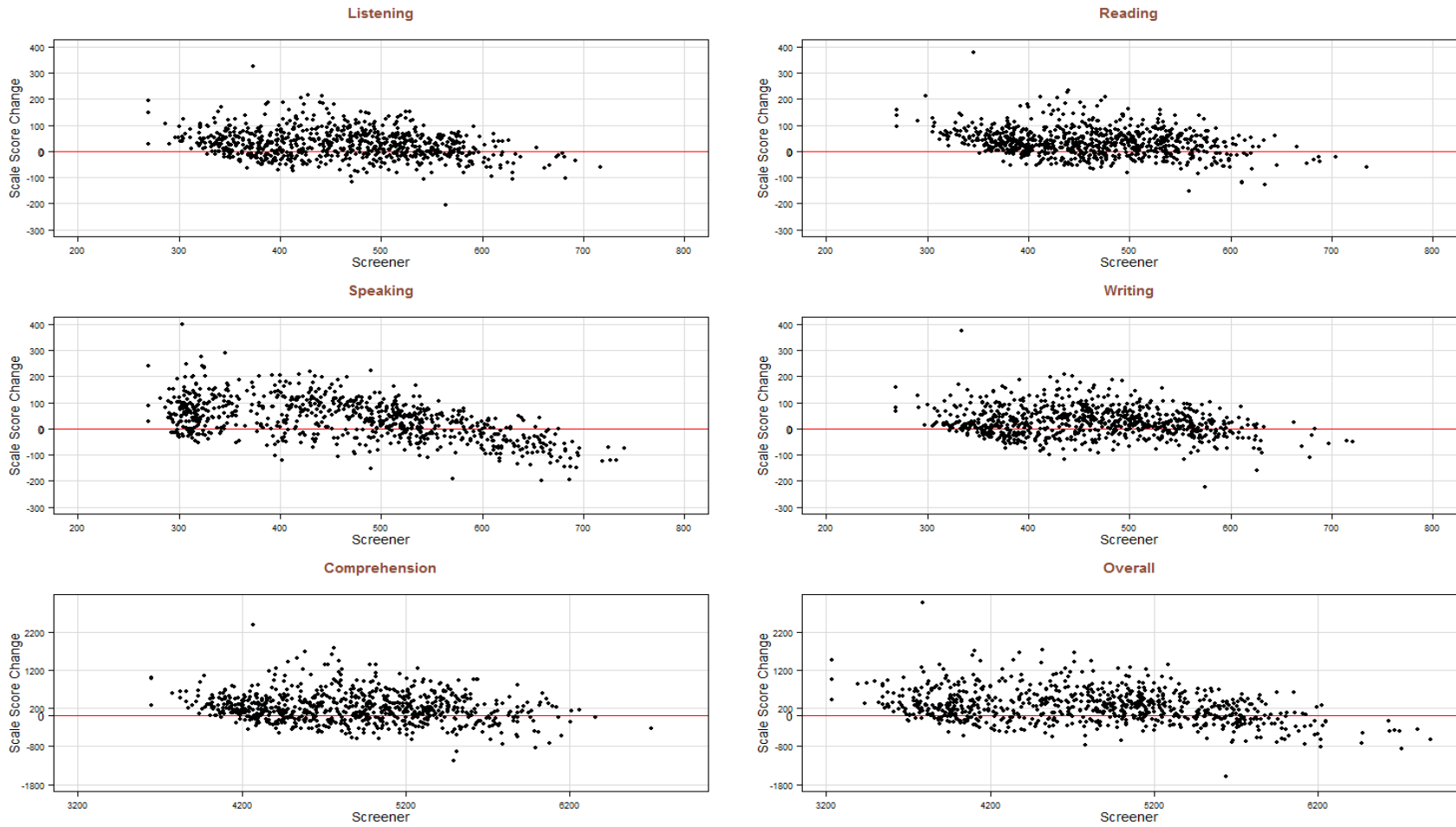


Figure S10.5: Scatter Plot of Scale Score Change from Screener to Summative for Grade Band 6-8

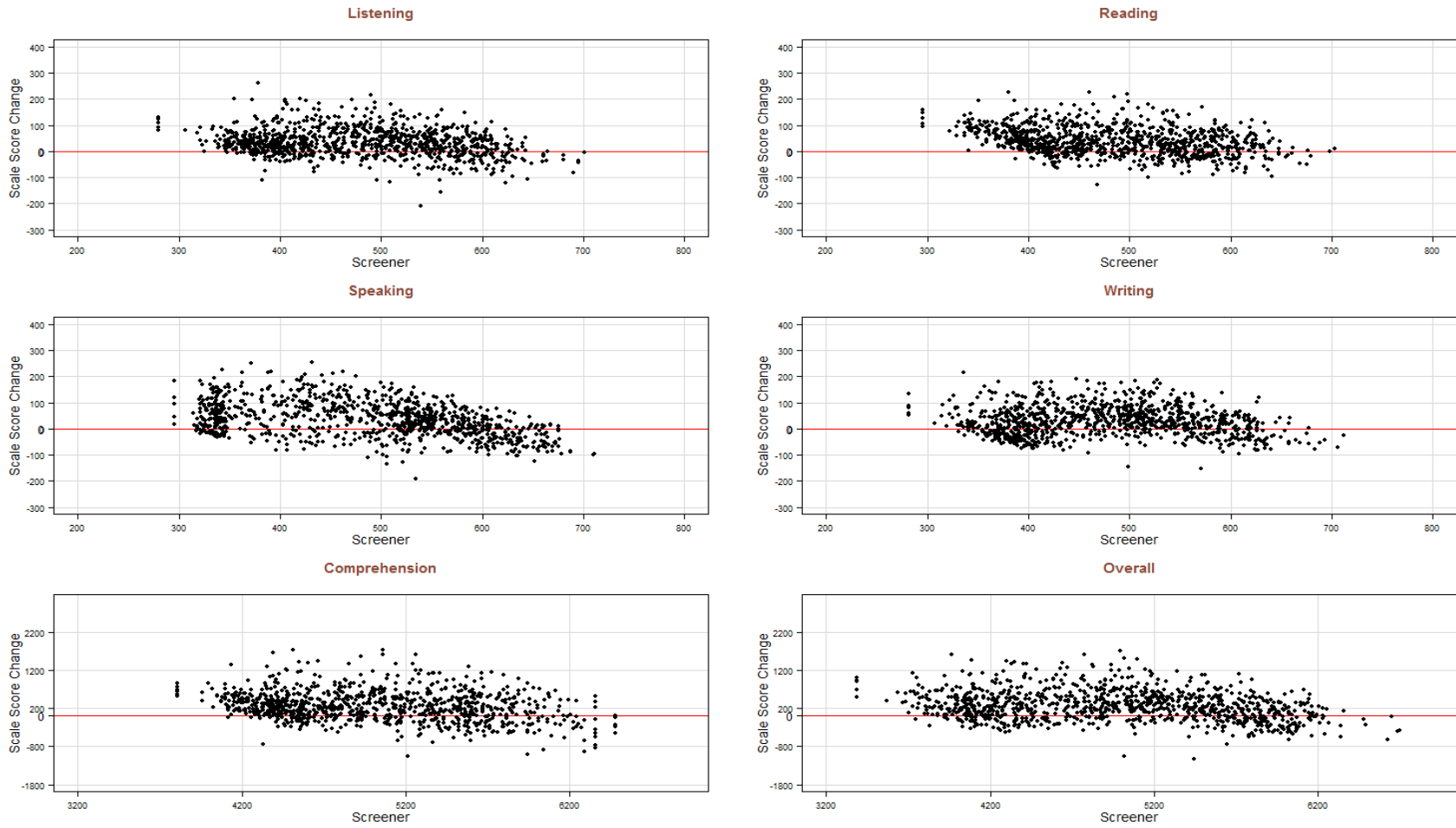
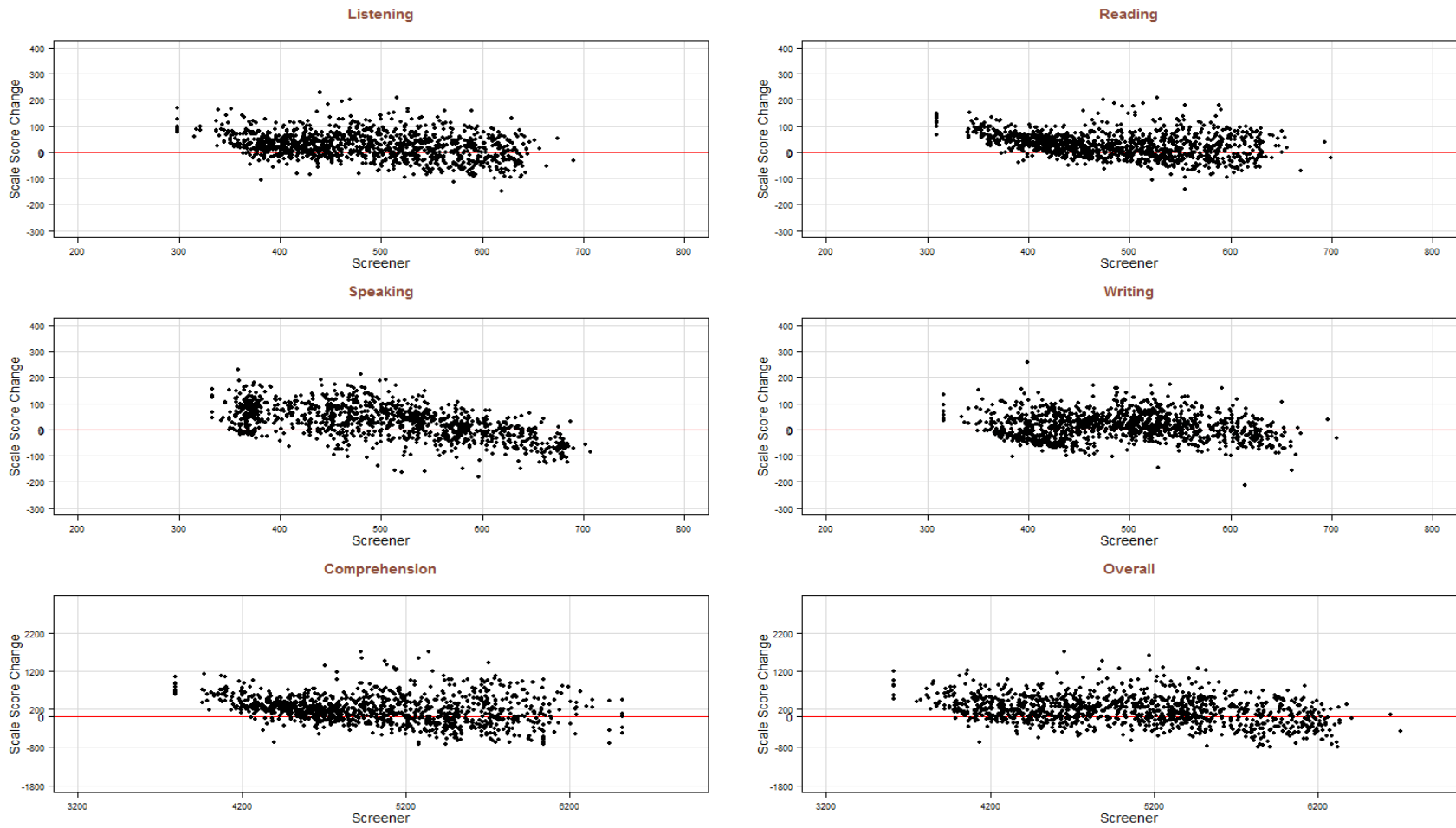


Figure S10.6: Scatter Plot of Scale Score Change from Screener to Summative for Grade Band 9-12



Section 11: Student Progress from Screener to Summative - Comparison of Scale Scores

Table S11.1: Comparison of Scale Scores Between Screener and Summative - Listening

Grade/Grade Band	Test	N	Mean	SD	Min	P10	P25	P50	P75	P90	Max
K	Screener	5027	524	58.0	319	440	490	535	563	583	715
	Summative	5027	567	77.0	310	467	512	567	622	669	753
1	Screener	941	510	75.8	288	397	460	523	566	603	662
	Summative	941	544	82.7	318	413	489	560	606	641	708
2-3	Screener	1190	482	71.8	286	386	416	490	535	576	710
	Summative	1190	518	85.7	313	393	448	530	583	624	742
4-5	Screener	785	460	84.6	270	349	389	463	523	571	717
	Summative	785	488	89.6	298	358	413	496	558	601	700
6-8	Screener	985	478	86.2	279	371	401	476	550	595	701
	Summative	985	511	90.4	271	391	428	519	582	631	731
9-12	Screener	1171	485	82.9	299	381	414	480	552	601	690
	Summative	1171	508	83.3	275	406	440	503	569	625	760

Table S11.2: Comparison of Scale Scores Between Screener and Summative - Reading

Grade/Grade Band	Test	N	Mean	SD	Min	P10	P25	P50	P75	P90	Max
K	Screener	5001	521	58.0	322	438	485	530	561	584	709
	Summative	5001	568	73.9	315	469	517	569	620	666	749
1	Screener	939	486	76.2	286	390	428	481	543	592	682
	Summative	939	543	82.8	350	437	479	536	608	665	751
2-3	Screener	1177	472	74.0	278	379	410	479	527	567	707
	Summative	1177	516	80.3	320	414	454	508	578	626	775
4-5	Screener	786	461	78.0	270	364	398	459	514	564	735
	Summative	786	495	81.2	328	394	427	486	554	611	723
6-8	Screener	982	483	83.0	296	384	413	477	552	597	703
	Summative	982	522	82.2	339	421	453	516	582	641	744
9-12	Screener	1163	488	78.1	310	387	426	484	548	597	699
	Summative	1163	514	77.1	352	428	454	497	561	631	770

Table S11.3: Comparison of Scale Scores Between Screener and Summative - Speaking

Grade/Grade Band	Test	N	Mean	SD	Min	P10	P25	P50	P75	P90	Max
K	Screener	4976	517	72.8	341	407	470	524	568	608	712
	Summative	4976	588	72.4	318	496	554	598	636	668	735
1	Screener	931	491	81.5	310	349	434	505	550	590	655
	Summative	931	562	90.0	303	420	525	580	622	656	721
2-3	Screener	1172	463	88.3	292	334	389	470	532	576	703
	Summative	1172	520	92.7	285	375	460	537	589	624	756
4-5	Screener	785	464	113.7	270	311	352	473	541	619	740
	Summative	785	500	98.5	267	350	424	523	577	613	712
6-8	Screener	970	488	102.6	296	337	403	508	569	620	711
	Summative	970	523	95.9	303	373	454	546	595	631	696
9-12	Screener	1128	503	93.4	333	369	430	506	575	633	707
	Summative	1128	530	79.1	343	418	473	542	590	624	719

Table S11.4: Comparison of Scale Scores Between Screener and Summative - Writing

Grade/Grade Band	Test	N	Mean	SD	Min	P10	P25	P50	P75	P90	Max
K	Screener	4991	486	49.2	366	432	454	478	506	548	686
	Summative	4991	552	77.4	307	455	492	545	614	659	727
1	Screener	936	484	76.9	283	386	425	480	544	592	665
	Summative	936	529	83.9	336	403	467	536	593	634	743
2-3	Screener	1176	468	74.6	276	372	407	469	524	566	712
	Summative	1176	508	87.8	305	385	437	510	577	619	773
4-5	Screener	780	459	85.0	268	351	386	458	521	571	721
	Summative	780	483	92.8	303	349	403	494	552	601	707
6-8	Screener	978	476	87.1	281	371	401	474	544	599	712
	Summative	978	504	99.6	311	354	419	521	579	628	747
9-12	Screener	1164	488	78.2	316	393	423	485	543	603	705
	Summative	1164	499	88.3	283	369	428	504	565	613	757

Table S11.5: Comparison of Scale Scores Between Screener and Summative - Comprehension

Grade/Grade Band	Test	N	Mean	SD	Min	P10	P25	P50	P75	P90	Max
K	Screener	5030	5409	460.9	3978	4720	5105	5491	5744	5964	6375
	Summative	5030	5610	535.3	3738	4886	5231	5618	5993	6328	6689
1	Screener	942	5183	530.3	3785	4461	4822	5182	5577	5888	6387
	Summative	942	5476	592.6	3879	4656	5041	5459	5890	6324	6672
2-3	Screener	1190	5050	558.5	3756	4316	4547	5070	5486	5806	6439
	Summative	1190	5269	597.4	3977	4473	4802	5250	5714	6063	6692
4-5	Screener	788	4870	547.5	3649	4177	4411	4875	5272	5585	6700
	Summative	788	5085	616.4	3878	4300	4564	5039	5534	5923	6681
6-8	Screener	989	5016	596.9	3803	4313	4494	4937	5516	5832	6476
	Summative	989	5268	628.3	3581	4497	4722	5241	5706	6170	6869
9-12	Screener	1175	5038	573.7	3797	4326	4573	4986	5490	5826	6523
	Summative	1175	5232	595.5	3732	4589	4769	5116	5597	6096	7132

Table S11.6: Comparison of Scale Scores Between Screener and Summative - Overall

Grade/Grade Band	Test	N	Mean	SD	Min	P10	P25	P50	P75	P90	Max
K	Screener	5034	5164	438.4	3702	4554	4908	5206	5449	5636	6769
	Summative	5034	5650	550.4	3560	4946	5277	5646	6041	6373	7056
1	Screener	943	5025	605.6	3364	4148	4600	5072	5482	5819	6456
	Summative	943	5453	663.8	3809	4484	5036	5508	5954	6286	6888
2-3	Screener	1190	4849	611.4	3326	4034	4312	4904	5295	5628	6880
	Summative	1190	5214	680.9	3607	4225	4701	5267	5738	6069	7208
4-5	Screener	792	4782	723.9	3237	3845	4121	4812	5318	5732	6889
	Summative	792	5031	712.2	3518	4016	4410	5102	5578	5917	6770
6-8	Screener	992	4948	711.9	3388	4042	4297	4949	5551	5911	6697
	Summative	992	5218	723.7	3582	4194	4590	5315	5764	6143	6824
9-12	Screener	1178	5019	661.7	3611	4163	4459	5014	5493	5968	6703
	Summative	1178	5193	643.1	3472	4334	4658	5182	5680	6072	6927